Published online 29 January 2018

Nucleic Acids Research, 2018, Vol. 46, No. 4 ***2107–2120***
doi: 10.1093/nar/gky031

The C-terminal region of translesion synthesis DNA polymerase η is partially unstructured and has high conformational flexibility

Kyle T. Powers, Adrian H. Elcock and M. Todd Washington*

Department of Biochemistry, University of Iowa College of Medicine, Iowa City, IA 52242-1109, USA

Received November 21, 2017; Revised January 10, 2018; Editorial Decision January 11, 2018; Accepted January 22, 2018

ABSTRACT

Eukaryotic DNA polymerase η catalyzes translesion synthesis of thymine dimers and 8-oxoguanines. It is comprised of a polymerase domain and a C-terminal region, both of which are required for its biological function. The C-terminal region mediates interactions with proliferating cell nuclear antigen (PCNA) and other translesion synthesis proteins such as Rev1. This region contains a ubiquitin-binding/zinc-binding (UBZ) motif and a PCNA-interacting protein (PIP) motif. Currently little structural information is available for this region of polymerase η. Using a combination of approaches—including genetic complementation assays, X-ray crystallography, Langevin dynamics simulations, and small-angle X-ray scattering—we show that the C-terminal region is partially unstructured and has high conformational flexibility. This implies that the C-terminal region acts as a flexible tether linking the polymerase domain to PCNA thereby increasing its local concentration. Such tethering would facilitate the sampling of translesion synthesis polymerases to ensure that the most appropriate one is selected to bypass the lesion.

INTRODUCTION

DNA damage, which arises from both endogenous and exogenous sources, interferes with normal DNA replication and threatens genome stability. Translesion synthesis is an important process that allows the replication machinery to bypass DNA lesions and mitigate their deleterious effects (1–11). This process is initiated when one of the classical DNA polymerases at the replication fork encounters a lesion in the template strand that it cannot accommodate within its active site. This causes stalling of the replication fork. Proliferating cell nuclear antigen (PCNA)—an essential replication accessory factor—is mono-ubiquitylated, which is believed to signal the recruitment of multiple non-classical DNA polymerases to the stalled fork (12–14). These non-classical polymerases form a large, multi-protein complex around the ubiquitin-modified PCNA (Ub-PCNA) (15–17). Ultimately, one or more of these non-classical polymerases replace the stalled classical polymerase on the DNA and synthesize DNA across from the template lesion.

Polymerase η is a well-characterized non-classical polymerase that is found in all eukaryotes. It is a member of the Y-family of DNA polymerases, which are all involved in translesion synthesis (2,4–7,9). It catalyzes the efficient and accurate bypass of thymine dimers—lesions induced by ultraviolet (UV) radiation—and 8-oxoguanine adducts—lesions induced by reactive oxygen species (18–20). Yeast strains lacking polymerase η are sensitive to UV radiation and have an increased rate of UV-induced mutagenesis (21,22). Likewise, the loss of polymerase η in humans is responsible for the variant form of xeroderma pigmentosum (XPV), a rare cancer-prone genetic disorder (23,24).

Polymerase η is comprised of a polymerase domain and a C-terminal region. The polymerase domain contains fingers, palm, and thumb sub-domains as well as another sub-domain referred to as either the little finger sub-domain or the polymerase-associated domain (PAD) (25). A series of X-ray crystal structures of the polymerase domain bound to various damaged DNA substrates have been determined (26,27). These structures show that the enzyme has a larger active site than classical DNA polymerases. This allows polymerase η to accommodate the geometric distortions in the DNA substrate caused by these lesions and incorporate nucleotides opposite them utilizing normal Watson–Crick base pairing. In fact, steady state and pre-steady state kinetic studies have shown that polymerase η incorporates nucleotides opposite thymine dimers and 8-oxoguanines with the same efficiency with which it incorporates nucleotides opposite non-damaged templates (19,20,28,29).

The C-terminal region of polymerase η contains approximately 120 amino acid residues. It is dispensable for enzymatic activity *in vitro*, but is required for the biological function of polymerase η *in vivo* (30). This is because the C-terminal region mediates several protein-protein interac-

*To whom correspondence should be addressed. Tel: +1 319 335 7518; Fax: +1 319 335 9570; Email: todd-washington@uiowa.edu

tions needed for the proper function of polymerase η in cells. The C-terminal region contains an ubiquitin-binding/zinc-binding (UBZ) motif that is believed to interact with the ubiquitin moiety on Ub-PCNA (14). This region also contains a PCNA-interacting protein (PIP)-like motif. The PIP-like motif has been shown to mediate its interactions with PCNA as well as other proteins involved in translesion synthesis such as Rev1—another non-classical Y-family DNA polymerase (31–33).

There is very little structural information available for the C-terminal region of polymerase η. A low-resolution (>20 Å) structure of a complex containing polymerase η, Ub-PCNA and DNA determined by single particle electron microscopy suggested that this region might be structured (34). By contrast, several, independent analyses of the amino acid sequence have predicted this region to be largely unstructured (9,35). To resolve this controversy and to obtain reliable structural information about the C-terminal region of polymerase η, we used a variety of approaches including genetic complementation assays, X-ray crystallography, Langevin dynamics (LD) simulations, and small-angle X-ray scattering (SAXS). We used a novel approach in which we directly compared full ensembles derived from LD simulations to experimental SAXS data without reliance on curve fitting. The best agreement between the simulations and experimental data was achieved using an ensemble derived from simulations in which the C-terminal region of polymerase η was partially unstructured and displayed high conformational flexibility. This implies that the C-terminal region of polymerase η acts as a flexible tether linking the polymerase domain to PCNA thereby increasing its local concentration. Such tethering is likely a feature of other non-classical Y-family polymerases. This would facilitate the sampling of translesion synthesis polymerases to ensure that the most appropriate one is selected to bypass the lesion.

MATERIALS AND METHODS

Ultraviolet (UV) radiation sensitivity assays

The sensitivity to UV radiation was assayed using a yeast complementation assay (22). The full length *RAD30* gene (which encodes polymerase η) with an additional 1000 bp of yeast genomic DNA both upstream and downstream of the open reading frame was cloned into the low copy number *CEN LEU2* yeast shuttle vector YCplac111 (36) to yield pKW692. The Q5 Site-Directed Mutagenesis Kit (New England Biolabs) was used to introduce base substitutions into this plasmid to generate constructs that encode polymerase η mutant proteins. Each mutant protein had 10 consecutive amino acid residues substituted to alanine and glycine residues in the following pattern: Ala-Ala-Gly-Ala-Ala-Gly-Ala-Ala-Gly-Ala. We made a total of 23 constructs with substitutions of the following blocks of 10 residues: 513–522 (pKW693), 518–527 (pKW694), 523–532 (pKW695), 528–537 (pKW696), 533–542 (pKW697), 538–547 (pKW698), 543–552 (pKW699), 548–557 (pKW700), 553–562 (pKW701), 558–567 (pKW702), 563–572 (pKW703), 568–577 (pKW704), 573–582 (pKW705), 578–587 (pKW706), 583–592 (pKW707), 588–597 (pKW708), 593–602 (pKW709), 598–607 (pKW710), 603–612 (pKW711), 608–617 (pKW712), 613–622 (pKW713), 618–627 (pKW714), 623–632 (pKW715). The YCplac111 empty vector and plasmids pKW692 to pKW715 were transformed into yeast strain EMY74.7 with the genomic *RAD30* and *RAD5* genes deleted. For each yeast strain, serial dilutions containing 10^8, 10^7, 10^6 and 10^5 cells/ml were spotted on synthetic complete media lacking leucine. These cells were exposed to 0, 10, 20 and 50 J/m^2 of UV radiation and incubated in the dark at 30°C for 3–4 days. Each mutant strain was scored blindly using a three-point scale: 1 corresponded to the sensitivity of the strain harboring the YCplac111 plasmid (the negative control), 2 corresponded to a sensitivity in between the negative control and the positive control and 3 corresponded to the sensitivity of the strain harboring the pKW692 plasmid (the positive control). Three technical replicates of three biological replicates for each strain were scored using this scale and all nine scores for each strain were averaged.

Protein expression and purification

Codon optimized genes (Invitrogen) encoding N-terminally 6-His-tagged full-length polymerase η, 6-His-tagged polymerase η polymerase domain (residues 1–513), and 6-His-tagged polymerase η extended polymerase domain (residues 1–528) were cloned into pET-11a (Novagene) to generate pKW593, pKW594 and pKW595, respectively. The base sequence of the codon optimized gene for full-length polymerases η is provided in Supplementary Figure S1. Proteins were over-expressed in bacterial strain BL21(DE3) grown in LB media to an OD_{600} of 0.6 and induced with 1 mM IPTG at 18°C for 16 h. Pelleted cells were re-suspended and lysed at 4°C using an EmulsiFlex (Avestin) in buffer A (50 mM TrisCl, pH 7.0, 500 mM NaCl and 10 mM imidazole) with 1 mM PMSF and Complete, EDTA-free Protease Inhibitor Cocktail (Roche). The crude extracts were clarified by ultracentrifugation. Each of the proteins was purified in three chromatography steps at 4°C. First, the protein was loaded onto a 3 ml Ni-NTA Superflow column (Qiagen) in buffer A and eluted with a gradient of 10–500 mM imidazole in buffer A. Second, the protein was loaded onto a 10 ml SP Sepharose Fast Flow cation exchange column (GE Healthcare) in buffer B (50 mM MES, pH 6.0, 100 mM NaCl and 1 mM DTT) and eluted with a gradient of 100 mM NaCl to 1 M NaCl in buffer B. Third, the protein was loaded onto a pre-packed 120 ml HiLoad SuperDex 200 size-exclusion column (GE Healthcare) in buffer C (20 mM HEPES, pH 7.5, 200 mM NaCl and 5 mM DTT) with 5 mM $MgCl_2$ for SAXS or in buffer C with 200 mM $MgCl_2$ for X-ray crystallography. Purified proteins were concentrated to 25 mg/ml and used immediately in crystal trials or stored at 4°C for up to 12 h for SAXS data collection.

