# The Influence of Local DNA Sequence and DNA Repair Background on the Mutational Specificity of 1-Nitroso-8-Nitropyrene in *Escherichia coli*: Inferences for Mutagenic Mechanisms

**Iain B. Lambert,*[,1] Alasdair J. E. Gordon,[†,2] Barry W. Glickman[†,3] and Dennis R. McCalla***

**Department of Biochemistry, McMaster University, Hamilton, Ontario, Canada L8N 3Z5, and [†]Department of Biology, York University, Toronto, Ontario, Canada M3J 1P3*

Manuscript received April 23, 1992
Accepted for publication August 29, 1992

ABSTRACT

We have examined the mutational specificity of 1-nitroso-8-nitropyrene (1,8-NONP), an activated metabolite of the carcinogen 1,8-dinitropyrene, in the *lacI* gene of *Escherichia coli* strains which differ with respect to nucleotide excision repair (±Δ*uvrB*) and MucA/B-mediated error-prone translesion synthesis (±pKM101). Several different classes of mutation were recovered, of which frameshifts, base substitutions, and deletions were clearly induced by 1,8-NONP treatment. The high proportion of point mutations (>92%) which occurred at G·C sites correlates with the percentage of 1,8-NONP-DNA adducts which occur at the C(8) position of guanine. The most prominent frameshift mutations were −(G·C) events, which were induced by 1,8-NONP treatment in all strains, occurred preferentially in runs of guanine residues, and whose frequency increased markedly with the length of the reiterated sequence. Of the base substitution mutations G·C → T·A transversions were induced to the greatest extent by 1,8-NONP. The distribution of the G·C → T·A transversions was not influenced by the nature of flanking bases, nor was there a strand preference for these events. The presence of plasmid pKM101 specifically increased the frequency of G·C → T·A transversions by a factor of 30–60. In contrast, the −(G·C) frameshift mutation frequency was increased only 2–4-fold in strains harboring pKM101 as compared to strains lacking this plasmid. There was, however, a marked influence of pKM101 on the strand specificity of frameshift mutation; a preference was observed for −G events on the transcribed strand. The ability of the bacteria to carry out nucleotide excision repair had a strong effect on the frequency of all classes of mutation but did not significantly influence either the overall distribution of mutational classes or the strand specificity of G·C → T·A transversions and −(G·C) frameshifts. Deletion mutations were induced in the Δ*uvr*, pKM101 strain. The endpoints of the majority of the deletion mutations were G·C rich and contained regions of considerable homology. The specificity of 1,8-NONP-induced mutation suggests that DNA containing 1,8-NONP adducts can be processed through different mutational pathways depending on the DNA sequence context of the adduct and the DNA repair background of the cell.

EXAMINATION of mutational spectra in forward mutation targets has shown that mutagens exhibit marked specificity both with respect to the type of mutation induced and the site of mutation. This indicates that mutagenesis is not a random process, but rather that certain types of DNA adducts within particular DNA sequences (mutational hotspots) are misrepaired or misreplicated by specific mechanisms (EISENSTADT 1987). Identification of the mutational hotspots, the types of mutations induced and their relative proportions through the accurate determination of DNA sequence alterations represents a powerful approach to the study of mutational mechanisms. Moreover, the differences in the mutational spectra associated with systematic changes in the cellular genotype provide information with respect to the roles of specific gene products in error avoidance and mutation fixation.

The mutagenic and carcinogenic compound 1,8-dinitropyrene (1,8-DNP) is metabolically activated by reduction of a single nitro group to a species capable of binding covalently to DNA. The partially reduced derivative 1-nitroso-8-nitropyrene (1,8-NONP) is an activated intermediate. The major DNA adduct (>95% of the total) formed by 1,8-DNP or 1,8-NONP in bacteria (ANDREWS *et al.* 1986; LAMBERT *et al.*, 1991), cultured rabbit tracheal epithelial cells (NORMAN *et al.* 1989a), and female CD rats (NORMAN *et al.* 1990) is the guanine C(8) adduct *N*-(2′-deoxyguanosin-8-yl)-1-amino-8-nitropyrene [dG-C(8)-ANP] (Figure 1). Two uncharacterized minor adducts (<5% of the total) have also been detected (NORMAN *et al.*

[1] Present address: Department of Biology, Carleton University, Ottawa, Canada K1S 5B6.
[2] Present address: Institut Curie, Section de Biologie, 26 rue d'Ulm, 75231 Paris Cedex 05, France.
[3] Present address: University of Victoria, Centre for Environmental Health, RR2, 9865 West Saanich Road, Sidney, British Columbia, Canada V8L 3S1.

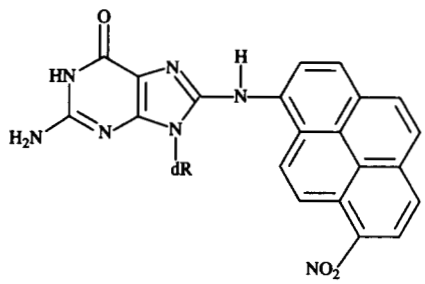


FIGURE 1.—Structure of 1-*N*-(2′-deoxyguanosin-8-yl)-amino-8-nitropyrene, the major DNA adduct formed by 1,8-NONP.

1989a; LAMBERT *et al.* 1991). In mutation assays 1,8-DNP is extremely active at the *Salmonella typhimurium* frameshift loci *hisC3076*, *hisD3052* and *hisD6580*. Induced reversion of the base substitution loci *hisG46* in *S. typhimurium* and *trpA* in *Escherichia coli* has also been observed, but only in strains which harbor plasmid pKM101 (ROSENKRANZ and MERMELSTEIN 1983; TOKIWA and OHNISHI 1986). DNA sequencing of 1,8-NONP-induced forward mutations in the *lacI* gene of nucleotide excision repair deficient (*ΔuvrB*) *E. coli* showed that frameshifts are the predominant mutation (LAMBERT *et al.* 1991).

In the present study we provide detailed mutational spectra of 1,8-NONP in the *lacI* gene of *E. coli* strains which differ with respect to: (a) nucleotide excision-repair capability and (b) the presence of plasmid pKM101 which promotes error-prone lesion bypass (WALKER 1984). The *lacI* genetic/M13 cloning system (MILLER 1978; SCHAAPER, DANFORTH and GLICKMAN 1985) used here allows all classes of mutation to be monitored in a well characterized target of over 1000 bp. The 652 mutations described in this paper, and 159 mutations reported previously (LAMBERT *et al.* 1991) represent a large data base of forward mutations which is likely to accurately reflect the mutational specificity of 1,8-NONP in *E. coli*.

## MATERIALS AND METHODS

**Bacteria and bacteriophage:** *LacI*$^-$ mutants were selected from *E. coli* strains NR6112 [F′*lacpro*; Δ(*lacpro*) *ara thi rfa*], EE125 [F′*lacpro*; Δ(*lacpro*) *ara thi rfa* pKM101], and CM6114 [same as EE125 except Δ(*bioFCD-uvrB-chlA*)]. Bacterial strains used for genetic characterization of the *lacI*$^-$ mutations, for recombinational transfer onto bacteriophage M13, and for detection of the recombinant phage have been described previously (SCHMEISSNER, GANEM and MILLER 1977; SCHAAPER, DANFORTH and GLICKMAN 1985), as has the bacteriophage mRS81 (M13*lacI*$^+$*Zα*$^-$; SCHAAPER, DANFORTH and GLICKMAN 1985). The F′*lac* used in these studies carries the I$^q$ and L8 promotor mutations to facilitate their use in the mutagenesis system (MILLER 1978).

**Media and chemicals:** Media were as described (MILLER 1972; COULONDRE and MILLER 1977). Phenyl-β-D-galactoside (PGal), 5-bromo-4-chloro-3-indoyl-β-D-galactoside and *O*-nitrophenyl-β-D-thiogalactoside were purchased from Research Organics Inc. (Cleveland, Ohio). The synthesis of 1,8-NONP (purity >99%) by oxidation of 1-amino-8-nitropyrene has been described previously (ANDREWS *et al.* 1986). All oligonucleotide probes and primers were purchased from the Central Facility of the Institute for Molecular Biology and Biotechnology (McMaster University, Hamilton).

**Selection and characterization of *lacI*$^-$ mutants:** Multiple cultures of NR6112 (wild-type), EE125 (pKM101) and CM6114 (*ΔuvrB*, pKM101) were grown overnight at 37° with shaking. Ampicillin (50 μg/ml) was added to EE125 and CM6114 cultures to maintain plasmid pKM101. Overnight cultures were centrifuged, and the pellet resuspended in Vogel-Bonner salt solution (VB). One-milliliter aliquots of each culture were treated with 1,8-NONP [25 nmol in 100 μl dimethyl sulfoxide (DMSO)], or DMSO alone (controls). Following incubation for 15 min at 37°, the bacteria were pelleted by centrifugation, washed, resuspended in VB and immediately placed on ice. The *lacI*$^-$ mutation frequency and cell survival were determined by plating appropriate dilutions of the cultures onto minimal plates containing PGal, and LB plates, respectively. PGal is a substrate for β-galactosidase but does not induce the enzyme. Therefore when PGal is present as the sole carbon source, only cells which constitutively express β-galactosidase (*i.e.*, which carry *lacI*$^-$ or *lacO*$^C$ mutations) form colonies. PGal plates were incubated for 2 days and LB plates for 16 hr at 37°. Mutation frequencies for both control (spontaneous) and 1,8-NONP-treated cultures were calculated as mutants per $10^6$ survivors. *LacI*$^-$ colonies were isolated from each independently treated culture. The combination of short treatment period in buffer, and direct plating onto PGal plates immediately following treatment ensured that all induced mutations were of independent origin.

Mutations were localized to one of seven segments of the *lacI* gene by deletion mapping as described by SCHMEISSNER, GANEM, and MILLER 1977. Mutations at the frameshift hotspot site, position 620–632 (FARABAUGH *et al.* 1978), were identified using an oligonucleotide/colony hybridization approach (HALLIDAY *et al.* 1990). Those mutants that did not hybridize to mutant-specific probes which were complementary to either an addition or deletion of the 4-bp repeat 5′-CTGG-3′ were subjected to further analysis. Mutant *lacI* genes were transferred from the F′ factor (*lacI*$^-$*lacZα*$^+$) to bacteriophage mRS81 (M13*lacI*$^+$*lacZα*$^-$) by *in vivo* recombination as described (SCHAAPER, DANFORTH and GLICKMAN 1985). Following plaque purification, viral DNA was purified, and sequenced using the dideoxy chain termination method of SANGER *et al.* (1980). Oligonucleotide primers complementary to specific positions of the *lacI* gene were used to sequence the regions indicated by the deletion mapping of the *lacI* mutations. These primers were 14 mers having 3′ ends at positions 148, 215, 302, 450, 604, 745, 901 and 1049 of the *lacI* gene and a 17-mer complementary to position 27–43 of the *lacZ* gene (REZNIKOFF and ABELSON 1978). The resulting DNA sequences were compared to that of the wild-type *lacI* gene (FARABAUGH 1978).

