How a homolog of high-fidelity replicases conducts mutagenic DNA synthesis

Young-Sam Lee1,2, Yang Gao1 & Wei Yang1

All DNA replicases achieve high fidelity by a conserved mechanism, but each translesion polymerase carries out mutagenic DNA synthesis in its own way. Here we report crystal structures of human DNA polymerase ν (Pol ν), which is homologous to high-fidelity replicases yet is error prone. Instead of a simple open-to-closed movement of the O helix upon binding of a correct incoming nucleotide, Pol ν has a different open state and requires the finger domain to swing sideways and undergo both opening and closing motions to accommodate the nascent base pair. A single–amino acid substitution in the O helix of the finger domain improves the fidelity of Pol ν nearly ten-fold. A unique cavity and the flexibility of the thumb domain allow Pol ν to generate and accommodate a looped-out primer strand. Primer loop-out may be a mechanism for DNA trinucleotide-repeat expansion.

RESULTS
A minimal active human Pol ν and four crystal structures

Human Pol ν is largely insoluble when expressed in E. coli26,27. To improve its solubility, we expressed human Pol ν tagged with an N-terminal tandem repeat of maltose-binding protein (MBP) in HEK293 cells (Online Methods). We also serially deleted the N- and C-terminal unstructured regions to create Pol v77 (E175–G863, 77 kDa)27, Pol v76, (A181–G863), Pol v75 (A192–G863) and Pol v74 (D199–G863) (Supplementary Fig. 1). All four Pol ν variants were soluble and readily purified. Except for Pol v74, the N-terminal MBP tag was easily removable by PreScission cleavage. Tag-free Pol v75, v76 and v77 were equally active in DNA binding and synthesis. We purified Pol v75 to homogeneity (Supplementary Fig. 2a) and used it for most of the following structural and biochemical studies.

DNA synthesis assays with Pol ν expressed in mammalian cells confirmed that Pol ν is an error-prone DNA polymerase that preferentially misincorporates dT regardless of template sequence (Supplementary Fig. 2b). With previously reported mutation-hotspot and mutation-coldspot DNA substrates26, the measured kcat and Km of Pol v75 for incorporating correct (dC) and incorrect (dT) nucleotides (Supplementary Table 1) agreed well with the kinetic parameters obtained from E. coli–expressed and refolded Pol v77 (ref. 26). In contrast to the majority of DNA polymerases, which carry out DNA synthesis with a similar efficiency and accuracy regardless of DNA sequence, Pol ν had up to a 20-fold difference in catalytic efficiency and fidelity depending on DNA sequence (Supplementary Table 1).

1Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, USA.
2Present address: Well-Aging Research Center, Samsung Advanced Institute of Technology, Suwon, Korea. Correspondence should be addressed to W.Y. (wei.yang@nih.gov).

Received 13 December 2014; accepted 10 February 2015; published online 16 March 2015; doi:10.1038/nsmb.2985
To understand Pol ν’s unusual error-prone DNA synthesis, we determined the crystal structure of a Pol ν-75–DNA complex, Ndna1, at 3.1-Å resolution (Online Methods), with a 14-nt template, an 11-nt primer DNA and a nonreactive nucleotide analog, dAMPNPP (Supplementary Fig. 3a, b). Owing to crystal-lattice interactions, the blunt end of each DNA was in the active site instead of the 3-nt 5′ overhang, and the incoming nucleotide was absent. After modifying the DNA substrate and changing the duplex length (Supplementary Fig. 3c, d), we obtained three additional crystals: Ndna2, Ndna3 and Ndna4, with Ndna2 and Ndna3 in the same space group as Ndna1. We solved these three structures by molecular replacement and refined them at the highest resolution of 2.95 Å (Ndna2) and lowest resolution of 3.3 Å (Ndna3) (Table 1). Perhaps owing to the high salt and low pH of the crystallization buffer (1.2 M (NH₄)₂SO₄, pH 5.9–6.3), all four structures were protein–DNA binary complexes without dAMPNPP. Interestingly, in the first three structures (Ndna1–3), the finger domain changed from the open, ajar conformation to a closed conformation correlating with the increases of the DNA length (Supplementary Fig. 3). In Ndna4, the finger domain was closed, and the thumb domain underwent a large rotation.

