Repair of a Mismatch Is Influenced by the Base Composition of the Surrounding Nucleotide Sequence

Madeleine Jones, Robert Wagner and Miroslav Radman

Institut Jacques Monod, Tour 43, 2 Place Jussieu, 75251 Paris-Cedex 05, France

Manuscript received October 27, 1986
Revised copy accepted January 12, 1987

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*. A series of transition (G:T and A:C) and transversion (G:A and C:T) mismatches located throughout most of the bacteriophage λ *cI* gene has been examined. The results suggest that the transition mismatches are generally better repaired than the transversion mismatches and that, at least for the transversion mismatches studied, repair efficiency increases with increasing G:C content in the neighboring nucleotide sequence. This specificity of the *E. coli* mismatch repair system can account, in part, for the similar frequencies of base substitution mutations throughout the *E. coli* genome.

THE *Escherichia coli* mismatch repair system is able to recognize noncomplementary base pairs in DNA and acts to replace misincorporated bases (for review see RADMAN and WAGNER [1986]). Regions of DNA in which GATC sequences are fully adenine methylated appear to be refractory to mismatch repair (PUKKILA *et al.* 1983; WAGNER *et al.* 1984) and it appears to be the transient undermethylation of newly synthesized GATC sequences in the region immediately following the replication fork which allows mismatch repair to operate only on newly synthesized strands and thereby remove replication errors (RADMAN and WAGNER 1986; PUKKILA *et al.* 1983; WAGNER *et al.* 1984; WAGNER and MESELSON 1976; RADMAN *et al.* 1980).

The *E. coli* mismatch repair system does not recognize and/or repair all mismatches with equal efficiency (DOHET, WAGNER and RADMAN 1985; KRAMER, KRAMER and FRITZ 1984). Both transition mismatches (G:T and C:A) are readily recognized and repaired, whereas three of the six transversion mismatches are not (DOHET, WAGNER and RADMAN 1985). The finding that two different mismatches at the same site in the DNA are repaired with different efficiencies indicates that context cannot be the sole determinant of mismatch repair efficiency. However, the results of studies of mismatch repair in *Streptococcus pneumoniae* (LACKS, DUNN and GREENBERG 1982; CLAVERYS *et al.* 1983) and reversion frequencies in the *trpA* gene of *E. coli* (CHOY and FOWLER 1985) suggest that mismatch repair efficiency is influenced, at least to some extent, by context.

In an attempt to determine which elements of context influence mismatch repair in *E. coli*, we have examined a series of transition and transversion mismatches located throughout most of the bacteriophage λ *cI* gene. The results suggest that, for a given mismatch, repair efficiency increases with increasing G:C content in the neighboring nucleotide sequence.

MATERIALS AND METHODS

Phage strains used in these experiments are listed in Table 1. Strand preparation, annealing, transfection, and scoring procedures have been described (DOHET, WAGNER and RADMAN 1985).

The majority of DNA used in these experiments was prepared from phages grown in a *dam* (deficient in adenine methylation) strain, GM33 (MARINUS and MORRIS, 1975), and, therefore, is unmethylated at GATC sequences.

Fully adenine-methylated DNA strands were prepared from phages grown in a *dam*-methylase overproducer strain (HERMAN and MODRICH 1981).

Bacteria used in transfection experiments, *E. coli* C600 and a mismatch repair deficient (*mutL*) derivative of it, have been described (DOHET, WAGNER and RADMAN 1985).

RESULTS AND DISCUSSION

The repair of individual transition mismatches (G:T and A:C) from 12 different locations in the bacteriophage λ *cI* gene and individual transversion mismatches (G:A and C:T) from 6 different locations was studied in transfection experiments using heteroduplexes with single base pair mismatches. The data (Table 2) are presented as percentages of mixed infective centers, *i.e.*, those infective centers containing phages derived from both strands of a single heteroduplex, obtained following transfection of mismatch repair proficient bacteria. The extent or efficiency of mismatch repair is inversely related to the fraction of mixed infective centers (DOHET, WAGNER and RADMAN 1985). To confirm that the heteroduplex preparations were not contaminated to an unusual or unequal extent by homoduplex molecules, which

