Published online 5 April 2011

Nucleic Acids Research, 2011, Vol. 39, No. 14 ***5945–5954***
doi:10.1093/nar/gkr159

High-throughput analysis of the mutagenic and cytotoxic properties of DNA lesions by next-generation sequencing

Bifeng Yuan[1,2], Jianshuang Wang[1], Huachuan Cao[1], Ruobai Sun[3] and Yinsheng Wang[1,*]

[1]Department of Chemistry, University of California, Riverside, CA 92521-0403, USA, [2]College of Chemistry and Molecular Sciences, Wuhan University, Wuhan, 430072, China and [3]Department of Botany and Plant Sciences, University of California, Riverside, CA 92521-0403, USA

Received December 23, 2010; Revised March 2, 2011; Accepted March 5, 2011

ABSTRACT

Human cells are constantly exposed to environmental and endogenous agents which can induce damage to DNA. Understanding the implications of these DNA modifications in the etiology of human diseases requires the examination about how these DNA lesions block DNA replication and induce mutations in cells. All previously reported shuttle vector-based methods for investigating the cytotoxic and mutagenic properties of DNA lesions in cells have low-throughput, where plasmids containing individual lesions are transfected into cells one lesion at a time and the products from the replication of individual lesions are analyzed separately. The advent of next-generation sequencing (NGS) technology has facilitated investigators to design scientific approaches that were previously not technically feasible or affordable. In this study, we developed a new method employing NGS, together with shuttle vector technology, to have a multiplexed and quantitative assessment of how DNA lesions perturb the efficiency and accuracy of DNA replication in cells. By using this method, we examined the replication of four carboxymethylated DNA lesions and two oxidatively induced bulky DNA lesions including (5′*S*) diastereomers of 8,5′-cyclo-2′-deoxyguanosine (cyclo-dG) and 8,5′-cyclo-2′-deoxyadenosine (cyclo-dA) in five different strains of *Escherichia coli* cells. We further validated the results obtained from NGS using previously established methods. Taken together, the newly developed method provided a high-throughput and readily affordable method for assessing quantitatively how DNA lesions compromise the efficiency and fidelity of DNA replication in cells.

INTRODUCTION

Human genome is constantly assaulted by endogenous and exogenous agents (1), among which reactive oxygen species (ROS) can be produced by normal aerobic metabolism, ionizing radiation and anti-tumoral agents (2). Aside from single-nucleobase lesions, ROS could also induce the formation of bulky DNA lesions including 8,5′-cyclo-2′-deoxyguanosine (cyclo-dG) and 8,5′-cyclo-2′-deoxyadenosine (cyclo-dA) (3). In addition to ROS, genomic DNA in living cells is susceptible to damage from exposure to *N*-nitroso compounds (NOCs) in diet, tobacco smoke and other environmental sources as well as from endogenous sources (4). The exposure to endogenous NOCs was found to be significantly associated with the risk of developing cancer (5). Some endogenously produced NOCs can be metabolized to give diazoacetate, which induces the carboxymethylation of DNA (6,7). The accumulation of ROS- and NOC-induced DNA lesions may bear important implications in the pathogenesis of a number of human diseases including cancer and neurodegeneration (5,8). However, the mutagenic properties of these DNA lesions in cells remain unexplored.

Shuttle vector technology has been widely used for examining how a structurally defined DNA lesion affects the efficiency and fidelity of DNA replication in cells (9,10). In this assay, a replicable plasmid harboring a site-specifically inserted and structurally defined lesion is allowed to replicate in host cells. The progeny plasmids are subsequently isolated and transfected into bacterial cells for further amplification and phenotypic selection. Although this type of assay can couple with DNA sequencing to determine the identities and frequencies of mutations, phenotypic assay is indirect and potentially affected by selection bias. It also necessitates scoring a sufficient number of mutations to obtain statistically robust information. Recently, Delaney and Essigmann (9,11) introduced the CRAB (competitive replication and adduct bypass) and REAP (restriction endonuclease and

*To whom correspondence should be addressed. Tel: +1 951 827 2700; Fax: +1 951 827 4713; Email: yinsheng.wang@ucr.edu

post-labeling) assays to assess quantitatively the cytotoxic and mutagenic properties of DNA lesions. In these assays, the entire population of progeny genome is interrogated, which affords statistically sound information about the bypass efficiencies and mutation frequencies, and no phenotypic selection is required. However, lesion-containing M13 genomes are transfected into *Escherichia coli* cells and analyzed one at a time, which is time-consuming.

The development of Sanger DNA sequencing method about 30 years ago has had a profound impact on biological research, and the recent introduction of next-generation sequencing (NGS) has made it feasible to produce a tremendous volume of sequencing data cheaply (12). NGS technology has had a significant impact on genomic research (13,14) and had many applications including whole-genome analysis of cancer cells (15), genome-wide DNA cytosine methylation mapping (16), DNA–protein interaction studies (ChIP-Seq) (17), etc. NGS technology has enabled investigators to design scientific approaches that were previously not technically feasible or affordable. We reason that NGS technology may render it possible to assess the mutagenic and cytotoxic properties of DNA lesions by sequencing a large number of DNA molecules without tedious phenotypic scoring. We also envision that, with the numerous reads produced cheaply and rapidly by NGS and with a bar-coding strategy, statistically sound results for the bypass efficiencies and mutation frequencies of multiple DNA lesions might be obtained from a single-sequencing experiment.

In this study, we established an NGS coupled with shuttle vector technology for high-throughput and cost-effective discovery of how DNA lesions compromise DNA replication in cells. Using this method, we assessed the mutagenic and cytotoxic properties of four carboxymethylated DNA lesions, N^4-carboxymethyl-2′-deoxycytidine (N^4-CMdC), N^6-carboxymethyl-2′-deoxyadenosine (N^6-CMdA), O^4-carboxymethylthymidine (O^4-CMdT) and *N*3-carboxymethylthymidine (*N*3-CMdT), and two oxidatively induced bulky DNA lesions, (5′*S*)-cyclo-dG and (5′*S*)-cyclo-dA (Figure 1).

MATERIALS AND METHODS

Chemicals and bacterial strains

Unmodified oligodeoxyribonucleotides (ODNs) used in this study were purchased from Integrated DNA Technologies (Coralville, IA, USA). [γ-^{32}P]ATP was obtained from Perkin Elmer (Piscataway, NJ, USA). Shrimp alkaline phosphatase was obtained from USB Corporation (Cleveland, OH, USA), and all other enzymes were from New England Biolabs (Ipswich, MA, USA). 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) was purchased from TCI America (Portland, OR, USA). Chemicals unless otherwise noted were obtained from Sigma-Aldrich (St Louis, MO, USA). M13mp7(L2) and wild-type AB1157 *E. coli* strains were kindly provided by Prof. John M. Essigmann, and polymerase-deficient AB1157 strains [Δ*pol B1*::spec (pol II-deficient), *ΔdinB* (pol IV-deficient), *ΔumuC*::kan (pol V-deficient) and *ΔumuC*::kan *ΔdinB* (pol IV, pol V-double knockout)] were generously provided by Prof. Graham C. Walker (18).

