Replication studies of carboxymethylated DNA lesions in human cells

Jun Wu1, Pengcheng Wang2, Lin Li1, Nicole L. Williams2, Debin Ji1, Walter J. Zahrancik3, Changjun You1, Jianshuang Wang1, Zucai Suo3 and Yinsheng Wang1,2,*

1Department of Chemistry, University of California, Riverside, CA 92521, USA, 2Environmental Toxicology Graduate Program, University of California, Riverside, CA 92521, USA and 3Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210, USA

Received December 13, 2016; Revised May 01, 2017; Editorial Decision May 03, 2017; Accepted May 04, 2017

ABSTRACT

Metabolic activation of some N-nitroso compounds (NOCs), an important class of DNA damaging agents, can induce the carboxymethylation of nucleobases in DNA. Very little was previously known about how the carboxymethylated DNA lesions perturb DNA replication in human cells. Here, we investigated the effects of five carboxymethylated DNA lesions, i.e. O6-CMdG, N6-CMdA, N3-CMdT, N3-CMdT and O4-CMdT on the efficiency and fidelity of DNA replication in HEK293T human embryonic kidney cells. We found that, while neither N6-CMdA nor N4-CMdC blocked DNA replication or induced mutations, N3-CMdT, O4-CMdT and O6-CMdG moderately blocked DNA replication and induced substantial frequencies of T→A (81%), T→C (68%) and G→A (6.4%) mutations, respectively. In addition, our results revealed that CRISPR-Cas9-mediated depletion of Pol η resulted in significant drops in bypass efficiencies of N4-CMdC and N3-CMdT. Diminution in bypass efficiencies was also observed for N6-CMdA and O6-CMdG upon depletion of Pol κ, and for O6-CMdG upon removal of Pol ζ. Together, our study provided molecular-level insights into the impacts of the carboxymethylated DNA lesions on DNA replication in human cells, revealed the roles of individual translesion synthesis DNA polymerases in bypassing these lesions, and suggested the contributions of O6-CMdG, N3-CMdT and O4-CMdT to the mutations found in p53 gene of human gastrointestinal cancers.

INTRODUCTION

The integrity of the human genome is constantly challenged by endogenous and exogenous chemical agents including N-nitroso compounds (NOCs), which form from dietary consumption and are present in tobacco smoke as well as other environmental sources (1–5). Exposure to NOCs is known to be associated with elevated risks of gastrointestinal cancers and other human diseases (1,3–7). In this vein, metabolic activation of many NOCs result in the generation of a common reactive intermediate, i.e. diazoacetate, which can induce carboxymethylation and, to a much lower degree, methylation of nucleobases in DNA (Figure 1) (8–12). In addition, it was shown that the types and frequencies of mutations found at non-CpG sites in p53 gene in human stomach and colorectal cancers exhibit close similarity to those observed in the p53 gene induced by diazoacetate, but not by N-methyl-N-nitrosourea (MNU), a DNA methylating agent (13). Thus, it was suggested that DNA carboxymethylation, but not methylation, may account for the mutations found in p53 gene of human gastrointestinal tumors (13,14).

A better understanding of the implications of carboxymethylated DNA lesions in the etiology of human gastrointestinal cancers requires detailed investigations about their occurrence, repair and biological endpoints. Along this line, treatment of isolated DNA with diazoacetate or other carboxymethylating agents was found to induce the carboxymethylation of O6 of guanine, N6 of adenine, N4 of cytosine, as well as N3 and O6 of thymine in DNA to give O6-CMdG, N6-CMdA, N4-CMdC, N3-CMdT and O4-CMdT, respectively (Figure 1) (8–12). In addition, O6-CMdG and N6-CMdA could be detected in cultured human cells, and treatment with azaserine, a pancreatic carcinogen that can be converted to diazoacetate by cellular esterases (Figure 1), could result in a dose-dependent induction of O6-CMdG, but not N6-CMdA (15). Furthermore, it was observed that humans consuming red meat have a higher level of nitrosated compounds in feces (16) and an elevated level of O6-CMdG in colonic exfoliated cells than the control group on vegetarian diet (17). The occurrence of N4-CMdC, N3-CMdT and O4-CMdT in cellular and tissue DNA, however, has yet been examined. Some repair studies have also been conducted for O6-CMdG, where the lesion was initially thought not to be a substrate for O6-
methylguanine DNA methyltransferase (MGMT) based on the observation that the azaserine-mediated cell killing cannot be rescued by the overexpression of MGMT (18). However, a recent study showed that $O^6$-CMdG in a double-stranded synthetic oligodeoxyribonucleotide (ODN) could be recognized by MGMT in vitro, where the carboxymethyl group could be transferred to the active site cysteine residue of the repair protein (19).

