Control of Large Chromosomal Duplications in Escherichia coli by the Mismatch Repair System

Marie-Agnès Petit,* Joan Dimpfl,* Miroslav Radman† and Harrison Echols*

*Division of Biochemistry and Molecular Biology, University of California, Berkeley, California 94720, and †Institut Jacques Monod, F-75251 Paris-Cedex 05, France

Manuscript received May 25, 1991
Accepted for publication July 2, 1991

ABSTRACT

Excessive recombination between repeated, interspersed, and diverged DNA sequences is a potential source of genomic instability. We have investigated the possibility that a mechanism exists to suppress genetic exchange between these quasi-homologous (homeologous) sequences. We examined the role of the general mismatch repair system of Escherichia coli because previous work has shown that the mismatch repair pathway functions as a barrier to interspecies recombination between E. coli and Salmonella typhimurium. The formation of large duplications by homeologous recombination in E. coli was increased some tenfold by mutations in the mutL and mutS genes that encode the mismatch recognition proteins. These findings indicate that the mismatch recognition proteins act to prevent excessive intrachromosomal exchanges. We conclude that mismatch repair proteins serve as general controllers of the fidelity of genetic inheritance, acting to suppress chromosomal rearrangements as well as point mutations.

E. coli and Salmonella typhimurium carry large duplications of genomic segments at frequencies that can be as high as 10^-4 of a bacterial population (ANDERSON and ROTH 1977; PETES and HILL 1988). These duplications are thought to arise by an unequal recombination event between homologous sequences dispersed in the genome. Two examples of such “gene families” are the rnr operons for ribosomal RNA and the rhs sequences (ANDERSON and ROTH 1977; PETES and HILL 1988). Seven rnr loci are present in the E. coli chromosome. More recently, up to five rhs loci have been described in E. coli; the rhs sequences lack a known genetic function (SADOSKY et al. 1989; FEULNER et al. 1990). In addition to duplication, rnr operons have been defined as sites for inversion, deletion, and transposition of chromosomal segments (HILL et al. 1977; HILL and HARNISH 1981, 1982).

Chromosomal rearrangements, especially duplications, can serve as valuable sources for environmental or evolutionary adaptation. For example, duplication in Salmonella appears to aid growth on limiting carbon source, probably by a gene dosage effect (SONTI and ROTH 1989). However, some genetic rearrangements, especially deletions, can be highly deleterious. Recent work has shown that the frequency of large duplications in E. coli increases tenfold when the SOS response is constitutively activated (DIMPFL and ECHOLS 1989). This observation indicates that duplication formation might be subject to some form of cellular regulation; the SOS response might allow more frequent rearrangements as a source of increased genetic variation for an endangered bacterial population (DIMPFL and ECHOLS 1989). Based on the SOS work, we were led to explore mechanisms that might normally suppress formation of large duplications and other chromosomal rearrangements.

A property of the rnr and rhs gene families is a slight divergence in DNA sequence (homeologous sequences) (HILL, HARVEY and GRAY 1990). We were intrigued with the possible regulatory significance of this property because the recombinational barrier between the homeologous DNA sequences of E. coli and S. typhimurium can be largely eliminated by mutations in the mismatch repair system (RAYSSIGUTER, THALER and RADMAN 1989). Thus, mismatch repair proteins might function as part of a general mechanism to limit intrachromosomal as well as interchromosomal recombination between duplicated, diverged DNA sequences.

To study the recombination events leading to duplication, we have used a bacterial strain that allows quantitative measurement of duplications between the rhsA and rhsB sequences some 140 kbp apart in the genome (LIN, CAPAGE and HILL 1984). The rhsA and rhsB sequences share a 3.7-kbp region of substantial homology, as judged by hybridization data, but differ by 22 mismatches among the 990 bp already available for sequence comparison (SADOSKY et al. 1989; FEULNER et al. 1990; C. W. HILL, personal communication). The formation of large duplications was increased some tenfold by mutations inactivating the mutL and mutS genes of the mismatch repair pathway. Thus, mismatch repair proteins are likely to function in
limiting intrachromosomal exchanges. Additional experiments indicate that the recombination events inhibited by the MutL and MutS mismatch proteins are RecBC-dependent.

