BIOCHEMISTRY

Structure and mechanism of human PrimPol, a DNA polymerase with primase activity

Olga Rechkoblit,[1]* Yogesh K. Gupta,[1]* Radhika Malik,[1]* Kanagalaghatta R. Rajashankar,[2,3] Robert E. Johnson,[4] Louise Prakash,[4] Satya Prakash,[4] Aneel K. Aggarwal[1]†

PrimPol is a novel human enzyme that contains both DNA primase and DNA polymerase activities. We present the first structure of human PrimPol in ternary complex with a DNA template-primer and an incoming deoxynucleoside triphosphate (dNTP). The ability of PrimPol to function as a DNA primase stems from a simple but remarkable feature—almost complete lack of contacts to the DNA primer strand. This, in turn, allows two dNTPs to bind initiation and elongation sites on the enzyme for the formation of the first dinucleotide. PrimPol shows the ability to synthesize DNA opposite ultraviolet (UV) lesions; however, unexpectedly, the active-site cleft of the enzyme is constrained, which precludes the bypass of UV-induced DNA lesions by conventional translesion synthesis. Together, the structure addresses long-standing questions about how DNA primases actually initiate synthesis and how primase and polymerase activities combine in a single enzyme to carry out DNA synthesis.

INTRODUCTION

DNA replication is one of the most demanding of all biological processes, requiring multiple DNA polymerases. In addition to replicative DNA polymerases for duplication of the leading and lagging DNA strands (*1*–*3*), another group of specialized enzymes, the translesion synthesis (TLS) polymerases (*4*), is required to bypass specific DNA lesions or other barriers to replication. However, because DNA polymerases are incapable of de novo DNA synthesis, primases are required to initiate DNA replication and maintain continuity of the replication fork (*5*, *6*). In humans, the PriS primase synthesizes a short RNA primer on the leading and lagging DNA strands, which is then extended with deoxynucleoside triphosphates (dNTPs) by the replicative polymerases (*7*). Until very recently, DNA polymerase and primase activities in human cells were considered the province of separate enzymes. This changed markedly with the discovery of PrimPol in human cells (*8*–*10*), which has both DNA polymerase and DNA/RNA primase activities within the same enzyme.

PrimPol is only the second primase to be discovered in human cells and is present in both nuclear and mitochondrial compartments (*8*). Like PriS, it belongs to the so-called archaeo-eukaryotic primase family of enzymes (*11*). However, unlike PriS, PrimPol has dual TLS polymerase and primase activities that lend it the flexibility to bypass DNA lesions via TLS as well as the capacity to "skip" DNA lesions and initiate DNA synthesis de novo downstream of the lesion (*8*–*10*, *12*, *13*). Intriguingly, PrimPol can reprime DNA synthesis using (and preferentially) dNTPs (*8*, *9*), as opposed to NTPs used by most other primases (*5*, *6*), including PriS. PrimPol has emerged as a new type of DNA damage tolerance enzyme, but a number of questions remain unanswered. First, how do the polymerase and primase activities coexist within the same enzyme? Currently, there is no structure of human PrimPol alone or in complex with a template-primer [only the structure of a distantly related archaeal enzyme in the apo form (*14*)]. Second, what is the underlying mechanism of primer synthesis and TLS bypass? Despite extensive effort over decades, there is still no structure of a primase catalytic domain in complex with a template-primer. Intriguingly, almost all current models of primase action postulate the existence of "initiation" and "elongation" dNTP/NTP binding sites on the catalytic core for the formation of the initial dinucleotide of the primer (*5*, *6*, *15*, *16*), but structural information has been lacking. Third, why does the enzyme prefer dNTPs over NTPs?

To address these prevailing questions, we present here the structure of the catalytic core of human PrimPol in ternary complex with a template-primer and an incoming nucleotide [deoxyadenosine triphosphate (dATP)]. The structure captures the PrimPol catalytic core in the act of inserting a nucleotide opposite a DNA template and provides the basis of TLS activity opposite ultraviolet (UV)–induced DNA lesions. Strikingly, the structure reveals an almost complete lack of contacts to the DNA primer strand, thus eliminating the need for a preexisting primer. We define the long-sought initiation site of nucleotide binding in a primase catalytic domain, which is essential for the formation of the initial dinucleotide. Together, the structure addresses long-standing questions related to the mechanism of DNA primases and how primase and polymerase activities are coordinated in a single DNA damage tolerance enzyme.

RESULTS

Structure determination

The catalytic core of human PrimPol (residues 1 to 354) was cocrystallized with a 13-nucleotide (nt) primer (5′-GGGTGTGGTAGCG^{dd}-3′)/17-nt template (5′-CATCGCTACCACACCCC-3′) and dATP from a mix containing Ca^{2+}. The cocrystals diffract to 2.2 Å resolution with synchrotron radiation (Advanced Photon Source) and belong to space group *P*1 with unit cell dimensions of $a = 50.72$ Å, $b = 65.21$ Å, $c = 72.46$ Å, $\alpha = 67.88°$, $\beta = 85.21°$, and $\gamma = 86.64°$ (table S1). There are two PrimPol-DNA-dATP ternary complexes in the crystallographic asymmetric unit. The structure was solved by the single-wavelength anomalous diffraction (SAD) method, using x-ray data measured at the Se-*K* absorption edge from two selenomethionine (SeMet) derivative crystal forms belonging to space groups *C*2 and $P2_1$. The two ternary complexes (A and B) in the crystallographic asymmetric unit are very similar in

[1]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, Box 1677, 1425 Madison Avenue, New York, NY 10029, USA. [2]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA. [3]Northeastern Collaborative Access Team, Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439, USA. [4]Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77755–1061, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: aneel.aggarwal@mssm.edu

structure (for example, root mean square deviation of ~0.41 Å over 249 Cαs), except for residues 1 to 17, which could be traced as an α helix in complex A but are apparently disordered in complex B. The final refined model (R_{free}, 25.4%; R_{work}, 20.9%) consists of two PrimPol molecules (A, residues 1 to 17, 35 to 200, and 261 to 348; B, residues 35 to 200 and 261 to 348), two DNA templates (A, 12 nt; B, 8 nt), two DNA primers (A, 7 nt; B, 8 nt), two dATP molecules, two calcium ions, and a total of 98 solvent molecules (table S1). We describe below the structure corresponding to ternary complex A.

