J Mol Evol (1991) 33:125–132

Journal of Molecular Evolution

Counterselection of GATC Sequences in Enterobacteriophages by the Components of the Methyl-Directed Mismatch Repair System

Patrick Deschavanne and Miroslav Radman

Laboratoire de Mutagénèse, Institut J. Monod, 2 place Jussieu, 75251 Paris Cedex 05, France

Summary. Weak to severe deficit of GATC sequences in the DNA of enterobacteriophages appears to be correlated with their undermethylation during growth in *dam*$^+$ (GATC ade-methylase) bacteria. This observation is corroborated by the sequence analysis showing no evidence for site-specific mutagenicity of 6meAde. The MutH protein of the methyl-directed mismatch repair system recognizes and cleaves the undermethylated GATC sequences in the course of mismatch repair. To enquire whether the MutH function of the methyl-directed mismatch repair system participates in counterselection of GATC sequences in enterobacteriophages, we have studied the yield of bacteriophage ΦX174 containing either 0, 1, or 2 GATC sequences, in wild type, *dam,* and *mut* (*H, L, S, U*) *Escherichia coli.* Following transfection with unmethylated DNA containing two GATC sequences, a net decrease in the yield of infective particles was observed in all bacterial *mutH*$^+$ *dam*$^-$ strains, whereas no detectable decrease was observed in bacteria infected by DNA without GATC sequence. This effect of the MutH function is maximum in wild type and *mutL* and *mutS* bacteria whereas the effect is not significant in *mutU* bacteria, suggesting an interaction of the helicase II with the MutH protein.

However, in *dam*$^+$ bacteria, the presence of GATC sequences leads to an increased yield of infective particles. The effect of GATC sequence and its Dam methylation system on phage yield in *mutH*$^-$ bacteria reveals that methylated GATC sequences are advantageous to the phage. These results suggest that the methyl-directed mismatch repair system, and in particular its MutH protein, may have participated in severe counterselection of GATC sequences from enterobacteriophages, presumably by DNA cleavage or by interfering with DNA replication or packaging when GATC sequences are undermethylated. Coevolution of the Dam and MutH proteins could then account for the loss of GATC sequences from DNA of bacteriophages growing in *dam*$^+$ hosts.

Key words: GATC sequence — Counterselection — Mismatch repair system — Enterobacteriophages — *Escherichia coli*

Introduction

Analysis of the incidence of diverse short DNA sequences in all available genomes or genomic fragments has so far revealed two striking sequence deficits: a nearly 10-fold deficit of CpG sequences in mammalian genomes (compared to any other dinucleotide sequence, e.g., GpC) (Cooper 1983; Cooper and Gerber-Huber 1985; Cooper and Krawczak 1989) and a mild to very severe deficit of GATC sequences in some bacteriophages having specific enterobacterial species as hosts (McClelland 1985). In the former case, the cytosine methylation appears to be responsible (Cooper 1983; Razin and Cedar 1984; Razin and Szyf 1984), as there exist regions of mammalian genomes rich in CpG sequences in which they are not methylated (Bird 1986). In addition, eukaryotic genomes devoid of cytosine methylation do not show the CpG deficit (Doerfler 1983). Where the CpG deficit is compensated by the excess of TpG and CpA sequences, it is likely that either deamination of 5-methylcytosine (5-meC) to thymine or the miscopying of 5-meC by adenine causes this pronounced CpG → TpG sequence evo-

Offprint requests to: P. Deschavanne

lution (Cooper and Gerber-Huber 1985; Cooper and Krawczak 1989; Yomo and Ohno 1989).

The deficit of GATC sequences is found only in the genomes of bacteriophages growing in enterobacteria endowed with the Dam-methylase activity that methylates the adenine in the GATC sequence to 6-methyladenine (6-meA) (Barbeyron et al. 1984; McClelland 1985) and is not related to a host restriction-modification system (McClelland 1985; Sharp 1986). If, in analogy to 5-meC, 6-meA were an unstable base, i.e., chemically unstable or frequently miscopied by a nonthymine base, one would expect a deficit of GATC sequences in bacterial DNA as well and an increase in GXTC tetranucleotide sequences (where X is T, G, or C) in bacteriophages with a severe GATC deficit. For example, ΦX174 single-stranded phage DNA has no GATC sequence (15 expected, see Materials and Methods section for the methods of calculation of the expected number of GATC sequences) and T7 phage has six GATC sequences (141 expected); yet we found, in agreement with McClelland (1985), no increase in GXTC sequences over other tetranucleotide sequences, suggesting that GATC sequences did not mutate away by a preferential and specific change of their modified A. Furthermore, single-stranded phages like M13 (3 GATC sequences, 17 expected) show a lower level of GATC methylation than the double-stranded phage λ (116 GATC sequences, 212 expected). The case of double-stranded phages T1, T3, and T7, which code for an inhibitor of the restriction-modification system of *Escherichia coli* (adenine methyltransferase in T1, S-adenosylmethionine hydrolase for T3, or the gene 0.3 product for T7) (Spoerel and Herrlich 1979; Spoerel et al. 1979; Wagner et al. 1979; Auer and Schweiger 1984; Bandyopadhyay et al. 1985; Krüger et al. 1985; Hughes et al. 1987), suggests that it is the absence rather than the presence of 6-meA in GATC sequences within *dam*$^+$ bacteria that causes the loss of these sequences. To examine how unmethylated GATC sequences could be counterselected, we have considered the fact that the *dam*$^+$ bacteria are also endowed with the methyl-directed mismatch repair system, components of which can recognize and cut unmethylated GATC sequences. We have examined whether these mismatch repair components are involved in counterselection of GATC sequences. The key components of the methyl-directed mismatch repair system are mismatch recognition (MutS protein), complex formation (MutL protein), helicase activity (MutU or UvrD protein), and the recognition and nicking of unmethylated GATC sequences (MutH protein) (Radman and Wagner 1986; Modrich 1989). To test for counterselection of GATC sequences in different mismatch repair and/or adenine methylation mutants, we have used ΦX174 phage devoid of GATC sequences (Sanger et al. 1977) and derivatives carrying either one or two GATC sequences (Längle-Rouault et al. 1986, 1987). Our results indicate that the mismatch repair components, in particular mutH function, may have caused counterselection of GATC sequences in bacteriophage DNA.