Preparation of ODN substrates containing a modified DNA lesion

The 12-mer lesion-containing ODNs 5′-ATGGCGXGCT AT-3′ ('X' represents modified nucleoside) were synthesized following previously published procedures (19–21). The identities of the modified ODNs were confirmed by electrospray ionization-mass spectrometry (ESI-MS) and tandem mass spectrometry (MS) analyses (Supplementary Figures S1–S3). To differentiate the progeny vectors for individual lesions after *in vivo* replication, a 10-mer ODN

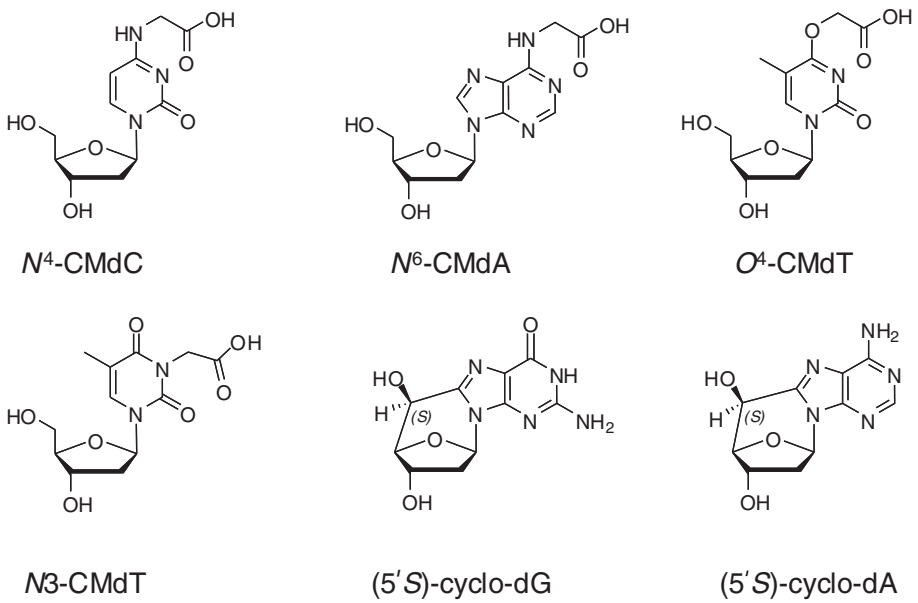

Figure 1. The structures of N^6-CMdC, N^6-CMdA, O^4-CMdT, *N*3-CMdT, (5′*S*)-cyclo-dG and (5′*S*)-cyclo-dA.

with a dinucleotide barcode (5′-GCAGGATGBB-3′, 'BB' represents barcode) was ligated to the 12-mer lesion-bearing ODN and the resulting ligation product was purified by denaturing PAGE (The 22-mer sequences are listed in Table 1). The identities of the modified 22-mer ODNs were again confirmed by ESI-MS and tandem MS analyses.

Construction of single-stranded M13 genomes harboring a site-specifically inserted DNA lesion

The M13mp7(L2) viral genomes, either lesion-free or carrying a site-specifically inserted DNA lesion, were prepared following the previously described procedures (11). Briefly, 20 pmol of single-stranded (ss) M13mp7(L2) was digested with 40 U EcoRI at 23°C for 8 h to linearize the vector. Two scaffolds, 5′-CATCCTGCCACTGAATCAT GGTCATAGCTTTC-3′ and 5′-AAAACGACGGCCAG TGAATTATAGC-3′ (25 pmol), each spanning one end of the cleaved vector and the modified ODN insert, were annealed with the linearized vector. The 22-mer insert (30 pmol, 5′-GCAGGATGBBATGGCGXGCTAT-3′, where 'X' and 'BB' represent modified nucleoside and the lesion-specific barcode, respectively) was 5′-phosphorylated with T4 polynucleotide kinase. The 5′-phosphorylated 22-mer inserts were ligated to the above vector by using T4 DNA ligase in the presence of the two scaffolds at 16°C for 8 h. T4 DNA polymerase (22.5 U) was subsequently added and the resulting mixture was incubated at 37°C for 4 h to degrade the scaffolds and residual unligated vector. The solution was extracted with phenol/chloroform/isoamyl alcohol (25:24:1, v/v), and the aqueous phase was passed through the QIAquick PCR Purification column (Qiagen) to remove residual phenol and salt. The constructed genomes were normalized against a lesion-free competitor genome, which was prepared by inserting a 25-mer unmodified ODN (Table 1) into the EcoRI-linearized genome, following the procedures described by Delaney and Essigmann (11)

Transfection of *E. coli* cells with ssM13 vectors containing a DNA lesion

Desalted N^4-CMdC-, N^6-CMdA-, O^4-CMdT-, *N*3-CMdT-, (5′*S*)-cyclo-dG- and (5′*S*)-cyclo-dA-containing as well as control M13 genomes were mixed at 1:1 ratio (25 fmol each) and transfected into SOS-induced wild-type AB1157 *E. coli* cells and the isogenic *E. coli* cells that are deficient in pol II, pol IV, pol V, or both pol IV and pol V. The electrocompetent SOS-induced cells were prepared following the previously published procedures (22). After transfection, the *E. coli* cells were grown in LB culture at 37°C for 6 h, after which the phage was recovered from the supernatant by centrifugation at 13 000 r.p.m. for 5 min. The resulting phage was further amplified in SCS110 *E. coli* cells to increase the progeny/lesion-genome ratio (11). The phage recovered from the supernatant was passed through a QIAprep Spin M13 column (Qiagen) to isolate the ssM13 DNA.

Table 1. The sequences of the 22-mer lesion-containing and the control lesion-free ODNs used for replication studies

Name of ODNs	Sequences
N^4-CMdC	5′-GCAGGATG**CG**ATGGCG**X**GCTAT-3′
N^6-CMdA	5′-GCAGGATG**AT**ATGGCG**X**GCTAT-3′
O^4-CMdT	5′-GCAGGATG**TC**ATGGCG**X**GCTAT-3′
*N*3-CMdT	5′-GCAGGATG**TG**ATGGCG**X**GCTAT-3′
Cyclo-dG	5′-GCAGGATG**GA**ATGGCG**X**GCTAT-3′
Cyclo-dA	5′-GCAGGATG**AG**ATGGCG**X**GCTAT-3′
Control	5′-GCAGGATG**AC**ATGGCG**C**GCTAT-3′
Competitor	5′-GCAGGATGCGATGGCGATAAGCTAT-3′

'X' in bold designates DNA lesion and the dinucleotide barcode is in underlined bold font.