An unusual open state of Pol ν

The Ndna2 and Ndna4 structures resemble the finger-closed form of A-family polymerases (Figs. 1 and 2a), which usually represents the blocked active site

![Image](https://example.com/figure1.png)

**Figure 1** Structure of Pol ν. (a) The primary structure. The N termini of four Pol ν variants are indicated by arrows. Structural domains and three insertions are marked with boundary residues. (b) Ndna2 structure (left) and its superposition with *Bacillus* Pol I [33] (gray, right). DNA is colored yellow (primer) and orange (template). The catalytic residues and incoming nucleotide in Pol I are shown in ball-and-stick model. (c) A zoomed-in view of the superimposed Exo domain. The Pol ν Exo site is mutated (equivalent residues in *Bacillus* Pol I are shown in sticks) and blocked by a long loop. (d) The thumb domain. (e) The palm domain. Ins1–3 are marked by arrows in d and e.

**Table 1** Data collection and refinement statistics

| Data collection | Ndna1 | Ndna2 | Ndna3 | Ndna4 | Ndna1 Br (SAD) |
|-----------------|-------|-------|-------|-------|----------------|
| **Space group** | H32   | H32   | H32   | I222  | H32            |
| **Cell dimensions** |       |       |       |       |                |
| a, b, c (Å) | 292.7, 292.7, 110.7 | 290.9, 290.9, 110.2 | 292.1, 292.1, 108.1 | 98.0, 90.0, 120.0 | 294.1, 294.1, 110.8 |
| α, β, γ (°) | 90.0, 90.0, 120.0 | 90.0, 90.0, 120.0 | 90.0, 90.0, 120.0 | 90.0, 90.0, 120.0 | 90.0, 90.0, 120.0 |
| Wavelength (Å) | 0.9099 | 1.0000 | 1.0000 | 1.0000 | 0.9099 (Br peak) |
| Resolution (Å)a | 30.0–3.10 (3.15–3.10) | 30.0–2.95 (3.00–2.95) | 30.0–3.30 (3.36–3.30) | 30.0–3.20 (3.26–3.20) | 30.0–3.20 (3.26–3.20) |
| Rmerge (%)a | 11.9 (61.2) | 7.9 (63.7) | 13.2 (69.6) | 9.8 (61.9) | 9.3 (70.4) |
| I / σIa | 11.9 (2.1) | 16.7 (2.2) | 13.3 (2.5) | 13.0 (2.0) | 16.3 (2.8) |
| Completeness (%)a | 99.5 (98.2) | 99.0 (100.0) | 99.4 (98.6) | 98.9 (98.6) | 100.0 (100.0) |
| Redundancya | 4.4 (3.6) | 4.7 (4.3) | 5.7 (5.3) | 5.2 (5.1) | 7.1 (7.1) |
| **Refinement** |       |       |       |       |                |
| Resolution (Å) | 30.0–3.10 | 30.0–2.95 | 30.0–3.30 | 30.0–3.20 | 30.0–3.20 |
| No. reflections | 32,894 | 37,333 | 26,294 | 25,063 |                |
| Rwork / Rfree (%) | 21.3 / 23.7 | 22.2 / 24.1 | 20.3 / 22.3 | 24.2 / 28.1 |                |
| No. atomsb | Protein | 5,112 | 4,971 | 5,025 | 4,952 |
| DNA | 505 | 550 | 489 | 449 |
| Ligand | – | 12 (MES) | 12 (MES) | – |                |
| B factors | Protein | 87.2 | 74.5 | 60.7 | 79.8 |
| DNA | 85.9 | 82.7 | 60.2 | 88.9 |
| Ligand | – | 69.5 | 58.7 | – |                |
| r.m.s. deviations | Bond lengths (Å) | 0.008 | 0.007 | 0.008 | 0.013 |
| Bond angles (°) | 1.112 | 1.061 | 1.076 | 1.683 |                |

