Effect of Sequence Context on O^6-Methylguanine Repair and Replication in Vivo[†]

James C. Delaney and John M. Essigmann*

Department of Chemistry and Division of Bioengineering and Environmental Health, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received July 16, 2001; Revised Manuscript Received October 4, 2001

ABSTRACT: Understanding the origins of mutational hotspots is complicated by the intertwining of several variables. The selective formation, repair, and replication of a DNA lesion, such as O^6-methylguanine (m^6G), can, in principle, be influenced by the surrounding nucleotide environment. A nearest-neighbor analysis was used to address the contribution of sequence context on m^6G repair by the *Escherichia coli* methyltransferases Ada or Ogt, and on DNA polymerase infidelity in vivo. Sixteen M13 viral genomes with m^6G flanked by all permutations of G, A, T, and C were constructed and individually transformed into repair-deficient and repair-proficient isogenic cell strains. The 16 genomes were introduced in duplicate into 5 different cellular backgrounds for a total of 160 independent experiments, for which mutations were scored using a recently developed assay. The Ada methyltransferase demonstrated strong 5′ and 3′ sequence-specific repair of m^6G in vivo. The Ada 5′ preference decreased in the general order: GXN > CXN > TXN > AXN (X = m^6G, N = any base), while the Ada 3′ preference decreased in the order: NX(T/C) > NX(G/A), with mutation frequencies (MFs) ranging from 35% to 90%. The Ogt methyltransferase provided MFs ranging from 10% to 25%. As was demonstrated by Ada, the Ogt methyltransferase repaired m^6G poorly in an AXN context. When both methyltransferases were removed, the MF was nearly 100% for all sequence contexts, consistent with the view that the replicative DNA polymerase places T opposite m^6G during replication irrespective of the local sequence environment.

The treatment of cells with DNA damaging agents results in the formation of a host of different DNA lesions, a subset of which can give rise to mutations. If one were to plot the frequency and type of mutation as a function of the position along a gene, a distribution is generated that is commonly referred to as a mutational spectrum. Early work by Benzer, Miller and later many others revealed that there is a distinct mutational spectrum for nearly all DNA damaging agents (*1*, *2*). That is, the mutational spectrum for an oxidizing agent is distinct from that of ultraviolet light, and that of a low molecular weight alkylating agent can be distinguished from that of aflatoxin (*3*−*6*). Mutational spectra are usually punctuated by sites at which mutagenic events occur more frequently than expected. These are mutational hotspots, and recent work has begun to address the mechanisms underlying the existence of such sites. The formation of hotspots can arise from three nonmutually exclusive natural scenarios, namely, (i) the higher than average reactivity of a particular base toward a chemical (or radiation), (ii) decreased repair of a damaged base, and (iii) increased insertion of incorrect deoxynucleotides during lesion replication. It is reasonable to speculate that these events are modulated by the local sequence environment surrounding the base to be modified, or the DNA lesion to be repaired or replicated. Many studies have recently addressed the biases in the formation of lesions along DNA sequences in vitro and in living cells (*7*, *8*). In the present work, we address biases in lesion processing after the lesion-formation event. Specifically, this work addresses how hotspots may arise as a consequence of preferential mutagenic replication or the preferential failure to repair a lesion in all possible three base contexts. To accomplish this goal, a panel of 16 viral genomes was constructed in which the promutagenic base, m^6G[1] (O^6-methylguanine), was present in each possible flanking base sequence. The 16 site-specifically modified genomes were allowed to replicate in cells that had different repair backgrounds in order to gain insight into the mechanistic basis for the differential mutagenic behavior of this base in cells.

m^6G was chosen for analysis because it forms frequently from exogenous (*9*) and endogenous (*10*−*12*) sources, and a correlation has been made between the accumulation of m^6G and organ-specific tumorgenicity in rats by DNA alkylating agents (*13*, *14*). If m^6G evades repair, it elicits G → A transition mutations (*15*), and in vitro primer extension studies show that the misincorporation frequency of deoxynucleotide triphosphates opposite m^6G, or polymerase extension past m^6G, can vary in different sequences (*16*, *17*). *Escherichia coli* has two main defense systems against alkylation at the O^6 position of guanine and the O^4 position of thymine. The inducible Ada and constitutive Ogt methyltransferases act to reverse the base damage directly by transferring the alkyl group to an active site cysteine residue

[†] This work was supported by Grants CA09112 and CA80024 from the National Institutes of Health.

* Correspondence should be addressed to this author at 77 Massachusetts Ave., Room 56-669, Cambridge, MA 02139. Phone: (617) 253-6227, FAX: (617) 253-5445, E-mail: jessig@mit.edu.

[1] Abbreviations: m^6G, O^6-methylguanine; MF, mutation frequency; REAP, Restriction Endonuclease And Postlabeling determination of mutation frequency; NER, nucleotide excision repair.

10.1021/bi015578f CCC: $20.00
Published on Web 11/13/2001

(*18*, *19*). Studies of m^6G repair performed in a few sequence contexts have also shown context-dependent repair of m^6G by both mammalian cell extracts and purified Ada (*20*, *21*). To date, there have been no studies that assess the effect of repair or replication of m^6G, or for that matter other DNA lesions placed at a specific location within a genome, in all nearest-neighbor contexts in vivo. Recently, a method was introduced whereby one can rapidly and with high yield prepare genomes with adducts at specific sites, and this genome construction procedure was coupled with a high-throughput analytical technique to detect mutations quantitatively (*22*). Here we have used this system to define the context-dependent processing of m^6G in vivo in cells having five different genetic backgrounds.

MATERIALS AND METHODS

Materials. Phosphoramidites were from Glen Research. *Eco*RI, *Bbs*I, *Hae*III, T4 polynucleotide kinase, and T4 DNA ligase were from New England Biolabs. ATP, exonuclease III, Sephracryl S-400, and Sephadex G-25 (Fine) were from Amersham Pharmacia Biotech. Shrimp alkaline phosphatase was from Roche Diagnostics Corp. Lysozyme and RNase A were from Sigma (product numbers L6876 and R5000, respectively). [γ-^{32}P]ATP (6000 Ci/mmol) was from Perkin-Elmer NEN Life Science Products. Snake venom phosphodiesterase I (40 units/mg) was from ICN Biomedical. PEI-TLC plates (20 × 20 cm) were from J. T. Baker.