MATERIALS AND METHODS

Bacterial strains: Bacterial strains are listed in Table 1 with their genotypes. New strains were prepared using transduction by phage P1 of transposon insertions in the gene of interest.

Assay of glySL duplications: The genetic test allowing the detection of duplications has been previously described (Lin, Capage and Hill 1984; Dimpfel and Echols 1989). The duplication assay relies on the glySL cold-sensitive mutant allele of the glycyl-tRNA synthetase gene, in conjunction with trpA36glyU(sup). The trpA36 mutation is effectively suppressed by the glyU(sup) allele at 37°C. However, at 20°C suppression of trpA36 is inefficient because of poor charging of the suppressor tRNA, leading to poor growth in the absence of tryptophan. In a glySL duplication, the glycine-inserting missense suppressors are charged more efficiently because of a gene dosage effect, leading to better growth and large colony formation in the absence of tryptophan. Southern hybridization has shown that this selection is specific for duplication of the rhsA-rhsB region, which includes the glySL gene (Lin, Capage and Hill 1984). Duplication frequency was deduced from the ratio of large colonies over total colonies (20 plates scored per experiment). As mismatch-deficient strains are known to favor transposon excision (Lundblad and Kleckner 1984), several large colonies from each experiment were checked for the antibiotic marker linked with the mismatch mutation; 45 among 45 had retained the antibiotic-resistant phenotype.

Segregation analysis of duplications: These measurements were based on techniques developed for Salmonella (Anderson, Miller and Roth 1976). To remove selective pressure favoring glySL duplications, individual large colonies were grown at 37°C in LB broth for 10 to 15 generations, and then plated on the original selective medium without tryptophan at 20°C for 4 days. If segregation occurred, a mixed population of large and small colonies could be detected, among which the proportion of small colonies represented the proportion of segregants.

RESULTS

Effect of mismatch repair mutations on duplication formation: In the duplication test strain CH1504, duplications of the rhsA-rhsB region containing the glySL gene are recognized as large colonies after growth under low temperature conditions (see MATERIALS AND METHODS for details). A schematic map of the CH1504 chromosome, before and after duplication, is presented in Figure 1. Duplication frequency was measured in strain CH1504 and its derivatives harboring a mutL, mutS, mutH or uvrD mutation. Each of these mutations inactivates the general mismatch repair pathway (Modrich 1989; Radman 1989). The results of duplication assays are presented in Table 2. In a mut+ context, the duplication frequency was 1.1 x 10^-4, which agrees with previously reported data (Lin, Capage and Hill 1984; Dimpfel and Echols 1989). In the absence of an active mutL or mutS gene product, duplication frequency was 10–15-fold higher (Table 2, lines 2 and 3). In contrast, the mutH and uvrD mutants did not show an increase in duplication frequency (Table 2, lines 4 and 5).

To verify the presence of a duplication in the large colonies, a segregation experiment was performed on several purified colonies. Duplication mutations are distinguished by their instability when the selective pressure for them is removed, presumably because of homologous recombination within the duplicated region (Campbell 1965; Anderson and Roth 1977). The results of the segregation analysis are presented in Table 3. Among 11 large colonies tested for each strain, the mutL, mutS and mutH strains were found to give 100% with unstable phenotype. In addition, the percentage of segregation was indistinguishable between the wild type and the mutL or mutS derivatives. This observation indicates that the frequency of homologous recombination between duplicated regions responsible for segregation was approximately the same for the three strains. Thus the enhanced duplication formation in the mismatch-defective mutants appears to be specific for mismatched DNA sequences (although our experiments do not demonstrate that the same sequences are recombined in the mismatch mutants and wild-type strains). The uvrD strain gave 0/11 unstable phenotypes; therefore, duplication frequency might be overestimated in the uvrD strain.