Protein crystallization and structural determination

The polymerase η extended polymerase domain was crystallized using the hanging drop method with drops comprised of 400 nl of protein (25 mg/ml) and 400 nl of various commercial crystallization buffers at 18°C. The best diffracting crystals were obtained with buffer containing 80 mM

sodium cacodylate, pH 6.5, 160 mM calcium acetate, 20% glycerol and 14.4% (w/v) PEG8000. Protein crystals formed after 5 days. The crystals were flash frozen in liquid nitrogen for shipment to the Advance Light Source (Lawrence Berkeley National Labs). Data were collected at the Molecular Biology Consortium's 4.2.2 beamline. Data were processed and scaled using XDS and SCALA (37,38). The structure of the polymerase domain of polymerase η lacking a bound magnesium ion (PDB ID: 1JIH) was used for molecular replacement using PHASER (25,38). Model building and refinement were carried out using COOT, PHENIX and FFX (39–41). The coordinates and structure factors for the polymerase domain of polymerase η with a bound magnesium ion were deposited in the Protein Database (PDB ID: 5VTP).

Construction of DNA polymerase models for simulations

Two separate models of full-length DNA polymerase η were built for the LD simulations. The first model used the previously determined X-ray crystal structure of the polymerase domain lacking a bound magnesium ion (PDB ID: 1JIH). In this model, 51 side chains were built because the deposited PDB file did not contain side chains for these residues. The second model used the structure determined here of the polymerase domain with a bound magnesium ion (PDB ID: 5VTP). 6-His tags were added to the N-termini of the polymerase domains using the loop-modeling program LOOPY (42). At this stage, these models were used in the LD simulations of the polymerase domain alone. For the simulations of the full-length protein, further model building was performed. A homology model of the UBZ motif (residues 549–582) was built from a template model (PDB ID: 2I5O) using SWISS-MODEL (43,44) and models of putative α-helix T (residues 514–525), α-helix V (residues 593–600), and α-helix W (residues 621–632) were built using the ProBuilder On-line webserver. Construction of the full-length models was done to connect the structured regions to one another by flexible linkers in four steps. First, the C terminus of the polymerase domain (residues 1–509) was positioned near the N terminus of α-helix T using VMD, and a linker corresponding to residues 510–513 was built connecting these two pieces using LOOPY. Second, the C terminus of the resulting model (residues 1–525) was positioned near the N terminus of the UBZ, and a linker corresponding to residues 526–548 was built connecting these pieces. Third, the C terminus of the resulting model (residues 1–582) was positioned near the N terminus of α-helix V, and a linker corresponding to residues 583–592 was built connecting these pieces. Finally, the C terminus of the resulting model (residues 1–600) was positioned near the N terminus of α-helix W, and a linker corresponding to residues 601–620 was built connecting these pieces.

Langevin dynamics simulations

Since simulating the dynamics of DNA polymerase η at full atomic resolution is currently not feasible, the atomic models were converted into coarse-grained (CG) models that retained most of the structural detail but contained fewer particles. We built CG models in which each residue is replaced by one to four pseudo-atoms (CG beads) depending on the size of the residue using methods reported previously (45). For example, glycine is modeled as a single GC bead, while tryptophan is modeled as four CG beads. Partial charges were added to ionizable residues using the Henderson-Hasselbalch equation and a simulation pH equal to 7.5 to match the pH used in the SAXS experiments. The entire charge was placed on the CG bead that most closely matched the position of the ionizable group. The hydrodynamic radii for all CG beads were set to 3.5 Å, as this value was shown previously to most accurately reproduce translational and rotational diffusion coefficients for a wide variety of proteins (45).

The function used to describe the energy, $E(r)$ of the polymerase as a function of the positions of its CG beads (r) consists of the following sum of terms:

$$\begin{aligned} E(r) = & \sum_{\text{bonds}} 0.5 k_{\text{bond}} (r_{ij} - r_{\text{eq}})^2 \\ & + \sum_{\text{angles}} 0.5 k_{\text{angles}} (\theta_{ijk} - \theta_{\text{eq}})^2 \\ & + \sum_{\text{dihedrals}} 0.5\, k^{(n)}_{\text{dihedrals}} \left[1 - \cos\left(n(\phi_{ijkl} - \phi_{\text{eq}})\right)\right] \\ & + \sum_{\text{structured}} (5\varepsilon\, \sigma_{ij}^{12}/r_{ij}^{12} - 6\varepsilon\, \sigma_{ij}^{10}/r_{ij}^{10}) \\ & + \sum_{\text{steric}} 5\, \sigma_{vdw}^{12}/r_{ij}^{12} + \sum_{\text{elec}} 332.08\, q_i q_j \exp(-\kappa\, r_{ij})/D\, r_{ij} \end{aligned}$$

The first three sums represent bonded interactions within the protein. They describe the energy associated with distortions to bond lengths, bond angles and bond dihedral angles, respectively. Each of these terms was implemented as described previously (45). For bonds and angles, purely harmonic potential functions were applied. In the first sum, r_{ij} is the current bond length between beads i and j; r_{eq} is the corresponding equilibrium bond length and k_{bond} is the bond's force constant. For all simulations reported here, k_{bond} was set to 20 kcal/mol/Å^2. In the second sum, θ_{ijk} is the current bond angle between beads i, j and k; θ_{eq} is the corresponding equilibrium bond angle and k_{angle} is the angle's force constant. For all simulations reported here, k_{angle} was set to 10 kcal/mol/Å^2. In the third sum, φ_{ijkl} is the current bond dihedral between beads i, j, k and l where n is 1 or 3; φ_1 and φ_3 are phase angles which define the energy maxima of the dihedral angles and $k_{\text{dihedral}}{}^{(1)}$ and $k_{\text{dihedral}}{}^{(3)}$ are the dihedral force constants. For all simulations reported here, $k_{\text{dihedral}}{}^{(1)}$ was set to 0.5 kcal/mol/Å^2 and $k_{\text{dihedral}}{}^{(3)}$ was set to 0.25 kcal/mol/Å^2.

The final three sums represent non-bonded interactions within the polymerase. They describe the energy associated with the non-bonded interactions of CG beads that are in contact in the structured domains of the initial model, the non-bonded interactions of all other pairs of CG beads, and the electrostatic interactions of all pairs of charged CG beads, respectively. All three of these energy functions depend on the distance, r_{ij}, between CG beads i and j. The energy function used in the fourth sum is a Gō potential of the type that is commonly used in simplified simulations of protein folding (46–48). The $1/r^{10}$ dependence of the attraction means that these interactions are energetically rewarded only at very short distances. The fifth sum is applied to all other pairs of CG beads and is used to prevent steric

overlap so that domains do not move through each other in physically impossible ways. The sixth sum describes the electrostatic interactions between pairs of CG beads, where κ is the Debye–Hückel screening parameter, D is the dielectric constant of the solvent (78.4 for water), and q_i and q_j are the partial charges assigned to the CG beads. κ is related to the square root of the ionic strength, which was set to 200 mM to match the ionic strength used in the SAXS experiments.

All simulations reported here were performed with the CG simulation code *uiowa_BD* (45,48,49). The conformational dynamics of the polymerase were simulated using the Langevin dynamics algorithm developed by Geyer and Winter (50), a simple extension of the Ermak-McCammon algorithm of Brownian dynamics (51). The former algorithm has the important advantage that it allows comparatively long time steps to be employed (50). In all simulations, a time step of 125 fs was used, and a list of pairs of beads engaged in non-bonded interactions was constructed every 100 time steps (every 12.5 ps). Short-range non-bonded interactions (those between CG beads less than 12.5 Å) were updated every time step, medium-ranged interactions (those between CG beads separated by >12.5 Å but <25 Å) were updated every 10 time steps. Debye–Hückel electrostatic interactions were computed using a treecode algorithm that ensures that interactions between all pairs of charged CG beads are computed regardless of their separation distance (52). Finally, to model the diffusive aspects of motion as accurately as possible and to increase the rate at which conformational space is explored, hydrodynamic interactions were computed between all pairs of CG beads at the Rotne–Prager–Yamakawa (RPY) level of theory (53,54). The RPY-approximated diffusion tensor was recomputed every 12.5 ps with a multi-threaded implementation of the Cholesky decomposition used to determine correlated random displacements of the CG beads (55). Each LD simulation was run for 10 μs with snapshots being saved every ns for subsequent analysis. Each simulation, therefore, yielded a total of 10 000 frames. All simulations were repeated three times using different seeds to initialize the random displacements used in the LD algorithm.

Small angle X-ray scattering

SAXS data were collected at Argonne National Laboratory using the Advance Photon Source on beamline 18-ID. Samples were passed over an inline Wyatt Technology WTC-030S5 analytical size exclusion column in buffer C with 5 mM $MgCl_2$. Samples were then passed through a 100 μl quartz capillary flow cell and exposed to the X-ray beamline with 0.5 s exposure every 3 s. The X-ray wavelength was 1.033 Å and data were recorded on a Pilatus3 1M detector at a sample-to-detector distance of 3.5 m. All data were collected at room temperature. Buffer subtraction was performed on all frames followed by evolving factor analysis using BioXTAS RAW to remove any trace contaminant peaks (56). The 50 frames from the elution peak were averaged and used for all subsequent analysis. The elution profiles showing the frames used in averaging and the frames used in buffer subtraction are provided in Supplementary Figure S2. PRIMUS and AUTORG were used to examine the linear Guinier region and calculate the radius of gyration (R_g) (57,58). Pairwise distance distribution plots ($P(r)$ plots) and D_{max} values were calculated using GNOM (58,59).

Analysis of the Langevin dynamics simulations and the SAXS data

Each of the LD simulations resulted in an ensemble of 10 000 structures output as sequential PDB files. CRYSOL was used to generate predicted scattering curves for each of the structures in the ensembles (60). χ^2 values were determined by comparing the predicted scattering curves for the individual structures in the ensembles to the experimental scattering curves. The individual predicted scattering curves from each frame of a given simulation were averaged to obtain a predicted scattering curve of the overall ensemble. χ^2 values were determined by comparing the predicted scattering curves for the full ensembles to the experimental scattering curve. $P(r)$ plots were generated from predicted scattering curves from individual structures in the ensemble using MOLEMAN (61). The $P(r)$ plots for the ensembles were generated by summing all the inter-atomic distances in all 10 000 structures using MATLAB (MathWorks) and generating histograms in SigmaPlot (Systat). The cut off for determine D_{max} was set to include 99.9% of the inter-atomic distances. The χ^2 values, R_g values and D_{max} values obtained from the three replicates for each simulation were nearly identical.

