Published online 14 June 2018

Nucleic Acids Research, 2018, Vol. 46, No. 13 ***6455–6469***
doi: 10.1093/nar/gky444

5-Formylcytosine mediated DNA–protein cross-links block DNA replication and induce mutations in human cells

Shaofei Ji[1], Iwen Fu[2], Spandana Naldiga[3], Hongzhao Shao[1], Ashis K. Basu[3], Suse Broyde[2] and Natalia Y. Tretyakova[4,5,*]

[1]Department of Chemistry, University of Minnesota, Minneapolis, MN 55455, USA, [2]Department of Biology, New York University, New York, NY 10003, USA, [3]Department of Chemistry, University of Connecticut, Storrs, CT 06269, USA, [4]Department of Medicinal Chemistry, University of Minnesota, Minneapolis, MN 55455, USA and [5]Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA

Received April 02, 2018; Revised April 27, 2018; Editorial Decision May 01, 2018; Accepted May 29, 2018

ABSTRACT

5-Formylcytosine (5fC) is an epigenetic DNA modification introduced via TET protein-mediated oxidation of 5-methyl-dC. We recently reported that 5fC form reversible DNA–protein conjugates (DPCs) with histone proteins in living cells (Ji et al. (2017) *Angew. Chem. Int. Ed.*, 56:14130–14134). We now examined the effects of 5fC mediated DPCs on DNA replication. Synthetic DNA duplexes containing site-specific DPCs between 5fC and lysine-containing proteins and peptides were subjected to primer extension experiments in the presence of human translesion synthesis DNA polymerases η and κ. We found that DPCs containing histones H2A or H4 completely inhibited DNA replication, but the replication block was removed when the proteins were subjected to proteolytic digestion. Cross-links to 11-mer or 31-mer peptides were bypassed by both polymerases in an error-prone manner, inducing targeted C→T transitions and –1 deletions. Similar types of mutations were observed when plasmids containing 5fC-peptide cross-links were replicated in human embryonic kidney (HEK) 293T cells. Molecular simulations of the 11-mer peptide-dC cross-links bound to human polymerases η and κ revealed that the peptide fits well on the DNA major groove side, and the modified dC forms a stable mismatch with incoming dATP via wobble base pairing in the polymerase active site.

INTRODUCTION

Genomic DNA is associated with a range of architectural and regulatory proteins including histones, HMG proteins and protein factors involved in transcription, replication, and repair. These dynamic DNA–protein complexes can be converted to covalent DNA–protein cross-links (DPCs) when proteins are trapped on DNA strands upon exposure to α,β-unsaturated carbonyls, free radicals, heavy metals, and UV radiation (1–5). A wide range of proteins participate in DPC formation, including histones, DNA polymerases, DNA repair proteins and transcription factors (4,6,7). Because of their unusually large size and their ability to disrupt DNA–protein interactions, covalent DPCs can interfere with chromatin architecture, inhibit DNA replication, transcription and repair, and induce mutations and toxicity in human cells (1–4,8–10). Cellular DPCs have been proposed to contribute to human diseases including cancer, aging, and neurodegenerative disorders (11,12).

We and others recently reported that 5-formylcytosine (5fC) residues of DNA form reversible DPCs with Lys side chains of histone proteins (Figure 1A) (13–16). 5fC bases are endogenous DNA epigenetic marks generated via oxidation of 5-methylcytosine (5mC) by ten eleven translocation dioxygenases (TET) (17–19). Although the global genomic levels of 5fC are relatively low (0.02–0.002% of total cytosines), they are found in all mammalian tissues (20) and are recognized by specific protein readers (21,22). 5fC are thought to play a role in epigenetic regulation due to their presence in regulatory elements of mammalian genes (18) and their association with gene expression levels (21,23). The recent reports of 5fC-mediated cross-links to histone proteins *in vitro* (16) and in human cells ($1.20 \pm 0.07 \times 10^{-4}$ % of all dCs) (15) have raised some important questions such as: *Do reversible 5fC-histone DPCs regulate gene expression?* And *Can 5fC-histone DPCs be tolerated by the DNA replication machinery?*

A recent report by the Balasubramanian group suggests that Schiff base complexes between histone residues and 5fC in the genome influence gene expression levels by changing nucleosomal organization and establishing dis-

*To whom correspondence should be addressed. Tel: +1 612 626 3432; Fax: +1 612 624 3869; Email: trety001@umn.edu

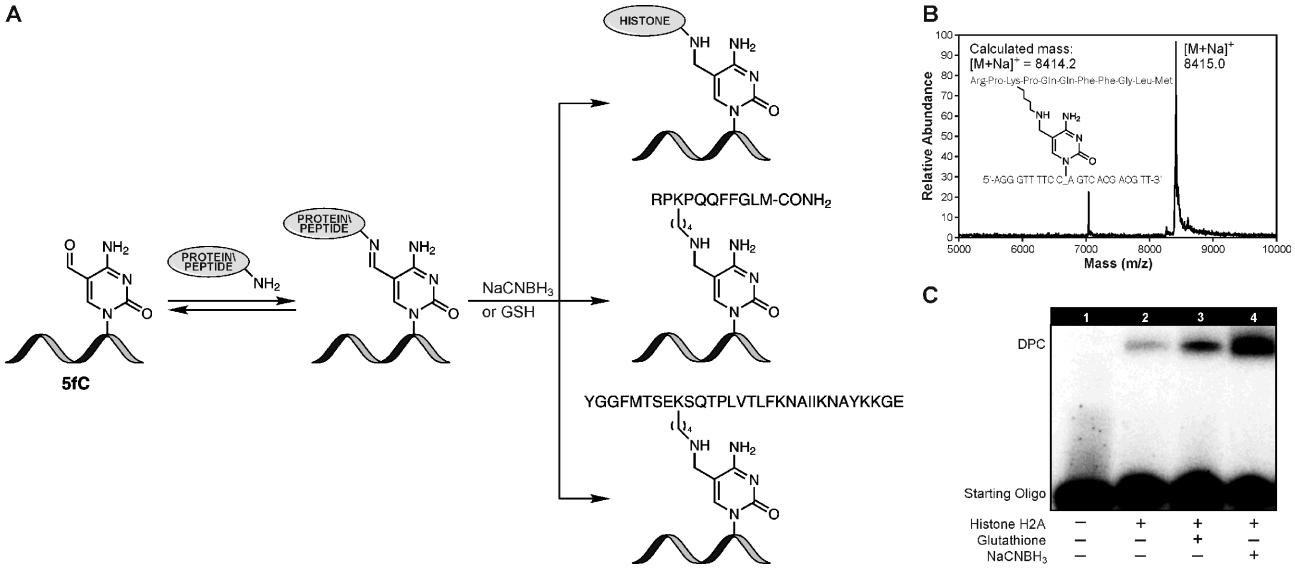

Figure 1. (**A**) The formation of DNA–protein and DNA–peptide cross-links at 5-formylcytosine of DNA. The aldehyde group of 5fC readily forms reversible Schiff base conjugates with the Arg or Lys side chains of proteins and polypeptides, which can be stabilized by reduction with NaCNBH3 or glutathione (GSH). (**B**) Representative MALDI MS spectrum of the conjugate between 5fC (X) containing DNA 23-mer 5′-AGG GTT TTC CXA GTC ACG ACG TT-3′ and 11-mer polypeptide RPKPQQFFGLM-CONH2 following NaCNBH3 reduction. (**C**) Denaturing PAGE analysis of the cross-links between 5fC (X) containing DNA 23-mer 5′-AGG GTT TTC CXA GTC ACG ACG TT-3′ and histone H2A. The reaction mixtures were heated prior to loading on gel, and samples in lanes 3 and 4 were reduced with NaCNBH3 or GSH, respectively, to stabilize DPCs.

tinct regulatory regions (bioRxiv doi: 10.1101/224444). However, to our knowledge, there are no previous studies that address the effects of 5fC-histone conjugates on DNA replication. Our previous studies with other types of DPCs have established a key role of translesion synthesis (TLS) polymerases such as human Pol κ and η in catalyzing DNA replication past DNA–peptide conjugates (24–26). Unlike replicative Pols which have tight active sites and undergo a conformational change upon binding correct dNTP, Pol κ and η have open sites that allow them to bypass bulky DNA lesions. These specialized polymerases are recruited to stalled replication forks and traverse the damaged sites, affording the cell additional time to repair the damage before the replicase returns to complete genome duplication (27,28).

In the present work, we evaluated the ability of human TLS polymerases κ and η to catalyze DNA synthesis past 5fC-mediated conjugates to histone H2A, histone H4 and shorter polypeptides representing possible proteolytic degradation products of larger DPCs. These two TLS polymerases were selected based on our previous studies demonstrating their ability to catalyze polymerase bypass past other DNA-polypeptide cross-links (25,26,29). For example, we have shown that hPol η and κ knockdown in human cells influences replication bypass fidelity of DPCs conjugated at the C7 position of 7-deaza-dG (24). In addition, our preliminary cellular experiments revealed that hPol η and κ knockdown strongly affects the replication fidelity of 5fC-mediated DPCs in human cells (A. Basu, unpublished observations). We employed a combination of gel electrophoresis and mass spectrometry based techniques to characterize the efficiency and the fidelity of polymerase bypass past 5fC-polypeptide lesions. Further, molecular modeling and molecular dynamics simulations were conducted to evaluate the structural effects of such endogenous DPCs on the ternary complex of hPol η and κ with DNA template–primer complex and the incoming dNTPs. Finally, cellular replication of 5fC-peptide containing plasmids was examined in human embryonic kidney (HEK) 293T cells.

MATERIALS AND METHODS

Full details of the experimental and modeling methods are given in the supporting information

Synthesis and characterization of 5fC-containing oligonucleotide. 5fC-containing 23-mer and 17-mer oligodeoxynucleotides (ODNs, 5′-AGG GTT TTC CXA GTC ACG ACG TT-3′ and 5′-TTC CXA GTC ACG ACG TT-3′, where X = 5fC, A and B in Table 1) were prepared using solid phase synthesis on an ABI 394 DNA synthesizer (Applied Biosystem, Foster City, CA, USA). 5-Formyl-dC-III-CE phosphoramidite was purchased from Glen Research (Sterling, VA, USA). Synthetic ODNs were deprotected in 30% ammonium hydroxide at room temperature for 17 h and then in 80% acetic acid at 20°C for 6 h. 5fC-containing ODNs were purified by semi-preparative HPLC, desalted by NAP-10 columns (GE Healthcare, Pittsburgh, PA, USA), and characterized by mass spectrometry (Table 1). Unmodified DNA oligodeoxynucleotides were purchased from IDT (Coralville, IA).

