Repair of Single- and Multiple-Substitution Mismatches During Recombination in *Streptococcus pneumoniae*

Anne-Marie Gasc, Armand-Michel Sicard and Jean-Pierre Claverys

Centre de Recherche de Biochimie et de Génétique Cellulaires du Centre National de la Recherche Scientifique, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex, France

Manuscript received February 20, 1988
Accepted for publication August 31, 1988

ABSTRACT

The use as genetic markers, during transformation of *Streptococcus pneumoniae*, of 19 sequences differing from wild type, located throughout the amiA locus, enabled us to examine the fate of 24 single- and 11 multiple-mismatches during recombination. Tentative mismatch ranking as a function of decreasing repair efficiency by the Hex mismatch repair system is G/T = A/C = G/G (maximum repair: 90–95%) > C/T (mostly 75 to 90% repair) > A/A (from 50 to 90% repair) > T/T (50–65% repair) > A/G (from 0 to 20% repair) > C/C. No indication of correction of the latter has been obtained. Over the limited number of samples examined, we observed no influence of the base composition of the surrounding sequence on correction efficiency for both transition mismatches and for G/G and C/C. Variations in the surrounding sequence affect repair of A/G and C/T, and, even more strongly, of A/A and T/T. No simple correlation to the G:C content of the surrounding sequence is apparent from our results, in contrast to the conclusion drawn for the Mut mismatch repair system of *Escherichia coli*. Examination of the fate of multiple mismatches suggests that C/C may sometimes impede recognition of otherwise corrected mismatches.

The *Streptococcus pneumoniae* mismatch repair system (Hex system) is able to recognize noncomplementary base pairs in DNA (for review see Claverys and Lacks 1986). The Hex system acts both during transformation, a recombination process that directly produces heteroduplex DNA, and after DNA replication to remove misincorporated bases.

DNA-mediated transformation provides a unique opportunity to study the specificity of mismatch repair. Indeed, as a result of processing upon entry (Lacks 1977), double stranded donor DNA is made available inside the cell in the form of single stranded fragments. Insertion of a major portion of the single stranded fragments into the recipient genome occurs by physically displacing a segment of the homologous recipient strand (Fox and Allen 1964; Gurney and Fox 1968; Mejean and Claverys 1984), giving rise to heteroduplex molecules. These heteroduplex molecules are subject to Hex-mediated mismatch repair in wild-type strains. Differential repair of the eight possible transition- and transversion-mismatches account for marker-specific variations in transforming efficiencies observed in studies involving mutations in the same locus (Claverys et al. 1981, 1982, 1983; Lacks, Dunn and Greenberg 1982). Efficient repair of the two complementary transition mismatches (A/C and G/T) leads to a low efficiency (LE) of transformation since the donor DNA strand is preferentially repaired. High efficiency (HE) (Claverys and Lacks 1986) of transformation is observed when both complementary mismatches are refractory to repair. LE markers are 20 times less efficient than HEs. The identification of base mismatches recognized by the Hex system has been carried out by the direct determination of DNA sequences of a limited set of markers belonging to the amiA locus (Claverys et al. 1981, 1982, 1983) and to the malM locus (Lacks, Dunn and Greenberg, 1982).

In this paper we present a survey of the results obtained with 19 sequences differing from wild type, located throughout the amiA locus. This study allows us to propose a tentative mismatch ranking as a function of repair efficiency. In addition we examined, for a given mismatch, whether the G:C content in the neighboring nucleotide sequence affects repair efficiency, as suggested for the *Escherichia coli* mismatch repair system (Jones, Wagner and Radman 1987). Finally, analysis of the fate of double-substitution mismatches suggests that the C/C mismatch could preclude recognition of otherwise corrected adjacent mismatches.

MATERIALS AND METHODS

Culture media and transformation procedures have been described (Claverys, Roger and Sicard 1980). Isolation of native DNA, resolution of complementary strands and preparation of heteroduplex DNA are as previously described (Roger 1972; Claverys, Roger and Sicard 1980).