## RESULTS

**Selection and characterization of *lacI*$^-$ mutants:** Treatment of *E. coli* strain CM6114 (*ΔuvrB*, pKM101) with 25 μM 1,8-NONP produced a *lacI*$^-$ mutation frequency of $265 \times 10^{-6}$ at about 70% survival. This represents an 88-fold increase over the spontaneous mutation frequency determined in parallel experiments. Similar treatment of strain EE125 (pKM101), which is proficient in DNA excision repair capability

**TABLE 1**

**Distribution of *lacI*⁻ mutants following 1,8-NONP treatment**

| | Occurrences | | | |
|---|---|---|---|---|
| | NR6112 (wild-type) | EE125 (pKM101) | NR6113[c] (*ΔuvrB*) | CM6114 (*ΔuvrB*,pKM101) |
| Mutation frequency (1,8-NONP)[a] | 11 (4.2)[b] | 36 (12) | 105 (30) | 265 (71) |
| Mutation frequency (control)[a] | 5.4 (2.8) | 7.3 (2.3) | 3.2 (1.6) | 3.0 (1.6) |
| Mutation | | | | |
| CTGG (+)[d] | 129 | 45 | 15 | 4 |
| CTGG (−)[d] | 41 | 16 | 4 | 2 |
| Base substitutions | | | | |
| G·C → T·A | 5 | 31 | 4 | 66 |
| G·C → C·G | 1 | 1 | 1 | 2 |
| A·T → T·A | 0 | 1 | 0 | 7 |
| A·T → C·G | 0 | 2 | 0 | 1 |
| G·C → A·T | 5 | 3 | 7 | 4 |
| A·T → G·C | 0 | 1 | 0 | 0 |
| Tandem doubles | 1 | 0 | 0 | 1 |
| Frameshifts | | | | |
| −(G·C) | 29 | 18 | 110 | 114 |
| −(A·T) | 5 | 2 | 0 | 2 |
| −2 | 3 | 0 | 2 | 3 |
| +1 | 3 | 2 | 1 | 2 |
| Deletions | 28 | 10 | 7 | 15 |
| Duplications | 2 | 1 | 1 | 1 |
| Complex | 2 | 0 | 1 | 0 |
| Uncharacterized | 9 | 15 | 6 | 17 |
| Total No. mutants | 263 | 148 | 159 | 241 |

[a] Mutation frequencies are expressed as the number of *lacI*⁻ mutants per $10^8$ survivors.

[b] Values given are means from several independent determinations. The number of plates counted for determination of mutation frequencies following 1,8-NONP treatment were: NR6112 and CM6114, 36 plates each; EE125 and NR6113, 24 plates each. The number of plates counted for determination of control mutation frequencies were: NR6112 and CM6114, 12 plates each; EE125 and NR6113, 8 plates each. Standard deviations are shown in parentheses.

[c] Data from LAMBERT *et al.* (1991).

[d] Frameshift hotspot at position 620-632 (5′-CTGGCTGGCTGGC-3′).

but otherwise isogenic to CM6114 resulted in an increase of only 5-fold over the spontaneous *lacI*⁻ mutation frequency at 81% survival. In NR6112 (wild-type), 25 μM 1,8-NONP treatment gave rise to a 2-fold increase over the spontaneous mutation frequency at a survival of 82%. A total of 652 *lacI*⁻ mutants [241 from CM6114, 148 from EE125 and 263 from NR6112] were selected from PGal plates, purified, and the mutations localized in the *lacI* gene by deletion mapping.

The nature of the 652 *lacI*⁻ mutations recovered after 1,8-NONP treatment is summarized in Table 1. Oligonucleotide probe analysis was performed to identify frameshift hotspot events (±5′-CTGG-3′ at position 620–632) in the mutant collection. Of the mutations, 2.5% (4 additions and 2 deletions) from the CM6114 (*ΔuvrB*, pKM101) collection, 41% (45 additions and 16 deletions) from the EE125 (pKM101) collection, and 65% (129 additions and 41 deletions) from the NR6112 (wild-type) collection occurred at this site. Frameshift hotspot events account for the majority of spontaneous events in the *lacI* gene (FARABAUGH *et al.* 1978) with additions and deletions comprising 56% and 14%, respectively, of spontaneous mutations in a wild-type background (FARABAUGH *et al.* 1978; SCHAAPER, DANFORTH and GLICKMAN 1986; HALLIDAY and GLICKMAN 1991), and 42% and 15%, respectively, of spontaneous mutations in an isogenic strain harboring pKM101 (A. J. E. GORDON and B. W. GLICKMAN, in preparation). Since frameshift hotspot events dominate spontaneous mutation spectra and the proportion of these events decreases as the 1,8-NONP-induced mutation frequencies increase, a significant proportion of the frameshift hotspot events recovered in this study are likely to be of spontaneous origin.

Of the non-hotspot mutants (containing mutations other than ±5′-CTGG-3′ at position 620–632), 221/235 CM6114 (*ΔuvrB*, pKM101), 79/87 EE125 (pKM101) and 83/93 NR6112 (wild-type) *lacI*⁻ mutations were cloned into mRS81 and the DNA was sequenced. The remaining 32 *lacI*⁻ mutations were not recovered onto vector mRS81 despite repeated attempts. However, genetic analysis and oligonucleo-

tide hybridization patterns indicated that 15 of these mutants [6 NR6112, 5 EE125; 4 CM6114 ] had undergone large deletions (data not shown). Of the 383 cloned *lacI*$^-$ mutants 25 were not characterized. The sequencing results are described in detail in Tables 2–5.

Base substitution mutations were prominent only in strains which harbored plasmid pKM101 (Tables 1 and 2). Single base substitutions accounted for 34% (80/235) and 45% (39/87) of the non-hotspot mutations recovered in CM6114 (Δ*uvrB*, pKM101) and EE125 (pKM101), respectively. Approximately 90% of the base substitutions in each pKM101 strain occurred at G·C base pairs. G·C → T·A transversions were the predominant base substitution mutation observed with 66 such events being characterized in CM6114; 31 G·C → T·A transversions were recovered in EE125. The most frequent base substitutions observed at A·T sites were A·T → T·A transversions, of which 7 events were characterized in CM6114, and 1 event was isolated in EE125. Only small numbers of each of the other possible single base substitutions, and one tandem base substitution were observed in pKM101-containing strains. In NR6112 (wild-type), which lacks pKM101, only 12% of the non-hotspot mutations (11/93) were single base substitutions, all of which occurred at G·C sites. Approximately equal numbers of transversions and transitions were recovered in the wild-type strain.

A total of 183 frameshift mutations involving the gain or loss of 1 or 2 bp were recovered following 1,8-NONP treatment [121 CM6114 (Δ*uvrB*, pKM101); 22 EE125 (pKM101); 40 NR6112 (wild-type)] (Tables 1 and 3). Frameshifts comprised 51% of the non-hotspot mutations in CM6114 (121/235), 25% (22/87) in EE125, and 43% (40/93) in NR6112. The vast majority of frameshifts [99% in CM6114; 87% in EE125; 85% in NR6112] involved G·C base pairs. Minus (G·C) events predominated and were observed in 114 CM6114 mutants, 18 EE125 mutants, and 29 NR6112 mutants. Other types of frameshifts characterized in the three *E. coli* strains included 4 +(G·C), 6 −(GC·CG), 9 −(A·T) and 3 +(A·T) events.

Deletions accounted for 53 of the mutants recovered following 1,8-NONP treatment. Of these, 11 CM6114 (Δ*uvrB*, pKM101), 5 EE125 (pKM101) and 22 NR6112 (wild-type) deletions were cloned onto mRS81. DNA sequencing of these mutations showed that the deletions ranged in size from 3 to 343 bp (Table 4) and that the endpoints of 34/38 of the deletions contained regions of considerable (73–100%) homology. It has been suggested that palindromic structures within the deleted sequence might play a role in deletion formation (GLICKMAN and RIPLEY 1984; SCHAAPER, DANFORTH and GLICKMAN 1986; WESTON-HAFER and BERG 1989). The ability of each of the deleted sequences to form energetically stable hairpin structures was examined using the computer program "FOLD" (ZUKER and STEIGLER 1981), part of the Genetics Computer Group software package (DEVEREUX, HAEBERLI and SMITHIES 1984). Although a few of the deleted sequences did contain palindromic structures (for example, deletion 146–268 (GLICKMAN and RIPLEY 1984), such a feature was not common in this collection of deletions.

Although homology within the endpoints was the most prominent feature of the deletions recovered from 1,8-NONP-treated strains, other interesting sequences were noted. (1) The 3′-endpoint of deletion 352–394, which does not contain direct repeats, and the 3′ end of the repeated sequence in deletion 267–282 (recovered nine times) are at 5′-GATC-3′ sequences; this sequence is the site of MutH endonuclease activity during methyl-directed mismatch repair (LAHUE, AU and MODRICH 1989). (2) The sequence 5′-GTGG-3′ which is frequently found at sites of deletion and frameshift mutations (FIX, BURNS and GLICKMAN 1987; HALLIDAY and GLICKMAN 1991; MO, MAKI and SEKIGUCHI 1991) is near the endpoints of 11 of the deletions characterized in this study. (3) In a previous study we noted that no homology was apparent between the endpoints of a 401-bp 1,8-NONP-induced deletion in a Δ*uvrB* strain, but that the 5′-endpoint was flanked by two sequences which resembled the DNA gyrase consensus sequence (LAMBERT *et al.* 1991).

Four duplications were recovered following 1,8-NONP treatment (Table 5). The 6-bp duplication in the 610–621 region recovered in the EE125 (pKM101) and NR6112 (wild-type) collections creates a 5′-$(GCGTCT)_3$-3′ sequence from the original 5′-$(GCGTCT)_2$-3′ sequence and is located immediately 5′ to the frameshift hotspot site (position 620–632). The 20 bp duplication (695–714) in NR6112 also contains direct repeats at either end of the duplicated sequence.

Two NR6112 (wild-type) mutants, K158 and K183, contained complex changes (Table 5) which were comprised of both base substitution and frameshift events.

**Induction of mutational events by 1,8-NONP:** The extent to which any mutational class was induced by 1,8-NONP can be estimated by comparing the frequency of that event with and without 1,8-NONP treatment. The expected spontaneous frequency of different classes of mutation was calculated using the overall frequency of spontaneous mutation determined in the present study and the specificity of 729 and 198 spontaneous *lacI* mutations characterized in wild-type and pKM101 *E. coli*, respectively (HALLIDAY and GLICKMAN 1991; A. J. E. GORDON and B. W.

