

# SCIENTIFIC REPORTS



OPEN

## Structural basis for polymerase $\eta$ -promoted resistance to the anticancer nucleoside analog cytarabine

Olga Rechkoblit<sup>1</sup>, Jayati Roy Choudhury<sup>2</sup>, Angeliki Buku<sup>1</sup>, Louise Prakash<sup>2</sup>, Satya Prakash<sup>2</sup> & Aneel K. Aggarwal<sup>1</sup>

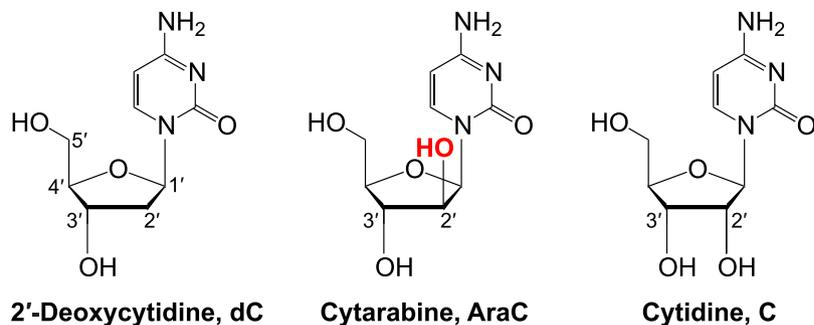
Cytarabine (AraC) is an essential chemotherapeutic for acute myeloid leukemia (AML) and resistance to this drug is a major cause of treatment failure. AraC is a nucleoside analog that differs from 2'-deoxycytidine only by the presence of an additional hydroxyl group at the C2' position of the 2'-deoxyribose. The active form of the drug AraC 5'-triphosphate (AraCTP) is utilized by human replicative DNA polymerases to insert AraC at the 3' terminus of a growing DNA chain. This impedes further primer extension and is a primary basis for the drug action. The Y-family translesion synthesis (TLS) DNA polymerase  $\eta$  (Pol $\eta$ ) counteracts this barrier to DNA replication by efficient extension from AraC-terminated primers. Here, we provide high-resolution structures of human Pol $\eta$  with AraC incorporated at the 3'-primer terminus. We show that Pol $\eta$  can accommodate AraC at different stages of the catalytic cycle, and that it can manipulate the conformation of the AraC sugar via specific hydrogen bonding and stacking interactions. Taken together, the structures provide a basis for the ability of Pol $\eta$  to extend DNA synthesis from AraC terminated primers.

Cytarabine (1- $\beta$ -D-arabinofuranosylcytosine, AraC) has been the mainstay therapy for acute myeloid leukemia (AML) for over 40 years<sup>1-3</sup>. Resistance to cytarabine is a major cause of treatment failure and only ~27% of adult patients survive more than 5 years after the initial diagnosis<sup>4</sup>. AraC is a nucleoside analog that differs from 2'-deoxycytidine only by the presence of an additional hydroxyl group at the C2' position of the 2'-deoxyribose (Fig. 1). This 2'-OH of the arabinose sugar moiety points in an opposite direction to that of the 2'-OH of the ribose sugar in ribonucleotides (Fig. 1).

AraC is a prodrug, which after entering the cell is phosphorylated by deoxycytidine and pyrimidine kinases to its active form, AraC 5'-triphosphate (AraCTP)<sup>5</sup>. AraC kills cells undergoing DNA synthesis in the S-phase of the cell cycle and its chemotherapeutic action is primarily associated with DNA fragmentation and chain termination<sup>5</sup>. Concordantly, high-fidelity replicative DNA polymerases  $\alpha$ ,  $\delta$  and  $\epsilon$  that replicate the bulk of genomic DNA in eukaryotes<sup>6,7</sup> can proficiently insert AraC at the 3' terminus of a growing DNA chain but the subsequent extension reaction is markedly impeded<sup>8-11</sup>. Nevertheless, a substantial fraction of AraC-terminated primers do get extended<sup>12</sup>, and thereby reducing the effectiveness of the chemotherapy. This has been attributed to specialized, lower fidelity translesion DNA synthesis (TLS) polymerases, which allow for the continuity of the replication fork by allowing replication through modified DNA bases that impede the high-fidelity replicative polymerase<sup>7,13</sup>. TLS polymerase  $\eta$  (Pol $\eta$ ) stands out in that human cells deficient in Pol $\eta$  are ~3-fold more sensitive to AraC than wild-type cells and the polymerase can efficiently extend AraC-terminated primers *in vitro*<sup>14</sup>.

To see how Pol $\eta$  can extend AraC-terminated primers we present here high-resolution crystal structures of human Pol $\eta$  in binary and ternary complexes with AraC. We show that Pol $\eta$  can accommodate AraC at different stages of the catalytic cycle, and that it can manipulate the conformation of the AraC sugar via specific hydrogen bonding and stacking interactions. Taken together, the structures provide an unexpected basis for the ability of Pol $\eta$  to extend DNA synthesis from AraC terminated primers.

<sup>1</sup>Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, Box 1677, 1425 Madison Avenue, New York, NY, 10029, USA. <sup>2</sup>Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX, 77555-1061, USA. Correspondence and requests for materials should be addressed to A.K.A. (email: [aneel.aggarwal@mssm.edu](mailto:aneel.aggarwal@mssm.edu))



**Figure 1.** Deoxycytidine, cytarabine, and cytidine nucleosides. Chemical structures of 2'-deoxycytidine (dC, 1- $\beta$ -D-2'-deoxy-ribofuranosylcytosine), cytarabine (AraC, 1- $\beta$ -D-arabinofuranosylcytosine), and cytidine (C, 1- $\beta$ -D-ribofuranosylcytosine). The 2'-OH group in arabinose and ribose sugar moieties points in opposite directions.