Minimal ensemble searches were carried out five times using the FoXS (Fast X-ray Scattering) server with 100 randomly chosen structures from the LD simulations in order to select the combination of one, two, three, four and five structures that best fit the experimental scattering data (62,63). We then recalculated the predicted scattering curves of the individual structures obtained from the minimal ensemble search using CRYSOL. This was done for the sake of consistency, because FoXS and CRYSOL calculate predicted scattering curves using a slightly different methodology. We then calculated a weighted average scattering curve for each minimal ensemble and compared them to the experimental scattering curve to determine χ^2 values.

RESULTS

Yeast polymerase η is comprised of a polymerase domain (residues 1–512) and a C-terminal region (residues 513–632). There are multiple X-ray crystal structures of the polymerase domain (25,27,64,65). By contrast there is very little structural information for the C-terminal region. This region contains a UBZ motif (residues 549–582). While there are no structures of the UBZ motif from yeast polymerase η, there are NMR and X-ray crystal structures for homologous UBZ motifs (44,66). The C-terminal region also contains a PIP-like motif (residues 621–628). Both the UBZ motif and PIP-like motif have been shown to be important for the function of polymerase η *in vivo* (30). Here we have used a combination of approaches including genetic complementation assays, X-ray crystallography, LD simulations, and SAXS experiments to examine the structure and conformational flexibility of the C-terminal region of polymerase η.

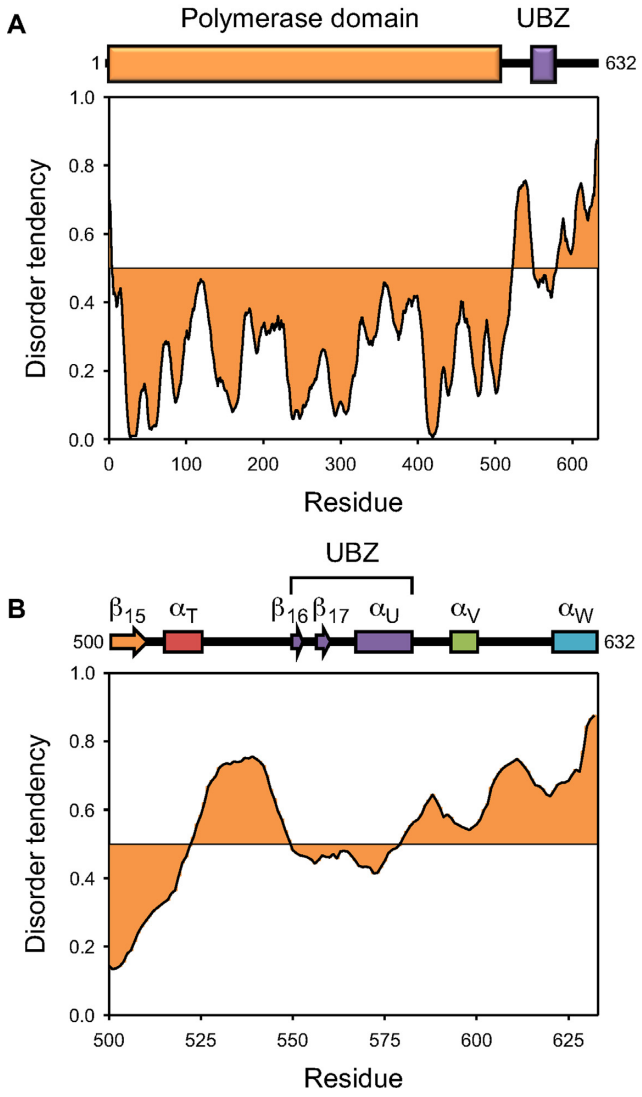

Figure 1. Predicted structured and unstructured regions in DNA polymerase η. (**A**) The disorder tendency obtained from the meta protein disorder prediction server is shown as a function of residue number for full-length polymerase η. The positions of the structured polymerase domain and the structured UBZ motif are indicated. (**B**) The disorder tendency is shown as a function of residue number for the C-terminal region of polymerase η (residues 500–632). The positions of the known and putative secondary structural elements are indicated.

The C-terminal region of DNA polymerase η is predicted to be partially unstructured

To better understand the structure of the C-terminal region of polymerase η, we examined the tendency of the protein to have unstructured/disordered regions using the meta approach (67). The protein was predicted to be structured between residues 4 and 522 and between 550 and 579 (Figure 1A). The former region approximately corresponds to the polymerase domain, and the latter region corresponds to the UBZ motif. Residues 523–549 and residues 580–632 in the C-terminal region of polymerase η are predicted to be unstructured.

Table 1. X-ray data collection and refinement statistics

Data collection	
Space group	$P3_121$
Diffraction source	ALS synchrotron beamline 4.2.2
Wavelength (Å)	1.00003
Cell dimensions	
a, *b*, *c* (Å)	130.68, 130.68, 93.91
α, β, γ (°)	90.00, 90.00, 120.00
Resolution (Å)	56–59-2.80 (2.95–2.80)[a]
R_{sym}	0.096 (1.024)
R_{pim}	0.032 (0.343)
R_{rim}	0.105 (1.126)
I/σ(I)	18.8 (2.2)
Completeness (%)	100.0 (100.0)
Redundancy	10.6 (10.7)
Refinement	
Resolution	43.37–2.80
No. reflections	23171
R-factor (%)	22.08
R-free (%)	24.90
No. atoms	
Protein	7829
Water	23
Ions	2
Ramachandran plot (%)	
Favored	95.0
Allowed	5.0
Disallowed	0.0
Wilson *B*-factor	69.230
Rmsd bonds (Å)	0.003
Rmsd angles (°)	0.619

[a]Highest-resolution shell is shown in parentheses.

We next examined the tendency of the potentially unstructured portions within the C-terminal region to form secondary structural elements using Phrye2 (68). We found three putative α-helices, which in keeping with the conventional nomenclature for polymerase η we named putative α-helix T (residues 514–525), putative α-helix V (residues 593–600), and putative α-helix W (residues 621–632) (Figure 1B). By this convention, the secondary structural elements of the UBZ are designated β-strand 16 (residues 549–552), β-strand 17 (residues 556–560) and α-helix U (residues 567–582). Combining the results from the meta approach and Phyre2, we predict that 63 residues of the 120 residues in the C-terminal region are unstructured.

Putative structured elements within the C-terminal region are required for biological function

To understand the importance of specific portions of the C-terminal region of polymerase η for its biological function, we measured the sensitivity to ultraviolet (UV) radiation of multiple yeast strains. These strains produced either wild-type polymerase η, no polymerase η, or 1 of 23 mutant forms of polymerase η in which ten consecutive residues within the C-terminal region have been substituted by alanine and glycine residues. Various dilutions of yeast cells were spotted on plates and exposed to either no UV radiation or UV radiation at 10, 20 and 50 J/m^2. The UV sensitivities of the yeast strains with mutant forms of polymerase η were scored into three categories: those with low UV sensitivities similar to the strain with wild-type polymerase η, those with high UV sensitivities similar to the strain lacking

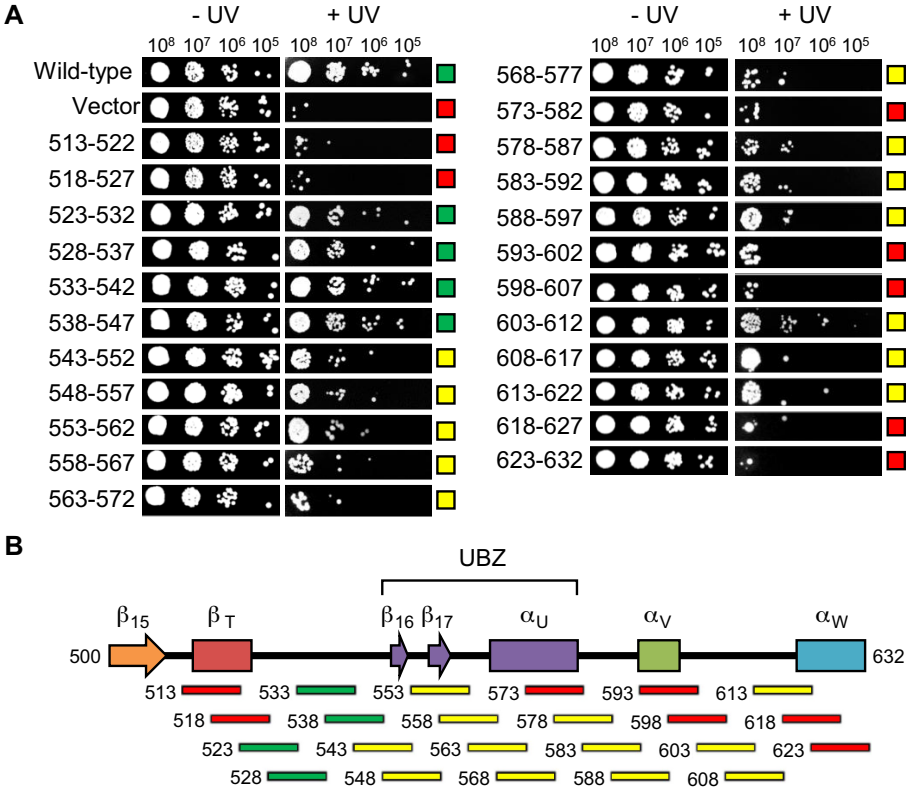

Figure 2. Functional analysis of the C-terminal region of DNA polymerase η. (**A**) Representative images of spotting assays are shown indicating the UV sensitivities of yeast strains producing wild-type polymerase η (positive control), no polymerase η (negative control), or 1 of 23 mutant forms of polymerase η each with ten consecutive substitutions in the C-terminal region. Cells were spotted at various concentrations from 10^8 to 10^5 cells/ml and exposed to no UV radiation or UV radiation at 20 J/m^2. The UV sensitivities of all strains were scored in three categories: normal (green), intermediate (yellow) and high (red). (**B**) A diagram summarizing the levels of UV sensitivity of the 23 mutant forms of polymerase η is shown. The colored lines indicate the UV sensitivities and the positions of the alanine substitutions for each of the mutant proteins. The positions of the known and predicted secondary structural elements are indicated.

polymerase η, and those with intermediate UV sensitivities in between these two strains (Figure 2A).

