5-Formylcytosine-induced DNA–peptide cross-links reduce transcription efficiency, but do not cause transcription errors in human cells

Received for publication, June 25, 2019, and in revised form, September 26, 2019. Published, Papers in Press, October 9, 2019, DOI 10.1074/jbc.RA119.009834

© Shaofei Ji, Daeyoon Park, Konstantin Kropachev, Marina Kolbanovskiy, Iwen Fu, Suse Broyle, Maram Essawy, Nicholas E. Geacintov, and Natalia Y. Tretyakova.

From the Departments of Chemistry, Biochemistry, Molecular Biology and Biophysics, and Pharmacology and the Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455 and the Departments of Chemistry and Biology, New York University, New York, New York 10003

Edited by Karin Musier-Forsyth

5-Formylcytosine (5fC) is an endogenous epigenetic DNA mark introduced via enzymatic oxidation of 5-methyl-dC in DNA. We and others recently reported that 5fC can form reversible DNA–protein conjugates with histone proteins, likely contributing to regulation of nucleosomal organization and gene expression. The protein component of DNA–protein cross-links can be proteolytically degraded, resulting in smaller DNA–peptide cross-links. Unlike full-size DNA–protein cross-links that completely block replication and transcription, DNA–peptide cross-links can be bypassed by DNA and RNA polymerases and can potentially be repaired via the nucleotide excision repair (NER) pathway. In the present work, we constructed plasmid molecules containing reductively stabilized, site-specific 5fC–polypeptide lesions and employed a quantitative MS-based assay to assess their effects on transcription in cells. Our results revealed that the presence of DNA–peptide cross-link significantly inhibits transcription in human HEK293T cells but does not induce transcription errors. Furthermore, transcription efficiency was similar in WT and NER-deficient human cell lines, suggesting that the 5fC–polypeptide lesion is a weak substrate for NER. This finding was confirmed by in vitro NER assays in cell-free extracts from human HEK293T cells, suggesting that another mechanism is required for 5fC–polypeptide lesion removal. In summary, our findings indicate that 5fC-mediated DNA–peptide cross-links dramatically reduce transcription efficiency, are poor NER substrates, and do not cause transcription errors.

In living cells, genomic DNA dynamically interacts with a range of architectural and regulatory proteins, allowing for chromatin packaging, DNA repair, gene expression, and propagation of genetic information. Exposure to free radicals, UV radiation, heavy metals, and common antitumor drugs can lead to irreversible cross-linking of proteins to DNA strands to form DNA–protein cross-links (DPCs)2 (1–3). In addition, topoisomerases, DNA polymerases, glycosylases, and meiotic recombinases can be trapped on their DNA substrates (4, 5). DPCs are among the most common DNA lesions in living cells and are known to accumulate in tissues with age, likely contributing to human disease (1).

We recently discovered that 5-formylcytosine (5fC) residues in DNA and lysine or arginine side chains of histone proteins form reversible DPCs that can be stabilized by reduction (6). 5fC is an epigenetic DNA mark generated via oxidation of 5-methylcytosine by ten eleven translocation dioxygenases (Tet) (7, 8). Li et al. (9) independently reported 5fC–histone DPC formation in chromosomal core particles. Although the resulting DNA–histone cross-links are reversible ($t_{1/2} \approx 1.8$ h), they are likely to have a large effect on chromatin structure and gene expression (6). A recent report suggests that 5fC-mediated covalent DNA–histone interactions play a role in controlling nucleosome positioning, establishing distinct regulatory regions for fine-tuning the levels of gene expression (10).

The presence of endogenous 5fC-mediated DNA–protein cross-links in living cells raises a question of how such bulky lesions are tolerated by cells and whether they influence the efficiency and the fidelity of DNA replication and transcription. We recently investigated the effects of reductively stabilized 5fC-mediated DPC on DNA synthesis and found that TLS DNA polymerase $\eta$ and $\kappa$ were completely blocked by covalent DNA–histone lesions, whereas smaller DNA–polypeptide lesions were readily bypassed (11). We have also shown that large DPCs completely inhibit T7 RNA polymerase (35).

It has been suggested that in vivo, the protein component of DPCs is proteolytically degraded by specialized metalloproteases (SPRTN/Wss1) and/or the ubiquitin/proteasome pathway to smaller polypeptide conjugates (DpCs) (12–16), which could be removed via the nucleotide excision repair (NER)

2 The abbreviations used are: DPC, DNA–protein cross-link; DpC, DNA–peptide cross-links; 5fC, 5-formylcytosine; NER, nucleotide excision repair; ESI, electrospray ionization; CTAB, competitive transcription and bypass assay; HRMS, high-resolution MS; XPA, group A xeroderma pigmentosum.