Materials and Methods

Escherichia coli Cell Strains. Table 1 gives the description of the bacterial strains used.

ΦX174 Mutant Phage. ΦX174 Eam3 J-F ins3 carrying no GATC sequence, ΦX174 Eam3 J-F ins6 carrying 1 GATC sequence, and ΦX174 Eam3 J-F ins5 carrying 2 GATC sequences were described previously (Muller and Wells 1980; Längle-Rouault et al. 1986, 1987). We obtained *am*$^+$ derivatives of these three phages by selection on *su*$^-$ bacteria HF 4704. They will be referred to throughout the text as 0 GATC, 1 GATC, and 2 GATC inserts, respectively. ΦX174 *Ts* (temperature sensitive) 116 was used as an internal 0 GATC control for all the experiments.

DNA Preparation. Single-strand ΦX174 DNA of diverse mutants was prepared following the method of Cunningham et al. (1980) and extracted with phenol at 55°C (Schekman et al. 1971) from CsCl-purified phages (Radman 1976).

Transfection and Scoring of Infective Centers. Spheroplasts for the transfection were prepared by the lysozyme–EDTA procedure (Benzinger 1978). To increase the precision of transfection efficiency measurement and of phage titering, we used an internal standard phage DNA carrying a *Ts* mutation in otherwise wild type (wt) phage. Thus the transfection was performed with a mixture of a roughly equal amount of DNA from a GATC-free ΦX174 *Ts* 116 mutant and a derivative carrying either 0, 1, or 2 GATC, at 32°C either for 8 min (killing experiments) or 4 h (yield experiments), the peak time for the yield was determined from the kinetics of the production of infective particles for each transfecting DNA, and the spheroplasts or the phage were plated on HF 4733 *su*$^+$ indicator and incubated at 32°C until plaques appeared. To discriminate between the *Ts* mutant plaques and the plaques originating from one of the three *Ts*$^+$ derivatives, plaques were selected and phage progeny spotted on replica plates incubated either at 32 or 42°C.

Data Analysis. The primary results for a given transfection were expressed as the *Ts*$^+$/*Ts* ratio, i.e., the ratio of the number of infective centers growing at 42°C (i.e., from phages containing the insert with 0, 1, or 2 GATC sequences) to the number of infective centers appearing only at 32°C (originating from the *Ts* 116 phages). In a second step this ratio was compared to that of the *mutH* experiment to give either the fraction of surviving phages or that of the proportion of infective phages produced by the bacteria transfected using *mutH* as the standard. When statistical tests were applied (*t*-test, ANOVA, Kruskal–Wallis, chi-square), the significance level was taken as the probability *P* corresponding to the risk associated with the rejection of the null hypothesis. A *P* value <5% permits the conclusion, at risk *P*, that the compared groups are not sampled from the same population.

Sequence Analysis. The sequences from genes of *E. coli*, *Bacillus subtilis*, and their related bacteriophages were obtained from GenBank using Bysance analysis system from CITI 2 (Paris, France). The same software package was also used for the homology study between *E. coli* and *Streptococcus* sequences. To