Overall structure

The PrimPol catalytic core enfolds the replicative end of the template-primer in a cavity lined by the antiparallel β sheets of two α/β modules, referred hereafter as ModN and ModC (Fig. 1 and fig. S1). ModN (residues 35 to 105; α2, α3, and β1 to β4) interacts primarily with the template DNA strand and the templating base (T), whereas ModC (residues 108 to 200 and 261 to 348; α4 to α6 and β5 to β13) harbors residues for catalysis and interactions with incoming nucleotide (dATP), as well as contacts with the template strand (Fig. 1). Notably, the primer DNA strand is almost completely devoid of protein contacts (described below).

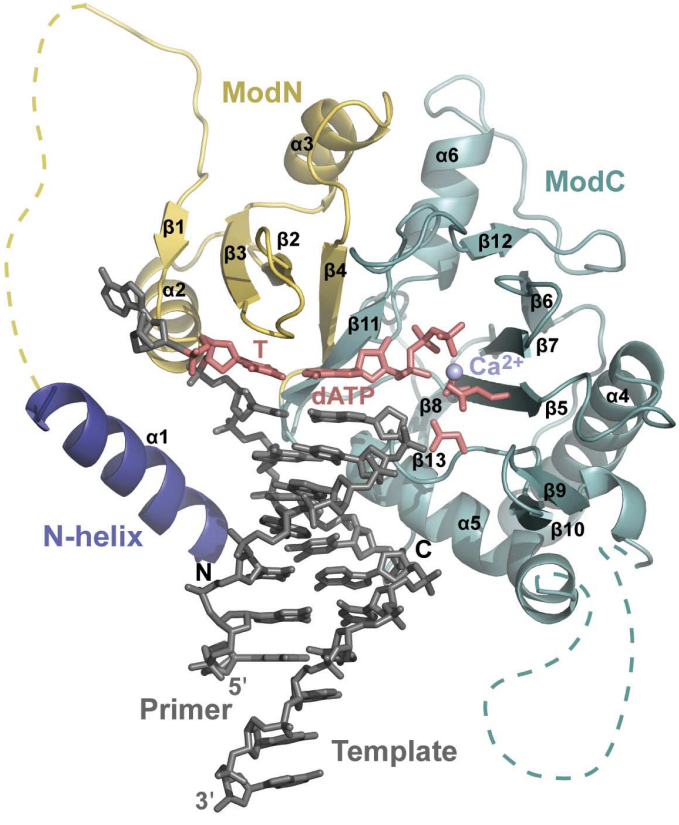

Fig. 1. Overall structure of human PrimPol ternary complex with template-primer DNA and incoming dATP. The N-helix and modules ModN and ModC are shown in cartoon representation in dark blue, yellow, and cyan, respectively. The DNA is shown as gray sticks, and the Ca^{2+} ion is shown as a light blue sphere. The templating base T and the incoming dATP are shown in red. Yellow and cyan dashed lines depict unstructured regions in the ModN and ModC, respectively. The side chains of key catalytic active-site residues Asp^{114}, Glu^{116}, and Asp^{280} are highlighted in red. Secondary structure elements (α helices and β strands) are labeled in black.

An N-terminal helix (N-helix; residues 1 to 17) connects to ModN via a long, flexible linker (residues 18 to 34) and juts into the DNA major groove (Fig. 1). The flexibility of the linker is indicated by the lack of density for this region in the electron density map. Residues 201 to 260, connecting helix α5 and strand β9 in ModC, are also undefined in the electron density map. This segment is most variable in sequence and length among PrimPol homologs and predicted to be largely unstructured (fig. S1). It may serve as a scaffold to recruit other components of the replication machinery.

The PrimPol catalytic core has been described as containing three conserved sequence motifs (I, II, and III) (*8*, *9*), wherein motifs I (DxE) and III (hDh) contain the acidic catalytic residues, and motif II (SxH) contains the incoming nucleotide binding motif. Accordingly, mutagenesis of Asp^{114} and Glu^{116} in motif I and His^{169} in motif II has been shown to abrogate the polymerase and primase activities of human PrimPol (*8*–*10*). All three motifs are located within ModC. However, motif III is much more extended in sequence and structure than previously predicted, containing not only an acidic catalytic residue (Asp^{280}) but also conserved residues that interact with the template strand and the incoming nucleotide (Fig. 1 and fig. S1). Here, we also highlight two new motifs, Ia (RQ) and Ib (QRhY/F), in ModN that harbor most of the residues that interact with the template strand.

The classical DNA polymerase fold has been likened to a partially open right hand with palm, fingers, and thumb domains (*17*–*19*), wherein the palm domain contains the active site, the finger domain interacts with the nascent base pair, and the thumb domain enfolds the duplex portion of the template-primer. The human PrimPol catalytic core does not bear any obvious resemblance to a polymerase fold (fig. S2). There is no thumb domain to grip the template-primer, and ModC encompasses functions of both the finger and palm domains in containing the active site and interacting with the nascent T-dATP base pair (Fig. 1 and fig. S2). The N-helix is spatially related to the polymerase-associated domain (PAD) in Y-family polymerases but makes far fewer contacts in the major groove (fig. S2) (*20*–*22*).

Nucleotide incorporation

The template T makes standard Watson-Crick (WC) hydrogen bonds with incoming dATP (Figs. 1 and 2A), in accord with the ~1000-fold bias toward correct versus incorrect WC base pairing (*8*). The dATP triphosphate moiety lies in a depression on the ModC surface, with residues Lys^{165}, Ser^{167}, and His^{169} from motif II and Arg^{291} and Lys^{297} from motif III, making direct hydrogen bonds with the phosphate groups (Fig. 2A and fig. S3). These residues are conserved in PrimPol homologs (fig. S1), and mutation of His^{169} has been previously shown to diminish the polymerase and primase activities of human PrimPol (*10*). The dATP sugar lies flush against the main-chain atoms of residues 289 to 291 and is buttressed in this position by a hydrogen bond between its 3′ hydroxyl and the main-chain amide of Arg^{291} and a hydrogen bond between its O4′ atom and the side chain of Arg^{288} (Fig. 2B). The close approach between the dATP sugar and main-chain atoms of residues 289 to 291 may explain the atypical preference of human PrimPol for dNTP over NTP (*8*, *9*). When we model ATP in place of dATP, the 2′ hydroxyl of ATP ribose sterically clashes with the Asn^{289} main-chain carbonyl (Fig. 2B). Human PrimPol can incorporate NTPs (at a lower efficiency) (*8*–*10*, *23*), but it would require a significant rearrangement of residues 289 to 291 and/or the sugar moiety.