We found that the strains with high UV sensitivities similar to the one lacking polymerase η produced mutant forms of the protein with substitutions in either α-helix U in the UBZ motif or one of the three putative secondary structural elements predicted by Phyre2: α-helix T, α-helix V and α-helix W (Figure 2B). The PIP-like motif is located on putative α-helix W. The strains with the other mutant forms of polymerase η have either low UV sensitivities similar to the one with wild-type polymerase η or intermediate UV sensitivities. These results show that the amino acid sequence of the portions of the C-terminal region that are essential for the biological function of polymerase η correspond to the known and putative secondary structural elements. The amino acid sequences of the putative unstructured stretches between these elements are not absolutely required for the biological function of polymerase η. Consistent with this, the amino acid sequences of the stretches between the actual and putative secondary structural elements have lower conservation than do the amino acid sequences of these structured elements (Supplementary Figure S3).

X-ray crystal structure of the extended polymerase domain of DNA polymerase η

All of the recombinant proteins used for determining X-ray crystal structures of the polymerase domain of polymerase η are truncated following β-strand 15 (residues 498–509), which is located within the little finger sub-domain (25,27,64,65). The putative α-helix T is located adjacent to this β-strand and may be part of the little finger sub-domain. We produced a form of polymerase η containing residues 1–525 that includes putative α-helix T. We solved the X-ray crystal structure of this longer form of polymerase η to determine whether this putative α helix is visible in the electron density map (Table 1). There was no clear electron density beyond residue 509 indicating that this putative α helix does not have a unique position and is not part of the little finger sub-domain.

This crystal structure had a magnesium ion bound in its active site coordinated to the side chain carboxylate groups of Asp-30 and Glu-156 at distances of 2.2 Å (Figure 3A). Surprisingly, the presence of magnesium in the active site shifted the position of the little finger sub-domain by ∼3.5 Å from its location in the structure of the DNA-free form of the enzyme lacking a bound magnesium ion (PDB ID: 1JIH) (Figure 3B). This movement of the little finger sub-domain narrows the DNA-binding cleft of the

Table 2. SAXS and simulation parameters

Simulation/experiment	Starting model	χ^2	R_g (Å)	D_{max} (Å)
Polymerase η catalytic domain				
Experimental SAXS data	N/A	N/A	30.5	111
LD simulation	1JIH	7.57	28.9	98.5
LD simulation	5VTP	4.63	29.9	104
Full-length polymerase η				
Experiment SAXS data	N/A	N/A	37.6	165
LD simulation	PD	6.55	39.6	166
LD simulation	PD-U	6.78	39.8	168
LD simulation	PD-HUHH	3.34	38.2	164

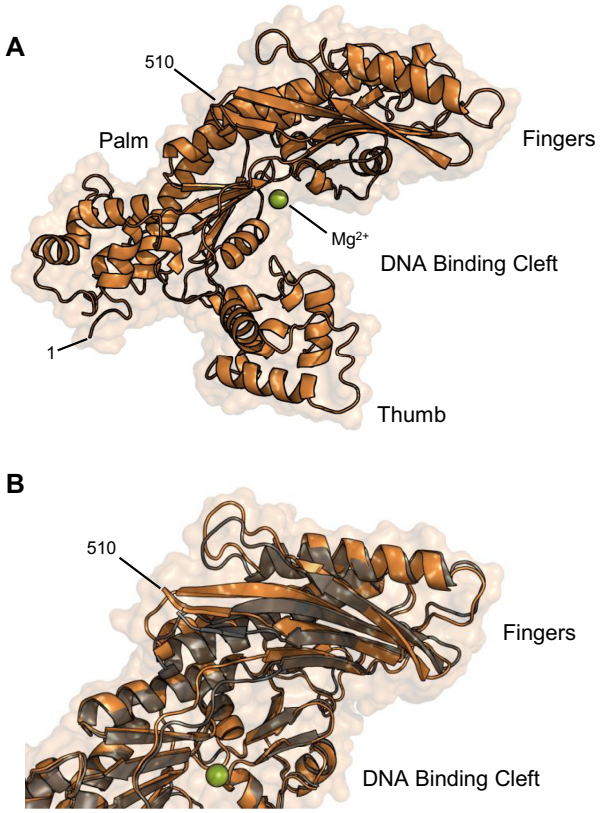

Figure 3. Structure of the polymerase domain of DNA polymerase η. (**A**) The X-ray crystal structure of the polymerase domain of DNA polymerase η with a bound magnesium ion (PDB ID: 5VTP) is shown. The fingers, palm, and thumb sub-domains are indicated. (**B**) An overlay of the structures of the polymerase domain with a bound magnesium ion (orange) and the polymerase domain lacking a bound magnesium ion (PDB ID: 1JIH) (gray) is shown.

polymerase domain. The position of the little finger sub-domain in the magnesium-bound structure is in between its positions in the more open magnesium-free structure and in the more closed DNA-bound structure. This is important because this magnesium-bound structure agreed better with the small-angle X-ray scattering data (see below).

Langevin dynamics simulations of the polymerase domain of DNA polymerase η

We carried out a series of LD simulations of full-length polymerase η to model the conformational flexibility of the C-terminal region. Before doing these simulations, however, we needed to ensure that we could accurately model the polymerase domain alone. Therefore, we carried out two LD simulations of the polymerase domain of polymerase η. One of the simulations used as a starting model the previously determined X-ray crystal structure of the polymerase domain lacking a bound magnesium ion (PDB ID: 1JIH) (25). The other used as a starting model the X-ray crystal structure of the polymerase domain determined here with a magnesium ion bound (PDB ID: 5VTP). Both simulations were coarse-grained LD simulations in which each amino acid residue was replaced by one to four CG beads. These simulations were carried out in triplicate for 10 μs of simulation time (Figure 4). In both LD simulations, the overall structures of the polymerase domains were unchanged with only very small movements of the sub-domains relative to one another effectively narrowing and broadening the DNA binding cleft by ∼2 Å.

Small-angle X-ray scattering of the polymerase domain of DNA polymerase η

To experimentally validate the LD simulations of the polymerase domain of polymerase η, we carried out SAXS experiments with the polymerase domain (Figure 5A and Supplemental Table ST1). The Guinier plot was linear, indicating that the polymerase domain did not aggregate. The R_g based on Guinier analysis was equal to 30.5 Å (Table 2). The pair-wise distance-distribution function shows a little tailing at high values of r and yielded a maximum distance (D_{max}) equal to 111 Å (Figure 5B, Table 2). The Porod coefficient was equal to 3.9, which is consistent with a compact structure, and the Porod volume was ∼134 000 Å^3.

We compared the experimental scattering curve to the predicted scattering curves for both starting models and for the entire ensembles for both LD simulations (Figure 5A, Table 2). The best fit was obtained using the full ensemble from the LD simulation with the model derived from the X-ray structure with the bound magnesium ion. The χ^2 values obtained from the three replicates for this simulation were nearly identical: 4.63, 4.63 and 4.64 (Supplementary Figure S4). For the sake of comparison, we found that using the coarse-grained starting model derived from this X-ray crystal structure yielded a poorer fit ($\chi^2 = 8.15$) than using the entire ensemble. Overall, the R_g and D_{max} values calculated from the full ensemble from this simulation, which were equal to 29.9 and 104 Å respectively, agreed with the experimentally determined values. Moreover, this full en-

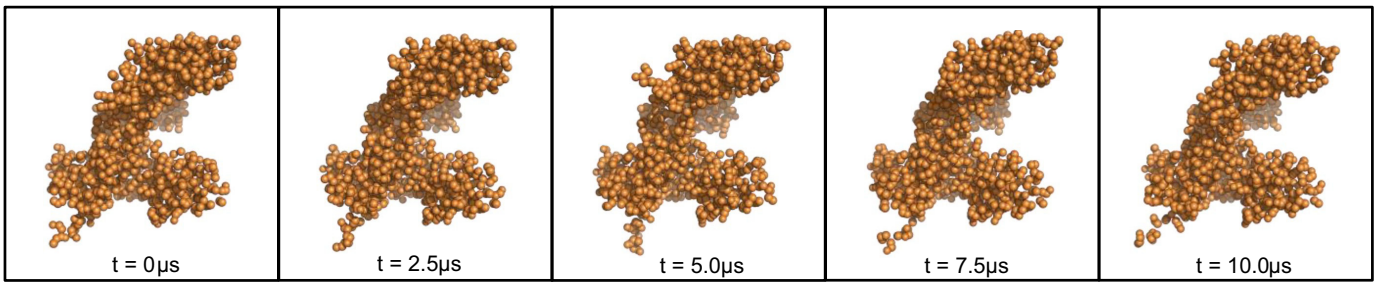

Figure 4. Langevin dynamics simulations of the polymerase η polymerase domain. (**A**) Individual structures of the polymerase domain obtained from the coarse-grained LD simulation using the structure of the polymerase domain with a bound magnesium ion (PDB ID: 5VTP) as a starting model are shown.

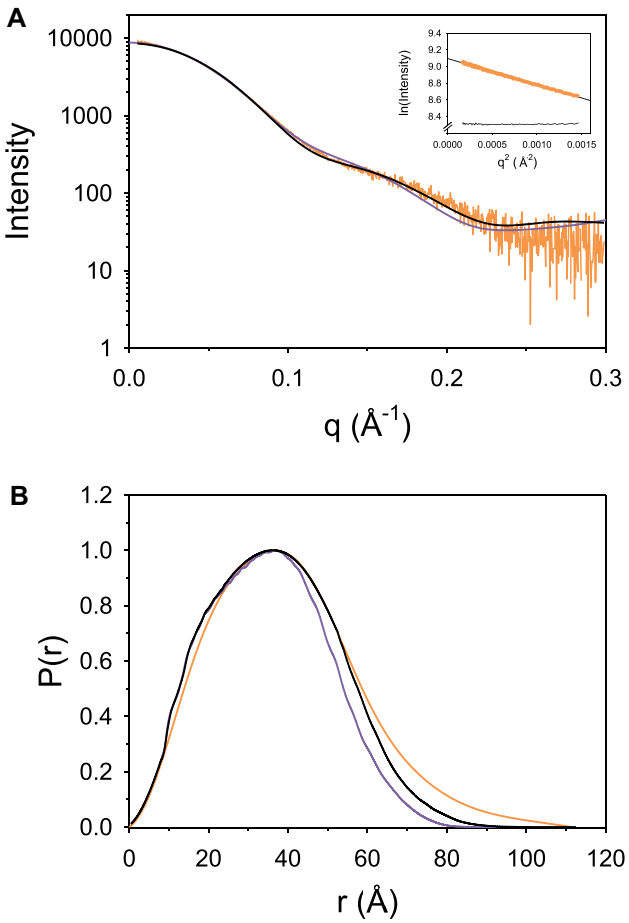

Figure 5. SAXS analysis of the polymerase η polymerase domain. (**A**) Experimental scattering curve of the polymerase domain (*orange*) and predicted scattering curves from the ensembles derived from the LD simulations using the structures of the polymerase domain with the bound magnesium ion (PDB ID: 5VTP, *black*) and the polymerase domain lacking a bound magnesium ion (PDB ID: 1JIH, *purple*) as starting models are shown. The Guinier plot of the polymerase domain and the residuals are shown as an inset. (**B**) Experimental $P(r)$ curves of the polymerase domain (*orange*) and predicted scattering curves from the entire ensemble derived from both simulations (*black* and *purple*) are shown.

semble reproduced well the experimental $P(r)$ function (Figure 5B).

