Mechanisms of mutagenesis in the *Escherichia coli* mutator *mutD5*: Role of DNA mismatch repair  

( *lacI* gene/conditional mutator/proofreading/transitions and transversions/dislocation mutagenesis)  

**ROEL M. SCHAAPER**  

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709  

Communicated by Franklin W. Stahl, July 20, 1988

ABSTRACT To investigate the mechanisms of spontaneous mutation in the *Escherichia coli* *mutD5* mutator strain, 502 mutations generated in this strain in the N-terminal part of the *lacI* gene were sequenced (i” mutations). Since the mutator strength of this strain depends on the medium in which it grows, mutations were analyzed in both minimal medium (moderate mutator activity) and rich medium (high mutator activity). In either case, 95% of all mutations were base substitutions and 5% were single-base deletions. However, the nature and site distribution of the base substitutions differed dramatically for the two conditions. In minimal medium (mutation frequency, 480-fold above background), 90% of the base substitutions were transitions. These observations suggest that different modes of mutagenesis operate under the two conditions. *mutD5* cells have been reported to be defective in exonucleolytic proofreading during DNA replication. The present data suggest that *mutD* cells in rich medium also suffer a defect in mutHLS-encoded mismatch correction. This hypothesis was confirmed by the direct measurement of mismatch repair in *mutD5* cells by transfection of M13mp2 heteroduplex DNA.

*Escherichia coli* mutator strains are useful tools for studying the mechanisms by which organisms maintain their mutation rates. In wild-type strains, these rates are generally very low (10-10 per base pair per cell division) (1). Most strong mutators affect some aspect of DNA replication (2). At least three major fidelity steps are identified (3, 4): (i) discrimination by the DNA polymerase against insertion of incorrect nucleotides; (ii) exonucleolytic removal of incorrectly inserted nucleotides by the 3'-5' exonuclease associated with the DNA polymerase (editing or proofreading); and (iii) DNA mismatch repair, which recognizes and corrects mismatches shortly after replication. This system, encoded by the *mutH*, *mutL*, and *mutS* genes, can distinguish the correct from the incorrect base in the mismatch based on the undermethylation of the newly synthesized strand (5-8).

*mutD5* is among the strongest mutants in *E. coli*, its mutation frequencies often being elevated 10^4-fold or higher (9, 10). The *mutD* (or *dnaQ*) gene encodes the z-subunit of the Pol III holoenzyme, which carries the 3'-5' proofreading exonuclease of this enzyme (11-15). Pol III from *mutD5* or *dnaQ* strains is known to be defective in 3'-5' exonuclease activity (11). These observations provide a logical basis to account for the *mutD5* mutator phenotype. However, it remains a question whether the 10^4 or more enhancement of mutagenesis in *mutD* cells can be ascribed entirely to a defect in proofreading. Ferst et al. (16) argued that the contribution of proofreading to the fidelity of DNA synthesis by Pol III holoenzyme may be limited to several hundredfolds because of its high cost (increased exonuclease leading to increased hydrolysis of correctly inserted nucleotides). The measured fidelity of the enzyme [10^-5 to 10^-4, depending on the mismatch (16, 17)] and an insertion fidelity (10^-4 to 10^-5) (4) also predict a value significantly smaller than the magnitude of the *mutD* mutator effect. A second feature of *mutD* strains requiring further explanation is that the high mutation frequency of *mutD5* strains is only attained in rich medium (9, 10). In minimal medium, the effect is 100 times less, hinting at the intriguing possibility that proofreading may be regulated (18, 19).

Precise knowledge of the nature and positions of *mutD5*-generated mutations may yield insights into the underlying mechanisms. We therefore undertook an investigation of the specificity of mutation in *mutD5* cells in both minimal and rich medium. The specificity of mutagenesis in the *mutD5* strain in rich medium has been studied before, measuring the reversion of a set of *trpA* alleles (20) or sequencing a number of forward mutations in the *lacI* gene (21). To obtain a more complete spectrum of errors, we now sequenced a total of 502 *mutD5*-induced mutations in the N-terminal part (nucleotides 30-240) of the *lacI* gene (i” mutations), about equally distributed over the two media. Two markedly different spectra were observed. Analysis of these spectra suggested that the strong mutator phenotype of *mutD5* strains results from at least two deficiencies: impaired proofreading and impaired postreplicative mismatch repair. The latter was also demonstrated directly by transfection of heteroduplex DNA.