Synthesis, purification and characterization of DNA–histone and DNA–peptide cross-links. DNA–protein or DNA–peptide cross-links were synthesized and characterized as previously described (15). Briefly, to form DNA–histone cross-links, DNA 23-mers containing site-specific 5fC (A in

Table 1. Nucleobase sequence and mass spectrometry characterization of synthetic DNA strands

		Expected mass (Da)	Observed mass (Da)
A	5′-AGG GTT TTC CXA GTC ACG ACG TT-3′, where X = 5fC*	7058.6	7059.1
B	5′-TTC CXA GTC ACG ACG TT-3′, where X = 5fC	5149.3	5149.4
C	5′-AGG GTT TTC CXA GTC ACG ACG TT-3′, where X contains 5fC cross-link to the K residue of RPKPQQFFGLM-$CONH_2$ (see Figure 1A for structure)	8391.2	8415.0 $[M+Na]^+$
D	5′-AGG GTT TTC CXA GTC ACG ACG TT-3′, where X = 5fC cross-link to YGGFMTSEKSQTPLVTLFKNAIIKNAYKKGE	10507.0	10532.0 $[M+Na]^+$
E	5′-TTC CXA GTC ACG ACG TT-3′, where X contains 5fC cross-link to the K residue of RPKPQQFFGLM-$CONH_2$	6480.9	6504.8 $[M+Na]^+$
F	5′-AGG GTT TTC CXA GTC ACG ACG TT-3′, where X contains 5fC cross-link to the K residue of histone H4	18 284	N/A
G	5′-AGG GTT TTC CXA GTC ACG ACG TT-3′, where X contains 5fC cross-link to the K residue of histone H2A	21 038	N/A

*DNA 23-mers **A, C, D, F, G** are complementary to M13mp2 (6254–6276). **B** and **E** is a truncated version of the same sequence. Substance P (11-mer, RPKPQQFFGLM-$CONH_2$) and β-endorphin (31-mer, YGGFMTSEKSQTPLVTLFKNAII KNAYKKGE) were selected as model polypeptides based on our earlier work (15).

Table 1, 300 pmol) were incubated with equimolar amounts of recombinant histones H2A or H4 (New England Biolabs, Beverly, MA, USA) in 16 μl sodium phosphate buffer (4.5 mM, pH 7.4) for 3 h at 37°C. These conjugates were either used directly or subjected to $NaCNBH_3$ reduction as described below.

To convert transient Schiff base conjugates to stable DPCs amenable to gel purification, structural characterization, and kinetic analyses, samples were treated with $NaCNBH_3$ (4 μl of 100 mM solution) at 37°C overnight (Figure 1A). The reaction mixtures were heated at 80°C for 10 min and separated by 4–12% denaturing SDS PAGE. Gel-purified DNA–histone conjugates were desalted on Micro Bio-spin columns (Bio-Rad, Hercules, CA, USA) and characterized by HPLC-ESI^+-MS and MS^2 (15).

Stable DNA–peptide cross-links containing 11-mer peptide, RPKPQQFFGLM-$CONH_2$ and 31-mer peptide, YGGFMTSEKSQ TPLVTL FKNAIIKNAYKKGE were prepared analogously, with the exception that 10–30 molar excess of synthetic peptides were used. The reaction mixtures were heated at 90°C for 15 min and purified by 20% (w/v) denaturing PAGE containing 7 M urea. Gel-purified DNA–peptide conjugates were desalted by solid phase extraction on Sep-Pak C18 cartridges (Waters Corporation, Milford, MA, USA) and characterized by MALDI-TOF-MS using 3-HPA matrix (see Figure 1B for representative spectra). DNA–protein and DNA–peptide conjugates were further purified by denaturing PAGE as needed to achieve >98% purity.

Preparation of template–primer complexes for primer extension assays. Synthetic DNA 12-mers (500 pmol, 5′-AAC GTC GTG ACT-3′) were radiolabelled by incubating with γ-^{32}P ATP (1 μl, PerkinElmer Life Sciences, Boston, MA) and T4 PNK (20 units) in 20 μl 1 × T4 PNK buffer (New England Biolabs, Beverly, MA, USA) at 37°C for 1 h. The enzyme was deactivated by heating at 65°C for 10 min, and excess γ-^{32}P ATP was removed using Microspin G25 columns (GE Healthcare, Pittsburgh, PA, USA). Radiolabeled 12-mer primers (30 pmol) were mixed with 1.5 eq. of 23-mer DNA template strands containing site-specific DNA–protein or DNA–peptide cross-links (5′-AGG GTT TTC CXA GTC ACG ACG TT-3′, where X is adducted position) in 10 mM Tris–HCl buffer (pH 8.0) and 50 mM NaCl. The mixtures were heated at 90°C for 10 min and allowed to slowly cool down to room temperature overnight to generate primer–template complexes for primer extension assays. To digest the protein component of DPCs, primer–template complexes containing histone H4-5fC cross-link (2 pmol) were incubated with proteinase K (2.4 units) at 37°C for 48 h in the presence of 1× T4 PNK buffer (New England Biolabs, Beverly, MA, USA). The reaction mixtures were heated at 80°C for 2 h to deactivate the enzyme and then allowed to slowly cool down to room temperature overnight.

Primer extension assays using human TLS DNA polymerases. ^{32}P-labeled primer–template duplexes containing unmodified dC, 5-formyl-dC, DNA–peptide or DNA–protein crosslinks, or proteinase K digested DPCs (1 pmol) were incubated with human DNA polymerase η or κ (2 pmol) in the presence of 50 mM Tris–HCl (pH 7.5), 5 mM DTT, 5 mM $MgCl_2$, 50 mM NaCl, 100 μg/mL BSA and 10% glycerol (v/v) (total volume, 30 μl) at 37°C. The polymerization was initiated by the addition of all four dNTPs at 1 mM final concentration. Aliquots of the reaction mixtures (4 μl) were taken at pre-selected time points (0, 5, 30, 90, 180 min) and quenched with a gel loading buffer (20 mM EDTA in 95 % formamide containing 0.05 % bromophenol blue and xylene cyanol). The extension products were loaded onto 20 % denaturing PAGE containing 7 M urea, run at 80 W for 2.5 h in 1× TBE buffer, and visualized using Typhoon FLA 7000 phosphorimager (GH Healthcare, Pittsburgh, PA, USA).

Single nucleotide incorporation assays. ^{32}P-labeled primer–template complexes containing unmodified dC or 5fC cross-links to 11-mer peptide or 31-mer peptide (1.5 pmol) were incubated with human DNA polymerases (0.5 pmol of hPol η or 2 pmol of hPol κ) in the presence of 50 mM Tris–HCl (pH 7.5), 5 mM DTT, 5 mM $MgCl_2$, 50 mM NaCl, 100 μg/ml BSA and 10% glycerol (v/v) at 37°C. Individual dNTPs (50 μM final concentration) were added to initiate the polymerization in a final volume of 20 μl. Aliquots (4 μl) were quenched with gel loading buffer

after 0, 5, 15 or 30 min, and the extension products were resolved by 20% (w/v) denaturing PAGE containing 7 M urea and visualized using Typhoon FLA 7000 system.

Steady-state kinetic analyses. Steady-state kinetics for the incorporation of individual dNTPs opposite unmodified dC or 11-mer peptide cross-links (RPKPQQFFGLM-$CONH_2$) was examined by performing single nucleotide insertion assays in the presence of human DNA polymerases η or κ and increasing concentrations of specific dNTPs. primer–template duplexes (30 nM) were incubated with 0.4–1.0 nM hPol η or 1.2 nM hPol κ in the presence of individual dNTPs (5, 10, 25, 50, 100, 150, 250, 500 μM) for specified time periods (0–30 min). Primer extension products were visualized with a Typhoon FLA 7000 system and quantified by volume analysis using the ImageQuant TL 8.0 software (GE Healthcare). Steady-state kinetic parameters were calculated by nonlinear regression analysis using one-site hyperbolic fits in Prism 4.0 (Graphpad Software, La Jolla, CA, USA).

Sequencing and quantification of primer extension products using HPLC-ESI$^-$-MS and HPLC-ESI$^-$-MS^2. Biotinylated primers (Biotin-5′-$(T)_8$ AAC GTC GUG ACT-3′, purchased from IDT) were annealed to 17-mer oligonucleotide containing either unmodified dC (negative control) or 11-mer peptide–DNA cross-links (5′-TTC CXA GTC ACG ACG TT-3′, where X = dC or fC-Lys cross-link to the 11-mer peptide RPKPQQFFGLM-$CONH_2$). The resulting template–primer complexes (200 pmol) were incubated with hPol κ (50 pmol) or hPol κ (70 pmol) at 37°C for 6 h in the presence of 50 mM Tris–HCl (pH 7.5), 5 mM DTT, 5 mM $MgCl_2$, 50 mM NaCl, 100 μg/ml BSA, 10% glycerol (v/v) and 1 mM of each dNTPs in a total volume of 90 μl. Following the addition of 20 mM sodium phosphate buffer (pH 7.0) containing 150 mM NaCl (400 μl), biotinylated primer extension products were captured on streptavidin–sepharose high performance beads (200 μl, GE Healthcare). To release truncated primer extension products, streptavidin beads were washed with water (3 × 400 μl) and then incubated at 37°C for 4 h with uracil-DNA glycosylase (UDG, 20 units) in the presence of 50 mM Tris–HCl (pH 7.5), 1 mM EDTA and 1 mM DTT in a final volume of 500 μl. After removing the supernatant, the beads were washed with water (3 × 400 μl), and UDG-induced abasic sites were cleaved with hot piperidine (250 mM, 95°C for 1 h). The supernatants were dried under vacuum and reconstituted in 25 μl of water for HPLC–ESI$^-$-MS analysis.

HPLC–ESI-MS/MS sequencing and quantification of primer extension products was performed using a Dionex UltiMate 3000 HPLC coupled to a LTQ Orbitrap Velos mass spectrometer (Thermo Scientific, Watham, MA, USA) as described previously (25,26). A Zorbax 300SB-C18 column (150 × 0.5 mm, 5 μm, Agilent Technologies, Inc., Wilmington, CA, USA) was eluted at a flow rate of 15.0 μl/min using 15 mM ammonium acetate in water (A) and acetonitrile (B). Solvent composition was linearly increased from 2 to 20% B in 25 min. Mass spectrometry analyses were performed at a resolution of 60 000 and a scan range of m/z 300–2000. Relative quantification of various primer extension products was performed by comparing the corresponding HPLC–ESI-MS peak areas in extracted ion chromatograms. MS/MS fragmentation patterns were compared with the expected CID fragmentation from Mongo Oligo mass calculator version 2.06 available online (The RNA Institute, College of Arts and Sciences, State University of New York at Albany).

Construction and characterization of a plasmid containing a single DNA–peptide cross-link and its replication in human embryonic kidney (HEK) 293T cells. The 23-mer oligonucleotide (A in Table 1) containing $NaCNBH_3$ stabilized cross-link to the 11-mer peptide (RPKPQQFFGLM-$CONH_2$) was enzymatically ligated to a gapped single-stranded pMS2 plasmid carrying neomycin and ampicillin resistance genes, following methods reported earlier (24). A control plasmid containing native C at position X was constructed analogously. HEK 293T cells were grown to ∼90% confluency and transfected with 50 ng of the engineered plasmid constructs using Lipofectamine (Invitrogen, Carlsbad, CA, USA). Subsequently, the cells were allowed to grow at 37°C in 5% CO_2 for 2 days. Plasmid DNA was isolated, purified, and used to transform *Escherichia coli* cells. The transformants were analyzed by oligonucleotide hybridization followed by DNA sequence analysis, as reported earlier (24).

Molecular dynamics simulations

Initial models. Our initial models of human polymerase κ (hPol κ) in complex with unmodified DNA were based on our earlier work by Lior-Hoffmann *et al.* (30), which employed the PDB (31) entry 2OH2 crystal structure of human pol κ ternary complex (32). The initial model of human polymerase η (hPol η) in complex with unmodified DNA was based on the ternary crystal structure PDB entry 3MR2 (33). In this hPol η structure, three residues in the palm domain (residues 155−157) and two residues in the little finger domain (residues 411−412) were not resolved and thus were modeled into the polymerase.