All genetic markers, except *strA1*, belong to the amiA

---

1 To whom correspondence should be addressed.
**Table 1**

<table>
<thead>
<tr>
<th>amiA mutations</th>
<th>Mutation position</th>
<th>DNA sequence</th>
<th>Base change</th>
</tr>
</thead>
<tbody>
<tr>
<td>amiA122</td>
<td>2505</td>
<td>GGCCTACATT</td>
<td>TAAGTTAACG</td>
</tr>
<tr>
<td>amiA129</td>
<td>2549-2551</td>
<td>CAAGTAGCAT</td>
<td>TTTGCTTCTCTCCTGTA</td>
</tr>
<tr>
<td>amiA14</td>
<td>2658</td>
<td>TGAAAGTGTT</td>
<td>TAGGTATATGC</td>
</tr>
<tr>
<td>amiA9 rev1</td>
<td>2568</td>
<td>TGGAGATTC</td>
<td>TACGTTACG</td>
</tr>
<tr>
<td>amiA14 rev36</td>
<td>2658-2560</td>
<td>AAAGTGTTGTA</td>
<td>CTTATGCTA</td>
</tr>
<tr>
<td>amiA154</td>
<td>3166</td>
<td>ATTACTGGTT AAGTGTGTTG</td>
<td>T A</td>
</tr>
<tr>
<td>amiA36</td>
<td>3346</td>
<td>CTTCCAGATT</td>
<td>AATTCCTATC</td>
</tr>
<tr>
<td>amiA124</td>
<td>3373</td>
<td>GCTTGAGATTA ACGTTCTTAC</td>
<td></td>
</tr>
<tr>
<td>amiA6</td>
<td>3374</td>
<td>CTTGAGATTTG ACGTTCTTACAG</td>
<td></td>
</tr>
<tr>
<td>amiA6 rev2</td>
<td>3374</td>
<td>CTGAGATTTG  CCGTTCTTACAG</td>
<td></td>
</tr>
<tr>
<td>amiA6 rev3</td>
<td>3374</td>
<td>CTGAGATTTG  TCGTTCTTACAG</td>
<td></td>
</tr>
<tr>
<td>amiA6 rev4</td>
<td>3373-3374</td>
<td>GCTGAGATTTG CAGTTCTTAC</td>
<td></td>
</tr>
<tr>
<td>amiA141</td>
<td>3383</td>
<td>GGCCTTCTTA GCGTTCTTAC</td>
<td></td>
</tr>
<tr>
<td>amiA141 rev100</td>
<td>3383</td>
<td>GCGGTTCTTA GTTTTACCA</td>
<td></td>
</tr>
<tr>
<td>amiA141 rev32</td>
<td>3382-3383</td>
<td>TGGGTTCTTAC  GCGTTCTTAC</td>
<td></td>
</tr>
<tr>
<td>amiA141 rev10</td>
<td>3381-3383</td>
<td>TGGGTTCTTAC CAGTTCTTAC</td>
<td></td>
</tr>
<tr>
<td>amiA3</td>
<td>4834</td>
<td>CGTTTCCTCTT TAAGTTACG</td>
<td></td>
</tr>
<tr>
<td>amiA3 rev1</td>
<td>4834</td>
<td>CGTTTCCTCTT CAAGTTACG</td>
<td></td>
</tr>
</tbody>
</table>

All amiA mutations are of spontaneous origin. rev refers to revertant to pseudo-wild type (see MATERIALS AND METHODS). As to the origin of the complex change which occurred spontaneously to generate the amiA29 mutation, we observed that this mutation creates a 12-bp long perfect palindrome (the sequence from position 2545 to position 2555 is inversely repeated from position 2567 to position 2578). We propose that the amiA29 change (located exactly in the middle of these 12 bp) occurred by spontaneous sequence-directed mutagenesis (GOLDING and GLICKMAN 1985), a mechanism which predicts multiple changes as a result of a single mutational event.

**Table 2**

<table>
<thead>
<tr>
<th>Marker</th>
<th>Transforming efficiency</th>
<th>Calculated strand efficiency</th>
<th>Percent repair</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L/H</td>
<td>L</td>
<td>H</td>
</tr>
<tr>
<td>amiA141</td>
<td>0.88 (±0.09)</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>amiA150</td>
<td>0.83 (±0.16)</td>
<td>0.16</td>
<td>0.11</td>
</tr>
<tr>
<td>amiA144</td>
<td>0.88 (±0.18)</td>
<td>6.8</td>
<td>0.77</td>
</tr>
</tbody>
</table>

*a* Calculated by combining overall transforming efficiency (taken as L+H) and individual single strand efficiencies (L/H).