Construction of Genomes Built with m^6G at Specific Sites. Oligonucleotides of sequence 5′ GAAGACCNm^6GNGCGTCC 3′ (N represents all permutations of G, A, T, and C) were individually made, deprotected, gel-purified, characterized, and incorporated into the genome as described (*22*). Briefly, 10 pmol of oligonucleotide containing m^6G was 5′-phosphorylated in a 30 μL volume with T4 polynucleotide kinase (15 units). Approximately 70 μL of a solution containing equimolar amounts (10 pmol) of the scaffold oligonucleotides 5′GGTCTTCCACTGAATCATGGTCATAGC3′,5′AAAAC-GACGGCCAGTGAATTGGACGC 3′, and single-stranded M13mp7(L2) that had been linearized at the *Eco*RI site was added to the tube containing the phosphorylated oligonucleotide. After heating the DNA components at equimolar concentration to 80 °C for 5 min and annealing, the solution was made 10 mM DTT, 1 mM ATP, and 5.4 units/μL T4 DNA ligase, after which it was incubated at 16 °C for 2 h. Using this technique, 90% of the insert was ligated to both the 5′ and 3′ ends of the vector as judged by PAGE (*23*). An 80% ligation efficiency was obtained if the phosphorylated insert were added to a preannealed linearized vector and scaffold set, and then ligated at 16 °C for 30 min. Scaffolds were degraded to approximately 4-mers by incubation with exonuclease III (1 unit/μL) at 37 °C for 2 h. The constructed genomes were desalted using Centricon YM-100 filter units and 2 × 2 mL rinses with 10 mM Tris-HCl, 1 mM EDTA (pH 8.0).

Transformation of E. coli with m^6G-Modified Genomes. Isogenic repair-deficient *E. coli* were made by P1 *vir* transduction (*24*). The drug-resistant alleles *uvr*B5::Tn*10*Tet^r (*25*), *ogt*-1::Kan^r (*26*), and Δ*ada*-25::Cam^r (*27*) had been previously used to construct the triple mutant *E. coli* strain CJM2 (*uvrB*, *ogt*, *ada*) (*28*) by P1 *vir* transduction into the wild-type strain FC215 (*29*). We used these alleles from CJM2 to construct the additional isogenic repair-deficient *E. coli* strains JDUvrB+ (*ogt*, *ada*), JDAda+ (*uvrB*, *ogt*), and JDOgt+ (*uvrB*, *ada*) by sequential P1 *vir* transduction using the appropriate alleles. Since tetracycline resistance from the *uvr*B5::Tn*10*Tet^r allele does not always result in the elimination of nucleotide excision repair (NER), wild-type cells were first transduced with *uvr*B5::Tn*10*Tet^r, after which they were characterized by their sensitivity to UV light. Electrocompetent cells were transformed with the individually constructed genomes as described (*22*). Approximately 10^6 electroporated cells were independently transformed with the M13 construct, yielding 10^8-10^{10} plaque forming units/mL after growth in liquid culture. Two independently constructed genomes bearing the 5′ Tm^{6}GG 3′ sequence that were electroporated approximately 2 h apart gave nearly identical mutation frequencies (MFs), suggesting negligible change in cellular repair status during serial electroporations.

Analysis of Mutations. The MFs for all electroporated genomes passaged through all cell strains were obtained using the REAP assay (Restriction Endonuclease And Postlabeling determination of mutation frequency) as described (*22*) and outlined (Figure 1). Briefly, progeny phage was used to infect SCS110 (JM110, *end A1*) cells. The isolated double-stranded M13 DNA was treated in 50 μL of buffer consisting of 10 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, and 1 mM DTT (pH 7.9) (1 × buffer 2), to which was added 10 units of *Bbs*I, 3 units of shrimp alkaline phosphatase, and 30 μg of RNase A at 37 °C for 4 h, followed by heating at 80 °C for 5 min. The linearized dephosphorylated vector was purified away from small contaminating oligonucleotides by size exclusion chromatography through Sephracryl S-400 resin. Fractions containing vector DNA were pooled and ethanol precipitated. Approximately 0.4 pmol of vector was used for ^{32}P-labeling. DNA was suspended in 7.5 μL of 1 × buffer 2 that was supplemented with 10 mM DTT, 10 pmol of ATP at a specific activity of 1000 Ci/mmol ([γ-^{32}P]ATP mixed with nonradioactive ATP), and 5 units of T4 polynucleotide kinase. This ATP addition provided a greater than 10-fold molar excess of ATP to 5′-phosphorylatable ends. Samples were incubated at 37 °C for 1 h followed by 65 °C for 20 min. An oligonucleotide fragment containing the 5′ lesion site for base composition analysis was generated by adding 2.5 μL of 1 × buffer 2 containing 30 units of *Hae*III (from a stock concentration of 50 units/μL) to the labeled mixture. After this solution was incubated at 37 °C for 2 h, 10 μL of denaturing dye was added, and the oligonucleotides were resolved using 20% denaturing polyacrylamide gels. After migration through ~4 cm of gel, the 18-mers were excised, eluted by crushing and soaking in 200 μL of water for ~2 h, and desalted with Sephadex G-25 Fine resin. To the lyophilized ^{32}P-end-labeled oligonucleotide was added a 5 μL solution of 100 mM Tris-HCl buffer (pH 8.8), 15 mM $MgCl_2$, and 5 × 10^{-2} units of snake venom phosphodiesterase I. After incubation of the solution at 37 °C for 1 h, ~0.5 μL of sample was spotted onto a 20 × 20 cm PEI-TLC plate, which was developed with 200 mL of saturated $(NH_4)_2SO_4$, dried, and exposed to a storage phosphor screen. MFs were calculated using ImageQuant 5.0 software (Molecular Dynamics) by drawing ellipses around the A (5′ dAMP) and G (5′ dGMP) spots and reporting the percentage

