Influence of Neighboring Base Sequence on the Distribution and Repair of N-Ethyl-N-nitrosourea-induced Lesions in _Escherichia coli_

Philip A. Burns, Alasdair J. E. Gordon, Kathy Kunsmann, and Barry W. Glickman

Department of Biology, York University, Toronto, Ontario M3J 1P3, Canada

ABSTRACT

N-Ethyl-N-nitrosourea-induced mutations occurring within a 180-base pair target in the _lac_ gene of _Escherichia coli_ were characterized by DNA sequencing. In total, 109 mutations were characterized in a wild-type background and 180 in an excision-repair-deficient (UvrB*) background. The majority of mutations induced in the two backgrounds (77% and 85%, respectively) were G:C = > A:T transitions, presumably resulting from miscoding O6-ethylguanine lesions. A significant proportion of the mutations (17% and 15%, respectively) were A:T = > G:C transitions, which probably result from miscoding O6-ethylthymine lesions. An analysis of the distribution of both types of mutation in the two backgrounds reveals two distinct influences of neighboring base sequence. These effects apply equally to both the G:C = > A:T and A:T = > G:C transitions. Firstly, miscoding lesions are most likely to occur at 5'-purine-G-3'/5'-purine-T-3' sites. Secondly, the excision-repair machinery is less efficient at removing both O6-ethylguanine and O6-ethylthymine lesions which are flanked on both sides by G:C base pairs. Thus, in the wild-type spectrum an overabundance of transitions occurs at a 5'-G-G/G/C-3'/5'-G-T/G/C-3' sequence (where the mutated base is underlined).

INTRODUCTION

One approach to the investigation of mutational mechanisms is to determine the nature of genetic alterations at the DNA sequence level. Such an analysis can provide information on the relative contribution of different premutational lesions and neighboring base sequence to the mutational process. Recent studies of the mutational specificity of the alkylating agents MNNG and EMS have revealed important influences of neighboring base sequence on the final distribution of mutations and the determination of mutational hotspots (1, 2). MNNG induces almost exclusively G:C = > A:T transitions; and, as these tend to occur at 5'-R-G-3' sites, this specificity is thought to reflect an influence by the preceding purine base on the reactivity of the O6 position of guanine. In contrast, EMS which also induces primarily G:C = > A:T transitions, does not display a specificity for 5'-R-G-3' sites. However, the distribution of such mutation is influenced by the UvrABC excision-repair system, which can remove O6-ethylguanine lesions from DNA (3-5); O6-ethylguanine lesions which are adjacent to A:T base pairs are more efficiently removed by the excision repair machinery than those which are flanked by G:C base pairs (2). Thus, in an excision-proficient strain EMS-induced mutations tend to occur at 5'-G/G/C-G/C-3' sites.

It was of interest to learn whether these two distinct influences on the distribution of alkylated-induction mutant could both be observed with an ethylating agent which is more closely related, chemically and mechanistically, to MNNG and MNU than EMS. We chose the carcinogenic alkylating agent ENU for this study. This agent also has the advantage of enabling us to examine these influences on the mutagenicity of the O4-EtT lesion as well. Here we report the mutational specificity of ENU in the _lac_ gene in both an excision-repair-proficient (wild-type) and excision-repair-deficient (UvrB*) strain of _Escherichia coli_.

MATERIALS AND METHODS

Strains and Media. _E. coli_ strains NR3835: F' lac-pro; _Δ(pro-lac ara th_ trpE9777; and NR3951: _Δ(bio-uvrB)_ derivative of NR3835 were used in the collection of _lac_ mutants. Strains used for mapping and cloning of mutants have been described (6, 7). Media were as described by Miller (8) and Coulondre and Miller (6).

Treatment. The treatment and selection procedures were as described previously (1, 2). Five independent cultures of each strain were grown to mid-log phase in LB broth, spun down, washed, and resuspended in Vogel-Bonner salt solution. Treatment with ENU (Sigma), dissolved in dimethyl sulphoxide, was at a concentration of 30 mM for 20 min at 37°C with occasional shaking. After washing, _lac_ mutants were selected by directly plating the treated cells onto agar plates containing phenyl-β-d-galactosidase (Research Organics, Inc.) as the sole carbon source. The exposure in buffer, in conjunction with the direct plating onto selective media immediately after washing, insures the independence of each mutant. Appropriate dilutions of treated cells were also plated on LB plates to determine survival. Phenyl-β-d-galactosidase is a substrate for β-galactosidase but does not induce synthesis of the enzyme. Hence when supplied as the sole carbon source, only cells that synthesize β-galactosidase constitutively (_lac_ or _lac_2 mutations) form colonies. Mutant colonies were isolated from each strain; mutations occurring within the initial 180 base pairs of the _lac_ gene were identified using a simple complementation test (8).