*b* Calculated as 1 - (calculated strand efficiency).

locus in which mutations confer resistance to 2 × 10^{-5} M amethopterin (SICARD 1964). These mutations are listed in Table 1 together with their map position and their DNA sequence. Spontaneous revertants to wild type occur and can be selected (SICARD 1964), but frequently are not true revertants; they are referred to as rev (Table 1). The str41 mutation which confers resistance to 2 mg/ml of streptomycin is used as reference donor marker for purposes of comparison of marker transformation efficiency (see below).

Marker transformation efficiency is defined as the ratio of transformants for a particular donor marker to transformants for the reference str41 donor marker. The overall transforming efficiency of a marker is likely to reflect the sum of individual strand integration efficiencies (GABOR and HOTCHKISS 1966). These can be measured using in vitro constructed complementary heteroduplexes, H wild type/L mutant and H mutant/L wild type (where H and L refer to the buoyant density of polyuridylate-polyguanylate-complexed strands in CsCl density gradients) (ROGER 1977; CLAVERYS, ROGER and SICARD 1980). L/H ratio allows a comparison of individual strand efficiencies for a given donor marker.

Repair efficiency has been tentatively calculated by taking into account the overall transforming efficiency and individual single strand efficiencies. High efficiency markers such as amiA142 or amiA258 exhibit an efficiency close to 2 as compared to our reference marker str41, this overall transforming efficiency being contributed equally by each strand (CLAVERYS, ROGER and SICARD 1980). Such markers are considered to escape repair by the Hex system. In contrast, low efficiency markers (LE, resulting from transition mutations; CLAVERYS et al. 1981, 1982, 1983), such as amiA9, exhibit an efficiency close to 0.1, both strands being equally poorly efficient. Thus, individual strand efficiencies are 5% of that observed for HE markers and consequently each strand (i.e., each LE mismatch) is 95% repaired. For a major portion of intermediate efficiency markers (IE) such as amiA141, amiA150 or amiA144, one mismatch is 90% repaired (Table 2). The complementary mismatch escapes almost completely repair (it is not sure whether calculated repair efficiencies close to 20% represent real repair; indeed, the use of L/H ratio to calculate strand efficiency may lead to overestimate the contribution of the less efficient strand which is unavoidably contaminated by a small amount of the complementary—and efficient—strand and consequently to underestimate the contribution of the most efficient strand).

Co-repair of flanking markers which would otherwise be refractory to repair is used to monitor the correction of phenotypically silent mismatches such as those between wild-type and amiA43rev sequences. Three IE markers, amiA141, amiA150 and amiA144, have been used as flanking markers;
Mismatch Repair Specificity

TABLE 3
Use of co-correction to estimate repair efficiency of silent mismatches

<table>
<thead>
<tr>
<th>Cross No</th>
<th>Strand</th>
<th>Silent mismatch</th>
<th>Flanking marker</th>
<th>Percent recombination</th>
<th>Transforming efficiency</th>
<th>Calculated strand efficiency</th>
<th>Percent repair</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L</td>
<td>+</td>
<td>amiA141</td>
<td>&lt;1%</td>
<td>0.91 (±0.14)</td>
<td>(0.08)</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>6rev2</td>
<td>+</td>
<td></td>
<td></td>
<td>(0.83)</td>
<td>&lt;5</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>6rev2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>+</td>
<td>amiA141</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>L</td>
<td>+</td>
<td>amiA150</td>
<td>31%</td>
<td>0.94 (±0.14)</td>
<td>(0.15)</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>6rev2</td>
<td>+</td>
<td></td>
<td></td>
<td>(0.79)</td>
<td>&lt;5</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>6rev2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>+</td>
<td>amiA150</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>L</td>
<td>6rev2</td>
<td>+</td>
<td>41%</td>
<td>0.42 (±0.06)</td>
<td>(0.30)</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>+</td>
<td>amiA150</td>
<td></td>
<td></td>
<td>(0.11)</td>
<td>&lt;5</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>6rev2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>+</td>
<td>amiA144</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>L</td>
<td>+</td>
<td>amiA150</td>
<td>31%</td>
<td>0.35 (±0.07)</td>
<td>(0.11)</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>6rev2</td>
<td>+</td>
<td></td>
<td></td>
<td>(0.24)</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>+</td>
<td>amiA150</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>6rev2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>L</td>
<td>+</td>
<td>amiA144</td>
<td>41%</td>
<td>0.87 (±0.13)</td>
<td>(0.76)</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>+</td>
<td>amiA144</td>
<td></td>
<td></td>
<td>(0.11)</td>
<td>&lt;5</td>
</tr>
</tbody>
</table>