aValues in parentheses are for highest-resolution shell. A single crystal was used for each data collection and structure refinement. bThere is no water included in the structure refinement. SAD, single-wavelength anomalous diffraction.
the reaction-ready state of a polymerase–DNA–dNTP ternary complex16–18,28,29 and occasionally a post-reaction-product state when DNA is not yet translocated for the next round of dNTP incorporation30,31. The catalytic core of Pol V75 superimposed well with those of *Bacillus* and *Taq* DNA polymerase I (abbreviated as Pol I) in the ternary complexes22,33, with 1.34-Å and 1.48-Å r.m.s. deviation over 380 pairs of Cα atoms, respectively. Residues 192–416 of Pol V formed a degenerate 3′–5′-exonuclease domain, which deviates from the equivalent domains in *E. coli* and *Bacillus* Pol I (2.6 Å over 145 pairs of Cα atoms) (Fig. 1b,c). The four catalytic essential carboxylates are replaced by T224, M226, F324 and L401 in the pseudo-Exo domain of Pol V, and the mutated active site is further blocked by an elongated loop (residues 229–241) (Fig. 1c). The first two insertions (Ins1 and Ins2) in Pol V and Pol θ are located at the tip and the base of the thumb domain (Fig. 1a,d) and probably influence DNA binding (details below). The third insertion (Ins3) in Pol V is on the backside of the palm domain distal from the DNA-binding surface (Fig. 1e). Although Ins3 in Pol θ is further toward the C terminus24,25, modeling shows that it also is on the backside of the palm domain and extends toward the pseudo-Exo domain (Fig. 1e).

In the absence of an incoming nucleotide, DNA Pol I of *E. coli*, *Bacillus* and *Taq* assume an open state with the O helix in the finger domain rotating ~40° away from the catalytic center16,34 (Supplementary Video 1). The Ndna1 structure of Pol V revealed a semiopen finger domain that resembles the ajar state of *Bacillus* DNA Pol I (PDB 3HP6 (ref. 35)), in which the O helix is halfway between the open and closed state, owing to a mismatched incoming nucleotide. The catalytic cores of these two structures superimposed reasonably well except for the absence of dNTP in Ndna1 (Fig. 2b).

The O helix of Ndna3 was the most open among the four Pol V75 structures, but Ndna3 was different from the open state of Pol I (Fig. 2c,d). The O helix in Ndna3 was rotated 25°, instead of 40°, away from the DNA substrate relative to that in Ndna2. Among the previously characterized A-family polymerases, helices Oa and Ob (sequentially after the O helix in the finger domain) move slightly outward in the opposite direction of the O helix from the closed state to accommodate the downstream (+1) template nucleotide (Fig. 2c). However, in Ndna3, Oa and Ob rotate 17.5° toward the DNA duplex in the same direction as the O helix and occlude both the +1 and templating nucleotide–binding site (Fig. 2c). Instead of the O helix opening, a large portion of the finger domain (helices Oa, Oa and Ob) rotate toward the DNA–dNTP− template.
Ob and O) swings over the DNA duplex (Supplementary Video 2). In this unusual open state of Pol ν, binding of a templating base is blocked. Although binding of an incoming nucleotide is possible, without a template base dNTP binding has to be unstable and non-specific. Pol ν needs to open the templating base–binding site for a nascent base pair to form and close over the incoming dNTP for DNA synthesis to take place. The unusual open state of Pol ν may explain its weak binding of dNTP, reduced catalytic efficiency and low fidelity.

Unique residues in Pol ν and the error signature
Several residues uniquely conserved in Pol ν appear to stabilize its unusual open state. First, Y682 on the O helix of Pol ν replaces a phenylalanine residue in high-fidelity Pol I (Fig. 3a) and favors the unusual open state by forming a hydrogen bond with R692 (on helix Oa). The Y682 and R692 pairs close off the binding site for the +1 template nucleotide, which is normally sandwiched between helices O and Oa in Pol I (Fig. 2c), and displaces the tyrosine (Y686 in Pol ν and Y714 in Pol I) that is a placeholder for the template base (Fig. 2c,d). Interestingly, Y682F-mutant Pol ν has greatly reduced catalytic activity but improved accuracy2. Second, G689 of Pol ν replaces a conserved serine in Pol I and allows a sharp turn between helices O and Oa, thus keeping them together and preventing Pol ν from reaching the normal open state. Nearby E691 of Pol ν replaces the aromatic residue in Pol I (Y719), which stacks with the downstream template base (+3) and stabilizes the template strand (Fig. 2c). The negatively charged E691 of Pol ν cannot form favorable interactions with DNA. As a result, the downstream template is disordered in all Ndna structures.

