JOURNAL OF BACTERIOLOGY, Nov. 1985, p. 932–937
0021-9193/85/110932-06$02.00/0

Vol. 164, No. 2

Common Evolutionary Origin of the Phage T4 *dam* and Host *Escherichia coli dam* DNA-Adenine Methyltransferase Genes

STANLEY HATTMAN,[1]* JENNIFER WILKINSON,[1] DAVID SWINTON,[1] SAMUEL SCHLAGMAN,[1] PAUL M. MACDONALD, [2]† AND GISELA MOSIG[2]

Department of Biology, University of Rochester, Rochester, New York 14627,[1] and Department of Molecular Biology, Vanderbilt University, Nashville, Tennessee 37232[2]

Received 31 May 1985/Accepted 6 August 1985

We compared the known DNA nucleotide and encoded amino acid sequences of the *Escherichia coli* and bacteriophage T4 *dam* (DNA-adenine methyltransferase) genes. Despite the absence of any DNA sequence homology, there were four regions (11 to 33 residues long) of amino acid sequence homology containing 45 to 64% identity. These results suggest that the genes for these two enzymes have a common evolutionary origin.

Bacteriophages T2 and T4 contain minor amounts of the methylated base N^6-methyladenine (m^6A) in their DNA (7, 9); however, the closely related phage T6 lacks any detectable m^6A. In accord with these findings is the fact that T2 and T4 both control the production of a DNA-adenine methyltransferase (Dam) activity, but T6 does not (8, 10, 12, 15). The T2 enzyme is chromatographically distinct from the major host DNA-adenine methyltransferase (29; J. E. Brooks, Ph.D. thesis, University of Rochester, Rochester, N.Y., 1977) specified by the *Escherichia coli dam* gene (22). Mutants of T-even phages have been isolated that are defective in glucosyl transferase production (25); these nonglucosylated forms are designated here simply as T2 *gt*, T4 *gt*, and T6 *gt*. Mutants have been isolated from T2 *gt* and T4 *gt* that abolish (C. P. Georgopoulos, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, 1969) or alter DNA methyltransferase activity (11, 16, 24); only in a *gt* background is the phenotype recognizable with respect to host restriction susceptibility. However, methylation appears to be a nonessential function because the unmethylated phages are viable (11; C. P. Georgopoulos, Ph.D. thesis). The ability to obtain unmethylated virion DNA demonstrates that the host Dam DNA methyltransferase is poorly able to methylate the phage genome in vivo. We presume that the presence of 5-hydroxymethylcytosine (hm^5C) in the phage DNA interferes with sequence recognition by the host enzyme; in addition, the intracellular level of the host enzyme is limiting (27). The host Dam enzyme methylates the palindromic sequence GATC to Gm^6ATC (13, 19); this sequence is also methylated by the T2 and T4 enzymes (14, 27, 29; J. E. Brooks, Ph.D. thesis). However, we should point out that the phage enzymes normally methylate the sequence $GAThm^5C$, which is found in place of GATC in the viral genome. Phage T1 also specifies a DNA-adenine methyltransferase capable of methylating GATC (2). Because of the high degree of sequence homology (6, 26) and the similarity in genetic maps (R. L. Russell, Ph.D. thesis, California Institute of Technology, Pasadena, 1967) it has been assumed that T2 and T4 produce similar enzymes. However, a number of observations suggest that the two proteins are different. (i) The T2 DNA methyltransferase appears to be a small protein, with an estimated molecular size of less than 15 kilodaltons (kDa) (M. Masurekar and S. Hattman, unpublished data; J. E. Brooks and S. Hattman, unpubished data), whereas the cloned T4 *dam* gene (27) produces a protein of about 30 kDa in maxicells (M. Myers, S. Schlagman, and S. Hattman, unpublished data). This value corresponds to the molecular size calculated for a T4 protein defined by one of two possible ribosome binding sites which could initiate translation in the same reading frame (20). This predicted molecular size is similar to that of the *E. coli* Dam enzyme (4). The second, shorter T4 translation frame (starting with GUG in place of AUG) codes for a polypeptide of 26 kDa (20). (ii) The m^6A content in virion DNA is lower for wild-type T4 than for wild-type T2, and this difference is also observed for T4 *gt* versus T2 *gt* (11). It is not yet clear whether the lower m^6A content in T4 is attributable to the lower level of DNA methyltransferase activity induced in T4 than in T2 infection (10, 15) or to a different sequence specificity or some other property. In this regard, although both phage enzymes methylate GATC, the T2 Dam enzyme can methylate other sequences (5, 14); recent studies suggest that the Dam enzyme produced by the cloned T4 *dam* gene methylates predominantly, if not exclusviely, GATC (S. Schlagman, S. Hattman, and M. G. Marinus, manuscript in preparation). In view of the similarities in sequence recognition and molecular weights (MWs) of the *E. coli* and T4 Dam^+ enzymes, we decided to use the DNA sequence data to make a computer-assisted comparison of the genes and the polypeptides they encode.

