Repair of defined single base-pair mismatches in *Escherichia coli*
(bacteriophage λ/heteroduplex/mutagenesis/gene conversion)

CHRISTIANE DOHET*, ROBERT WAGNER*, AND MIROSLAV RADMAN†

*Institut J. Monod, Centre National de la Recherche Scientifique, Université Paris 7, Tour 43, 2 place Jussieu, 75251 Paris, Cedex 05, France; and †Département de Biologie Moléculaire, Université libre de Bruxelles, B 1640 Rhode-St-Genèse, Belgium

Communicated by Matthew Meselson, September 27, 1984

ABSTRACT

Heteroduplexes with single base-pair mismatches of known sequence were prepared by annealing separated strands of bacteriophage λ DNA and used to transfect *Escherichia coli*. Each of the eight possible single base-pair mismatches was constructed. Genetic analysis of the progeny phages obtained from transfected bacteria indicates that the E. coli mismatch repair system does not recognize (or does not repair) all single base-pair mismatches with equal efficiency. In particular, the A→G, C→T, and C→C mismatches appear to be less repaired than any others. The mutator character of mismatch repair-deficient mutants suggests that such unrepaired mismatches should occur infrequently during E. coli DNA replication.

Mismatch repair is believed to correct errors arising during DNA replication by removing incorrect—i.e., mispaired—bases from newly synthesized DNA strands (1, 2). The mismatch repair system in *Escherichia coli* does not appear to act in regions of DNA where G-A-T-C sequences are fully adenine methylated (3) and it appears to be the transient undermethylation of newly synthesized G-A-T-C sequences in the region immediately following the replication fork that allows mismatch repair to occur only on newly synthesized strands (1–3). Repair appears to involve localized excision and resynthesis (1, 4, 5).

Some mutator mutants of *E. coli* [mutU, mutL, mutS, mutU (uvrD, uvrE)] have been shown to be deficient in mismatch repair (3, 6–9). The mut-dependent increase of forward mutation or reversion frequencies observed for several markers depends on both the marker and the mut mutation (8, 10), suggesting that the *E. coli* mismatch repair system may recognize or repair certain mismatches more efficiently than others. It appears that such is the case for the mismatch repair system of *Streptococcus pneumoniae* where two mismatches, A→G and C→C, do not appear to be repaired (11). The results of experiments using heteroduplexes prepared from separated strands of bacteriophage λ DNA to transfect *E. coli* cells suggest that at least one mismatch may not be repaired in *E. coli* as well, although the nature and sequence of the mismatch were not known (3).

To determine the specificity of the *E. coli* mismatch repair system, we have measured the amount of repair of mismatches in heteroduplexes prepared from separated strands of unmethylated bacteriophage λ DNA containing sequenced mutations in the *cl* gene. Mutants were chosen to allow the formation, in mutant/wild-type heteroduplexes, of each of the eight possible single base-pair mismatches—A→C, A→G, A→A, T→G, T→C, T→T, C→C, and G→G.

The results indicate that all mismatches are not recognized (or not repaired) with equal efficiency. In particular, the A→G, C→T, and C→C mismatches appear to be less repaired than any others.

MATERIALS AND METHODS

λ phages with sequenced *cl* mutations were obtained from Franklin Hutchinson (Yale University) (Fig. 1). The original phages were *c* + (or *cl*) and also carried the markers cl857 *ind-* Oam 29. The phages used in these experiments were made *O* + by recombination. *cl* mutations were originally sequenced, in the laboratory of Franklin Hutchinson (12), by the method of Maxam and Gilbert (20). DNA strands were prepared and annealed as described (1). Transfection was accomplished by the method of Mandel and Higa (13). Transfection efficiency was in the range of 5 × 10⁻⁸ to 5 × 10⁻⁶ transfected cells per DNA heteroduplex at a multiplicity of about 0.1 DNA molecule per cell. Transfection efficiency varies with DNA preparation and bacterial strain but does not appear to affect the results—i.e., transfection efficiency does not appear to be related to the extent of mismatch repair. Transfected cells were plated before lysis, giving rise to infective centers. To minimize any selective advantage or disadvantage associated with the *cl* mutations, infective centers were grown under conditions (37°C) in which all phages used in these experiments grow lytically due to the presence of the temperature-sensitive *cl*857 mutation. Individual infective centers were picked and the phages were replated at 32°C to determine their genotype at *cl*. Infective centers were scored as clear (C)—i.e., containing only phages with the *cl* mutation; turbid (C+)—i.e., containing only phages with the wild-type allele of the *cl* mutation; or mixed (C+/C)—i.e., containing both wild-type and mutant phages. The results of control experiments in which infective centers derived from cells infected with phages having only the wild-type allele of the *cl* gene were analyzed reveal that ~1% of turbid infective centers will be scored as mixed because of forward mutation of C+ to C. This sets the lower limit of mixed infective centers that will be recovered in these experiments somewhere between 0.5% and 1%. Reversions of the *cl* mutations are much too infrequent to interfere with scoring. At least 350 infective centers were tested for each heteroduplex. Different DNA preparations of the same heteroduplex give essentially identical results.