the two former mutations escape repair almost completely when introduced on the H strand, and the latter, when introduced on the L strand (Table 2). Correction of any of two complementary (phenotypically silent) mismatches can be inferred from their depressing effect (due to corepair) on the transforming efficiency of such linked IE markers. This phenomenon attributed to long-patch mismatch correction has been called exclusion of donor markers (for review see CLAVERYS and LACKS 1986); it is exemplified in Table 3 for two phenotypically silent mismatches involving amiA6rev2 and wild-type sequences. When amiA6rev2 is transformed with DNA carrying amiA141, amiA150 or amiA144, only the transforming efficiency of the latter is significantly decreased (Table 3, first three crosses). This indicates that the LamiA144/H+ mismatch which normally escapes repair (Table 2) is subject to co-correction initiated at the L+/HamiA6rev2 site (G/G mismatch). A similar effect promoted by this silent G/G mismatch is 95% repaired. The same conclusion can be drawn from quantitative results obtained with amiA150 (Table 3). A basically similar reasoning has been applied for other silent mismatches.

Rapid cloning of DNA fragments carrying amiA mutations has been carried out by vector integration into the chromosome followed by endonucleolytic excision (MEJANE et al. 1981). DNA sequencing procedures have been described (CLAVERYS et al. 1981).

RESULTS AND DISCUSSION

Repair of single substitution mismatches: The repair of 26 different substitution mismatches located throughout the amiA locus was investigated in transformation experiments. The data obtained for seven different transition mutations indicate that the two complementary mismatches, A/C and G/T, are repaired with maximum efficiency (90–95%, Table 4). Fine G:C to C:G transversions have been studied (Table 5). In each case we conclude that the G/G mismatch is repaired out as efficiently as both transition mismatches, whereas the complementary C/C mismatch escapes repair. The picture is more complex for the two other transversion mismatches. Eight crosses involving A:T to C:G mutational changes have been studied (Table 6). It appears that the A/G mismatch is not—or only slightly—repaired (from 0 to 20%); the complementary C/T mismatch is most frequently repaired with an efficiency ranging from 65 to 90%, with the exception of the mismatch at position 4834 (40% repair). It should be pointed out that some C/T mismatches are corrected as efficiently as transition
TABLE 4
Repair of single transition (A:T to G:C) mismatches in various positions along the amiA locus

<table>
<thead>
<tr>
<th>Mismatch position (cross)*</th>
<th>Flanking marker</th>
<th>Complementary mismatches</th>
<th>Transformation efficiency (L/H)%</th>
<th>Repair efficiency *</th>
</tr>
</thead>
<tbody>
<tr>
<td>2558 A/+</td>
<td>No</td>
<td>(+)C (m)T</td>
<td>0.16</td>
<td>92</td>
</tr>
<tr>
<td>(9+/)</td>
<td></td>
<td>(m)A (+)G</td>
<td>0.09 (ref. 1)</td>
<td>91</td>
</tr>
<tr>
<td>3373 (124+/)</td>
<td>No</td>
<td>(+)G (m)A</td>
<td>0.14</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)T (+)C</td>
<td>0.15 (ref. 4)</td>
<td>95</td>
</tr>
<tr>
<td>3374 (6+/)</td>
<td>No</td>
<td>(+)G (m)A</td>
<td>0.15</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)T (+)C</td>
<td>0.15 (ref. 3)</td>
<td>94</td>
</tr>
<tr>
<td>3374 (rev2/rev3)</td>
<td>Yes</td>
<td>(2)C (3)T</td>
<td>0.86</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3)A (2)G</td>
<td>0.86 (ref. 3)</td>
<td>95</td>
</tr>
<tr>
<td>3381 (141/rev10)</td>
<td>No</td>
<td>(10)C (m)T</td>
<td>0.20</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)A (10)G</td>
<td>0.20 (ref. 4)</td>
<td>90</td>
</tr>
<tr>
<td>3383 (+/rev100)</td>
<td>Yes</td>
<td>(+)C (100)T</td>
<td>6.09 (144)*</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(100)A (+)G</td>
<td>0.28 (150)*</td>
<td>95</td>
</tr>
<tr>
<td>4834 (3/rev1)</td>
<td>No</td>
<td>(1)C (m)T</td>
<td>0.14</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)A (1)G</td>
<td>0.14 (ref. 3)</td>
<td>90</td>
</tr>
</tbody>
</table>