When located on a single-stranded M13 plasmid, neither $N^6$-CMdA nor $N^4$-CMdC blocks DNA replication or induces mutations in *Escherichia coli* cells, whereas $N^3$-CMdT and $O^4$-CMdT are both blocking and miscoding during DNA replication in *E. coli* cells (20). Among the three SOS-induced DNA polymerases, Pol V, but not Pol II or Pol IV, promotes the replicative bypass of $N^3$-CMdT and $O^4$-CMdT in *E. coli* cells (20). In addition, $N^3$-CMdT and $O^4$-CMdT could reduce the efficiency and fidelity of DNA transcription mediated by T7 RNA polymerase or human RNA polymerase II in vitro and in cells (21). So far, it remains unknown how the carboxymethylated DNA lesions compromise the efficiency and accuracy of DNA replication in human cells.

Herein, we prepared double-stranded plasmids containing a site-specifically inserted $O^6$-CMdG, $N^3$-CMdT, $O^4$-CMdT, $N^4$-CMdC or $N^6$-CMdA (Figure 1) and examined how these lesions impede DNA replication and induce mutations in cultured human cells that are proficient in translesion synthesis (TLS) or deficient in one of major TLS DNA polymerases, i.e. Pols η, τ, κ and ζ.

**MATERIALS AND METHODS**

**Materials**

All chemicals, if not specifically mentioned, were from Sigma-Aldrich (St. Louis, MO, USA) or EMD Millipore, and all enzymes, unless otherwise noted, were obtained from New England Biolabs (Ipswich, MA, USA), 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) was obtained from Oakwood Products Inc. (West Columbia, SC, USA) and [$\gamma^{32}$P]ATP was purchased from Perkin Elmer (Piscataway, NJ, USA). All unmodified ODNs were from

**Figure 1.** Formation of diazoacetate and its induction of carboxymethylated DNA lesions.
Integrated DNA Technologies (Coralville, IA, USA). The 12-mer ODNs harboring a site-specifically incorporated carboxymethylated lesion were synthesized using conventional phosphoramidite chemistry, as described previously (8,9,11). The identities and purities of all the lesion-harboring ODNs were confirmed by liquid chromatography–mass spectrometry (LC–MS) and tandem MS (MS/MS) analyses prior to their insertion into double-stranded plasmids. Exonuclease-deficient human Pol ε (with amino acid residues 1–1189), which harbors substitutions in three amino acids (D275A, E277A, and E277E) and is capable of being replicated in SV40 large T- and pX-encoding plasmid and for assessing the qualities of the successfully prepared dT- and dCmT-bearing plasmids. The amounts of the lesion-containing double-stranded vectors were independently normalized against that of the lesion-free competitor vector following published procedures (23,25).

Primer extension assay

The 20-mer lesion-containing ODNs were generated by ligating the 12-mer lesion-containing ODN to an 8-mer ODN (GATCCCTAG) in the presence of a 27-mer scaffold (dGTAGCTAGGATCATCGACTCAATTACAGT), where ‘X’ designates an unmodified dN, as previously described (24). Primer extension assays were performed under standing-start conditions, where the primer stops right before the site of the lesion or its corresponding unmodified nucleoside. The primer–template complex, which contained the 20-mer lesion-containing or control ODN and a 13-mer primer (at a final concentration of 10 nM each, Supplementary Figure S1), was incubated in a reaction buffer at 37°C for 30 min with various concentrations of human Pol ε and all four dNTPs (250 μM each). The reaction buffer contained 50 mM Tris–HCl (pH 7.4), 8 mM MgCl₂, 1 mM DTT and 10% glycerol. An equal volume of formamide gel-loading buffer [80% formamide, 10 mM EDTA (pH 8.0), 1 mg/ml xylene cyanol and 1 mg/ml bromophenol blue] was added to terminate the reaction. The reaction mixtures were subsequently resolved on a 20% (19:1) denaturing polyacrylamide gel and the gel band intensities analyzed using a Typhoon 9410 Variable Mode Imager (Amersham Biosciences Co.).