This work was supported by the NIH, National Institutes of Health Grants R01-ES-023350 (to N. Y. T.), R01-ES-025987 (to S. B.), and R01-ES-024050 (to N. E. G.). This work was also supported in part by a Wayland E. Noland graduate student fellowship (to S. J.) and a doctoral dissertation fellowship from the University of Minnesota (to S. J.). Computational resources were provided by Extreme Science and Engineering Discovery through Grant MCB0600372 (to S. B.) and by New York University IT High Performance Computer Resources and Services. The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article contains Figs. S1–S8.

1 To whom correspondence should be addressed: Masonic Cancer Center, University of Minnesota, 2231 6th St. SE, 2–147 CCRB, Minneapolis, MN 55455. Tel.: 612-626-3432; Fax: 612-624-3869; E-mail: trety001@umn.edu.

This is an open access article under the CC BY license.
Transcription past 5-formylcytosine DNA–peptide cross-links

A.

B.

C.

Figure 1. A, formation of DNA–protein and DNA–peptide cross-links via reductive amination reactions between 5-formyl-dC in DNA and the -NH$_2$ group of Lys. B, denaturing PAGE analysis of reaction mixtures after reductive amination reaction between 5'-ATGGCGGGXTAT-3', where X = 5fC and RPKQFFGLM-CONH$_2$. The 5' ends of oligonucleotides were radiolabeled with $[^32]$P]ATP and T4 PNK. C, representative MALDI-MS spectrum of synthetic DNA–polypeptide conjugate.

pathway (17, 18). In Xenopus extracts, DPC proteolysis is replication-dependent (12, 14). However, evidence is accumulating for additional mechanisms of DPC recognition and repair in human cells. For example, Chesner and Campbell (19) examined the kinetics of DPC repair in human cells and reported that hOGG1-abasic site lesions present on nonreplicating plasmids are rapidly repaired. This suggests that DPC recognition and repair in human cells can take place in the absence of DNA replication. Interestingly, DPC lesions on the transcribed strand of DNA were removed faster than the same DPC lesions on the nontranscribed strands (19), suggesting that DPC-mediated blockage of transcription complex may facilitate their repair.

In the present work, reductively stabilized, site-specific 5fC–polypeptide lesions were incorporated into the coding region of pTGFP-T7-Hha10 plasmids under the control of the cytomegalovirus promoter. The resulting constructs were transfected in WT or NER-deficient human cells, and the effects of DpC lesions on transcription were evaluated using the new competitive transcription and bypass assay developed and fully validated by You and Wang et al. (20). The potential role of NER in repair of 5fC–polypeptide lesions was further investigated using in vitro assays employing cell-free extracts from human HeLa cells.

Results

Construction and characterization of plasmid vectors containing site-specific 5fC–polypeptide and 5-formylC-Lys cross-links

A gapped plasmid strategy (21) was employed to create plasmid vectors containing site-specific 5fC–polypeptide cross-links. Non-replicative pTGFP-T7-Hha10 plasmids containing cytomegalovirus promoter were used for this purpose. First, 5fC-containing DNA 12-mers (5'-ATGGCGGGXTAT-3', where X = 5fC) were conjugated to an 11-mer peptide (RPKPQFFGLM-CONH$_2$) using reductive amination in the presence of NaCNBH$_3$, generating covalent DpCs (Fig. 1A) (6, 11). DpCs were purified by denaturing PAGE (Fig. 1B), characterized by MALDI-TOF-MS (Fig. 1C), and ligated into gapped plasmids (Fig. 2 and Fig. S1).

The corresponding cross-links to single Lys were generated analogously using l-lysine, purified by HPLC, and characterized by MS (Fig. S2) prior to incorporation into gapped plasmid. In the resulting plasmid constructs, DNA lesions were placed on the transcribed strand of DNA 56 nucleotides downstream of the transcription start site (21). In addition to control plasmid containing standard dC in place of DpC lesion, a competitor plasmid was prepared containing three additional bp near the modification site (Fig. 3B). Co-transfection with competitor
Transcription past 5-formylcytosine DNA–peptide cross-links

Human embryonic kidney (HEK293T) cells were transfected with a mixture of DpC-containing vectors and competitor plasmids. In control experiments, unmodified plasmids containing unmodified dC instead of the DpC lesion were employed. 24 h post-transfection, RNA transcripts were isolated and amplified using RT-PCR (Fig. S3). The resulting RT-PCR products were digested with SfaNI/NcoI to release short 13-mer oligonucleotides containing the region of interest from lesion-bearing or control plasmids while at the same time generating 16-mer oligonucleotides from competitor plasmids (Figs. 3 and 4A and Fig. S3). HPLC-ESI-HRMS/MS on an Orbitrap Velos mass spectrometer was employed to quantify and sequence RT-PCR products. We found that the use of high-resolution MS on an Orbitrap MS allows for accurate and sensitive detection of RT-PCR products with minimal background from the sample matrix. The use of competitor plasmid as an internal control accounts for any variations in transfection efficiency or plasmid degradation (19).