## Results

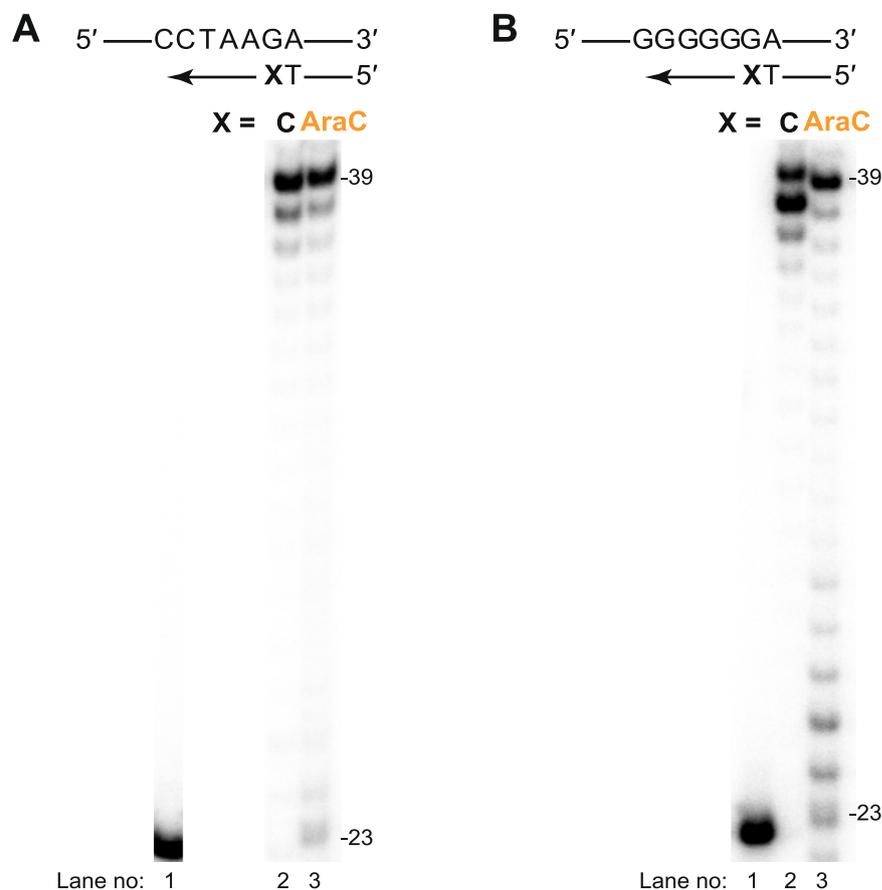
**Biochemical Analysis.** We first carried out biochemical analysis to compare the ability of human Pol $\eta$  to extend DNA synthesis from an unmodified C and AraC 3'-terminated DNA primers. As shown in Fig. 2A (lane 2), with a 5'...CCTAAGA...-3' DNA template (5'-GTCTAATACTTCCTAAGATGCCTACACTGGAGTACCGGA-3'), elongation from unmodified C paired with the cognate template G base results in nearly complete extension to the full products in 10 min of the reaction time. A slightly lesser amount of the extension products is observed with the AraC-terminated primer (Fig. 2A, lane 3). Likewise, with a 5'...GGGGGA...-3' DNA template that contains six consecutive G residues, extension from the 3'-terminal unmodified C or AraC opposite the first G of the template strand results in the nearly complete products (Fig. 2B, lanes 2 and 3). We conducted primer elongation on the G-repeat template in the presence of all four dNTPs with the unmodified primer, whereas, with the AraC-terminated primer, we used AraCTP, dATP, dGTP and dTTP nucleotides. Thus, the efficient formation of the fully extended products (Fig. 2B, lane 3) suggest that Pol $\eta$  can not only extend from a single AraC residue but also can consequently incorporate and extend from it to create a run of AraC bases. Overall, our results demonstrate that Pol $\eta$  is proficient in extension from AraC-terminated primers in various sequence contexts and are in accord with previous observations<sup>14</sup>.

**Structure determination.** We crystallized the human Pol $\eta$  catalytic core (residues 1 to 432) with a DNA 12-nucleotide (nt) template (5'-CATGACAGTGCT-3')/8-nt primer (5'-AGCACTGT-3') and AraCTP. However, when we solved the structure and refined it to 1.83 Å resolution, we discovered that despite the suboptimal reaction conditions in the crystallization drop (low pH 6.0 and presence of Ca<sup>2+</sup> ions) Pol $\eta$  had covalently added the AraC residue to the 3' end of the primer strand. This represents a “post-insertion” binary complex in which polymerase has not yet translocated to the next templating position. To capture the polymerase in the act of inserting dNTP from an AraC terminated primer, we first incubated the polymerase with a 12-nt template (5'-CATTGCAGTGCT-3')/7-nt primer (5'-AGCACTG-3') and AraCTP at pH 8.0 and in the presence of Mg<sup>2+</sup> to incorporate AraC at the primer terminus, followed by removal of unincorporated AraCTP, and then addition of nonreactive dATP analog dAMPNPP (2'-deoxyadenosine-5'-[( $\alpha,\beta$ )-imido]triphosphate) to the crystallization mix. This represents a ternary complex in which AraC is incorporated at 3' end of the primer and dAMPNPP is the incoming nucleotide. The structure was refined to high 1.75 Å resolution and provides atomic details on the conformation of AraC and its interactions with the polymerase. The crystal data, data collection statistics, and refinement statistics for both complexes are summarized in Table 1.

**Overall Arrangement.** In both complexes (Figs 3A and 4A), Pol $\eta$  embraces the template–primer with its palm (residues 1–13 and 90–238), fingers (residues 17–87), thumb (residues 241–301) domains and the PAD (polymerase associated domain or the little finger domain; residues 319–432) that is unique to Y-family polymerases. The palm domain provides the catalytic residues Asp13, Asp115, and Glu116, while the fingers domain is positioned above the templating base. The thumb and the PAD grip the template–primer DNA duplex from the minor and major groove surfaces, respectively (Figs 3A and 4A).