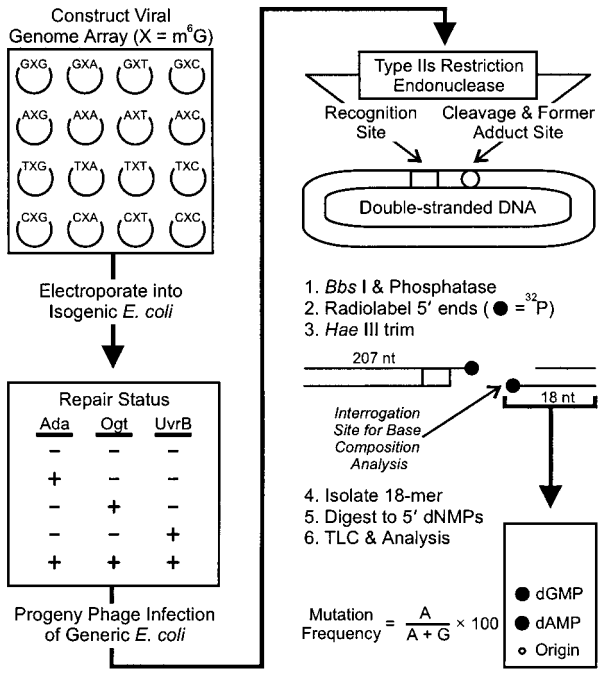

FIGURE 1: Overview of approach for determining the flanking base influence on m^{6}G processing in vivo. Sixteen viral genomes bearing m^{6}G in all nearest-neighbor sequence contexts were individually constructed and electroporated into isogenic *E. coli* strains with differing elements for m^{6}G repair. The adduct in each case was located at the site at which the Type IIs restriction endonuclease *Bbs*I cleaves (a Type IIs enzyme cleaves a defined number of nucleotides away from its recognition sequence). Phage progeny was used to transfect a generic strain of *E. coli* from which double-stranded DNA was harvested for mutation analysis. REAP analysis was performed by cleaving the progeny genomes with *Bbs*I and ultimately radiolabeling the vector 5′ overhanging ends. Treatment with *Hae*III generated two radioactive oligonucleotide fragments 207 nucleotides (nt) and 18 nt long. The 5′ terminal base of the 18 nt fragment corresponds to the lesion site; therefore, an analysis of the fractional base composition at this site will generate the MF and specificity. These numbers are based on the entire population of progeny from a given transfection. The 18-mer was isolated and digested with snake venom phosphodiesterase to 5′ deoxynucleotide monophosphates, which were resolved using one-dimensional TLC and quantified by PhosphorImager analysis.

value of 100 × [A/(A + G)], while using the program's local average background correction function. The similarity of MFs between independently constructed and electroporated genomes instills confidence in the reproducibility of the technique.

RESULTS

Sixteen viral genomes with m^{6}G in all nearest-neighbor contexts were made and individually transfected into five isogenic strains of *E. coli* having different repair properties for m^{6}G (Figure 1). Progeny from each transfection was analyzed for mutations by the REAP assay, which exploits the ability of a Type IIs restriction enzyme to cut duplex DNA a defined number of bases away from its recognition site. The cutting site was programmed to occur immediately 5′ to the site that had contained the adduct. The fractional base composition at the former adduct site (and, hence, the MF and specificity) was quantified by radiolabeling this site and resolving the labeled deoxynucleotide monophosphates via TLC (Figure 2).

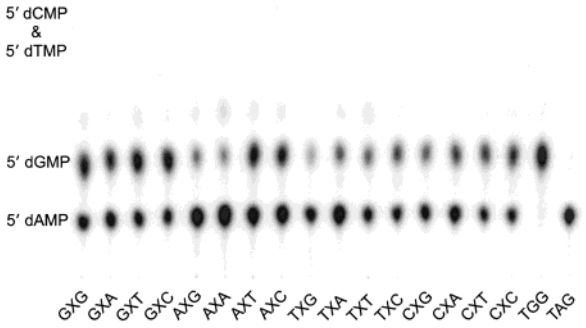

FIGURE 2: Representative TLC chromatogram used in determining the fractional base composition at a preselected genome site based on the ensemble of progeny from a site-specific mutagenesis experiment. This particular chromatogram was used to determine the MFs of m^{6}G in Ada$^+$, Ogt$^-$, UvrB$^-$ cells (Figure 4, panel A).

Effect of Sequence Context on m^{6}G Replication. A *uvrB, ogt, ada* cell lacks all known DNA repair proteins for m^{6}G. In the absence of all known repair factors, there was little effect of the sequence surrounding m^{6}G on its MF and specificity, as evidenced by the insertion of thymine opposite the adduct in all contexts (Figure 3, panels A and B). In addition, it is noteworthy that the MFs for m^{6}G in all contexts were nearly 100%, indicating that the replicative DNA polymerase paired a template m^{6}G with T in all sequence contexts. This observation provided a unique opportunity to address directly and thoroughly the impact of specific proteins on m^{6}G repair in vivo. For example, when a single repair system is expressed, an observed decrease in MF can be attributed solely to that repair system, and not to a combination of differential repair and replication.

Effect of Sequence Context on m^{6}G Repair by the NER Complex. Our main reason for using cells lacking NER in the aforementioned experiments was to eliminate the possibility of the NER complex shielding m^{6}G from repair by the Ada or Ogt methyltransferases as a function of sequence. Evidence for this shielding phenomenon exists, since the G → A MF from m^{6}G in an NER-deficient background is lower than in wild-type cells (*30*). The NER complex has also been implicated in directly repairing m^{6}G (*31*−*33*). That point notwithstanding, the MFs of cells expressing NER but lacking the Ada and Ogt methyltransferases (Figure 3, panels C and D) were strikingly similar to those found in cells lacking NER and both methyltransferases (Figure 3, panels A and B). This result is not surprising, as our lesions were in a single-stranded vector, and repair by the NER complex (requiring double-stranded DNA for excision of an oligonucleotide containing the damaged base) would occur after a round of replication. The data presented in Figure 3 (panels A and B) suggest that an m^{6}G:T pair would form, ultimately resulting in an A:T pair after NER processing and gap fill-in by a DNA polymerase, regardless of sequence context. It is conceivable that the highly miscoding nature of m^{6}G would reveal context preferences for m^{6}G repair by NER, should such preferences exist, with the use of double-stranded vectors bearing m^{6}G:C pairs. With modification, our system could address that issue.