Generation of sequencing library and determination of the bypass efficiency and mutation frequency using NGS

The sequencing library was generated using NEBNext® DNA Sample Prep Master Mix Set 1 (New England Biolabs, Ipswich, MA, USA; Figure 2). Briefly, 15 sets of primers each housing a unique dinucleotide barcode (Supplementary Table S1), which designated host cell lines or individual biological replicates, were employed to generate polymerase chain reaction (PCR) products from the progeny vectors. PCR amplification of the region of interest in the resulting progeny genome was performed by using Phusion high-fidelity DNA polymerase (New England Biolabs) and running at 98°C for 60 s and 15 cycles at 98°C for 10 s, 46°C for 30 s and 72°C for 5 s, with a final extension at 72°C for 5 min. The 15 sets of PCR products were purified by QIAquick Nucleotide Removal Kit (Qiagen) and then mixed at equal amounts. The PCR mixture was phosphorylated at the 5′-end using T4 polynucleotide kinase. A single 'A' nucleotide was added to the 3′ end of the PCR products and the resulting purified PCR mixture was ligated to two paired-end (PE) Adapters (Table S1). The ligation products were further amplified using PE PCR primers (Supplementary Table S1). The PCR amplification was performed at 98°C for 60 s and 15 cycles at 98°C for 10 s, 70°C for 30 s and 72°C for 5 s, with a final extension at 72°C for 5 min. The resulting PCR products (166 bp) were gel-purified and subjected to NGS using Illumina Genome Analyzer IIe system (Illumina, San Diego, CA, USA).

After obtaining the raw sequencing data, the reads that failed to pass the Illumina chastity filter were removed. The low-quality reads which contained >1 nt with a quality score below 20 or any undefined nucleotide 'N' were further filtered and removed. Only the reads with perfect match to characteristic strings 'ATTCAGTGGC AGGATG' from the 3 rd–18th nucleotides and 'ATGG' from the 21st–24th nucleotides for forward sequence reads, or the reads with perfect match to characteristic strings 'CGGCCAGTGAATTATAG' from the 3 rd–19th nucleotides and 'CCAT' from the 24th–27th nucleotides for reverse sequence reads were selected for analysis of barcode distribution. An 'R' script was used to specify cell line/biological replicate barcode or lesion-related barcode at a given position, and to calculate the nucleobase (A, T, C or G) frequencies at the specific lesion site. The bypass efficiency was calculated using

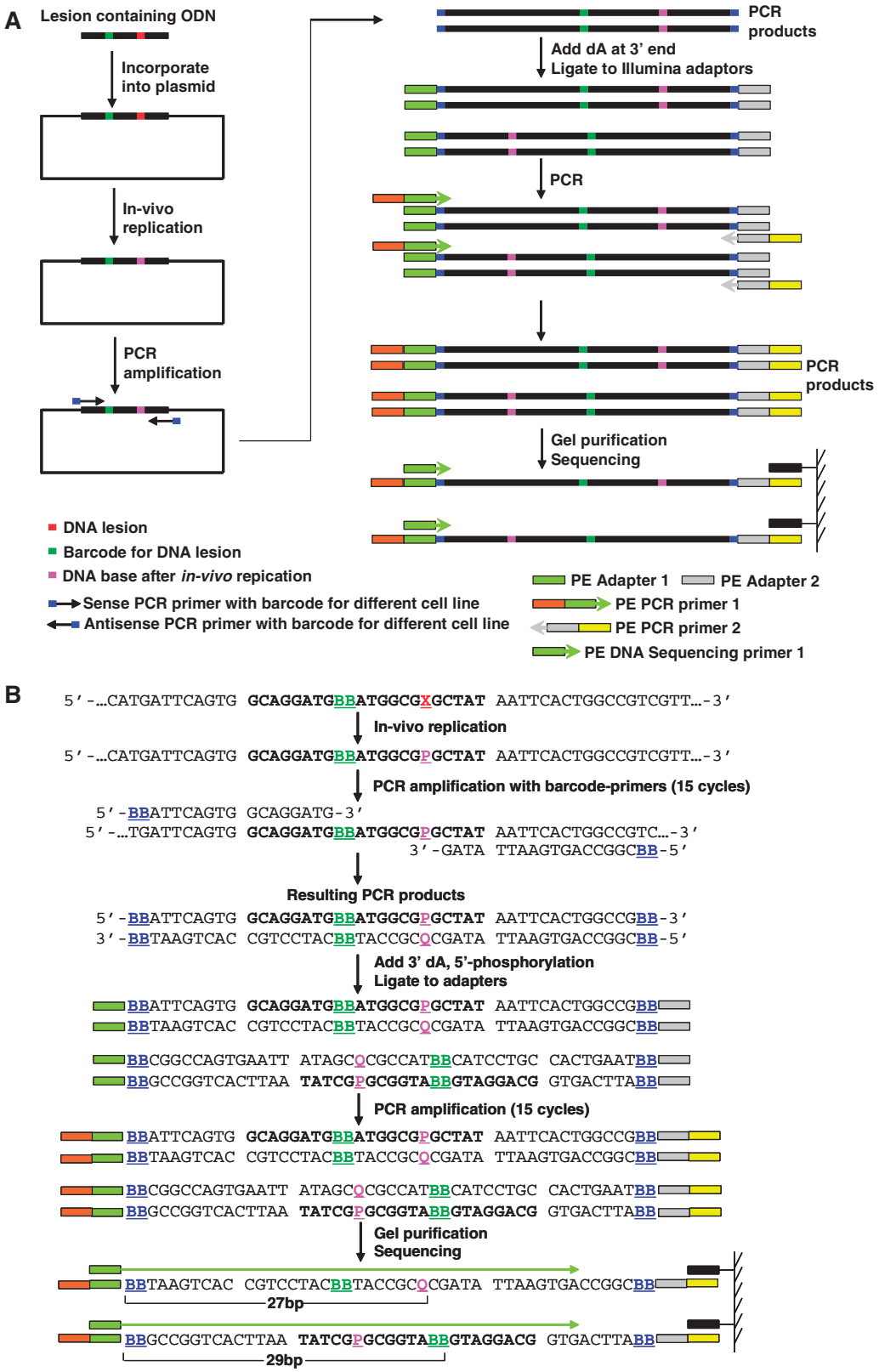

Figure 2. (**A**) A schematic diagram shows the experimental procedures. The 22-mer lesion-containing ODNs with barcodes were ligated to the EcoR I-linearized M13 vector, mixed at equal amounts and subjected to *in vivo* replication. The harvested M13 progenies were amplified with barcoded PCR primers, and equal amounts of PCR products from different cell lines were mixed and ligated with PE adapters. The ligation products were further amplified using PE PCR primers, and the resulting PCR products were purified and subjected to NGS analysis. (**B**) Detailed sequence information for the PCR amplification of progeny genome, adapter ligation, and PCR amplification for the construction of sequencing library. 'X', 'P' and 'Q' represent the lesion, the DNA base after *in vivo* replication of DNA lesion, and the paired nucleobase of 'P' in the complementary strand, respectively. 'BB' in green and blue represent lesion-specific barcode and the barcode for host cell lines and biological replicates, respectively.

the following formula, %bypass = (total number of reads from lesion genome) / (total number of reads from control genome) × 100%. The percentages of base substitution at lesion site were calculated using the following formula, %base substitution = (total number of reads of A, T, C or G at original lesion site from lesion genome) / (total number of reads from lesion genome) × 100%.