Langevin dynamics simulations of full-length DNA polymerase η

The agreement between the experimental scattering curves and the predicted scattering curves from the entire coarse-grained LD simulations for the polymerase domain of polymerase η gave us confidence that we could accurately model the polymerase domain alone. Thus, to understand the structure and conformational flexibility of the C-terminal domain, we carried out coarse-grained LD simulations of full-length polymerase η. We used the X-ray crystal structure of the polymerase domain with a bound magnesium ion (PDB ID: 5VTP) as the starting model for this domain. We then constructed three builds of the C-terminal region of polymerase eta (Figure 6A): one in which there were no structural elements (referred to as **PD**), one in which only the UBZ motif was structured (referred to as **PD-U**), and one in which the UBZ motif and the three putative α helices were structured (referred to as **PD-HUHH**). For each of these three models of full-length polymerase eta, we carried out coarse-grained LD simulations in triplicate for 10 μs of simulation time (Figure 6B). In all simulations, the C-terminal region displayed a similarly high degree of flexibility as this region adopted many conformations. These conformations are described in more detail below.

Small-angle X-ray scattering of full-length DNA polymerase η

To experimentally validate the LD simulations of full-length polymerase η, we carried out SAXS experiments with the full-length protein (Figure 7A and Supplemental Table ST1). The linear Guinier plot indicated that the full-length protein did not aggregate, and the R_g based on this analysis was equal to 37.6 Å. The pair-wise distance-distribution function showed much more tailing at high values of r than observed for the polymerase domain and yielded a D_{max} equal to 165 Å (Figure 7B). The Porod coefficient was equal to 3.7, which is consistent with a somewhat less compact structure than that of the polymerase domain alone, and the Porod volume was ∼204 000 Å^3.

We compared the experimental scattering curve to predicted scattering curves for the entire ensembles derived from all three LD simulations (Figure 7A). The best fit was obtained using the full ensemble derived from the simulations with the model containing the UBZ motif and the three putative α helices (PD-HUHH). The χ^2 values obtained from the three replicates for this simulation were

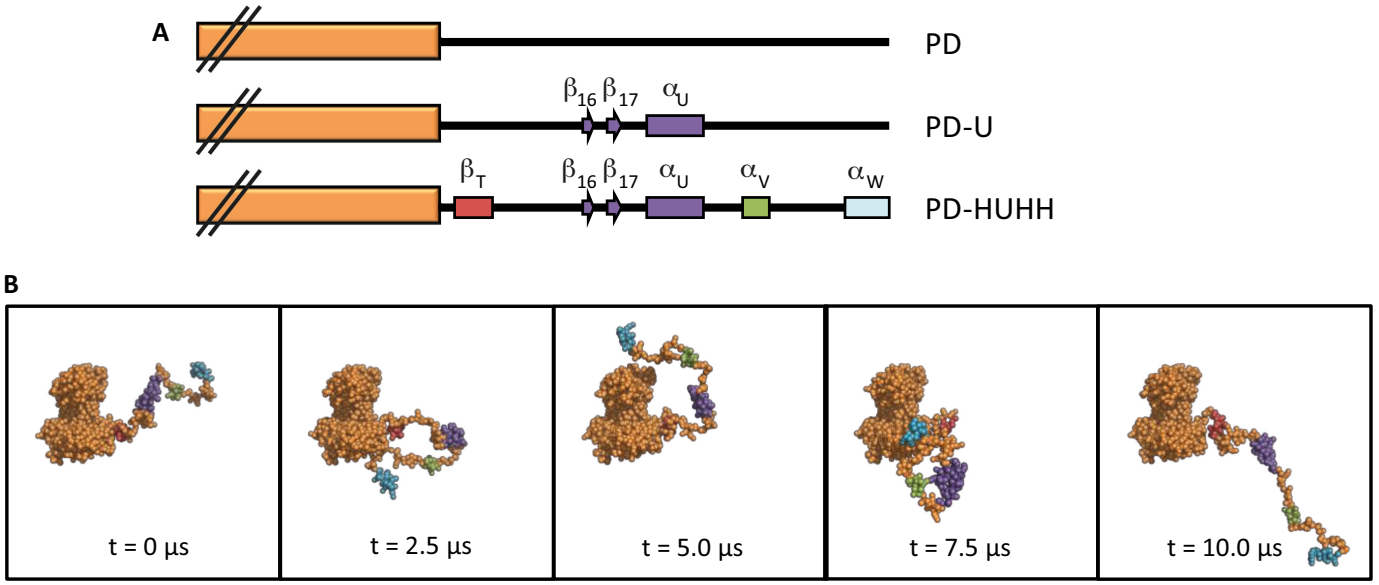

Figure 6. Langevin dynamics simulations of full-length polymerase η. (**A**) Diagrams of the three builds of polymerase η (PD, PD-U and PD-HUHH) are shown. (**B**) Individual structures of full-length polymerase η obtained from the coarse-grained LD simulation using the PD-HUHH build as a starting model are shown.

nearly identical: 3.34, 3.65 and 3.85 (Supplementary Figure S5). The R_g and D_{max} values calculated from the full ensemble for this simulation, which were equal to 38.2 Å and 164 Å respectively, agreed with the experimentally determined values (Table 2). Moreover, this full ensemble reproduced well the experimental pair-wise distance distribution function (Figure 7B).

Conformational flexibility of the C-terminal region of DNA polymerase η

The full ensemble derived from the LD simulation of full-length polymerase η containing the UBZ motif and the three putative α helices (PD-HUHH) best fit the experimental SAXS data. We therefore examined the conformational flexibility of this ensemble in more detail. We graphed the R_g as a function of time (Figure 8A) and as a histogram showing the relative frequencies of each R_g value in the ensemble (Figure 8B). The R_g values for each snapshot of the ensemble ranged from 30.5 to 55.9 Å. We also graphed the D_{max} as a function of time (Figure 8C) and as a histogram showing the relative frequencies of each D_{max} value in the ensemble (Figure 8D). The D_{max} values for each snapshot of the ensemble ranged from 87.0 to 234 Å.

The plots of R_g as a function of time and D_{max} as a function of time show that the conformation of full-length polymerase η changes significantly during the simulation. There are portions of the LD simulation in which the R_g and D_{max} parameters are near their minimal values. These represent relatively compact states in which the C-terminal region remains largely unstructured, but is occupying space close to the polymerase domain (Figure 9A). These compact states account for ∼5% of the structures in the ensemble. There are portions of the LD simulations in which the R_g and D_{max} values are near their average values. These represent states in which the C-terminal region is partially extended (Figure 9B), and they account for approximately 90% of the ensemble. There are other portions of the LD simulations in which the R_g and D_{max} are near their maximal values. These represent states in which the C-terminal region is relatively extended (Figure 9C). These extended states also account for ∼5% of the structures in the ensemble.

Despite the fact that the compact states and fully extended states represent only a small fraction of the ensemble, they sometimes persisted for up to 500 ns of simulation time. We examined the individual snapshots from this LD simulation to determine whether there was a structural basis for the persistence of these states during the LD simulation. We found that in the compact states, the UBZ motif is often interacting with little finger sub-domain of the polymerase domain (Figure 10). Interactions between the UBZ and other sub-domains of the polymerase domain were not observed. There appears to be no preferred binding site for the UBZ on the little finger sub-domain. Both of these surfaces, however, have multiple charged residues conserved across eukaryotes that can mediate transient interactions between the UBZ motif and this part of the polymerase domain. In the extended states, the UBZ is not interacting with the little finger sub-domain.

DISCUSSION

All four eukaryotic members of the Y-family of non-classical polymerases are predicted to have C-terminal regions containing a significant amount of intrinsic disorder (35). To our knowledge, this has not been experimentally demonstrated and the range of conformational states that these regions can adopt has not been explored. Thus, we set out to determine whether the C-terminal region of polymerase η is partially unstructured and to examine its conformational flexibility. The principle motivation for this work was to address biological questions about the struc-