Construction of lesion-containing and lesion-free plasmids

The lesion-containing as well as the lesion-free control and competitor genomes were prepared following the previously published procedures (25). First, a parent vector was constructed by modifying the sequence of the original pTGFPHha10 plasmid, which contains an SV40 replication origin and is capable of being replicated in SV40 large T antigen-transformed mammalian cells (25). The parental vector was subsequently digested with Nt.BstNBI to generate a 25-mer complementary ODN through annealing with a 25-mer single-stranded ODN through annealing with a 25-mer complementary ODN in large excess. The gapped plasmid was then isolated from the mixture by using 100 kDa cutoff ultracentrifugal filter units (Millipore).

To ensure complete removal of the 25-mer restriction fragment, the steps of annealing with its complementary strand and centrifugation were repeated. The gapped vector was filled with a 5'-phosphorylated 13-mer lesion-free ODN (5'-AATTGAGTCGATG-3') and a 5'-phosphorylated 12-mer lesion-carrying ODN (5'-ATGGCGXCTAT-3') (X = N₆-CmdA, N₄-CmdC, N₀-CmdG, N₃-CmdT, O₄-CmdT) or the corresponding lesion-free ODN by using T4 DNA ligation and ATP in the ligation buffer (Figure 2A). The ligation mixture was separated by using agarose gel electrophoresis in the presence of ethidium bromide to purify the successfully ligated supercoiled plasmid, and Supplementary Figure S2 displays the image of a representative agarose gel for monitoring the construction of O₄-CmdT-containing plasmid and for assessing the quality of the successfully prepared dT- and dCmT-bearing plasmids. The amounts of the constructed lesion-containing double-stranded vectors were independently normalized against that of the lesion-free competitor vector following published procedures (23,25).

Cellular DNA replication and plasmid isolation

The lesion-bearing and the corresponding non-lesion control plasmids were mixed individually with the competitor genome at molar ratios of 3:1 and 1:1, respectively. A relatively high lesion/competitor genome ratio (i.e. 3:1) was chosen so as to obtain a relatively large amount of replication products (and thus PCR products) for the lesion-containing plasmids, thereby facilitating the identification and quantification of mutagenic products arising from replication past the lesion sites. The HEK293T cells were cultured at 37°C in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal bovine serum (Invitrogen, Carlsbad, CA, USA) and 100 U/ml penicillin in an incubator with 5% CO₂. The HEK293T cells and the CRISPR/Cas9 genome-engineered cells (1 × 10⁵) were seeded in a 24-well plate and cultured for 24 h before they were transfected with 300 ng of the above-described control/competitor and lesion/competitor genome mixtures by using Lipofectamine 2000 following the manufacturer’s recommended procedures. The cells were harvested at 24 h following the transfection, and the progeny genomes were isolated using Qiagen Spin kit (Qiagen, Valencia, CA, USA) (25). The residual unreplicated plasmids were further digested by restriction endonuclease DpnI, followed by removal of the resulting linear DNA with exonuclease III digestion, as described elsewhere (26). In this vein, the parent plasmid carried 25 DpnI recognition sites and cleavage at any of these sites would lead to degradation of the entire plasmid by exonuclease III and prohibit the subsequent PCR amplification of the parent vector.