Cloning and Sequencing. The _lac_ mutants were cloned from the _F_ plasmid by _in vivo_ recombination onto a specifically constructed M13 bacteriophage strain according to Schaper et al. (7). Single stranded phage DNA was sequenced using the dideoxynucleotide chain termination method (9), from a specific 14-base primer which anneals at positions 302–315 in the _lac_ sequence (P-L Biochemicals).

RESULTS

Mutagenesis in the 180-base pair target region of the _lac_ gene has been well characterized. This region encodes the DNA-binding domain of the repressor protein, and appears to be particularly sensitive to base substitution mutation. From a data base of over 6000 sequence-characterized _lac_ mutations generated within this laboratory, a total of 106 different base substitutions have been characterized in this sequence (10). These changes involve 39 of the 60 codons and roughly equal proportions of the six possible base substitution classes. This extensive data base allows us to consider sites at which mutations were not recovered following ENU exposure.

Treatment of the wild-type strain resulted in a Lac- mutation frequency of 2.7 × 10^-4 at 13% survival compared with a mutation frequency of 1.5 × 10^-3 at 15% survival in the UvrB* strain (averages of five independent treatments). This represents a 135-fold increase over spontaneous mutation frequency in the
ENU MUTATIONAL SPECIFICITY IN E. coli

The same treatment thus yields a 5- to 6-fold enhancement of mutation in the UvrB" strain over that seen in the wild-type strain.

A total of 400 mutants were picked from five independently treated cultures of each strain and subjected to a simple complementation test (see Reference 8) which identifies mutations lying within the initial 180 base pairs (I~d dominant-negative mutations). In the case of the wild-type strain, 22 I~d mutants were chosen randomly from each of the five cultures and characterized. Of these 110 mutants, one was lost during the course of the study. For the UvrB" strain, 100 I~d mutants were chosen in a similar fashion and characterized further. The classification, numbers, and local DNA sequence of 109 and 100 mutations induced by ENU in a wild-type and UvrB" background, respectively, are presented in Table 1. The distribution of ENU-induced mutation in the two backgrounds is presented in Fig. 1.

In the wild-type strain, the majority of mutations were G:C = >A:T transitions, accounting for 77% (84/109) of the total. These changes were distributed over 18 of the 23 sites known to produce a selectable LacI" phenotype via this transition. The mutations were not randomly distributed over these sites. For example, one site (position 185) accounted for 20% of the G:C = >A:T transitions. The next most frequently recovered class of mutation was the A:T = >G:C transition, which made up 17% (19/109) of the total. Mutations were recovered at seven of the 17 known sites of A:T = >G:C transition within this target sequence. Two of the sites (positions 54 and 83) were hotspots for mutation, accounting for over 60% (12/19) of this class. The remaining 6% (6/109) of the mutations was comprised of five transversions and one 15-base pair deletion.

In the UvrB" spectrum, all the characterized mutations were transitions. The majority were G:C = >A:T transitions, accounting for 85% (85/100) of the total. These changes were distributed over 17 of the 23 known potential sites of G:C = >A:T transition within the target sequence. There was considerable site to site variation in numbers of mutations, with two of the sites (positions 90 and 120) contributing 25% (25/100) of the total spectrum. The site specificity of mutation within the UvrB" spectrum did not appear to correlate with that found in the wild-type spectrum; of the five most mutable sites from each spectrum only two were in common (positions 90 and 185). The remaining 15% (15/100) of the mutations in the UvrB" spectrum were A:T = >G:C transitions, distributed over eight of the 17 known potential sites within the target sequence. In contrast to the wild-type spectrum, there were no obvious hotspots for A:T = >G:C transition in the UvrB" spectrum.

DISCUSSION

We have sequenced 209 ENU-induced mutations in the lacI gene recovered in an excision-repair-proficient (wild-type) and an excision-repair-deficient (UvrB" strain). In the wild-type background, most of the mutations (94%) were transitions, with G:C = >A:T favored over A:T = >G:C by a factor of 17 to 1. In contrast to the wild-type spectrum, there were no obvious hotspots for A:T = >G:C transition in the UvrB" spectrum.
of 4.4:1 (84:19). This ratio correlates well with the ratio of \(O^6\)-EtG to \(O^4\)-EtT lesions induced by ENU in \textit{E. coli} (11) and supports the model that G:C = \(\rightarrow\)A:T and A:T = \(\rightarrow\)G:C transitions arise as a consequence of the misrepairing properties of the \(O^6\)-EtG and \(O^4\)-EtT lesions, respectively (12–15). This observation confirms and extends the findings of Richardson et al. (11), who studied the mutational consequences of ENU in excision-repair-proficient \textit{E. coli} using the \textit{gpt} gene as the genetic target.