Effect of Sequence Context on m^{6}G Repair by the Ada Methyltransferase. To address the context dependence of Ada for repair of m^{6}G, genomes containing the lesion were replicated in an Ada$^+$, Ogt$^-$, UvrB$^-$ cell. As shown in Figure

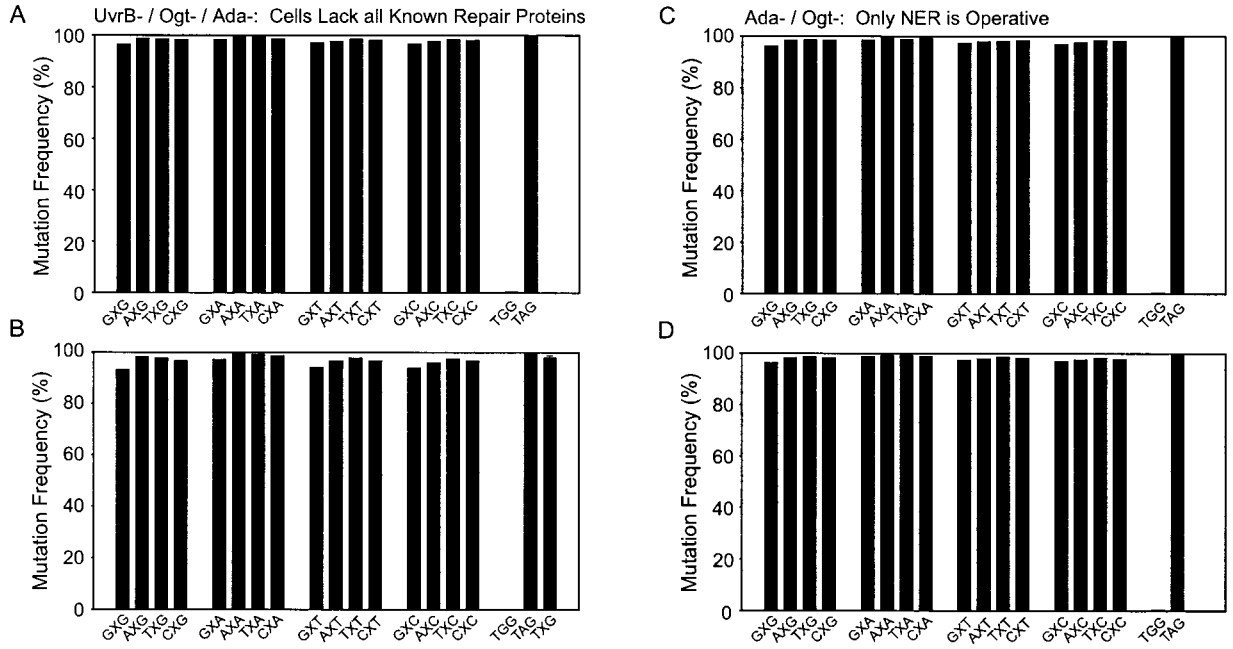

FIGURE 3: m^6G shows a nearly 100% MF in all sequence contexts, and nucleotide excision repair does not reduce G → A transition mutations from m^6G:T pairs. The MF vs nearest-neighbor sequence context plots for m^6G in the absence of both Ada and Ogt methyltransferases and of UvrB, as well as in the absence of just the Ada and Ogt methyltransferases, are presented. Sequences are read 5′ to 3′, with the m^6G lesion labeled as X. Data presented in panels A and B were obtained from two independently constructed and electroporated sets of genomes using cells lacking both Ada and Ogt methyltransferases, as well as the UvrB protein necessary for nucleotide excision repair. Panel B contains a duplicate independent construction of a control genome, 5′ TXG 3′ (far right lane), which was electroporated with an approximate 2 h time difference between the other 5′ TXG 3′ genome within the series. This control genome was analyzed in triplicate using the REAP assay with ± 2 standard deviations shown. Data presented in panels C and D were obtained from two independent analyses using the REAP assay from one constructed and electroporated set of genomes using cells lacking just the Ada and Ogt methyltransferases.

4, repair of m^6G by Ada generated a wide range of MFs which, in this case, were clearly sequence-dependent. At physiologically normal levels of the Ada protein, the MF of m^6G was suppressed by 10−65% depending upon the sequence surrounding the lesion. A consensus sequence for the bias in repair of m^6G by Ada was demonstrated for both 5′ and 3′ contexts (Figure 4). An m^6G lesion in a 5′ (G/C)m^6GN 3′ context was repaired better than in a 5′ (A/T)-m^6GN 3′ context, while the adduct in a 5′ Nm^6G(T/C) 3′ context was repaired better than in a 5′ Nm^6G(G/A) 3′ context. Thus, while the replicative polymerase does not display context-dependent variations in coding properties as it traverses m^6G, it is clear that the methyltransferase indeed does detect structural variations that lead to context-dependent repair.

Panels A and B in Figure 4 represent duplicate experiments in that two sets of genomes were individually made and individually transfected into Ada^+, Ogt^-, $UvrB^-$ cells. It is not surprising that for duplicate genome constructions and electroporations, the absolute MFs shown in Figure 4 for panel A were not identical to those for panel B. The variations observed would not be expected based on the established precision of the REAP assay (*22*), so it is concluded that they are due to differences in intracellular methyltransferase concentrations between the two batches of cells. It has been demonstrated that the concentration of uninduced Ada in *E. coli* can vary from sample to sample (*10*). Presumably, the concentration of any repair protein, including Ogt, can also vary depending on slight differences in the preparation of cells to be transformed. For that reason, we avoid not combining the data from individual cell preparations and transformations. We note, however, that the desired rank-order of MFs with respect to sequence context was maintained between independent experiments.

Effect of Sequence Context on m^6G Repair by the Ogt Methyltransferase. In a cell that had only the Ogt methyltransferase expressed, a strikingly large reduction in mutagenesis was observed (Figure 5, panels A and B). Ogt was the dominant repair element for m^6G in physiologically normal *E. coli*, since cells lacking both NER and Ada had sequence context profiles similar to those of wild-type cells (Figure 5, panels C and D). The fact that MF patterns in wild-type cells mirrored those of Ogt^+ cells (*uvrB*, *ada*) is probably owed to the fact that there are 30−40 molecules of Ogt and only 1−2 molecules of Ada per cell (*34*), and both methyltransferases react at about the same rate on m^6G (*35*).