PCR and polyacrylamide gel electrophoresis (PAGE) analyses

The progeny genomes emanating from cellular replication were amplified by PCR with the use of GoTaq Hot Start DNA polymerase (Promega, Madison, WI, USA). The two primers were 5’-GCTAGCGGATGATCGACTCAATTACAG-3’ and
Figure 2. Schematic diagrams showing the procedures for the preparation of the lesion-bearing plasmid (A) and the SSPCR-CRAB assay (B). ‘X’ indicates the carboxymethylated lesions. The C/C mismatch site is underlined. ‘P1’ represents one of primers for PCR and contains a G as the terminal 3′-nucleotide corresponding to the C/C mismatch site of the lesion-bearing genome. It also contains a C/A mismatch three bases away from its 3′-end for improving PCR specificity. ‘M’ designates the nucleobase incorporated at the lesion site after replication, and ‘N’ represents the paired nucleobase of ‘M’ in the complementary strand.

5′-GCTGATTATGATCTAGTTGCGGCCGC-3′, and the PCR amplifications started at 95°C for 2 min, followed by 35 cycles at 95°C for 30 s, 64°C for 30 s, and 72°C for 1 min, and a final 5-min extension at 72°C. The PCR products were purified using Cycle Pure Kit (Omega, Norcross, GA, USA) and stored at −20°C until use. For PAGE analysis, a portion of the PCR products was treated with 5 U NcoI and 1 U shrimp alkaline phosphatase (SAP) at 37°C in 10 μl of NEB buffer 3 for 1 h, followed by heating at 80°C for 20 min to deactivate the SAP (Figure 3A). The above mixture was then treated with 5 U of T4 polynucleotide kinase (T4 PNK) in 15 μl of NEB buffer 3 containing 5 mM dithiothreitol and ATP (10 pmol cold and 1.66 pmol [γ-32P]ATP). The reaction was continued at 37°C for 30 min, followed by heating at 65°C for 20 min to deactivate the polynucleotide kinase. To the reaction mixture was subsequently added 2 U of SfaNI in 5 μl NEB buffer 3 (Figure 3A), and the solution was incubated at 37°C for 1.5 h, followed by quenching with 20 μl of formamide gel-loading buffer containing xylene cyanol FF and bromophenol blue dyes. The mixture was loaded onto 30% native polyacrylamide gel (acrylamide/bis-acrylamide = 19:1), and the gel band intensities were quantified using phosphorimager analysis (Figure 3B and Supplementary Figure S3). The effects of DNA lesions on replication efficiency and fidelity are represented by the relative bypass efficiency (RBE) and mutation frequency, respectively. The RBE values were calculated using the following formula, %RBE = (lesion signal/competitor signal)/(non-lesion control signal/competitor signal) × 100%. The mutation frequency was calculated from the percentage of mutagenic product among the sum of all products arising from replication of the lesion-containing genome.

Identification and quantification of mutagenic products by LC–MS/MS

The PCR products were digested with 30 U SfaNI restriction endonuclease and 15 U SAP in 150 μl NEB buffer 3 at 37°C for 2 h, followed by deactivation of the phosphatase at 80°C for 20 min. To the mixture was added 50 U NcoI restriction endonuclease in 50 μl NEB buffer 3 and the solution was incubated at 37°C for another 2 h. The resulting solution was extracted with phenol/chloroform/isoamyl alcohol (25:24:1, v/v/v). To the aqueous layer were subsequently added 2.5 volumes of 100% ethanol and 0.1 volume of 3.0 M sodium acetate, and the solution was incubated at −20°C overnight to precipitate the DNA.
pellet was then dissolved in doubly distilled water for LC–MS/MS analysis (Figure 4 and Supplementary Figure S4). An Agilent 1200 capillary HPLC system (Agilent Technologies, Santa Clara, CA, USA) and an LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) were used for all the LC–MS and MS/MS experiments. An Agilent Zorbax SB-C18 column (0.5 x 250 mm, 5 µm in particle size) was employed, and the gradient for LC-MS/MS analysis was 5 min of 5–20% methanol followed by 50 min of 20–45% methanol in 400 mM HFIP (pH was adjusted to 7.0 with the addition of triethylamine). The temperature for the ion-transport tube was maintained at 300°C. The mass spectrometer was set up for monitoring the fragmentations of the [M−3H]− ions of the 13-mer ODNs, d(AATTACAGCACGC) and d(AATTACAGCGCGC), which are the complementary strand sequences for the unmutated product and that with a T→C mutation at the initial damage site, respectively. To correct for the differences in ionization and fragmentation efficiencies of the two ODNs, we constructed a calibration curve by analyzing mixtures of the two synthetic ODNs at different molar ratios under the same conditions as those used for analyzing the restriction digestion mixture of replication products. The calibration curve was established by plotting the peak area ratios vs. the molar ratios for the two ODNs (Supplementary Figure S5).