The acidic catalytic residues Asp^{114}, Glu^{116}, and Asp^{280} are clustered between the dATP triphosphate moiety and the primer terminus (Fig. 2, A and B). A Ca^{2+} ion occupies a position corresponding to "metal B" in

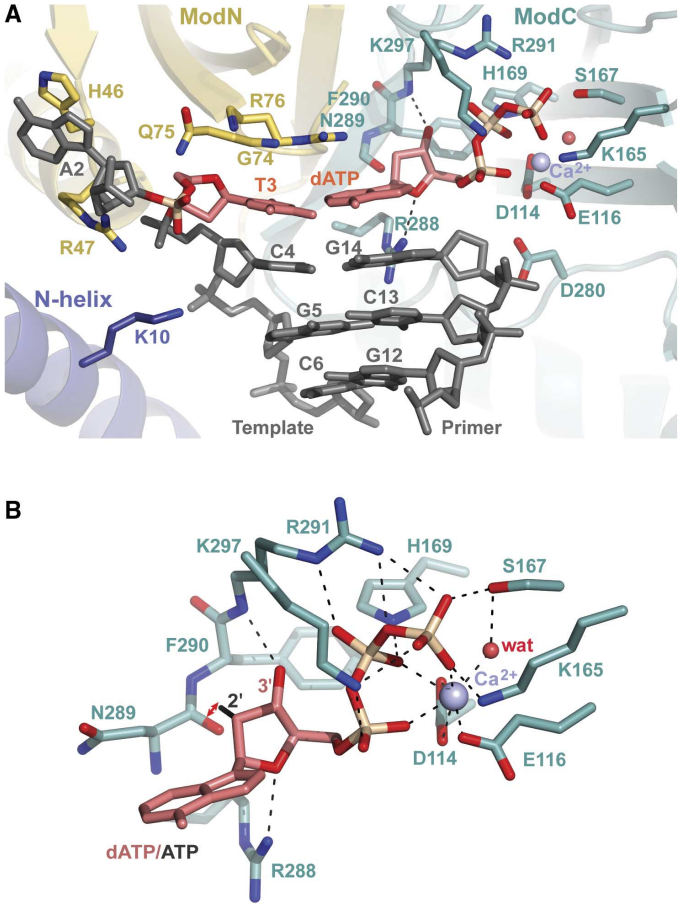

Fig. 2. PrimPol active-site region. (**A**) Close-up view of the PrimPol active-site region. The N-helix and the modules ModN and ModC are shown in dark blue, yellow, and cyan, respectively. The DNA is colored gray, the templating base and the incoming dATP are in red, and the Ca^{2+} ion is in light blue. The key catalytic residues (Asp[114], Glu[116], and Asp[280]), the residues contacting the incoming dATP (Lys[165], Ser[167], His[169], Arg[288], Asn[289], Phe[290], Arg[291], and Lys[297]), the templating base (Gly[74], Gln[75], and Arg[76]), and the rest of the template strand (Lys[10], His[46], and Arg[47]) are shown in sticks and with oxygen atoms in bright red and nitrogen atoms in blue. (**B**) Details of PrimPol interactions with dATP. The 3′-OH of the dATP forms a hydrogen bond with the backbone amino group of Arg[291]. Modeling of ATP (2′-OH shown in black) indicates an unfavorable clash between 2′-OH and the backbone carboxyl oxygen of Asn[289].

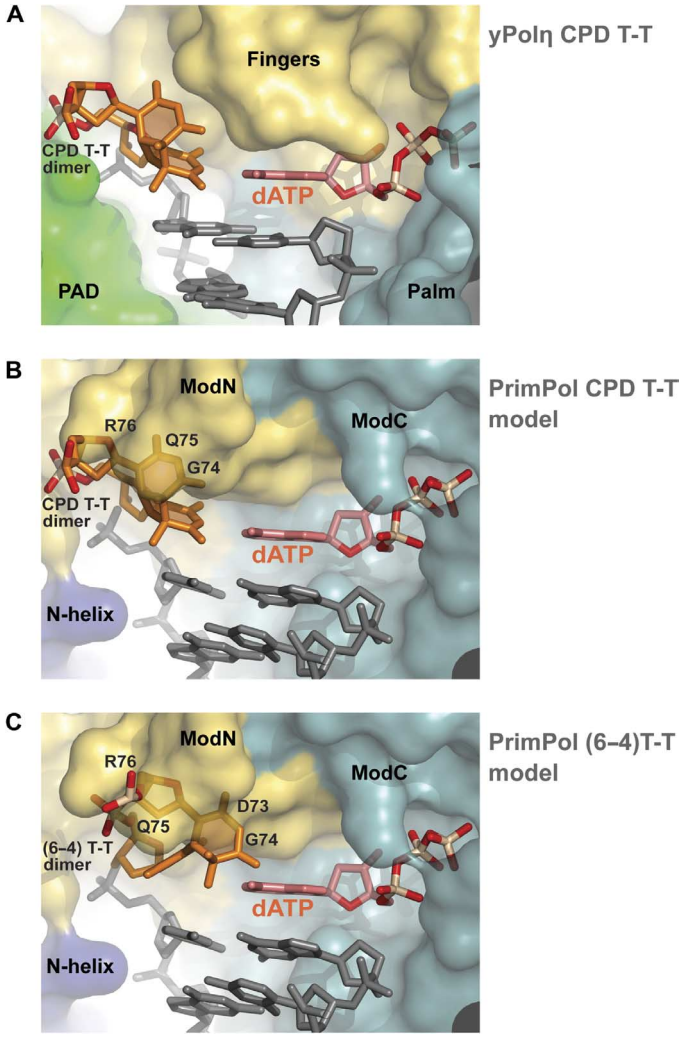

Fig. 3. PrimPol lacks space in its active-site cleft to accommodate UV-induced DNA lesions. (**A**) Structure of the yeast Polη inserting dATP opposite the 3′ base of the *cis-syn* thymine dimer in the template strand [Protein Data Bank (PDB) ID: 3MFI] (*21*). The fingers, palm, and PAD domains of Polη are colored yellow, cyan, and green, respectively; DNA is in gray, dATP is in red, and the *cis-syn* thymine dimer is in orange. This structure reveals an enlarged active site that accommodates the two covalently linked thymine bases. (**B**) Model of PrimPol with a CPD T-T dimer. The model is derived upon a superposition of PrimPol and yPolη (PDB ID: 3MFI) by the key active-site residues. The 5′ base of the CPD T-T dimer severely collides with the backbone atoms of Gly[74], Gln[75], and Arg[76] of the PrimPol's ModN. The N-helix, ModN, and ModC subdomain residues are shown in dark blue, yellow, and cyan, respectively. (**C**) Model of PrimPol with a (6–4) pyrimidine-pyrimidone T-T dimer. This model has been produced by a superposition of the undamaged portion of the (6–4) T-T–containing DNA duplex (PDB ID: 1CFL) (*43*) to the DNA in the PrimPol active site. The 5′ base of the (6–4) T-T dimer severely collides with the backbone atoms of Asp[73], Gly[74], Gln[75], and Arg[76] of the PrimPol's ModN.