**MATERIALS AND METHODS**

**Bacterial Strains.** Strain NR9232 (ara, thi, zaf3::Tn10 *mutD5*, *Δprolac*, F' *prolac*), used for selection of *lacI* mutations, was derived from KA796 (ara, thi, *Δprolac*) (22) as described (21). NR9102 is the isogenic *mutA* strain. The F' *prolac* carries the P' (*lacI*) and L8 (lacZ) promoter mutator genes. In addition, it carries two silent base substitutions in the *lacI* gene (positions 622 and 623; unpublished data) that abolish the spontaneous frameshift hot spot (loss or gain of the 4-base sequence TGGC at nucleotides 620–632) (23).

**Mutagenesis Experiments.** LB broth (rich) and glucose minimal medium containing Vogel-Bonner salts (MM) have been described (23), as have the phenylgalactoside (P-Gal) plates used to select *lacI* mutants (23). NR9232 was grown overnight in minimal medium or in LB broth and diluted in fresh medium of the same variety to a concentration of 100 cells per ml. Aliquots (200 μl) of the diluted cultures were distributed to the wells of 96-well microtiter dishes, which were incubated at 37°C for 1 (LB) or 2 (MM) days on a rotary platform. *lacI* mutants were selected by spreading 5 μl from each well (undiluted for MM cultures; diluted 1:10 for LB cultures) on quarter sections of P-Gal plates, which
were incubated for 48 hr at 37°C. Only lacI' colonies grow on P-Gal plates. Mutant frequencies were obtained by plating appropriate dilutions from 40 independent LB and MM cultures on P-Gal and MM plates to determine mutant and total cell count, respectively. To determine the frequencies of rifampicin- and nalidixic acid-resistant cells, aliquots were plated on MM plates containing rifampicin (100 μg/ml) or nalidixic acid (40 μg/ml).

Selection of i-d Mutations and DNA Sequence Analysis. One lacI' mutant per culture was picked randomly and regridded on P-Gal plates. The i-d mutations among them were scored (24) by transfer of their F' into CSH52 (Apr, ara, strA, thi, recA, 880dellac') by replica-mating and replication of the conjugants onto MM plates containing 5-bromo-4-chloro-3-indolyl β-D-galactoside, where lacI diploids with a wild-type mutation form blue patches. They are of two classes: i-d and lacO (operator mutants). lacO hot spot mutants (T → C at position +6 of lacZ) (21, 23) were screened out by a colony-hybridization method (25). The remaining dominant F'lacI' (almost exclusively i-d) were then transferred into strain 90C for transfer of their lacI gene to phage mRS81 for DNA sequencing (22, 23).

Heteroduplex Transfection Experiments. M13mp2 replicative form II heteroduplex DNA (nicked duplex) containing a T-G mismatch at position 90 of lacZo was prepared by a published procedure (26). Phage M13mp2 T90, an M13mp2 (ochre) derivative incapable of α-complementation (27), provided the viral strand; one of its blue plaque revertants [TAA → CAA (17)] provided the complementary strand. mp2T90 was grown in a dam" host, the revertant in a dam" host, providing for a hemimethylated heteroduplex. The DNA was transfected into Hanahan cells (28) and mismatch repair was monitored by directly plating aliquots of the transfection (to produce infective centers) in soft agar layers containing isopropyl β-D-thiogalactoside, 5-bromo-4-chloro-3-indolyl β-D-galactoside, and a complementation strain CSH50, scoring for colorless, blue, or mixed plaques.

RESULTS

mutD5 Mutation Frequencies in Minimal and Rich Media. Representative mutD-induced mutation frequencies for several mutational targets in minimal and rich medium are presented in Table 1. In minimal medium, the mutD5-induced forward mutation frequency to lacI' is enhanced 175-fold. In rich medium, an additional 75-fold increase occurs for an overall 13,000-fold effect.