We wished to study ternary complexes containing DNA–peptide cross-links at the templating base, human pol κ and η, and incoming dGTP or dATP nucleotides. To construct the DNA–peptide conjugates of interest, we first replaced the template base in our initial models with a C. This cytosine base was then modified at its C5 atom by conjugation to the side chain of the single lysine residue via a methylene linker. We then added three amino acid residues to this lysine to create the four amino acid peptide adduct NH_2-RPKP-$CONH_2$, and this peptide was energy minimized. Subsequently, we added one additional amino acid residue at a time to the C terminus, to create the 11-mer peptide RPKPQQFFGLM-$CONH_2$, with minimization at each step; the C terminus of the growing peptide was directed toward the 3′-end of the template strand to place it in the major groove, where the cytosine C5 atom is located in B-DNA. For the Watson-Crick partner, we remodeled the incoming dNTP to dGTP. These final models were named 11mer-dGTP in both pol η and pol κ. To study the mismatch of A with the modified C, we replaced the dGTP with N1-protonated dATP and aligned it with the templating C to form a C–A+ wobble pair with two hydrogen bonds

in pol η and pol κ; these two models are termed 11mer-dATP(η) and 11mer-dATP(κ). In addition, four analogous models without 11-mer peptide were constructed as controls, named dC-dGTP(κ), dC-dATP(κ), dC-dGTP(η) and dC-dATP(η). Thus, a total of eight polymerase models were prepared for the full MD simulations studies. Exploratory studies were also conducted for DPC models with incoming unprotonated dATP in pol η and pol κ.

Simulation methods. Molecular dynamics simulations were performed for each of the eight polymerase models. Following minimization and equilibration, we conducted ∼400 ns unrestrained production MD simulations using the AMBER14 package (34) with force field ff14SB (35,36) for proteins and incorporated modifications for DNA (37–39). We used the Joung-Cheatham model (40) for the Na^+ and Cl^- ions, and the TIP3P model (41) for water. For non-standard residues, the dGTP, dATP, N1-protonated dATP and DNA–protein cross-link lesion, parameters were assigned according to the GAFF (42) and the AMBERff14SB (35,36) force fields using the antechamber module of AMBER 14 (34). Partial charges for dGTP and dATP were taken from the literature (43–45); partial charges for the DNA–protein cross-link lesion and the N1-protonated dATP were determined using the restrained electrostatic potential (RESP) fit procedure at the HF/6-31G(d) level of theory using Gaussian 09 (46). Details on partial charge calculations and MD simulation protocol are given in the Methods Section in Supplementary Data.

Molecular topology and coordinate files for the initial polymerase models were constructed using the *tleap* module of AmberTools14 (34). Sufficient numbers of Na^+ ions were added to neutralize the system, and then additional Na^+ and Cl^- ions were added to bring the salt concentration close to the physiological value (0.15 M). The systems were solvated in the rectangular TIP3P water box with at least a 10 Å buffer, resulting in ∼70 Na^+ ions, 60 Cl^- ions and ∼21 000 water molecules.

Initial structures were modeled using Discovery Studio software v2.5.5 (Accelrys), and all molecular images and movies were generated with PyMOL (The PyMOL Molecular Graphic System, version 1.3x Schrodinger, LLC). Post-processing of all simulations was carried out using the CPPTRAJ (47) module of AMBER14 (34) and Curves+ (48).

RESULTS

Bypass of DNA–protein and DNA–peptide cross-links by human DNA polymerases *in vitro*

To induce DPC formation, 5fC containing DNA strands (A and B in Table 1) were incubated with recombinant human histones H2A or H4 or model polypeptides (RPKPQQFFGLM-$CONH_2$, YGGFMTSEKSQTPLVTLFKN AIIKNAYKKGE) The resulting DNA–protein/polypeptide conjugates were either used directly or stabilized by reduction in the presence of $NaCNBH_3$ (Figure 1). Reversible Schiff base DPCs were used directly, while reductively stabilized DPCs (D–G in Table 1) were purified by denaturing PAGE and characterized by ESI-MS or MALDI mass spectrometry (Table 1) (15).

To investigate the effects of C5-dC conjugated DNA–protein and DNA–peptide cross-links on DNA replication, primer–template complexes containing unmodified dC, 5fC, DNA–peptide, or DNA–protein cross-links (A, C, D, F, and G in Table 1) were subjected to *in vitro* replication in the presence of human DNA polymerases η or κ (enzyme: substrate = 2:1), and the primer extension products were analyzed by denaturing PAGE (Figures 2 and 3). These experiments revealed that the presence of reversible 5fC-histone cross-links dramatically reduces the amount of full length primer extension products in reactions catalyzed by human lesion bypass polymerases κ and η (Figure 2). However, kinetic parameters for polymerase bypass could not be established due to the reversible nature of Schiff base histone-DNA linkages (calculated half-life, 1.8 h) (15). Because such imino conjugates cannot be isolated in pure form and can spontaneously dissociate under physiological conditions, the template strands contain a mixture 5fC and 5fC-histone cross-links (15).

To allow for further mechanistic and kinetic studies, primer extension experiments were repeated using gel purified DPC templates that were stabilized by $NaCNBH_3$ reduction to form irreversible amino linkages (Figure 1). We found that reductively stabilized C5-dC DPC lesions to full size histone proteins completely inhibited DNA replication: no full length products were observed for histone H2A and H4 DPCs (Figure 3E and F). However, Pol κ and Pol η were able to extend the primer opposite smaller cross-links containing 11-mer peptide (RPKPQQFFGLM-$CONH_2$ and a 31-mer peptide (YGGFMTSEKSQTPLVTL FKNAIIKN AYKKGE) (Figure 3C and D). When the protein component of histone H4 DPC was digested to single amino acids using proteinase K, the ability of polymerase η to bypass the lesion was restored (Figure 3G).

To determine whether TLS polymerases Pol κ and Pol η replicate 5fC-peptide conjugates in an error free or error prone manner, single nucleotide incorporation assays with individual dNTPs were conducted (see example in Figure 4). PAGE analysis of the extension products has revealed that while hPol η inserted either one of the four dNTPs opposite 11-mer peptide-dC conjugates, pol κ predominantly incorporated an incorrect base, dA (Figure 4). Similar results were obtained for substrates containing a larger cross-link to 31-mer peptide, although incorporation efficiency was reduced (Supplementary Figure S1). These results provided initial evidence that 5fC mediated DNA–peptide lesions interfere with fidelity of TLS DNA polymerases.

To determine the catalytic efficiency for incorporation of G, A, C, and T opposite 5fC-peptide conjugates, steady state kinetics experiments were conducted using the same primer-substrate complexes (Figure 3) and hPol η or κ in the presence of individual dNTPs (5, 10, 25, 50, 100, 150, 250, 500 μM). The reaction mixtures were quenched at preselected time points (0–30 min). To ensure steady state conditions, 30–80 molar excess of DNA substrate over polymerases was employed, and time points were selected in such a way that < 35% of extension products were observed (see Supplementary Figure S2 for representative gel). Kinetic parameters (k_{cat}, K_m,Table 2) were calculated by plotting reaction velocities against the concentration of dNTPs using Michaelis–Menten equation (49).

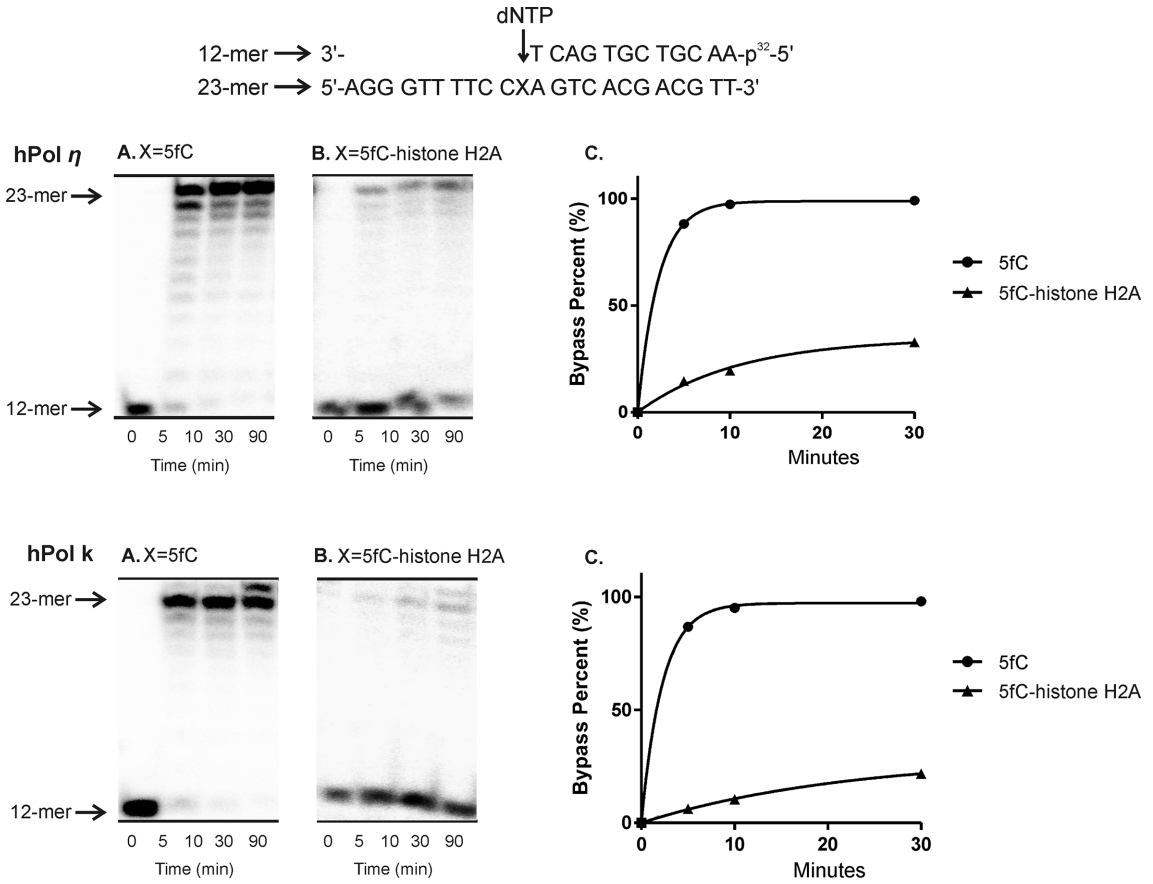

Figure 2. Primer extension assays for hPol η and hPol κ catalyzed bypass of reversible DNA–histone cross-links generated via a Schiff base formation between histone proteins and 5fC in DNA. 32P-labeled 12-mer primer was annealed with 23-mer template containing 5-formyl-dC (**A**), or 5-formyl-dC conjugated to histone H2A via Schiff base formation (**B**). Polymerase reactions were initiated by addition of a mixture of hPol η or κ and a mixture of dNTP, quenched at pre-selected time points, and loaded onto 20% denaturing PAGE. Percent bypass was plotted over time (**C**).

Table 2. Steady-state kinetics parameters for single nucleotide incorporation opposite unmodified dC or 11-mer peptide (RPKPQQFFGLM-$CONH_2$) conjugated to C5 position of cytosine by hPol η or κ

Polymerase	Template	Incoming nucleotide	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (μM^{-1} min^{-1})	*f
hPol η	dC	dATP	1.54 ± 0.42	354 ± 174	0.0043	0.013
		dGTP	2.50 ± 0.30	7.5 ± 3.3	0.33	1
		dCTP	0.52 ± 0.11	232 ± 98	0.0022	0.007
		dTTP	2.38 ± 0.90	427 ± 276	0.0056	0.017
	11-mer peptide	dATP	2.82 ± 0.32	59 ± 22	0.048	1.92
		dGTP	1.50 ± 0.15	61 ± 14	0.025	1
		dCTP	1.20 ± 0.13	179 ± 45	0.0066	0.26
		dTTP	1.12 ± 0.08	108 ± 21	0.01	0.4
hPol κ	dC	dGTP	0.075 ± 0.006	6.3 ± 2.7	0.012	1
	11-mer peptide	dATP	0.068 ± 0.008	80 ± 31	0.0008	1

*f (misinsertion frequency) = $(k_{cat}/K_m)_{incorrect\ dNTP}/(k_{cat}/K_m)_{correct\ dNTP}$.