As was exhibited by Ada, Ogt demonstrated a 2−3-fold sequence preference for m^6G repair. However, Ogt gave a somewhat altered mutational signature with respect to Ada. As shown in Figure 5, there was no apparent consensus sequence for the 3′ base; nevertheless, as was observed for Ada, the lesion in a 5′ Am^6GN 3′ context was highly refractory to repair. The least repairable context for Ogt was the same as that for Ada, in that the highest MF was observed for 5′ Am^6GA 3′ in both Ogt^- and Ada^- cells.

DISCUSSION

A mutational hotspot could be due to (i) the preferential formation of an adduct in a specific sequence context, (ii) the context-dependent inability to repair the adduct, and (iii) the context-dependent infidelity of a DNA polymerase. With

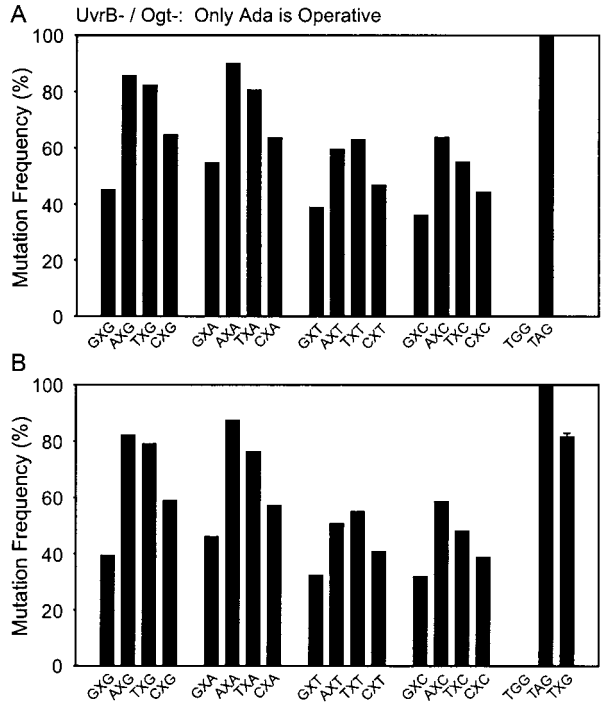

FIGURE 4: Cells expressing only the Ada methyltransferase exhibit a wide range of MFs. The MF vs nearest-neighbor sequence context plots for m^6G in cells lacking both NER and the Ogt methyltransferase are presented. Sequences are read 5′ to 3′, with the m^6G lesion labeled as X. Data presented in panels A and B were obtained from two independently constructed and electroporated sets of genomes. Panel B contains a duplicate independent construction of a control genome, 5′ TXG 3′ (far right lane), which was electroporated with an approximate 2 h time difference between the other 5′ TXG 3′ genome within the series. This control genome was analyzed in triplicate using the REAP assay with ± 2 standard deviations shown.

the exception of studies on an aromatic amide and amine (*36*−*40*) and an oxidized base (*41*), there have been few attempts to explain the molecular basis of mutational hotspots in detailed mechanistic terms. This difficulty arises because all three factors listed above could contribute in a significant way to the landscape of a mutational spectrum, and it is difficult to measure each factor experimentally. The use of a site-specifically modified genome in which an adduct can be programmed into any context makes the first point, preferential formation of a lesion, controlled because one knows exactly how much lesion is present (100%) in a given context at the time that DNA enters a cell. The use of a cell in which all known DNA repair systems are inactivated allows one to probe context-dependent polymerase infidelity, and the use of DNA repair mutants allows one to probe the influence of context on specific repair reactions in vivo. Taking the above points into consideration, a direct means to address the origin of hotspots is provided by using viral or plasmid genomes with adducts at specific sites introduced into cells of known DNA repair and replication backgrounds. In the past, the ability to perform such experiments has been constrained by technical issues. Specifically, as evidenced in the present study, a thorough analysis of context-dependent mutagenesis and repair for one lesion, performed in duplicate, required 32 genome constructions, and 160 transfections and MF analyses. The scale of this experiment demanded an efficient protocol for genome construction and mutation analysis (*22*).

The experimental scheme (Figure 1) provides for the rapid analysis of the mutational outcome from any DNA lesion. m^6G was chosen for detailed analysis because partial data on it are known. For example, it is known to induce G → A mutations (*15*), Ada, Ogt, and NER are known to play a role in its repair (*26*, *31*, *35*), and there are literature data that its mutagenic replication (*16*) and repair (*20*, *21*) are context-dependent. The high MFs we observed in methyltransferase-deficient cells (Figure 3) were comparable to that found by Pauly et al. for one context in repair-deficient *E. coli* (*42*). Our result concerning the mutagenic potential of m^6G within each possible nearest-neighbor context contrasts with in vitro primer extension studies showing that the DNA polymerase I misincorporation frequency of deoxynucleotide triphosphates opposite m^6G and extension past m^6G are dependent on the surrounding sequence (*16*, *17*).

Although experiments were performed using m^6G in single-stranded vectors, the lesion may have been repaired in a double-stranded DNA environment (generated after replication). Indeed, the Ada repair protein reacts with m^6G 1000 times faster in double-stranded rather than in single-stranded DNA (*43*). If duplex formation precedes repair, an m^6G:T pair would be the repair substrate, due to the highly miscoding nature of m^6G (Figure 3). We observed that the MF of m^6G was suppressed by 10−65% in cells expressing only the Ada methyltransferase (Figure 4) as a function of the sequence surrounding the lesion. This strong 2−3-fold context dependence on m^6G repair by Ada may be significant when the amount of *O*-alkylated products exceeds or is nearly equal to the amount of limiting methyltransferases, prior to any adaptation response (the "adaptive response" is an induced expression of Ada following exposure of the cell to an alkylating agent). The Ada methyltransferase binds to approximately 7 base pairs of DNA (*44*), and therefore it is possible, or even likely, that tracts of sequences beyond nearest-neighbors would be determinants for binding. As evidence of this fact, oligonucleotide duplexes containing m^6G in identical nearest-neighbor contexts, but with different next to nearest-neighbor bases, exhibit a 4-fold variation in the rate of repair (*21*). It should also be mentioned that the sequence context effects on m^6G mutagenesis that we observed by Ada and Ogt may have been modulated by another active DNA repair system, such as mismatch repair (*42*, *45*, *46*). For example, the sequence-dependent binding of such proteins to the lesion may shield m^6G from methyltransferase repair. In the case of mismatch repair, it is also possible that futile cycles of repair of the base opposing m^6G in certain sequences may give the methyltransferase a better chance of repairing the lesion prior to fixation of the bypass event. While these optional hypotheses cannot be ruled out for all contexts, preliminary work (Delaney and Essigmann, unpublished experiments) performed in eight contexts (5′ Nm^6GG 3′ and 5′ Nm^6GA 3′) showed similar context effects of repair by Ada in isogenic *E. coli* that lack NER and Ogt and either have or lack mutS. These experiments suggested a negligible impact of mismatch repair on our observed MFs.