The Nature of mutD-Induced Mutations. The lacI gene, coding for the repressor of the lac operon, contains >1000 base pairs and therefore the DNA sequencing was limited to its N-terminal region (operator-binding domain, nucleotides 30-240). Mutations in this region are dominant over lacI' +. They are termed i-d and are easily detected among lacI' mutants by a genetic test (24). Dominance results from the tetrameric nature of the repressor: monomers defective in DNA binding still aggregate through their attachment site in the C terminus, but the mixed multimers are inactive. The N

### Table 1. E. coli mutD5 mutant effect in minimal (MM) and rich (LB) media

<table>
<thead>
<tr>
<th>Strain</th>
<th>Medium</th>
<th>lacI'</th>
<th>i-d</th>
<th>rif+</th>
<th>nal'</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR9102 (mut')</td>
<td>LB/MM</td>
<td>0.4*</td>
<td>0.04</td>
<td>0.016</td>
<td>0.012</td>
</tr>
<tr>
<td>NR9232 (mutD5)</td>
<td>MM</td>
<td>69 (175)</td>
<td>19 (480)</td>
<td>2.1 (130)</td>
<td>0.15 (13)</td>
</tr>
<tr>
<td>NR9232 (mutD5)</td>
<td>LB</td>
<td>5100 (13,000)</td>
<td>1500 (37,000)</td>
<td>48 (3000)</td>
<td>53 (4400)</td>
</tr>
</tbody>
</table>

Frequencies are means for 40 independent cultures. Numbers in parentheses indicate the -fold increases over the wild-type frequency. No significant differences were observed for strain NR9102 grown in either MM or LB (results combined).

*This frequency (0.4 ± 0.2) × 10^-3 is ~2.5 times lower than previously observed. The -fold increases for mutD-induced lacI' and i-d mutations may therefore be somewhat overestimated.

### Table 2. The nature of mutD5-induced i-d mutations in minimal and rich media

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Minimal medium</th>
<th>Rich medium</th>
<th>Ratio*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base substitutions</td>
<td>232</td>
<td>245</td>
<td>79</td>
</tr>
<tr>
<td>A→T → G/C</td>
<td>34</td>
<td>133</td>
<td>310</td>
</tr>
<tr>
<td>G+C → A/T</td>
<td>54</td>
<td>85</td>
<td>125</td>
</tr>
<tr>
<td>Total transitions</td>
<td>88</td>
<td>218</td>
<td>197</td>
</tr>
<tr>
<td>A→T → T-A</td>
<td>111</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>A→T → C-G</td>
<td>13</td>
<td>10</td>
<td>61</td>
</tr>
<tr>
<td>G+C → T-A</td>
<td>18</td>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td>G+C → C-G</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total transversions</td>
<td>144</td>
<td>27</td>
<td>14</td>
</tr>
<tr>
<td>Single-base frameshifts</td>
<td>12</td>
<td>12</td>
<td>79</td>
</tr>
<tr>
<td>Deletions†</td>
<td>1</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Total</td>
<td>245</td>
<td>257</td>
<td>79</td>
</tr>
</tbody>
</table>

The 502 mutations were derived from 498 dominant lacI' mutants, a limited number of double mutations being observed (see legend to Fig. 1).

*Ratio of frequencies in rich versus minimal medium based on the i-d frequencies of Table 1, corrected by the relative abundance of each class.

†Deletion of 67 base pairs (nucleotides 130–196).

termus contains a high density of detectable sites and is highly useful for studies of mutational specificity (24). Thus far, 112 different base-substitution pathways have been identified (ref. 24; unpublished data). Deletions, duplications, and frameshifts can be scored as well.

The i-d mutations comprised 30% of the mutD5-induced lacI' mutants in either medium. Taking into account the lower percentage of i-d mutations in a wild-type strain (10%) (23, 24), the mutator effect for i-d mutations can be calculated as 480-fold in minimal medium and 37,000-fold in rich medium (Table 1). The sequencing results for 502 mutations are summarized in Table 2. In either medium, 75% of the mutations were base substitutions with the remaining 5% mostly single-base frameshifts. All but 1 of the 24 frameshifts was the loss of a single base (data not shown). They all occurred in a run of two or more identical bases and probably arose through slippage during DNA synthesis, as first suggested by Streisinger et al. (29). No obvious differences were discerned between the two frameshift spectra.