In the UvrB\(^{-}\) strain the mutation frequency increased 5-fold over the wild-type frequency, and the spectrum of mutations was composed exclusively of transition mutations. These observations are consistent with previous studies on mutation induction by ethylating agents (2, 4, 16) and support the proposal that the UvrABC excision repair complex is involved in the removal of the misrepairing lesions \(O^6\)-EtG and \(O^4\)-EtT from the DNA (3, 4). In the UvrB\(^{-}\) strain the ratio of G:C = \(\rightarrow\)A:T transitions to A:T = \(\rightarrow\)G:C transitions increases to 5.7:1 compared to 4.4:1 in the wild-type strain. Although this difference is not statistically significant \((P > 0.40)\), it may reflect a slight bias for the repair of \(O^6\)-EtG lesions over \(O^4\)-EtT by the excision repair complex.

The transversions recovered in the wild-type strain may be the result of error-prone SOS bypass of such noncoding lesions as \(N^1\)- and \(N^7\)-ethyalted purines and abasic sites (17, 18). The absence of transversions in the UvrB\(^{-}\) spectrum may reflect the overabundance of the \(O^6\)-EtG and \(O^4\)-EtT lesions which swamp the transversion based events.

A single 15-base pair deletion was recovered in the wild-type strain. Its origin remains unclear. It is not simply a consequence of directly repeated or palindromic sequences, a feature characteristic of deletions found in a spontaneous mutation spectra in the \textit{lacI} gene (19). This identical deletion was recovered, independently, three times in a similar-sized spectrum induced by MNU in the same strain,\(^5\) suggesting it may however be a consequence of attempted repair of alkylation damage. As has been suggested earlier, attempts at DNA repair involving the production of a strand break may lead to frame-shift events (20, 21).

From previous studies with EMS (2) and MNNG (1) we demonstrated that neighboring base sequence can influence the final distribution of mutation at two separate levels: \((a)\) at the level of repair and \((b)\) at the level of initial interaction with the DNA. The ethylating agent EMS induced almost exclusively \((97\%)\) G:C = \(\rightarrow\)A:T transitions and an analysis of the distribution recovered in a wild-type strain revealed a bias in favor of mutation at G:C base pairs which are flanked on both sides by G:C base pairs. This bias, however, disappeared in a UvrB\(^{-}\) strain and we concluded that the flanking bases influence the repairability of \(O^6\)-EtG lesions by the UvrABC excision complex.

In Table 2A we analyze the distribution of ENU-induced G:C = \(\rightarrow\)A:T transitions with respect to flanking base sequence. We classified the mutations among the 23 known available sites into those which are flanked by an A:T base pair on at least one side (total of 15), and sites which are flanked on both sides by G:C base pairs (total of eight). In the UvrB\(^{-}\) background there was no significant difference between the observed and the expected numbers of mutations in these classes. However, in the wild-type background, there was a significant departure \((P < 0.05)\) of the observed numbers from the expected numbers.


An almost twofold bias in mutation at sites flanked by G:C base pairs is observed. This bias, like that seen for EMS (2) reveals a significant influence of the neighboring base sequence on the repairability of ENU-induced \(O^6\)-EtG lesions by the UvrABC excision repair complex.

In Table 2B we analyze the spectrum of A:T = \(\rightarrow\)G:C transitions in an analogous manner. Although the number of mutations involved is small, there appeared to be a significant bias \((P < 0.01)\) in favor of mutation at A:T base pairs flanked on both sides by G:C base pairs in the wild-type spectrum. Again this bias disappeared in the UvrB\(^{-}\) background. It thus appears that like the \(O^6\)-EtG lesion, the \(O^4\)-EtT lesion is significantly more repairable when adjacent to one or more A:T base pairs. The preferential repair of both the ethylated adducts of guanines and thymines may be explained if there is a greater distortion of the regular helical structure by ethylation lesions in the immediate vicinity of relatively weak A:T base pair bonds, hence making them more "apparent" to the excision repair complex. As a consequence, both the \(O^6\)-EtT and \(O^4\)-EtT lesions would be more efficiently repaired in a region flanked by A:T base pairs.

An investigation of MNNG-induced mutagenesis of the \textit{lacI} gene (1) revealed a strong influence by the 5' flanking base on the distribution of mutation. MNNG induced almost exclusively \((98\%)\) G:C = \(\rightarrow\)A:T transitions in \textit{lacI}, and these occurred at an 8-fold higher frequency at G:C base pairs preceded (5') by a purine as opposed to a pyrimidine. The putative alkyllating species of MNNG is thought to be a methyl diazonium ion (22), and the bias in favor of 5'-R-G-3' sites may reflect an influence by the preceding purine on the reactivity of the O\(^6\) position of guanine. This same bias has been observed with MNU in the \textit{gpt} (11) and \textit{lacI}\(^2\) genes of \textit{E. coli}. ENU resembles MNNG and MNU in that it is a nitrosamine which is predicted to require activation to an alkyl diazonium ion before reacting with DNA. We have analyzed the distribution of G:C = \(\rightarrow\)A:T transitions with respect to the 5' flanking base in the UvrB\(^{-}\)