From the data of Table 2 and Table 3, we infer that the MutL and MutS proteins are likely to have an active role in limiting the frequency of large duplications. The MutL and MutS proteins are involved in the initial recognition of mismatched heteroduplex DNA (Grilley et al. 1989; Modrich 1989). MutS binds to the mismatch, and MutL associates with MutS. The MutH protein nicks at an unmodified GATC methylation sequence, and the UvrD helicase (helicase II) presumably evicts the unwanted DNA strand (Modrich 1989). Our data indicate a special role for the mismatch recognition proteins.

The genetic analysis reported in Table 2 reflects two processes: the formation of duplications by homologous recombination and the resolution of dupli-
cations by homologous recombination. Thus the MutL and MutS proteins might destabilize the duplicated state rather than inhibit duplication formation. If this alternative were true, segregation frequency should be significantly reduced in mutL or mutS mutants compared to wild type. The data in Table 3
show that the segregation frequency was not reduced for mutL or mutS strains compared to wild type. We believe therefore that this alternative explanation is unlikely. We conclude that the MutL and MutS proteins probably act to reduce formation of large duplications.

**Recombination genes affecting enhanced duplication formation:** We wanted to determine which recombination pathway of E. coli was stimulated by the mutL and mutS mutations. For this purpose, we introduced additional mutations inactivating the RecBC or RecF pathways (CLARK 1973; SMITH 1987) (Table 4). Duplication formation in mutL and mutS strains was reduced by a factor of 10 when a recB mutation was introduced. Therefore, mutL and mutS gene products seem to interfere with a recombination event promoted by the RecBC pathway of E. coli. In contrast, introduction of a recF mutation did not reduce duplication formation in mutL and mutS strains.

Our data on the RecBC pathway indicate that this recombinational route is highly efficient for intrachromosomal exchanges in the absence of intervention by the mismatch proteins. Thus, genomic stability may depend on the ability of the mismatch repair system to exert a negative control on the activity of the RecBC pathway in homologous exchanges.

**Additivity with SOS-enhanced duplication frequency:** In a recent study, another route to enhanced duplication formation was described; the same duplication studied here became some tenfold more frequent when the SOS region was derepressed by an SOS-constitutive mutant of recA (DIMPFL and ECHOLS 1989). We investigated the combined effect of a mutL or mutS defect and SOS constitutive expression (Table 5); the resultant strains exhibited a 25–30-fold increase in duplication frequency. Therefore, the two mutational routes to enhanced duplication formation appear to be additive. In additional experiments, we found that the mutL mutation conferred a ten-fold increase in duplication formation in a strain carrying a noninducible allele of the LexA repressor of SOS (data not shown). The SOS pathway to induced duplication depends on both RecF and RecBC (DIMPFL and ECHOLS 1989; and our unpublished data). This dependence on RecF indicates that SOS-inducible duplications involve at least some differences in mechanism from those stimulated by a defect in mismatch repair.

**DISCUSSION**

**Role of mismatch recognition proteins in homologous recombination:** From the data reported in this paper, we conclude that the mismatch recognition proteins MutS and MutL act to prevent excessive intrachromosomal recombination between duplicated, diverged (homologous) DNA sequences. Interestingly, our data on mismatch repair mutants are highly similar to those obtained in E. coli-S. typhimurium crosses, in which mutS and mutL mutations were by far the most effective in stimulating recombination (RAYSSIGUIER, Thaler and Radman 1989). SHEN and