**TABLE 2**

**Base substitutions recovered from *E. coli* following 1,8-NONP treatment**

| Site | Mutation[a] | Amino acid | Sequence[b] | Occurrences | | |
|---|---|---|---|---|---|---|
| | | | | NR6112 (wild-type) | EE125 (pKM101) | CM6114 (*ΔuvrB*, pKM101) |
| 49 | C → A | Tyr → Oc3 | ACATC G TATAA | 0 | 1 | 0 |
| 53 | G → T | Val → Phe | ACGAT G TCGCA | 0 | 2 | 0 |
| 54 | T → G | Val → Gly | CTGCG A CATCG | 0 | 1 | 0 |
| 54 | T → C | Val → Ala | CTGCG A CATCG | 0 | 1 | 0 |
| 57 | C → A | Ala → Glu | ACTCT G CGACA | 0 | 2 | 2 |
| 59 | G → T | Glu → Am2 | TCGCA G AGTAT | 0 | 0 | 1 |
| 65, 66 | GC → AA | Ala → Asn | AGTAT GC CGGTG | 1 | 0 | 0 |
| 66 | C → A | Ala → Asp | CACCG G CATAC | 0 | 0 | 1 |
| 72 | T → A | Val → Asp | AAGAG A CACCG | 0 | 1 | 0 |
| 75 | C → A | Ser → Tyr | GATAA G AGACA | 0 | 0 | 2 |
| 80 | C → A | Gln → Lys | GGTCT G ATAAG | 0 | 0 | 3 |
| 81, 82 | AG → TT | Gln → Leu | TTATC AG ACCGT | 0 | 0 | 1 |
| 84 | C → A | Thr → Asn | AAACG G TCTGA | 1 | 0 | 0 |
| 86 | G → C | Val → Leu | AGACC G TTTCC | 0 | 0 | 1 |
| 89 | T → A | Ser → Thr | GCGGG A AACGG | 0 | 0 | 1 |
| 90 | C → A | Ser → Tyr | CGCGG G AAACG | 0 | 1 | 0 |
| 90 | C → T | Ser → Phe | CGCGG G AAACG | 0 | 0 | 1 |
| 92 | C → A | Arg → Ser | CACGC G GGAAA | 0 | 0 | 2 |
| 93 | G → T | Arg → Leu | TTCCC G CGTGG | 0 | 3 | 1 |
| 93 | G → A | Arg → His | TTCCC G CGTGG | 1 | 0 | 0 |
| 102 | A → T | Asn → Ile | GGTGA A CCAGG | 0 | 0 | 1 |
| 104 | C → T | Gln → Am6 | GGCCT G GTTCA | 1 | 1 | 0 |
| 105 | A → C | Gln → Pro | GAACC A GGCCA | 0 | 0 | 1 |
| 107 | G → A | Ala → Thr | ACCAG G CCAGC | 0 | 1 | 0 |
| 110 | A → T | Ser → Cys | AGGCC A GCCAC | 0 | 0 | 1 |
| 111 | G → A | Ser → Asn | GGCCA G CCACG | 1 | 0 | 0 |
| 116 | G → T | Val → Phe | GCCAC G TTTCT | 0 | 1 | 1 |
| 120 | C → A | Ser → Tyr | TCGCA G AAACG | 0 | 0 | 2 |
| 134 | G → T | Glu → Och | CGCGG G AAAAA | 0 | 0 | 1 |
| 143 | G → T | Glu → Och | AAGTG G AAGCG | 0 | 0 | 3 |
| 150 | C → A | Ala → Glu | CCATC G CCGCT | 0 | 1 | 0 |
| 158 | G → T | Glu → Am7 | TGGCG G AGCTG | 0 | 2 | 0 |
| 169 | C → A | Tyr → Oc8 | GGAAT G TAATT | 0 | 0 | 1 |
| 174 | C → A | Pro → His | GGTTG G GAATG | 0 | 0 | 2 |
| 174 | C → G | Pro → Arg | GGTTG G GAATG | 0 | 0 | 1 |
| 180 | G → C | Arg → Pro | CAACC G CGTGG | 1 | 0 | 0 |
| 185 | G → A | Ala → Thr | GCGTG G CACAA | 0 | 1 | 0 |
| 185 | G → C | Ala → Pro | GCGTG G CACAA | 0 | 1 | 0 |
| 185 | G → T | Ala → Ser | GCGTG G CACAA | 0 | 0 | 1 |
| 186 | C → A | Ala → Glu | GTTGT G CCACG | 1 | 0 | 2 |
| 195 | T → A | Leu → Gln | CCGCC A GTTGT | 0 | 0 | 1 |
| 200 | G → T | Gly → Cys | TGGCG G GCAAA | 0 | 0 | 1 |
| 201 | G → T | Gly → Val | GGCGG G CAAAC | 0 | 0 | 2 |
| 203 | A → T | Lys → Och | CGGGC A AACAG | 0 | 0 | 1 |
| 210 | C → A | Ser → Am10 | GCAAC G ACTGT | 1 | 0 | 4 |
| 221 | G → T | Gly → Cys | TGATT G GCGTT | 0 | 0 | 3 |
| 222 | G → T | Gly → Val | GATTG G CGTTG | 0 | 0 | 1 |
| 222 | G → A | Gly → Asp | GATTG G CGTTG | 1 | 0 | 0 |
| 228 | C → A | Ala → Asp | AGGTG G CAACG | 0 | 2 | 0 |
| 258 | C → A | Ser → Am12 | TTTGC G ACGGC | 0 | 0 | 2 |
| 270 | C → A | Ala → Glu | TCGCC G CGACA | 0 | 0 | 1 |
| 318 | C → A | Ser → Am13 | CCATC G ACACC | 0 | 0 | 4 |
| 326 | G → T | Glu → Oc14 | TGGTA G AACGA | 0 | 1 | 1 |
| 341 | G → T | Glu → Oc15 | GCGTC G AAGCC | 0 | 0 | 4 |
| 381 | G → A | Arg → His | GCAAC G CGTCA | 0 | 0 | 1 |
| 437 | G → T | Glu → Oc19 | CTGTG G AAGCT | 1 | 0 | 0 |
| 487 | G → T | Gln → His | GACCA G ACACC | 0 | 1 | 0 |
| 518 | G → T | Glu → Oc20 | CCCAT G AAGAC | 0 | 0 | 1 |
| 528 | C → A | Thr → Lys | GTCGC G TACCG | 0 | 0 | 2 |
| 537 | G → A | Gly → Asp | ACTGG G CGTGG | 0 | 0 | 1 |
| 623 | G → T | Ala → Ser | GTCTG G CTGGC | 0 | 1 | 0 |

*(Continued on p. 916)*

**TABLE 2**

*Continued*

| Site | Mutation[a] | Amino acid | Sequence[b] | Occurrences | | |
|---|---|---|---|---|---|---|
| | | | | NR6112 (wild-type) | EE125 (pKM101) | CM6114 (Δ*uvrB*, pKM101) |
| 677 | G → T | Glu → Oc26 | AACGG G AAGGC | 0 | 0 | 1 |
| 693 | C → A | Ala → Asp | ACATG G CACTC | 0 | 1 | 1 |
| 731 | G → T | Glu → Am25 | TGAAT G AGGGC | 0 | 1 | 1 |
| 750 | C → A | Ala → Glu | GCATC G CAGTG | 0 | 0 | 3 |
| 783 | G → T | Gly → Val | GCTGG G CGCAA | 0 | 1 | 1 |
| 783 | G → A | Gly → Asp | GCTGG G CGCAA | 1 | 0 | 0 |
| 795 | C → A | Ala → Asp | TAATG G CGCGC | 0 | 0 | 2 |
| 803 | G → T | Glu → Am27 | TTACC G AGTCC | 0 | 1 | 0 |
| 834 | C → A | Ser → Am28 | CTACC G AGATA | 0 | 1 | 0 |
| 842 | G → T | Gly → Opal | TAGTG G GATAC | 0 | 2 | 0 |
| 843 | G → T | Gly → Val | AGTGG G ATACG | 0 | 0 | 2 |
| 857 | G → T | Glu → Oc31 | ATACC G AAGAC | 0 | 1 | 0 |
| 867 | C → A | Ser → Oc32 | AACAT G AGCTG | 1 | 1 | 3 |
| 896 | A → T | Lys → Och | CCATC A AACAG | 0 | 0 | 1 |
| 928 | C → A | Ser → Arg | TCCAC G CTGGT | 0 | 1 | 0 |
| 939 | T → A | Leu → Am32 | GCAGC A AGCGG | 0 | 0 | 1 |
| 959 | C → T | Gln → Am34 | CGCCT G GCCCT | 0 | 0 | 1 |
| 993 | C → A | Ser → Oc36 | CCAGT G AGACG | 0 | 2 | 1 |
| 1005 | G → T | Arg → Ile | GAAAA G AAAAA | 0 | 1 | 0 |
| 1013 | A → C | Thr → Pro | AAACC A CCCTG | 0 | 1 | 0 |

[a] The nature of the mutation as determined on the nontranscribed strand.
[b] The sequence shown is that of strand containing the mutated purine and is read 5′ → 3′. Nonsense mutations are numbered according to MILLER, COULONDRE and FARABAUGH (1978). The numbering of the *lacI* gene is according to FARABAUGH (1978).

GLICKMAN, in preparation). Table 6 shows that while the frequency of some mutational classes was only slightly increased other classes were strongly induced by 1,8-NONP treatment. G·C → T·A transversions and −(G·C) frameshifts were clearly the most prominent induced events in both pKM101 strains. In CM6114 (Δ*uvrB*, pKM101) the frequency of these mutations was increased 810- and 3200-fold respectively, while in EE125 (pKM101) the frequency of G·C → T·A transversions and −(G·C) frameshifts was increased 36- and 39-fold, respectively. Other events whose frequency was increased appreciably above the spontaneous level in the Δ*uvrB*, pKM101 strain included A·T → T·A transversions, G·C → A·T transitions, and deletions. In NR6112 (wild-type), −(G·C) frameshift events were induced 12-fold above the spontaneous level. Much smaller increases were observed for other mutational events, which did not exceed the observed induction of ±CTGG events.

**Specificity of 1,8-NONP-induced bases substitutions:** Base substitutions were specifically induced by 1,8-NONP in those strains which contained pKM101 (Table 6). Of the 119-base substitution mutations which were characterized in EE125 (pKM101) and CM6114 (Δ*uvrB*, pKM101), 107 occurred at G·C sites. Fully 91% of the base substitution mutations at these sites were G·C → T·A transversions.

There are 75 known sites in the *lacI* gene at which a G·C → T·A transversion yields a selectable $I^-$ phenotype on PGal medium (MILLER, COULONDRE and FARABAUGH 1978; GORDON *et al.* 1988; and our unpublished results). We assume here that these 75 sites give a reasonable approximation of the total *lacI* G·C → T·A target. The influence of flanking bases on the mutability of particular guanine sites was examined by comparing the observed distribution of independent mutations to the distribution of the detectable sites with a $\chi^2$ goodness of fit test . With respect to the bases immediately 3′ or 5′ the mutated guanine residue there was no significant difference ($P > 0.05$ in all cases) between the distribution of G·C → T·A transversions recovered in EE125 (pKM101) or CM6114 (Δ*uvrB*, pKM101) which had been treated with 1,8-NONP and the expected distribution.

In CM6114 (Δ*uvrB*, pKM101), eight base substitution mutations occurred at A·T sites, of which seven were A·T → T·A transversions. It is striking that of the 80-base substitution mutations collected from the Δ*uvrB*, pKM101 strain following 1,8-NONP treatment, 73 were transversions that could arise by misincorporation of adenine across from the purine during DNA synthesis (66 G·C → T·A and 7 A·T → T·A events).

**Specificity of 1,8-NONP-induced frameshift mutation:** The frequency of −(G·C) frameshift mutations was significantly elevated in all strains by 1,8-NONP treatment (Table 6). Of the 183 frameshift mutations recovered following 1,8-NONP treatment, 176 involved base loss [161 −(G·C) events, 9 −(A·T) events, and 6 −(GC·CG) events].