RESULTS

We previously examined the influence of four carboxymethylated DNA lesions (i.e. N6-CmdA, N6-CmdC, O6-CmdT and N3-CmdT) situated on a single-stranded M13 plasmid on the efficiency and fidelity of DNA replication in E. coli cells (20). In this study, we incorporated these four carboxymethylated DNA lesions along with O6-CmdG into double-stranded plasmids and asked how these lesions impede DNA replication and induce mutations in HEK293T cells. We also investigated how the replicative bypass of these lesions is modulated by TLS DNA polymerases by conducting the replication experiments in the isogenic cells where the major TLS polymerases were individually depleted by the CRISPR/Cas9 genome editing method (23). We employed a previously established strand-specific PCR-based competitive replication and adduct bypass (SSPCR-CRAB) assay for the cellular replication experiments (23,25). To this end, we inserted site-specifically a single carboxymethylated lesion into a double-stranded shut-
In this respect, one of the primers (P1) contains a G as the terminal 3'-nucleotide corresponding to the C/C mismatch locus (Figure 2B), which is used to selectively amplify the progeny genomes arising from the replication of the bottom, lesion-containing strand. Furthermore, a C/A mismatch was strategically placed in the P1 primer three nucleotides from its 3' end to improve the specificity of strand-specific PCR, as previously described (27). The resulting PCR products were digested with NcoI and SfaNI (Figure 2B), and the restriction fragments were subsequently analyzed using native PAGE and LC–MS/MS (Figures 3-4 and Supplementary Figures S3 and S4). The quantification data from these analyses were then employed for calculating the bypass efficiencies and mutation frequencies, as described in Materials and Methods.

It turned out that neither N6-CMdA nor N4-CMdC substantially blocked DNA replication in HEK293T cells, with the bypass efficiencies being ~66% and 81%, respectively. O6-CMdT, N3-CMdT and O6-CMdG, on the other hand, moderately blocked DNA replication in HEK293T cells, with the bypass efficiencies being ~39%, 40% and 38% respectively. Replication experiments conducted in HEK293T cells with Pol η, Pol ι, Pol κ and Pol ζ being individually depleted with the CRISPR-Cas9 method revealed the roles of these polymerases in bypassing the carboxymethylated DNA lesions. In particular, we observed significant decreases in bypass efficiencies for N4-CMdC and N3-CMdT in cells depleted of Pol η, for N6-CMdA and O6-CMdG in cells lacking Pol κ, and for O6-CMdG in cells deficient in Pol ζ (Figure 5A), supporting the roles of these polymerases in bypassing the relevant carboxymethylated DNA lesions. It is worth noting that, despite being statistically significant, the extent of decreases in bypass efficiencies were modest upon depletion of these TLS polymerases. The lack of pronounced effects on the bypass efficiencies upon the depletion of these TLS polymerases suggest that the carboxymethylated DNA lesions might be bypassed by replicative DNA polymerases. To test this, we conducted an in-vitro replication study with the use of exonuclease-deficient human Pol ε. Our results from primer extension assay showed that, as expected, human Pol ε could bypass readily the lesion-free 20-mer template ODNs. In addition, the polymerase could also bypass 20-mer template ODNs housing a site-specifically inserted N6-CmdA, N4-CmdC, O6-CmdD, N3-CmdT and O6-CmdT and produce full-length replication products (Supplementary Figure S1). In line with the more substantial blockage effects of O6-CmdG, N3-CmdT and O6-CmdT on cellular DNA replication, these lesions constitute stronger impediments toward human Pol ε-mediated replication in vitro than N6-CmdA and N6-CmdC (Supplementary Figure S1).