**Binary complex.** AraC is covalently incorporated into the primer strand and its base makes a normal Watson–Crick base pair with a template G. The sugar of AraC stacks against the planar phenyl ring of Phe18 in the fingers domain (Fig. 3B,C). Phe18 occurs at a position in common with an aromatic residue in many other DNA polymerases, functioning as a “steric gate” for discrimination against the insertion of ribonucleotides (NTPs). That is, the orientation of 2'-OH on a NTP sugar is such that it would clash with the aromatic residue and the NTP would be excluded from incorporation. Because the 2'-OH on an AraC sugar points in an opposite direction to that on an NTP sugar, it does not generate any steric clashes and provides a basis by which Pol $\eta$  can incorporate AraCTP but not an NTP into a growing DNA chain. AraC is further stabilized in the Pol $\eta$  active site by a hydrogen bond between its 3'-OH and the backbone N atom of Phe17 (Fig. 3C).

Interestingly, the conformation of AraC sugar is C3'-*endo* (Fig. 3C,D). This was a surprise because an A-form RNA-like C3'-*endo* sugar pucker is considered unfavorable for an arabinonucleoside<sup>15–17</sup>. However, the C3'-*endo* conformation is stabilized in the Pol $\eta$  active site by a hydrogen bond and stacking interactions with Arg61, a



**Figure 2.** Pol $\eta$ -catalyzed extension from AraC 3'-terminated primers. **(A)** Extension of  $^{32}\text{P}$  5'-end-labeled 23-nt primer with the unmodified C or AraC residues at the 3'-end bound to the 5'-...CCTAAGA...-3' 39-nt template in the presence of dATP, dGTP, and dTTP. Lane 1, unmodified DNA substrate; Lane 2, extension from the unmodified C; Lane 3, extension from AraC. Lanes 1, 2 and 3 were run on the same gel, but lane 1 was cropped from a different part of the gel. Thus, lane 1 is separated with white space from lanes 2 and 3. The original gel is shown in Supplementary Fig. 1. **(B)** Extension with the 5'-...GGGGGGA...-3' 39-nt template. Lane 1, unmodified DNA substrate; Lane 2, extension from the unmodified C in the presence of dATP, dGTP, dTTP and dCTP; Lane 3, extension from the AraC in the presence of dATP, dGTP, dTTP and AraCTP. Lanes 1–3 were run together on the same gel. The original gel is shown in Supplementary Fig. 2. All extension reactions were conducted at 37 °C for 10 min.

residue unique to Pol $\eta$ . In addition, there is a direct hydrogen bond between 2'-OH of the AraC sugar and the main chain carbonyl of Ile48. Together, these interactions facilitate and stabilize the C3'-*endo* pucker even though it is unfavorable for an arabinonucleoside. Overall, it further supports the idea that DNA polymerases in general mold DNA into an A-form-like conformation at the primer terminus<sup>18–26</sup>.

**Ternary complex.** The ternary complex provides a basis for the addition of dNTP from an AraC primer terminus. As expected, 1) AraC has incorporated at the 3' end of the primer strand, 2) the polymerase has translocated to the next templating position, and 3) the dATP nonreactive analog dAMPNPP has entered the active site (Fig. 4A–C). The dAMPNPP sugar is juxtaposed against the phenyl ring of Phe18, and two  $\text{Mg}^{2+}$  ions (A and B) complete the active site.  $\text{Mg}_A^{2+}$  is ligated by the  $\alpha$ -phosphate of dAMPNPP, the carboxylates of Asp115 (2.55 Å) and Glu116 (2.22 Å), and a water molecule (Fig. 4B).  $\text{Mg}_B^{2+}$  is coordinated in the basal octahedral plane by the dAMPNPP  $\beta$ - and  $\gamma$ -phosphates, the carboxylates of Asp13 (2.15 Å) and Asp115 (2.08 Å) and by the backbone carboxyl oxygen of Met14 (2.25 Å). Thus, the overall configuration of the active site is very similar to that observed in previous Pol $\eta$  ternary complexes with unmodified DNAs and nonreactive dNTP analogs<sup>27</sup>, except that the 3'-OH of the AraC primer terminus is shifted by  $\sim 1.1$  Å compared to the unmodified structures and is now too far away to interact with  $\text{Mg}_A^{2+}$ . This small shift of the 3'-OH is due to a slight tilt of the AraC sugar moiety (Fig. 3D). Hence, the  $\chi$  torsion angle around the N-glycosidic bond that connects the base to the sugar (O4'-C1'-N1-C2) is changed from  $-115.0^\circ$  in the unmodified 3' primer T residue to  $-149.2^\circ$  in the 3' primer AraC. The altered position of the 3'-OH of AraC might underlie the slightly lower efficiency in Pol $\eta$ -catalyzed extension from AraC 3'-terminus relative to the unmodified primer base.

Compared to the binary complex, the AraC sugar at the primer terminus has switched to the normal for DNA B-form C2'-*endo* conformation (Figs 4B,C). It appears to be driven into this conformation by a direct hydrogen

	AraC post-insertion binary complex	AraC extension ternary complex
<b>Data collection</b>		
Space group	P6 <sub>1</sub>	P6 <sub>1</sub>
Cell dimensions:		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	98.6 98.6 81.5	98.8 98.8 81.8
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Resolution range (Å) <sup>a</sup>	45.0–1.83 (1.86–1.83)	85.6–1.75 (1.78–1.75)
<i>R</i> <sub>merge</sub> (%)	7.4 (61.2)	8.5 (26.5)
<i>I</i> / $\sigma$ <i>I</i>	31.4 (2.3)	17.4 (1.1)
Completeness (%)	99.7 (95.7)	100 (100)
Redundancy	12.0 (7.3)	16.6 (16.4)
CC <sub>1/2</sub> (%)	100 (80.7)	100 (60.5)
<b>Refinement</b>		
Resolution range (Å)	42.7–1.83	59.1–1.75
No. reflections	39,435	45,739
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	16.0/19.6	16.8/19.5
No. atoms		
Protein	3,411	3,371
DNA	431	438
Ligand (dAMPNPP)	0	1
Ligand (other)	6	2
Ion (Mg <sup>2+</sup> )	0	2
Water	436	374
<i>B</i> -factors		
Protein	26.5	42.8
DNA	30.1	44.8
Ligand (dAMPNPP)	–	29.1
Ligand (other)	16.3	45.6
Ion (Mg <sup>2+</sup> )	–	35.0
Water	33.2	43.6
R.m.s. deviations		
Bond length (Å)	0.006	0.010
Bond angles (°)	0.83	1.04

**Table 1.** X-ray data collection and refinement statistics. <sup>a</sup>Values in parentheses are for highest-resolution shell.