polymerases (*24*), coordinated via all three phosphates of dATP, the carboxylates of Asp[114] and Glu[116], and a water molecule (Fig. 2, A and B). There is no density for a Ca^{2+} ion at a position corresponding to "metal A" in replicative or TLS polymerases, which likely reflects the fact that the primer in our structure contains a dideoxy terminus and that metal A tends to be much more labile than metal B (*25*). Nonetheless, the putative 3′ hydroxyl in our structure is located ~3.3 Å from the dATP α-phosphate and is aligned with respect to the Pα-O3′ bond (at an angle of ~150°) for an in-line nucleophilic attack, leading to the incorporation of A opposite template T (Fig. 2, A and B).

The ability of human PrimPol to function as a TLS polymerase

Human PrimPol has been shown to bypass UV- and oxidative-damaged DNA templates (*8*, *9*, *13*). Accordingly, we expected the active-site cleft of PrimPol to be "open" in the same manner as TLS polymerase Polη, which can efficiently replicate through UV-induced cyclobutane pyrimidine dimers (CPDs) and oxidative 8-oxoguanine DNA damage (Fig. 3A and fig. S2) (*21*, *22*, *26*). Unexpectedly, the PrimPol active-site cleft is relatively constrained with respect to the templating base (Fig. 1). In particular, residues 74 to 76, in a loop between strands β2 and β3 in ModN, bear down on the templating base, with the side chain of

Arg[76] draped over the base (Fig. 2A). This prevents the next 5′ nucleotide from stacking above the templating base and, consequently, only templating base T is held in the active-site cleft, whereas the rest of the 5′ unpaired template strand is directed out of the active-site cleft (with the 5′ unpaired nucleotide stabilized via stacking interactions with His[46] from the newly ascribed motif Ia) (Fig. 2A). Accordingly, when we model the 3′T of a UV-induced *cis-syn* T-T dimer at the templating position in the PrimPol active-site cleft (Fig. 3B), the 5′T of the T-T dimer (covalently linked to the 3′T by a cyclobutane ring) overlaps with Gly[74] and the side chain of Arg[76]. This may explain why human PrimPol in some experiments only extends a primer when nucleotides have already been incorporated opposite a *cis-syn* T-T dimer (*9*). Perhaps most surprising is the finding that human PrimPol can insert a T opposite the 3′T of a UV-induced (6–4) T-T photoproduct (*9*, *13*). A (6–4) T-T photoproduct is much more distorted than a *cis-syn* T-T dimer, and when we model the 3′T photoproduct at the templating position (Fig. 3C), the 5′T (covalently linked to the 3′T via a single C6–C4 bond) again sterically overlaps with Gly[74] and Arg[76]. It is conceivable that PrimPol bypasses the (6–4) T-T photoproduct and other bulky DNA lesions by looping them out (*23*), possibly in the space between ModN and N-helix.

The ability of human PrimPol to function as a primase

Inherent to primer synthesis are long-standing ideas that (i) binding of a primase to a template should not depend on a preexisting primer and (ii) the primase catalytic core should contain two distinct nucleotide binding sites, so-called initiation and elongation sites, for the formation of the initial dinucleotide (*5*, *6*). The primer DNA strand in our structure is almost completely devoid of protein contacts (Fig. 1). One reason for the lack of contacts is the absence of a thumb domain in PrimPol, which, in replicative and TLS polymerases, makes many contacts with the primer strand (fig. S3). Consequently, only ~90 Å^2 solvent accessible surface area is buried at the PrimPol-primer interface, as compared to ~683 Å^2 in the replicative Polδ-primer interface and ~475 Å^2 in the TLS Polη-primer interface (*21*, *27*).

From the structure, we define for the first time the nucleotide initiation site in a primase catalytic core. The incoming dATP is at the elongation site of a primase, whereas the nucleotide at the 3′ end of the primer strand is at the initiation site (Fig. 4). We can model β- and γ-phosphates on the primer 3′ nucleotide to provide the closest picture yet of how two initial nucleotide triphosphates would be bound to a primase catalytic core to initiate de novo DNA synthesis (Fig. 4). The putative β- and γ-phosphates extend over to strands β9 and β10 in ModC but do not partake in hydrogen bonds with the nearest residues (Fig. 4). Thus, one reason why it has proved so difficult to capture a nucleotide at the initiation site of a primase (from soaking experiments) appears to be the inherent weak binding to this site.

The ability of human PrimPol to function as both a primase and a polymerase

A comparison of human PrimPol and PriS catalytic cores provides some insights into the ability of PrimPol to function as both a polymerase and a primase. Despite the low sequence identity (<15%), the catalytic cores of the two enzymes can be structurally aligned (fig. S4) (*16*). Many of the residues identified as important for catalysis and interaction with dNTP in PrimPol have counterparts in PriS. For example, the catalytic residues Asp[114], Glu[116], and Asp[280] in PrimPol are related to Asp[109], Asp[111], and Asp[306] in PriS, and residues Lys[165], Ser[167], His[169], Arg[288], and Arg[291] that secure dNTP in the PrimPol active site are related to Arg[163], Ser[160], His[166], His[315], and Lys[318] in PriS (*16*). The two catalytic cores