The k_{cat}/K_m for incorporation of correct base (G) opposite unmodified dC and 5dC-11-mer peptide conjugate by hPol η were 0.33 and 0.025 μM^{-1} min^{-1}, respectively, indicating that the presence of 11-mer peptide conjugate strongly interfered with accurate DNA replication. Furthermore, the specificity constants for insertion of incorrect bases (A, C or T) opposite dC-11-mer peptide conjugate were 2–10 times higher than those observed for unmodified dC, suggesting that this lesion can be mutagenic. Overall, the preference order for nucleotide insertion opposite 5fC-11mer peptide DPC was A > G > T >C. hPol η preferentially incorporated A (incorrect base) opposite the peptide adduct, rather than the correct base (G), with misinsertion frequency (*f*) of 1.92 (Table 2). In reactions catalyzed by hPol κ, high fidelity was observed when replicating native DNA, incorporating only correct bases (G) opposite C (Table 2). However, only the incorrect base (A) was inserted opposite dC-11-mer peptide conjugate, although the catalytic

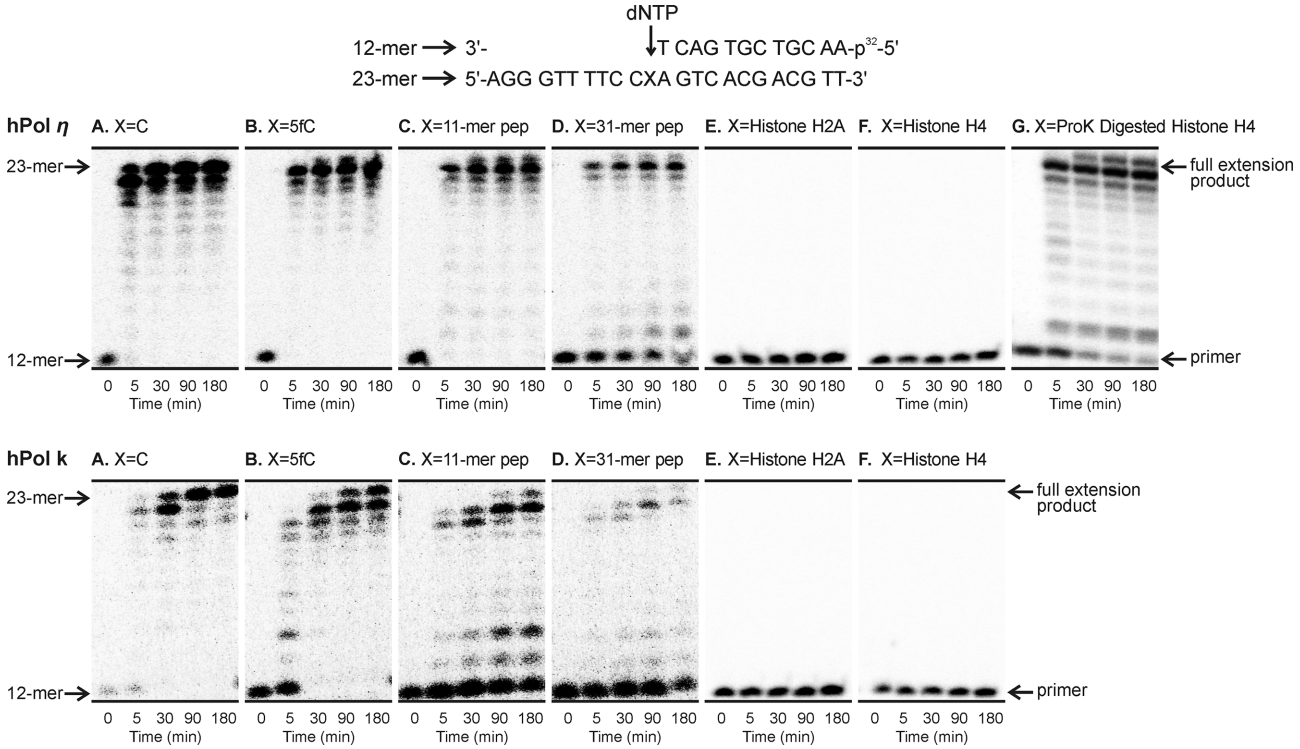

Figure 3. Primer extension assays for replication bypass of reductively stabilized DNA–protein or DNA–peptide cross-links by hPol η (top panel) and hPol κ (bottom panel). 32P-labeled 12-mer primer was annealed with 23-mer template containing unmodified dC (**A**), 5-formyl-dC (**B**), or C5-dC cross-links to 11-mer peptide RPKPQQFFGLM-CONH2 (**C**) 31-mer peptide YGGFMTSEKSQTPLVTLFKNAIIKNA YKKGE (**D**), full size histone H2A (**E**), histone H4 (**F**), or proteinase K digested histone H4 (**G**). Polymerase reactions were initiated by addition of DNA polymerases and a mixture of dNTP and quenched at pre-selected time points prior to loading onto 20% denaturing PAGE.

efficiency was 30- to 100-fold lower as compared to hPol η (Table 2).

To gain additional insight into the effects of dC-11-mer peptide cross-links on DNA replication, primer extension products were sequenced by HPLC-ESI-MS/MS (Figure 5A) (25,26,50). In brief, biotinylated primers (Biotin-5′-$(T)_8$ AAC GTC GUG ACT-3′) were annealed to 17-mer oligonucleotides (5′-TTC CXA GTC ACG ACG TT-3′) containing unmodified dC or dC-11-mer peptide crosslinks at position X. The resulting primer–template complexes were incubated with hPol η or κ in the presence of all four dNTPs, and the extension products were captured on streptavidin beads (Figure 5A). The extended primers were site-specifically cleaved with uracil DNA glycosylase/hot piperidine to release 9-mer oligonucleotide products that contain the region of interest and are amenable for sequencing by HPLC–ESI^--MS/MS (Figure 5A). HPLC–ESI^--FTMS analyses of the reaction mixtures using Orbitrap Velos mass spectrometer at $R = 60\,000$ and mass accuracy <10 ppm allowed for detection and relative quantification of primer extension products from peak areas in extracted ion chromatograms (Supplementary Figure S3), while their identities were confirmed by HPLC–ESI^--MS/MS (Figure 5B–D). In control experiments with unmodified DNA, both hPol η and κ preferentially generated error-free primer extension products (5′-pGACTGGGAA-3′, $[M–3H]^{3-} = 954.5$, Table 3). In contrast, replication bypass of dC-11-mer peptide conjugates by hPol η led to large numbers of C → T mutations (20.1 %) and (-1) deletions (12.4 %). Primer extension in the presence of hPol κ resulted in C →T transitions (13.9 %) and deletions 78.6 %, with only 7.5 % of error-free products (Table 3). No other replication products were detected. Overall, these results are in agreement with our steady-state kinetics data (Table 2), indicating that C5-cytosine conjugated DNA–peptide conjugates induce targeted C →T mutations. In addition, significant numbers of (–1) deletions were detected upon replication of DPC containing template (Table 3).

Replication of 5fC DNA–peptide cross-links in human cells

To determine whether 5fC mediated DNA–peptide cross-links can induce mutations in living cells, single stranded pMS2 plasmid containing site-specific 5fC DPC to an 11-mer peptide was constructed using the previously reported strategy (Figure 6A) (51,52). In brief, 23-mer oligonucleotide containing reductively stabilized 11-mer peptide to 5fC (C in Table 1) was enzymatically ligated into a gapped pMS2 plasmid, and the plasmids bearing site-specific DPC lesions were purified by agarose gel. Upon replication of DPC containing plasmid construct in HEK293T cells, we found that the number of progeny was comparable to that from a control plasmid, suggesting that the 5fC-peptide cross-links did not block DNA replication in human cells. However, sequencing analyses revealed that 11–16% of the progeny plasmids contained mutants, which included both targeted and semi-targeted mutations (Figure 6B). Consis-

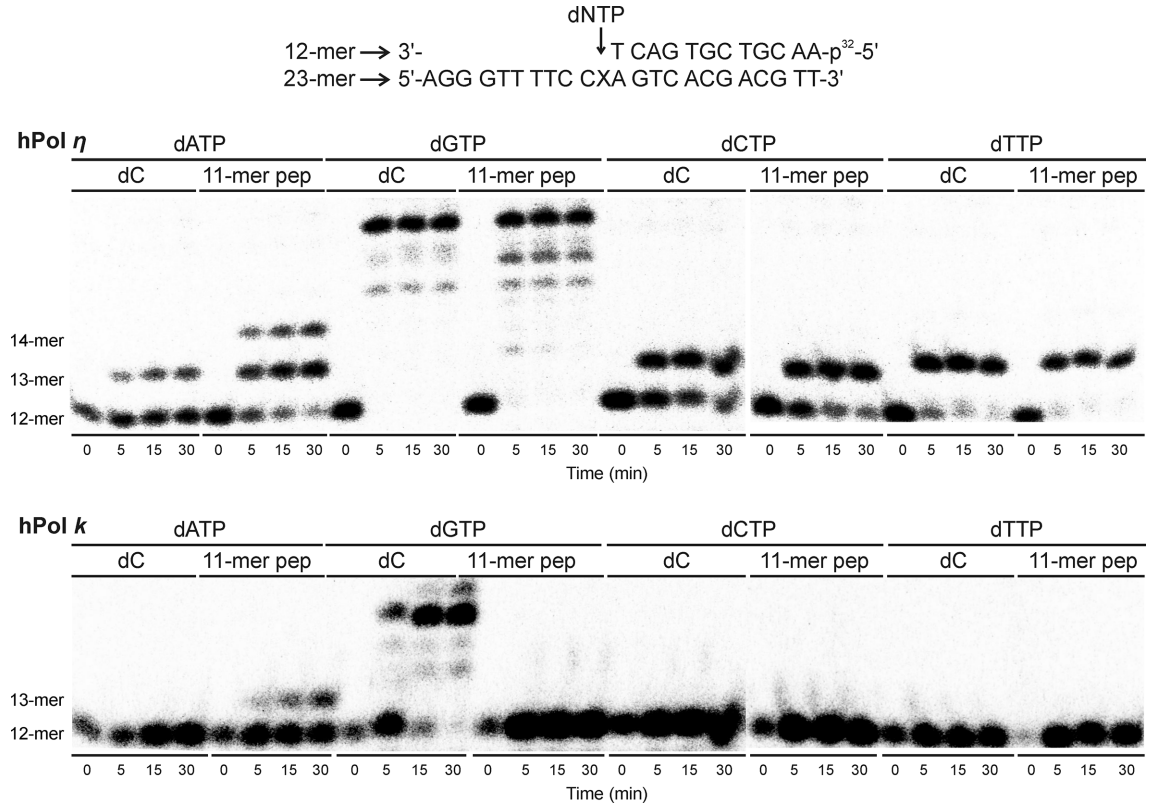

Figure 4. Single nucleotide insertion opposite unmodified dC (control) and 5fC-11 mer peptide cross-links (X = RPKPQQFFGLM-CONH2). template–primer complexes (top) were incubated with hPol η (top panel) or hPol κ (bottom panel) in the presence of 50 μM individual dNTP and then quenched at pre-selected time points.