E675 and K679 are unique to Pol ν and are located near the nascent base pair (Fig. 3a,b), and thus they may favor dT misincorporation5. We generated Pol ν mutants with these residues replaced by the corresponding amino acids in Pol I (Supplementary Fig. 2c,d). The E675R mutation reduced polymerase activity but did not change nucleotide preference (Fig. 3c and Supplementary Table 1). In contrast, K679A-mutant Pol ν75 was as active as the WT protein and exhibited better accuracy (by eight- to ten-fold) in nucleotide selection than wild type (WT), regardless of whether the DNA substrate sequence was a hotspot or coldspot for mutation (Fig. 3c,d). The positively charged K679 potentially forms hydrogen bonds with O6 of dG (template) and O4 of the incoming dTTP, thus favoring the dT misincorporation opposite dG (Fig. 3c). This role would be similar to those of the positively charged R61 of human Pol η, as well as the R61K mutation, both of which promote G-T mismatches and A-to-G hypermutation in somatic cells16.

The flexible thumb and primer strand loop out
Ins2 (residues 592–606) in Pol ν forms an extended β-hairpin replacing a short α-helix in replicative A-family polymerases (Fig. 4a,b). Interestingly, the addition of Ins2 creates a cavity in the DNA-binding surface of Pol ν. As a result, the primer strand upstream from its 3’ end (~2 to ~4 nt) is exposed to solvent (Fig. 4a). Furthermore, the thumb domain undergoes a 29.3° rigid-body rotation toward the upstream DNA in the Ndna4 structure (Fig. 4c and Supplementary Video 3). In other A-family polymerases, the α helix that is replaced by the β hairpin of Ins2 in Pol ν would prevent the thumb movement. While maintaining contacts with the template strand, the rotated thumb in Ndna4 leaves the primer strand solvent exposed from ~2 nt upwards.

The cavity in the Pol ν thumb domain and exposed primer strand are reminiscent of the cavity on the template side of the DNA-binding surface in E. coli DNA Pol II that results in template-strand loop-out and deletional frameshifts18. We suspect that Pol ν can accommodate looped-out primers in the cavity and lead to either insertional frameshifts or mutations if realignment ensues. To determine the size and location of primer loop-outs, we constructed seven DNA substrates with 1 to 4 nt looped out in the primer strand at the ~2, ~3 or ~4 position (Supplementary Fig. 4). Indeed Pol ν75 could extend primers with a 1-nt loop-out 3-bp upstream from the 3’ end (P[3,1])
Deletion of Ins1 leads to reduced processivity of DNA synthesis only, but deletion of Ins2 or Ins3 appears to destabilize Pol θ and eliminate the translesion-synthesis activities of Pol θ. To examine how Ins3 influences Pol v’s tolerance of primer loop-out, we replaced 25 residues of Pol v (residues 583–607), including Ins2, with the corresponding ten residues of E. coli Pol I (residues 680–689) (Online Methods and Supplementary Fig. 2c,d). The resulting ΔIns2 Pol v75 was stable and retained one-third of WT Pol v75’s catalytic efficiency on normal DNA. However, its catalytic efficiency with P[3,1] and P[3,2] loop–out substrates was reduced to less than one-fifth that of WT (Fig. 4f and Supplementary Tables 1 and 2). It is not surprising that the ΔIns2 protein is not a perfect mimic of WT Pol v or Pol I, but the differential reduction of its catalytic activity on normal versus loop–out DNA is consistent with a role of the structural cavity and the primer–loop–out model.