RESULTS AND DISCUSSION

Heteroduplexes were prepared from separated strands of four different *cl* mutants and their respective wild-type alleles to create each of the eight possible single base-pair mismatches (Table 1). The mutations were chosen to fall within a small (35-base-pair) region of the *cl* gene (Fig. 1). The DNA was isolated from phages grown in *dam E. coli* (GM32-deficient in adenine methylation; ref. 17) so that repair on either strand could be observed, allowing any strand preference for repair to be detected.

Transfections of *dam* and wild-type cells give essentially identical results (not shown), suggesting that, in this system, methylation during transfection is not rapid enough to affect the results. A similar result was obtained by others utilizing helper phage-mediated transfection (3).

The results of transfections with all the eight different heteroduplexes are presented in Table 1. Three types of infective
Table 1. Genetic analysis of infective centers derived from heteroduplexes containing defined mismatches

<table>
<thead>
<tr>
<th>Mismatch position</th>
<th>Mismatch</th>
<th>Transfected bacteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Wild type*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mixed, % (C+/C)</td>
</tr>
<tr>
<td>26</td>
<td>c⁺→G⁻</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>c⁻→T⁻</td>
<td>10.3</td>
</tr>
<tr>
<td>11</td>
<td>c⁺→C⁻</td>
<td>6.9</td>
</tr>
<tr>
<td></td>
<td>c⁻→T⁻</td>
<td>12.2</td>
</tr>
<tr>
<td>45</td>
<td>c⁺→G⁻</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>c⁻→T⁻</td>
<td>39.4</td>
</tr>
<tr>
<td></td>
<td>c⁺→C⁻</td>
<td>80.0</td>
</tr>
<tr>
<td>15</td>
<td>c⁻→A⁻</td>
<td>72.7</td>
</tr>
</tbody>
</table>

The source of "c⁺" strands (both l and r) for all heteroduplexes was λ cI857 ind l (sequence, ref. 14). λ cI857 ind l phages with additional cI mutations were used to provide "c⁻" strands. The mutants are spontaneous single base substitutions isolated and sequenced in the laboratory of Franklin Hutchinson (12). Mismatch positions are as follows: position 26, SP27, G-C → A/T transition; position 11, SP62, A/T → T-A transversion; position 45, SP51, G-C → C-G transversion; position 15, SP57, G-C → T-A transversion.

*E. coli C600 (15).
†E. coli ES871 (16).

centers are observed: (i) those yielding only c⁺ phages (pure turbid), (ii) those yielding only cI phages (pure clear), and (iii) those yielding both c⁺ and cI phages (mixed). Pure infective centers may result from mismatch repair, from strand loss (1), or from contaminated strand preparations (transfections with single-stranded single-strand pre annealing reveal that strand preparations are frequently contaminated with r strands). Mixed infective centers reflect either the absence of both mismatch repair and strand loss, the chance superposition of two transfected cells in one infective center, or the uptake by one cell of more than one DNA molecule. Earlier work with multiply mismatched heteroduplexes indicates that the chance superposition of transfected cells is a relatively rare event (1). Transfecting with an equal mixture of cI and c⁺ homoduplex DNA molecules at a multiplicity 10 times higher than that used for heteroduplexes produced no more than 1% mixed infective centers, most of which presumably resulted from forward mutations of c⁺ to cI (see Materials and Methods). Thus, mixed infective centers derive primarily from heteroduplexes that have escaped mismatch repair.