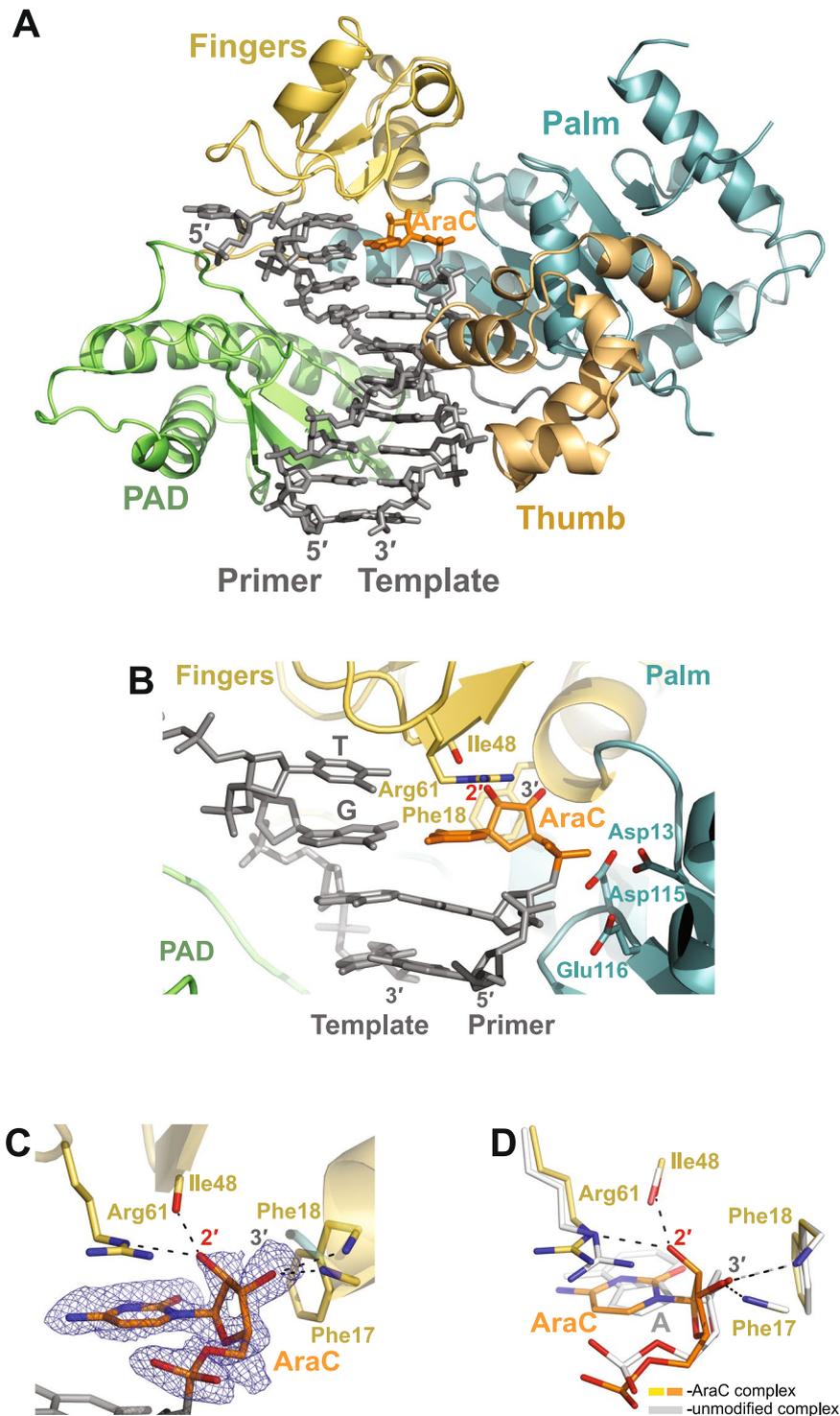
bond between 2'-OH of the AraC sugar and the non-bridging oxygen atom of the  $\alpha$ -phosphate of dAMPNPP. Thus, in the C2'-*endo* conformation, the AraC sugar plays a direct role in binding of the incoming nucleotide. The 2'-OH of AraC sugar is also involved in a water-mediated hydrogen bond with Arg61. This has the effect of stabilizing Arg61 into a single rotameric conformation, as compared to multiple conformations of Arg61 in Pol $\eta$  ternary complex with unmodified primer terminus and nonreactive dNTP analogs (Fig. 4D)<sup>27</sup>.