**DISCUSSION**

Until now, a simple open–and–closed rotation of the finger domain during each cycle of nucleotide incorporation has been universally observed in A–, B–, C– and X–family DNA polymerases16,18,28,29, reverse transcriptases and RNA polymerases37,38 (Fig. 5a). The Pol v75 structures, however, reveal an unprecedented open state, which is open for an incoming nucleotide but closed for template-base binding. The unusual open state would change the kinetic and dynamic processes of nascent-base-pair formation, incoming-nucleotide selection and incorporation. A handful of residues unique to Pol v (Fig. 3a) are likely to stabilize the unusual open state, and conservation of Y682 and G689 in Pol v and Pol θ (Fig. 3a) indicates that the unusual open state is probably also present in Pol θ.

The error signature of Pol v in misincorporating dTTP is also due to the unique K679 on the O helix (Fig. 3) Our analysis of K679A mutant Pol v on mutation hotspots and coldspots confirmed the earlier finding that replacement of one amino acid can substantially increase discrimination against dT misincorporation (Supplementary Table 1). Different amino-acid replacements, E675Q, K679Q and E691R, are found in Pol θ, and these may result in mutation spectra and translesion properties different from those of Pol v. The key feature of Y-family polymerases, which specialize in translesion synthesis, is the small and nonconserved fingerprint domain that accommodates different DNA lesions in a preformed ‘closed’ active site39. The finger–closed conformation in the post–reaction–product state was previously observed only among low–fidelity X– and Y–family polymerases30,31. Whether the closed conformation revealed in Ndna2 and Ndna4 bears functional significance for Pol v in error–prone and translesion DNA synthesis awaits future studies. The dramatically different finger–open state of Pol v and Pol θ with substitutions of amino acids that contact the nascent base pair (Fig. 5a) may provide the answer as to how these homologs of high–fidelity replicases carry out mutagenic and translesion DNA synthesis.

 Primer loop–out allows the template and primer to misalign and DNA synthesis to continue when the downstream template is not usable or blocked (Fig. 5b). Misalignment of primer and template, however, would depend on the local DNA sequence and require adjustment of the DNA position. The near–30° rotation of the thumb domain in Ndna4 (Fig. 4c) has not been observed among DNA polymerases, but it is reminiscent of the thumb movement observed with T7 RNA polymerase during transcription initiation to elongation40. The mobility and the cavity in the thumb domain, both of which appear to correlate with Ins2, may allow Pol v to sample DNA for alternative template–primer alignments and accommodate nucleotide loop–out in the primer strand (Fig. 5b). Removal of Ins2 does not alter the fidelity of Pol v but decreases its tolerance for primer loop–out (Supplementary Tables 1 and 2). Pol θ, which contains a

with 42% of the catalytic efficiency of normal DNA substrate, whereas the Klenow fragment of E. coli Pol I (Klenow) could not (Fig. 4d,e and Supplementary Table 2). The loop–out location at the −3 position matches perfectly with the observed cavity in Pol v. The nucleotide insertion efficiency of Pol v75 inversely correlated with increase of the loop size from 1 to 3 nt at the −3 position (Supplementary Fig. 4). When the 1–nt loop–out was 4 to 5 bp upstream (P[4 or 5,1]), Pol v75 became quite efficient, and even Klenow showed weak activity. Because Ins2 is conserved between Pol θ and Pol v, we tested whether Pol θ could extend a loop–out–containing primer. We generated and purified an active polymerase domain of Pol θ, Pol 086 (Online Methods and Supplementary Fig. 5). Pol 086 was ~4 times more active than Pol v75 and efficiently extended P[3,1] to full–length (25 nt) and also to 1–nt–longer products (26 nt) (Fig. 4e). Interestingly, with normal DNA substrate Pol 086 produced both the full–length (26 nt) and 1–nt–shorter (25 nt) products (Fig. 4e), results agreeing with the report that Pol 086 is prone to generating both types of frameshift mutations3.

The functional importance of the three inserted regions in Pol θ has been examined previously25. Deletion of Ins1 leads to reduced processivity of DNA synthesis but deletion of Ins2 or Ins3 appears to destabilize Pol θ and eliminate the translesion-synthesis activities of Pol θ.