TABLE 1

Bacteriophage λ strains

Mutation	Mutation[a] position	Base change	
		Transition	Transversion
BP81[b]	2	T:A → C:G	
UA9	6		C:G → A:T
SP57[b]	15		C:G → A:T
GM752[b,c]	31	C:G → T:A	
UV23	43	C:G → T:A	
LP206	43		C:G → A:T
BL80	58	T:A → C:G	
SP39	58		T:A → G:C
GM751[b,c]	98	C:G → T:A	
BL51	101	T:A → C:G	
UA77	133		C:G → A:T
MG13[c]	161	C:G → T:A	
SP44	208	C:G → T:A	
SP40	208		C:G → A:T
BU10[b]	221	T:A → C:G	
BP138[b]	311	T:A → C:G	
BL46	403	T:A → C:G	
BU15[b]	632	T:A → C:G	

All phage strains are from the laboratory of Dr. F. HUTCHINSON (Yale University) where the *cI* mutants were sequenced (WOOD and HUTCHINSON 1984). Except where noted, phage mutants were derived from λ*CI857 ind*⁻ *Oam29* which was used to provide "c^+" strands for heteroduplexes.

[a] Number of base pairs from the start of the *cI* PRM transcript.

[b] Phage made O^+ by recombination. "c^+" strands for heteroduplexes from λ*cI857 ind*⁻, except for *GM752* and *GM751* (see below).

[c] Mutants derived from λ *b515 b519 cI857 ind*⁻ *Oam29* Tn*2* (WOOD and HUTCHINSON 1984) which was used to provide "c^+" strands for heteroduplexes.

would lower the percentage of mixed infective centers or make comparisons between preparations impossible, control transfections were performed using mismatch repair deficient (*mutL*) bacteria. The fraction of mixed infective centers obtained from these transfections ranged from 65% to 90% (data not shown; see DOHET, WAGNER and RADMAN 1985).

The data confirm that neighboring nucleotide sequence or location within a gene cannot be the sole determinant of mismatch repair efficiency. At each of the three sites where we can prepare four different mismatches (bp 43, 58 and 208), a wide range of mismatch repair efficiencies is observed, with the G:A and C:T transversion mismatches consistently less well repaired than transition mismatches. In addition, for the majority of transition mismatch pairs, the G:T mismatch is better repaired than the A:C mismatch.

It is clear that there is a considerable range of repair efficiencies observed for the same mismatch in different locations. This variation is much greater than the variation observed between different transfections with the same heteroduplex preparation or between different DNA preparations of the same heteroduplex (data not shown). To confirm that the differences observed in repair efficiencies are not some artifact of using unmethylated DNA or a result of differences in the extent of mismatch-stimulated killing, which occurs only in unmethylated DNA (DOUTRIAUX, WAGNER and RADMAN, 1986), heteroduplexes with some of the mismatches were prepared with one strand adenine-methylated. The fractions of mixed infective centers obtained from transfections with these hemimethylated heteroduplexes (Table 2) are not significantly different from those obtained with unmethylated heteroduplexes. Thus, the variation observed in the extent of mismatch repair is unlikely to be an artifact of the experimental system.

There is no obvious dependence of mismatch repair on the distance of the mismatch from the beginning of the *cI* gene or from a GATC sequence (GATC sequences within or flanking the region studied are located −167, 535 and 639 bp from the beginning of the *cI* gene (SANGER *et al.* 1982).

It may be that, while different mismatches at one site are repaired with different efficiencies, the mismatch repair efficiency is also influenced by the base composition of the nucleotide sequence surrounding the mismatch. This hypothesis was tested by comparing the fraction of mixed infective centers to the G:C content of segments of DNA of varying length on either side of the mismatch. When the region considered includes 4 bp on either side of the mismatch, it is evident that the repair efficiency of the two transversion mismatches studied increases (*i.e.*, the fraction of mixed infective centers decreases) with increasing G:C content (Figure 1). This tendency is less evident when the region considered is smaller than 4 bp on either side of the mismatch, but is still apparent when the region is as large as 10 bp on either side of the mismatch (data not shown).