The results from PAGE and LC–MS/MS analyses of restriction fragments of PCR products from progeny genomes also allowed us to assess the mutation frequencies of the carboxymethylated DNA lesions. Our results showed that neither N6-CmdD nor N6-CmdA was mutagenic in HEK293T cells that are TLS-proficient or deficient in any of the four TLS polymerases (Figure 5B and Supplementary Figure S4). However, both O6-CmdT and N3-CmdT were highly mutagenic in HEK293T cells, with T→C transition and T→A transversion occurring at frequencies of 72% and
that the types and relative frequencies of mutations observed at non-CpG sites in p53 gene in human gastrointestinal cancer and those induced from the replication of diazoacetate-treated human p53 gene-containing plasmid in yeast cells are strikingly similar (13). Thus, it was advocated that N-nitroso compounds may constitute important etiological agents for the development of gastrointestinal cancers (13). Diazoacetate is known to induce the carboxymethylation and, to a much lesser degree, methylation of nucleobases in DNA (8–12). Thus, understanding the implications of carboxymethylated DNA lesions in the etiology of human cancer development requires a rigorous assessment about their occurrence, repair and biological endpoints. In the present study, we examined the biological consequences of the carboxymethylated DNA lesions by assessing how these lesions inhibit DNA replication and induce mutations in human cells, and by investigating how the replicative bypass of these lesions is modulated by TLS DNA polymerases.

Our results showed that N^6-CMdA and N^4-CMdC did not perturb the efficiency or fidelity of DNA replication in HEK293T cells (Figure 5). In addition, replication past these two lesions are also highly accurate in HEK293T cells depleted of any of the four TLS polymerases. These results are in agreement with the observation that neither lesion strongly impedes human Pol ε-mediated primer extension in-vitro (Supplementary Figure S1) and the previous findings made from replication studies of these two lesions in E. coli cells (20). The lack of perturbation of these two lesions on the fidelity of DNA replication perhaps can be attributed to the absence of alterations in base-pairing properties of the nucleobases arising from either carboxymethylation event (Figure 6). Nevertheless, the bypass of N^6-CMdA and N^4-CMdC requires Pol κ and Pol η, respectively, which differs from the previous study showing no difference in bypass efficiencies for these lesions upon depletion of Pol IV and Pol V (orthologues of Pol κ and Pol η, respectively) in E. coli (20). In addition, an in-vitro study showed that yeast pol η was inefficient in extending the primer across N^4-CMdC in the template strand, and the polymerase preferentially inserted the incorrect dCMP opposite the lesion (24). These differences could be due to the variations in TLS polymerases in different organisms, and/or the fact that the replication with a single polymerase in-vitro does not fully recapitulate that with the entire replication machinery in cells (30,31).

This is the first study about the effect of O^6-CMdG on DNA replication in human cells. Our result revealed that the lesion moderately blocks DNA replication (with a 38% bypass efficiency) and induces exclusively G→A transition at a frequency of 6.4% in HEK293T cells (Figure 5). In addition, significant decreases in bypass efficiency of the lesion were observed when Pol κ or Pol ζ was genetically depleted, and a significant reduction in G→A mutation was only found upon depletion of Pol ζ. The induction of G→A mutation is in keeping with a recent crystal structure study of O^6-CMdG-containing duplex DNA showing that the lesion can form base pair with thymidine through Watson–Crick base pairing (Figure 6) (32). In addition, a recent study showed that, among the several mammalian TLS polymerases tested, Pol η catalyzed the most efficient
Figure 6. Proposed base-pairings involved in correct nucleotide incorporations opposite N6-CMdA and N4-CMdC, along with the incorrect nucleotide insertions opposite other carboxymethylated DNA lesions.

N6-CMdA:dT  
N4-CMdC  
O6-CMdG:dT  
O4-CMdT  
N3-CMdT

bypass and extension over O6-CMdG adduct in vitro (33), which differs from our finding about the lack of involvement of Pol η in bypassing the O6-CMdG lesion in human cells. This difference could again be attributed to the difference between replication study with a single purified polymerase in vitro and that with the entire replication machinery in cells.