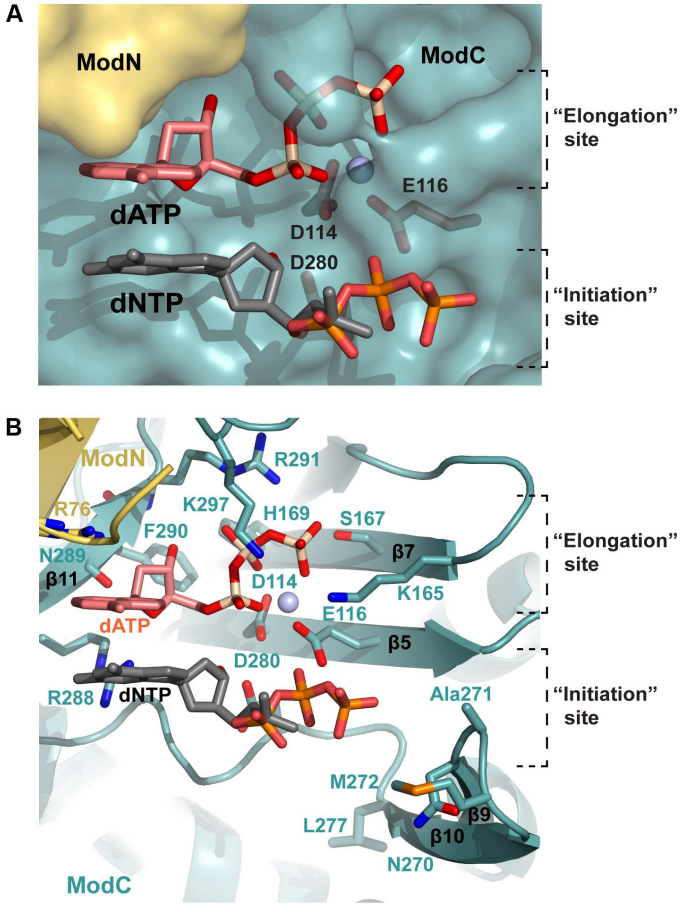

Fig. 4. PrimPol has room for dNTP at the initiation site. (**A**) A dNTP can be modeled at the initiation site (the oxygen atoms are colored bright red and the phosphorus atoms are in orange) without any clashes. This initiating dNTP becomes the 5′ end of the primer. To model the dNTP, we superimposed the base and sugar moieties of dATP on the base and sugar moieties of the 3′-terminal primer base (gray sticks). Thus, a phosphodiester bond between the incoming dNTP at the elongation site and the "initiating" dNTP can be formed to produce a primer strand with a terminal 5′-triphosphate. (**B**) A triphosphate of the initiating dNTP does not collide with any residues of PrimPol's ModC.

diverge primarily over the ModN and N-helix regions that carry most of the residues that interact with template strand in PrimPol (Fig. 5 and fig. S1). For example, invariant Lys[10] (N-helix), Arg[47] (motif Ia), and Gln[75] (motif Ib) make direct hydrogen bonds with successive phosphates on the template strand, whereas Arg[76] and Tyr[78] on motif Ib make van der Waals contacts with the template base and sugar moiety (Fig. 5). By contrast, there is no equivalent of N-helix in PriS and only weak correspondence to motifs Ia and Ib. Together, it is feasible that these conserved residues that latch human PrimPol onto the template strand provide the necessary hold for polymerase activity. Intriguingly, this is a smaller "footprint" on the template strand than with most other polymerases and may underlie the highly distributive nature of PrimPol, whereby it adds only ~4 nt before dissociating (*28*).

DISCUSSION

An unavoidable aspect of life is the threat posed by external and internal DNA damaging agents to the integrity of genomic DNA. Although a

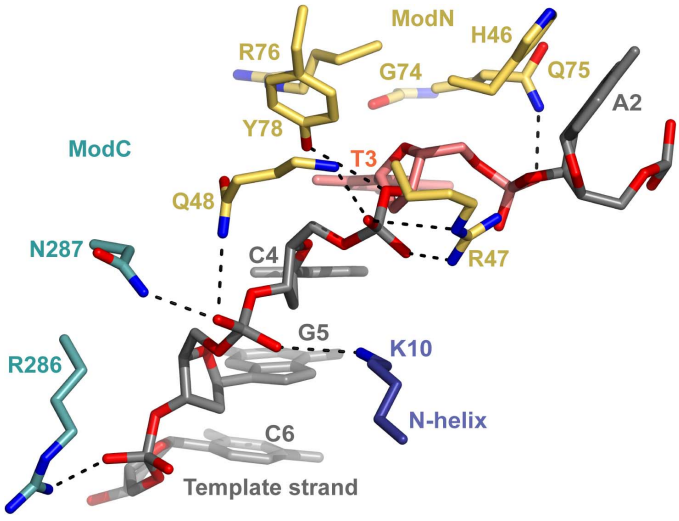

Fig. 5. PrimPol interactions with the template strand. The N-helix, ModN, and ModC residues are shown in dark blue, yellow, and cyan, respectively. The DNA is colored gray, and the templating base T3 is in red. The oxygen atoms of the backbone phosphate groups and O4′ of sugars are highlighted in bright red.

variety of repair mechanisms exist to remove the resulting DNA lesions, some lesions escape repair and block the replication machinery. One solution to maintain continuity of the replication fork is TLS, whereby a specialized polymerase(s) is recruited to bypass the lesion. Another solution is to skip the lesion altogether and reprime DNA synthesis downstream of the lesion. Both of these seemingly disparate activities are embodied in PrimPol in human cells. Here, we present the first crystal structure of PrimPol in complex with its DNA and incoming nucleotide substrates. The structure is also the first, to our knowledge, of a primase catalytic core in association with a template-primer.

PrimPol plays a prominent role in the bypass of UV-induced DNA adducts, such as a *cis-syn* T-T dimer and a (6–4) T-T photoproduct (*9*, *13*), as well as in the bypass of 8-oxoguanine resulting from oxidative stress (*8*). The ability of PrimPol to bypass the distorted (6–4) T-T photoproduct suggested that its active-site cleft might be exceptionally spacious, perhaps even more so than in Polη, which can readily bypass a *cis-syn* T-T dimer but is partially inhibited by a (6–4) T-T photoproduct (*29*, *30*). Surprisingly, the PrimPol active-site cleft is constrained with respect to the templating base. Hence, it is easy to see how the PrimPol can bypass a small 8-oxoguanine type lesion, but not so easy to see how it can accommodate a *cis-syn* T-T dimer or a (6–4) T-T photoproduct. It is possible that PrimPol uses a different mechanism to bypass a (6–4) T-T photoproduct or another bulky DNA lesion, whereby it loops it out in the space between ModN and N-helix, with the linker in between (residues 18 to 34) possibly stabilizing the displaced lesion.

The repriming of DNA synthesis to rescue a stalled replication fork is well established in prokaryotes and eukaryotes (*31*–*33*), but the enzyme responsible for it in human cells has remained unclear. PrimPol has emerged as perhaps the best candidate, which is capable of repriming DNA synthesis downstream of DNA lesions or other obstructions (*8*–*10*, *13*, *34*). The fact that PrimPol can reprime with dNTPs also minimizes the amount of subsequent RNA processing to complete the replication cycle. From the structure, we highlight two key features that lend to the ability of PrimPol to initiate de novo DNA synthesis. First, we note an almost complete lack of contacts to the DNA primer strand, making PrimPol much less dependent on a preexisting primer to initiate DNA synthesis. Second, we identify for the first time a site on the catalytic core where the initiating dNTP (or NTP) would bind. Almost all current models of primase action postulate two dNTP/NTP binding sites on the catalytic core, initiation and elongation, for the formation of the initial dinucleotide (*5*, *6*, *15*, *16*). The first dNTP/NTP is proposed to bind the elongation site followed by a second dNTP/NTP to the initiation site, with the initiating nucleotide triphosphate eventually becoming the 5′ end of the primer (*35*). The elongation nucleotide in our structure is held much more tightly to the enzyme than to the initiation nucleotide, which may dictate the order of nucleotide binding.