Capillary HPLC-ESI-HRMS/MS analysis of RT-PCR–amplified transcript products from in vivo transcription experiments revealed that 5fC–polypeptide conjugates significantly inhibited transcription in HEK293T cells. By normalizing the amounts of transcripts generated from lesion-bearing plasmids (13-mer oligonucleotides) to the amounts of transcription products from the competitor plasmids (16-mer oligonucleotides), we found that the relative transcription efficiency of the templates containing DNA–peptide conjugates was only ~10% as compared with unmodified dC in the control experiments (Fig. 4B). Transcription levels of probes containing 5fC–lysine conjugates were ~40%, indicating that these smaller lesions are less blocking as compared with the 11-mer peptide conjugates (Fig. 4B).

Despite their blocking effects on human RNAP II, DpC conjugates did not induce transcription errors. Only error-free products (5′-CCGAGATAGCAGGC, [M-4H]4− = 974.7), where the correct base (G) was inserted opposite the modified dC during the RNA synthesis, were observed. We tried but failed to detect any transcription products containing nucleobase substitutions, deletions, insertions, or post-lesion synthesis errors (Fig. 5A). MS/MS spectrum of the error-free product (5′-CCGAGATAGCAGGC) was consistent with the theoretical spectrum predicted from Mongo Oligo Mass Calculator (Fig. 5B). Similar results were observed for 5fC conjugated to Lys residue, indicating that the fidelity of in vivo transcription was unaffected.

Potential role of NER in 5fC DpC removal in human cells

To investigate potential role of nucleotide excision repair in the removal of DpC lesions conjugated to 5fC in DNA, in vivo transcription assays were repeated using XPA-deficient cells and XPA-complemented human fibroblast cells. NER deficiency in XPA cells and competency in complemented cells was

plasmid makes it possible to account for any differences in transfection efficiency by normalizing transcript amounts from test plasmid relative to transcripts amounts from competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-

tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three additional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
tional bp in the competitor plasmid (21). The use of DNA sequence containing three addi-
...
Transcription past 5-formylcytosine DNA–peptide cross-links

A. Parent vector

![Diagram of DNA and peptide cross-links]

B. Competitor vector

![Diagram of DNA and peptide cross-links]

Figure 3. A and B, Ncol/SfaNI-mediated restriction digestion and post-labeling assays for characterization of the plasmid vectors containing unmodified dC or DNA lesions (A) and the competitor plasmids (B). C, denaturing PAGE analysis of restriction-digested undamaged control plasmids (lane 1), the competitor plasmids (lane 2), 1:1 mixture of the control and competitor plasmids (lane 3), and plasmids containing site-specific 5fC conjugated to 11-mer peptide (lane 4).

confirmed by the strand-specific primer extension-quantitative polymerase chain reaction assay using a known NER substrate (cholesterol-dR) (19), UV sensitivity (Fig. S8), and in vitro assays with known NER substrates (22). Our CTAB experiments revealed that the relative transcription efficiencies of 5fC-conjugated 11-mer peptide cross-links in XPA-deficient cells and XPA-complemented cells 24 h post-transfection were ~12% and ~18%, respectively (Fig. 4B). This difference was not statistically significant (p > 0.1, n = 3). Similarly, the transcription efficiency of 5fC–polypeptide-containing templates was ~49 and 63% relative to control in XPA-deficient and XPA-complemented cells, respectively, and this difference was not statistically significant (Fig. 4B, p > 0.1, n = 3). Analogous results were obtained when RNA was isolated from transfected cells at a longer time point (48 h after transfection; Fig. S4). These results indicate that unlike other DpC lesions reported in the literature (17, 18), 5fC-conjugated peptide cross-links are poor substrates of nucleotide excision repair.

In vitro NER assay for DpC lesions using HeLa cell extracts

To confirm our results of the repair of 5fC conjugates in human cells, in vitro NER assays in cell-free extracts were performed (Fig. 6). Briefly, DNA–peptide and DNA–lysine cross-links were site-specifically incorporated into 135-mer DNA duplexes with an internal 32P-radiolabel near the lesion (Fig. 6A). The resulting 135-mer DNA duplexes were incubated with NER-competent nuclear extracts from human HeLa cells (23). The time-dependent appearance of characteristic 24–30-mer dual-incision products resulting from lesion-bearing templates indicates successful NER activity (Fig. 6B). In positive control experiments performed with a known NER substrate, (+)-cis-BPDE-dG, time-dependent formation of the characteristic NER products at ~26–28 nt in length was observed (lanes 10 and 11 in Fig. 6C). In the corresponding NER experiments employing 5fC–polypeptide templates, a series of weak bands ~35 nt in length were detected (lanes 6 and 7 in Fig. 6C) that were not observed in negative controls employing unmodified dC (lanes 12 and 13 in Fig. 6C). These NER products moved slower as compared with typical NER products (lane 11 in Fig. 6C), probably because of the presence of a covalently attached 11-mer peptide. When treated with proteinase K to digest the polypeptides to single amino acids, the mobility of these bands was increased, but they were still slightly shifted as compared with BPDE control because of the presence of 5fC–Lys lesions in their structure (lanes 1–5 in Fig. 6C).