**TABLE 3**

**Frameshift mutations recovered from *E. coli* following 1,8-NONP treatment**

| Site | Mutation[a] | Sequence[b] | Occurrences | | |
|---|---|---|---|---|---|
| | | | NR6112 (wild-type) | EE125 (pKM101) | CM6114 (*ΔuvrB*, pKM101) |
| 66-67 | −C | CACC (GG) CATA | 0 | 0 | 1 |
| 90-92 | −C | ACGC (GGG) AAAC | 1 | 4 | 10 |
| 90-92 | +C | ACGC (GGG) AAAC | 1 | 0 | 0 |
| 108-109 | −C | GGCT (GG) CCTG | 0 | 0 | 1 |
| 112-113 | −C | ACGT (GG) CTGG | 0 | 0 | 3 |
| 132-134 | −G | ACGC (GGG) AAAA | 0 | 0 | 3 |
| 132-134 | +G | ACGC (GGG) AAAA | 0 | 1 | 1 |
| 135-139 | −A | CGGG (AAAAA) GTGG | 1 | 0 | 0 |
| 135-139 | +A | CGGG (AAAAA) GTGG | 0 | 0 | 1 |
| 142-143 | −G | AAGT (GG) AAGC | 0 | 0 | 1 |
| 160-161 | −GC | CGGA (GC) TGAA | 1 | 0 | 0 |
| 171-172 | −T | TGGG (AA) TGTA | 0 | 0 | 1 |
| 173-175 | −C | GGTT (GGG) AATG | 3 | 0 | 4 |
| 186 | −C | TTGT (G) CCAC | 0 | 0 | 1 |
| 187 | +A | TGGC (A) CAAC | 0 | 1 | 0 |
| 188 | −C | TGTT (G) TGCC | 0 | 0 | 1 |
| 189-190 | −A | GCAC (AA) CAAC | 3 | 0 | 0 |
| 199-201 | −G | TGGC (GGG) CAAA | 0 | 0 | 1 |
| 202 | −C | GTTT (G) CCCG | 0 | 0 | 1 |
| 221-222 | −G | GATT (GG) CGTT | 2 | 0 | 0 |
| 228-229 | −C | AGGT (GG) CAAC | 1 | 0 | 0 |
| 234-235 | −C | GACT (GG) AGGT | 0 | 0 | 4 |
| 241-242 | −G | GTCT (GG) CCCT | 1 | 1 | 0 |
| 243-245 | −C | TGCA (GGG) CCAG | 0 | 1 | 3 |
| 250-254 | −CG | TGCA (CGCGC) CGTC | 1 | 0 | 2 |
| 298-300 | −G | AACT (GGG) TGCC | 0 | 0 | 2 |
| 303-304 | −C | CGCT (GG) CACC | 0 | 0 | 2 |
| 385 | −C | CACT (G) ACGC | 0 | 1 | 0 |
| 389-391 | −G | CAGT (GGG) CTGA | 0 | 2 | 3 |
| 412-413 | −G | CGCT (GG) ATGA | 1 | 0 | 1 |
| 421-422 | −G | ACCA (GG) ATGC | 0 | 0 | 2 |
| 426-427 | −C | CAAT (GG) CATC | 0 | 1 | 0 |
| 484-485 | −C | GTCT (GG) TCAG | 1 | 0 | 0 |
| 491-493 | −C | TGAT (GGG) TGTC | 1 | 1 | 7 |
| 496 | −C | TGTT (G) ATGG | 0 | 0 | 1 |
| 513-515 | −C | TCAT (GGG) AGAA | 0 | 1 | 1 |
| 524-525 | −G | AGAC (GG) TACG | 1 | 1 | 0 |
| 559-561 | −G | CATT (GGG) TCAC | 1 | 0 | 0 |
| 574-578 | −CG | AAAT (CGCGC) TGTT | 1 | 0 | 1 |
| 586-588 | −G | TAGC (GGG) CCCA | 1 | 0 | 2 |
| 589-591 | −C | TAAT (GGG) CCCG | 1 | 1 | 4 |
| 604 | +C | CCGA (G) ACAG | 1 | 0 | 0 |
| 670-671 | −G | TAGC (GG) AACG | 0 | 0 | 1 |
| 675-677 | −G | GAAC (GGG) AAGG | 0 | 1 | 1 |
| 680-681 | −G | GGAA (GG) CGAC | 0 | 0 | 1 |
| 733-735 | −G | ATGA (GGG) CATC | 0 | 0 | 2 |
| 743-745 | −C | CAGT (GGG) AACG | 1 | 1 | 10 |
| 756-758 | −G | TGCT (GG) TTGC | 0 | 0 | 2 |
| 781-783 | −G | CGCT (GGG) CGCA | 2 | 0 | 5 |
| 809-811 | −G | GTCC (GGG) CTGC | 2 | 0 | 1 |
| 841-843 | −G | TAGT (GGG) ATAC | 4 | 1 | 1 |
| 877-879 | −C | CGGC (GGG) ATAT | 0 | 0 | 9 |
| 888-889 | −C | TGGT (GG) TTAA | 0 | 0 | 1 |
| 891-892 | −C | TGAT (GG) TGGT | 0 | 0 | 1 |
| 896-898 | −A | CATC (AAA) CAGG | 1 | 0 | 1 |
| 916-919 | −G | TGCT (GGGG) CAAA | 3 | 0 | 12 |
| 927 | −G | ACCA (G) CGTG | 0 | 1 | 0 |
| 955-957 | −G | CTCA (GGG) CCAG | 0 | 0 | 1 |
| 970-972 | −G | TGAA (GGG) CAAT | 1 | 0 | 4 |
| 985 | −G | TGTT (G) CCCG | 0 | 0 | 2 |
| 986-988 | −C | AGAC (GGG) CAAC | 1 | 0 | 0 |
| 1006-1010 | −A | AAAG (AAAAA) CCAC | 0 | 2 | 0 |
| 1006-1010 | +A | AAAG (AAAAA) CCAC | 1 | 0 | 0 |
| 1011-1012 | −C | GGGT (GG) TTTT | 0 | 0 | 1 |

[a] The nature of the mutation as determined on the nontranscribed strand.

[b] The sequence given (5′ 3′) is taken from the purine containing strand; for −2 frameshifts the sequence is taken from the nontranscribed strand. The frameshift occurs in the sequence shown in parentheses.

**TABLE 4**

**Deletions recovered from *E. coli* following 1,8-NONP treatment**

| Site | Occurrence | Size (bp) | Sequence 5′-[deleted sequence]-3′ |
|---|---|---|---|
| NR6112 (wild-type) | | | |
| (−88)–78 | 1 | 165 | AGGCTATTC [TGGTGGCC . . . TGTCTCTTA] TCAGAC |
| 146–268 | 3 | 123 | AAGTGGAA [GCGGCGATGGC . . GCAAATTGTC] GCGGCGATTAA |
| 267–282 | 4 | 16 | GCAAATTG [TCGCGGCGATTAAATC] TCGCGCCGATC |
| 301–423 | 2 | 123 | CAACTGGG [TGCCAGCGT . . . TGGATGACCAGGA] TGCCAT |
| 306–316 | 1 | 11 | GGGTGCCA [GCGTGGTGGTG] TCGATGGTAGAA |
| 331–350 | 1 | 20 | GTAGAACG [AAGCGGCGTCGAAGCCTGTA] AAGCGGCGGT |
| 339–357 | 1 | 19 | GAAGCGGCG [TCGAAGCCTGTAAAGCGGC] GGTGCACA |
| 352–394 | 1 | 43 | GAAGCCTGTAA [AGCGGCG . . . CAGTGGGCTG] ATCATTA |
| 418–425 | 1 | 8 | CTGGATGA [CCA GGATG] CCATTGCTG |
| 454–796 | 1 | 343 | CTGCACTAA [TGTTCCGGC . . . ATGCGCGCC] ATTACCG |
| 572–874 | 1 | 303 | ACCAGCAA [ATCGCGCTGTTA . . . TGTTAT] ATCCCGCCGTAAC |
| 776–815 | 1 | 40 | AGATG [GCGCTGGGCGCA . . . CCGGGCTGC] GCGTTGGTGCGG |
| 788–796 | 1 | 9 | GGGCGCA [ATGCGCGCC] ATTACCGAG |
| 917–969 | 2 | 53 | CCTGCTG [GGGCAAACCAG . . . GCGGTGAA] GGGCAATCAGCTG |
| 943–969 | 1 | 27 | CCGCTTGCT [GCAACTCT . . . GCGGTGAA] GGGCAATCAGC |
| EE125 (pKM101) | | | |
| 143–329 | 1 | 187 | AAAAAGTG [GAAGCGGCGAT . . . GTAGAAC] GAAGCGGCGTC |
| 250–353 | 2 | 104 | CCTGCA [CGCGCCGTCGCAAATT . . . TAAAG] CGGCGGTGCA |
| 267–282 | 1 | 16 | GCAAATTG [TCGCGGCGATTAAATC] TCGCGCCGATC |
| 331–350 | 1 | 20 | GTAGAACG [AAGCGGCGTCGAAGCCTGTA] AAGCGGCGGT |
| CM6114 (Δ*uvrB*, pKM101) | | | |
| 187–189 | 1 | 3 | GCGTGGC [ACA] ACAACTGGC |
| 267–282 | 4 | 16 | GCAAATTG [TCGCGGCGATTAAATC] TCGCGCCGATC |
| 339–357 | 1 | 19 | GAAGCGGCG [TCGAAGCCTGTAAAGCGGC] GGTGCACA |
| 828–837 | 1 | 10 | GTTGGTGCGG [ATATCTCGGT] AGTGGGAT |
| 917–969 | 1 | 53 | CCTGCTG [GGGCAAACCAG . . . GCGGTGAA] GGGCAATCAGC |
| 929–943 | 2 | 15 | GGGGCAAACCAGC [GTGGACCGCTTGCTG] CAACTCTCT |
| 980–1022 | 1 | 43 | GCAATCAG [CTGTTGCCC . . . TGGCGC] CCAATACGC |

The sequence shown is that of the nontranscribed strand. Repeated bases at deletion endpoints are underlined.

**TABLE 5**

**Complex mutations and duplications recovered from *E. coli* following 1,8-NONP treatment**

| Mutant | Position | Mutation | Sequence change[a] |
|---|---|---|---|
| NR6112, K183 | 483/485 | Complex | TG ACCAGA → TG GCAGA |
| NR6112, K158 | 189 | Complex | GCAC AACAACT → GCAC GCCCAACT |
| NR6112, K191 | 610–615 | Duplication | CTCGGC GCGTCT GCGTCTGG |
| NR6112, K8 | 695–714 | Duplication | GTGCC ATGTCCGGTTTTCAACAAAC CATGC |
| EE125, L70 | 610–615 | Duplication | CTCGGC GCGTCT GCGTCTGG |
| CM6114, N165 | 863–883 | Duplication | AGAC AGCTCATGTTATATCCCGCCG TTAA |

[a] The sequence shown is that of the nontranscribed strand. The mutation involves the underlined sequence.

The dependence of mutation frequency on the length of the reiterated sequence was examined by expressing the number of −(G·C) events as a function of the target size (*S*) (*S* = length of the run (*R*) multiplied by the number of detectable sites of size *R* in the *lacI* gene) (Table 7). To obtain *S*, only G·C runs at, or prior to, position 1012 were considered since −1 frameshift mutations after this point do not yield a selectable phenotype (CALOS and MILLER 1981; LAMBERT *et al.* 1991). These calculations show that the number of observed frameshifts per guanine in a run of given length was 0.016, 0.12, 0.95 and 2.3 for $(G)_1$, $(G)_2$, $(G)_3$ and $(G)_4$ sites, respectively. These values are consistent with similar calculations in a Δ*uvrB* strain in which the number of 1,8-NONP-induced −(G·C) frameshifts in a run of given length was 0.019, 0.18, 0.87 and 3.5 for $(G)_1$, $(G)_2$, $(G)_3$ and $(G)_4$, respectively (LAMBERT *et al.* 1991). Thus the frequency of frameshift mutation increases markedly with the length of the reiterated sequence.

The influence of neighboring bases on the relative mutability of particular frameshift sites was addressed by classifying the detectable −(G·C) frameshift sites in the *lacI* gene and the sites of 1,8-NONP-induced −(G·C) mutation according to the nature of their flanking bases. The influence of the 5′ or 3′ flanking base on the occurrence −(G·C) frameshifts was analyzed using the combined data for all strains with a $\chi^2$ goodness of fit test. Minus (G·C) mutations at 5′-NGGGN-3′ sites were recovered 106 times. The distribution of the frameshift mutations recovered from 1,8-NONP treated strains was not significantly differ-

**TABLE 6**

**Induction of mutation events by 1,8-NONP**

A. Induction of mutational classes by 1,8-NONP

| Mutation | NR6112 (wild-type) frequency (×10⁻⁶) 1,8-NONP | Spontaneous | Fold increase | EE125 (pKM101) frequency (×10⁻⁶) 1,8-NONP | Spontaneous | Fold increase | CM6114 (ΔuvrB, pKM101) frequency (×10⁻⁶) 1,8-NONP | Spontaneous | Fold increase |
|---|---|---|---|---|---|---|---|---|---|
| G·C → T·A | 0.21 | 0.09 | 2.3 | 7.5 | 0.21 | 36 | 73 | 0.09 | 810 |
| G·C → A·T | 0.21 | 0.33 | 0.6 | 0.70 | 0.25 | 2.8 | 4.4 | 0.10 | 44 |
| A·T → T·A | Insufficient # mutants | | | Insufficient # mutants | | | 7.7 | 0.09 | 86 |
| −(G·C) | 1.2 | 0.10 | 12 | 4.3 | 0.11 | 39 | 130 | 0.04 | 3200 |
| −(A·T) | 0.21 | 0.10 | 2.1 | Insufficient # mutants | | | Insufficient # mutants | | |
| Deletions | 1.2 | 0.52 | 2.3 | 2.4 | 0.85 | 2.8 | 17 | 0.35 | 49 |
| CTGG (+) | 5.5 | 3.0 | 1.8 | 11 | 3.0 | 3.7 | 4.4 | 1.2 | 3.6 |
| CTGG (−) | 1.8 | 0.77 | 2.3 | 3.9 | 1.1 | 3.5 | Insufficient # mutants | | |
| Total | 11 | 5.4 | 2.0 | 36 | 7.3 | 4.9 | 265 | 3.0 | 88 |

B. Influence of plasmid pKM101 and ΔuvrB on the frequency of mutational events.

| | Effect of plasmid pKM101[a] | | Effect of ΔuvrB[b] | |
|---|---|---|---|---|
| | EE125/NR6112 | CM6114/NR6113[c] | NR6113[c]/NR6112 | CM6114/EE125 |
| G·C → T·A transversions | 61 | 29 | 21 | 10 |
| −(G·C) frameshifts | 3.8 | 1.8 | 66 | 31 |

The calculations in A are based on the frequencies for spontaneous and 1,8-NONP-induced mutations as obtained in this study, the specificity data of Table 1, and the specificity of spontaneous mutation. For NR6112, the distribution of spontaneous mutations is that described in HALLIDAY and GLICKMAN (1991). For calculations involving EE125 and CM6114 we have used the distribution of spontaneous mutation in EE125 (A. J. E. GORDON and B. W. GLICKMAN, in preparation). It is assumed that excision repair does not influence the observed distribution of mutation (J. A. HALLIDAY, F. A. ALLEN, and B. W. GLICKMAN, unpublished results). Other mutational classes were not observed in sufficient numbers to warrant inclusion here.