mutD-Induced Base Substitutions. In addition to the ~80-fold difference in i-d frequency, the two conditions produce different specificities of base substitutions (Table 2). In minimal medium, 144/232 (62%) are transversions, mostly (77%) due to A→T → T-A. A→T → C-G and G+C → T-A transversions contribute moderately; G→C→C are rare. Among the transitions, G→C→A are more frequent than A→T→G in rich medium, substitution frequencies are dominated (89%) by transitions, as seen before (20, 21). A→T→G are also more frequent than A→T→G. Among the transversions, A→T → T-A do not dominate and appear with similar frequency as A→T→C-G and G+C→T-A. These differences
suggest that the increased mutagenesis upon medium shift represents not merely the enhancement of one mechanism but, more probably, the involvement of different mechanisms. The frequency ratios (Table 2) quantitatively express the effect of the medium shift on the individual mutations: transitions are increased \( \approx 200\)-fold; transversions are increased only 14-fold.

The complete spectra are presented in Fig. 1. In minimal medium, several interesting features emerge. First, the excess A-T \( \rightarrow \) T-A transversions arise at a specific sequence, GTGG (positions 96, 141, and 183), as already observed by Fowler et al. (21). A fourth hot spot (position 195) has a related sequence, CTGG. Together these sites contribute 103/111 (93%) of all A-T \( \rightarrow \) T-A. They are true hot spots because the \( i^{-d} \) target contains a total of 18 detectable A-T \( \rightarrow \) T-A sites. The four hot spots comprise all available GTGG or CTGG sequences in the target (the complement CCAC or CCAG is present several times but A-T \( \rightarrow \) T-A cannot be scored here). Second, the majority (64%) of transitions is found at sequences with an intriguing context: TTTCC (positions 89), TTTTCC (position 90), and TTTC (position 119) (mutated base underlined). In each case, the mutated base is one of several repeated bases and the substitution appears directed by the base that follows (or precedes) the series. They may represent a case of "dislocation" mutagenesis (see Discussion).

In rich medium, the transitions dominate with a distribution very different from that in minimal medium. In fact, the predominance and distribution of the rich-medium transitions were similar to those observed previously for \( \text{mutH}, \text{mutL}, \) and \( \text{mutS} \) strains (24). These strains are defective in post-replicative mismatch repair and, as a result, mutators (250- to 300-fold for \( i^{-d} \)) (24). On this similarity, it was considered possible that \( \text{mutD} \) in rich medium might be defective in mismatch repair (see below).

**Mismatch Repair of Heteroduplex Molecules in \( \text{mutDS} \) Cells.**

\( \text{M13mp2} \) replicative form DNA containing a single T-G mismatch was transfected into competent cells of a wild-type, \( \text{mutDS} \), or \( \text{mutL} \) strain (Table 3). This heteroduplex contains a TAA ochre codon in the viral (+)-strand and the complement of a revertant codon (CAA) in the (-)-strand. The TAA codon abolishes the \( \alpha \)-complementing ability (colorless plaque) (27), while the CAA codon restores it (blue plaque) (17), permitting a visual distinction between the products of two strands. Repair of the mismatch prior to replication leads to a pure plaque, either white [\( (+) \)-strand expression] or blue [\( (-) \)-strand expression], whereas absence of repair will be evidenced by a mixed plaque. Table 3 shows that \( \text{mutDS} \) cells grown in rich medium are indeed deficient in mismatch repair. They produce a high proportion (39.1%) of mixed plaques, whereas wild-type cells yield only a few (2.6%) mixed plaques. The percentage in \( \text{mutDS} \) cells approaches that of the \( \text{mutL} \) strain (48.7%) known to be deficient in mismatch repair (5-8). In contrast, \( \text{mutDS} \) cells, grown in minimal medium prior to conversion into competent cells, are proficient in repair, yielding only 3.8% mixed infective centers.