Although PrimPol uses the same active site for polymerase and primase activities, the catalytic core by itself is sufficient for polymerase activity but not for primase activity. The primase activity requires a zinc finger module (Znf; residues ~372 to 487), which is located C-terminal to the catalytic core and is shown capable of binding to single-stranded DNA (*28*). Hence, Znf may provide an additional grip on the template during the formation of the initial dinucleotide. Alternatively, Znf may function in a manner analogous to the Fe-S domain in PriL, a regulatory subunit that acts in association with PriS. A recent crystal structure of the Fe-S domain shows that it interacts with a nucleotide triphosphate that would be at the 5′ end of a growing primer (*36*). Thus, Znf could also act as a "translocation" site to capture the initiating nucleotide triphosphate following its translocation from the active site. In which case, it could offer a simple means to "count" the primer length, whereby once the primer has grown to a certain length, the initiating nucleotide triphosphate at the 5′ end of the primer may no longer be able to reach and bind Znf.

In conclusion, we present here the first structure of the catalytic core of human PrimPol in the act of inserting a nucleotide opposite a DNA template. We uncover structural features that afford PrimPol the ability to function as both a TLS polymerase and a DNA primase in maintaining genomic integrity.

MATERIALS AND METHODS

Cloning, expression, and purification of recombinant proteins

Human PrimPol catalytic domain (residues 1 to 354) was expressed in *Saccharomyces cerevisiae* as an N-terminally tagged glutathione *S*-transferase (GST) fusion protein and then purified as described for Polη (*20*). Briefly, the yeast cell pellets were lysed, and the proteins were precipitated by ammonium sulfate, resuspended, and loaded onto a glutathione-Sepharose column. The GST tag was cleaved on-column with PreScission protease, and the resulting PrimPol protein (carrying an extra GPGGDPH peptide on its N terminus due to the expression construct design) was further purified by sequential chromatography on HiTrap Heparin and Superdex 75 columns (GE Healthcare). The protein was concentrated to ~35 mg/ml in 25 mM tris (pH 7.5), 250 mM NaCl, and 2 mM tris(2-carboxyethyl) phosphate (TCEP) and stored in aliquots at −80°C.

For the expression of SeMet-labeled PrimPol (residues 1 to 354), a codon-optimized gene sequence was synthesized (GenScript) and inserted into a modified pET28b(+) vector (Novagen) to produce His_6-SUMO (small ubiquitin-like modifier) N-terminally tagged fusion protein. In addition, to increase the number of methionines, a triple $Leu^{97}Met$ $Leu^{250}Met$ $Ile^{315}Met$ mutant was generated by site-directed mutagenesis from the template plasmid (GenScript). The clones were transformed into *Escherichia coli* BL21 (DE3) codon Plus RIL

(Stratagene). The cells were grown in M9 minimal medium at 37°C and induced by the addition of 1 mM isopropyl-β-D-1-thiogalactopyranoside in the presence of a cocktail of inhibitory amino acids and SeMet according to the manufacturer's specifications (Shanghai Medicilon Inc.). Induction was performed at 18°C overnight. The bacterial cells were pelleted, resuspended, and lysed by sonication. The proteins were purified from the soluble fraction by a nickel-chelating affinity column, followed by cleavage of the His_6-SUMO tag with Ulp1 protease. An additional pass through HisTrap High Performance was performed to bind the cleaved His_6-SUMO tag. The SeMet-labeled proteins were further purified as described for the native protein. SeMet incorporation was verified by mass spectrometry.

DNA preparation

For crystallography, the 13-mer DNA oligonucleotide primer was synthesized with a 2′,3′-dideoxyguanosine at its 3′ end (5′-GGGT-GTGGTAGCG^{dd}-3′) (W.M. Keck Foundation, Yale), purified by ion exchange chromatography, and desalted by dialysis. The 12-mer primer (5′-GGTGTGGTAGCG-3′) used for crystallization of SeMet-containing proteins, the 17-mer template (5′-CATCGC-TACCACACCCC-3′), and all other DNA oligonucleotides used for crystallography and the activity and binding studies were purified using reversed-phase high-performance liquid chromatography (Integrated DNA Technologies).

Crystallization, data collection, and structure determination

Initial crystallization conditions for the native and SeMet-labeled PrimPol (residues 1 to 354)–DNA–dNTP ternary complexes were determined with matrix screens (Hampton Research and Qiagen). Template-primer DNAs were annealed and mixed with PrimPol in a 1.2:1 molar ratio to a final concentration of 0.3 mM in 25 mM tris (pH 7.5), 155 mM NaCl, 1.5 mM TCEP, 12 mM $MgCl_2$, and 7 mM dATP or nonhydrolyzable dApNHpp analog (Jena Bioscience). The complexes were incubated at room temperature for 5 min and then centrifuged at 10,000 rpm for 10 min at 4°C. The crystals were grown at 20°C by a hanging drop method against a reservoir solution containing 200 mM $CaCl_2$ (for native PrimPol) or 300 mM NaSCN (for SeMet-labeled proteins) and 17 to 23% PEG 3350. The rod-like crystals appeared in 1 to 3 weeks; microseeding was required in some conditions to produce crystals up to ~0.2 mm in length. For data collection, crystals were cryoprotected stepwise in reservoir solution supplemented with 24% PEG 3350 and 10% and then 20% glycerol and flash-frozen in liquid nitrogen.