Determination of bypass efficiency using CRAB assay

The bypass efficiencies of O^4-CMdT, (5′*S*)-cyclo-dG and (5′*S*)-cyclo-dA were further evaluated by employing CRAB assay developed by Delaney and Essigman (11) The transfection and *in vivo* replication of lesion-containing M13 vectors were conducted using previously described methods (11). PCR amplification of the region of interest in the resulting progeny genome was performed by using Phusion high-fidelity DNA polymerase. The primers were 5′-YCAGCTATGACCATGATTC AGTGCCATG-3′ and 5′-YTCGGTGCGGGCCTCTTC GCTATTAC-3′ (Y is an amino group), and the amplification cycle was 30, each consisting of 10 s at 98°C, 30 s at 62°C and 15 s at 72°C, with a final extension at 72°C for 5 min. The PCR products were purified by using QIAquick PCR purification kit (Qiagen).

For the bypass efficiency assay, a portion of the above PCR fragments was treated with 10 U Tsp509I in 10-μl NEB buffer 2 at 65°C for 30 min and 1 U shrimp alkaline phosphatase at 37°C for 30 min, followed by heating at 65°C for 20 min to deactivate the shrimp alkaline phosphatase. The above mixture was then treated in a 15-μl NEB buffer 2 with 5 mM DTT, ATP (50 pmol cold, premixed with 1.66 pmol [γ-^{32}P]ATP) and 10 U T4 polynucleotide kinase. The reaction was continued at 37°C for 30 min, followed by heating at 65°C for 20 min to deactivate the T4 polynucleotide kinase. To the reaction mixture was subsequently added 10 U BtsCI, and the solution was incubated at 37°C for 30 min, followed by quenching with 15-μl formamide gel loading buffer containing xylene cyanol FF and bromophenol blue dyes. The mixture was loaded onto a 30% native polyacrylamide gel (acrylamide:*bis*-acrylamide = 19:1) and products were quantified by phosphorimager analysis. After the restriction cleavages, the original lesion site was housed in a 12-mer/18-mer duplex, d(pATGGCGPGCTAT)/ d(p*A ATTATAGCQCGCCATBB), where 'P' represents the nucleobase incorporated at the initial damage site during *in vivo* DNA replication, 'Q' is the paired nucleobase of 'P' in the complementary strand, and 'p*' designates the 5′-radiolabeled phosphate (Supplementary Figure S4). The 18-mer products were monitored instead of the 12-mer products because the latter products co-migrated with non-specific bands. The bypass efficiency was calculated using the following formula, %bypass = (lesion signal/competitor signal)/(non-lesion control signal/competitor signal) (11). The mutation frequencies were determined by liquid chromatography-tandem mass spectrometry (LC-MS/MS) since the 18-mers bearing a single nucleobase difference could not be well-resolved by PAGE.

Determination of bypass efficiency and mutation frequency using LC-MS/MS

In order to identify the replication products using LC-MS/MS, PCR products were treated with 50 U BtsCI and 20 U shrimp alkaline phosphatase in 250-μl NEB buffer 2 at 37°C for 2 h, followed by heating at 65°C for 20 min. To the resulting solution was added 50 U of Tsp509I, and the reaction mixture was incubated at 65°C for 1 h followed by extraction once with phenol/chloroform/isoamyl alcohol (25:24:1, v/v). The aqueous portion was dried with Speed-vac, desalted with high-performance liquid chromatography (HPLC) and dissolved in 12-μl water. The ODN mixture was subjected to LC-MS/MS analysis. A 0.5 × 150 mm Zorbax SB-C18 column (5 μm in particle size, Agilent Technologies) was used for the separation and the flow rate was 8.0 μl/min, which was delivered by using an Agilent 1100 capillary HPLC pump. A 5-min gradient of 0–20% methanol followed by a 35-min of 20–50% methanol in 400 mM 1,1,1,3,3,3-HFIP, (pH was adjusted to 7.0 by the addition of triethylamine) was employed for the separation. The effluent from the LC column was coupled directly to an LTQ linear ion trap mass spectrometer (Thermo Electron, San Jose, CA, USA), which was set up for monitoring the fragmentation of the $[M\text{-}3\,H]^{3-}$ ions of the 12-mer [d(ATGGCGPGCTAT), where 'P' designates A, T, C or G] and the $[M\text{-}4\,H]^{4-}$ ion of the 15-mer [i.e. d(ATGGCGATAAGCTAT)] ODNs.

RESULTS

Experimental strategy

Our strategy for high-throughput mutagenesis study involves a combination of NGS with shuttle vector technology, as depicted in Figure 2. Following previously published procedures (23–25), we constructed the ssM13 shuttle vectors carrying structurely defined lesions at a specific site and normalized the relative amounts of the lesion-containing genomes. Six lesion-bearing and one control M13 genomes were mixed together and transfected simultaneously into *E. coli* cells. To illustrate the roles of various translesion synthesis DNA polymerases in bypassing these lesions *in vivo*, we employed wild-type AB1157 *E. coli* cells as well as the isogenic strains deficient in pol II, pol IV, pol V or both pol IV and pol V as the host cells for the replication experiments. After *in vivo* replication, the ssM13 progeny vectors were isolated. Fifteen pairs of barcoded primers (Supplementary Table S1), which designated 15 distinct sets of progeny genomes arising from triplicate replication experiments in five different host cell lines, were employed to generate PCR products from the progeny vectors. The 15 sets of PCR products were then mixed at equal amounts and the resulting PCR product mixture was phosphorylated at the 5′-end, adenylated at the 3′-end, and ligated to PE adapters 1 and 2 (Supplementary Table S1). The ligation products were further amplified using PE PCR primers (Supplementary Table S1), and the resulting PCR products (166bp) were gel-purified and subjected to

NGS analysis using Illumina Genome Analyzer IIe system. From the sequencing results, we determined the mutagenic and cytotoxic properties of multiple DNA lesions in different bacterial hosts by interrogating the distribution of barcodes and nucleobase (A, T, C or G) frequencies at the specific lesion site. In addition, the sequencing reads obtained for the lesion-containing genomes relative to the lesion-free genome allowed for the calculation of bypass efficiencies for the lesions.

Synthesis of lesion-containing ODNs

Previous studies demonstrated that potassium diazoacetate was capable of inducing N^4-CMdC, N^6-CMdA, O^4-CMdT and *N*3-CMdT in isolated DNA (20,21). In addition, ROS-induced bulky DNA lesions including cyclo-dG and cyclo-dA could be detected in mammalian cells (26–28) (Figure 1), though a recent study suggested that the cellular levels of cyclo-dG and cyclo-dA might be lower than those measured previously (29). However, it remains unexplored how these lesions compromise the fidelity and efficiency of DNA replication *in vivo*. Such studies necessitate the availability of ODNs containing site-specifically incorporated DNA lesions. To this end, we employed traditional phosphoramidite chemistry and synthesized N^4-CMdC-, N^6-CMdA-, O^4-CMdT-, *N*3-CMdT-, cyclo-dG- and cyclo-dA-containing ODNs, 5′-ATGGCGXGCTAT-3′ ('X' represents modified nucleoside) (19–21). After HPLC purification, the purities and identities of these lesion-bearing ODNs were confirmed by ESI-MS and tandem MS (MS/MS) analyses (Figures S1–S3). The lesion-bearing 12-mer ODNs were then ligated with barcode-containing 10-mer ODNs to yield the 22-mer lesion-containing ODNs (Table 1).