The pattern is less striking with the transition mismatches, although there may be a tendency toward increasing repair efficiency with increasing G:C content of the nucleotide neighborhood (Figure 2). The fact that transition mismatches are frequently well repaired, *i.e.*, transfections yield a small fraction of mixed infective centers, may make it difficult to detect a clear pattern.

The finding that increasing the number of G:C base pairs close to a mismatch appears to increase the extent of mismatch repair suggests that mismatch repair operates most efficiently in regions of stable double helix. The results of an NMR study of the structure of synthetic oligonucleotides in solution, consisting of the 5 bp on either side of position 43 or position 208 and containing either a G:A or G:T mismatch, indicate that the least-repaired mismatch, *i.e.*, the G:A mismatch at position 208, can exist in an extrahelical configuration (FAZAKERLEY *et al.* 1986). The three better-repaired mismatches are found only in a helical form. The suggestion that the intrahelical mismatches may be preferentially repaired was made earlier (RADMAN *et al.* 1985; WERNTGES *et al.* 1986).

TABLE 2

Mixed infective centers derived from heteroduplexes of bacteriophage λ DNA with single defined mismatches in various positions along the *cI* gene

Mismatch position	DNA sequence		Unmethylated DNA				Hemimethylated DNA			
			Transitions		Transversions		Transitions		Transversions	
			Mixed (c/c^+)	Total no.	Mixed (c/c^+)	Total no.	Mixed (c/c^+)	Total no.	Mixed (c/c^+)	Total no.
	ATTTAACGTATGAGCACAAAA	(+)								
	TAAATTGCATACTCGTGTTTT	(+)								
2	-T-	(+)	9%	509						
	-G-	(*c*)								
2	-C-	(*c*)	10%	609						
	-A-	(+)								
	AACGTATGAGCACAAAAAAGA	(+)								
	TTGCATACTCGTGTTTTTTCT	(+)								
6	-A-	(*c*)			65%	344				
	-G-	(+)								
6	-C-	(+)			38%	299				
	-T-	(*c*)								
	GCACAAAAAAGAAACCATTAA	(+)								
	CGTGTTTTTTCTTTGGTAATT	(+)								
15	-G-	(+)			73%	980				
	-A-	(*c*)								
15	-T-	(*c*)			80%	496				
	-C-	(+)								
	ATTAACACAAGAGCAGCTTGA	(+)								
	TAATTGTGTTCTCGTCGAACT	(+)								
31	-G-	(+)	5%	233						
	-T-	(c)								
31	-A-	(*c*)	20%	241						
	-C-	(+)								
	GCAGCTTGAGGACGCACGTCG	(+)								
	CGTCGAACTCCTGCGTGCAGC	(+)								
43	-G-	(+)	5%	269						
	-T-	(*c*)								
43	-A-	(*c*)	15%	234						
	-C-	(+)								
43	-G-	(+)			38%	249				
	-A-	(*c*)								
43	-T-	(*c*)			30%	250				
	-C-	(+)								
	ACGTCGCCTTAAAGCAATTTA	(+)								
	TGCAGCGGAATTTCGTTAAAT	(+)								
58	-G-	(*c*)	9%	211						
	-T-	(+)								
58	-A-	(+)	29%	222						
	-C-	(*c*)								
58	-A-	(+)			37%	252			42%	206
	-G-	(*c*)								
58	-C-	(*c*)			50%	312			54%	225
	-T-	(+)								
	CTTGGCTTATCCCAGGAATCT	(+)								
	GAACCGAATAGGGTCCTTAGA	(+)								
98	-T-	(*c*)	6%	263						
	-G-	(+)								
98	-C-	(+)	32%	282						
	-A-	(*c*)								
101	-G-	(*c*)	4%	337						
	-T-	(+)								