### TABLE 2

Effect of the mismatch repair genes on the duplication of the rhsA-rhsB region

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Mismatch repair genotype</th>
<th>Duplication frequency ($\times 10^7$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH1504</td>
<td>+</td>
<td>1.1 (±0.5)</td>
</tr>
<tr>
<td>JCD1400</td>
<td>mutL</td>
<td>16.0 (±6)</td>
</tr>
<tr>
<td>JCD1402</td>
<td>mutS</td>
<td>11.0 (±3)</td>
</tr>
<tr>
<td>JCD1404</td>
<td>mutH</td>
<td>0.3 (±0.2)</td>
</tr>
<tr>
<td>JCD1406</td>
<td>uvrD</td>
<td>0.6 (±0.3)</td>
</tr>
</tbody>
</table>

* Average and standard deviation are given for three separate experiments.

### TABLE 4

Effect of recombination mutations on duplication frequency

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Mismatch repair genotype</th>
<th>Recombination genotype</th>
<th>Duplication frequency ($\times 10^7$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH1504</td>
<td>+</td>
<td>+</td>
<td>1.1 (±0.5)</td>
</tr>
<tr>
<td>JCD1408</td>
<td>mutL</td>
<td>+</td>
<td>10.0 (±6)</td>
</tr>
<tr>
<td>JCD1410</td>
<td>mutS</td>
<td>+</td>
<td>6.4 (±4)</td>
</tr>
<tr>
<td>JCD1412</td>
<td>mutL recB</td>
<td>+</td>
<td>0.9 (±1.5)</td>
</tr>
<tr>
<td>JCD1414</td>
<td>mutS recB</td>
<td>+</td>
<td>0.5 (±0.5)</td>
</tr>
<tr>
<td>JCD1416</td>
<td>+</td>
<td>recB</td>
<td>0.3 (±0.1)</td>
</tr>
<tr>
<td>JCD1418</td>
<td>mutL recF</td>
<td>+</td>
<td>100.0 (±7.0)</td>
</tr>
<tr>
<td>JCD1420</td>
<td>mutS recF</td>
<td>+</td>
<td>17.0 (±12)</td>
</tr>
<tr>
<td>JCD1422</td>
<td>+</td>
<td>recF</td>
<td>1.6 (±1.1)</td>
</tr>
</tbody>
</table>

* Average and standard deviation are given from three or four separate experiments.

### TABLE 3

Segregation of rhsA-rhsB duplications

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Mismatch repair genotype</th>
<th>No. of unstable colonies/total</th>
<th>Percent segregation of unstable colonies</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH1504</td>
<td>+</td>
<td>12/12</td>
<td>26 (±5)</td>
</tr>
<tr>
<td>JCD1400</td>
<td>mutL</td>
<td>11/11</td>
<td>39 (±8)</td>
</tr>
<tr>
<td>JCD1402</td>
<td>mutS</td>
<td>11/11</td>
<td>20 (±12)</td>
</tr>
<tr>
<td>JCD1404</td>
<td>mutH</td>
<td>11/11</td>
<td>41 (±6)</td>
</tr>
<tr>
<td>JCD1406</td>
<td>uvrD</td>
<td>0/11</td>
<td>0</td>
</tr>
</tbody>
</table>

* Average and standard deviation are given from three separate experiments.

### TABLE 5

Additive effect on duplication frequency of mismatch repair deficiency and SOS activation

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Mismatch repair genotype</th>
<th>SOS</th>
<th>Duplication frequency ($\times 10^7$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JCD1000</td>
<td>+</td>
<td>Constitutive</td>
<td>9 (±6)</td>
</tr>
<tr>
<td>JCD1422</td>
<td>mutL</td>
<td>Constitutive</td>
<td>26 (±8)</td>
</tr>
<tr>
<td>JCD1424</td>
<td>mutS</td>
<td>Constitutive</td>
<td>28 (±3)</td>
</tr>
</tbody>
</table>

* Average and standard deviation are given from three separate experiments.
HUANG (1989) have previously found that a mutS mutation increased recombination involving homelogous sequences between $\lambda$ phage and a plasmid. FEINSTEIN and LOW (1986) have noted that mutations in the mismatch repair genes confer increased recombination in Hfr $\times$ F$^-$ crosses; in these experiments involving highly homologous sequences, the major effect was noted for a mutant wrpD gene. These results indicate that a somewhat different, but mismatch-related mechanism can limit recombination in the case of very limited sequence divergence (the two point mutations in the hisF gene studied in the cross).