The fractions of excised 24–30-nt NER products were determined from the gel autoradiographs by comparing the relative intensities of the bands corresponding to intact and incised products. Quantitative assessment of individual gel lanes revealed that 5fC conjugated to 11-mer peptide is a poor NER substrate (<5% of the (+)-cis-BPDE-dG used as a positive control; Fig. 6, D–F). Furthermore, in vitro NER assays using the 5fC conjugated to a single lysine residue did not generate any detectable NER products (Fig. 6E and Fig. S5). Overall, our in vitro NER experiment confirmed that DpCs conjugated to the C5 position of cytosine are strongly resistant to NER in human cell extracts.

Discussion

The primary goal of this study was to elucidate the effects of 5fC-mediated DNA–polypeptide lesions on transcription in human cells. Although the global levels of 5fC in living cells are relatively low (0.002–0.02% of all cytosine) (7, 8), these epigenetic marks are preferentially found in the regulatory elements of genes and are thought to influence gene expression levels (24, 25). We and others recently reported that 5fC can form revers-
**Figure 4.** A, CTAB assay to investigate the effects of DpC lesions on transcription in human cells. B, relative bypass efficiencies of unmodified dC, 5fC conjugated to Lys, and 5fC–11-mer peptide conjugates (RPKPQQFFGLM-CONH$_2$) in human embryonic kidney cells (HEK293T), human NER-deficient cells (GM04429, XPA), and XPA-complemented cells (XPA-corrected). Each value represents the average of at least three independent experiments. **ns**, not significant. ****, $p < 0.0001$.  

Transcription past 5-formylcytosine DNA–peptide cross-links
Transcription past 5-formylcytosine DNA–peptide cross-links

Figure 5. A, capillary HPLC-ESI-HRMS/MS analysis of in vivo transcription products generated from site-specific 5fC-conjugated to 11-mer peptide (RPKPOQFF-GLM-CONH₂). 13-mer error-free products (5′-CCGAAATGCCGC-3′, [M-4H]⁴⁺ = 974.665); C to T (5′-CCGAAATACCCGC-3′, [M-4H]⁴⁺ = 970.667); C to A (5′-CCGAAATATCCGC-3′, [M-4H]⁴⁺ = 968.414); C to G (5′-CCGAAATACCCGC, [M-4H]⁴⁺ = 964.664); deletion (5′-CCGAAATCCCACGC-3′, [M-4H]⁴⁺ = 892.402); and 16-mer products from competitor plasmids (5′-CACAATAGCATATCGC-3′, [M-4H]⁴⁺ = 1206.957). B, CID MS² spectrum of the 13-mer error-free products.

Transcription past 5-formylcytosine DNA–protein conjugates with histone proteins in nucleosome core particles and in human cells (6, 9). Because of their endogenous formation, it is of critical importance to establish how 5fC-mediated DPCs influence the efficiency and the fidelity of transcription in human cells. Other types of DPCs have been reported to be processed by specialized proteases, giving rise to smaller DNA–peptide cross-links, which may be better tolerated by DNA replication and transcription machineries and are ultimately repaired by NER (12, 13, 17, 18).
Our earlier in vitro study revealed that T7 RNA polymerase is blocked by 5fC-conjugated protein lesions but that smaller polypeptide lesions were readily bypassed (35). However, T7 RNA polymerase is a small, single subunit enzyme that differs greatly from RNA polymerase II employed by human cells and lacks the proofreading activity (26, 27).

In the present work, we for the first time addressed the effects of 5fC–polypeptide conjugates on transcription in human cells.
Transcription past 5-formylcytosine DNA–peptide cross-links

For this purpose, we engineered plasmid vectors containing DpC or DNA–Lys lesions 56 nucleotides downstream from transcription start site. We found that presence of reductively stabilized 5fC–polypeptide conjugates on DNA significantly inhibited transcription in human cells (Fig. 4B). Bulky peptide lesions likely represent a physical impediment to transcription during RNA elongation, resulting in the stalling of RNA polymerase II at the lesion site. 5fC conjugated to a single lysine is less blocking to RNA polymerases as compared with an 11-mer peptide (Fig. 4B). However, it is not known whether the DNA–peptide lesions can be cleaved to single amino acid lesions in cells.