Table 3. Sequencing and quantification of primer extension products by hPol η or κ using high resolution HPLC-ESI-MS/MS on Orbitrap Velos

	Biotin-5′-(T)8-AA CGT CGU GAC T-3′ 3′-TT GCA GCA CTG AXC CTT-5′				
Polymerase	Template	Extension product	Percent product	Base opposite X	Comment
hPol η	dC	GACTGGGAA	95.6	G	error free
		GACTAGGAA	1.9	A	C to T transition
		GACT_GGAA	2.5	_	deletion
	11-mer peptide	GACTGGGAA	67.5	G	error free
		GACTAGGAA	20.1	A	C to T transition
		GACT_GGAA	12.4	_	deletion
hPol κ	dC	GACTGGGAA	97.3	G	error free
		GACT_GGAA	2.7	_	deletion
	11-mer peptide	GACTGGGAA	7.5	G	error free
		GACTAGGAA	13.9	A	C to T transition
		GACT_GGAA	78.6	_	deletion

tent with our *in vitro* results (Figure 4, Tables 2 and 3), the major type of mutation observed was C → T transitions, which occurred at a frequency of 6.7%. In addition, we noted a low level of targeted deletions (< 2%) and C → G transversions (0.7%). Semi-targeted mutations 3′ and 5′ to the lesion site were observed at 7% frequency, in addition to 3–6 base deletions at 1.7% frequency. Overall, our results of adduct replication in human cells are consistent with the *in vitro* data (Tables 2 and 3), indicating that both hPol η and κ can insert an incorrect base (A) opposite 5fC mediated DNA–peptide conjugates, leading to C → T mutations.

Molecular dynamics

In order to gain structural understanding of the mutagenesis observed for 5fC-11mer peptide cross-links (Tables 2 and 3, Figure 6B), 400 ns molecular dynamics simulations were carried out. To help explain the preferential misinsertion of A opposite 5fC-peptide lesions, we created a molecular model of the 11-mer peptide (RPKPQQFFGLM-$CONH_2$) cross-linked to the cytosine template base in the human DNA polymerase η ternary complex with incoming dATP nucleotide (Figure 1B and Supplementary Figure S4). Our investigation has focused on the C–A+ mismatch structure, in which the N1 atom in the adenine of the dATP incoming nucleotide is protonated (Figure 7A). This wobble base

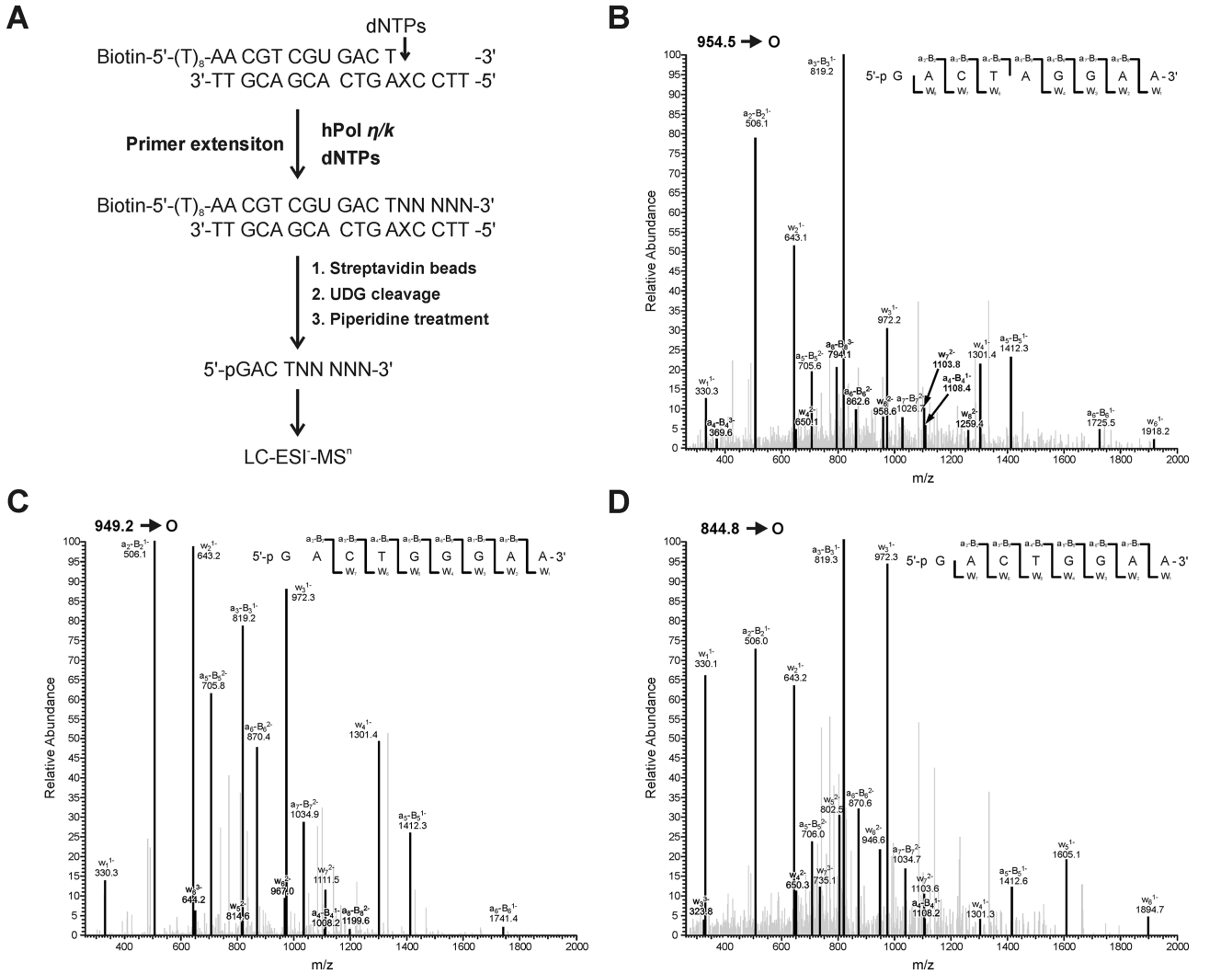

Figure 5. HPLC–ESI-MS/MS sequencing of primer extension products via streptavidin capture- nanoHPLC–ESI-MS/MS (**A**) and MS/MS spectra of primer extension products 5′-pGAC TGG GAA-3′ (**B**, error free), 5′-pGAC TAG GAA-3′ (**C**, C → T transition), and 5′-pGAC T_G GAA-3′ (**D**, –1 deletion).

pairing scheme affords two hydrogen bonds, namely C(N3)-A+(N6H2) and C(O2)-A+(N1H) with glycosidic bonds *anti* in both C and A+ (Figure 7A). Exploratory studies had indicated that a C–A mismatch structure without the N1 atom protonated and hence containing only one hydrogen bond, C(O2)-A(HN6), was much less stable in MD simulations.

Our simulations have revealed a stable (Supplementary Figure S5), nearly undistorted 11-mer peptide cross-linked to C, which fits well on the DNA major groove side and forms a C–A+ mismatched base-pair (Supplementary Figure S6 and Supplementary Table S1). Intriguingly, our single nucleotide incorporation data (Table 2) shows that the insertion of dA opposite the C5-dC cross-linked to the 11-mer peptide is more favorable than dA incorporation opposite unmodified C. Our simulations provide a possible structural explanation of this finding: the C–A+ mismatched base-pair is much more distorted, mainly through buckle, in the case of unmodified C. These structures suggest that the 11-mer peptide anchors the conjugated cytosine templating base to be ideally positioned for the formation of hydrogen bonds between the C and the A+ in the wobble pair. This anchoring is revealed by the inversion of the buckle (i.e. the peptide causes the buckle to invert direction and sign) as shown in Supplementary Figure S6 and Supplementary Table S1.

Our experimental data (Table 2) show that nucleotide insertion efficiency in hPol κ is much lower than by hPol η for both control and modified substrates. Furthermore, only mismatched base (A) is inserted opposite the peptide cross-link. To explain this data, we carried out MD simulations in pol κ with N1-protonated dATP, as in the pol η case, opposite to C5-dC cross-linked to 11-mer peptide (Figure 7B), which fits comfortably on the DNA major groove side. Our simulations revealed a stable (Supplementary Figure S5), modestly distorted C–A+ mismatched base-pair (Supplementary Table S1). The C–A+ mismatch is quite well formed, with two hydrogen bonds having occupancies >94 %.

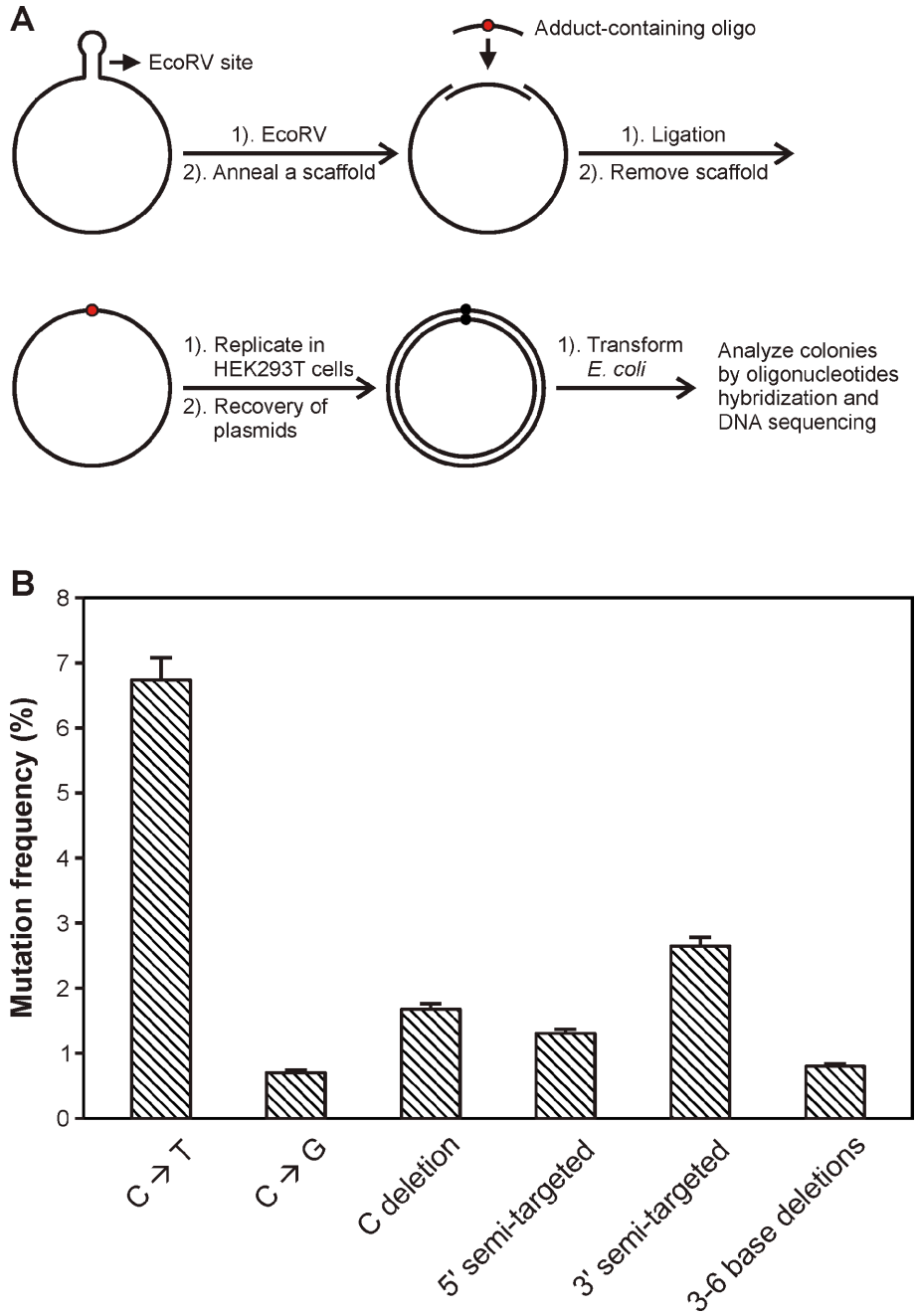

Figure 6. (**A**) Construction of DPC-containing pMS2 plasmid and its replication in HEK 293T cells. (**B**)Types and frequencies of mutations induced by the DNA–peptide cross-link in HEK293T cells. The data represent the mean and standard deviation of three independent experiments. Targeted C → T, C → G, C deletion, 3–6 base deletions as well as a variety of semi-targeted mutations near the lesion sites.