## Discussion

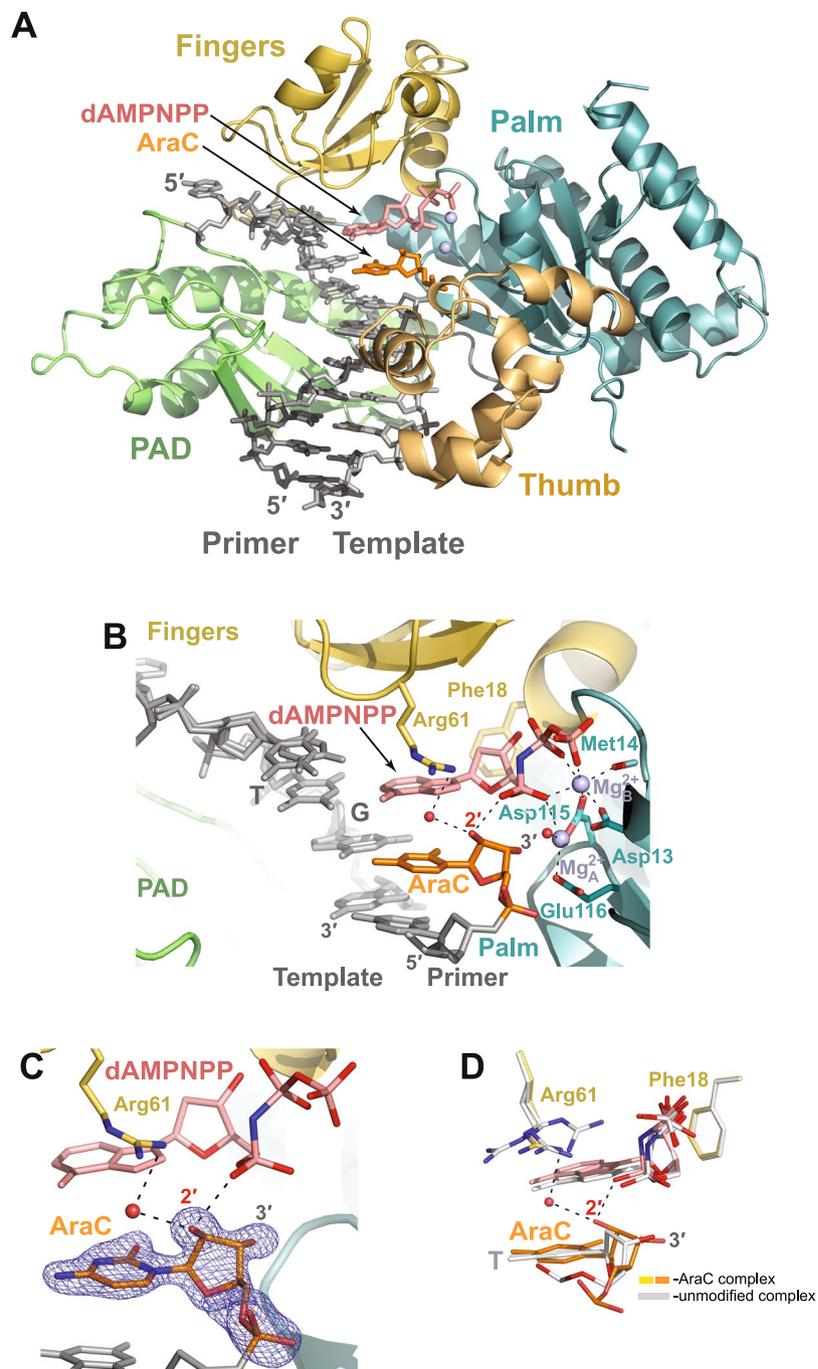
Pol $\eta$  is currently the only known human DNA polymerase capable of efficiently extending DNA synthesis from an AraC-terminated primer. We show here that AraC is stabilized in the Pol $\eta$  active site via specific hydrogen bonding and stacking interactions. By contrast, human X-family polymerases Pol $\lambda$  and Pol $\beta$  that function in base excision repair (BER) can insert AraC in a gapped DNA substrate, but there is no evidence to suggest that these polymerases can extend DNA synthesis from AraC-terminated primers<sup>28,29</sup>.

Interestingly, a time-resolved crystallographic study of Pol $\eta$ 's catalytic cycle has revealed that the deoxyribose sugar of the 3'-terminal primer base changes its conformation from C2'-*endo* to C3'-*endo* during the phosphoryl transfer reaction<sup>30</sup>. This is apparently necessary to avoid clashes with the non-bridging oxygen of incoming nucleotide. We envisage a similar C2'-*endo* to C3'-*endo* transition of the AraC sugar during the catalytic reaction. Indeed, when we superimpose the C3'-*endo* AraC sugar (observed in our binary complex) onto the penultimate primer base in the time resolved structure (PDB ID: 4ECX, 300 sec reaction time)<sup>30</sup> (Fig. 5), the “extra” 2'-OH on the AraC sugar is not only accommodated without any steric overlap, but can potentially make hydrogen bonds with the phosphate group and adenine base of the incorporated nucleotide (Fig. 5). Whether this creates sequence bias or mutagenic incorporation of purines during the extension of AraC remains to be determined.

Overall, the ability of the arabinose sugar in AraC to adopt or transition to the C3'-*endo* pucker is surprising. This is because the sugar in AraC (and other arabinonucleosides) is considered relatively restricted in conformation and preferentially adopts the C2'-*endo*<sup>16</sup> or C1'-*exo*<sup>17</sup> puckers. However, we show that this conformational

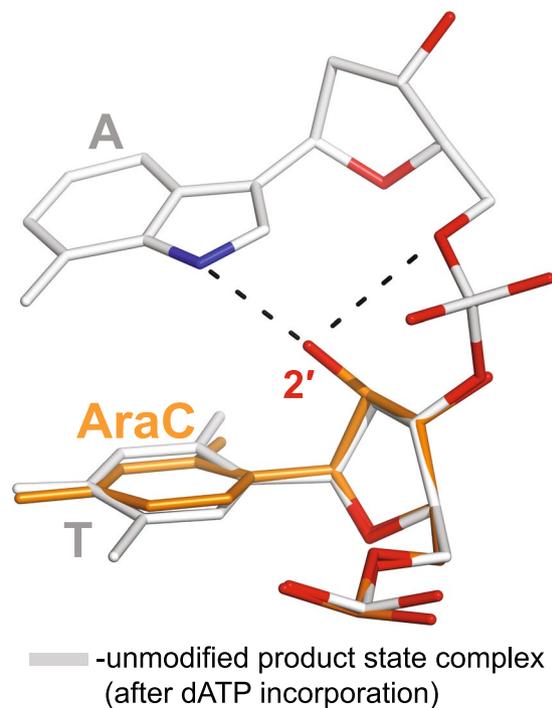


**Figure 3.** The Pol $\eta$  AraC post-insertion binary complex. **(A)** Overall structure of the complex; the palm, fingers, thumb and PAD domains are shown in cartoon representation in cyan, yellow, light orange, and green, respectively. The DNA template–primer duplex is shown in gray sticks with the 3′-terminal primer AraC residue in orange. **(B)** A close-up view of the AraC in the active site of Pol $\eta$ . Asp13, Asp115, and Glu116 are the catalytic residues. **(C)** A simulated annealing  $F_o - F_c$  omit map (contoured at  $3.0\sigma$ -level at 1.83 Å resolution and colored in blue) showing the clear electron density for the entire AraC residue. The sugar moiety of the AraC is stacked against the steric gate residue Phe18 and its 2′-OH group points away from it. **(D)** Superposition of the AraC and the unmodified (PDB ID: 4J9P)<sup>36</sup> post-insertion binary complexes. The unmodified complex is in white and has an unmodified A residue at the 3′-end of the primer strand; the structures are superimposed by the palm and fingers domains of the polymerase. Both, the AraC and the unmodified A sugar moieties have C3′-endo conformations.