[a] Values refer to the ratio of the 1,8-NONP-induced frequency of each type of mutation in a strain containing plasmid pKM101 divided by the frequency of that type of mutation in an otherwise isogenic strain lacking pKM101.

[b] Values refer to the ratio of the 1,8-NONP-induced frequency of each type of mutation in a strain deficient in nucleotide excision repair as a result of a chromosomal ΔuvrB mutation divided by the frequency of that type of mutation in an otherwise isogenic $uvrB^+$ strain.

[c] Frequency data for calculations involving NR6113 is from LAMBERT *et al.* (1991).

**TABLE 7**

**The influence of reiterated sequence length on the frequency of 1,8-NONP-induced −(G·C) frameshift mutations**

| Length[a] | No. sites[b] | Target[c] | Occurrences/target size[d] NR6112 (wild-type) | EE125 (pKM101) | CM6114 (ΔuvrB, pKM101) | Weighted average[e] |
|---|---|---|---|---|---|---|
| 1 | 312 | 312 | | 0.035[d] | 0.017 | 0.016 |
| 2 | 81 | 162 | 0.15 | 0.10 | 0.13 | 0.12 |
| 3 | 23 | 69 | 0.95 | 1.0 | 0.94 | 0.95 |
| 4 | 1 | 4 | 2.6 | | 2.6 | 2.3 |

[a] The number of G·C base pairs in the reiterated sequence.

[b] The number of sites which occur at or prior to position 1012.

[c] The target size is the length of the run multiplied by the number of detectable sites of similar size in the *lacI* gene.

[d] Values are expressed as occurrences/target size for a hypothetical population of 100 frameshift mutants.

[e] The weighted average is calculated by considering the total number of −(G·C) mutations in the three strains which occur at sites of the specified length, and then normalizing the value to a hypothetical population of 100 frameshift mutants. Therefore the strains with the largest number of frameshift mutations contribute most to this weighted average.

ent from the distribution of detectable $(G)_3$ sites with respect to the nature of 5′-flanking base. However, analysis of the 3′-flanking base showed that 5′-NGGGA-3′ sites were represented more often in the 1,8-NONP-induced spectrum (53%) than would be expected on the basis of the number of detectable 5′-NGGGA-3′ sites (35%). This difference is statistically significant ($P < 0.005$).

A total of 32 frameshift mutations occurred at 5′-NGGN-3′ sites. The majority (78%) of these mutations occurred at 5′-TGGN-3′ sites: 78% in CM6114 (ΔuvrB, pKM101); 67% in EE125 (pKM101); 86% in NR6112 (wild-type). The bias toward −(G·C) frameshifts at 5′-TGGN-3′ sites, and away from 5′-CGGN-3′ sites, is significant ($P < 0.05$). When the 3′-flanking base was examined, 5′-NGGA-3′ sites were represented to a greater extent than was expected on the basis of the number of detectable sites, but this difference was not statistically significant.

While too few −(G·C) frameshift mutations occurred at sites containing lone guanine residues to allow statistical analysis it might be noted that 7/8

−(G·C) frameshifts at such sites contained a 5′-T.

The 12 frameshift mutations which occurred at A·T base pairs consisted of 9 losses and 3 additions. Eleven of the 12 A·T frameshifts occurred in runs of contiguous adenine residues. The relatively high proportion of these mutations (9/12) which were recovered from the *uvr*$^+$ strains NR6112 (wild-type) and EE125 (pKM101) suggest that some have spontaneous origins (Table 6).

## DISCUSSION

Nature of the premutational lesion: More than 95% of the total DNA adduct formed by 1,8-NONP treatment of *E. coli* is the C(8) guanine lesion dG-C(8)-ANP (LAMBERT *et al.* 1991). The DNA sequencing data presented here shows that more than 92% (289/313) of all point mutations recovered from NR6112 (wild-type), EE125 (pKM101) and CM6114 (Δ*uvrB*, pKM101) following 1,8-NONP treatment occurred at G·C base pairs. Similarly, in a spectrum of sequenced 1,8-NONP-induced mutations in a Δ*uvrB* strain, 122/123 point mutations occurred at G·C base pairs (LAMBERT *et al.* 1991). Thus, in total, about 95% of the point mutations which we have characterized in the *lacI* gene following 1,8-NONP treatment occur at G·C base pairs which correlates well with the proportion of DNA adduct formed at the C(8) position of guanine. Although these observations strongly suggest that the dG-C(8)-ANP adduct is the principal premutagenic lesion, confirmation of this hypothesis will require determination of the mutagenic potential of 1,8-NONP adducts using site specific single adduct mutagenesis experiments.

A small number of other point mutations at A·T bases were characterized (Table 1) including A·T → T·A transversions whose frequency was increased approximately 86-fold above the spontaneous level in CM6114 (Δ*uvrB*, pKM101) (Table 6). Some of the mutations may be targeted by the minor dA or dT adduct which comprises some 2–3% of the total adduct formed as determined by $^{32}$P-postlabeling analysis (LAMBERT *et al.* 1991). Although this minor lesion has not been unambiguously identified there is evidence (MCCOY *et al.* 1985) that it could result from a reaction between the aryl nitrenium/carbonium ion derived from 1,8-NONP and dA.

**Influence of genetic background on the mutational specificity:** Our data show that the most prominent classes of mutation induced by 1,8-NONP in *E. coli* are −(G·C) frameshifts and G·C → T·A transversions (Table 6). The extent to which the frequency of these mutations was increased above the spontaneous frequency was strongly influenced by the host genetic background. In the wild-type strain NR6112 −(G·C) frameshifts were the only class of mutation whose frequency was increased appreciably (12-fold) above the spontaneous level. An increase in the frequency of −(G·C) frameshifts (39-fold) and G·C → T·A transversions (36-fold) occurred in EE125 (pKM101). An elevation of both the overall frequency of mutation and the extent to which 1,8-NONP induced specific classes of mutation was observed in strains deficient in nucleotide excision repair. In CM6114 (Δ*uvrB*, pKM101) −(G·C) frameshifts were induced 3200-fold and G·C → T·A transversions were induced 810-fold by 1,8-NONP treatment. In the nucleotide excision repair deficient strain NR6113 (Δ*uvrB*), which is isogenic to CM6114 except that it lacks the plasmid pKM101 (Table 1; LAMBERT *et al.* 1991), −(G·C) frameshifts were induced 730-fold while G·C → T·A transversions were induced 26-fold. Comparison of the induced mutation frequency of G·C → T·A transversions and −(G·C) frameshifts in strains which are isogenic except for the presence of plasmid pKM101, which encodes the *mucA/B* homologs of the *umuD/C* genes which are believed to facilitate error-prone bypass of DNA lesions (WALKER 1984; BRIDGES and WOODGATE 1985), demonstrates that the extent to which 1,8-NONP induces G·C → T·A transversions is increased by a factor of 30–60 in the presence of pKM101. In contrast, the presence of pKM101 effected only a slight increase in the frequency of −(G·C) frameshifts (Table 6). The influence of nucleotide excision repair capability on the distribution of mutation was also analyzed (Table 6). The deficiency in nucleotide excision repair resulted in a significant increase in the frequency of both 1,8-NONP-induced −(G·C) frameshifts and G·C → T·A transversions events with frameshift mutations perhaps increased to a greater extent than G·C → T·A transversions.

With respect to the influence of plasmid pKM101 the present data are consistent with the overall mutational profile of 1,8-DNP inferred from reversion studies in *S. typhimurium*. 1,8-DNP is nonmutagenic at the base substitution locus *hisG46* in the absence of plasmid pKM101 (TA1535) but is extremely mutagenic in the isogenic strain TA100 which contains the plasmid pKM101. Frameshift mutations at both the *hisC3076* and *hisD3052* loci are induced efficiently in strains lacking pKM101 (ROSENKRANZ and MERMELSTEIN 1983; TOKIWA and OHNISHI 1986). Similar results have been reported for aflatoxin $B_1$ ($AFB_1$). MucA/B function has been shown to be required for $AFB_1$-induced base substitution; increasing the number of copies of UmuDC by providing *umuD/C*$^+$ on a plasmid failed to increase the frequency of $AFB_1$-induced base substitutions (FOSTER, GROOPMAN and EISENSTADT 1988). Although the frequency of $AFB_1$-induced −1 frameshifts was increased by the presence of MucA/B, such events were nevertheless observed

frequently in *mucA/B*$^-$bacteria (REFOLO, BENNET and HUMAYUN 1987). It should be noted that the bacterial strains used both in our studies with 1,8-NONP and the $AFB_1$ studies (FOSTER, GROOPMAN and EISENSTADT 1988; REFOLO, BENNET and HUMAYUN 1987) are *umuD/C*$^+$. Thus it appears that with $AFB_1$, and possibly also with 1,8-NONP, that the plasmid-encoded MucA/B is more active at enhancing base substitution mutagenesis than endogenous UmuD/C. This might reflect different properties or activities of the UmuD/C and MucA/B proteins. In this context we note that although RecA-cleaved fragments of both MucA and UmuD are active in mutagenesis, the cleavage event is not obligatory for MucA-mediated mutagenesis activity (SHIBA *et al.* 1990).

The small effect of pKM101 on the frequency of −(G·C) frameshift mutations indicates either that high levels of frameshift mutagenesis are mediated by endogenous UmuD/C protein, or alternatively, that these mutations occur independently of MucA/B and UmuD/C function. The former possibility is supported by the observations that −(G·C) frameshifts induced by both the dG-C(8) adduct of 2-acetylaminofluorene (AAF) (KOFFEL-SCHWARTZ *et al.* 1984),and the dG-N(7) adduct of AFB1 (BENNET, LUO and HUMAYUN 1991) are *umuC*-dependent.

Preferential repair of lesions from the transcribed strand of expressed genes has been documented in a variety of systems (see TERLETH, VAN DE PUTTE and BROUWER 1991), and in *E. coli* appears to be mediated by a transcription-repair coupling factor encoded by the *mfd* locus (SELBY, WITKIN and SANCAR 1991). The existence of differential repair may also have consequences for the distribution of mutations between strands. We have examined this possibility by comparing the strand distribution of 1,8-NONP-induced G·C → T·A and −(G·C) events in strains which differ with respect to the Δ*uvrB* mutation [*i.e.*, NR6113 (Δ*uvrB*) *vs.* NR6112 (wild-type) and CM6114 (Δ*uvrB*, pKM101) *versus* EE125 (pKM101)]. We make the assumption that these mutations arise as the result of a lesion on a guanine residue. Of the 75 sites in the *lacI* gene at which G·C → T·A transversions are known to produce an I$^-$ phenotype, 38 sites are in the nontranscribed strand and 37 are in the transcribed strand. Comparison of the strand distribution of G·C → T·A transversions recovered from CM6114 and EE125 with the known distribution of detectable sites using a $\chi^2$ goodness of fit test shows that there is no difference in either strain between the observed and expected distributions. For the analysis of the strand distribution of −(G·C) frameshifts we used the 23 $(G)_3$ sites in the *lacI* gene which give a detectable phenotype, of which 14 sites are on the nontranscribed strand and 9 are on the transcribed strand. The strand distribution of mutation at $(G)_3$ sites in both NR6113 and NR6112 is identical to the known distribution of detectable sites. Although significantly more −(G·C) frameshifts occur on the transcribed strand in both CM6114 and EE125 than are expected on the basis of the distribution of detectable sites ($P < 0.05$ for EE125 data; $P < 0.005$ for CM6114 data) the observed strand specificity is independent of nucleotide excision repair capability. Thus, for those classes of mutation where the number of mutations is large enough to warrant statistical analysis there is no evidence that a deficiency/proficiency in excision repair conferred by the Δ*uvrB/uvr*$^+$ genotype alters the strand distribution of mutation.