Table 1. Bacterial strains used in this study

E. coli K 12	Description	Relevant genotype	Source (reference)
HF 4733	F⁻, *thy, thr, gal, lac, endo*I⁻, *su*II⁺, ΦX^s, *strp^s*, *dra/m*⁻	*su*⁺	P. Howard-Flanders
HF 4733 *dam*	As HF 4733 but *dam*3	*su*⁺	ES548 × GM81 (Marinus 1973, conjugation G Maenhaut-Michel and G Längle-Rouault)
HF 4704	F⁻, *thy, thr, uvrA*13, ΦX^s, *strp^s*, *su*⁻	su⁻	P. Howard-Flanders (Lindqvist and Sinsheimer 1967)
HF 4704 *dam*	As HF 4704 but *dam*3	*dam*	P1 transduction of *dam*13::Tn9 from GM2199 (Marinus et al. 1983), this study
ES 548	*trp*450, *str*A, *thi, lacZ*(ICR 36)	Wild type	E. Siegel (Vaccaro and Siegel 1977)
ES 871	As ES 548 but *mutL* 25	*mutL*	E. Siegel (Vaccaro and Siegel 1977)
RH 4447	As ES 548 but *mutH* 101, *thy*⁺	*mutH*	ES548 × KMBL3773 (Glickman and Radman 1980, conjugation G Maenhaut-Michel and G Längle-Rouault)
ES 548 *dam*	As ES 548 but *dam*3, *strp^s*, *thy*⁺, *spc^r*	*dam*	ES548 × GM81 (Marinus 1973, conjugation G Maenhaut-Michel and G Längle-Rouault)
ES 548 *dam mutH*	As ES 548 but *mutH* 101, *dam*3	*dam mutH*	(Caillet-Fauquet et al. 1984)
KMBL 3790	(λc⁺), F⁻, *thyA*306, *lysA*65, *argA*103 *bio*37, *pheA*97, *cysC, endoA*101, *dam*3, *mutU* (*uvrE*)	*dam mutU*	(Glickman and Radman 1980)
GM 150	F⁻, *thi*1, *leuB*6, *proA*2, *his4, metB*1 *lacY*1, *galK*2, *dam*3, *mutL*451	*dam mutL*	M. Marinus
GM 169	F⁻, *ara*14, *thi*1, *thyA*12, *strA, tsx*3 *deoB*16, *dam*3, *mutS*453	*dam mutS*	M. Marinus

estimate the expected number of a particular tetranucleotide in a gene of a genome, we used a second-order Markov chain predictor that takes into account the observed frequency of tri- and dinucleotides (McClelland 1985; Sharp 1986; Phillips et al. 1987a,b). We tested that this predictor gave a good estimation of the tetranucleotide frequency on our data set. We tested also the zero and first orders of the Markov chain analysis and confirmed that they do not apply to our set of data. This estimator of the expected number of tetranucleotides, for the GATC sequence for example, is equal to

$$p(GAT) \times p(ATC)/p(AT)$$

The use of the trinucleotide frequencies such as p(GAT) and p(ATC) removes some of the bias due to variation in di- and trinucleotide frequency and takes into account preferences in codon usage.

Results

GATC Sequences and the Formation of Infective Centers in Mismatch Repair Mutants

The data in Table 2 show that there is no significant detectable *mut*-dependent inactivation of unmethylated DNA in either *dam*⁺ or *dam* bacteria, with the possible exception of *dam*⁺ *mutL* bacteria strains. In this strain we observed a slight but significant killing of the 1 GATC DNA ($P = 0.023$) and a slight protection in the same bacteria for the 0 GATC DNA ($P = 0.048$). Although formally significant, the effect of *MutL* mutation cannot be interpreted in view of current knowledge (Modrich 1989), therefore, at this stage, we cannot draw any conclusion based on this observation. Nevertheless, with the possible exception of 1 GATC in *mutL* background (S = 0.9) no killing could be observed in *dam*⁻ or *dam*⁺ background for unmethylated DNA whatever the number of GATC in the molecule. These results indicate that the phage DNA, with or without GATC sequences, is not inactivated as it enters bacteria and that some replication cycles are allowed (Kornberg 1980).

Table 2. Survival ratio, relative to *mutH* bacteria, of the number of infective centers obtained after transfection of different bacterial strains by unmethylated DNA of phages containing different numbers of GATC sequences

		0 GATC	1 GATC	2 GATC
dam⁺	Wild type	1.13 (1103)	1.22 (1270)	1.19 (608)
	mutH	1.0 (626)	1.0 (737)	1.0 (681)
	mutL	1.35 (1632)*	0.9 (1542)*	1.45 (1151)
dam	Wild type	1.05 (1012)	1.19 (1557)	0.96 (1016)
	mutH	1.0 (897)	1.0 (1226)	1.0 (1079)
	mutU	1.21 (827)	1.11 (450)	1.11 (1577)

The number of infective centers analyzed is given in parentheses. The results are the mean of two to four experiments
* Significantly different at the 5% level from the *mutH* control

GATC Sequences and the Yield of Infective Phage Particles in Different mut Strains

The described killing experiments with infective center assay do not take into account the number

Table 3. Yield of infective centers relative to *mutH* bacteria, after transfection of different bacterial strains with DNA (unmethylated in *dam* and methylated in *dam*$^+$ experiments) containing different numbers of GATC sequences

		0 GATC	1 GATC	2 GATC
dam	*mutH*	1.0 (3179)	1.0 (4348)	1.0 (3316)
	wt	0.93 (3147)	0.76 (3216)	0.58 (2991)*
	mutU	1.09 (1994)	1.26 (1833)	0.81 (1981)
	mutL	1.20 (799)	0.61 (1451)	0.37 (1476)*
	mutS	0.80 (759)	0.92 (1399)	0.63 (1510)*
dam$^+$	*mutH*	1.0 (2544)	1.0 (1750)	1.0 (1781)
	wt	0.74 (1827)*	0.79 (1622)	1.21 (1289)
	mutL	0.48 (2209)*	0.80 (1804)	0.80 (1620)