* The two sequences involved in the cross are indicated between parentheses as follows: (+) refers to wild type, (number) to amiA mutation and (rev) followed by a number to amiA/rev mutation.

Footnotes a-g and reference s as in Table 4. See also Table 2, first line.

TABLE 5
Repair of single transversion (G:C to C:G) mismatches in various positions along the amiA locus

<table>
<thead>
<tr>
<th>Mismatch position (cross)*</th>
<th>Flanking marker</th>
<th>Complementary mismatches</th>
<th>Transformation efficiency (L/H)%</th>
<th>Repair efficiency *</th>
</tr>
</thead>
<tbody>
<tr>
<td>2558 (+/rev10)</td>
<td>Yes</td>
<td>(+)C (81)G</td>
<td>0.93 (144)*</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(81)C (+)G</td>
<td>0.25 (144)*</td>
<td>95</td>
</tr>
<tr>
<td>3373 (6/rev4)</td>
<td>No</td>
<td>(4)C (m)G</td>
<td>0.86</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)C (4)G</td>
<td>0.13 (ref. 3)</td>
<td>90</td>
</tr>
<tr>
<td>3374 (rev2/rev2)</td>
<td>Yes</td>
<td>(+)G (2)C</td>
<td>0.65</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2)G (+)C</td>
<td>0.65 (ref. 3)</td>
<td>95</td>
</tr>
<tr>
<td>3383 (141/rev10)</td>
<td>No</td>
<td>(+)C (m)G</td>
<td>0.88</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)C (0.99)</td>
<td>0.88 (ref. 3)</td>
<td>93</td>
</tr>
<tr>
<td>4834 (+/rev1)</td>
<td>Yes</td>
<td>(+)G (1)C</td>
<td>0.55 (144)*</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)G (+)C</td>
<td>1.06 (150)*</td>
<td>95</td>
</tr>
</tbody>
</table>

TABLE 6
Repair of single transversion (A:T to C:G) mismatches in various positions along the amiA locus

<table>
<thead>
<tr>
<th>Mismatch position (cross)*</th>
<th>Flanking marker</th>
<th>Complementary mismatches</th>
<th>Transformation efficiency (L/H)%</th>
<th>Repair efficiency *</th>
</tr>
</thead>
<tbody>
<tr>
<td>2505 (22+/)</td>
<td>No</td>
<td>(+)G (m)T</td>
<td>1.0</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)A (+)C</td>
<td>0.27 (ref. 1)</td>
<td>79</td>
</tr>
<tr>
<td>2658 (9/rev8)</td>
<td>No</td>
<td>(8)G (m)T</td>
<td>1.14</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)A (81)C</td>
<td>0.42 (ref. 1)</td>
<td>66</td>
</tr>
<tr>
<td>2660 (9/rev1)</td>
<td>No</td>
<td>(1)C (m)A</td>
<td>1.0</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)T (1)G</td>
<td>7.1 (ref. 1)</td>
<td>88</td>
</tr>
<tr>
<td>3346 (36+/)</td>
<td>No</td>
<td>(+)C (m)A</td>
<td>1.12</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)T (+)G</td>
<td>3.3 (ref. 1)</td>
<td>74</td>
</tr>
<tr>
<td>3374 (6/rev2)</td>
<td>No</td>
<td>(2)C (m)A</td>
<td>1.02</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)T (2)G</td>
<td>9.3 (ref. 3)</td>
<td>90</td>
</tr>
<tr>
<td>3374 (+/rev3)</td>
<td>Yes</td>
<td>(+)G (3)T</td>
<td>0.86</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3)A (+)C</td>
<td>0.86 (ref. 3)</td>
<td>95</td>
</tr>
<tr>
<td>3383 (141/rev10)</td>
<td>No</td>
<td>(100)T (m)G</td>
<td>1.2</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)C (100)A</td>
<td>1.2 (14Y)</td>
<td>92</td>
</tr>
<tr>
<td>4834 (3+/)</td>
<td>No</td>
<td>(+)G (m)T</td>
<td>1.45</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(m)A (+)C</td>
<td>0.7 (ref. 1)</td>
<td>40</td>
</tr>
</tbody>
</table>

Footnotes a-g and reference s as in Table 4. See also Table 2, first line.