Several studies on cells treated with alkylating agents suggest that the distribution of alkylation-induced mutations is modulated by the repair methyltransferases (*47*−*49*). These data are, for the most part, consistent with our observations in which site-specifically modified vectors were used in a

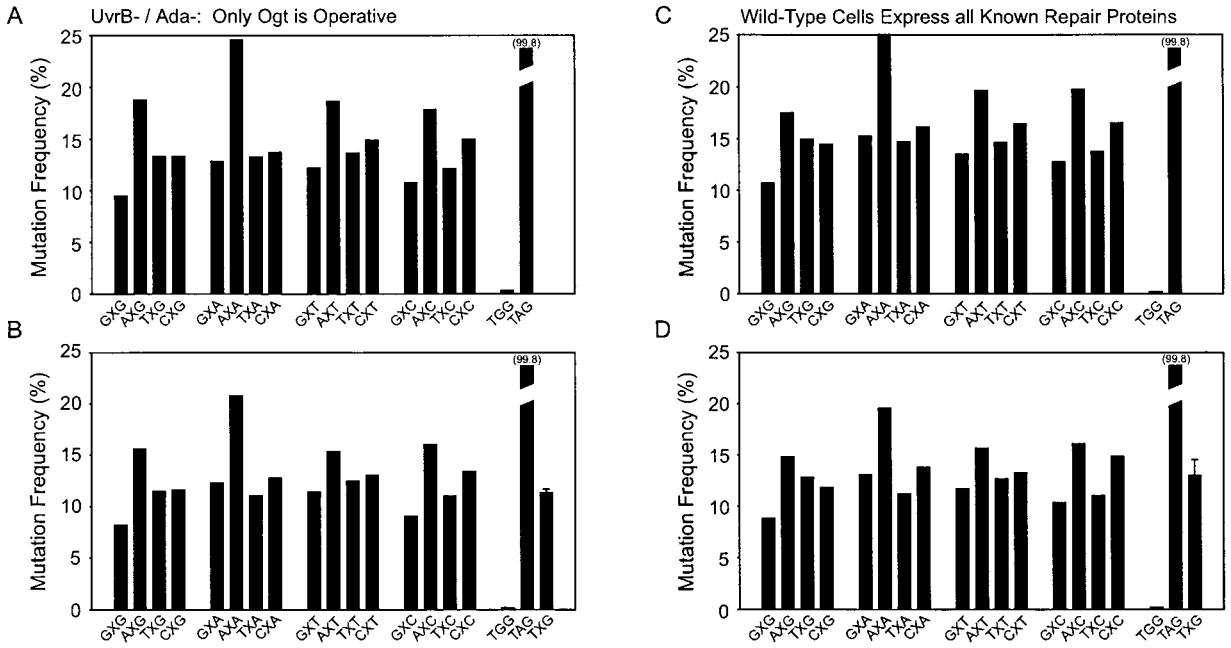

FIGURE 5: Cells expressing only the Ogt methyltransferase exhibit a wide range of MFs, and Ogt is the dominant repair element for m^6G in wild-type cells. The MF vs nearest-neighbor sequence context plots for m^6G in cells lacking both NER and the Ada methyltransferase, and in wild-type cells are presented. Sequences are read 5′ to 3′, with the m^6G lesion labeled as X. Data presented in panels A and B were obtained from two independently constructed and electroporated sets of genomes using cells lacking both NER and the Ada methyltransferase. Data presented in panels C and D were obtained from two independently constructed and electroporated sets of genomes using wild-type cells. Panels B and D contain duplicate independent constructions of control genomes, 5′ TXG 3′ (far right lanes), which were electroporated with an approximate 2 h time difference between the other 5′ TXG 3′ genomes within the series. These control genomes were analyzed in triplicate using the REAP assay with ± 2 standard deviations shown.

more controlled in vivo system. The alkylating agent 1-(2-chloroethyl)-3-cyclohexyl-1-nitrosourea induces G → A mutations ~2.8 times more frequently when a target guanine is flanked by A:T rather than G:C pairs in the presence of Ada and Ogt rather than in their absence (*47*). Likewise, in the case of ethyl methanesulfonate, guanines flanked on the 5′ side by an A:T rather than a G:C pair are 3.2 times more likely to mutate in cells expressing Ogt as compared to cells lacking this methyltransferase (*48*). A similar study using *N*-ethyl-*N*-nitrosourea also implicates preferential repair by Ogt when a target guanine is flanked by a 5′ G or 5′ C base; however, there is no statistical difference in repair preference by Ogt when the system was exposed to *N*-methyl-*N*-nitrosourea (*49*). Guanines in 5′ AGA 3′ contexts also have elevated levels of mutation, while less mutagenesis is seen for the second guanine in 5′ GG contexts (*47*, *49*), corresponding to the influence of Ada and Ogt we observed (Figures 4 and 5). Since ethylation is a poor inducer of the adaptive response (*31*), the aforementioned context effects on alkylation repair were most likely as a result of the Ogt methyltransferase.

Some studies in which cells were treated with specific alkylating agents have failed to show context rules that would have been predicted by our study. In these cases, it is possible that preferential modification of DNA (rather than context-preferred repair) caused the observed mutational asymmetry. For example, when cells are treated with *N*-methyl-*N*-nitrosourea or *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine, the majority of G → A transition mutations occur at 5′ (purine)-GN 3′ sites (*5*, *50*−*53*). It appears from these published studies that the second guanine in 5′ GG sequences exhibits a higher degree of mutation than a G flanked by a 5′ adenine. A mechanism has been proposed to explain why the 3′ G of 5′ GG sequences would be a preferred site of alkylation by alkylnitrosoureas (*54*).