The number of infective centers analyzed is given in parentheses; *dam* wt, *dam mutU*, *dam mutL*, *dam mutS*, *dam*$^+$ wt, and *dam*$^+$ *mutL* are all *mutH*$^+$ strains

* Significantly different at the 5% level from the *mutH* control

of phage particles produced in the different test bacteria (burst size). It was previously observed that phages containing 2 GATC sequences in *dam* bacteria always produced a smaller yield of infective particles than the ones containing 0 or 1 GATC sequence (F. Längle-Rouault and G. Maenhaut-Michel, personal communication). The phage yield, after transfection by methylated or unmethylated DNA from the three derivative phages in the different bacteria, relative to the phage yield in *mutH* bacteria is given in Table 3. This phage yield takes into account only the progeny of the surviving phage and so is independent of the killing experiments. In a *dam* background, unmethylated 0 or 1 GATC DNA yield a similar number of infective phage particles (as compared to the *mutH* control strain) for all the strains (Table 3), although there is a nonsignificant smaller phage yield of 1 GATC phages in *dam* wt and *dam mutL* bacteria. However, the fate of 2 GATC unmethylated DNA is somewhat different: compared with the *mutH* strain, the yield of infective particles by this unmethylated DNA is decreased in the four bacterial strains. This decrease is significant in all bacterial strains except the *dam mutU*. This decrease represents about 40% in the wt and *mutS* strains and 60% in *mutL*. In the case of the *mutU* bacteria the decrease in the phage yield is small, about 20%, and is not significant. The situation is different for the yield of infective phages by methylated DNA in *dam*$^+$ bacteria (Table 3): a decrease in the yield of infective phages was observed for 0 GATC DNA relative to a similar phage yield for the 1 and 2 GATC DNA. This result is not related to the state of methylation of DNA as similar results were obtained with unmethylated DNA (results not shown). To understand this result we tested the effect of the *dam* gene product by transfecting the three DNAs into *mutH* strains, either *dam* or *dam*$^+$ (Table 4). A high phage yield was observed

Table 4. Yield of infective centers in *dam*$^+$ *mutH* bacteria relative to *dam mutH* bacteria after transfection with unmethylated DNA containing different numbers of GATC sequences

	0 GATC	1 GATC	2 GATC
dam$^+$ *mutH*	0.97 (1014)	1.57 (1380)	1.47 (1340)*
dam mutH	1.0 (1228)	1.0 (1385)	1.0 (1181)

The number of infective centers analyzed is given in parentheses

* Significantly different at the 5% level from the *dam* control

in *dam*$^+$ bacteria for 1 and 2 GATC DNA as seen by the increased yield of the respective phages (however, the only significant difference is observed for 2 GATC DNA), whereas the yield was similar in *dam* and *dam*$^+$ bacteria for the 0 GATC DNA.

Discussion

Initially, we have studied the fate of the GATC sequence in enterobacteriophages. Several authors (McClelland 1985; Sharp 1986; Phillips et al. 1987a,b) have previously addressed this question but none have verified whether the methylation of the adenine leads to an increased mutability of this base, converting the GATC sequence into GXTC where X is T, G, or C. To address this specific question, we used a small set of data as the basic findings were already published (McClelland 1985; Phillips et al. 1987a,b). We used a second-order Markov chain predictor, which takes into account the observed frequency of tri- and dinucleotides (McClelland 1985; Sharp 1986; Phillips et al. 1987a,b), in order to calculate the predicted number of a given sequence. This comparison was done on a gene-by-gene basis for *E. coli*, *B. subtilis*, T3, and T4 or for complete genome for the other bacteriophages. For 16 genes of *E. coli* we found no statistical difference (by a chi-square test) between the observed number and the expected number of GXTC sequences (Table 5). This was also found for three genes of *B. subtilis* and for *B. subtilis* phages (Table 5). These two results are in agreement with previously published results (McClelland 1985; Phillips et al. 1987a,b). However, in *E. coli* bacteriophages, besides the great deficit in GATC sequences already reported (McClelland 1985) (Table 5 and Fig. 1), we found a slight but significant deficit in all the derived sequences except the GTTC (which exhibits a slight increase in the observed number, which is not observed for the complementary GAAC sequence) (Table 5 and Fig. 1). Thus, the GXTC sequences seem to be counterselected in general in all bacteriophages (Table 6), supporting the hypothesis that the deficit in GATC sequence does not derive from a hypermutability of the methylated adenine in such a sequence.

Table 5. Comparison between observed and expected number of GATC and related sequences in *Escherichia coli* (16 genes accounting for 42,959 bases), bacteriophages having *E. coli* as host (complete genome for F1, FD, M13, ΦX174, G4, λ, T7, 11 genes from T4 and 3 genes from T3 accounting for 166, 133 bases), *Bacillus subtilis* (3 genes accounting for 3892 bases) and sequences from phages having *B. subtilis* as host (7 genes from sp and Φ29 accounting for 10,477 bases

	GATC	GGTC	GCTC	GTTC	GACC	GAAC	GAGC
E. coli	=	=	=	=	=	=	=
E. coli phages	− −	−	=	+	−	−	−
B. subtilis	=	=	=	=	=	=	=
B. subtilis phages	=	=	=	+	=	=	=