DISCUSSION

Because of their endogenous formation in human cells (15), it is of critical importance to establish the effects of 5fC mediated DNA–histone protein conjugates on DNA replication. Our data presented herein (Figures 2 and 3) indicate that the presence of reductively stabilized histone protein-5fC conjugates completely blocks human lesion bypass polymerases η and κ. These two polymerases were selected for the present study because of their ability to catalyze replication past DNA-polypeptide cross-links (24,25,29) and other bulky DNA adducts (28,51,53), as well as our preliminary data showing that hPol η and κ knockdown in human cells influences replication fidelity (Naldiga and Basu, unpublished observations). Specifically, we previously reported that model DPCs generated by Huisgen cycloaddition reaction between alkyne-containing DNA (at the C5 position of dT) and azide-functionalized protein completely blocked DNA replication (29).

Although reversible histone protein-5fC Schiff base lesions also blocked DNA replication *in vitro* (Figure 2), some

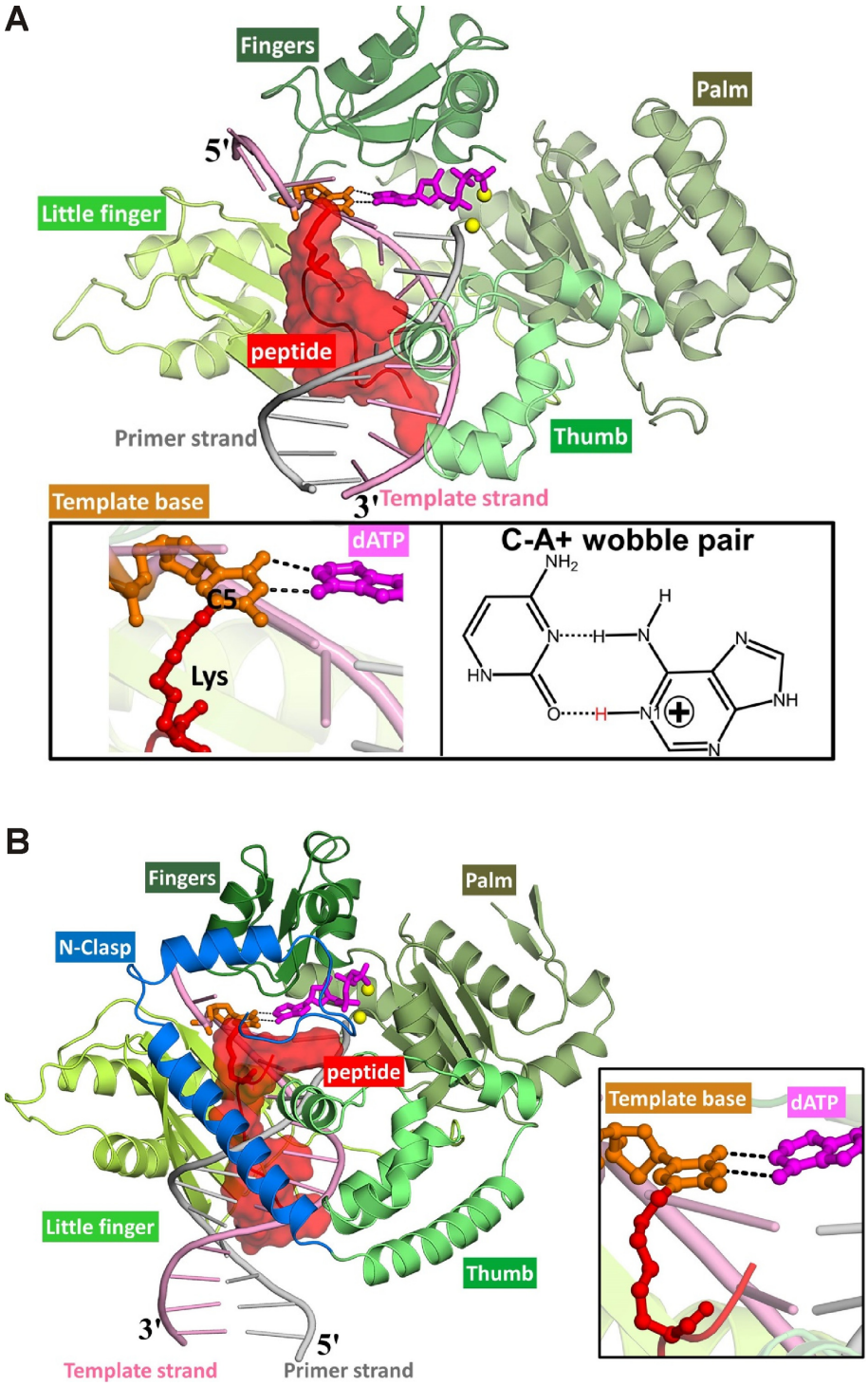

Figure 7. (**A**) The most representative structure from the 400 ns MD simulation of the pol η-DNA ternary complex, containing the 11-mer peptide cross-linked to the templating C opposite to the N1 protonated incoming dATP. Inset: (left), a zoomed-in view of the 11-mer peptide cross-linked to the templating C opposite the incoming dATP; (right), hydrogen bonding scheme of the C–A+ wobble pair. The protonated hydrogen at N1 in adenine is shown in red. A rotating view is given in the Supplementary Movie (Movie S1). (**B**) The most representative structure from the 400 ns MD simulation of the pol κ–DNA ternary complex, containing the 11-mer peptide cross-linked to the templating C opposite to the N1 protonated dATP. Inset: a zoomed-in view. A rotating view is given in the Supplementary Movie (Movie S2). In both structures, the complex is displayed in cartoon diagram except for the dATP and the cross-link-containing template C, which are in stick, and Mg^{2+} ions which are shown as yellow balls. The 11-mer peptide is shown in red surface. The two hydrogen bonds between the template base and incoming nucleotide are shown with dotted lines.

polymerase bypass was observed, probably due to transient protein dissociation from DNA. Therefore, endogenous formation of DNA–histone conjugates at 5fC marks is anticipated to interfere with DNA replication, unless a separate cellular mechanism facilitates their bypass and/or removal. Previous studies have revealed that the histone octamer can be released from the parental DNA by the action of histone chaperone FACT (facilitates chromatin transcription) which collaborates with MCM helicase (54,55). The loss of the histone octamer from the parental DNA ahead of the replication fork is accompanied by the dissociation of H3/H4 from H2A/H2B. FACT is a two subunit complex containing SPT16 and SSRP1 (structure specific recognition protein) capable of binding to both H2A/H2B and H3/H4 (56). Future studies will establish whether the same protein chaperones can dissociate reversibly bound histones from DNA in order to prevent mutagenesis and to allow for replication fork progression.

Alternatively, the protein component of DPC may be proteolytically degraded by specialized proteases such as Wss1/Spartan to form the corresponding DNA-small peptide conjugates (57,58). Recent reports indicate that Spartan can be recruited to stalled replication forks at the site of DPCs (59). The resulting DNA–peptide cross-links may be repaired by nucleotide excision pathway or be bypassed by translesion synthesis polymerases (9,60). We found that 5fC conjugates to 11-mer and 31-mer peptides can be readily bypassed by DNA polymerases η and κ (Figure 3C and D), although with a lower efficiency as compared with unmodified dC (Table 2). As compared to similar lesions containing a peptide cross-link to 7-deaza-dG or the N^2 position of dG (25,26,53), 5fC-peptide lesions represent a weaker block to DNA replication, allowing for their bypass.

Although lesion bypass by TLS polymerases avoids the deleterious consequence of stalled replication forks, this damage-tolerance mechanism may be error-prone, and cell survival can be associated with an increased risk of mutagenesis and carcinogenesis (51,61). We found that the presence of reductively stabilized 5fC-peptide lesions caused targeted C → T transitions and deletion mutations both *in vitro* (Figure 4, Tables 2 and 3) and in human cells (Figure 6B). The mutation frequency was lower in living cells, which is not surprising since human cells are proficient in DNA repair and contain many DNA polymerases that can cooperate with each other using 'polymerase switching' (62). Interestingly, *in vivo* mutagenesis studies also revealed both 3′ and 5′ semi-targeted mutations (Figure 6B), probably a result of peptide interfering with nucleotide incorporation by DNA polymerases. Semi-targeted mutations were also noted in our previous study with DNA–peptide cross-links conjugated to the N7 position of guanine (24).

It is not surprising that the results of *in vitro* and *in vivo* replication experiments for DNA–peptide adducts differ. In our previous studies, we found that other bulky DNA lesions that block DNA polymerases in vitro only marginally reduce translesion synthesis (TLS) in cells. For example, the C8-dG adduct formed by the carcinogen 3-nitrobenzanthrone was a strong replication block of hpol η, κ, ι and Rev1 *in vitro* (63), but when replicated in HEK 293T cells, allowed for 80% lesion bypass (64). Likewise, when we investigated the replicative properties of another DNA–peptide crosslink (DpC) in which the peptide was covalently attached to 7-deazaguanine, TLS in HEK293T cells was 76 ± 3% (24) as compared to significant blockage of both hpol η and κ by the same lesion *in vitro* (26). In a cellular system, multiple DNA polymerases can cooperate with each other to allow for facile bypass of a lesion that blocks single DNA polymerases *in vitro*.

Molecular modeling and molecular dynamics simulations of template–primer complexes revealed that modified C forms a stable mismatch with dATP in the active sites of polymerases η and κ, providing a plausible structural origin for C → T mutations. The peptide is stably positioned on the DNA major groove side when cross-linked to the templating C in both cases. In our model of hPol η ternary complex, the 11-mer peptide anchors the templating conjugated C in the spacious active site, so as to optimize the hydrogen bonding geometry when the peptide is present (Figure 7 and Supplementary Figure S6). The observed preference for incorporation of dATP rather than the correct base (dGTP) opposite C that is cross-linked to 11-mer peptide by hPol η can be explained by the weakened dC-dGTP base pair, so that only one full and two half occupancy hydrogen bonds can be formed (Supplementary Figure S7).