The following tentative mismatch ranking as a function of decreasing repair efficiency by the Hex system can be proposed: G/T = A/C = G/G (maximum repair: 90-95%) > C/T (mostly 75 to 90% repair) > A/A (from 50 to 90% repair) > T/T (50-65% repair) > A/G (from 0 to 20% repair) > C/C.

Influence of the base composition of the neighboring sequence: In the case of the mismatch repair system of E. coli, (Mut system), it has been suggested that repair of a given mismatch increases with increasing G:C content in the neighboring nucleotide sequence (Jones, Wagner and Radman 1987). This
Table 7

Repair of single transversion (A:T to T:A) mismatches in various positions along the amiA locus

<table>
<thead>
<tr>
<th>Mismatch position (cross)</th>
<th>Flanking marker</th>
<th>Complementary mismatches</th>
<th>Transformation efficiency (L/H)</th>
<th>Repair efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A/A (%)</td>
<td>T/T (%)</td>
</tr>
<tr>
<td>2660</td>
<td>No</td>
<td>(36)T (m)A</td>
<td>0.65</td>
<td>86</td>
</tr>
<tr>
<td>(9/rev36)</td>
<td></td>
<td>(m)T (36)A</td>
<td>(0.28)</td>
<td></td>
</tr>
<tr>
<td>3166</td>
<td>No</td>
<td>(+)T (m)A</td>
<td>0.89</td>
<td>ND</td>
</tr>
<tr>
<td>(54/+)</td>
<td></td>
<td>(m)T (+)A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3374</td>
<td>No</td>
<td>(3)T (m)A</td>
<td>0.47</td>
<td>87</td>
</tr>
<tr>
<td>(6/rev3)</td>
<td></td>
<td>(m)T (3)A</td>
<td>(0.37)</td>
<td>ref. 3</td>
</tr>
<tr>
<td>3382</td>
<td>No</td>
<td>(32)T (m)A</td>
<td>0.98</td>
<td>47</td>
</tr>
<tr>
<td>(141/rev32)</td>
<td></td>
<td>(m)T (32)A</td>
<td>(1.2)</td>
<td></td>
</tr>
</tbody>
</table>

Footnotes a–f and reference 3 as in Table 4.

ND, not determined.

was most apparent for the two transversion mismatches A/G and C/T when considering a region including 4 bp on either side of the mismatch (Figure 1A, bottom). To determine whether this is true for S. pneumoniae, we present repair efficiency of A/G and C/T mismatches by the Hex system as a function of the G:C content in the surrounding nucleotide sequence. We have considered a region including 4 bp (Figure 1A, top) or 10 bp (Figure 1B, top) on either side of the mismatch. In addition to our results obtained with nine different A:T to C:G transversions in the amiA locus, we included two malM markers (malM594, see above, and malM506, from the overall transforming efficiency of which it has been deduced that A/G is corrected) (LACKS, DUNN and GREENBERG 1982). There is no obvious correlation between repair efficiency by the Hex system and the G:C content of nucleotide sequences surrounding the mismatch. It is worth noting that at the very same site in the amiA locus (position 4834) where a C/T mismatch is only partly repaired (Table 6) a G/G mismatch is very efficiently corrected (Table 4b). Thus, although mismatch repair efficiency is indeed influenced by context, it appears that variations in repair efficiency from site to site result from more subtle changes than the overall G:C richness.

The difference between the Hex and Mut repair systems may result from some adaptation of the Hex mismatch-recognition enzyme(s) to the rather low G:C content of S. pneumoniae (close to 40% G:C) (EHRLICH 1984).