In vivo transcription of DNA containing 5fC-mediated DNA–peptide/lysine cross-links predominately generated error-free transcription products (Fig. 5). This is different from our in vitro T7 RNA polymerase bypass assays, where we observed preferential incorporation of A opposite modified dC (35). This is probably due to the lack of proofreading activity in viral T7 RNAP as compared with RNAP II, because the latter can detect and remove incorporated incorrectly added nucleotides via “backtracking” (27–29). Our results differ from a report by Nakano et al. (30) showing that in vivo T7 RNAP bypass of oxamine-induced DPCs produced large amounts of substitution and deletion transcriptional mutations. Collectively, these results suggest that the cross-linking site within DNA and the identity of the RNA polymerase significantly affect transcription fidelity.

In the case of transcription coupled repair, the arrest of transcription by bulky DNA lesions is the initiation step that subsequently leads to the recruitment of the NER proteins that lead to the excision and generation of the characteristic 24–30-nt excision repair products. There are previous data showing that the NER-deficient cells are hypersensitive to DPC-inducing excision repair products. There are previous data showing that the NER-deficient cells are hypersensitive to DPC-inducing excision repair products. There are previous data showing that the NER-deficient cells are hypersensitive to DPC-inducing excision repair products. There are previous data showing that the NER-deficient cells are hypersensitive to DPC-inducing excision repair products. There are previous data showing that the NER-deficient cells are hypersensitive to DPC-inducing excision repair products. There are previous data showing that the NER-deficient cells are hypersensitive to DPC-inducing excision repair products. There are previous data showing that the NER-deficient cells are hypersensitive to DPC-inducing excision repair products.

In summary, this study for the first time investigated the effects of 5fC-mediated DNA–peptide cross-links on transcription in human cells and found that these lesions dramatically reduce transcription efficiency but do not cause transcription errors. Because reversible 5fC–polypeptide conjugates naturally form in cells (6), it is not surprising that human RNA polymerases have evolved to allow for error-free bypass of such lesions by the transcription machinery. Furthermore, because the 5fC–histone conjugates formed in cells are reversible (6), it is possible that other mechanisms exist for the removal of transiently attached proteins. However, if present in a coding region of a gene, such structures are expected to reduce the levels of gene expression.

Experimental procedures

Synthesis and characterization of 5fC-containing oligonucleotide

DNA oligodeoxynucleotides (5’-ATGGCGGCGGXTAT-3’), where X = 5fC) were synthesized using solid-phase synthesis on an ABI 394 DNA synthesizer as previously described (6).
Nucleoside phosphoramidites and other reagents for solid-phase synthesis were purchased from Glen Research (Sterling, VA). 5FC-containing synthetic oligodeoxynucleotides were cleaved off the solid support using 30% ammonium hydroxide at room temperature overnight and deprotected in 80% acetic acid at room temperature for 6 h. The resulting DNA strands were purified by semi-preparative HPLC and characterized by MS as described previously (11). Standard unmodified oligonucleotides were purchased from IDT (Coralville, IA) and purified by semi-preparative HPLC or denaturing PAGE before use.

**Synthesis, purification, and characterization of DNA conjugated to 11-mer peptides or lysine residues**

DNA–peptide cross-links were synthesized and characterized by gel electrophoresis and MS as described in earlier publications (Fig. 1A) (6, 11). Briefly, 5FC-containing DNA 12-mers (5’-ATGGCGGGXAT-3’, 300 pmol) were incubated with 20-fold molar excess of 11-mer peptides (RPKPQQFFGLM-CONH₂) in 16 μl of sodium phosphate buffer (4.5 mM, pH 7.4) for 3 h at 37°C. To stabilize DNA–peptide linkages, Schiff base conjugates were reduced by adding 4 μl of 100 mM NaCNBH₃ and incubated at 37°C overnight. The reaction mixtures were heated at 90°C for 15 min and purified by 20% denaturing PAGE containing 7 M urea (Fig. 1B). Gel-purified DNA–peptide cross-links were desalted by solid-phase extraction on Sep-Pak C18 cartridges and characterized by MALDI-TOF-MS using 3-HPA matrix as previously described (Fig. 1C) (6).

DNA strands containing 5FC–Lys cross-links were synthesized analogously, with the exception that 25 mM synthetic 1-lysine was used in the reaction. The reaction mixture was purified by HPLC on an Xbridge BEH C18 column (2.5 μm) with a gradient of 100 mM TEAA at pH 7.0 (A) and acetonitrile (B). Solvent composition was changed linearly from 6% to 15% over 40 min, increased to 20% in 10 min, further to 75% in 1 min, and held at 75% for 5 min. HPLC-purified oligodeoxynucleotides containing 5FC–Lys conjugates were characterized by HPLC-ESI−MS on a Zorbax 300SB C18 column with a gradient of 15 mM ammonium acetate (A) and acetonitrile (B) (Fig. S2A).