NGS analysis of the bypass efficiencies and mutation frequencies of DNA lesions

In this study, we mixed six lesion-containing M13 genomes and a control lesion-free M13 genome and allowed them to replicate in five different *E. coli* strains. We obtained a total of 9.6 million valid sequencing reads for the replication products of these genomes and Supplementary Table S2 shows the typical number of reads obtained for replication products isolated from wild-type *E. coli* cells. Even with the most blocking DNA lesion, i.e. (5′*S*)-cyclo-dG, we still obtained about 10 000 reads in a single replicate experiment (Supplementary Table S2), which is much more than what can be achieved with traditional colony picking and Sanger sequencing method.

The bypass efficiencies were calculated from the ratio of the total number of reads from lesion genome over the total number of reads from the control genome. It turned out that N^4-CMdC and N^6-CMdA did not block DNA replication in wild-type AB1157 *E. coli* cells, with the bypass efficiencies being ~83% and 98%, respectively (Figure 3A). In addition, deficiency in pol II, pol IV or pol V in the isogenic AB1157 background did not affect considerably the bypass efficiencies for these two lesions (Figure 3A). O^4-CMdT and *N*3-CMdT, on the other hand, block appreciably DNA replication in wild-type AB1157 *E. coli* cells, with the bypass efficiencies being ~49% and 55%, respectively (Figure 3A). Deficiency in pol II, pol IV or pol V in the isogenic AB1157 background did not compromise the bypass efficiency for *N*3-CMdT (Figure 3A). Although deficiency in pol II or pol IV in the isogenic AB1157 background did not alter substantially the bypass efficiencies of O^4-CMdT, deficiency in pol V alone or in combination with pol IV decreased the bypass efficiencies to 22% and 19%, respectively, indicating that pol V may be involved in the bypass of O^4-CMdT (Figure 3A).

Cyclo-dG and cyclo-dA inhibited substantially the DNA replication in wild-type AB1157 cells, with the bypass efficiencies being ~11% and 31%, respectively (Figure 3B). Deficiency in pol II or pol IV did not affect considerably the bypass efficiencies for these two lesions; however, depletion of pol V alone or in conjunction with pol IV gave rise to further declines in bypass efficiencies of cyclo-dG to 6% and 4%, respectively. Likewise, the bypass efficiencies of cyclo-dA dropped to 13% and 10% in pol V-deficient and pol IV, pol V-double knockout cells, respectively. These data supported that pol V is the major DNA polymerase involved in the bypass of cyclo-dG and cyclo-dA in *E. coli* cells (Figure 3B).

The results from NGS data also allowed us to assess the mutation frequencies of DNA lesions in wild-type and bypass polymerase-deficient *E. coli* strains. The quantification data showed that: (i) Neither N^4-CMdC nor N^6-CMdA was mutagenic; (ii) both O^4-CMdT and *N*3-CMdT were highly mutagenic in wild-type *E. coli* cells, with $T \rightarrow C$ transition and $T \rightarrow A$ transversion occurring at frequencies of 86% and 66%, respectively; (iii) cyclo-dG and cyclo-dA were mutagenic in wild-type *E. coli* cells, with the major types of mutations being $G \rightarrow A$ transition and $A \rightarrow T$ transversion at frequencies of 20% and 11%, respectively. The deficiency in SOS-induced polymerases did not confer significant alteration in the mutation frequencies of all these DNA lesions except for *N*3-CMdT, where the deficiency in pol V, by itself or along with pol IV, resulted in significant increases in $T \rightarrow A$ mutation (Figure 3C–H).

It is worth noting that deficiency in pol V led to a decreased bypass efficiency, but did not give rise to an appreciable change in mutation frequency of O^4-CMdT. A lack of alteration in mutation frequency was also observed previously for other DNA lesions including S^6-methythioguanine and guanine-S^6-sulfonic acid in pol V-deficient background, whereas decreased bypass efficiencies were found for both lesions in pol V-deficient cells (24). The exact reason behind these observations is unclear, though it is possible that the coding property of O^4-CMdT might be interpreted similarly by pol V in the wild-type background and other polymerase(s) that are involved in bypassing this lesion in the pol V-deficient background.

We also found an appreciable drop in $T \rightarrow A$ mutation for *N*3-CMdT in pol IV-deficient cells, whereas the bypass efficiency was not perturbed by the deficiency in this polymerase. This result suggests that pol IV might be involved in the mutagenic bypass of this lesion in wild-type background. A very similar finding was made by a previous