The crystal structures of Ada and human O^6-alkylguanine-DNA alkyltransferase (AGT) have been solved (*55*−*57*), and a model has been proposed in which repair involves the rotation of m^6G out of the DNA helix into the proximity of a buried active site cysteine, which will serve as the recipient of the m^6G methyl group in the repair reaction (*57*, *58*). This process is commonly known as "base flipping" and was first discovered as the mechanism by which the *Hha*I cytosine-5 methyltransferase gains access to its substrate (*59*). It would be useful if structural and biophysical data could provide a framework within which one could explain the 3-fold differences in repair observed in our work for m^6G repair in vivo. Unfortunately, such biophysical studies have not been performed with m^6G in all nearest-neighbor contexts, so we can only speculate on the reasons for the bias in the repair of m^6G by Ada and Ogt. The two most plausible (and nonmutually exclusive) explanations for the repair bias are a sequence-dependent recognition of the alkylated base and a sequence-dependent initiation of base flipping. We noticed that Ada more poorly repaired m^6G when the modified base had a 3′ flanking purine residue (Figure 4). It is possible that the bias of Ada against repair of a 5′ Nm^6G(G/A) 3′ site may have been due to better stacking between m^6G and its 3′ neighboring base. However, this explanation alone is not sufficient, since Ada demonstrated a 5′ bias against repair of the mixed purine/pyrimidine 5′ (A/T)m^6GN 3′ sequence (Figure 4). Since homology models of repair alkyltransferases indicate that an arginine enters the major groove and may assist in extruding m^6G out of the helix and into the buried active site (*57*, *58*), the lesion most likely passes over its 3′ base as m^6G is swung out of the helix from the minor groove.

The transient increase in base stacking interactions between m^6G and the 3′ purine may create a thermodynamic impediment to the base flipping process, in agreement with our observations of the 3′ base influence of m^6G repair by Ada. Interesting work by the Singer group has proposed that sites of low melting temperature would be more poorly repaired by enzymes that use the "base flipping" mechanism than sites of relatively high melting temperature (*60*, *61*). We have not determined the thermodynamic stabilities of m^6G in each of our 16 contexts as yet, so it is not possible to indicate whether our observed mutational preferences would correlate with the hypothesis based upon the work of Singer et al. Data by Allawi and SantaLucia caution that the melting temperatures of various sequence contexts are difficult to predict, since a G:C pair on one side of a C:T mismatch can actually decrease the stability of the duplex, even though G:C pairs have one extra hydrogen bond (*62*).

Interestingly, the MF pattern exhibited by Ogt (Figure 5) would match that of Ada (Figure 4) had m^6G in a 5′ Tm^6-GN 3′ sequence not been as well repaired by Ogt. Although there is homology between the Ada and Ogt proteins, the amino acid N-terminal to the conserved arginine implicated in actively extruding m^6G is valine for Ada and alanine for Ogt (*63*, *64*). This subtle change could be just one of many causes for the differences we observed between the DNA sequence context preferences of Ada and Ogt for m^6G in vivo.

Finally, most studies done in the past have focused on finding the MF and specificity of a particular DNA adduct in one convenient sequence context. While those studies have been important for providing the qualitative aspects of mutagenesis by single DNA adducts, they have not contributed toward defining the quantitative features of mutational landscapes. The work presented here is an attempt to define systematically the pattern of mutagenesis by m^6G in DNAs replicated in cells of different DNA repair capacities. From such site-specific sequence context studies, empirical data are being generated that may tease out the important variables involved in lesion processing as a function of its sequence environment, which may ultimately explain in molecular terms the observed mutational spectra from exogenous and/or endogenous DNA damaging agents.