**Figure 4.** The Pol $\eta$  AraC extension ternary complex. **(A)** Overall structure of the complex. The incoming dAMPNPP residue is in red. The Mg $^{2+}$  ions A and B are represented as light blue spheres. The other colors are as described in Fig. 3. **(B)** The active site of the AraC extension ternary complex. Asp13, Asp115, and Glu116 are the catalytic residues. The water molecules coordinated by the 2'-OH of the AraC and by the Mg $^{2+}$  B-ion are shown as red spheres. **(C)** A simulated annealing Fo-Fc omit map (contoured at 3.0 $\sigma$ -level at 1.75 Å resolution and colored in blue) showing the clear electron density for the entire AraC residue. **(D)** Superposition of the AraC and the unmodified (PDB ID: 3MR2)<sup>27</sup> ternary complexes. The unmodified complex is in white and has an unmodified T residue at the 3'-end of the primer strand; the structures are superimposed by the palm and fingers domains of the polymerase. Both, the AraC and the unmodified T sugar moieties have C2'-endo conformations.

restriction is easily surmounted in the Pol $\eta$  active site by interactions that 2'-OH of the C3'-endo sugar is able to establish with the polymerase. It would be interesting to know whether a more “rigid” sugar in AraC reduces the ability of Pol $\eta$  to extend DNA synthesis: lending to increased efficacy in the treatment of AML. In particular, locked conformation nucleotide analog chemistry may be helpful in the design and synthesis of such conformationally restricted AraC-like analogs<sup>31</sup>.



**Figure 5.** Model of the Pol $\eta$  AraC extension product complex. The C3'-endo AraC is superimposed onto the penultimate primer base in the product state of the unmodified complex after covalent insertion of an adenine base (PDB ID: 4ECX, 300 sec reaction time)<sup>30</sup>. The 2'-OH of the AraC makes hydrogen bonds with the O5' of the phosphate group and with the N7 acceptor atom of the just incorporated adenine base.

## Methods

**Preparation of protein and DNA for crystallization.** An N-terminal His<sub>6</sub> tagged catalytic core of human Pol $\eta$  (residues 1–432) with a C406M mutation was overexpressed in *Escherichia coli* and purified as previously described<sup>27,32</sup>. Briefly, the His<sub>6</sub> tag was removed by overnight incubation with PreScission protease, and the protein was purified by ion-exchange (MonoS) chromatography followed by size-exclusion (Superdex 75). The protein was concentrated to ~1.3 mg/ml in 25 mM tris (pH 8.0), 250 mM NaCl, and 2 mM tris(2-carboxyethyl) phosphate (TCEP) and stored in aliquots at –80 °C.

**Crystallization.** The crystals of the post-insertion AraC binary complex were obtained by incubating the human Pol $\eta$  catalytic core with a template-primer (5'-CATGACAGTGCT-3'/5'-AGCACTGT-3') (Glen Research, Inc) in the presence of AraCTP (TriLink Biotechnologies inc.) by the hanging drop method against a reservoir solution containing 0.1 M MES pH 6.0 buffer and 10–14% PEG1500. Briefly, the template-primer DNAs were annealed and mixed with Pol $\eta$  in a 1.2:1 molar ratio to ~0.02 mM concentration of the complex in 25 mM tris (pH 8.0), 125 mM NaCl and 1 mM TCEP. The complex was concentrated with Amicon Ultra centrifugal filter (cut-off 3KDa) to a final complex concentration of ~0.105 mM. AraCTP and CaCl<sub>2</sub> were then added to the complex to 2 mM and 4 mM concentrations, respectively. The complex was centrifuged at 8,000 rpm for 2 min at 4 °C. Crystallization drop was formed by mixing 1  $\mu$ L of the complex with 1  $\mu$ L of the reservoir solution and the crystals were grown at 20 °C. (The AraC residue was covalently added to the 3' end of the primer strand under these conditions). A round of microseeding was necessary to produce larger diffraction-quality crystals. The crystals were cryoprotected in the reservoir solution supplemented with 24% PEG1500 and, sequentially, with 5%, 10% and 20% glycerol and flash frozen in liquid nitrogen for X-ray data collection.

To crystallize the extension ternary complex we prepared the binary polymerase complex with 12-nt template (5'-CATTGACAGTGCT-3')/7-nt primer (5'-AGCACTG-3') DNA as described above. We then added 2 mM AraCTP with 5 mM MgCl<sub>2</sub> and incubated the reaction for 20 min at 16 °C. Following removal of the unincorporated AraCTP with Amicon Ultra centrifugal filter, we added a nonreactive dATP analog dAMP-NPP (2'-deoxyadenosine-5'-[( $\alpha,\beta$ )-imido]triphosphate) at 2 mM and 5 mM MgCl<sub>2</sub>. The crystals were grown as described for the post-insertion binary complex.