TABLE 1. Regions of amino acid sequence homology

Region	Amino acid positions in Dam		No. homologous/ total
	T4	*E. coli*	
I	9–41	12–44	15/33
II	91–106	95–110	8/16
III	116–128	124–136	7/13
IV	164–174	174–184	7/11

In a previous study (27), Southern hybridizatioin analyses at low stringency failed to reveal any DNA sequency homol-

* Corresponding author.
† Present address: Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 20138.

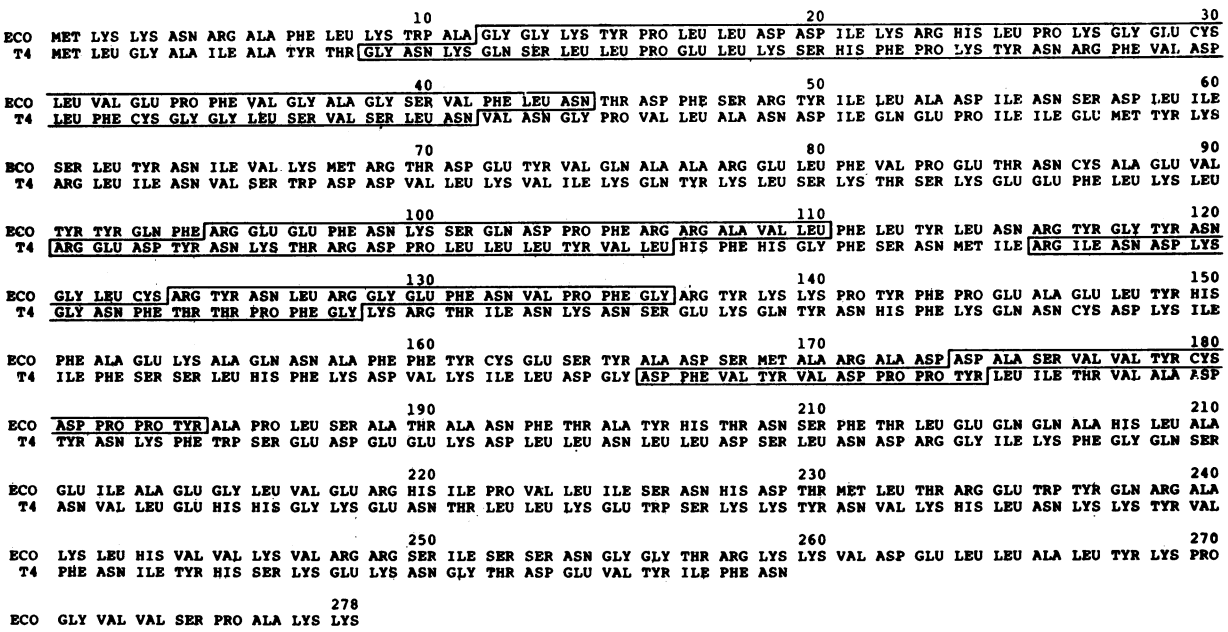

FIG. 1. Amino acid sequences of the *E. coli* (ECO) and T4 Dam proteins as deduced from the known nucleotide sequences. Regions of sequence homology identified by a computer homology matrix analysis are indicated by connecting lines.