TABLE 2

Continued

Mismatch position	DNA sequence		Unmethylated DNA				Hemimethylated DNA			
			Transitions		Transversions		Transitions		Transversions	
			Mixed (c/c^+)	Total no.	Mixed (c/c^+)	Total no.	Mixed (c/c^+)	Total no.	Mixed (c/c^+)	Total no.
101	-A-	(+)	16%	336						
	-C-	(c)								
	GGGGATGGGGCAGTCAGGCGT	(+)								
	CCCCTACCCCGTCAGTCCGCA	(+)								
133	-A-	(c)			35%	369			41%	222
	-G-	(+)								
133	-C-	(+)			19%	295			18%	218
	-T-	(c)								
	TTATTTAATGGCATCAATGCA	(+)								
	AATAAATTACCGTAGTTACGT	(+)								
161	-G-	(+)	25%	325						
	-T-	(c)								
161	-A-	(c)	33%	640						
	-C-	(+)								
	TGCAAAAATTCTCAAAGTTAG	(+)								
	ACGTTTTTAAGAGTTTCAATC	(+)								
208	-T-	(c)	34%	215			31%	217		
	-G-	(+)								
208	-C-	(+)	40%	234			35%	205		
	-A-	(c)								
208	-A-	(c)			83%	235			90%	239
	-G-	(+)								
208	-C-	(+)			82%	320			89%	242
	-T-	(c)								
	AAAGTTAGCGTTGAAGAATTT	(+)								
	TTTCAATCGCAACTTCTTAAA	(+)								
221	-T-	(+)	10%	546						
	-G-	(c)								
221	-C-	(c)	13%	512						
	-A-	(+)								
	GAGTATGAGTACCCTGTTTTT	(+)								
	CTCATACTCATGGGACAAAAA	(+)								
311	-G-	(c)	3%	544			4%	284		
	-T-	(+)								
311	-A-	(+)	4%	526			6%	323		
	-C-	(c)								
	AAGCACAACCAAAAAAGCCAG	(+)								
	TTCGTGTTGGTTTTTTCGGTC	(+)								
403	-G-	(c)	2%	531						
	-T-	(+)								
403	-A-	(+)	1%	540						
	-C-	(c)								
	AACCCACAGTACCCAATGATC	(+)								
	TTGGGTGTCATGGGTTACTAG	(+)								
632	-G-	(c)	6%	504			6%	241		
	-T-	(+)								
632	-A-	(+)	6%	507			4%	317		
	-C-	(c)								

Upper strand of all sequences is the *r*-strand. The wild-type sequence (c^+) is shown above the mismatches. Only the fraction of mixed infective centers is shown, which represents unrepaired heteroduplexes. Pure infective centers (c and c^+) showed random strand repair of mismatches on unmethylated DNA, as reported (DOHET, WAGNER and RADMAN 1985) and preferential mismatch repair on the unmethylated strand in hemimethylated heteroduplexes.

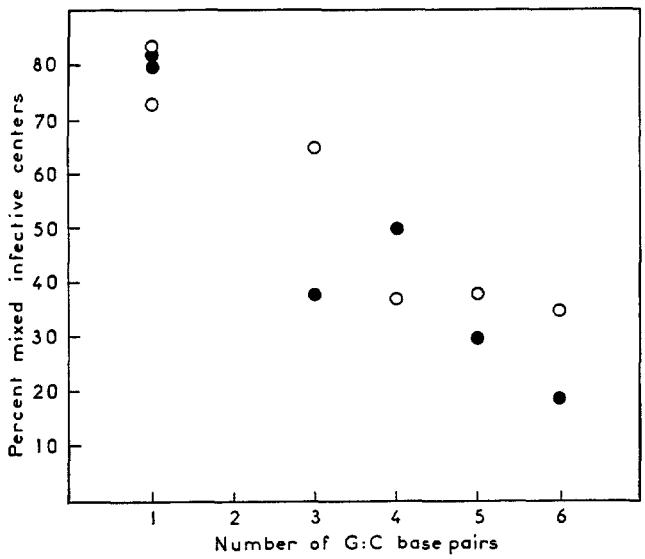

FIGURE 1.—Percent mixed infective centers as a function of the number of G:C base pairs in the region surrounding the mismatch: transversion mismatches. The mismatches are G:A (open symbols) and C:T (closed symbols). The region considered includes four base pairs on either side of the mismatch. The data are from Table 2.