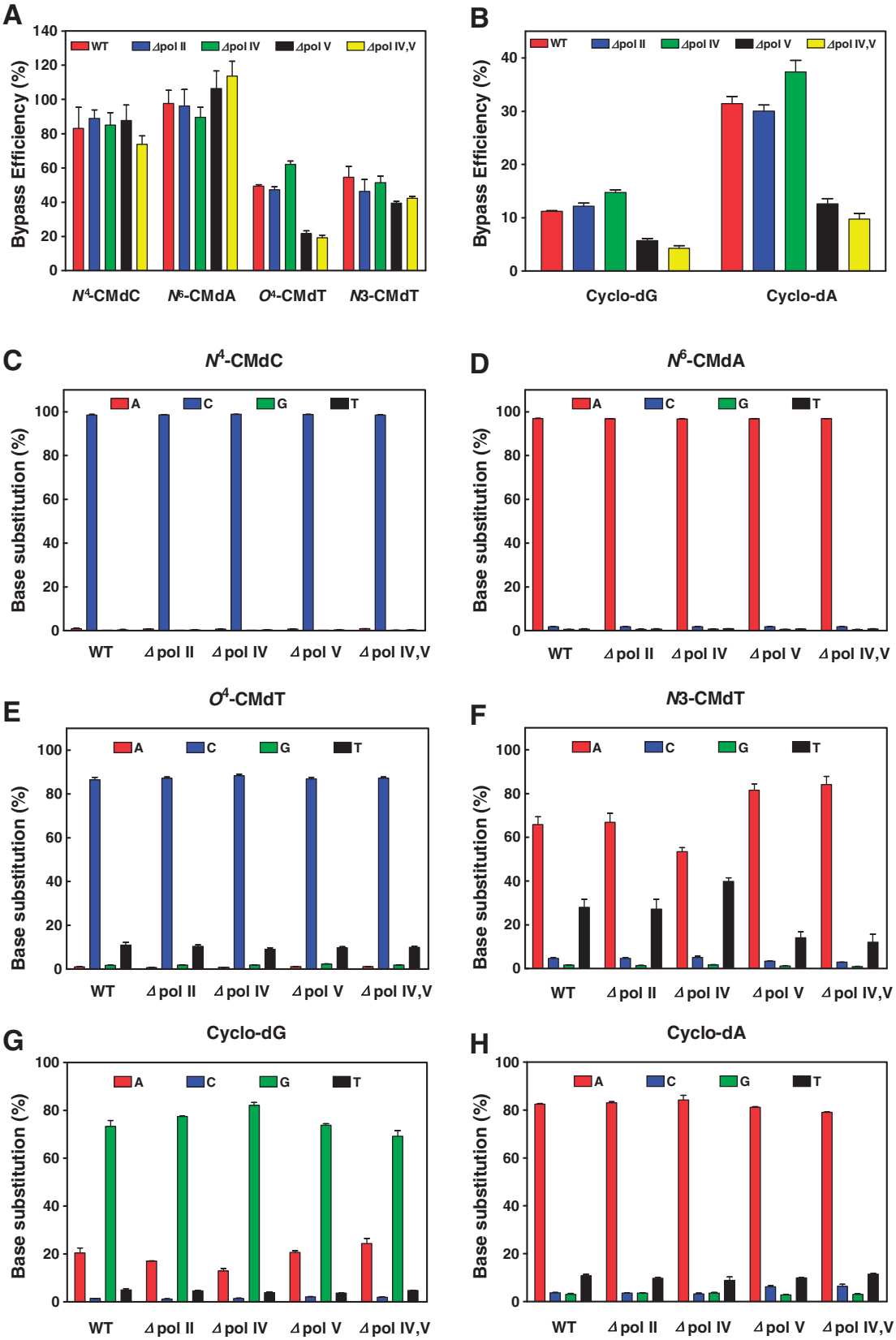

Figure 3. (**A**) Bypass efficiencies of N^4-CMdC, N^6-CMdA, O^4-CMdT and *N*3-CMdT. (**B**) Bypass efficiencies of (5′*S*)-cyclo-dG and (5′*S*)-cyclo-dA. (**C–H**) Mutation frequencies of N^4-CMdC, N^6-CMdA, O^4-CMdT, *N*3-CMdT, (5′*S*)-cyclo-dG and (5′*S*)-cyclo-dA. The data represent the means and standard deviations of results from three independent experiments.

elegant study by Neeley *et al.* (22), where the mutation spectra of 5-guanidino-4-nitroimidazole (NI) were observed to be significantly different in SOS-induced wild-type strain versus the corresponding pol II-deficient strain; however, the bypass efficiencies for this lesion were very similar in these two strains. The exact reason behind our observation is unclear, though we speculate that other polymerase(s) might be induced at a higher level in pol IV-deficient background than in the wild-type background, which may compensate for the decrease in bypass efficiency induced by the absence of pol IV.

Next, we compared the bypass efficiencies of O^4-CMdT, cyclo-dG and cyclo-dA under uninduced and SOS-induced conditions in wild-type AB1157 cells by using CRAB assay (Supplementary Figure S4). Compared to uninduced conditions, the quantitative results showed that the bypass efficiencies of O^4-CMdT, cyclo-dG and cyclo-dA in SOS-induced cells increased from 10% to 52%, 3% to 14%, 5% to 29%, respectively, which are corroborated by results obtained from LC-MS/MS measurements (Figure 4A and Supplementary Figure S4C). The 4- to 6-fold elevation in bypass efficiencies for these lesions in SOS-induced cells supported that higher level of expression of SOS-induced polymerases stimulated the bypass of these lesions. In addition, the mutation rates and patterns of O^4-CMdT and cyclo-dA are similar in uninduced and SOS-induced wild-type cells. However, the G→A mutation induced by cyclo-dG decreased from ~40% in uninduced cells to ~20% in SOS-induced cells.

Determination of bypass efficiencies using CRAB assay

To confirm the results obtained by NGS, we further examined the bypass efficiencies of O^4-CMdT, cyclo-dG and cyclo-dA by employing CRAB assay (Supplementary Figure S4). It turned out that the bypass efficiencies of O^4-CMdT, cyclo-dG and cyclo-dA in wild-type AB1157 *E. coli* cells, obtained using CRAB assay, were ~52%, 14% and 29%, respectively, which were very similar to those obtained from NGS analysis (49%, 11% and 31%, respectively; Figure 4A).

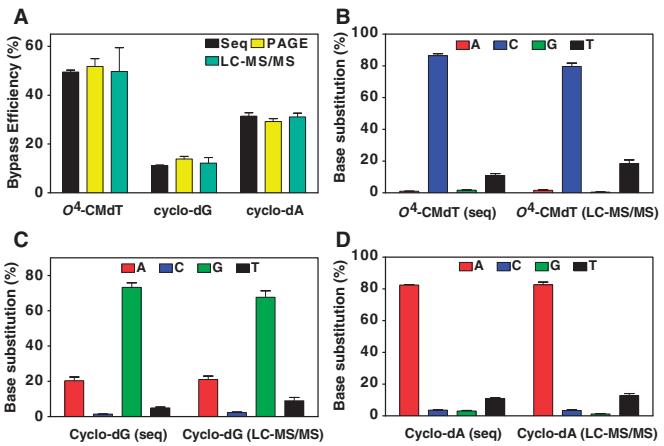

Figure 4. (**A**) Bypass efficiencies of O^4-CMdT, (5′*S*)-cyclo-dG and (5′*S*)-cyclo-dA measured by three different methods, NGS, CRAB assay and LC-MS/MS. (**B–D**) Mutation frequencies of O^4-CMdT, cyclo-dG and cyclo-dA measured by NGS and LC-MS/MS. The data represent the means and standard deviations of results from three independent experiments.

Determination of bypass efficiencies and mutation frequencies using LC-MS/MS

We also validated the bypass efficiencies and mutation frequencies obtained from NGS using our previously reported LC-MS/MS method (23–25). In this respect, the restriction digestion mixture was analyzed by LC-MS/MS and we monitored the fragmentation of the $[M - 3H]^{3-}$ ions of d(ATGGCGPGCTAT), where 'P' is an A, T, C or G, and the $[M - 4H]^{4-}$ ion of d(ATGGCG ATAAGCTAT) (Supplementary Figures S5 and S6). We then quantified the mutation frequencies and bypass efficiencies based on the relative amounts of different replication products with the consideration of differences in ionization and fragmentation efficiencies for different ODNs [LC-MS/MS data are shown in Figures S5 and S6, and calibration curves are depicted in Supplementary Figure S7]. It turned out that the bypass efficiencies and mutation frequencies for O^4-CMdT, cyclo-dG and cyclo-dA were consistent with what we found from NGS analysis (Figure 4).

DISCUSSION

NGS technology has found its applications in many aspects of biological research; however, it has not been employed for assessing how DNA lesions compromise DNA replication in cells. In the current study, we developed, for the first time, a high-throughput and cost-effective method by employing NGS in conjunction with shuttle vector technology for examining how carboxymethylated DNA adducts and ROS-induced bulky DNA lesions impede the progression of DNA replication and induce mutations in *E. coli* cells.

With this method, we demonstrated that N^4-CMdC and N^6-CMdA did not block DNA replication or induce mutations in *E. coli* cells (Figure 3). Our previous primer extension experiments showed that N^4-CMdC, but not N^6-CMdA, inhibited markedly primer extension mediated by the Klenow fragment of *E. coli* DNA polymerase I (20). Klenow fragment incorporated readily the wrong nucleotide, dAMP, opposite N^6-CMdA, and the enzyme also induced the misinsertion of dAMP and dTMP opposite N^4-CMdC (20). Several factors may contribute to the observed differences in nucleotide incorporation opposite N^4-CMdC and N^6-CMdA with Klenow fragment and in *E. coli* cells. First, DNA replication in *E. coli* cells may require both pol I and pol III (30). Second, the *in vitro* measurements were carried out in the presence of one kind of nucleotide at a time, which is different from *in vivo* replication conditions where all four nucleotides are mutually present. Third, *in vivo* DNA replication often involves the participation of auxiliary protein factors, which can alter both the efficiency and accuracy of nucleotide insertion by DNA polymerases (31).