The relative fractions of mixed and pure infective centers are determined by (i) the extent of mismatch repair, (ii) the amount of strand loss in heteroduplexes that have not been repaired, and (iii) the amount of contaminating homoduplex molecules. An upper limit for the amount of strand loss and strand contamination can be estimated from the results of heteroduplex transfections in mutL bacteria, in which mismatch repair has been shown to be absent or at least much less efficient than in wild-type bacteria (3, 5, 7, 9). In these experiments, pure infective centers account for only 15–23% of the total number of infective centers, compared to values as high as 94% when the transfected cells are wild type with respect to mismatch repair (Table 1). The finding of mismatches that do not appear to be repaired—i.e., ones yielding a large fraction of mixed infective centers in wild-type transfections—indicates that the difference between mutL and wild-type transfections is not due to a difference in the extent of strand loss. Thus, even if it is assumed that all pure infective centers from transfections in mutL cells are the result of strand loss or strand contamination, it does not appear that these phenomena substantially diminish the fraction of mixed infective centers recovered under conditions in which mismatch repair can occur. The amount of strand loss observed in these experiments (no more than 15–25% of transfecting heteroduplexes) is less than that observed in earlier experiments (1) using helper phage-mediated transfection (60–70%). We have no explanation for this difference except to suggest that it might reflect some difference in the replication (symmetric vs. asymmetric) of the heteroduplexes in the two transfection procedures—i.e., it may be that the preinfecting helper phages produce λ replication proteins such that transfecting heteroduplexes initiate rolling circle replication more rapidly than in CaCl₂-mediated transfections.

The data presented in Table 1 indicate that the mismatches T-G, C-A, T-T, A-A, and G-G are repaired at reasonably high efficiency in wild-type bacteria—i.e., <13% mixed infective centers, G-A, T-C, and C-C mismatches are poorly repaired; G-A and T-C mismatches may be completely refractory to

![Fig. 1. DNA sequence of the region of the λ cI gene coding for the amino-terminal end of the λ repressor (14). Locations and base changes of the mutations used in this study are indicated. Numbers indicate the number of bases from the start of the cI P<a sup><sub>RM</sub></a> transcript.](image-url)
repair. Thus, it appears that the E. coli mismatch repair system does not recognize or repair all mismatches with equal efficiency.

It is not clear to what extent the results are affected by nucleotide sequence context. However, the data indicate that neighboring nucleotide sequence cannot be the sole determinant of mismatch repair efficiency since there exists a reciprocal pair of mismatches, which are necessarily at the same site in the DNA, in which one mismatch (G-G) is well repaired and the other (C-C) is not (Table 1).

It may be that the specificity of the E. coli and S. pneumoniae mismatch repair systems is identical, as it appears that in neither system are G-A and C-C mismatches well repaired (11). However, there are differing reports of the amount of repair of C-T mismatches in S. pneumoniae (11, 18).

It is possible that mismatches differ with respect to the amount of time required to recognize or repair them. It may be that during E. coli DNA replication more time is available for mismatch repair than in our experimental system, such that all mismatches are recognized and repaired. However, it seems most likely that the time available for mismatch repair during E. coli DNA replication—i.e., the time before adenine methylation of newly synthesized G-A-T-C sequences—is no longer than the time available for mismatch repair in these experiments—i.e., the time between entry of heteroduplex DNA into the transfected cells and initiation of the first round of λ replication.

The data reveal no mismatch-dependent strand preference for repair. However, there does appear to be a consistent bias among pure bursts in favor of λ strand genotypes (Table 1). The source of this bias is not known. In the case of transfections of mutL bacteria, it may be accounted for by λ strand contamination, preferential strand loss, or residual repair. In transfections of wild-type bacteria with heteroduplexes containing repairable mismatches, in which pure bursts are primarily the result of mismatch repair, the bias may be due to preferential repair of the non-transcribed strand—i.e., it may be that transcription enzymes compete with mismatch repair enzymes for space on the transcribed strand such that repair is more common on the non-transcribed strand (r strand in the cI region).

The finding that the E. coli mismatch repair system readily recognizes and repairs only three of the six transversion mismatches (A-A, T-T, and G-G) accounts for the findings that the mutator effects observed in mutH, mutL, mutS, and mutU mutants, which are deficient in mismatch repair, and dam mutants, which have undirected mismatch repair, are due primarily to an increase in transition and frameshift mutations (10, 19). [We have found that a single base addition frameshift mismatch can be recognized and repaired on either strand by the E. coli mismatch repair system (21).] Given the observed pattern of mismatch repair and the fact that mut and dam mutants are mutators, it appears that, in E. coli, transversion mismatches either arise much less frequently during DNA replication than transition mismatches or are corrected by some repair system, perhaps polymerase-associated (e.g., proofreading), which acts before the mismatch repair we are studying. A similar situation may exist in S. pneumoniae, in which mismatch repair has also been found to repair the transition mismatches but only some of the transversion mismatches (11, 18).

Dr. F. Hutchinson generously provided the sequenced mutants used in this study. R.W. is the recipient of a senior fellowship from Ligue Nationale Française contre le Cancer. This work was supported by Ligue Nationale Française contre le Cancer and a grant from Association pour le Développement de la Recherche sur le Cancer.