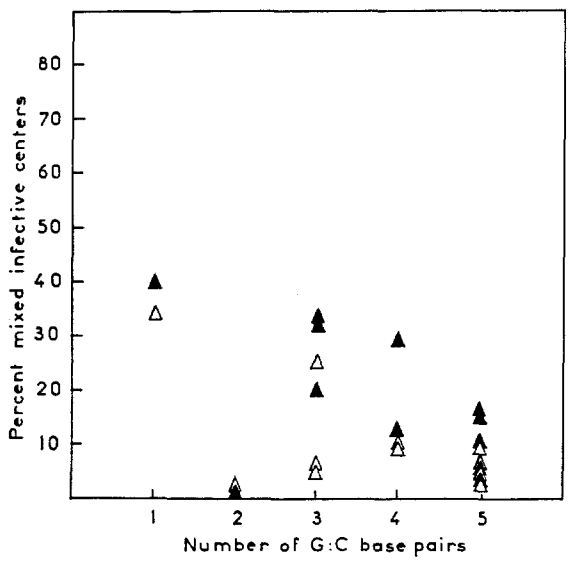

FIGURE 2.—Percent mixed infective centers as a function of the number of G:C base pairs in the region surrounding the mismatch: transition mismatches. The mismatches are G:T (open symbols) and A:C (closed symbols). The region considered includes four base pairs on either side of the mismatch. The data are from Table 2.

Recent work on crystal structures of mismatch-containing oligonucleotides (HUNTER *et al.* 1986) indicates that the changes in the glycosidic bond angles, in the base stacking and in the exposure of functional groups into the major and minor grooves, could all contribute to the recognition of mismatches by the mismatch repair enzymes; such structural signals could be lost upon opening up the double helix in the case of unrepaired mismatches (FAZAKERLEY *et al.* 1986).

It has been suggested that, because mismatch repair deficient mutants of *E. coli* are strong mutators, most mismatches that arise during replication are repaired (WAGNER *et al.* 1984; DOHET, WAGNER and RADMAN 1985; KRAMER, KRAMER and FRITZ 1984). It may be that the polymerase infrequently attempts to incorporate incorrect bases in A:T-rich regions of DNA where mismatch repair functions poorly. However, there is evidence suggesting that the proofreading exonuclease activity associated with the T4 phage polymerase is most active in A:T-rich regions (PLESS and BESSMAN 1983; PETRUSKA and GOODMAN 1985). Thus, the findings that mismatches that can exist in extrahelical configuration are poorly repaired and that the T4 phage DNA polymerase makes fewer mistakes in A:T-rich regions than in G:C-rich regions (PETRUSKA and GOODMAN 1985) suggest that, when mismatches that are poorly repaired by the mismatch repair system are formed in the course of DNA replication, they may disrupt base pairing between the newly synthesized and template strands such that the incorrect bases are most often removed by the single-strand-specific "proofreading" exonuclease associated with the DNA polymerase. It may be that the net effect of the differences in specificity between mismatch repair and polymerase proofreading is that all substitution mutations arise with similar frequencies throughout the *E. coli* genome (for further discussion, see RADMAN and WAGNER [1986]).

The sequenced mutants used in this study were provided by Drs. R. WOOD and F. HUTCHINSON (Yale University). We thank CHRISTIANE DOHET for providing the results for the mismatches at base pair positions 6 and 15 and Dr. R. D'ARI for critically reading the manuscript. M.J. was supported by fellowships from Institut pour l'Encouragement de la Recherche Scientifique dans l'Industrie et l'Agriculture (Belgium) and Centre National de la Recherche Scientifique (C.N.R.S.). R.W. was supported by senior fellowships from Association pour la Recherche sur le Cancer (A.R.C.). This work was supported by grants from A.R.C. and C.N.R.S. The Institut Jacques Monod is a part of the Centre National de la Recherche Scientifique and of Université Paris 7.

LITERATURE CITED

CHOY, H. E. and R. G. FOWLER, 1985 The specificity of base pair substitutions induced by the *mutL* and *mutS* mutators in *E. coli*. Mutat. Res. **142:** 93–97.

CLAVERYS, J. P., V. MEJEAN, A. M. GASC and A. M. SICARD, 1983 Mismatch repair in *Streptococcus pneumoniae*: relationship between base mismatches and transformation efficiencies. Proc. Natl. Acad. Sci. USA **80:** 5956–5960.