![Figure 1](image-url)

**Figure 1.**—Repair efficiency as a function of G:C content in the neighboring nucleotide sequence. The mismatches are C/T (A) and A/G (B). The region considered includes 4 (A) and 10 (B) bp on either side of the mismatch. Upper panel, the data for S. pneumoniae are from Tables 1 and 6; we have included the malM506 and malM594 mutations (arrows) from LACKS, DUNN and GREENBERG 1982. Due to the mode of calculation (see MATERIALS AND METHODS), repair efficiency cannot be higher than 95%. Lower panel, data for E. coli (from JONES, WAGNER and RADMAN 1987; percent pure infective centers replaces repair efficiency. Maximum repair results in 95–99% pure infective centers, whereas 10% pure infective centers are obtained in the absence of repair).
et al. 1987). The Hex system may thus be more efficient than the Mut system in correcting mismatches in AT-rich regions. Alternatively, it remains possible that the correlation observed for E. coli is only fortuitous, due to the limited number of mutations analyzed. Indeed, it has been recently reported that the rule that the repair efficiency increases with increasing G:C content does not apply in all instances (Lu and Chang 1988).

**Correspondence between S. pneumoniae and E. coli repair systems:** Both systems exhibit striking similarity for mismatch recognition specificities. Indeed, the two systems efficiently process transition mismatches and the G/T transversion mismatch (for E. coli, see Kramer, Kramer and Fritz 1984; Dohet, Wagner and Radman 1985; Jones, Wagner and Radman 1987; Lu and Chang 1988). Both systems also repair the two complementary A/A and T/T transversion mismatches with variable efficiency. However, the three remaining transversion mismatches could be processed differently. Concerning the C/C mismatch which has been found so far to escape repair by the Hex system (see above), there is one report of intermediate repair efficiency by the Mut system (Dohet, Wagner and Radman 1985).

Comparison of repair efficiencies for A/G and C/T mismatches by the Hex system of S. pneumoniae (Figure 1, top) and by the Mut system of E. coli (Figure 1, bottom) suggests that the Mut system exhibits a similar repair efficiency for both mismatches, whereas the Hex system corrects more efficiently C/T than A/G. As this comparison involves one of the largest samples examined in both systems, this difference may be significant and reflect differential affinities of Hex and Mut mismatch-recognition enzyme(s) for these mismatches. It should be pointed out that only one instance of repair for the A/G mismatch has been reported so far (corresponding to the malM506 marker, see above). It would be interesting to check if this repair is Hex-dependent or attributable to another mismatch repair system similar to the one recently described in E. coli (Lu and Chang 1988).

**Repair of multiple-substitution mismatches:** Eleven different double-substitution mismatches were investigated (Table 8). These are either adjacent or separated by one or two Watson-Crick base pairs. In crosses 1 to 7, each double mismatch contains a transition repair. Only in cross 4 is repair efficiency of the transition mismatch notably reduced by the presence of another substitution mismatch. In this cross the C/C mismatch completely inhibits repair of the flanking G/T mismatch. In cross 3 the double CA/CC (position 3373:3374) is recognized as efficiently as its transition mismatch component. It should be noticed that the very same double mismatch partially escaped recognition when a C/C mismatch was introduced 8 bp away (at position 3383; Claverys et al. 1985), possibly indicating local destabilization of the DNA helix (see further discussion below).

Double-mismatches in crosses 8 to 11 contain transversion mismatches. In crosses 8 to 10 the component containing the C/C mismatch is either partially repaired (crosses 8 and 10) or subject to very limited repair (50% repair, cross 9). We tentatively attribute this to the presence of the C/C mismatch which has, so far, been found to be completely refractory to repair when present alone. The multiple-mismatches in the last cross is complex, containing three substitutions and a 1 bp addition-deletion. The component containing two adjacent A/G mismatches is almost completely refractory to repair.