**Figure 5** Diagram of DNA synthesis by Pol v and a primer–loop–out model for TNR expansion. (a) The high–fidelity Pol I and TLS Pol v differ in the open states while being identical in the closed states. When the O helix in Pol v is open for dNTP binding, helices Oa and Ob exclude the template base from the active site. The unique K679 in Pol v further promotes dTTP misincorporation. The 3′–5′ exonuclease (Exo, pink stars), which proofreads and improves the accuracy of Pol I, is inactivated in Pol v (pseudo–Exo, black stars). (b) A primer–loop–out model. When a downstream template base is unusable (indicated by a red dot), Pol v can loop out 1 or 2 nt of the primer strand at the −3 position to reuse the normal template base(s) for lesion–by–DNA synthesis. The mobile thumb of Pol v (shown as semitransparent green) may facilitate DNA translocation and misalignment. (c) Repetitive DNA sequence such as TNRs (CNG), would ease loop–out of repeat units, as shown here, and result in repeat expansion.

© 2015 Nature America, Inc. All rights reserved.
larger Ins2 (ref. 25) and is potentially more flexible, is more efficient than Pol ν at synthesizing DNA with a primer loop-out (Fig. 4e).

Short DNA sequence repeats make it easy for primer and template to misalign by one repeat unit (Fig. 5c). If not realigned, looping out of the primer strand would result in insertions during replication. Trinucleotide repeat (TNR) expansion, a common cause of a number of human diseases including Huntington’s disease, fragile X syndrome, Friedreich’s ataxia and myotonic dystrophy, must involve DNA synthesis for these repeats to expand41,42. Various models have been hypothesized to account for the expansions. Prior to the discovery of Pol ν and Pol θ, no human DNA polymerase was known to loop out primer and generate expanded DNA products, however. The expansion has thus been attributed to formation of DNA secondary structure (hairpins or triplex) in the expanded strand outside of the DNA polymerase context33,44. Because of its tendency to form secondary structures, TNRs are inherently unstable and often require specialized TLS polymerases to replicate in normal cells. The ability of Pol ν and Pol θ to loop out primer strands raises the possibility that specialized TLS polymerases may also lead to repeat expansions when unrestrained or misguided.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 4XVI (Ndna1), 4XVK (Ndna2), 4XVL (Ndna3) and 4XVM (Ndna4).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank D. Leahy and R. Craigie for editing the manuscript. The research was supported by the intramural research program of the US National Institutes of Health (DK036146-08, W.Y.).

AUTHOR CONTRIBUTIONS

Y.-S.L. carried out most of the experiments. Y.G. helped with kinetic measurement of K678A-mutant Pol ν and with data deposition. Y.-S.L and W.Y. designed the project and prepared the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
ONLINE METHODS

Expression of human Pol ν in HEK293 cells. Human POLN cDNA was obtained from HEK293 cells by RT-PCR. On the basis of sequence alignment with bacterial homologs, and to remove nonconserved and mammalian recombination, four N-terminal– and C-terminal–deletion clones POLN77, POLN76, POLN75 and POLN74, encoding residues from E175, A181, A192 or D199 to G863, respectively, were constructed (Supplementary Fig. 1a). POLN77 was previously reported27. Amplified POLN variants were cloned into the mammalian expression vector pLEXm45. To improve protein solubility and ease purification, the His8 tag and two tandem copies of maltose-binding protein (2MBP)46 followed by a PreScission protease site were fused to the N termini of Pol ν variants. The His8-2MBP–tagged Pol ν variants were overexpressed in HEK293GNT1 cells according to a protocol described previously21.

Generation of Pol ν mutants. Mutations of E675R, K679A, and ΔIns2, in which 25 residues in Pol ν (residues 583–607), including Ins2, were replaced by the corresponding ten residues in E. coli Pol I (residues 680–689, PVRNIEGRRRI) by PCR-based mutagenesis with the following primers (with altered site underlined): 5′-GACACACGCACGACAGAGGCAAACCAAGAAGGTG and its reverse complement for E675A; 5′-GAGAGCAAACCAAGGCGGTGGTGTACGCCG and its reverse complement for K679A; and 5′-CCAGTGAGAAACGAGGAAG and its reverse complement for ΔIns2.