**Construction of plasmids containing site-specific 5FC–11-mer peptide and 5FC–Lys conjugates**

pTGFP-T7-Hha10 (control) and pTGFP-T7-Hha10comp (competitor) plasmids were a kind gift from Prof. Yinheng Wang at the University of California–Riverside (20). As compared with control plasmid, competitor plasmid contains three additional bp near the modification site. Plasmids were amplified in TOP10F’ chemically competent Escherichia coli, characterized by gel electrophoresis, and sequenced prior to use. Plasmid vectors containing site-specific DpC lesions were prepared using the gapped vector strategy described elsewhere (20, 21, 34) as shown in Fig. 2. Briefly, pTGFP-T7-Hha10 plasmids were treated with Nt.BstNBI to introduce two nicks 25 nucleotides apart, and the resulting 25-mer oligonucleotides (5’-pATGGCGGGGCTA-TTCGGGAGTCGATG-3’) were removed from the plasmid by incubating with ∼70 μl molar excess of the complementary strand (5’-CATCGACTCCCGAATAGCCCAGCAT-3’). The resulting gapped plasmids were purified via agarose gel and ligated with the 13-mer oligodeoxynucleotide, 5’-pTCGGGATCGATG-3’, and the 12-mer oligodeoxynucleotide containing site-specific DpC lesions (5’-ATGGCGGGXAT-3’, where X = 5FC conjugated to 11-mer peptide RPKPQQFFGLM-CONH₂ or single Lys) (Fig. 2). Lesion-bearing supercoiled plasmids were purified by agarose gel prior to use as shown in Figs. S1 and S6.

The presence of site-specific DpC lesions on engineered plasmid vectors was confirmed via NcoI/SfNI-mediated restriction digestion and post-labeling assays (Fig. 3). In brief, lesion-bearing plasmids (40 ng) were treated with NcoI (5 units) and shrimp alkaline phosphatase (0.5 unit). The 5’ ends of dephosphorylated fragments were radiolabeled with [γ-³²P]ATP and T4 PNK (5 units). Treatment with SfNI (2 units) released radiolabeled DpC-containing 13-mers. The reaction mixtures were analyzed by 20% denaturing PAGE. The presence of a DpC lesion was detected via a characteristic gel shift (Fig. 3C and Fig. S2B). The competitor vector was characterized similarly and produced DNA 16-mers because of the presence of three additional bases in its sequence (Fig. 3).

**CTAB assay**

In vivo CTAB assays were performed as described elsewhere (20). Briefly, human embryonic kidney cells (HEK293T), human XPA-deficient fibroblast cells (GM04429), and XPA-complemented fibroblasts (1.25 × 10⁶) obtained from the NIGMS Human Genetic Cell Repository (Camden, NJ) were seeded into 24-well plates in Dulbecco’s modified Eagle’s medium supplemented with 9% fetal bovine serum and cultured in a humidified atmosphere of 5% carbon dioxide and 95% air at 37°C overnight. NER deficiency in XPA cells was confirmed by the strand-specific primer extension-quantitative polymerase chain reaction assay using a known NER substrate (cholesterol-dr) (19), UV sensitivity (Fig. S8), and an in vitro NER assay (22). Lesion-bearing or unmodified control plasmids were mixed with the competitor plasmids at a 3:1 molar ratio, and 50 ng of the plasmid mixture was transfected into cells using Lipofectamine 2000 (Thermo Fisher Scientific). The cells are harvested after 24 h, and RNA was extracted with total RNA extraction kit (Omega BioTek, Norcross, GA). The purified RNA was treated twice with Ambion DNA-free kit (Thermo Fisher Scientific). RT-PCR was used to confirm that there was no residual DNA contamination (Fig. S7). cDNA was synthesized using a specific primer (5’-TCGGTGTTGCTGTGAT-3’) and Moloney murine leukemia virus reverse transcriptase. The resulting cDNA was amplified via PCR with Phusion DNA polymerases (New England Biolabs, Ipswich, MA) using primers 5’-CTAGCGGATCGACGTAC-3’ and 5’-TGCTCGCGA TGATCTTGCG-3’. PCR amplification started with incubation at 98°C for 3 min, followed by 37 cycles at 98°C for 20 s, 58°C for 30 s, 72°C for 20 s, and a final extension at 72°C for 5 min. PCR products were purified by E.Z.N.A. Cycle Pure kit (Omega BioTek, Norcross, GA), followed by treatment with SfNI (30 units), shrimp alkaline phosphatase (15 units), and NcoI (50 units; New England Biolabs) (Fig. S3). Proteins were extracted with an equal volume of phenol:chloroform:isoamyl alcohol solution (25:24:1), and the DNA was precipitated with 2.5 volumes of ethanol and 0.1 volumes of 3 M sodium acetate at
Transcription past 5-formylcytosine DNA–peptide cross-links

−80 °C overnight. DNA was reconstituted in LC-MS grade water for the HPLC-ESI-MS/MS analysis.