The bypass efficiencies for O^4-CMdT and *N*3-CMdT in wild-type AB1157 *E. coli* cells are ~49% and 55%, respectively. Both O^4-CMdT and *N*3-CMdT are highly mutagenic in wild-type AB1157 cells, with the major types of mutations being T → C transition and T → A transversion at frequencies of 86% and 66%, respectively (Figure 3). Previous studies revealed that diazoacetate could lead to the formation of O^4-CMdT and *N*3-CMdT in isolated DNA (21). In addition, the passage of diazoacetate-treated, human *p53* gene-containing plasmid in yeast cells could give rise to a mutation spectrum where the types and frequencies of mutations observed at non-CpG sites were strikingly similar to those found for *p53* gene mutations in human gastrointestinal tumors (32). This result suggests that diazoacetate might constitute an important etiological agent for gastrointestinal cancer development. In addition, ~43% of all mutations occurred at AT base pairs, with AT → TA, AT → GC and AT → CG substitutions occurring at frequencies of 20%, 12%, 10%, respectively (32). The high frequencies of T → C and T → A mutations found for O^4-CMdT and *N*3-CMdT suggest that these lesions may contribute to *p53* mutations induced by diazoacetate and found in human gastrointestinal tumors.

Our NGS data also revealed that (5′*S*)-cyclo-dG and (5′*S*)-cyclo-dA blocked strongly the DNA replication in *E. coli* cells, which is in line with previous studies showing that cyclo-dA is a strong blockade to T7 DNA polymerase and mammalian DNA polymerase δ *in vitro* (33), and to RNA polymerase II in mammalian cells (34). Additionally, cyclo-dG and cyclo-dA were mutagenic in *E. coli* cells, with the major types of mutations being G → A transition and A → T transversion at frequencies of 20% and 11%, respectively. Cyclo-dG and cyclo-dA, in which C8 of a purine base is covalently bonded to the C5′ of 2-deoxyribose in the same nucleoside, are formed from hydroxyl radical attack (33,35). Because of the presence of a covalent bond between 2-deoxyribose and purine moieties, these lesions are not repaired by base excision repair system but by the nucleotide excision repair pathway (33,34). This additional covalent bond causes local structural distortion to the DNA helix (33), which may compromise the base pairing capabilities of these DNA lesions thereby inducing mutations during DNA replication *in vivo*. It is worth noting that we attempted but failed to find deletion or off-target mutation for (5′*S*)-cyclo-dG and (5′*S*)-cyclo-dA. Considering that cyclo-dG and cyclo-dA can be detected in mammalian cells (26–28), the cytotoxic and mutagenic properties of cyclo-dG and cyclo-dA suggest that the formation and accumulation of these lesions *in vivo* may bear significant pathological consequences.

Introducing barcodes into M13 genome allowed for the high-throughput evaluation of the cytotoxic and mutagenic properties of DNA lesions. In the present study, dinucleotide barcodes were used to represent different DNA lesions and cell line hosts, and the replication of up to 16 different lesion-containing genomes in up to 16 different cell lines can be analyzed simultaneously. Expanding the length of the barcode sequence can further increase the numbers of lesions investigated and the number of host cells studied, thereby improving further the throughput of the method. On the Illumina platform, the single-end read-lengths are typically up to 40 bp; longer reads are possible but may incur a higher error rate. In our case, a valid read should include a cell line barcode, the lesion barcode and the lesion site. In this regard, 27 bp of forward sequence and 29 bp of reverse sequence (Figure 2B), which were fully covered in the 40-bp sequencing range, satisfied the requirement for a valid read.

Lastly, it is worth noting that we only assessed the mutagenic and cytotoxic properties of the six DNA lesions in a single-sequence context, and it is possible that the bypass efficiencies and mutation frequencies of DNA lesions may differ in different sequence contexts. The NGS-based method developed in the present study can be applicable for assessing the effects of sequence context on DNA replication in the future. Similar as traditional CRAB assay (9), the NGS method reported here can also be employed for investigating how DNA lesions are repaired in cells. Additionally, with the use of a double-stranded vector, the method can be adapted for examining the cytotoxic and mutagenic properties of DNA lesions in mammalian cells (36). Taken together, our current study demonstrated that NGS, combined with shuttle vector technology, provided a high-throughput and cost-effective method to uncover the cytotoxic and mutagenic properties of DNA lesions.

It is worth noting that the method reported in this study also bears some limitations, which arises primarily from the error rate introduced by the NGS method. In this context, we observed an error rate of 1.2% for the control genome (Supplementary Table S2), which is higher than the error rate obtained by the Sanger sequencing method and slightly higher than the average sequencing error rate on the Illumina platform (37). The latter could be attributed, in part, to the sequencing error produced at the barcode sites. Considering the error rate of 1.2%, a lesion with an induced mutation frequency that is >3–4% could not be investigated with the strategy described in this paper. Nevertheless, future improvements in sequencing accuracy of NGS, which has been continuously improving since its inception, and the use of longer barcode sequence are expected to improve the accuracy in determining the mutation frequency.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. John M. Essigmann and Prof. Graham C. Walker for providing the *E. coli* strains used in the present study and Dr Glenn Hicks, Dr John Weger and Dr Thomas Girke for assistance with the next-generation sequencing and data processing.