Our observation that hPol η is more efficient at bypassing 5fC DNA–peptide cross-links as compared with hPol κ is in agreement with previous experiments performed with other DPCs and structurally related lesions (25,26,65). Structurally, both human (33) and yeast (66) pol η have a large active site cleft with the fingers domain elevated, so that two nucleotides can be accommodated at the same time, notably the two covalently linked thymines in the *cis-syn* cyclobutane pyrimidine dimer (CPD). In contrast, pol κ has a constricted active site cleft with the fingers lowered onto the templating base, so that an additional nucleotide cannot be accommodated (32) (Supplementary Figure S8A). Furthermore, pol κ possesses the unique L-shaped N-clasp at its N-terminus, composed of short (αN1) and long (αN2) α helices, with the αN2 helical axis diagonally crossing the DNA duplex. With this N-clasp, pol κ encircles the DNA by holding the catalytic core (fingers, palm and thumb subdomains) and little finger together on the major groove side; therefore, the major groove side is occluded (Supplementary Figure S8B). On the other hand, pol κ contains a gap between the catalytic core and little finger on the DNA minor groove side that pol η lacks (Supplementary Figure S8C). The absence of this opening in pol η allows the newly synthesized duplex to retain a stable B-form, 'splinted' by the ungapped polymerase (33), even in the presence of a CPD lesion. These structural differences between pol η and pol κ may help explain our experimental data on a molecular level. Overall, the lower nucleotide insertion efficiency of pol κ relative to pol η (Table 2) can be rationalized by its more constricted active site region and the presence of the major groove-occluding N-clasp in pol κ, which present steric obstacles in nucleotide insertion and translocation.

In addition to C →T transitions, significant numbers of –1 deletion products were detected by LC–MS-based strategy (Table 3). Deletion mutations were a significant feature in the primer extension experiments for both pol η (12.4%) and pol κ (78.6%) in the presence of the peptide adduct (Table 3). The structure of pol η can provide an insight into

this observation. With an active site region that can accommodate the two nucleotides of the CPD lesion (Supplementary Figure S9A) (33), the occurrence of a deletion in the primer extension data can be explained by a mechanism in which the adducted 3′-nucleotide is paired with the incoming dNTP, while the 5′ neighboring base on the templating strand is skipped. This is analogous to the Type II crystal structure of Dpo4 (61), where two nucleotides occupied the spacious Dpo4 active site and the incoming dNTP is aligned with the 5′-templating base, while its 3′-neighbor is skipped (Supplementary Figure S9B). This structure provides a molecular view of a -1 deletion, and has been previously observed for 1, N^2-ε-guanine (50) and an abasic site (61) in the active site of Dpo4. While no such -1 deletion crystal structure is available for pol η, a similar mechanism is plausible for pol η, in view of the expanded active site similarity.

In contrast, pol κ has a more constricted active site and thus must use a different mechanism to induce -1 deletions in the presence of a peptide lesion. For instance, the adducted nucleotide may assume a 'slipped' or extrahelical conformation (50,67). Alternatively, slippage may occur at the subsequent nucleotide insertion opposite the adduct (50,67). For example, pol κ has been shown to use a template-slippage deletion mechanism (68). A single bulged out base that is skipped would be responsible for the deletion, but the exact location of the bulge is unknown. For example, in *E.coli* DNA pol II which does not have an expanded active site region, multiple looped out structures have been noted (69). The generation of one base or two base deletion mutations by TLS polymerases is well documented, especially in bypassing bulky DNA adducts (50). Our primer extension data for pol κ in the presence of the peptide adduct also revealed significant incorporation of the normal incoming dGTP (Table 3). Our simulations with ternary complex-containing dGTP and the 11-mer peptide in pol κ showed a well-aligned C-dGTP Watson–Crick pair, whose extension could explain this observation (Supplementary Figure S6 and Supplementary Table S1).

Our modeling studies of C–A mismatch mutations as well as our above discussion of slippage mechanisms consider template-directed nucleotide incorporation; moreover, our work reveals that template-directed dATP incorporation via wobble pairing is feasible with the bulky peptide cross-linked to the templating C. However, it is also possible that some C–A mismatches or deletion events occur via a non-instructive, untemplated mechanism, governed by the 'A' rule; this rule determined that 'damaged nucleotides produce 'non-instructive' sites at which normal base pairing was not possible, and at which As were incorporated due to inherent polymerase preference' (70,71). This mechanism is a consideration in the bypass of bulky lesions.

Together with our earlier findings (26,29), these results indicate that the cross-linking site with DNA has a strong impact on the biological outcomes resulting from DNA–peptide lesions. C5-Thymidine peptide conjugates examined in our earlier work (25,29) induced high numbers of frameshift mutations and base substitutions, while the TLS bypass of a peptide conjugated to the C7 position of 7-deaza-dG was extremely inefficient, but essentially error-free (26). The C5-dC adducts examined here are less blocking and more mutagenic than C5-dT and 7-deaza-dG lesions examined in these earlier studies, inducing large numbers of C to T transitions (Figure 4 and Table 2). Similarly, Minko *et al.* have shown that peptide cross-linked to the N^2 position of dG completely blocked DNA replication, while the same peptide at the N^6 position of dA was bypassed, introducing small amounts of mutations (53,72). However, every full size DNA–protein conjugate examined completely blocked DNA replication as shown in (25,26,29) and in Figures 2 and 3E, F.

Taken together, our data suggest that reversible conjugation between 5fC in DNA and histone proteins has a profound effect on DNA replication catalyzed by lesion bypass polymerases Pol κ and Pol η. Future studies are needed to establish whether such endogenous DPC contribute to C to T mutations frequently observed at CpG islands within the human genome (73), to test their effects on additional DNA polymerases, and to identify cellular mechanism and protein factors that prevent polymerase stalling of these super-bulky lesions observed at epigenetic marks of DNA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Xun Ming and Dr Peter W. Villalta (University of Minnesota) for their help with mass spectrometry analysis, and Robert Carlson (University of Minnesota) for his help with figure preparation.

FUNDING

National Institute of Environmental Health Sciences [R01 ES-023350 to N.T., R01 ES-025987 to S.B.]; National Cancer Institute [R01 CA-075449 to S.B.]; Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation [MCB060037 to S.B.]; NYU IT High Performance Computing Resources and Services; Wayland E. Noland Graduate Student fellowship from the Chemistry Department at the University of Minnesota (in part). Funding for open access charge: National Institute of Environmental Health Sciences.

Conflict of interest statement. None declared.

REFERENCES

1. Barker,S., Weinfeld,M. and Murray,D. (2005) DNA–protein crosslinks: their induction, repair, and biological consequences. *Mutat. Res.*, **589**, 111–135.
2. Ide,H., Shoulkamy,M.I., Nakano,T., Miyamoto-Matsubara,M. and Salem,A.M.H. (2011) Repair and biochemical effects of DNA–protein crosslinks. *Mutat. Res.*, **711**, 113–122.
3. Tretyakova,N.Y., Groehler,A. and Ji,S. (2015) DNA–protein cross-links: formation, structural identities, and biological outcomes. *Acc. Chem. Res.*, **48**, 1631–1644.
4. Barker,S., Weinfeld,M., Zheng,J., Li,L. and Murray,D. (2005) Identification of mammalian proteins cross-linked to DNA by ionizing radiation. *J. Biol. Chem.*, **280**, 33826–33838.
5. Speit,G., Schütz,P. and Merk,O. (2000) Induction and repair of formaldehyde-induced DNA–protein crosslinks in repair-deficient human cell lines. *Mutagenesis*, **15**, 85–90.