The comparison was made by a chi-square test. The equal sign (=) indicates no significant difference between the observed and the expected number of sequences, a single (+) or (−) sign indicates a significant difference at the 5% level with an indication of whether the observed number was greater (+) or lower (−) than the expected number. A double sign has the same meaning as a single sign but the difference is significant at the 1% level. GACC, GAAC, and GAGC are the complementary sequence of GGTC, GTTC, and GCTC sequences, respectively

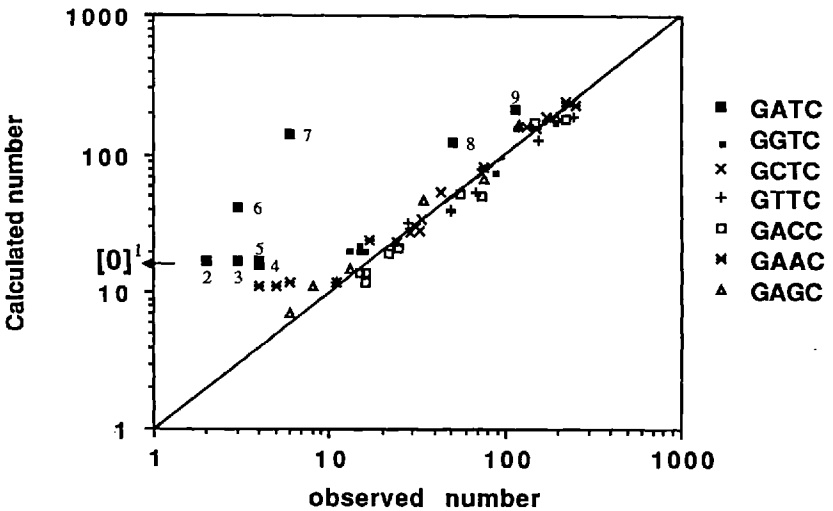

Fig. 1. Comparison between the observed number and the expected number of different sequences GXTC (where X = A, T, G, or C) for nine coliphages [ΦX174 (1), G4 (2), M13 (3), Fd (4), F1 (5), T3 (6), T7 (7), T4 (8), λ (9)]. The line drawn represents identity between observed and expected values. The points above the line indicate a deficit in the sequence under consideration whereas those under the line represent an excess of observed number as compared to the expected number (see Table 6). Due to the logarithmic scale used, we cannot represent the zero number of GATC sequence observed in ΦX174 and therefore we add a zero point outside the frame of the figure to represent this particular situation.

We have shown that transfection of *mutH*$^+$ bacteria by unmethylated 2 GATC phage DNA leads to a decreased yield of infective bacteriophage particles. This could result from a decreased efficiency of DNA replication and/or packaging, or from DNA inactivation processes due to MutH function. In either case, targets for these effects would be the unmethylated GATC sequences in either single- or double-stranded DNA. The latter hypothesis could be supported by considering the known activities of the MutH protein: MutH, in conjunction with MutS and MutL, is known to cut DNA heteroduplexes containing mismatched bases at unmethylated GATC sequences (Modrich 1989), but a weak endonuclease activity of the free MutH protein [cleaving d(GATC) sequences 5′ to the dG] on the unmethylated DNA strands was detected irrespective of the presence of mismatches (Welsh et al. 1987). This endonuclease activity could act, in mismatch repair, at the strand discrimination stage of mismatch correction, which involves scission of the unmethylated strand (Längle-Rouault et al. 1987; Modrich 1989). Thus, the decreased yield of infective particles containing 2 GATC sequences observed in *dam mutH*$^+$ (whether *mut*$^+$, *mutU, mutL,* or *mutS*) bacteria for unmethylated DNA could be due to the cleavage of this DNA by the MutH protein. In the same strains, transfection with DNA containing only one GATC sequence led to a decrease (though not significant) in the yield of infective particles (the yield of phage containing an insert with no GATC sequence is the same in *dam mutH* and *dam mutH*$^+$ bacteria). There is a significant inverse correlation between the number of GATC sequences of the transfecting DNA and the yield of phage resulting from transfection by this DNA. In *dam* background, all bacteria that are defective in another protein of the *mut* system show a significant decrease in the yield of 2 GATC phages, except the *mutU* (Table 3). For *dam mutU* we observed a small decrease in phage yield of 2 GATC phage and no decrease at all for 0 and 1 GATC DNA. It is possible that there is an interaction between the MutU and MutH proteins as suggested by Radman and Wagner (1986). It was proposed that helicase II opens the double-stranded DNA due to the presence of mismatch, creating single-stranded DNA, and then "walks along" the DNA in conjunction with MutH

Table 6. Observed (o) and expected (e) number of GXTC sequences (where X = A, T, G, or C) in different bacteriophages

Sequences	Genomes							
	ΦX o	ΦX e	G4 o	G4 e	fdo	fde	M13 o	M13 e
GATC	0	15	2	17	4	16	3	17
GGTC	15	20	16	20	15	22	13	20
GCTC	24	23	34	34	29	27	31	30
GTTC	28	32	23	22	49	40	50	39
GACC	22	19	25	21	16	14	16	12
GAAC	11	12	17	24	6	12	5	11
GAGC	13	15	6	7	8	11	11	12
Sum	124	151	129	154	141	153	141	154