UV sensitivity of XPA and XPA corrected cells

Human XPA-deficient fibroblasts (GM04429) and the corresponding XPA-complemented fibroblasts obtained from NIGMS Human Genetic Cell Repository (Camden, NJ) were grown in Dulbecco’s modified Eagle’s medium containing 15% FBS and 1× antibiotic. The cells were trypsinized and counted. 100,000–200,000 cells were plated in 35-mm dishes. The cells were incubated at 37 °C overnight or until they reached 60% confluence. To perform UV dosing experiments, the dishes were removed from the incubator, the medium was removed, and the dishes were immediately placed in a UVP HybriLinker HL-2000 apparatus. Samples (in triplicate) were irradiated with specified UV dosage (0–8 J/cm²). Immediately after irradiation, fresh medium was added to the dishes, and they were returned to the 37 °C incubator. 20 h after UV dosing, the cells were trypsinized and counted.

HPLC-ESI-MS/MS analysis of RT-PCR products

RT-PCR products generated from in vivo transcription were analyzed and quantified using a Dionex Ultimate 3000 HPLC coupled to a LTQ Orbitrap Velos mass spectrometer. A Zorbax 300SB-C18 column (150×0.5 mm, 5 μm) 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 changed 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. The mass spectrometer was set to monitor 13-mer oligonucleotides corresponding to transcription products (5′-CCGAAATAXCCCGC-3′, where X = A, T, C, or G) and 16-mer oligonucleotides originating from competitor plasmids (5′-CACAATAGCATATCCG-3′) as an internal standard for quantification. Relative quantification was performed by comparing HPLC-ESI-MS peak areas in extracted ion chromatograms corresponding to each RT-PCR product relative to the internal standard. MS/MS spectra were used to confirm the sequence of each oligodeoxynucleotide product.

In vitro NER experiments in HeLa cell extracts

DNA 135-mer duplexes containing site-specific DpC lesions were constructed as described in detail by Kropachev et al. (23) Briefly, gel-purified 12-mer oligonucleotides containing reductively stabilized 5fC−polypeptide or 5fC−Lys conjugates were radiolabeled with [γ-32P]ATP and T4 PNK and ligated to produce 135-mer oligonucleotide duplexes (see Fig. 6A for DNA sequence). Nuclear extracts were prepared using human HeLa cells and utilized for in vitro NER assays as described earlier (23). Internally 32P-labeled 135-mer oligonucleotide duplex containing site-specific DNA lesion at the 70th nucleotide from the 5′ end (1 pmol; Fig. 6A) was mixed with 17.5 μl of 1 M KCl, 20 μl of Tris-ATP (10 mM at pH 7.9), 10 μl of freshly prepared HeLa cell extracts, and sufficient dialysis buffer containing 12 mM MgCl₂, 25 mM HEPES-KOH, pH 7.9, 2.5 mM DTT, 1 mM EDTA, and 10% glycerol in a total volume of 50 μl. After pre-selected incubation times and appropriate preparation, aliquots of the reaction mixture were quenched with gel loading buffer and loaded onto 12% denaturing PAGE for analysis.

Model of the DpC-containing B-DNA duplex

The structure of the 11-mer peptide (RPKPQQFFGLM-CONH₂) cross-linked to the C5-dC in B-DNA (5′-CAT-GACGGCT-3′) was extracted from the most representative structure of our previous MD simulation for this DpC bound to human DNA polymerase η (11). Similarly, the unmodified B-DNA structure was extracted from our previous simulation of DNA polymerase η without DpC (11).

Author contributions—S. J. formal analysis; S. J., D. P., K. K., M. K., and I. F. investigation; S. J. methodology; S. J. writing-original draft; I. F. software; S. B. data curation; S. B., N. E. G., and N. Y. T. supervision; S. B., N. E. G., and N. Y. T. writing-review and editing; M.E. validation; N. Y. T. conceptualization; N. Y. T. funding acquisition; N. Y. T. project administration.

Acknowledgments—We gratefully acknowledge Prof. Yinsheng Wang at the University of California at Riverside for providing the pTGFP-T7-Hha10 and pTGFP-T7-Hha10comp plasmids as kind gifts. We thank Xun Ming and Dr. Peter W. Villalta (University of Minnesota) for help with MS analyses, Dr. Colin Campbell for advice with viability essays, and Robert Carlson (University of Minnesota) for help with figure preparation. We thank Dr. Lihua Wang for dedicated support of computational infrastructure in the Broyde laboratory.