O4-CMdT and N3-CMdT moderately impede DNA replication in HEK293T cells, with the bypass efficiencies being ~39% and 40%, respectively, and both are highly mutagenic, with the major types of mutations being T→C transition and T→A transversion at frequencies of 72% and 67%, respectively (Figure 5). These observations are in line with the findings made from replication studies conducted using E. coli cells (20). It is also worth noting that O4-CMdT and N3-CMdT could significantly inhibit DNA transcription mediated by T7 RNA polymerase in vitro or mammalian RNA polymerase II in cells, and it was revealed that N3-CMdT was a strong miscoding lesion and predominantly induced mutant transcripts containing a uridine opposite the lesion, whereas O4-CMdT was a moderately miscoding lesion that only directed the misinsertion of guanosine opposite the lesion during transcription (34). Thus, our results supported that these two modified thymidine derivatives exhibit similar miscoding properties during DNA replication and transcription. The highly specific misincorporation of thymidine opposite N3-CMdT is perhaps due to the placement of a carboxyl group at the Watson-Crick hydrogen bonding face, which may foster the base-pairing of the modified nucleobase with thymine (Figure 6).

The G→A, T→C and T→A mutations observed for O6-CMdG, O4-CMdT and N3-CMdT suggest that these lesions may confer severe biological consequences. In this vein, G→A mutation constitutes the most frequent type of mutation observed at non-CpG sites in p53 gene induced by treatment with diazoacetate or in human gastrointestinal cancers (13). In addition, ~43% of all diazoacetate-induced mutations in p53 gene occur at AT base pairs, with AT→TA, AT→GC and AT→CG substitutions occurring at frequencies of 20%, 12%, 10%, respectively (13). These results together suggest that O6-CMdG, O4-CMdT and N3-CMdT may contribute significantly to the GC→AT, AT→TA and AT→GC mutations induced by diazoacetate, and those observed in p53 gene in human gastrointestinal tumors.

In summary, our study provided important new knowledge about the influence of carboxymethylated DNA adducts on perturbing the efficiency and fidelity of DNA replication in cultured human cells, and revealed the roles of different TLS polymerases in bypassing these lesions. Thus, the present study offered significant novel insights into the biological consequences of carboxymethylated DNA lesions by revealing their impact on DNA replication in human cells.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

FUNDING
National Institutes of Health [R01 DK082779]. Funding for open access charge: National Institutes of Health [R01 DK082779].

Conflict of interest statement. None declared.
REFERENCES

1. Louis, P., Hold, G.L., and Flint, H.J. (2014) The gut microbiota, bacterial metabolites and colorectal cancer. *Nat. Rev. Microbiol.*, 12, 661–672.

2. Tricker, A. (1997) *N*-nitroso compounds and man: sources of exposure, endogenous formation and occurrence in body fluids. *Eur. J. Cancer Prev.*, 6, 226–268.

3. Mirvish, S.S. (1995) Role of *N*-nitroso compounds (NOC) and *N*-nitrosation in etiology of gastric, esophageal, nasopharyngeal and bladder cancer and contribution to cancer of known exposures to NOC. *Cancer Lett.*, 93, 17–48.

4. Jakszyn, P., Bingham, S., Pera, G., Agudo, A., Luben, R., Welch, A., Boeing, H., Del Gaudio, G., Palli, D. and Saieva, C. (2006) Endogenous versus exogenous exposure to *N*-nitroso compounds and gastric cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST) study. *Carcinogenesis*, 27, 1497–1501.

5. Newton, C., Graham, A., Heptinstall, L., Powell, S., Summers, C., Tannenbaum, S.R. and Wogan, G.N. (1985) Carcinogenicity in rats of *N*-nitroso compounds (NOC) and *N*-nitrosation in the human colon and occurrence in body fluids. *Chem. Biol.*, 11, 1332–1338.

6. Newton, C., Graham, A., Heptinstall, L., Powell, S., Summers, C., Kalshke, N., Smith, J. and Markham, A. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.*, 17, 2503–2516.

7. Buchy, W.F., Shuker, D.E., Charnley, G., Newberne, P.M., Tannenbaum, S.R. and Wogan, G.N. (1985) Carcinogenicity in rats of *N*-nitrosated glycine derivatives. *Toxicol. Appl. Pharmacol.*, 80, 17–48.