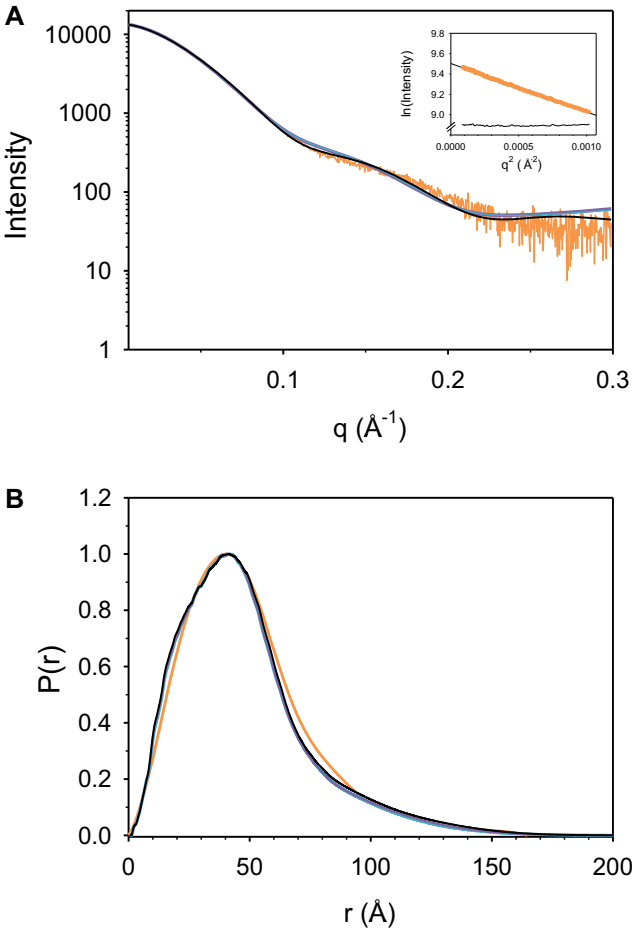

Figure 7. SAXS analysis of full-length polymerase η. (**A**) Experimental scattering curve of full-length polymerase domain (*orange*) and predicted scattering curves from the ensembles derived from the LD simulations using the PD (*purple*), the PD-U (*blue*), and the PD-HUHH (*black*) builds as a starting model are shown. The Guinier plot of full-length polymerase η and the residuals are shown as an inset. (**B**) Experimental *P*(*r*) curves of the polymerase domain (*orange*) and predicted scattering curves from the ensembles derived from all three simulations (*purple*, *blue*, and *black*) are shown.

ture of non-classical polymerases in order to better understand how their structures explain the way in which non-classical polymerases are selected to carry out translesion synthesis. Before discussing the biological implications of our finding, however, we will discuss the novel combination of computational and experimental approaches that we used to study the properties of the C-terminal region of polymerase η.

We combined LD simulations of full-length polymerase η and SAXS measurements to generate a structural model of polymerase η that matched the experimental data. This model is an ensemble of 10 000 structures, each one representing a snapshot from a LD simulation sampled every 1 ns. The simulation that best fit the experimental data was one in which the polymerase domain was structured, the UBZ motif in the C-terminal region was structured, and the amino acid residues that comprised putative α-helix T, α-helix V, and α-helix W in the C-terminal region were structured. All other portions of the C-terminal region were un-

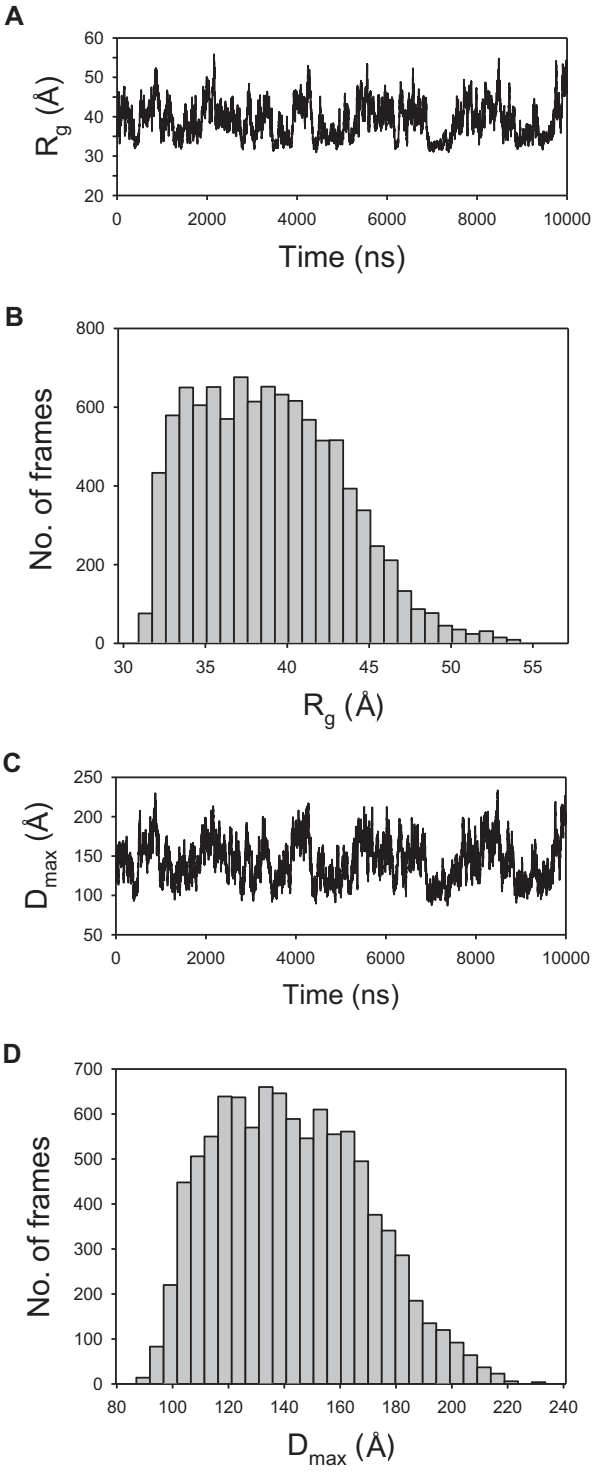

Figure 8. Conformational flexibility of full-length polymerase η. (**A**) The R_g values of each structure of the ensemble derived from the coarse-grained LD simulations of full-length polymerase η using the PD-HUHH build are shown as a function of time. (**B**) A histogram of the number of structures in the ensemble with a given R_g value is shown. (**C**) The D_{max} values of each structure of the ensemble derived from the coarse-grained LD simulations of full-length polymerase η using the PD-HUHH build are shown as a function of time. (**B**) A histogram of the number of structures in the ensemble with a given D_{max} value is shown.

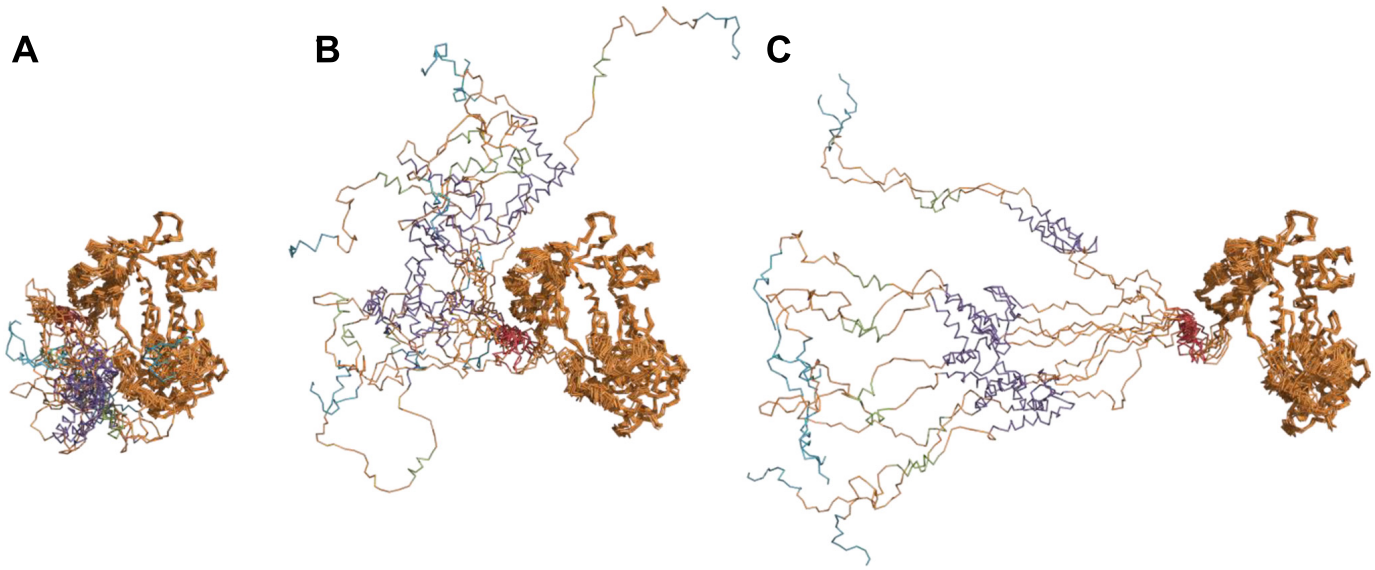

Figure 9. Compact, partly extended, and extended conformations of full-length polymerase η. (**A**) Ten structures of full-length polymerase η derived from the coarse-grained LD simulations using the PD-HUHH build all with R_g and D_{max} values near their minimums are shown. The polymerase domains of these structures are overlaid, and the UBZ motif is shown in *purple*. (**B**) Ten structures of full-length polymerase η all with intermediate R_g and D_{max} values are shown. The polymerase domains of these structures are overlaid, and the UBZ motif is shown in *purple*. (**C**) Ten structures of full-length polymerase η all with R_g and D_{max} values near their maximums are shown. The polymerase domains of these structures are overlaid, and the UBZ motif is shown in *purple*.

structured. This ensemble agreed well with the experimental SAXS data, as well as the functional *in vivo* data, providing confidence that these secondary structural elements are indeed present in polymerase η and that the conformational flexibility observed here accurately reflects the behavior of polymerase η in solution.

The combination of SAXS measurements and either molecular dynamics (MD) simulations, Brownian dynamics (BD) simulations or LD simulations has been a powerful approach to studying the conformational flexibility of intrinsically disordered regions of proteins (69). Typically, this has been done very differently than the strategy applied here. Normally, MD, BD or LD simulations are used to generate a large ensemble of possible structures of the intrinsically disordered regions of a protein. This large ensemble is then used to carry out a minimal ensemble search. One searches through the large ensemble for a small number of structures, usually one to five, that when combined in certain proportions best fit the data. The result is often three or four conformational states with accompanying weighting factors that when combined match the experimental data. In this strategy, one must seek the minimum number of states that fits the SAXS data because there are two degrees of freedom for each state (one for its structure and one for its weighting factor). Using more than the minimum number of states needed to fit the data increases the degrees of freedom leading to less meaningful models.

In order to compare the approach that we used in this study with the more traditional approach described above, we carried out an analysis of our SAXS data using a minimal ensemble search. We randomly selected five different sets of 100 structures derived from the best LD simulation and carried out five different minimal ensemble searches. The χ^2 value from the best minimal ensemble search was equal to 6.16 and occurred when the minimal ensemble contained three states (Supplemental Table ST2). These three structures and their corresponding weighting factors are shown in Supplementary Figure S6. These structures all had intermediate R_g and D_{max} values.