X-ray data on the native and SeMet-labeled PrimPol-DNA-dNTP cocrystals were measured at the Advanced Photon Source in Chicago (beamlines 24-ID and 23-ID). The data sets were processed using XDS (*37*), iMosflm (*38*), and aimless and pointless in the CCP4 program package (*39*). The native crystals diffracted to ~2.2 Å resolution and belonged to space group *P*1, whereas the SeMet crystals diffracted to ~3 to 3.5 Å resolution and belonged to space groups *C*2 and $P2_1$ (table S1). The *P*1 and $P2_1$ crystals contained two PrimPol-DNA-dNTP ternary complexes in the crystallographic asymmetric unit, whereas the *C*2 crystals contained only one complex. To calculate experimental phases for structure determination, we used SAD data measured at Se-*K* absorption edge (~0.9791 Å) on both forms of SeMet crystals (*C*2 and $P2_1$). The phases were calculated with program SHARP (*40*), and the density-modified, noncrystallographic symmetry–averaged maps were calculated using the program DM (*39*). The experimental maps showed reasonably clear electron density for most regions of the protein-DNA complexes. A model was built manually in Coot (*41*), refined in Phenix (*42*), and then positioned in the native *P*1 space group by molecular replacement. The model was further built in Coot and refined in Phenix in an iterative manner (to the 2.2 Å diffraction limit of the native data). As an additional aid to model building, we labeled the template strand with 5-iododeoxyuracil and measured another datum on these crystals at low energy (~7 keV or ~1.77 Å). The location of iodine peaks in the anomalous difference map calculated using anomalous data from these iodinated DNA crystals unambiguously confirmed the register of both DNA and primer strands in our model. The final model was refined to ~2.2 Å resolution with R_{free} and R_{work} values of ~25.4 and 20.9%, respectively, and displayed good stereochemistry (table S1). The two ternary complexes (A and B) in the crystallographic asymmetric unit of native space group *P*1 were very similar in structure, except for residues 1 to 17, which could be traced as an α helix in complex A but not in complex B. We considered whether this α helix in complex A was from a symmetry-related molecule (residues 244 to 260), but the orientation in that case would be reversed and did not match the electron density. Two other regions (residues 18 to 34 and 201 to 259) could not be traced in either complex A or complex B. These disordered regions lacked any predictable secondary structure and might only become ordered upon interaction with other protein partners.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/2/10/e1601317/DC1

fig. S1. Structure-based sequence alignment of the catalytic core domains of human, mouse, gecko, and zebrafish PrimPols.

fig. S2. Overall structures of PrimPol and yPolη ternary complexes with template-primer DNA and dNTP.

fig. S3. Simulated annealing F_o-F_c omit map for dATP, templating base T3, and Ca^{2+} residues within the PrimPol active-site region.

fig. S4. Overall structures of human PrimPol-DNA-dNTP complex and human primase PriS complex with uridine triphosphate.

table S1. Crystallographic parameters and refinement statistics.

REFERENCES AND NOTES

1. S. Waga, B. Stillman, The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* **67**, 721–751 (1998).
2. E. Johansson, S. A. Macneill, The eukaryotic replicative DNA polymerases take shape. *Trends Biochem. Sci.* **35**, 339–347 (2010).
3. R. E. Johnson, R. Klassen, L. Prakash, S. Prakash, A major role of DNA polymerase δ in replication of both the leading and lagging DNA strands. *Mol. Cell* **59**, 163–175 (2015).
4. S. Prakash, R. E. Johnson, L. Prakash, Eukaryotic translesion synthesis DNA polymerases: Specificity of structure and function. *Annu. Rev. Biochem.* **74**, 317–353 (2005).
5. R. D. Kuchta, G. Stengel, Mechanism and evolution of DNA primases. *Biochim. Biophys. Acta* **1804**, 1180–1189 (2010).
6. D. N. Frick, C. C. Richardson, DNA primases. *Annu. Rev. Biochem.* **70**, 39–80 (2001).
7. W. C. Copeland, T. S. Wang, Enzymatic characterization of the individual mammalian primase subunits reveals a biphasic mechanism for initiation of DNA replication. *J. Biol. Chem.* **268**, 26179–26189 (1993).
8. S. García-Gómez, A. Reyes, M. I. Martínez-Jiménez, E. S. Chocrón, S. Mourón, G. Terrados, C. Powell, E. Salido, J. Méndez, I. J. Holt, L. Blanco, PrimPol, an archaic primase/polymerase operating in human cells. *Mol. Cell* **52**, 541–553 (2013).
9. J. Bianchi, S. G. Rudd, S. K. Jozwiakowski, L. J. Bailey, V. Soura, E. Taylor, I. Stevanovic, A. J. Green, T. H. Stracker, H. D. Lindsay, A. J. Doherty, PrimPol bypasses UV photoproducts during eukaryotic chromosomal DNA replication. *Mol. Cell* **52**, 566–573 (2013).
10. L. Wan, J. Lou, Y. Xia, B. Su, T. Liu, J. Cui, Y. Sun, H. Lou, J. Huang, hPrimpol1/CCDC111 is a human DNA primase-polymerase required for the maintenance of genome integrity. *EMBO Rep.* **14**, 1104–1112 (2013).
11. L. M. Iyer, E. V. Koonin, D. D. Leipe, L. Aravind, Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: Structural insights and new members. *Nucleic Acids Res.* **33**, 3875–3896 (2005).