REFERENCES

1. Benzer, S. (1961) *Proc. Natl. Acad. Sci. U.S.A. 47*, 403−415.
2. Coulondre, C., and Miller, J. H. (1977) *J. Mol. Biol. 117*, 577−606.
3. Sargentini, N. J., and Smith, K. C. (1994) *Mutat. Res. 309*, 147−163.
4. Miller, J. H. (1985) *J. Mol. Biol. 182*, 45−65.
5. Richardson, K. K., Richardson, F. C., Crosby, R. M., Swenberg, J. A., and Skopek, T. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 344−348.
6. Cariello, N. F., Cui, L., and Skopek, T. R. (1994) *Cancer Res. 54*, 4436−4441.
7. Denissenko, M. F., Cahill, J., Koudriakova, T. B., Gerber, N., and Pfeifer, G. P. (1999) *Mutat. Res. 425*, 205−211.
8. Cloutier, J. F., Castonguay, A., O'Connor, T. R., and Drouin, R. (2001) *J. Mol. Biol. 306*, 169−188.
9. Loveless, A. (1969) *Nature 223*, 206−207.
10. Taverna, P., and Sedgwick, B. (1996) *J. Bacteriol. 178*, 5105−5111.
11. Shuker, D. E., and Margison, G. P. (1997) *Cancer Res. 57*, 366−369.
12. Sedgwick, B. (1997) *Carcinogenesis 18*, 1561−1567.
13. Margison, G. P., and Kleihues, P. (1975) *Biochem. J. 148*, 521−525.
14. Wiestler, O., von Deimling, A., Kobori, O., and Kleihues, P. (1983) *Carcinogenesis 4*, 879−883.
15. Loechler, E. L., Green, C. L., and Essigmann, J. M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6271−6275.
16. Singer, B., Chavez, F., Goodman, M. F., Essigmann, J. M., and Dosanjh, M. K. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 8271−8274.
17. Dosanjh, M. K., Galeros, G., Goodman, M. F., and Singer, B. (1991) *Biochemistry 30*, 11595−11599.
18. Samson, L. (1992) *Mol. Microbiol. 6*, 825−831.
19. Pegg, A. E. (2000) *Mutat. Res. 462*, 83−100.
20. Dolan, M. E., Oplinger, M., and Pegg, A. E. (1988) *Carcinogenesis 9*, 2139−2143.
21. Georgiadis, P., Smith, C. A., and Swann, P. F. (1991) *Cancer Res. 51*, 5843−5850.
22. Delaney, J. C., and Essigmann, J. M. (1999) *Chem. Biol. 6*, 743−753.
23. Bailey, E. A., Iyer, R. S., Harris, T. M., and Essigmann, J. M. (1996) *Nucleic Acids Res. 24*, 2821−2828.
24. Miller, J. H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. Backendorf, C., Spaink, H., Barbeiro, A. P., and van de Putte, P. (1986) *Nucleic Acids Res. 14*, 2877−2890.
26. Rebeck, G. W., and Samson, L. (1991) *J. Bacteriol. 173*, 2068−2076.
27. Shevell, D. E., Abou-Zamzam, A. M., Demple, B., and Walker, G. C. (1988) *J. Bacteriol. 170*, 3294−3296.
28. Samson, L., Han, S., Marquis, J. C., and Rasmussen, L. J. (1997) *Carcinogenesis 18*, 919−924.
29. Mackay, W. J., Han, S., and Samson, L. D. (1994) *J. Bacteriol. 176*, 3224−3230.
30. Chambers, R. W., Sledziewska-Gojska, E., Hirani-Hojatti, S., and Borowy-Borowski, H. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7173−7177.
31. Samson, L., Thomale, J., and Rajewsky, M. F. (1988) *EMBO J. 7*, 2261−2267.
32. Rossi, S. C., Conrad, M., Voigt, J. M., and Topal, M. D. (1989) *Carcinogenesis 10*, 373−377.
33. Voigt, J. M., Van Houten, B., Sancar, A., and Topal, M. D. (1989) *J. Biol. Chem. 264*, 5172−5176.
34. Rebeck, G. W., Smith, C. M., Goad, D. L., and Samson, L. (1989) *J. Bacteriol. 171*, 4563−4568.
35. Wilkinson, M. C., Potter, P. M., Cawkwell, L., Georgiadis, P., Patel, D., Swann, P. F., and Margison, G. P. (1989) *Nucleic Acids Res. 17*, 8475−8484.
36. Koffel-Schwartz, N., and Fuchs, R. P. P. (1995) *J. Mol. Biol. 252*, 507−513.
37. Napolitano, R. L., Lambert, I. B., and Fuchs, R. P. (1994) *Biochemistry 33*, 1311−1315.
38. Belguise-Valladier, P., and Fuchs, R. P. (1995) *J. Mol. Biol. 249*, 903−913.
39. Mu, D., Bertrand-Burggraf, E., Huang, J. C., Fuchs, R. P., Sancar, A., and Fuchs, B. P. (1994) *Nucleic Acids Res. 22*, 4869−4871.
40. Shibutani, S., Suzuki, N., Tan, X., Johnson, F., and Grollman, A. P. (2001) *Biochemistry 40*, 3717−3722.
41. Hatahet, Z., Zhou, M., Reha-Krantz, L. J., Morrical, S. W., and Wallace, S. S. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 8556−8561.
42. Pauly, G. T., Hughes, S. H., and Moschel, R. C. (1994) *Biochemistry 33*, 9169−9177.
43. Lindahl, T., Demple, B., and Robins, P. (1982) *EMBO J. 1*, 1359−1363.
44. Takahashi, M., Sakumi, K., and Sekiguchi, M. (1990) *Biochemistry 29*, 3431−3436.
45. Rasmussen, L. J., and Samson, L. (1996) *Carcinogenesis 17*, 2085−2088.
46. Brown, J., Brown, T., and Fox, K. R. (2001) *Biochem. J. 354*, 627−633.

47. Jurado, J., Ferrezuelo, F., and Pueyo, C. (1995) *Mol. Carcinog. 14*, 233−239.
48. Vidal, A., Abril, N., and Pueyo, C. (1995) *Carcinogenesis 16*, 817−821.
49. Vidal, A., Abril, N., and Pueyo, C. (1997) *Environ. Mol. Mutagen. 29*, 180−188.
50. Richardson, K. K., Crosby, R. M., Richardson, F. C., and Skopek, T. R. (1987) *Mol. Gen. Genet. 209*, 526−532.
51. Burns, P. A., Gordon, A. J., and Glickman, B. W. (1987) *J. Mol. Biol. 194*, 385−390.
52. Glickman, B. W., Horsfall, M. J., Gordon, A. J., and Burns, P. A. (1987) *Environ. Health Perspect. 76*, 29−32.
53. Gordon, A. J., Burns, P. A., and Glickman, B. W. (1990) *Mutat. Res. 233*, 95−103.
54. Buckley, N. (1987) *J. Am. Chem. Soc. 109*, 7918−7920.
55. Moore, M. H., Gulbis, J. M., Dodson, E. J., Demple, B., and Moody, P. C. (1994) *EMBO J. 13*, 1495−1501.
56. Wibley, J. E., Pegg, A. E., and Moody, P. C. (2000) *Nucleic Acids Res. 28*, 393−401.
57. Daniels, D. S., Mol, C. D., Arvai, A. S., Kanugula, S., Pegg, A. E., and Tainer, J. A. (2000) *EMBO J. 19*, 1719−1730.
58. Verdemato, P. E., Brannigan, J. A., Damblon, C., Zuccotto, F., Moody, P. C., and Lian, L. Y. (2000) *Nucleic Acids Res. 28*, 3710−3718.
59. Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994) *Cell 76*, 357−369.
60. Hang, B., Sági, J., and Singer, B. (1998) *J. Biol. Chem. 273*, 33406−33413.
61. Sági, J., Hang, B., and Singer, B. (1999) *Chem. Res. Toxicol. 12*, 917−923.
62. Allawi, H. T., and SantaLucia, J. (1998) *Nucleic Acids Res. 26*, 2694−2701.
63. Nakabeppu, Y., Kondo, H., Kawabata, S., Iwanaga, S., and Sekiguchi, M. (1985) *J. Biol. Chem. 260*, 7281−7288.
64. Potter, P. M., Wilkinson, M. C., Fitton, J., Carr, F. J., Brennand, J., Cooper, D. P., and Margison, G. P. (1987) *Nucleic Acids Res. 15*, 9177−9193.

BI015578F