In this study, we combined LD simulations and SAXS measurements in, to our knowledge, a novel way. We started by creating several of builds of full-length polymerase η and ran coarse-grained LD simulations of these builds. We then directly compared the full ensembles derived from these simulations to the experimental SAXS data without any weighting factors or curve fitting. We assessed agreement between the model and the experimental data by calculating χ^2 values, and the best ensemble yielded a χ^2 equal to 3.34. Thus, the quality of agreement of the complete ensemble to the experimental data without the reliance on curve fitting was better than the minimal ensemble search that relies on curve fitting.

We believe that the approach for combining simulations and scattering data used in this study is preferable to the traditional minimal ensemble search method for several reasons and ought to be used where possible. First, this approach does not rely on curve fitting to obtain weighting factors. In fact, the tendency of similar states to be represented in the full ensemble to various extents is determined by the force field parameters used in the simulation, not by curve fitting. Second, a full ensemble containing 10 000 structures each related to one another by a time step in a LD simulation is a far more realistic model than one obtained from a minimal ensemble search. The few structures that comprise the minimal ensemble are each kinetically unstable and would rapidly change conformation in solution

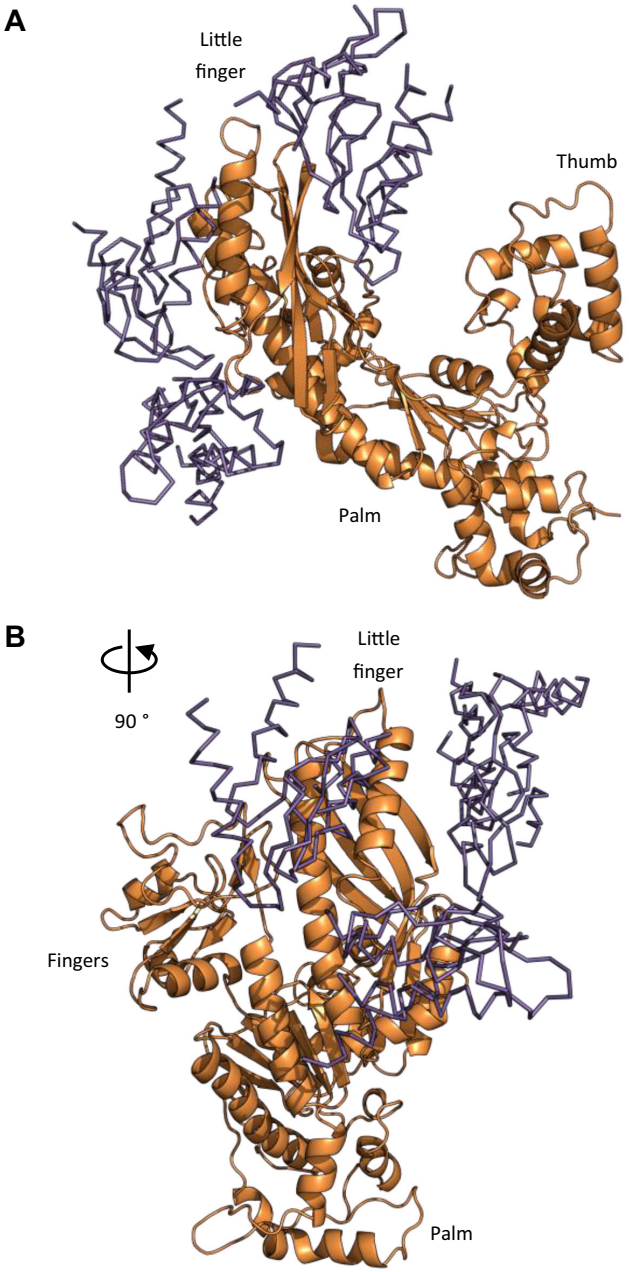

Figure 10. Interactions between the UBZ motif and the little finger subdomain. (**A**) The positions of the UBZ motif (*purple*) relative to the polymerase η polymerase domain (*orange*) are shown for ten structures with near minimum R_g and D_{max} values derived from the coarse-grained LD simulation using the PD-HUHH build. (**B**) These same ten structures shown from a side view.

and in simulations to a state not represented in the minimal ensemble. The advantage of the full ensemble is that when a similar change in conformation occurs, each structure changes into another structure represented in the full ensemble.

Implications for translesion synthesis

The conformational flexibility of the C-terminal region of polymerase η that we demonstrated here is likely important for facilitating polymerase sampling and polymerase switching during translesion synthesis. Cells possess multiple non-classical polymerases, each capable of synthesizing DNA opposite one or more specific types of DNA lesions—its cognate lesions (5,9–11,70). A key step in translesion synthesis is the selection of the most appropriate polymerase to bypass the lesion. This is important, because translesion synthesis is only efficient and accurate when a non-classical polymerase bypasses a cognate lesion. If an inappropriate non-classical polymerase is used, an incorrect nucleotide is likely to be incorporated. This misincorporation will result in a mutation once the damage has been repaired. It is unclear exactly how the most appropriate non-classical polymerase is selected during translesion synthesis.

The most straightforward model of non-classical polymerase selection is the kinetic sampling model (11). According to this model, multiple non-classical polymerases directly compete with one another for binding the DNA substrate. When a polymerase binds the DNA substrate, one of two things can occur: it can either dissociate with a first-order rate constant k_{off} or it can incorporate a nucleotide with a first-order rate constant k_{pol}. The relative magnitudes of the k_{off} and k_{pol} rate constants determine the frequency of dissociation and the frequency of nucleotide incorporation. If a polymerase binds to DNA containing a non-cognate lesion, the k_{off} rate constant will be substantially greater than the k_{pol} rate constant. This will favor dissociation rather than nucleotide incorporation. If a polymerase binds to DNA containing a cognate lesion, the k_{pol} will be greater than the k_{off}. This will favor nucleotide incorporation rather than dissociation. This way a competition among multiple non-classical polymerases should maximize the likelihood that the most appropriate one will bypass the DNA lesion, no matter what type of lesion is present.

A requirement for the kinetic selection model is that multiple non-classical polymerases compete for the DNA substrate. One way this can occur is by the formation of large, multi-protein bypass complexes containing multiple non-classical polymerases (15–17). The formation of such complexes would increase the local concentrations of these polymerases and increase the rates at which they bind the DNA substrate. Moreover, the competition among the non-classical polymerases would be facilitated by the polymerases having intrinsically disordered regions that act as flexible tethers. The polymerase domains of these enzymes can take turns binding and releasing the DNA template without dissociating from the complex and without altering any protein-protein interactions within the complex. This would facilitate the sampling of non-classical polymerases to ensure that the most appropriate one is selected to bypass the lesion and thereby maximize the efficiency and accuracy of translesion synthesis.

AVAILABILITY

PDB ID: 5VTP.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Christine Kondratick, Melissa Gildenberg, Brittany Ripley, Maria Spies, Lokesh Gakhar, Bret Freudenthal and Lynne Dieckman for discussions. We thank Casey Andrews, Lokesh Gakhar, Melissa Gildenberg and Nick Schnicker for technical assistance. We thank Jonathan Hogg for his contributions to parallelizing the Cholesky factorization for multicore simulations. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

FUNDING

National Institute of General Medicine [GM081433 and GM087290 to M.T.W., A.H.E.]. Funding for open access charge: National Institute of General Medicine [GM081433].

Conflict of interest statement. None declared.