[a] In the case of T3 and T4, the sequences of a number of their genes were pooled for this analysis

until a nonmethylated GATC sequence, which is then cut by MutH. This interaction would explain the small decrease in the yield observed in *mutU* bacteria. We do not rule out a decrease in the yield of 1 GATC phages in *dam* bacteria, as the yield of these phages in wt and *mutL* bacteria was intermediate between the 0 GATC control and the 2 GATC DNA. This observation mirrors that of mismatch repair measurement with the same three DNAs and agrees with the notion that at least one unmethylated GATC sequence is required for mismatch repair (Längle-Rouault et al. 1986, 1987; Lahue et al. 1987; Lu 1987; Welsh et al. 1987) but that 2 GATC sequences lead to a higher efficiency.

In *dam*$^+$ bacteria, we observed no decrease in the yield for GATC-containing DNA, which is in agreement with observations that MutH protein does not interact with methylated GATC sequences (Welsh et al. 1987). However, we have observed a MutH-dependent decrease in the phage yield for the DNA that does not contain any GATC sequence (Table 3). This observation could reflect an involvement of MutH, which is irrespective of GATC sequences, e.g., a binding or nuclease activity, although such an activity has not been demonstrated (Welsh et al. 1987). It cannot be excluded that 6-meAde residues provide some selective advantage to the phage DNA. It is possible also that Dam protein itself protects the DNA containing GATC sequences (Szyf et al. 1986). To test these hypotheses, we compared the phage yield from unmethylated DNA in *dam mutH* and *dam*$^+$ *mutH* bacteria (Table 4): no decrease in phage yield was detected for 0 GATC DNA in *dam*$^+$ as compared to *dam* bacteria. However, the phage yield for DNA containing 1 or 2 GATC sequences in *dam*$^+$ bacteria is about 50% higher than that observed for 0 GATC DNA (although only the result for 2 GATC DNA is significantly different from the control). This selective advantage appears to be due to the Dam protein alone as it is independent of the MutH function (Table 4). Thus, it may be that the Dam protein or its methylase activity protects DNA containing GATC from the effects of the MutH and some other unknown functions that decrease the phage yield.

The above results could account for the counterselection of GATC sequences observed in bacteriophages specific to enterobacteria (McClelland 1985), as even small effects on phage yield can have large long-term fitness effects. It is known that the Dam methylation is delayed relative to DNA synthesis (Marinus 1976; Geier and Modrich 1979; Lyons and Schendel 1984; Szyf et al. 1984; Ogden et al. 1988) and that it never methylates all GATC sequences in either double-stranded (Dreiseikelmann et al. 1979) or single-stranded (Lu et al. 1983) phages. Thus, DNA containing unmethylated GATC sequences could be inactivated by the MutH protein, leading to a significant decrease in the yield of phages containing numerous GATC sequences, therefore providing a selective advantage for phages that lose GATC sequences (an extreme case being ΦX174, which is completely devoid of GATC). Besides this specific activity toward unmethylated GATC sequences, MutH (and perhaps some other proteins) is probably able to decrease the phage yield irrespective of GATC sequence (Table 3). However, the Dam protein itself seems to provide some selective advantage to DNA containing GATC sequences. This protection could not be complete as on the one hand there is a lag between synthesis of DNA and methylation by Dam protein, leaving GATC sequences unmethylated for a short time (Lyons and Schendel 1984; Szyf et al. 1984; Ogden et al. 1988) and on the other hand because the Dam protein is not able to complete the methylation of all GATC sequences in phage DNA (Lu et al. 1983).

Our results indicate that the components of the methyl-directed mismatch repair system could be responsible for the counterselection of GATC sequence in enterobacteriophages. A speculation is that the counterselection effect is formally analogous to a weak restriction system, as propositions were made by Claverys and Lacks (Claverys and Lacks 1986; Lacks 1988) that the Hex mismatch repair system and the Dpn II methylase and endonuclease of

Table 6. Extended

Genomes									
f1 o	f1 e	λ o	λ e	T7 o	T7 e	T3[a] o	T3 e	T4[a] o	T4 e
4	17	116	212	6	141	3	42	51	125
15	20	118	157	194	173	71	51	88	75
33	28	133	161	174	187	43	54	75	75
49	39	181	190	246	188	69	54	157	129
16	14	149	173	222	184	56	52	74	50
4	11	224	242	250	226	78	81	150	155
11	12	194	186	121	165	35	48	76	68
145	153	1316	1558	1309	1371	380	410	753	739