The intervention of the MutL and MutS proteins in the recombination process might occur immediately after the RecA-promoted strand exchange between rhsA and rhsB. As judged by experiments in vitro, RecA will catalyze heteroduplex formation through small sequence heterologies (DAS GUPTA and RADDING 1982). The consecutive mismatches would offer a target for the MutL and MutS proteins, whose activity could result in elimination of the heteroduplex intermediate, hence preventing its resolution in duplication. The ability of the MutL and MutS proteins to detect a mispaired region of a chromosome seems to be utilized in at least two ways. As shown previously, the general mismatch repair pathway contributes to replicational fidelity through repair of errors by DNA polymerase (MODRICH 1989; RADMAN 1989). Mismatch repair proteins also contribute to recombinational fidelity, by preventing the introduction of homologous sequences by horizontal transfer between species (RAYSSIGUIER, THALER and RADMAN 1989) and by limiting the introduction of vertical divergence by large rearrangements (this work).

**Importance of controlled homeologous recombination between gene families:** Recombination between members of the rrn gene family is another known possibility for making large rearrangements of the E. coli chromosome (HILL, HARVEY and GRAY 1990). The rrr operons are known to contain sequence divergence that could appear as mismatch protein targets in the course of the recombination process (JINKS-ROBERTSON and NOMURA 1987); for example, about 1% divergence was observed in a sequence comparison between 16S RNA of rrrG and rrrB (SHEN, SQUIRES and SQUIRES 1982). In addition, the spacer tRNA sequences might serve as targets for a special recognition system. For the single rearrangement investigated in this study, the occurrence can be as high as 0.1% in the absence of mutL or mutS gene products; under these conditions, the cumulated probability of a rearrangement between any two members of a gene family could reach 5%. Thus, we presume that long-term genomic stability requires a general mechanism for negative control of intrachromosomal recombination and unequal crossing over between homologous chromosomes. The likelihood of such a recombination-limiting system has also been inferred from another point of view: the much lower frequency of duplication by homologous recombination compared to the frequency of segregation by homologous recombination (HILL, HARVEY and GRAY 1990). Our experiments define the mismatch recognition proteins as likely components of this recombinational fidelity system. Eukaryotic cells appear to carry a general mismatch repair system similar to that found in E. coli (HOLMES, CLARK and MODRICH 1990; THOMAS, ROBERTS and KUNKEL 1991). This mismatch pathway might also prevent aberrant recombination (RAYSSIGUIER, THALER and RADMAN 1989; BAILIS and ROTHSTEIN 1990).

For eukaryotic organisms, recombination between homologous chromosomes is almost exclusively limited to meiosis. The recombinational fidelity mechanism proposed here might participate in this limitation on mitotic recombination. The frequent occurrence of homologous sequences in eukaryotes could trigger the mismatch-directed system to abort attempted mitotic recombination between homologous chromosomes; suppression of this pathway would then be a component of meiotic development. The mismatch system could be the regulatory balance limiting recombination to the proper meiotic pathway.

We thank CHARLES HILL for unpublished information, RICHARD EISNER for editorial help, MARCIE ROSENBERG and MALINI RAJAGOPALAN for advice about the manuscript, and GRAHAM WALKER, NANCY KLECKNER, ROBERT LLOYD and MARGARET LIEB for strains. This work was supported by a grant from the National Institutes of Health (CA41655).

**LITERATURE CITED**


Communicating editor: J. W. Drake