12. T. Helleday, PrimPol breaks replication barriers. *Nat. Struct. Mol. Biol.* **20**, 1348–1350 (2013).
13. S. Mourón, S. Rodriguez-Acebes, M. I Martínez-Jiménez, S. García-Gómez, S. Chocrón, L. Blanco, J. Méndez, Repriming of DNA synthesis at stalled replication forks by human PrimPol. *Nat. Struct. Mol. Biol.* **20**, 1383–1389 (2013).
14. G. Lipps, A. O. Weinzierl, G. von Scheven, C. Buchen, P. Cramer, Structure of a bifunctional DNA primase-polymerase. *Nat. Struct. Mol. Biol.* **11**, 157–162 (2004).
15. D. N. Frick, S. Kumar, C. C. Richardson, Interaction of ribonucleoside triphosphates with the gene 4 primase of bacteriophage T7. *J. Biol. Chem.* **274**, 35899–35907 (1999).
16. M. L. Kilkenny, M. A. Longo, R. L. Perera, L. Pellegrini, Structures of human primase reveal design of nucleotide elongation site and mode of Pol α tethering. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15961–15966 (2013).
17. D. L. Ollis, P. Brick, R. Hamlin, N. G. Xuong, T. A. Steitz, Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* **313**, 762–766 (1985).
18. S. Doublié, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251–258 (1998).
19. P. J. Rothwell, G. Waksman, Structure and mechanism of DNA polymerases. *Adv. Protein Chem.* **71**, 401–440 (2005).
20. J. Trincao, R. E. Johnson, C. R. Escalante, S. Prakash, L. Prakash, A. K. Aggarwal, Structure of the catalytic core of *S. cerevisiae* DNA polymerase η: Implications for translesion DNA synthesis. *Mol. Cell* **8**, 417–426 (2001).
21. T. D. Silverstein, R. E. Johnson, R. Jain, L. Prakash, S. Prakash, A. K. Aggarwal, Structural basis for the suppression of skin cancers by DNA polymerase η. *Nature* **465**, 1039–1043 (2010).
22. C. Biertümpfel, Y. Zhao, Y. Kondo, S. Ramón-Maiques, M. Gregory, J. Y. Lee, C. Masutani, A. R. Lehmann, F. Hanaoka, W. Yang, Structure and mechanism of human DNA polymerase η. *Nature* **465**, 1044–1048 (2010).
23. M. I. Martínez-Jiménez, S. García-Gómez, K. Bebenek, G. Sastre-Moreno, P. A. Calvo, A. Díaz-Talavera, T. A. Kunkel, Luis Blanco, Alternative solutions and new scenarios for translesion DNA synthesis by human PrimPol. *DNA Repair* **29**, 127–138 (2015).
24. T. A. Steitz, DNA polymerases: Structural diversity and common mechanisms. *J. Biol. Chem.* **274**, 17395–17398 (1999).
25. S. Lone, S. A. Townson, S. N. Uljon, R. E. Johnson, A. Brahma, D. T. Nair, S. Prakash, L. Prakash, A. K. Aggarwal, Human DNA polymerase κ encircles DNA: Implications for mismatch extension and lesion bypass. *Mol. Cell* **25**, 601–614 (2007).
26. T. D. Silverstein, R. Jain, R. E. Johnson, L. Prakash, S. Prakash, A. K. Aggarwal, Structural basis for error-free replication of oxidatively damaged DNA by yeast DNA polymerase η. *Structure* **18**, 1463–1470 (2010).
27. M. K. Swan, R. E. Johnson, L. Prakash, S. Prakash, A. K. Aggarwal, Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase δ. *Nat. Struct. Mol. Biol.* **16**, 979–986 (2009).
28. B. A. Keen, S. K. Jozwiakowski, L. J. Bailey, J. Bianchi, A. J. Doherty, Molecular dissection of the domain architecture and catalytic activities of human PrimPol. *Nucleic Acids Res.* **42**, 5830–5845 (2014).
29. R. E. Johnson, S. Prakash, L. Prakash, Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Polη. *Science* **283**, 1001–1004 (1999).
30. S.-L. Yu, R. E. Johnson, S. Prakash, L. Prakash, Requirement of DNA polymerase η for error-free bypass of UV-induced CC and TC photoproducts. *Mol. Cell. Biol.* **21**, 185–188 (2001).
31. R. C. Heller, K. J. Marians, Replication fork reactivation downstream of a blocked nascent leading strand. *Nature* **439**, 557–562 (2006).
32. M. Lopes, M. Foiani, J. M. Sogo, Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol. Cell* **21**, 15–27 (2006).
33. I. Elvers, F. Johansson, P. Groth, K. Erixon, T. Helleday, UV stalled replication forks restart by re-priming in human fibroblasts. *Nucleic Acids Res.* **39**, 7049–7057 (2011).
34. D. Schiavone, S. K. Jozwiakowski, M. Romanello, G. Guilbaud, T. A. Guilliam, L. J. Bailey, J. E. Sale, A. J. Doherty, PrimPol is required for replicative tolerance of G quadruplexes in vertebrate cells. *Mol. Cell* **61**, 161–169 (2016).
35. R. J. Sheaff, R. D. Kuchta, Mechanism of calf thymus DNA primase: Slow initiation, rapid polymerization, and intelligent termination. *Biochemistry* **32**, 3027–3037 (1993).
36. A. G. Baranovskiy, N. D. Babayeva, Y. Zhang, J. Gu, Y. Suwa, Y. I. Pavlov, T. H. Tahirov, Mechanism of concerted RNA-DNA primer synthesis by the human primosome. *J. Biol. Chem.* **291**, 10006–10020 (2016).
37. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
38. T. G. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, A. G. W. Leslie, *iMOSFLM*: A new graphical interface for diffraction-image processing with *MOSFLM*. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281 (2011).
39. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, Overview of the *CCP*4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
40. E. de La Fortelle, G. Bricogne, Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
41. P. Emsley, K. Cowtan, *Coot*: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
42. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *PHENIX*: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
43. J.-H. Lee, G.-S. Hwang, B.-S. Choi, Solution structure of a DNA decamer duplex containing the stable 3′ T·G base pair of the pyrimidine(6–4)pyrimidone photoproduct [(6–4) adduct]: Implications for the highly specific 3′ T → C transition of the (6–4) adduct. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6632–6636 (1999).

Acknowledgments

Funding: This work is based on research conducted at the Northeastern Collaborative Access Team (NECAT) and GM/CA @ APS beamlines at the Argonne National Laboratory. The NECAT beamlines were funded by the National Institute of General Medical Sciences (NIGMS) from the NIH (P41 GM103403). The GM/CA @ APS beamlines have been funded in whole or in part with funds from the National Cancer Institute (ACB-12002) and the NIGMS (AGM-12006). The Pilatus 6M detector on 24-ID-C beamline was funded by an NIH–Office of Research Infrastructure Programs High-End Instrumentation grant (S10 RR029205). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. **Author contributions:** O.R. and A.K.A. designed the experiments. O.R., R.M., and R.E.J. performed protein expression and purification. L.P. and S.P. guided protein expression in yeast. O.R. and R.M. performed crystallization. O.R. R.M., and Y.K.G. performed x-ray data collection. Y.K.G., K.R.R., O.R., and R.M. performed phasing and refinement. A.K.A. and O.R. wrote the paper. All authors reviewed the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Atomic coordinates and structure factors have been deposited in the PDB under accession code 5L2X. All other data needed to evaluate the conclusions in the manuscripts are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 10 June 2016
Accepted 19 September 2016
Published 21 October 2016
10.1126/sciadv.1601317

Citation: O. Rechkoblit, Y. K. Gupta, R. Malik, K. R. Rajashankar, R. E. Johnson, L. Prakash, S. Prakash, A. K. Aggarwal, Structure and mechanism of human PrimPol, a DNA polymerase with primase activity. *Sci. Adv.* **2**, e1601317 (2016).

Structure and mechanism of human PrimPol, a DNA polymerase with primase activity

Olga Rechkoblit, Yogesh K. Gupta, Radhika Malik, Kanagalaghatta R. Rajashankar, Robert E. Johnson, Louise Prakash, Satya Prakash and Aneel K. Aggarwal

Sci Adv **2** (10), e1601317.
DOI: 10.1126/sciadv.1601317