Streptococcus, on one hand, and the Mut system from *E. coli,* on the other hand, would have a common ancestor some 10^9 years ago. The Dam and MutH proteins would be the counterparts of the Dpn II system of *Streptococcus,* whereas Hex A would be the equivalent of MutS. Indeed, there is a 36% identity between Hex A and MutS and 30% identity between Dam and Dpn II methylases (Manarelli et al. 1985) (our analysis showed two large domains conserved up to 40–45%, results not shown), however, we found only about 20% identity between MutH and Dpn II endonucleases, although they have the same substrate. The molecular weight of the monomer of Dpn II endonuclease and that of MutH are very close, but the latter acts as a monomer (Manarelli et al. 1985). This could account for the difference in substrate for these two enzymes, Dpn II recognizing unmethylated GATC in double-stranded DNA, whereas the best substrate for MutH seems to be hemimethylated GATC in double-stranded DNA (Welsh et al. 1987) or, as we postulate here, unmethylated GATC in single-stranded DNA. The involvement of Dam methylation in the methyl-directed mismatch repair of *E. coli* suggests a coevolution of the Dam and MutH functions. In *dam* bacteria the methyl-directed mismatch repair system would act by cutting DNA at GATC sequence, thus inactivating phages and leading to a lower yield of phages containing numerous GATC sequences (other means of interfering with phage yield is not excluded as discussed). In dam^+ bacteria the counterselection mechanism could be the same except that the number of unmethylated GATC sequences is smaller and therefore the selective process would take a longer time to lead to a significant decrease in the number of GATC sequences in a given genome. The coevolution of the MutH and Dam functions could account for the puzzling observations that the counterselection of GATC sequences correlates with their undermethylation in dam^+ $mutH^+$ bacteria.

Acknowledgments. We thank C. Dohet for her help in initial stages of this work, G. Maenhaut-Michel and P. Caillet-Fauquet for strains, and J. Filipski and R. D'Ari for comments on the manuscript. This work was supported by C.N.R.S. and by contract No. BI6-E-154F from European Communities.

References

Auer B, Schweiger M (1984) Evidence that *Escherichia coli* virus T1 induces a DNA methyltransferase. J Virol 49:588–590

Bandyopadhyay PK, Studier FW, Hamilton DL, Yuan R (1985) Inhibition of the type I restriction-modification enzymes *ECO*B and *Eco*iK by the gene 0.3 protein of bacteriophage T7. J Mol Biol 182:567–578

Barbeyron T, Kean K, Forterre P (1984) DNA adenine methylation of GATC sequences appeared recently in the *Escherichia coli* lineage. J Bacteriol 160:586–590

Benzinger R (1978) Transfection of Enterobacteriaceae and its applications. Microbiol Rev 42:194–236

Bird A (1986) CpG-rich islands and the function of DNA methylation. Nature 321:209–213

Caillet-Fauquet P, Maenhaut-Michel G, Radman M (1984) SOS mutator effect in *E. coli* mutants deficient in mismatch correction. EMBO J 3:707–712

Claverys JP, Lacks SA (1986) Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. Microbiol Rev 50:133–165

Cooper DN (1983) Eukaryotic DNA methylation. Hum Genet 64:315–333

Cooper DN, Gerber-Huber S (1985) DNA methylation and CpG suppression. Cell Diff 17:199–205

Cooper DN, Krawczak M (1989) Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. Hum Genet 83:181–188

Cunningham RP, Dasgupta C, Shibata T, Radding CM (1980) Homologous pairing in genetic recombination: recA protein makes joint molecules of gapped circular DNA and closed circular DNA. Cell 20:223–235

Doerfler W (1983) DNA methylation and gene activity. Annu Rev Biochem 52:93–124

Dreiseikelmann B, Eichenlaub R, Wackernagel W (1979) The effect of differential methylation by *Escherichia coli* of plasmid DNA and phage T7 and λ DNA on the cleavage by restriction endonuclease *Mbo*I from *Moraxella bovis.* Biochim Biophys Acta 562:418–428

Geier GE, Modrich P (1979) Recognition sequence of the *dam* methylase of *Escherichia coli* K12 and mode of cleavage of *DpnI* endonuclease. J Biol Chem 254:1408–1413

Glickman BW, Radman M (1980) *Escherichia coli* mutator mutants deficient in methylation-instructed DNA mismatch correction. Proc Natl Acad Sci USA 77:1063–1067

Hughes JA, Brown LR, Ferro AJ (1987) Expression of the cloned coliphage T3 S-adenosylmethionine hydrolase gene inhibits DNA methylation and polyamine biosynthesis in *Escherichia coli*. J Bacteriol 169:3625–3632

Kornberg A (1980) DNA replication. WH Freeman, San Francisco

Krüger DH, Shroeder C, Reuter M, Bogdarina IG, Buryanov YI, Bickle TA (1985) DNA methylation of bacterial viruses T3 and T7 by different DNA methylases in *Escherichia coli* K12 cells. Eur J Biochem 150:323–330

Lacks SA (1988) Discussion summary: mismatch repair of DNA. In: Friedberg EC, Hanawalt PC (eds) Mechanisms and consequences of DNA damage processing. Liss, New York, pp 391–395

Lahue RS, Su SS, Modrich P (1987) Requirement for d(GATC) sequences in *Escherichia coli mutHLS* mismatch correction. Proc Natl Acad Sci USA 84:1482–1486

Längle-Rouault F, Maenhaut-Michel G, Radman M (1987) GATC sequences, DNA nicks and the mutH function in *Escherichia coli* mismatch repair. EMBO J 6:1121–1127

Längle-Rouault F, Maenhaut-Michel G, Radman M (1986) GATC sequence and mismatch repair in *Escherichia coli*. EMBO J 5:2009–2013