References

1. Barker, S., Weinfield, M., and Murray, D. (2005) DNA–protein crosslinks: their induction, repair, and biological consequences. Mutat. Res. 589, 111–135 CrossRef Medline
2. Tret'yakova, N. Y., Groehler, A., 4th, and Ji, S. (2015) DNA–protein cross-links: formation, structural identities, and biological outcomes. Acc. Chem. Res. 48, 1631–1644 CrossRef Medline
3. Wickramaratne, S., Ji, S., Mukherjee, S., Su, Y., Penc, M. G., Lior-Hoffmann, L., Fu, I., Broyde, S., Gaugenerich, F. P., Distefero, M., Schärer, O. D., Sham, Y. Y., and Tret’yakova, N. (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 CrossRef Medline
4. Neale, M. J., Pan, J., and Keeney, S. (2005) Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. Nature 436, 1053–1057 CrossRef Medline
5. Yang, S. W., Burgin, A. B., Jr., Huizenga, B. N., Robertson, C. A., Yao, K. C., and Nash, H. A. (1996) A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. Proc. Natl. Acad. Sci. U.S.A. 93, 11534–11539 CrossRef Medline
6. Li, S., Shao, H., Han, Q., Seiler, C. L., and Tret’yakova, N. Y. (2017) Reversible DNA–protein cross-linking at epigenetic DNA marks. Angew. Chem. Int. Ed. Engl. 56, 14310–14314 CrossRef Medline
7. Aso, T., Shiraishi, I., Adachi, M., Sakugawa, K., Tsuchida, Y., Inoue, T., and Yang, S. W. (2010) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333, 1300–1303 CrossRef Medline
8. Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C. A., Hagemeyer, C., and Carell, T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew. Chem. Int. Ed. Engl. 50, 7008–7012 CrossRef Medline
9. 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 CrossRef Medline
10. Raiber, E. A., Portella, G., Martínez Cuesta, S., Hardisty, R., Murat, P., Li, Z., Iurlaro, M., Dean, W., Spindel, J., Beraldi, D., Liu, Z., Dawson, M. A., et al.
Transcription past 5-formylcytosine DNA–peptide cross-links

Reik, W., and Balasubramanian, S. (2018) 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. Nat. Chem. 10, 1258–1266 CrossRef Medline

10. You, C., Wang, P., Dai, X., and Wang, Y. (2014) Transcriptional bypass of DNA–protein cross-links by human excision nuclease. Proc. Natl. Acad. Sci. U.S.A. 103, 4056–4061 CrossRef Medline

9. You, C., Dai, X., Yuan, B., Wang, J., Wang, J., Brooks, P. J., Niedernhofer, L. J., and Wang, Y. (2012) A quantitative assay for assessing the effects of DNA lesions on transcription. Nat. Chem. Biol. 8, 817–822 CrossRef Medline

8. Shafirovich, V., Kropachev, K., Kolbanovskii, M., and Geacintov, N. E. (2019) Excision of oxidatively generated guanine lesions by competing base and nucleotide excision repair mechanisms in human cells. Chem. Res. Toxicol. 32, 753–761 CrossRef Medline

7. Cramer, P. (2002) Multisubunit RNA polymerases. Curr. Opin. Struct. Biol. 12, 89–97 CrossRef Medline

6. Nakano, T., Ouchi, R., Kawazoe, J., Pack, S. P., Makino, K., and Ide, H. (2012) T7 RNA polymerases backed up by covalently trapped proteins catalyze highly error prone transcription. J. Biol. Chem. 287, 6562–6572 CrossRef Medline

5. de Graaf, B., Clore, A., and McCullough, A. K. (2009) Cellular pathways for DNA repair and damage tolerance of formaldehyde-induced DNA–protein cross-links. DNA Repair (Anst.) 8, 1207–1214 CrossRef Medline

4. You, C., Wang, P., Dai, X., and Wang, Y. (2014) Transcriptional bypass of regioisomeric ethylated thymidine lesions by T7 RNA polymerase and human RNA polymerase II. Nucleic Acids Res. 42, 13706–13713 CrossRef Medline

3. Kropachev, K., Kolbanovskii, M., Cai, Y., Rodríguez, F., Kolbanovskii, A., Liu, Y., Zhang, L., Amin, S., Patel, D., Brody, S., and Geacintov, N. E. (2009) The sequence dependence of human nucleotide excision repair efficiencies of benzo[a]pyrene-derived DNA lesions: insights into the structural factors that favor dual incisions. J. Mol. Biol. 386, 1193–1203 CrossRef Medline

2. Song, C. X., Szulwach, K. E., Dai, Q., Fu, Y., Mao, S. Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., Gao, J., Liu, P., Li, L., Xu, G. L., Jin, P., and He, C. (2013) Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell 153, 678–691 CrossRef Medline

1. Spruijt, C. G., Gnerlich, F., Smits, C. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikov, K., Müller, U., et al. (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell 152, 1146–1159 CrossRef Medline