**Further discussion:** Striking similarities exist between mismatch recognition specificities of the Mut system of E. coli and the Hex system of S. pneumoniae. This is not totally surprising in light of the efficient role of both mismatch repair systems in postreplication correction of biosynthetic errors (review see Claverys and Lacks 1986). However, we have also pointed out a number of differences between the two mismatch repair systems. Taking into account the influence of neighboring sequences on mismatch recognition specificity (see above), these differences may only reflect the still limited available information. Direct binding studies with purified mismatch recognition enzymes using identical mismatched DNA heteroduplexes, as initiated with the mutS-encoded protein (Su and Modrich 1986), should help clarify this point. If both the Hex system of S. pneumoniae and the Mut system of E. coli have been evolved to correct biosynthetic errors, the differences in mismatch recognition specificities discussed above would imply that different type of biosynthetic errors are produced in these bacteria. From in vitro studies carried out with DNA polymerase III holoenzyme from E. coli (Fersht and Knill-Jones 1981), it would be expected that mismatches G/T, A/C, A/A, and A/G frequently occur as errors in DNA synthesis compared with C/T. This fits better with mismatch recognition specificity of the Mut system of E. coli than with that of the Hex system of S. pneumoniae. Indeed, we observed rather efficient correction of C/T for the latter, together with inefficient correction of A/G. One could imagine that A/G is an unfrequent biosynthetic error in S. pneumoniae. It would be interesting to analyze which types of errors are made by the pneumococcal replication machinery. Alternatively, an A/G-specific mismatch repair system may function in this organism.

A comprehensive discussion of the structures of base pair mismatches and their implications for mismatch correction can be found in a recent review (Modrich 1987). It has been suggested that all repairable mismatches can adopt an intrahelical form
TABLE 8
Repair of multiple-substitution mismatches in various positions along the ami4 locus

<table>
<thead>
<tr>
<th>Cross No.</th>
<th>Mismatch position (cross)</th>
<th>Flanking markers</th>
<th>Complementary mismatches</th>
<th>Transformation efficiency</th>
<th>Repair efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>2nd</td>
<td>1st (%)</td>
</tr>
<tr>
<td>1</td>
<td>2658–2660 (+/rev36)</td>
<td>Yes</td>
<td>(+)CaA</td>
<td>(36)TaT</td>
<td>0.55 (144)</td>
</tr>
<tr>
<td>2</td>
<td>3373–3374 (124/rev2)</td>
<td>No</td>
<td>(2)GC</td>
<td>(m)AG</td>
<td>0.11</td>
</tr>
<tr>
<td>3</td>
<td>3373–3374 (+/rev4)</td>
<td>Yes</td>
<td>(+)GG</td>
<td>(4)CA</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>3381–3383 (+/rev10)</td>
<td>Yes</td>
<td>(+)TaC</td>
<td>(10)CaG</td>
<td>0.83 (144)</td>
</tr>
<tr>
<td>5</td>
<td>2658–2660 (+/rev1)</td>
<td>Yes</td>
<td>(+)CaA</td>
<td>(1)TaG</td>
<td>0.83</td>
</tr>
<tr>
<td>6</td>
<td>3373–3374 (124/rev3)</td>
<td>No</td>
<td>(3)GT</td>
<td>(m)AG</td>
<td>0.12</td>
</tr>
<tr>
<td>7</td>
<td>3373–3374 (124/rev4)</td>
<td>No</td>
<td>(4)CA</td>
<td>(m)AG</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>3373–3374 (rev3/rev4)</td>
<td>Yes</td>
<td>(3)GT</td>
<td>(4)CA</td>
<td>—</td>
</tr>
<tr>
<td>9</td>
<td>3382–3383 (+/rev32)</td>
<td>Yes</td>
<td>(+)AC</td>
<td>(32)TG</td>
<td>0.74 (144)</td>
</tr>
<tr>
<td>10</td>
<td>3373–3374 (rev3/rev4)</td>
<td>Yes</td>
<td>(2)GC</td>
<td>(4)CA</td>
<td>—</td>
</tr>
<tr>
<td>11</td>
<td>2549—2551 (+/299)</td>
<td>No</td>
<td>(+)GG</td>
<td>(m)TTC</td>
<td>1.1</td>
</tr>
</tbody>
</table>

* Refers to mismatches shown in "complementary mismatches," first column (1st) or second column (2nd).

Footnotes b-g and references 1-3 as in Table 4.

We are grateful to Michael Chandler for editing the manuscript.

LITERATURE CITED


GABOR, M., and R. D. HOTHKISS, 1966 Manifestation of linear


JONES, M., R. WAGNER and M. RADMAN, 1987 Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. Genetics 115: 605-610.


Communicating editor: N. KLECKNER