Lindqvist BH, Sinsheimer RL (1967) Process of infection with bacteriophage ΦX174. XIV. Studies on macromolecular synthesis during infection with a lysis-deficient mutant. J Mol Biol 28:87–94

Lu AL (1987) Influence of GATC sequences on *Escherichia coli* DNA mismatch repair *in vitro*. J Bacteriol 169:1254–1259

Lu AL, Clark S, Modrich P (1983) Methyl-directed repair of DNA base-pair mismatches *in vitro*. Proc Natl Acad Sci USA 80:4639–4643

Lyons SM, Schendel PF (1984) Kinetics of methylation in *Escherichia coli* K-12. J Bacteriol 159:421–423

Manarelli BM, Balganesh TS, Greenberg B, Springhorn SS, Lacks SA (1985) Nucleotide sequence of the *Dpn* II DNA methylase gene of *Streptococcus pneumoniae* and its relationship to the *dam* gene of *Escherichia coli*. Proc Natl Acad Sci USA 82:4468–4472

Marinus MG (1973) Location of DNA methylation genes on the *Escherichia coli* K-12 genetic map. Mol Gen Genet 127: 47–55

Marinus MG (1976) Adenine methylation of ogazaki fragments in *Escherichia coli*. J Bacteriol 128:853–854

Marinus MG, Carraway M, Frey AZ, Brown L, Arraj JA (1983) Insertion mutations in the *dam* gene of *Escherichia coli* K-12. Mol Gen Genet 192:288–289

McClelland M (1985) Selection against *dam* methylation sites in the genomes of DNA of enterobacteriophages. J Mol Evol 21:317–322

Modrich P (1989) Methyl-directed DNA mismatch correction. J Biol Chem 264:6597–6600

Muller UR, Wells RD (1980) Intercistronic regions in ΦX174 DNA. II Biochemical and biological analysis of mutants with altered intercistronic regions between genes J and F. J Mol Biol 141:25–41

Ogden GB, Pratt MJ, Schaechter M (1988) The replicative origin of the *E. coli* chromosome binds to cell membranes only when hemimethylated. Cell 54:127–135

Phillips GJ, Arnold J, Ivarie R (1987a) Mono- through hexanucleotide composition of the *Escherichia coli* genome: a Markov chain analysis. Nucleic Acids Res 15:2611–2626

Phillips GJ, Arnold J, Ivarie R (1987b) The effect of codon usage on the oligonucleotide composition of the *E. coli* genome and identification of over- and underrepresented sequences by Markov chain analysis. Nucleic Acids Res 15: 2627–2638

Radman M (1976) An endonuclease from *Escherichia coli* that introduces single polynucleotide chain scissions in ultraviolet-irradiated DNA. J Biol Chem 251:1438–1445

Radman M, Wagner R (1986) Mismatch repair in *Escherichia coli*. Annu Rev Genet 20:523–538

Razin A, Cedar H (1984) DNA methylation in eukaryotic cells. Int Rev Cytol 92:159–185

Razin A, Szyf M (1984) DNA methylation patterns: formation and function. Biochim Biophys Acta 782:331–342

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, Hutchison CA III, Slocombe PM, Smith M (1977) Nucleotide sequence of bacteriophage ΦX174 DNA. Nature 265: 687–695

Sharp PM (1986) Molecular evolution of bacteriophages: evidence of selection against the recognition sites of host restriction enzymes. Mol Biol Evol 3:75–83

Schekman RW, Iwaya M, Bromstrup K, Denhardt DT (1971) The mechanism of replication of ΦX174 single-strand DNA. III. An enzymic study of the structure of the replicative form II DNA. J Mol Biol 57:177–199

Spoerel N, Herrlich P (1979) Colivirus-T3-coded S-adenosylmethionine hydrolase. Eur J Biochem 95:227–233

Spoerel N, Herrlich P, Bickle TA (1979) A novel bacteriophage defense mechanism: the anti-restriction protein. Nature 278: 30–34

Szyf M, Avraham-Haetzni K, Reifman A, Shlomai J, Kaplan F, Oppenheim A, Razin A (1984) DNA methylation pattern is determined by the intracellular level of the methylase. Proc Natl Acad Sci USA 81:3278–3282

Szyf M, Meisels E, Razin A (1986) Biological role of DNA methylation: sequence-specific single-strand breaks associated with hypomethylation of GATC sites in *Escherichia coli* DNA. J Bacteriol 168:1487–1490

Vaccaro KK, Siegel EC (1977) The frameshift mutability of *PolA1* and *recA1* derivatives of mutator strains of *Escherichia coli*. Mutat Res 42:443–446

Wagner EF, Auer B, Schweiger M (1979) Development of *Escherichia coli* virus T1: escape from host restriction. J Virol 29: 1229–1231

Welsh KM, Lu AL, Clark S, Modrich P (1987) Isolation and characterization of the *Escherichia coli mutH* gene product. J Biol Chem 262:15624–15629

Yomo T, Ohno S (1989) Concordant evolution of coding and noncoding regions of DNA made possible by the universal rule TA/CG deficiency–TG/CT excess. Proc Natl Acad Sci USA 86:8452–8456

Received September 24, 1990/Revised February 5, 1991