---

### Table 2

#### Comparison of the influence of flanking base pairs on transition mutations in excision proficient and deficient backgrounds

<table>
<thead>
<tr>
<th>Flanking bases (^a)</th>
<th>Number of sites (^b)</th>
<th>Number of transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Expected</td>
</tr>
<tr>
<td>(Uvr^+) background</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3' and/or 5' A:T</td>
<td>15</td>
<td>46</td>
</tr>
<tr>
<td>3' and 5' G:C</td>
<td>8</td>
<td>38</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>84</td>
</tr>
<tr>
<td>(Uvr^B) background</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3' and/or 5' A:T</td>
<td>15</td>
<td>61</td>
</tr>
<tr>
<td>3' and 5' G:C</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>85</td>
</tr>
</tbody>
</table>

\(^a\) Refers to the 3' and 5' flanking base pairs of the mutated base pair.

\(^b\) Includes sites at which no mutations were recovered in this study, but which are known from other studies to produce a Laci' phenotype if mutated. See References 1, 2, and 10.

The \(\chi^2\) is calculated with 1 degree of freedom.
ENU MUTATIONAL SPECIFICITY IN E. coli

Table 3 Influence of the 5′ flanking base on ENU-induced transition mutations in a UvrB− background

<table>
<thead>
<tr>
<th>5′ Base</th>
<th>Number of sites*</th>
<th>Number of mutations</th>
<th>Average per site</th>
</tr>
</thead>
<tbody>
<tr>
<td>G:C to A:T transitions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3</td>
<td>27</td>
<td>9.0</td>
</tr>
<tr>
<td>G</td>
<td>5</td>
<td>32</td>
<td>6.4</td>
</tr>
<tr>
<td>Purine</td>
<td>8</td>
<td>59</td>
<td>7.4</td>
</tr>
<tr>
<td>C</td>
<td>8</td>
<td>12</td>
<td>1.5</td>
</tr>
<tr>
<td>T</td>
<td>7</td>
<td>14</td>
<td>2.0</td>
</tr>
<tr>
<td>Pyrimidine</td>
<td>15</td>
<td>26</td>
<td>1.7</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>85</td>
<td>3.7</td>
</tr>
<tr>
<td>A:T to G:C transitions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purine</td>
<td>8</td>
<td>12</td>
<td>1.5</td>
</tr>
<tr>
<td>Pyrimidine</td>
<td>9</td>
<td>3</td>
<td>0.3</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>15</td>
<td>0.9</td>
</tr>
<tr>
<td>χ²</td>
<td>6.6</td>
<td>&lt;0.05</td>
<td></td>
</tr>
</tbody>
</table>

*Includes sites at which no mutations were recovered in this study, but which are known from other studies to produce a Lac− phenotype if mutated. See References 1, 2, and 10.

A 4.4-fold bias was found in favor of G:C to A:T transitions occurring at 5′-R-G-3′ sites while an 8.5-fold bias was found for A:T = >G:C transitions occurring at 5′-R-G-3′ sites while an 8.5-fold bias was found for A:T = >G:C transitions occurring at 5′-R-T-3′ sites. It thus appears that the final distribution of mutations induced by ENU in the lacI gene of E. coli is heavily influenced by the neighboring base sequence of the potential sites of mutation.

From previous studies of a different nitrosamine (MNNG) and a different ethylating agent (EMS) we were able to identify two separate levels at which sequence context effects could operate in the determination of the ENU-induced spectrum of mutations. From those studies one predicts that 5′-G-G-C-3′ sequences would be hotspots for ENU-induced mutation in a wild-type strain. This was indeed observed. The hottest site for mutation in the wild-type strain was a 5′-G-G-C-3′ sequence (site 185). It was more than twice as mutable as any other site and accounted for over 15% of the total mutations. Similarly, all three known potential 5′-G-G-C-3′ sequences in this target, 174, 185, and 201, are recovered at relatively high frequencies and together account for nearly 30% of all the mutations recovered. In an analogous manner, the A:T = >G:C transition is most frequently recovered at the three potential 5′-G-T-G-C-3′ sequences (positions 54, 83, and 141 in the target sequence) which together account for about 14% of total mutation.

REFERENCES


Downloaded from cancerres.aacrjournals.org on May 28, 2017. © 1988 American Association for Cancer Research.
Influence of Neighboring Base Sequence on the Distribution and Repair of \( N \)-Ethyl-\( N \)-nitrosourea-induced Lesions in \textit{Escherichia coli}

Philip A. Burns, Alasdair J. E. Gordon, Kathy Kunsmann, et al.