8. Wang, J. and Wang, Y. (2010) Synthesis and characterization of oligodeoxyribonucleotides containing 6′- and 6′,6′-carboxymethylthymidine and their formation in DNA. *Nucleic Acids Res.*, 38, 6774–6784.

9. Wang, J. and Wang, Y. (2009) Chemical synthesis of oligodeoxyribonucleotides containing 6′- and 6′,6′-carboxymethylthymidine and 6′- and 6′,6′- carboxymethyl guanine: Implications for colorectal cancer. *Cancer Res.*, 66, 1859–1865.

10. You, C., Wang, P., Nay, S.L., Wang, J., Dai, X., O’Connor, T.R. and Wang, Y. (2016) Roles of Aag, alkBh2, and alkBh3 in the repair of carboxymethylated and ethylated thymidine lesions. *ACS Chem. Biol.*, 11, 1332–1338.

11. You, C., Swanson, A.L., Dai, X., Yuan, B., Wang, J. and Wang, Y. (2013) *N*-alkylthymidine lesions in human cells. *Nucleic Acids Res.*, 41, 9256–9265.

12. You, C., Swanson, A.L., Dai, X., Yuan, B., Wang, J. and Wang, Y. (2011) In vitro replication studies of carboxymethylated DNA lesions with *Saccharomyces cerevisiae* polO and polE. *Biochemistry*, 50, 7666–7673.

13. You, C., Swanson, A.L., Dai, X., Yuan, B., Wang, J. and Wang, Y. (2013) Synthesis of 6′- and 6′,6′-deoxyguanosine derivatives by DNA polymerases η and ι. *J. Biol. Chem.*, 288, 28548–28556.

14. Zhang, F., Tsunoda, M., Suzuki, K., Kikuchi, Y., Wilkinson, O., Ohmori, H., Hanafusa, T., Ohashi, E. and Vaziri, C. (2009) Separate pathways for the occurrence and biological consequences of DNA adducts. *ACS Chem. Biol.*, 4, 5524–5532.

15. You, C. and Wang, Y. (2016) Mass spectrometry-based quantitative strategies for assessing the biological consequences and repair of DNA adducts. *Acc. Chem. Res.*, 49, 205–213.

16. Liu, S. and Wang, Y. (2015) Mass spectrometry for the assessment of the occurrence and biological consequences of DNA adducts. *Chem. Soc. Rev.*, 44, 7829–7854.

17. Gan, G.N., Wittsieben, J.P., Wittsieben, B.O. and Wood, R.D. (2008) DNA polymerase zeta (polζ) in higher eukaryotes. *Cell Res.*, 18, 174–183.

18. Zhang, F., Tsunoda, M., Suzuki, K., Kikuchi, Y., Wilkinson, O., Millington, C.L., Margison, G.P., Williams, D.M., Morishita, E.C. and Takáneká, A. (2013) Structures of DNA duplexes containing *O*′-carboxymethylguanine, a lesion associated with gastrointestinal cancer, reveal a mechanism for inducing pyrimidine transition mutations. *Nucleic Acids Res.*, 41, 5524–5532.

19. Rüž, M.H., Dextor, H.R., Millington, C.L., Van Loon, B., Williams, D.M. and Sturla, S.J. (2016) Bypass of mutagenic *O*′-carboxymethylguanine DNA adducts by human Y- and B-family polymerases. *Chem. Res. Toxicol.*, 29, 1493–1503.

20. You, C., Wang, J., Dai, X. and Wang, Y. (2015) Transcriptional inhibition and mutagenesis induced by N-nitroso compound-derived carboxymethylated thymidine adducts in DNA. *Nucleic Acids Res.*, 43, 1012–1018.

21. Williams, N.L., Wang, P., Wu, J. and Wang, Y. (2016) In vitro Lesion Bypass Studies of *O*′-alkylthymidines with Human DNA Polymerase η. *Chem. Res. Toxicol.*, 29, 669–675.

22. Wang, P., Amato, N.J., Zhao, Q. and Wang, Y. (2015) Cytotoxic and mutagenic properties of *O*′-alkylthymidines in *Escherichia coli* cells. *Nucleic Acids Res.*, 43, 10795–10803.