ogy between the T4 *dam* and *E. coli dam* genes. This is not surprising, since the T4 *dam* gene is 72% A+T and the *E. coli dam* gene is 50% A+T. Attempts to visually detect DNA sequence homology in homology matrices generated by the DNA sequence programs of Pustell and Kafatos (23) were also unsuccessful. The longer of the two possible T4 *dam* gene translation frames encodes 259 amino acids (MW, 30,414), compared to 278 amino acids (MW, 32,097) for the *E. coli* Dam enzyme. Because of this similarity, as well as the similarity in methylation sequence specificity, we com-

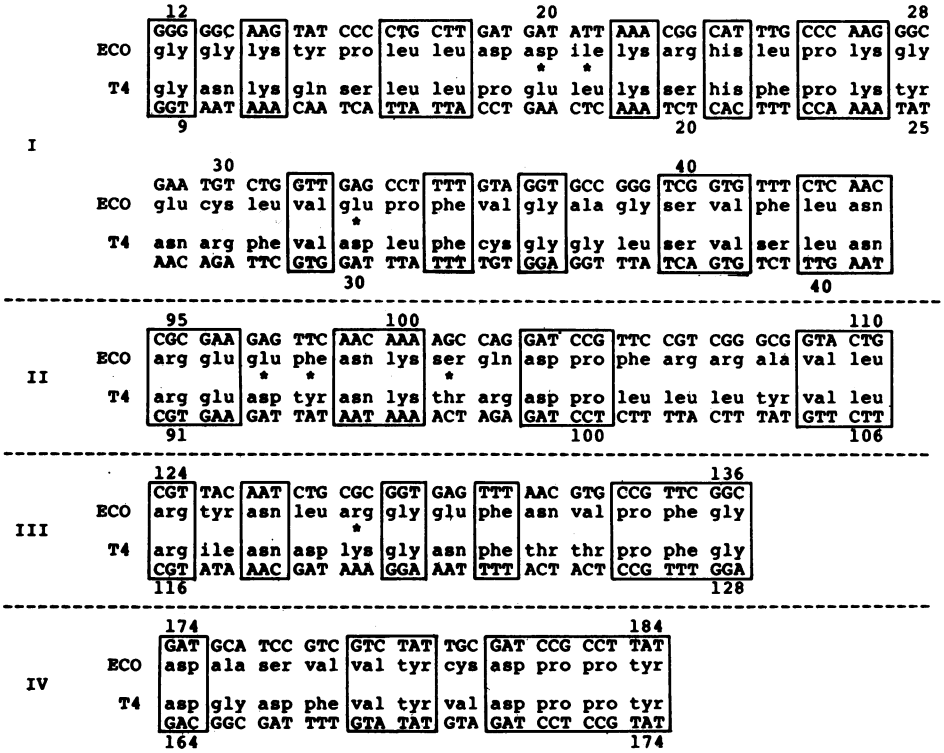

FIG. 2. Comparison of nucleotide and amino acid sequences in the four homologous regions. The homologous regions indicated in Fig. 1 have been aligned; boxes denote identical amino acids, and asterisks represent chemically similar amino acids. ECO, *E. coli*.