REFERENCES

1. Prakash,S. and Prakash,L. (2002) Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. *Genes Dev.*, **16**, 1872–1883.
2. Prakash,S., Johnson,R.E. and Prakash,L. (2005) Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu. Rev. Biochem.*, **74**, 317–353.
3. Lehmann,A.R. (2005) Replication of damaged DNA by translesion synthesis in human cells. *FEBS Lett.*, **579**, 873–876.
4. Guo,C., Kosarek-Stancel,J.N., Tang,T.S. and Friedberg,E.C. (2009) Y-family DNA polymerases in mammalian cells. *Cell. Mol. Life Sci.: CMLS*, **66**, 2363–2381.
5. Waters,L.S., Minesinger,B.K., Wiltrout,M.E., D'Souza,S., Woodruff,R.V. and Walker,G.C. (2009) Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol. Mol. Biol. Rev.: MMBR*, **73**, 134–154.
6. Washington,M.T., Carlson,K.D., Freudenthal,B.D. and Pryor,J.M. (2010) Variations on a theme: eukaryotic Y-family DNA polymerases. *Biochim. Biophys. acta*, **1804**, 1113–1123.
7. Sale,J.E., Lehmann,A.R. and Woodgate,R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.*, **13**, 141–152.
8. Zhao,L. and Washington,M.T. (2017) Translesion synthesis: insights into the selection and switching of DNA polymerases. *Genes*, **8**, E24.
9. Pryor,J.M., Dieckman,L.M., Boehm,E.M. and Washington,M.T. (2014) Eukaryotic Y-family polymerases: a biochemical and structural perspective. *Nucleic Acids Mol. Biol.*, **30**, 85–108.
10. Lehmann,A.R., Niimi,A., Ogi,T., Brown,S., Sabbioneda,S., Wing,J.F., Kannouche,P.L. and Green,C.M. (2007) Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair*, **6**, 891–899.
11. Powers,K.T. and Washington,M.T. (2017) Analyzing the catalytic activities and interactions of eukaryotic translesion synthesis polymerases. *Methods Enzymol.*, **592**, 329–356.
12. Hoege,C., Pfander,B., Moldovan,G.L., Pyrowolakis,G. and Jentsch,S. (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature*, **419**, 135–141.
13. Stelter,P. and Ulrich,H.D. (2003) Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. *Nature*, **425**, 188–191.
14. Bienko,M., Green,C.M., Crosetto,N., Rudolf,F., Zapart,G., Coull,B., Kannouche,P., Wider,G., Peter,M., Lehmann,A.R. *et al.* (2005) Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science*, **310**, 1821–1824.
15. Wojtaszek,J., Liu,J., D'Souza,S., Wang,S., Xue,Y., Walker,G.C. and Zhou,P. (2012) Multifaceted recognition of vertebrate Rev1 by translesion polymerases zeta and kappa. *J. Biol. Chem.*, **287**, 26400–26408.
16. Wojtaszek,J., Lee,C.J., D'Souza,S., Minesinger,B., Kim,H., D'Andrea,A.D., Walker,G.C. and Zhou,P. (2012) Structural basis of Rev1-mediated assembly of a quaternary vertebrate translesion polymerase complex consisting of Rev1, heterodimeric polymerase (Pol) zeta, and Pol kappa. *J. Biol. Chem.*, **287**, 33836–33846.
17. Boehm,E.M., Spies,M. and Washington,M.T. (2016) PCNA tool belts and polymerase bridges form during translesion synthesis. *Nucleic Acids Res.*, **44**, 8250–8260.
18. Johnson,R.E., Prakash,S. and Prakash,L. (1999) Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Poleta. *Science*, **283**, 1001–1004.
19. Haracska,L., Yu,S.L., Johnson,R.E., Prakash,L. and Prakash,S. (2000) Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase eta. *Nat. Genet.*, **25**, 458–461.
20. Washington,M.T., Johnson,R.E., Prakash,S. and Prakash,L. (2000) Accuracy of thymine-thymine dimer bypass by Saccharomyces cerevisiae DNA polymerase eta. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 3094–3099.
21. McDonald,J.P., Levine,A.S. and Woodgate,R. (1997) The Saccharomyces cerevisiae RAD30 gene, a homologue of Escherichia coli dinB and umuC, is DNA damage inducible and functions in a novel error-free postreplication repair mechanism. *Genetics*, **147**, 1557–1568.
22. Johnson,R.E., Prakash,S. and Prakash,L. (1999) Requirement of DNA polymerase activity of yeast Rad30 protein for its biological function. *J. Biol. Chem.*, **274**, 15975–15977.
23. Johnson,R.E., Kondratick,C.M., Prakash,S. and Prakash,L. (1999) hRAD30 mutations in the variant form of xeroderma pigmentosum. *Science*, **285**, 263–265.
24. Masutani,C., Kusumoto,R., Yamada,A., Dohmae,N., Yokoi,M., Yuasa,M., Araki,M., Iwai,S., Takio,K. and Hanaoka,F. (1999) The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. *Nature*, **399**, 700–704.
25. Trincao,J., Johnson,R.E., Escalante,C.R., Prakash,S., Prakash,L. and Aggarwal,A.K. (2001) Structure of the catalytic core of S. cerevisiae DNA polymerase eta: implications for translesion DNA synthesis. *Mol. Cell*, **8**, 417–426.
26. Silverstein,T.D., Johnson,R.E., Jain,R., Prakash,L., Prakash,S. and Aggarwal,A.K. (2010) Structural basis for the suppression of skin cancers by DNA polymerase eta. *Nature*, **465**, 1039–1043.
27. Silverstein,T.D., Jain,R., Johnson,R.E., Prakash,L., Prakash,S. and Aggarwal,A.K. (2010) Structural basis for error-free replication of oxidatively damaged DNA by yeast DNA polymerase eta. *Structure*, **18**, 1463–1470.
28. Washington,M.T., Prakash,L. and Prakash,S. (2003) Mechanism of nucleotide incorporation opposite a thymine-thymine dimer by yeast DNA polymerase eta. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 12093–12098.
29. Carlson,K.D. and Washington,M.T. (2005) Mechanism of efficient and accurate nucleotide incorporation opposite 7,8-dihydro-8-oxoguanine by Saccharomyces cerevisiae DNA polymerase eta. *Mol. Cell. Biol.*, **25**, 2169–2176.
30. Kondratick,C.M., Washington,M.T., Prakash,S. and Prakash,L. (2001) Acidic residues critical for the activity and biological function of yeast DNA polymerase eta. *Mol. Cell. Biol.*, **21**, 2018–2025.
31. Haracska,L., Kondratick,C.M., Unk,I., Prakash,S. and Prakash,L. (2001) Interaction with PCNA is essential for yeast DNA polymerase eta function. *Mol. Cell*, **8**, 407–415.
32. Boehm,E.M., Powers,K.T., Kondratick,C.M., Spies,M., Houtman,J.C. and Washington,M.T. (2016) The proliferating cell nuclear antigen (PCNA)-interacting protein (PIP) motif of DNA polymerase eta mediates Its interaction with the C-terminal domain of Rev1. *J. Biol. Chem.*, **291**, 8735–8744.
33. Boehm,E.M. and Washington,M.T. (2016) R.I.P. to the PIP: PCNA-binding motif no longer considered specific: PIP motifs and other related sequences are not distinct entities and can bind multiple proteins involved in genome maintenance. *BioEssays*, **38**, 1117–1122.
34. Lau,W.C., Li,Y., Zhang,Q. and Huen,M.S. (2015) Molecular architecture of the Ub-PCNA/Pol eta complex bound to DNA. *Sci. Rep.*, **5**, 15759.
35. Ohmori,H., Hanafusa,T., Ohashi,E. and Vaziri,C. (2009) Separate roles of structured and unstructured regions of Y-family DNA polymerases. *Adv. Protein Chem. Struct. Biol.*, **78**, 99–146.

36. Gietz,R.D. and Sugino,A. (1988) New yeast-Escherichia coli shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. *Gene*, **74**, 527–534.
37. Kabsch,W. (2010) Xds. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 125–132.
38. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D, Biol. Crystallogr.*, **50**, 760–763.
39. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D, Biol. Crystallogr.*, **60**, 2126–2132.
40. Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K. and Terwilliger,T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D, Biol. Crystallogr.*, **58**, 1948–1954.
41. Fenn,T.D. and Schnieders,M.J. (2011) Polarizable atomic multipole X-ray refinement: weighting schemes for macromolecular diffraction. *Acta Crystallogr. D, Biol. Crystallogr.*, **67**, 957–965.
42. Xiang,Z., Soto,C.S. and Honig,B. (2002) Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 7432–7437.
43. Biasini,M., Bienert,S., Waterhouse,A., Arnold,K., Studer,G., Schmidt,T., Kiefer,F., Cassarino,T.G., Bertoni,M., Bordoli,L. *et al.* (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, **42**, W252–W258.
44. Bomar,M.G., Pai,M.T., Tzeng,S.R., Li,S.S. and Zhou,P. (2007) Structure of the ubiquitin-binding zinc finger domain of human DNA Y-polymerase eta. *EMBO Rep.*, **8**, 247–251.
45. Frembgen-Kesner,T. and Elcock,A.H. (2009) Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *J. Chem. Theory Computat.*, **5**, 242–256.
46. Hills,R.D. Jr and Brooks,C.L. 3rd (2009) Insights from coarse-grained Go models for protein folding and dynamics. *Int. J. Mol. Sci.*, **10**, 889–905.
47. Go,N. (1983) Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.*, **12**, 183–210.
48. Elcock,A.H. (2006) Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput. Biol.*, **2**, e98.
49. Frembgen-Kesner,T. and Elcock,A.H. (2010) Absolute protein-protein association rate constants from flexible, coarse-grained Brownian dynamics simulations: the role of intermolecular hydrodynamic interactions in barnase-barstar association. *Biophys. J.*, **99**, L75–L77.
50. Winter,U. and Geyer,T. (2009) Coarse grained simulations of a small peptide: effects of finite damping and hydrodynamic interactions. *J. Chem. Phys.*, **131**, 104102.
51. Ermak,D.L. and Mccammon,J.A. (1978) Brownian Dynamics with Hydrodynamic Interactions. *J. Chem. Phys.*, **69**, 1352–1360.
52. Li,P.J., Johnston,H. and Krasny,R. (2009) A Cartesian treecode for screened coulomb interactions. *J. Comput. Phys.*, **228**, 3858–3868.
53. Rotne,J. and Prager,S. (1969) Variational treatment of hydrodynamic interaction in polymers. *J. Chem. Phys.*, **50**, 4831.
54. Yamakawa,H. (1970) Transport properties of polymer chains in dilute solution - hydrodynamic interaction. *J. Chem. Phys.*, **53**, 436.
55. Hogg,J.D. (2008) *Technical Report RF-RAL-2008-029*. Rutherford-Appleton Laboratory, UK VMD.
56. Nielsen,S.S., Toft,K.N., Snakenborg,D., Jeppesen,M.G., Jacobsen,J.K., Vestergaard,B., Kutter,J.P. and Arleth,L. (2009) BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. *J. Appl. Crystallogr.*, **42**, 959–964.
57. Konarev,P.V., Volkov,V.V., Sokolova,A.V., Koch,M.H.J. and Svergun,D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **36**, 1277–1282.
58. Petoukhov,M.V., Konarev,P.V., Kikhney,A.G. and Svergun,D.I. (2007) ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.*, **40**, S223–S228.
59. Svergun,D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, **25**, 495–503.
60. Svergun,D., Barberato,C. and Koch,M.H.J. (1995) CRYSOL - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, **28**, 768–773.
61. Kleywegt,G.J. (1997) Validation of protein models from C-alpha coordinates alone. *J. Mol. Biol.*, **273**, 371–376.
62. Schneidman-Duhovny,D., Hammel,M., Tainer,J.A. and Sali,A. (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.*, **105**, 962–974.
63. Pelikan,M., Hura,G.L. and Hammel,M. (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.*, **28**, 174–189.
64. Alt,A., Lammens,K., Chiocchini,C., Lammens,A., Pieck,J.C., Kuch,D., Hopfner,K.P. and Carell,T. (2007) Bypass of DNA lesions generated during anticancer treatment with cisplatin by DNA polymerase eta. *Science*, **318**, 967–970.
65. Schorr,S., Schneider,S., Lammens,K., Hopfner,K.P. and Carell,T. (2010) Mechanism of replication blocking and bypass of Y-family polymerase {eta} by bulky acetylaminofluorene DNA adducts. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 20720–20725.
66. Suzuki,N., Rohaim,A., Kato,R., Dikic,I., Wakatsuki,S. and Kawasaki,M. (2016) A novel mode of ubiquitin recognition by the ubiquitin-binding zinc finger domain of WRNIP1. *FEBS J.*, **283**, 2004–2017.
67. Ishida,T. and Kinoshita,K. (2008) Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics*, **24**, 1344–1348.
68. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
69. Kikhney,A.G. and Svergun,D.I. (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.*, **589**, 2570–2577.
70. Pryor,J.M. and Washington,M.T. (2011) Pre-steady state kinetic studies show that an abasic site is a cognate lesion for the yeast Rev1 protein. *DNA Repair*, **10**, 1138–1144.