Although plasmid pKM101 only moderately influenced −(G·C) frequency (2–4-fold) it had a significant influence on the strand specificity of such mutation. As noted above, in both EE125 (pKM101) and CM6114 (Δ*uvrB*, pKM101) there were significantly more −(G·C) frameshift events at $(G)_3$ sites on the transcribed strand than expected on the basis of the number of detectable $(G)_3$ sites. This was in contrast to the isogenic strains which lacked pKM101 in which the distribution of mutations at $(G)_3$ sites was similar to the expected distribution. Although a similar tendency was observed for −(G·C) events at $(G)_2$ sites [CM6114 (Δ*uvrB*, pKM101) *vs.* NR6113 (Δ*uvrB*)] the difference was not significant. As noted above, strand specificity of mutation could result in strand specific repair associated with transcription. However, we have found no evidence that differential nucleotide excision repair imparted by a Δ*uvrB* mutation alters the strand specificity of 1,8-NONP mutation. Moreover it would be expected that such strand specific repair would result in a preponderance of mutation on the nontranscribed strand which is clearly not the case. Thus, the strand specificity observed here may not be associated with transcriptionally coupled repair, but rather, may be a consequence of DNA replication, perhaps reflecting a different interaction of pKM101-encoded mutagenesis functions with leading or lagging strands during translesion synthesis.

**Base substitution mutations:** G·C → T·A changes represented a significant proportion of the mutations induced by 1,8-NONP in pKM101 strains (Table 1) suggesting that the mutagenic pathway involves error-prone DNA synthesis. The nature of these mutations can be considered in the context of (1) insertion opposite the lesion, (2) removal of terminal nucleotides by the exonucleolytic proofreading activity of DNA polymerase and (3) extension of a particular terminus to produce a mutation. The ability of the mismatch repair system to differentially recognize and repair certain base:adduct mismatches might also be considered. In principle, the fact that mutations result most frequently from adenine being inserted opposite the DNA adduct could reflect the ability of the lesion

to influence the selectivity of any of these steps.

Studies which have examined base insertion opposite other dG-C(8) lesions have reported that cytosine is the most commonly inserted base (MICHAELS *et al.* 1991; RABKIN and STRAUSS 1984). This might be attributable to the fact that DNA adduct formation at the C(8) position of guanine does not directly interfere with the three normal hydrogen bonding positions of the base and consequently insertion of the proper base could be directed by the adducted base. At lower frequency, misinsertion of adenine might also be directed by guanine adducts in either the *anti* or *syn* conformations. Computer modeling (BROYDE *et al.* 1990) and NMR (NORMAN *et al.* 1989b) studies have suggested that dG-C(8) adducts formed with 2-aminofluorene can pair with either C or A when the modified guanine residue is in the *syn* conformation. The structures require that the base paired to the adduct be in ionized or tautomeric forms which are, in fact, more frequent in DNA duplexes than was anticipated from the pK of free bases in solution (QUIGLEY *et al.* 1986; SOWERS *et al.* 1987). LOECHLER (1989) has proposed that bulky C(8) lesions might shift guanine in its *anti* form toward the major groove, a perturbation termed adduct-induced base wobble. This would result in the guanine becoming more "pyrimidine-like" and might facilitate the formation of a G:A mismatch in which the $O^6$ and H-N(1) of guanine pair with the exocyclic amino group and N(1), respectively, of adenine (KAN *et al.* 1983; LOECHLER 1989). A final possibility for a G·A base pair is suggested by the crystal structure of a G(unmodified, *anti*)·A(*syn*) pair in which the $O^6$ and H-N(1) of guanine are paired with $H_2N^2$ and N(7), respectively, of adenine in the *syn* conformation (BROWN *et al.* 1986).

An alternative possibility is that the dG-C(8)-ANP lesion represents a template site which is noninformational to a DNA polymerase; it has been postulated that such sites are bypassed with the preferential incorporation of adenine (STRAUSS 1991). One such noninformational lesion, the apurinic/apyrimidinic (AP) site, can arise enzymatically or from increased hydrolysis of the *N*-glycosylic bond of some unstable DNA adducts. We note however, that adduction at the C(8) position of guanine does not alter the charge of the purine ring at physiological pH, and is unlikely to enhance the lability of the *N*-glycosylic bond. Moreover, the known specificity of mutation across from AP sites cannot account for the distribution of base substitutions observed here in that bases are inserted opposite AP sites in the order A (56%) > T (29%) > G (15%) (KUNKEL 1984). Stable noninformational adducts might also result from an adduct-induced *anti* → *syn* conformational change, such as that observed following binding of AAF to the C(8) position of guanine (EVANS, MILLER and BELAND 1980).

Recent studies have provided evidence that bases adjacent to the site of base substitution can direct incorporation of an incorrect base as a result of transient misalignment of template and primer strands during DNA synthesis (KUNKEL and SONI 1988). However, the specificity of the G·C → T·A transversions recovered in this study, which suggests that there is no nearest base influence on G·C → T·A events, argues strongly against such a model for 1,8-NONP-induced events.

None of the eight A·T → T·A transversions recovered following 1,8-NONP treatment occurred at sites which could produce this mutation by a transient misalignment mechanism. Potential mechanisms for A·T → T·A transversion are similar to those previously discussed for G·C → T·A transversions. There are three potential A·A pairs: (1) A*anti*·A*anti*, (2) A(*anti*)·A(*anti*)(imino) and (3) A(*anti*)(imino)·A(*syn*).

**1,8-NONP-induced frameshift mutations:** Reiterated sequences are important sites for 1,8-NONP-induced −(G·C) frameshifts with the likelihood of such a mutation, when expressed per guanine, increasing with the number of contiguous guanine residues (Table 7). STREISINGER *et al.* (1966) and STREISINGER and OWEN (1985) have postulated that misaligned replication intermediates derived from slippage of one strand relative to the other might be stabilized within reiterated sequences. It has been proposed that chemicals which bind to the C(8) position of guanine might promote frameshifting through an "incorporation-slippage" model (Figure 2A) (LAMBERT *et al.* 1990, 1991; SCHAAPER, KOFFEL-SCHWARTZ and FUCHS 1990) in which the dG-C(8) adduct pairs with cytosine (RABKIN and STRAUSS 1984; MICHAELS *et al.* 1991) but then hinders the progression of the DNA polymerase thereby providing increased opportunity for strand slippage. Within reiterated sequences a guanine residue 5′ to the adduct on the template strand might pair with the nascent cytosine to form a misaligned frameshift intermediate with a normal G·C template/primer terminus which could be extended to form the frameshift. This hypothesis was tested recently by examining the mutagenicity of plasmids containing a single dG-C(8)-AAF adduct at defined positions in a run of contiguous guanine residues. Although the observation that dG-C(8)-AAF adducts at the 3′-end of guanine runs are much more mutagenic than lesions at the 5′-end of such sequences (LAMBERT, NAPOLITANO and FUCHS 1992) was consistent with the model in Figure 2A, the additional observations (a) that the mutagenicity of dG-C(8)-AAF adducts increases as a function of the number of guanine residues 5′ to the lesion and (b) that lesions can enhance slippage even after the lesion has been bypassed in an error-free manner (LAMBERT, NAPOLITANO and FUCHS 1992), provided strong evidence

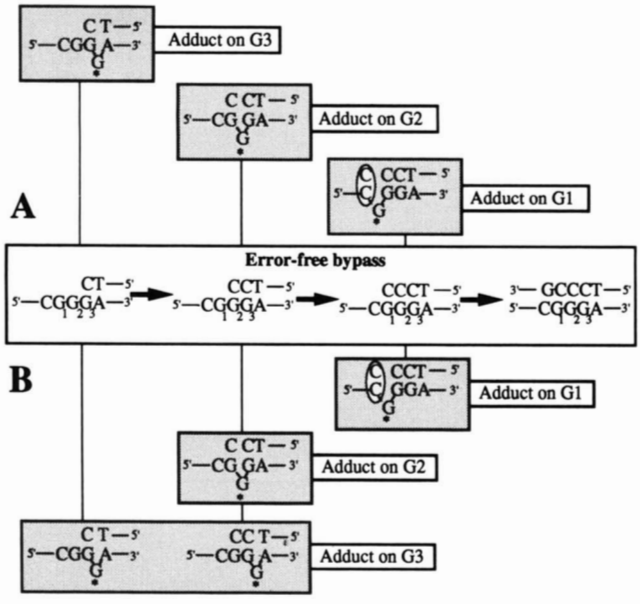


FIGURE 2.—Models for frameshift mutation in contiguous sequences. Error-free bypass of the lesion is shown in the open box. (**A**) The consequences of slippage occurring immediately subsequent to accurate insertion opposite the adduct. The intermediates derived from adducts on $G_2$ and $G_3$ are stabilized by the formation of a normal G·C base pair between the terminal cytosine on the nascent strand and the guanine 5′ to the adduct on the template strand. The mismatched terminus resulting from slippage following incorporation opposite an adduct on $G_1$ is circled. (**B**) The mutagenic intermediates formed if slippage can occur both immediately following incorporation opposite the adduct, and also after the polymerase has successfully bypassed the lesion. The increased number of slipped intermediates and the increased stability of intermediates containing more than one normal terminal G·C base pair might account for the fact that the frequency of frameshift mutation increases with the length of the reiterated sequence.

for the mechanism depicted in Figure 2B. This latter mechanism is entirely consistent with the specificity of 1,8-NONP as shown in Table 7. We therefore postulate that 1,8-NONP adducts promote slippage immediately following insertion of cytosine opposite the lesion, or at positions 5′ to the lesion on the template strand. The number of potential misaligned intermediates derived from slippage when an adduct is situated at any given position within the reiterated sequence increases with the length of the homopolymeric run 5′ to the adduct; in addition, the stability of the intermediates might differ depending on the number of normal base pairs at the primer/template terminus.

Minus (G·C) frameshifts at $(G)_1$ and $(G)_2$ sites occur preferentially at those sites which are flanked on the 5′-end by thymine. Such specificity might be explained by misincorporation of adenine opposite a 5′-T**G***-3′ lesion (**G*** is the adduct) followed by strand slippage yielding a stable frameshift intermediate containing a T·A template/primer terminus (see also Figure 3D). Indeed, the specificity of base substitution indicates that adenine can be inserted opposite the lesion during DNA synthesis, and recent studies of *in vitro* DNA synthesis have demonstrated that frameshifting can occur as a consequence of nucleotide misincorporation (BEBENEK and KUNKEL 1990).

The efficiency with which frameshift mutations form might be determined in part by the ability of flanking bases to accommodate extrahelical bases. Favorable intrastrand stacking of flanking purine bases might favor sequences such as R**G***R (where R is a purine, **G*** is the adduct). In this context it is notable that guanine sequences with a 3′-flanking adenine represent favored frameshift sites.

Although very few 1,8-NONP-induced frameshift mutations were observed at $(G)_1$ sites, we noted a novel sequence context effect at positions 751 [5′-**CCAC**TGC(G)ATGCTG-3′] (LAMBERT *et al.* 1991) and 927 [5′-AACCA(G)C**GTGG**AC-3′] (the deleted G is shown in parentheses). Each of these sites exhibit the following characteristics: (1) the mutant sequences exhibit perfect palindromy (underlined), (2) the G·C base pair which is deleted is one base removed from the axis of symmetry of the palindrome and (3) each site is adjacent to a 5′-GTGG-3′ or 5′-CCAC-3′ sequence (bold type), a motif which is frequently observed at sites of deletion and frameshift mutation (FIX, BURNS and GLICKMAN 1987; HALLIDAY and GLICKMAN 1991; MO, MAKI and SEKIGUCHI 1991).

The addition of a G·C base pair occurred at three sites: positions 90–92 (5′-ACGC**GGG**AAAC-3′), 132–134 (5′-ACGC**GGG**AAAA-3′), and 604 (5′-CCGA**G**ACAG-3′). Interestingly, the sites at position 90–92 and 132–134 share 10 bases of homology a: the site where the +1 mutation occurred, and in each case the guanine run to which a base was added (underlined) is flanked on the 3′ side by a run of adenine residues. This might be related to the increased propensity of such sequences to support slippage intermediates.