**DISCUSSION**

In minimal medium, \( \text{mutD} \) is a relatively moderate mutator (refs. 9 and 10 and present results). Nevertheless, for the \( i^{-d} \) target, the increase in mutation frequency is \( \approx 480\)-fold (Table 1). Mutations under this condition are characterized by an excess of A-T \( \rightarrow \) T-A transversions and the frequent occurrence of transitions at sequences reminiscent of a "dislocation" type of mutagenesis (see below). Together, these two comprise 70% of all mutations. The A-T \( \rightarrow \) T-A transversions arise at specific GTGG (or CTGG) sites. It is not immediately obvious why these are so active. GTGG is included in a number of larger sequences that serve as recognition signals for DNA-protein interactions (30-32). Frequent strand interruptions at or near such sites were postulated as an explanation for high frequencies of frameshifts, deletions, and duplications near these sites in a polA1 strain (30), but the implication of this for the origin of base substitutions is not clear. Possibly, the same features that make such DNA sequences useful for DNA-protein recognition also cause an inherent inaccuracy of DNA replication, an effect perhaps dramatically compounded by the \( \text{mutDS} \) defect.

The transitions in minimal medium have interesting DNA contacts: 64% are found at three sequences, TTTCC, TTTC, and TTTC (mutated base underlined). In each case, the mutated base is part of a run of identical bases and its replacement appears directed by the base that follows or precedes the series. This is reminiscent of the dislocation mechanism proposed for certain base substitutions by DNA polymerases in vitro (33). The model postulates misalignment (or slippage) during synthesis across the run, incorporation of a correct nucleotide at the base that follows the run, and realignment placing the bases back in register but creating a mispair at the final base of the run. It is an interesting possibility that such a mechanism may be responsible for a substantial portion of the \( \text{mutDS} \)-induced transitions. It is noted that Fowler et al. (20) already proposed such a model for \( \text{mutDS} \)-induced mutagenesis, although irrespective of runs.

The precise mechanisms responsible for mutagenesis in minimal medium cannot be deduced from the specificity data, but some hypotheses may be advanced. The total increase for \( i^{-d} \) mutations is \( \approx 480\)-fold (Table 1), but 70% of this is caused by the two classes discussed above. It cannot be excluded that these classes, in view of their "DNA structural" aspects, represent a somewhat special case. For the remaining 30% (56% transversions and 44% transitions) the increase is \( \approx 150\)-fold. In view of the 3'-exonuclease defect of \( \text{mutDS} \) cells (11), it seems reasonable to assign at least this segment to the proofreading defect. The \( \text{mutD} \) mutation could also indirectly affect the polymerase subunit of Pol III holoenzyme, as suggested (11, 34). The \( \alpha \)- and \( \epsilon \)-subunits bind together to form (with the \( \theta \)-subunit) the Pol III core (18) and functionally interact (35, 36). The possibility that the insertion fidelity of the polymerase is affected as well can therefore not be excluded.

In rich medium, \( \text{mutD} \) is a very strong mutator, \( i^{-d} \) mutations arising at a frequency 37,000-fold above wild-type level. In view of the possibility that proofreading might not be
FIG. 1. i-th base substitutions observed in mutD5 strain after growth in minimal and rich media. Red, transversions; yellow, G-C → A-T transitions; blue, A-T → G-C transitions. Double mutations were observed for two mutants in minimal medium: C→T at position 84 and C→A at position 129, each in combination with the lacO mutation T→C at position +6, and for four mutants in rich medium: (T→A at position 72 and G→A at position 107), (C→T at position 150 and T→C at position 54), (G→A at position 201 and A→G at position 205), and (A→G at position 83 and T→C at position 238).
The solely and mismatch frequency mutation to frequencies mutD5 proved in repair temporal and MutHLS by Pol repair (21). However, why this E-subunit mutation $=\Delta$ shift. Other added the binding to Regulation proofreading such that the error-saturation of the same through leading mechanism. $\Delta$-subunit, model. Mismatch discovery mediated by thymidine of $\Delta$-subunit, or alternatively, the one leading increase in mutagenesis (ii) demonstrated in mismatch deficient, increasing the apparent excess of transversions. In rich medium, however, mismatch repair becomes deficient, resulting in a strong increase in mutagenesis and a shift of the mutations in favor of transversions. 

Ronnie Dunn, RobinCornacchio, and Norris Armstrong are acknowledged for their excellent and enthusiastic technical assistance throughout this work. I thank J. Drake, T. Kunkel, and J. Roberts of this institute and L. Loeb (University of Washington) for their critical reading of this manuscript and helpful suggestions.