6. Loeber,R.L., Michaelson-Richie,E.D., Codreanu,S.G., Liebler,D.C., Campbell,C.R. and Tretyakova,N.Y. (2009) Proteomic analysis of DNA−protein cross-linking by antitumor nitrogen mustards. *Chem. Res. Toxicol.*, **22**, 1151–1162.
7. Groehler,A., Villalta,P.W., Campbell,C. and Tretyakova,N. (2016) Covalent DNA–protein cross-linking by phosphoramide mustard and nornitrogen mustard in human cells. *Chem. Res. Toxicol.*, **29**, 190–202.
8. Sanchez,A.M., Minko,I.G., Kurtz,A.J., Kanuri,M., Moriya,M. and Lloyd,R.S. (2003) Comparative evaluation of the bioreactivity and mutagenic spectra of acrolein-derived α-HOPdG and γ-HOPdG regioisomeric deoxyguanosine adducts. *Chem. Res. Toxicol.*, **16**, 1019–1028.
9. Baker,D.J., Wuenschell,G., Xia,L., Termini,J., Bates,S.E., Riggs,A.D. and O'Connor,T.R. (2007) Nucleotide excision repair eliminates unique DNA–protein cross-links from mammalian cells. *J. Biol. Chem.*, **282**, 22592–22604.
10. Tretyakova,N.Y., Michaelson-Richie,E.D., Gherezghiher,T.B., Kurtz,J., Ming,X., Wickramaratne,S., Campion,M., Kanugula,S., Pegg,A.E. and Campbell,C. (2013) DNA-reactive protein monoepoxides induce cell death and mutagenesis in mammalian cells. *Biochemistry*, **52**, 3171–3181.
11. Wu,F.-Y., Lee,Y.-J., Chen,D.-R. and Kuo,H.-W. (2002) Association of DNA–protein crosslinks and breast cancer. *Mutat. Res.*, **501**, 69–78.
12. Izzotti,A., Cartiglia,C., Taningher,M., De Flora,S. and Balansky,R. (1999) Age-related increases of 8-hydroxy-2-deoxyguanosine and DNA–protein crosslinks in mouse organs. *Mutat. Res.*, **446**, 215–223.
13. Henle,E.S. and Linn,S. (1997) Formation, prevention, and repair of DNA damage by iron/hydrogen peroxide. *J. Biol. Chem.*, **272**, 19095–19098.
14. Voitkun,V. and Zhitkovich,A. (1999) Analysis of DNA–protein crosslinking activity of malondialdehyde in vitro. *Mutat. Res.*, **424**, 97–106.
15. Ji,S., Shao,H., Han,Q., Seiler,C.L. and Tretyakova,N.Y. (2017) Reversible DNA–protein cross-linking at epigenetic DNA marks. *Angew. Chem. Int. Ed.*, **56**, 14130–14134.
16. Li,F., Zhang,Y., Bai,J., Greenberg,M.M., Xi,Z. and Zhou,C. (2017) 5-Formylcytosine yields DNA–protein cross-links in nucleosome core particles. *J. Am. Chem. Soc.*, **139**, 10617–10620.
17. Ito,S., Shen,L., Dai,Q., Wu,S.C., Collins,L.B., Swenberg,J.A., He,C. and Zhang,Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
18. Pfaffeneder,T., Hackner,B., Truß,M., Münzel,M., Müller,M., Deiml,C.A., Hagemeier,C. and Carell,T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed.*, **123**, 7146–7150.
19. Song,C.-X., Szulwach,Keith E., Dai,Q., Fu,Y., Mao,S.-Q., Lin,L., Street,C., Li,Y., Poidevin,M., Wu,H. *et al.* (2013) Genome-wide profiling of 5-Formylcytosine reveals its roles in epigenetic priming. *Cell*, **153**, 678–691.
20. Gackowski,D., Zarakowska,E., Starczak,M., Modrzejewska,M. and Olinski,R. (2015) Tissue-Specific differences in DNA modifications (5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxylcytosine and 5-hydroxymethyluracil) and their interrelationships. *PLoS One*, **10**, e0144859.
21. Bachman,M., Uribe-Lewis,S., Yang,X., Burgess,H.E., Iurlaro,M., Reik,W., Murrell,A. and Balasubramanian,S. (2015) 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.*, **11**, 555–557.
22. Spruijt,Cornelia G., Gnerlich,F., Smits,Arne H., Pfaffeneder,T., Jansen,Pascal W.T.C., Bauer,C., Münzel,M., Wagner,M., Müller,M., Khan,F. *et al.* (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, **152**, 1146–1159.
23. Neri,F., Incarnato,D., Krepelova,A., Rapelli,S., Anselmi,F., Parlato,C., Medana,C., Dal Bello,F. and Oliviero,S. (2015) Single-base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA methylation dynamics. *Cell Rep.*, **10**, 674–683.
24. Pande,P., Ji,S., Mukherjee,S., Schärer,O.D., Tretyakova,N.Y. and Basu,A.K. (2017) Mutagenicity of a model DNA–peptide cross-link in human cells: roles of translesion synthesis DNA polymerases. *Chem. Res. Toxicol.*, **30**, 669–677.
25. Wickramaratne,S., Boldry,E.J., Buehler,C., Wang,Y.-C., Distefano,M.D. and Tretyakova,N.Y. (2015) Error-prone translesion synthesis past DNA–peptide cross-links conjugated to the major groove of DNA via C5 of thymidine. *J. Biol. Chem.*, **290**, 775–787.
26. Wickramaratne,S., Ji,S., Mukherjee,S., Su,Y., Pence,M.G., Lior-Hoffmann,L., Fu,I., Broyde,S., Guengerich,F.P., Distefano,M. *et al.* (2016) Bypass of DNA–protein cross-links conjugated to the 7-Deazaguanine position of DNA by translesion synthesis polymerases. *J. Biol. Chem.*, **291**, 23589–23603.
27. 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.
28. Vaisman,A. and Woodgate,R. (2017) Translesion DNA polymerases in eukaryotes: what makes them tick? *Crit. Rev. Biochem. Mol. Biol.*, **52**, 274–303.
29. Yeo,J.E., Wickramaratne,S., Khatwani,S., Wang,Y.-C., Vervacke,J., Distefano,M.D. and Tretyakova,N.Y. (2014) Synthesis of site-specific DNA–protein conjugates and their effects on DNA replication. *ACS Chem. Biol.*, **9**, 1860–1868.
30. Lior-Hoffmann,L., Wang,L., Wang,S., Geacintov,N.E., Broyde,S. and Zhang,Y. (2012) Preferred WMSA catalytic mechanism of the nucleotidyl transfer reaction in human DNA polymerase kappa elucidates error-free bypass of a bulky DNA lesion. *Nucleic Acids Res.*, **40**, 9193–9205.
31. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
32. Lone,S., Townson,S.A., Uljon,S.N., Johnson,R.E., Brahma,A., Nair,D.T., Prakash,S., Prakash,L. and Aggarwal,A.K. (2007) Human DNA polymerase kappa encircles DNA: implications for mismatch extension and lesion bypass. *Mol. Cell*, **25**, 601–614.
33. Biertumpfel,C., Zhao,Y., Kondo,Y., Ramon-Maiques,S., Gregory,M., Lee,J.Y., Masutani,C., Lehmann,A.R., Hanaoka,F. and Yang,W. (2010) Structure and mechanism of human DNA polymerase eta. *Nature*, **465**, 1044–1048.
34. Case,D.A., Darden,T.A., Cheatham,T.E. 3rd, Simmerling,C.L., Wang,J., Duke,R.E., Luo,R., Walker,R.C., Zhang,W., Merz,K.M. *et al.* (2014) *AMBER 14 [computer software]*. University of California, San Francisco.
35. Maier,J.A., Martinez,C., Kasavajhala,K., Wickstrom,L., Hauser,K.E. and Simmerling,C. (2015) ff14SB: Improving the cccuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.*, **11**, 3696–3713.
36. Hornak,V., Abel,R., Okur,A., Strockbine,B., Roitberg,A. and Simmerling,C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, **65**, 712–725.
37. Perez,A., Marchan,I., Svozil,D., Sponer,J., Cheatham,T.E. 3rd, Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.
38. Zgarbova,M., Luque,F.J., Sponer,J., Cheatham,T.E 3rd, Otyepka,M. and Jurecka,P. (2013) Toward improved description of DNA Backbone: Revisiting epsilon and zeta torsion force field parameters. *J. Chem. Theory Comput.*, **9**, 2339–2354.
39. Cheatham,T.E. 3rd and Case,D.A. (2013) Twenty-five years of nucleic acid simulations. *Biopolymers*, **99**, 969–977.
40. Joung,I.S. and Cheatham,T.E. 3rd (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, **112**, 9020–9041.
41. Jorgensen,W.L., Chandreskhar,J., Madura,J.D., Imprey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
42. Wang,J., Wolf,R.M., Caldwell,J.W., Kollman,P.A. and Case,D.A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.
43. Perlow,R.A. and Broyde,S. (2001) Evading the proofreading machinery of a replicative DNA polymerase: induction of a mutation by an environmental carcinogen. *J. Mol. Biol.*, **309**, 519–536.
44. Perlow,R.A. and Broyde,S. (2002) Toward understanding the mutagenicity of an environmental carcinogen: structural insights into nucleotide incorporation preferences. *J. Mol. Biol.*, **322**, 291–309.
45. Jia,L., Geacintov,N.E. and Broyde,S. (2008) The N-clasp of human DNA polymerase kappa promotes blockage or error-free bypass of adenine- or guanine-benzo[a]pyrenyl lesions. *Nucleic Acids Res.*, **36**, 6571–6584.

46. Frisch,M.J., Trucks,G.W., Schlegel,H. B., Scuseria,G. E., Robb,M. A., Cheeseman,J. R., Scalmani,G., Barone,V., Mennucci,B., Petersson,G. A. *et al.* (2013) *Gaussian 09 [computer software]*. Gaussian, Inc.
47. Roe,D.R. and Cheatham,T.E. (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory Data. *J. Chem. Theory Comput.*, **9**, 3084–3095.
48. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
49. Fersht,A. (1999) *Structure and Mechanism in Protein Science*. W. H. Freeman & Co., NY, pp. 103–131.
50. Zang,H., Goodenough,A.K., Choi,J.-Y., Irimia,A., Loukachevitch,L.V., Kozekov,I.D., Angel,K.C., Rizzo,C.J., Egli,M. and Guengerich,F.P. (2005) DNA adduct bypass polymerization by sulfolobus solfataricus DNA polymerase Dpo4: analysis and crystal structures of mulple base pair substitutions and frameshift products with the addut 1,N2-ethenoguanine. *J. Biol. Chem.*, **280**, 29750–29764.
51. Colis,L.C., Raychaudhury,P. and Basu,A.K. (2008) Mutational specificity of γ-radiation-induced guanine−thymine and thymine−guanine intrastrand cross-links in mammalian cells and translesion synthesis past the guanine−thymine lesion by human DNA polymerase η. *Biochemistry*, **47**, 8070–8079.
52. Bose,A., Surugihalli,C., Pande,P., Champeil,E. and Basu,A.K. (2016) Comparative error-free and error-prone translesion synthesis of N2-2′-deoxyguanosine adducts formed by mitomycin C and its metabolite, 2,7-Diaminomitosene, in human cells. *Chem. Res. Toxicol.*, **29**, 933–939.
53. Minko,I.G., Yamanaka,K., Kozekov,I.D., Kozekova,A., Indiani,C., O'Donnell,M.E., Jiang,Q., Goodman,M.F., Rizzo,C.J. and Lloyd,R.S. (2008) Replication bypass of the acrolein-mediated deoxyguanine DNA–peptide cross-links by DNA polymerases of the DinB family. *Chem. Res. Toxicol.*, **21**, 1983–1990.
54. Tan,B.C., Chien,C.T., Hirose,S. and Lee,S.C. (2006) Functional cooperation between FACT and MCM helicase facilitates initiation of chromatin DNA replication. *EMBO J.*, **25**, 3975–3985.
55. Belotserkovskaya,R., Oh,S., Bondarenko,V.A., Orphanides,G., Studitsky,V.M. and Reinberg,D. (2003) FACT facilitates transcription-dependent nucleosome alteration. *Science*, **301**, 1090–1093.
56. Ransom,M., Dennehey,B.K. and Tyler,J.K. (2010) Chaperoning histones during DNA replication and repair. *Cell*, **140**, 183–195.
57. Duxin,Julien P., Dewar,James M., Yardimci,H. and Walter,Johannes C. (2014) Repair of a DNA–protein crosslink by replication-coupled proteolysis. *Cell*, **159**, 346–357.
58. Stingele,J., Schwarz,Michael S., Bloemeke,N., Wolf,Peter G. and Jentsch,S. (2014) A DNA-Dependent protease involved in DNA–protein crosslink repair. *Cell*, **158**, 327–338.
59. Vaz,B., Popovic,M., Newman,J.A., Fielden,J., Aitkenhead,H., Halder,S., Singh,A.N., Vendrell,I., Fischer,R., Torrecilla,I. *et al.* (2016) Metalloprotease SPRTN/DVC1 orchestrates replication-coupled DNA–protein crosslink repair. *Mol. Cell*, **64**, 704–719.
60. Minko,I.G., Zou,Y. and Lloyd,R.S. (2002) Incision of DNA–protein crosslinks by UvrABC nuclease suggests a potential repair pathway involving nucleotide excision repair. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 1905–1909.
61. Ling,H., Boudsocq,F., Woodgate,R. and Yang,W. (2001) Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell*, **107**, 91–102.
62. Friedberg,E.C., Lehmann,A.R. and Fuchs,R.P.P. (2005) Trading places: how do DNA polymerases switch during translesion DNA synthesis? *Mol. Cell*, **18**, 499–505.
63. Tokarsky,E.J., Gadkari,V.V., Zahurancik,W.J., Malik,C.K., Basu,A.K. and Suo,Z. (2016) Pre-steady-state kinetic investigation of bypass of a bulky guanine lesion by human Y-family DNA polymerases. *DNA Repair (Amst.)*, **46**, 20–28.
64. Pande,P., Malik,C.K., Bose,A., Jasti,V.P. and Basu,A.K. (2014) Mutational analysis of the C8-guanine adduct of the environmental carcinogen 3-nitrobenzanthrone in human cells: critical roles of DNA polymerases eta and kappa and Rev1 in error-prone translesion synthesis. *Biochemistry*, **53**, 5323–5331.
65. Cho,S.-H. and Guengerich,F.P. (2013) Replication past the butadiene diepoxide-derived DNA adduct S-[4-(N6-deoxyadenosinyl)-2,3-dihydroxybutyl]glutathione by DNA Polymerases. *Chem. Res. Toxicol.*, **26**, 1005–1013.
66. 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.
67. Becherel,O.J. and Fuchs,R.P.P. (2001) Mechanism of DNA polymerase II-mediated frameshift mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 8566–8571.
68. Mukherjee,P., Lahiri,I. and Pata,J.D. (2013) Human polymerase kappa uses a template-slippage deletion mechanism, but can realign the slipped strands to favour base substitution mutations over deletions. *Nucleic Acids Res.*, **41**, 5024–5035.
69. Wang,F. and Yang,W. (2009) Structural insight into translesion synthesis by DNA Pol II. *Cell*, **139**, 1279–1289.
70. Strauss,B., Rabkin,S., Sagher,D. and Moore,P. (1982) The role of DNA polymerase in base substitution mutagenesis on non-instructional templates. *Biochimie*, **64**, 829–838.
71. Strauss,B.S. (1991) The 'A rule' of mutagen specificity: a consequence of DNA polymerase bypass of non-instructional lesions? *Bioessays*, **13**, 79–84.
72. Minko,I.G., Kozekov,I.D., Kozekova,A., Harris,T.M., Rizzo,C.J. and Lloyd,R.S. (2008) Mutagenic potential of DNA–peptide crosslinks mediated by acrolein-derived DNA adducts. *Mutat. Res.*, **637**, 161–172.
73. Cooper,D.N. and Youssoufian,H. (1988) The CpG dinucleotide and human genetic disease. *Hum. Genet.*, **78**, 151–155.