Protein purification. WT and mutant His8-2MBP–tagged Pol v75 proteins were first purified with amylose resin (NEB). After removal of the His8-2MBP tag by PreScission protease, Pol ν75 was purified over a Hitrap heparin HP column and Superdex 200 (GE Healthcare). Purified proteins were concentrated to ~2 mg/ml and stored at −25 °C in 20 mM Tris, pH 8.0, 0.3 M NaCl, 2 mM TCEP and 30% glycerol. All purification steps were performed at 4 °C.

Crystallization and structure determination. Pol ν75–DNA–dAMPNPP complexes with the Pol ν75 concentration at 5–6 µg/ml were crystallized by the hanging-drop vapor–diffusion method at 20 °C against a reservoir containing ~1.2–1.5 M (NH4)2SO4 and 100 mM MES, pH 5.9–6.3. The crystals were cryoprotected by the addition of 1.5 M sodium malonate, pH 6.0, to the mother liquor. For phase determination, four thymines in the DNA substrate were substituted by 5-bromodeoxyuridines (Supplementary Fig. 3b). Diffraction data of native crystals and the single-wavelength anomalous diffraction (SAD) at the Br absorption edge were collected at the SER-CAT beam lines at the Advanced Photon Source and processed with HKL2000 (ref. 47) (Table 1). Phases were determined by combination of SAD data and molecular replacement with Bacillus DNA Pol I (PDB 3PV8)33 as a search model. COOT48 and PHENIX49 were used for model building, phase determination and structure refinement. All protein residues are in the favored and allowed regions of the Ramachandran plot, and none are in disallowed regions. All structural figures and videos were made with PyMOL (http://www.pymol.org).

DNA synthesis assay. DNA polymerase activities (Figs. 3c,d and 4e,f) were measured as previously described21. Briefly, the DNA primer was 6-FAM–labeled at the 5′ end. Reactions were carried out in 10 µl of reaction buffer (20 mM Tris, pH 8.3, 130 mM NaCl, 0.1 mg/ml BSA, 5 mM DTT, 3% glycerol and 5 mM MgCl2). The final concentrations were: DNA substrates, 100 nM; Pol ν75, 10 nM; and all four dNTPs or each dNTP, 100 µM. After preincubation of the protein–DNA mixture at 37 °C for 5 min, reactions were initiated by addition of desired nucleotides and MgCl2, and were allowed to proceed at 37 °C for 10 min. Reactions shown in supplementary figures were carried out under similar conditions; differences (if any) are specified in figure legends. Primer-extension products were resolved on 15% polyacrylamide–7.5 M urea gels in 1× TBE buffer, visualized by Typhon Trio (GE Healthcare), and quantified with ImageQuant TL (GE Healthcare).

Steady-state kinetic assay of single-nucleotide incorporation. $k_{cat}$ and $K_m$ of single-nucleotide incorporation were measured in 10 µl of the reaction buffer including 5 µM DNA substrates (1% primer was 6-FAM–labeled), 10–200 nM Pol ν75, and a correct (0.1 to 250 µM) or an incorrect dNTP (1 to 1,000 µM) for the template base. Reactions were initiated and terminated as described above except for the 5-min duration of reaction instead of 10 min. Quantification and curve fitting to the Michaelis–Menten equation for calculation of $k_{cat}$ and $K_m$ were carried out as previously described21.

Preparation of active human Pol θ polymerase domain. Human POLQ cDNA was obtained from HEK293 cells by RT-PCR. According to a previous publication25 and predicted secondary structures, two N-terminal–deletion clones, POLQ90 (G1792–V2590, the same as described in ref. 25) and a smaller POLQ86 (S1822–V2590), were constructed and inserted into the pLEXm-His8-2MBP expression vector as for POLN75. Pol θ90 and Pol θ86 were overexpressed in HEK293 cells and purified as described for Pol ν75.

45. Aricescu, A.R., Lu, W. & Jones, E.Y. A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr. D Biol. Crystallogr. 62, 1243–1250 (2006).
46. Jensen, R.B., Carreira, A. & Kowalczykowski, S.C. Purified human BRCA2 stimulates RAD51-mediated recombination. Nature 467, 678–683 (2010).
47. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).
48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132 (2004).
49. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221 (2010).