Minus 2 frameshift mutations resulting from the loss of GC·CG occurred at three different sites. Two of these sites contain alternating GC sequences: positions 250–254 (5′-TGCA**CGCGC**CGT-3′) and 574–578 (5′AAT**CGCGC**TGTT-3′). Another site of 1,8-NONP-induced −(GC·CG) mutation, position 790–795 (5′-CAAT**GCGCGC**CATT-3′) was identified previously (LAMBERT *et al.* 1991). Mutations at these sites could occur through a slippage-incorporation mechanism involving slippage of the GC repeat. Other laboratories have postulated that −2 frameshifts may be generated through mechanisms other than strand slippage. First, RIPLEY, CLARK and DEBOER (1986) observed that, in the rIIB gene of bacteriophage T4, a large proportion of −2 frameshifts occurred at sites immediately adjacent to potential palindromes. We note that sequences 108–158 and 252–274 of the *lacI* gene, which are predicted (ZUKER and STEIGLER 1981)

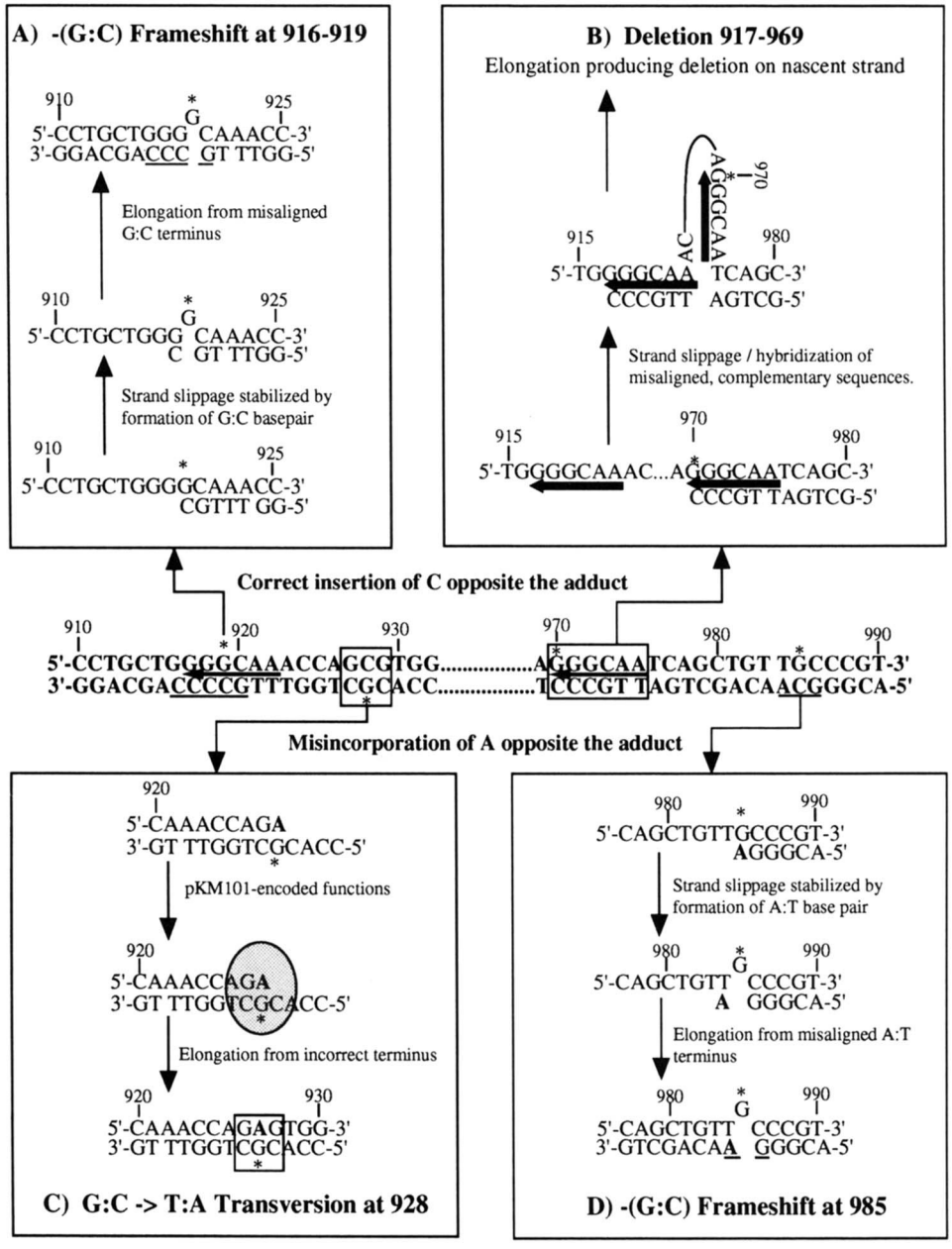


FIGURE 3.—The influence of DNA sequence on mutations induced by dG-C(8)-ANP adducts. We use as examples different types of mutations recovered in sequence 910–990. (**A**) Correct incorporation of cytosine opposite an adduct at position 919 followed by strand slippage could produce a misaligned intermediate which would be stabilized by the formation of normal G·C base pairs at the template/primer terminus. Elongation of such a terminus would produce a −(G·C) frameshift at position 916–919. According to the incorporation-slippage mechanism depicted in Figure 2 such a mutation could also result from an adduct formed at position 918 or 917, but not at position 916. (**B**) Deletion endpoints contain homologous G·C-rich sequences. Deletion 917–969 might be initiated by the correct insertion of cytosine opposite a dG-C(8)-ANP adduct at positions 970. This would create a 5′-TTGCCC-3′ terminus on the nascent strand which would be complementary to the sequence at 917–922. Strand slippage would produce a misaligned intermediate which could be stabilized by base pairing between the complementary sequences. Elongation of such an intermediate would lead to the observed deletion. (**C**) Misincorporation of adenine opposite a dG-C(8)-ANP adduct at position 928 would produce an incorrect G(adduct)·A terminus. The probability that this terminus would be elongated to yield a detectable base substitution mutation would be increased markedly in the presence of pKM101-encoded MucA/B. The nature of the bases flanking the adduct does not appear to influence the site specificity of the base substitutions. (**D**) The observed preference for −(G·C) frameshifts to occur at $(G)_1$ or $(G)_2$ sequences with a 5′-T residue might result from misincorporation of adenine opposite the dG-C(8)-ANP adduct followed by strand slippage. The thymine residue 5′ to the adduct on the template strand could form a normal A·T base pair with the nascent adenine on the primer strand thus stabilizing the misaligned structure.

to be capable of forming stable hairpin structures, are present immediately adjacent to the −2 mutations observed at positions 160–161 and 250–253, respectively. Second, FUCHS and co-workers have postulated that AAF adduction at dG-C(8) induces an altered conformation within alternating G·C sequences which

is processed to a −(GC·CG) mutation though a SOS-inducible mutagenic pathway involving neither RecA nor UmuC/D (BURNOUF, KOEHL and FUCHS 1989; FREUND, BICHARA and FUCHS 1989). In the *lacI* gene, the ratio of −(G·C) to −(GC·CG) frameshift mutations recovered following AAF treatment was 1:3 (SCHAAPER, KOFFEL-SCHWARTZ and FUCHS 1990). In contrast, in the present study the ratio of −(G·C) to −(GC·CG) frameshift mutations was 27:1. The low proportion of 1,8-NONP-induced frameshifts which are −(GC·CG) events might therefore indicate that the dG-C(8)-ANP adduct is relatively inefficient in inducing conformational changes relevant to −2 frameshifts.

**Deletions:** The frequency of deletions was increased by 1,8-NONP treatment in CM6114 (*ΔuvrB*, pKM101), providing the first evidence that nitropyrenes might stimulate deletion formation. The majority of the deletions which have been characterized contain significant homology within the deletion endpoints. These observations are consistent with a model in which slippage of the primer strand between two direct repeats on the template strand during replication results in the deletion of one copy of the repeat plus the intervening sequence (SINGER and WESTLYE 1988). In the collection of deletions obtained following 1,8-NONP treatment there is a notable degree of G·C richness within the repeat. G·C rich sequences would offer a relatively favorable target for 1,8-NONP DNA adduct formation. Thus, induction of deletions might result from the dG-C(8)-ANP adduct blocking DNA synthesis within the repeats, allowing an increased opportunity for misalignment (Figure 3B).

**Conclusions:** In this paper we have discussed several types of mutations which could result from attempted DNA synthesis past a dG-C(8) adduct. We postulate that several classes of 1,8-NONP-induced mutation might be derived from resolution of a replication intermediate containing a stalled DNA polymerase at the site of an adduct (Figure 3). Insertion of the correct base opposite the adduct followed by slippage of the newly synthesized strand could be stabilized by complementary sequences on the template, producing frameshift mutations and deletions. Incorporation of a wrong base without slippage would yield a base substitution mutation in the presence of certain cellular functions. Similarly, misincorporation opposite the adduct followed by strand slippage might account for some frameshift mutations. Our studies identify two principal factors which influence the mutagenic pathway through which such an intermediate is resolved. First, the presence of the replication accessory proteins encoded by *mucA/B*, which are believed to promote elongation from incorrect primer/template termini, strongly increases the probability that a base substitution will result from mispaired intermediates. Second, the extreme site specificity of frameshift mutations suggests that the precise sequence context of the DNA adduct is critical since this determines whether strand slippage can yield misaligned intermediates containing correctly paired primer/template termini.

We thank ERIC EISENSTADT for bacterial strains and ROBERT FUCHS for his comments on this manuscript. This work was supported by grants from the National Cancer Institute of Canada (D.R.M.) and the Natural Science and Engineering Research Council of Canada (B.W.G.). I.B.L. was supported by an Ontario Graduate Scholarship.

## LITERATURE CITED

ANDREWS, P. J., M. A. QUILLIAM, B. E. MCCARRY, D. W. BRYANT and D. R. MCCALLA, 1986 Identification of the DNA adduct formed by metabolism of 1,8-dinitropyrene in *Salmonella typhimurium*. Carcinogenesis **7:** 105–110.

BEBENEK, K., and T. A. KUNKEL, 1990 Frameshift errors initiated by nucleotide misincorporation. Proc. Natl. Acad. Sci. USA **87:** 4946–4950.

BENNETT, C. B., X. LUO and M. Z. HUMAYUN, 1991 Genetic requirements for frameshift reversion induced by bulky DNA adducts in M13 DNA. Mutat. Res. **249:** 19–27.

BRIDGES, B. A., and R. WOODGATE, 1985 Mutagenic repair in *Escherichia coli*: products of the *recA* gene and the *umuD* and *umuC* genes act at different steps in UV-induced mutagenesis. Proc. Natl. Acad. Sci. USA **82:** 4193–4197.

BROWN, T., W. N. HUNTER, G. KNEALE and O. KENNARD, 1986 Molecular structure of the G:A base pair in DNA and its implications for the mechanism of transversion mutations. Proc. Natl. Acad. Sci. USA **83:** 2402–2406.

BROYDE, S., B. E. HINGERTY, R. SHAPIRO and D. NORMAN, 1990 Unusual hydrogen bonding patterns in 2-aminofluorene (AF) and 2-acetylaminofluorene (AAF) modified DNA, pp. 113–123 in *Nitroarenes*, edited by P. C. HOWARD, S. S. HECHT and F. A. BELAND. Plenum Press, New York.

BURNOUF, D., P. KOEHL and R. P. P. FUCHS, 1989 Single adduct mutagenesis: strong effect of the position of a single acetylaminofluorene adduct within a mutation hotspot. Proc. Natl. Acad. Sci. USA **86:** 4147–4151.

CALOS, M. C., and J. H. MILLER, 1981 Genetic and sequence analysis of frameshift mutation induced by ICR-191. J. Mol. Biol. **153:** 39–66.

COULONDRE, C., and J. H. MILLER, 1977 Genetic studies of the *lac* repressor. III. Additional correlation of mutational sites with specific amino acid residues. J. Mol. Biol. **117:** 525–567.