**Structure determination and refinement.** The X-ray diffraction data were collected at the NSLS X25 beam line at the Brookhaven National Laboratory and at the 24-ID NE-CAT beamline at Advanced Photon Source in Chicago. The data from the NSLS X25 beam line were processed and scaled using the HKL2000 suite<sup>33</sup> and the data from the 24-ID NE-CAT beamline were processed by RAPD pipeline (<http://necat.chem.cornell.edu/>). We solved the structure by the molecular replacement method (Phaser)<sup>34</sup> in the CCP4 program package<sup>35</sup> using the Pol $\eta$  postinsertion binary complex structure (PDB ID: 4J9P) as a search model<sup>36</sup>. The model building, including substitution of the DNA sequence, was finished manually in Coot<sup>37</sup> based on the electron density maps

calculated in PHENIX Refine<sup>38</sup>. The final model was refined in PHENIX Refine<sup>38</sup> and belongs to P6<sub>1</sub> space group with unit cell dimensions of a = b = 98.6 Å, c = 81.5 Å,  $\alpha = \beta = 90^\circ$ , and  $\gamma = 120.0^\circ$ . The structure has been refined to 1.83 Å resolution with R<sub>free</sub> of 19.6% and R<sub>work</sub> of 16.0% and consists of one Pol $\eta$  molecule (residues 1 to 432), one DNA template (residues 2 to 12), one DNA primer (residues 1 to 9), and a total of 436 solvent molecules. The placement of the AraC residue was verified using simulated annealing omit maps calculated in PHENIX<sup>38</sup> with the AraC omitted from the model before heating to 2,000 K and then slowly cooling.

We solved the structure of the AraC-modified extension ternary complex by MR using the structure of Pol $\eta$  ternary complex (PDB ID 3MR2)<sup>27</sup> as a search model. We refined the model as described above to 1.75 Å resolution in P6<sub>1</sub> space group with unit cell dimensions of a = b = 98.8 Å, c = 81.8 Å,  $\alpha = \beta = 90^\circ$ , and  $\gamma = 120.0^\circ$ ; with R<sub>free</sub> of 19.6% and R<sub>work</sub> of 16.8%. The model consists of one Pol $\eta$  molecule (residues 1 to 432), one DNA template (residues 2 to 12), one DNA primer (residues 1 to 8), one dAMPNPP, two Mg<sup>2+</sup> ions and a total of 383 solvent molecules.

The crystal data, together with the data collection and refinement statistics, are summarized in Table 1.

**Primer extension.** We used two different DNA templates, G1 and G6, to examine extension from a C or AraC paired with an unmodified template G. The G1 template oligonucleotide contains the sequence 5'-GTCTAATACTTCTAAGATGCCTACACTGGAGTACCGGA-3'. The G6 template oligonucleotide contains a run of 6 G residues and it has the sequence 5'-GTCTAATACTTGGGGGGATGCCTACACTGGAGTACCGGA-3'. We employed the 5'-<sup>32</sup>P labeled 23 nt primers 5'-TCCGGTACTCCAGTGTAGGCATX-3' where X is a C or an AraC. Thus, DNA substrates for primer extension consisted of a <sup>32</sup>P-radiolabeled oligonucleotide primer (23 nt) annealed to a 39 nt oligonucleotide DNA template by heating a mixture of primer/template at a 1:1.5 molar ratio to 95 °C and allowing it to cool to room temperature for several hours. The standard DNA polymerase reaction (5  $\mu$ l) contained 25 mM tris (pH 7.5), 5 mM MgCl<sub>2</sub>, 1 mM dithiothreitol, 100  $\mu$ g/ml BSA, 10% glycerol, 10 nM DNA substrate, and 1 nM of Pol $\eta$ . Primer extension was assayed at 37 °C for 10 min in the presence of 25  $\mu$ M of dATP, dGTP, and dTTP for G1 template and 50  $\mu$ M each of dATP, dGTP, dTTP and dCTP (Roche Biochemicals, Indianapolis) for extension from C and 50  $\mu$ M each of dATP, dGTP, dTTP and araCTP for extension from AraC on G6 template DNA. Reactions were stopped by the addition of loading buffer (95% formamide, 0.05% cyanol blue and 0.05% bromophenol blue) and the reaction products were then resolved on a 12% polyacrylamide gel containing 8 M urea. Gels were dried before autoradiography with a PhosphorImager.

**Data availability.** Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 6D0M and 6D0Z for the AraC post-insertion binary and extension ternary complexes, respectively.

## References

- Magina, K. N. *et al.* Cytarabine dose in the consolidation treatment of AML: a systematic review and meta-analysis. *Blood* **130**, 946–948, <https://doi.org/10.1182/blood-2017-04-777722> (2017).
- Kadia, T. M., Ravandi, F., O'Brien, S., Cortes, J. & Kantarjian, H. M. Progress in acute myeloid leukemia. *Clin Lymphoma Myeloma Leuk* **15**, 139–151, <https://doi.org/10.1016/j.clml.2014.08.006> (2015).
- Lichtman, M. A. A historical perspective on the development of the cytarabine (7days) and daunorubicin (3days) treatment regimen for acute myelogenous leukemia: 2013 the 40th anniversary of 7 + 3. *Blood Cells Mol Dis* **50**, 119–130, <https://doi.org/10.1016/j.bcmd.2012.10.005> (2013).
- Cancer Facts and Figures 2017. *American Cancer Society* (2017).
- Grant, S. Ara-C: cellular and molecular pharmacology. *Adv Cancer Res* **72**, 197–233 (1998).
- Lujan, S. A., Williams, J. S. & Kunkel, T. A. DNA Polymerases Divide the Labor of Genome Replication. *Trends Cell Biol* **26**, 640–654, <https://doi.org/10.1016/j.tcb.2016.04.012> (2016).
- Jain, R., Aggarwal, A. K. & Rechko, B. Eukaryotic DNA polymerases. *Current Opinion in Structural Biology* **53**, 77–87 (2018).
- Mikita, T. & Beardsley, G. P. Functional consequences of the arabinosylcytosine structural lesion in DNA. *Biochemistry* **27**, 4698–4705 (1988).
- Perrino, F. W. & Mekosh, H. L. Incorporation of cytosine arabinoside monophosphate into DNA at internucleotide linkages by human DNA polymerase alpha. *J Biol Chem* **267**, 23043–23051 (1992).
- Perrino, F. W., Mazur, D. J., Ward, H. & Harvey, S. Exonucleases and the incorporation of arabinucleotides into DNA. *Cell Biochem Biophys* **30**, 331–352, <https://doi.org/10.1007/BF02738118> (1999).
- Tsuda, M. *et al.* The dominant role of proofreading exonuclease activity of replicative polymerase epsilon in cellular tolerance to cytarabine (Ara-C). *Oncotarget* **8**, 33457–33474, <https://doi.org/10.18632/oncotarget.16508> (2017).
- Major, P. P., Egan, E. M., Herrick, D. J. & Kufe, D. W. Effect of Ara-C incorporation on deoxyribonucleic acid synthesis in cells. *Biochem Pharmacol* **31**, 2937–2940 (1982).
- Prakash, S., Johnson, R. E. & Prakash, L. Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu Rev Biochem* **74**, 317–353, <https://doi.org/10.1146/annurev.biochem.74.082803.133250> (2005).
- Chen, Y. W., Cleaver, J. E., Hanaoka, F., Chang, C. F. & Chou, K. M. A novel role of DNA polymerase eta in modulating cellular sensitivity to chemotherapeutic agents. *Mol Cancer Res* **4**, 257–265, <https://doi.org/10.1158/1541-7786.MCR-05-0118> (2006).
- Teng, M. K., Liaw, Y. C., van der Marel, G. A., van Boom, J. H. & Wang, A. H. Effects of the O2' hydroxyl group on Z-DNA conformation: structure of Z-RNA and (araC)-[Z-DNA]. *Biochemistry* **28**, 4923–4928 (1989).
- Gmeiner, W. H., Konerding, D. & James, T. L. Effect of cytarabine on the NMR structure of a model okazaki fragment from the SV40 genome. *Biochemistry* **38**, 1166–1175, <https://doi.org/10.1021/bi981702s> (1999).
- Li, F. *et al.* 2'-Fluoroarabino- and arabinonucleic acid show different conformations, resulting in deviating RNA affinities and processing of their heteroduplexes with RNA by RNase H. *Biochemistry* **45**, 4141–4152, <https://doi.org/10.1021/bi052322r> (2006).
- Doublet, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251–258, <https://doi.org/10.1038/34593> (1998).
- Li, Y., Korolev, S. & Waksman, G. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J* **17**, 7514–7525, <https://doi.org/10.1093/emboj/17.24.7514> (1998).
- Franklin, M. C., Wang, J. & Steitz, T. A. Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell* **105**, 657–667 (2001).