FUNDING

The National Institutes of Health (R01 CA101864 and R01 DK082779). Funding for open access charge: The National Institute of Diabetes and Digestive and Kidney Diseases/NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Lindahl,T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
2. Finkel,T. and Holbrook,N.J. (2000) Oxidants, oxidative stress and the biology of ageing. *Nature*, **408**, 239–247.
3. Wang,Y. (2008) Bulky DNA lesions induced by reactive oxygen species. *Chem. Res. Toxicol.*, **21**, 276–281.
4. Tricker,A.R. (1997) *N*-nitroso compounds and man: sources of exposure, endogenous formation and occurrence in body fluids. *Eur. J. Cancer Prev.*, **6**, 226–268.
5. Jakszyn,P., Bingham,S., Pera,G., Agudo,A., Luben,R., Welch,A., Boeing,H., Del Giudice,G., Palli,D., Saieva,C. *et al.* (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.
6. Shuker,D.E. and Margison,G.P. (1997) Nitrosated glycine derivatives as a potential source of O^6-methylguanine in DNA. *Cancer Res.*, **57**, 366–369.
7. Harrison,K.L., Jukes,R., Cooper,D.P. and Shuker,D.E. (1999) Detection of concomitant formation of O^6-carboxymethyl- and O^6-methyl-2′-deoxyguanosine in DNA exposed to nitrosated glycine derivatives using a combined immunoaffinity/HPLC method. *Chem. Res. Toxicol.*, **12**, 106–111.
8. Marnett,L.J. (2000) Oxyradicals and DNA damage. *Carcinogenesis*, **21**, 361–370.
9. Delaney,J.C. and Essigmann,J.M. (2004) Mutagenesis, genotoxicity and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **101**, 14051–14056.
10. Moriya,M. (1993) Single-stranded shuttle phagemid for mutagenesis studies in mammalian cells: 8-oxoguanine in DNA induces targeted G•C → T•A transversions in simian kidney cells. *Proc. Natl Acad. Sci. USA*, **90**, 1122–1126.
11. Delaney,J.C. and Essigmann,J.M. (2006) Assays for determining lesion bypass efficiency and mutagenicity of site-specific DNA lesions in vivo. *Methods Enzymol.*, **408**, 1–15.
12. Metzker,M.L. (2009) Sequencing technologies—the next generation. *Nat. Rev. Genet.*, **11**, 31–46.
13. Huang,W. and Marth,G. (2008) EagleView: a genome assembly viewer for next-generation sequencing technologies. *Genome Res.*, **18**, 1538–1543.
14. Mardis,E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet.*, **24**, 133–141.
15. Stratton,M.R., Campbell,P.J. and Futreal,P.A. (2009) The cancer genome. *Nature*, **458**, 719–724.
16. Laird,P.W. (2010) Principles and challenges of genome-wide DNA methylation analysis. *Nat. Rev. Genet.*, **11**, 191–203.
17. Park,P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.*, **10**, 669–680.
18. Jarosz,D.F., Beuning,P.J., Cohen,S.E. and Walker,G.C. (2007) Y-family DNA polymerases in *Escherichia coli*. *Trends Microbiol.*, **15**, 70–77.
19. Romieu,A., Gasparutto,D. and Cadet,J. (1999) Synthesis and characterization of oligonucleotides containing 5′,8-cyclopurine 2′-deoxyribonucleosides: (5′*R*)-5′,8-cyclo-2′-deoxyadenosine, (5′*S*)-5′,8-cyclo-2′-deoxyguanosine, and (5′*R*)-5′,8-cyclo-2′-deoxyguanosine. *Chem. Res. Toxicol.*, **12**, 412–421.
20. Wang,J. and Wang,Y. (2010) Synthesis and characterization of oligodeoxyribonucleotides containing a site-specifically incorporated N^6-carboxymethyl-2′-deoxyadenosine or N^4-carboxymethyl-2′-deoxycytidine. *Nucleic Acids Res.*, **38**, 6774–6784.
21. Wang,J. and Wang,Y. (2009) Chemical synthesis of oligodeoxyribonucleotides containing *N*3- and O^4-carboxymethylthymidine and their formation in DNA. *Nucleic Acids Res.*, **37**, 336–345.
22. Neeley,W.L., Delaney,S., Alekseyev,Y.O., Jarosz,D.F., Delaney,J.C., Walker,G.C. and Essigmann,J.M. (2007) DNA polymerase V allows bypass of toxic guanine oxidation products *in vivo*. *J. Biol. Chem.*, **282**, 12741–12748.
23. Yuan,B., Jiang,Y. and Wang,Y. (2010) Efficient formation of the tandem thymine glycol/8-oxo-7,8-dihydroguanine lesion in isolated DNA and the mutagenic and cytotoxic properties of the tandem lesions in *Escherichia coli* cells. *Chem. Res. Toxicol.*, **23**, 11–19.
24. Yuan,B. and Wang,Y. (2008) Mutagenic and cytotoxic properties of 6-thioguanine, S^6-methylthioguanine, and guanine-S^6-sulfonic acid. *J. Biol. Chem.*, **283**, 23665–23670.
25. Yuan,B., Cao,H., Jiang,Y., Hong,H. and Wang,Y. (2008) Efficient and accurate bypass of N^2-(1-carboxyethyl)-2′-deoxyguanosine by DinB DNA polymerase *in vitro* and *in vivo*. *Proc. Natl Acad. Sci. USA*, **105**, 8679–8684.
26. Dizdaroglu,M., Dirksen,M.L., Jiang,H.X. and Robbins,J.H. (1987) Ionizing-radiation-induced damage in the DNA of cultured human cells. Identification of 8,5′-cyclo-2-deoxyguanosine. *Biochem. J.*, **241**, 929–932.
27. Kirkali,G., de Souza-Pinto,N.C., Jaruga,P., Bohr,V.A. and Dizdaroglu,M. (2009) Accumulation of (5′*S*)-8,5′-cyclo-2′-deoxyadenosine in organs of Cockayne syndrome complementation group B gene knockout mice. *DNA Repair*, **8**, 274–278.
28. Wang,J., Yuan,B., Guerrero,C., Bahde,R., Gupta,S. and Wang,Y. (2011) Quantification of oxidative DNA lesions in tissues of long-evans cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal. Chem.*, **83**, 2201–2209.
29. Belmadoui,N., Boussicault,F., Guerra,M., Ravanat,J.L., Chatgilialoglu,C. and Cadet,J. (2010) Radiation-induced formation of purine 5′,8-cyclonucleosides in isolated and cellular DNA: high stereospecificity and modulating effect of oxygen. *Org. Biomol. Chem.*, **8**, 3211–3219.
30. Sutton,M.D. and Walker,G.C. (2001) Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination. *Proc. Natl Acad. Sci. USA*, **98**, 8342–8349.
31. Maga,G., Villani,G., Crespan,E., Wimmer,U., Ferrari,E., Bertocci,B. and Hubscher,U. (2007) 8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins. *Nature*, **447**, 606–608.
32. Gottschalg,E., Scott,G.B., Burns,P.A. and Shuker,D.E. (2007) Potassium diazoacetate-induced p53 mutations in vitro in relation to formation of O^6-carboxymethyl- and O^6-methyl-2′-deoxyguanosine DNA adducts: relevance for gastrointestinal cancer. *Carcinogenesis*, **28**, 356–362.
33. Kuraoka,I., Bender,C., Romieu,A., Cadet,J., Wood,R.D. and Lindahl,T. (2000) Removal of oxygen free-radical-induced 5′,8-purine cyclodeoxynucleosides from DNA by the nucleotide excision-repair pathway in human cells. *Proc. Natl Acad. Sci. USA*, **97**, 3832–3837.
34. Brooks,P.J., Wise,D.S., Berry,D.A., Kosmoski,J.V., Smerdon,M.J., Somers,R.L., Mackie,H., Spoonde,A.Y., Ackerman,E.J., Coleman,K. *et al.* (2000) The oxidative DNA lesion 8,5′-(*S*)-cyclo-2′-deoxyadenosine is repaired by the nucleotide excision repair pathway and blocks gene expression in mammalian cells. *J. Biol. Chem.*, **275**, 22355–22362.
35. Dizdaroglu,M. (1986) Free-radical-induced formation of an 8,5′-cyclo-2′-deoxyguanosine moiety in deoxyribonucleic acid. *Biochem. J.*, **238**, 247–254.
36. Yuan,B., O'Connor,T.R. and Wang,Y. (2010) 6-Thioguanine and S^6-methylthioguanine are mutagenic in human cells. *ACS Chem. Biol.*, **5**, 1021–1027.
37. Quail,M.A., Kozarewa,I., Smith,F., Scally,A., Stephens,P.J., Durbin,R., Swerdlow,H. and Turner,D.J. (2008) A large genome center's improvements to the Illumina sequencing system. *Nat. Methods.*, **5**, 1005–1010.