DOHET, C., R. WAGNER and M. RADMAN, 1985 Repair of defined single base-pair mismatches in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **82:** 503–505.

DOUTRIAUX, M. P., R. WAGNER and M. RADMAN, 1986 Mismatch-stimulated killing. Proc. Natl. Acad. Sci. USA **83:** 2576–2578.

FAZAKERLEY, G. V., E. QUIGNARD, A. WOISARD, W. GUSCHLBAUER, G. A. VAN DER MAREL, J. H. VAN BOOM, M. JONES and M. RADMAN, 1986 Structures of mismatched base pairs in DNA and their recognition by the *E. coli* mismatch repair system. EMBO J. 5: 3697–3703.

HERMAN, G. E. AND P. MODRICH, 1981 *Escherichia coli* K12 clones

that overproduce *dam* methylase are hypermutable. J. Bacteriol. **145**: 644–646.

HUNTER, W. N., T. BROWN, N. N. ANAND and O. KENNARD, 1986 Structure of an adenine-cytosine base pair in DNA and its implications for mismatch repair. Nature **320:** 552–555.

KRAMER, B., W. KRAMER and H. J. FRITZ, 1984 Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch repair system of *E. coli*. Cell **38:** 879–887.

LACKS, S. A., J. J. DUNN and B. GREENBERG, 1982 Identification of base mismatches recognized by the heteroduplex-DNA-repair system of *Streptococcus pneumoniae*. Cell **31:** 327–336.

MARINUS, M. G. and N. R. MORRIS, 1975 Pleiotropic effects of a DNA adenine methylation mutation (*dam3*) in *Escherichia coli* K12. Mutat. Res. **28:** 15–26.

PETRUSKA, J. and M. F. GOODMAN, 1985 Influence of neighboring base sequences on DNA polymerase insertion and proofreading fidelity. J. Biol. Chem. **260:** 7533–7539.

PLESS, R. C. and M. J. BESSMAN, 1983 Influence of local nucleotide sequence on substitution of 2-aminopurine for adenine during deoxyribonucleic acid synthesis in vitro. Biochemistry **22:** 4905–4915.

PUKKILA, P. J., J. PETERSON, G. HERMAN, P. MODRICH and M. MESELSON, 1983 Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. Genetics **104:** 571–582.

RADMAN, M. and R. WAGNER, 1986 Mismatch repair in *Escherichia coli*. Annu. Rev. Genet. **20:** 523–538.

RADMAN, M., R. E. WAGNER, B. W. GLICKMAN and M. MESELSON, 1980 DNA methylation, mismatch correction and genetic stability. pp. 121–130. In: *Progress in Environmental Mutagenesis*, Edited by M. ALACEVIC. Elsevier/North Holland, Amsterdam.

RADMAN, M., C. DOHET, M. JONES, M. P. DOUTRIAUX, F. LAENGLE-ROUAULT, G. MAENHAUT-MICHEL and R. WAGNER, 1985 Processing of mispaired and unpaired bases in heteroduplex DNA in *E. coli*. Biochimie **67:** 745–752.

SANGER, F. AND A. R. COULSON, 1982 Nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. **162:** 729–773.

WAGNER, R. and M. MESELSON, 1976 Repair tracts in mismatched DNA heteroduplexes. Proc. Natl. Acad. Sci. USA **73:** 4135–4139.

WAGNER, R., C. DOHET, M. JONES, M. P. DOUTRIAUX, F. HUTCHINSON and M. RADMAN, 1984 Involvement of *Escherichia coli* mismatch repair in DNA replication and recombination. Cold Spring Harbor Symp. Quant. Biol. **49:** 611–615.

WERNTGES, H., G. STEGER, D. RIESNER and H. F. FRITZ, 1986 Mismatches in DNA double strands: thermodynamic parameters and their correlation to repair efficiencies. Nucleic Acids Res. **14:** 3773–3790.

WOOD, R. D. and F. HUTCHINSON, 1984 Non-targeted mutagenesis of unirradiated lambda phage in *Escherichia coli* host cells irradiated with ultraviolet light. J. Mol. Biol. **173:** 293–305.

Communicating editor: J. W. DRAKE