21. Johnson, S. J., Taylor, J. S. & Beese, L. S. Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc Natl Acad Sci USA* **100**, 3895–3900, <https://doi.org/10.1073/pnas.0630532100> (2003).
22. Batra, V. K. *et al.* Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure* **14**, 757–766, <https://doi.org/10.1016/j.str.2006.01.011> (2006).
23. Berman, A. J. *et al.* Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. *EMBO J* **26**, 3494–3505, <https://doi.org/10.1038/sj.emboj.7601780> (2007).
24. Garcia-Diaz, M., Bebenek, K., Krahn, J. M., Pedersen, L. C. & Kunkel, T. A. Role of the catalytic metal during polymerization by DNA polymerase lambda. *DNA Repair (Amst)* **6**, 1333–1340, <https://doi.org/10.1016/j.dnarep.2007.03.005> (2007).
25. Wang, F. & Yang, W. Structural Insight into Translesion Synthesis by DNA Pol II. *Cell* **139**, 1279–1289, <https://doi.org/10.1016/j.cell.2009.11.043> (2009).
26. Swan, M. K., Johnson, R. E., Prakash, L., Prakash, S. & Aggarwal, A. K. Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase delta. *Nat Struct Mol Biol* **16**, 979–986, <https://doi.org/10.1038/nsmb.1663> (2009).
27. Biertumpfel, C. *et al.* Structure and mechanism of human DNA polymerase eta. *Nature* **465**, 1044–1048, <https://doi.org/10.1038/nature09196> (2010).
28. Chrencik, J. E., Burgin, A. B., Pommier, Y., Stewart, L. & Redinbo, M. R. Structural impact of the leukemia drug 1-beta-D-arabino-furanosylcytosine (Ara-C) on the covalent human topoisomerase I-DNA complex. *J Biol Chem* **278**, 12461–12466, <https://doi.org/10.1074/jbc.M212930200> (2003).
29. Prakash Gowda, A. S., Polizzi, J. M., Eckert, K. A. & Spratt, T. E. Incorporation of gemcitabine and cytarabine into DNA by DNA polymerase beta and ligase III/XRCC1. *Biochemistry* **49**, 4833–4840, <https://doi.org/10.1021/bi100200c> (2010).
30. Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y. J. & Yang, W. Watching DNA polymerase eta make a phosphodiester bond. *Nature* **487**, 196–201, <https://doi.org/10.1038/nature11181> (2012).
31. Singh, S. K., Nielsen, P., Koshkin, A. A. & Wengel, J. LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chem Commun*, 455–456, <https://doi.org/10.1039/a708608c> (1998).
32. Ummat, A. *et al.* Structural basis for cisplatin DNA damage tolerance by human polymerase eta during cancer chemotherapy. *Nat Struct Mol Biol* **19**, 628–632, <https://doi.org/10.1038/nsmb.2295> (2012).
33. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307–326 (1997).
34. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674, <https://doi.org/10.1107/S0021889807021206> (2007).
35. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242, <https://doi.org/10.1107/S0907444910045749> (2011).
36. Zhao, Y. *et al.* Mechanism of somatic hypermutation at the WA motif by human DNA polymerase eta. *Proc Natl Acad Sci USA* **110**, 8146–8151, <https://doi.org/10.1073/pnas.1303126110> (2013).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132, <https://doi.org/10.1107/S0907444904019158> (2004).
38. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221, <https://doi.org/10.1107/S0907444909052925> (2010).

## Acknowledgements

We thank the staff at beam lines X25 at the Brookhaven National Laboratory and Northeastern Collaborative Access Team (NECAT) at the Argonne National Laboratory for facilitating x-ray data collection. This work was supported by NIH grant R01 CA200575. The NECAT beamline was funded by the National Institute of General Medical Sciences (NIGMS) from the NIH (P41 GM103403). 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., S.P. and A.K.A. designed the experiments. A.B. performed protein expression and purification. O.R. performed crystallization, X-ray data collection and refinement; J.R.C. performed primer extension experiments under supervision of S.P. and L.P.; O.R. and A.K.A. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-30796-w>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018