DEVEREUX, J., P. HAEBERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

EISENSTADT, E., 1987 Analysis of mutagenesis, pp. 1016–1033 in *Escherichia coli and Salmonella typhimurium*, edited by F. C. NEIDHARDT, J. L. INGRAHAM, K. B. LOW, B. MAGASANIK, M. SCHAECHTER and H. E. UMBARGER. American Society for Microbiology, Washington, D.C.

EVANS, F. E., D. W. MILLER and F. A. BELAND, 1980 Sensitivity of the conformation of deoxyguanosine to binding at the C-8 position of *N*-acetylated and unacetylated 2-aminofluorene. Carcinogenesis **1:** 955–959.

FARABAUGH, P. J., 1978 Sequence of the *lacI* gene. Nature **274:** 765–769.

FARABAUGH, P. J., U. SCHMEISSNER, M. HOFER and J. H. MILLER, 1978 Genetic studies of the *lac* repressor. VII. On the molec-

ular nature of the spontaneous hotspots in the *lacI* gene of *Escherichia coli*. J. Mol. Biol. **126:** 847–863.

FIX, D. F., P. A BURNS and B. W. GLICKMAN, 1987 DNA sequence analysis of spontaneous mutation in a *polA1* strain of *Escherichia coli* indicates sequence specific effects. Mol. Gen. Genet. **207:** 267–272.

FOSTER, P. L., J. D. GROOPMAN and E. EISENSTADT, 1988 Induction of base substitution mutations by aflatoxin B1 is MucAB dependent in *Escherichia coli*. J. Bacteriol. **170:** 3415–3420.

FREUND, A.-M., M. BICHARA and R. P. P. FUCHS, 1989 Z-forming sequences are spontaneous deletion hotspots. Proc. Natl. Acad. Sci. USA. **86:** 7465–7469.

GLICKMAN, B. W., and L. S. RIPLEY, 1984 Structural intermediates of deletion mutagenesis: a role for palindromic DNA. Proc. Natl. Acad. Sci. USA **81:** 512–516.

GORDON, A. J. E., P. A. BURNS, D. F. FIX, F. YATAGAI, F. ALLEN, M. J. HORSFALL, J. A. HALLIDAY, J. GRAY, C. BERNELOT-MOENS and B. W. GLICKMAN, 1988 Missense mutation in the *lacI* gene of *Escherichia coli*: inferences on the structure of the repressor protein. J. Mol. Biol. **200:** 239–251.

HALLIDAY, J. A., and B. W. GLICKMAN, 1991 Mechanisms of spontaneous mutation in DNA repair-proficient *Escherichia coli*. Mutat. Res. **250:** 55–71.

HALLIDAY, J. A., M. ZIELINSKA, S. S. AWADALLAH and B. W. GLICKMAN, 1990 Colony hybridization in *Escherichia coli*: a rapid procedure for determining the distribution of specific classes of mutations among a number of preselected sites. Environ. Mol. Mutagen. **16:** 143–148.

KAN, L. S., S. CHANDRASEGARAN, S. M. PULFORD and P. S. MILLER, 1983 Detection of a guanine.adenine base pair in a decadeoxyribonucleotide by proton magnetic resonance spectroscopy. Proc. Natl. Acad. Sci. USA **80:** 4263–4265.

KOFFEL-SCHWARTZ, N., J.-M. VERDIER, M. BICHARA, A.-M. FREUND, M. P. DAUNE and R. P. P. FUCHS, 1984 Carcinogen-induced mutation spectrum in wild-type, *uvrA* and *umuC* strains of *Escherichia coli*. J. Mol. Biol. **177:** 33–51.

KUNKEL, T. A., 1984 Mutational specificity of depurination. Proc. Natl. Acad. Sci. **81:** 1494–1498.

KUNKEL, T. A., and A. SONI, 1988 Mutagenesis by transient misalignment. J. Biol. Chem. **263:** 14784–14789.

LAHUE, R. S., K. G. AU and P. MODRICH, 1989 DNA mismatch correction in a defined system. Science **245:** 160–164.

LAMBERT, I. B., R. L. NAPOLITANO and R. P. P. FUCHS, 1992 Carcinogen-induced frameshift mutagenesis in repetitive sequences. Proc. Natl. Acad. Sci. USA **89:** 1310–1314.

LAMBERT, I. B., A. J. E. GORDON, T. A. CHIN, D. W. BRYANT, B. W. GLICKMAN and D. R. MCCALLA, 1990 Mutations induced in the *lacI* gene of *E.coli* by 1-nitroso-8-nitropyrene and furylfuramide: the influence of plasmid pKM101 and excision repair on the mutational spectrum, pp. 167–180 in *Nitroarenes* edited by P. C. HOWARD, S. S. HECHT and F. A. BELAND. Plenum Press, New York.

LAMBERT, I. B., A. J. E. GORDON, D. W. BRYANT, B. W. GLICKMAN and D. R. MCCALLA, 1991 The action of 1-nitroso-8-nitropyrene in *Escherichia coli*: DNA adduct formation and mutational consequences in the absence of nucleotide excision-repair. Carcinogenesis **12:** 879–884.

LOECHLER, E. L., 1989 Adduct-induced base-shifts: a mechanism by which the adducts of bulky carcinogens might induce mutations. Bioploymers **28:** 909–927.

MCCOY, E. C., M. HOLLOWAY, M. FRIERSON, G. KLOPMAN, R. MERMELSTEIN and H. S. ROSENKRANZ, 1985 Genetic and quantum chemical basis of the mutagenicity of nitroarenes for adenine-thymine base pairs. Mutat. Res. **149:** 311–319.

MICHAELS, M. L., T. M. REID, C. M. KING and L. J. ROMANO, 1991 Accurate *in vitro* translesion synthesis by *Escherichia coli* DNA polymerase I (large fragment) on a site-specific, amino-fluorene-modified oligonucleotide. Carcinogenesis **12:** 1641–1646.

MILLER, J. H., 1972 *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MILLER, J. H., 1978 The *lacI* gene: its role in *lac* operon control and its use as a genetic system, pp. 31–88 in *The Operon* edited by J. H. MILLER and W. S REZNIKOFF. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MILLER, J. H., C. COULONDRE and P. J. FARABAUGH, 1978 Correlation of nonsense sites in the *lacI* gene with specific codons in the nucleotide sequence. Nature **274:** 770–775.

MO, J.-Y., H. MAKI and M. SEKIGUCHI, 1991 Mutational specificity of the *dnaE173* mutator associated with a defect in the catalytic subunit of DNA polymerase III of *Escherichia coli*. J. Mol. Biol. **222:** 925–936.

NORMAN, C. A., I. B. LAMBERT, L. M. DAVISON, D. W. BRYANT and D. R. MCCALLA, 1989a DNA adduct formation in primary rat tracheal epithelial cells following treatment with 1,8-dinitropyrene and its partially reduced derivative, 1-nitroso-8-nitropyrene. Carcinogenesis **10:** 1323–1327.

NORMAN, D., P. ABUAF, B. E. HINGERTY, D. LIVE, D. GRUNBERGER, S. BROYDE and D. PATEL, 1989b NMR and computational characterization of the *N*-(deoxyguanosin-8-yl)aminofluorene adduct [(AF)G] opposite adenosine in DNA: (AF)G[syn]-A[anti] pair formation and its pH dependence. Biochemistry **28:** 7462–7476.

NORMAN, C. A., I. B. LAMBERT, L. M. DAVISON, D. W. BRYANT and D. R. MCCALLA, 1990 Formation and persistence of DNA adducts in rats following intraperitoneal administration of 1,8-dinitropyrene. Carcinogenesis **11:** 1037–1040.

QUIGLEY, G. J., G. UGHETTO, G. A. VAN DER MAREL, J. H. VAN BOOM and A. RICH, 1986 Non Watson-Crick G·C and A·T base pairs in a DNA-antibiotic complex. Science **232:** 1255–1258.

RABKIN, S. D., and B. S. STRAUSS, 1984 A role for DNA polymerase in the the specificity of nucleotide incorporation opposite *N*-acetyl-2-aminofluorene adducts. J. Mol. Biol. **178:** 569–594.

REFOLO, L. M., C. B. BENNETT and M. Z. HUMAYUN, 1987 Mechanisms of frameshift mutagenesis by aflatoxin $B_1$-2,3-dichloride J. Mol. Biol. **193:** 609–636.

REZNIKOFF, W. S. and J. N. ABELSON, 1978 The *lac* promotor, pp. 221–243 in *The Operon* edited by J. H. Miller and W. S Reznikoff. Cold Spring Harbor Laboratory, New York.

RIPLEY, L. S., A. CLARK and J. G. DEBOER, 1986 Spectrum of spontaneous frameshift mutations. Sequences of bacteriophage T4 rII gene frameshifts. J. Mol. Biol. **191:** 601–613.

ROSENKRANZ, H. S., and R. MERMELSTEIN, 1983 Mutagenicity and genotoxiicty of nitroarenes: all nitro-containing chemicals were not created equal. Mutat. Res. **101:** 217–267.

SANGER, F., A. R. COULSON, B. J. BARREL, A. J. H. SMITH and B. A. ROE, 1980 Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143:** 161–178.

SCHAAPER, R. M., B. N. DANFORTH and B. W. GLICKMAN, 1985 Rapid repeated cloning of mutant *lac* repressor genes. Gene **39:** 181–189.

SCHAAPER, R. M., B. N. DANFORTH and B. W. GLICKMAN, 1986 Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli lacI* gene. J. Mol. Biol. **189:** 273–284.

SCHAAPER, R. M., N. KOFFEL-SCHWARTZ and R. P. P. FUCHS, 1990 *N*-acetoxy-*N*-acetyl-2-aminofluorene-induced mutagenesis in the *lacI* gene of *Escherichia coli*. Carcinogenesis **11:** 1087–1095.

SCHMEISSNER, U., D. GANEM and J. H. MILLER, 1977 Genetic studies of the *lac* repressor. II. Fine structure deletion map of the *lacI* gene, and its correlation with the physical map. J. Mol. Biol. **109:** 303–326.

SELBY, C. P., E. M. WITKIN and A. SANCAR, 1991 *Escherichia coli mfd* mutant deficient in "mutation frequency decline" lacks strand-specific repair: *in vitro* complementation with purified coupling factor. Proc. Natl. Acad. Sci. USA **88:** 11574–11578.

SHIBA, T., H. IWASAKI, A. NAKATA and H. SHINAGAWA, 1990 Proteolytic processing of MucA protein in SOS mutagenesis: both processed and unprocessed MucA may be active in mutagenesis. Mol. Gen. Genet. **224:** 169–176.

SINGER, B. S., and J. WESTLYE, 1988 Deletion formation in bacteriophage T4. J. Mol. Biol. **202:** 233–243.

SOWERS, L. C., B. RAMSEY SHAW, M. L. VEIGL and W. D. SEDWICK, 1987 DNA base modification: ionized base pairs and mutagenesis. Mutat. Res. **177:** 201–218.

STRAUSS, B. S., 1991 The 'A rule' of mutagen specificity: a consequence of DNA polymerase bypass of non-instructional lesions? Bioessays **13:** 79–84.

STREISINGER, G., and J. OWEN, 1985 Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4. Genetics **109:** 633–659.

STREISINGER, G., Y. OKADA, J. EMRICH, J. NEWTON, A. TSUGITA, E. TERZAGHI and M. INOUYE, 1966 Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:** 77–84.

TERLETH, C., P. VAN DE PUTTE and J. BROUWER, 1991 New insights in DNA repair: preferential repair of transcriptionally active DNA. Mutagenesis **6:** 103–111.

TOKIWA, H., and Y. OHNISHI, 1986 Mutagenicity and carcinogenicity of nitroarenes and their sources in the environment. CRC Crit. Rev. Toxicol. **17:** 23–60.

WALKER, G. C., 1984 Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. Microbiol. Rev. **48:** 60–93.

WESTON-HAFER, K., and D. E. BERG, 1989 Palindromy and the location of deletion endpoints in *Escherichia coli*. Genetics **121:** 651–658.

ZUKER, M., and P. STEIGLER, 1981 Optimal computer folding of large RNA sequences using thermodynamics and auxillary information. Nucleic Acids Res. **9:** 133–148.

Communicating editor: N. R. DRINKWATER