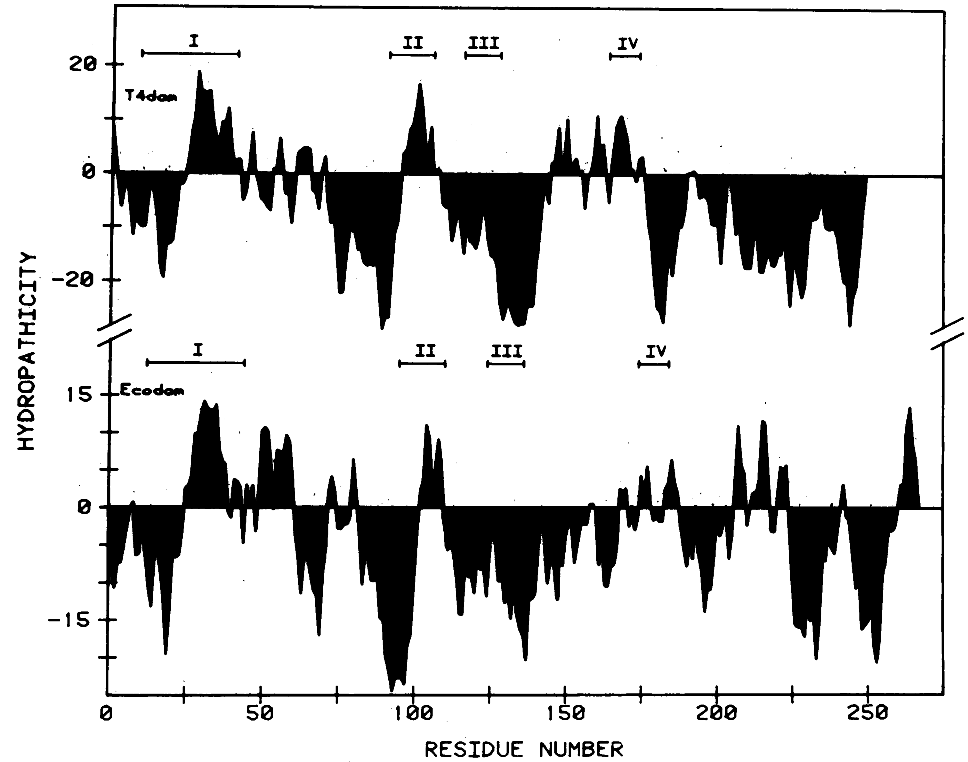

FIG. 3. Hydropathicity analysis of the T4 (top) and *E. coli* (bottom) Dam proteins. The hydropathy of the deduced amino acid sequences was analyzed by using a width of 11 amino acids and a jump of 1 (18).

Region I

T4					$*^{8}$	*	asn	*	gln	ser	*	*	*	glu	leu	lys	ser	his	phe
DPN	phe^{14}	thr	lys	trp	thr	gly	gly	lys	arg	gln	leu	leu	pro	val	ile	arg	glu	leu	ile
ECO	$*^{7}$	leu	*	*	ala	*	*	*	tyr	pro	*	*	asp	asp	*	lys	arg	his	leu

T4	*	*	-	*	*	*	phe	val	asp	leu	*	cys	*	*	leu	ser	val	ser	leu	asn
DPN	pro	lys	thr	tyr	asn	arg	tyr	phe	glu	pro	phe	val	gly	gly	gly	ala	leu	phe	phe	asp
ECO	*	*	-	gly	glu	cys	leu	val	*	*	*	*	*	ala	*	ser	val	*	leu	asn

Region III

T4	ser^{112}	asn	met	ile	*	ile	*	asp	*	gly	glu	*	thr	thr	*	phe	*	lys	arg	thr	ile	val
DPN	asn^{130}	gly	leu	tyr	arg	val	asn	ser	lys	asn	gln	phe	asn	val	pro	tyr	gly	arg	tyr	lys	asn	pro
ECO	$*^{120}$	*	*	cys	*	tyr	*	leu	arg	gly	glu	*	*	*	*	phe	*	*	*	*	lys	*

Region IV

T4	asp^{164}	*	*	*	*	*	val	*	*	*	*	leu	ile	thr	val	ala	asp
DPN	thr^{187}	gly	asp	phe	val	tyr	phe	asp	pro	pro	tyr	ile	pro	leu	ser	glu	thr
ECO	asp^{174}	ala	ser	val	*	*	cys	*	*	*	*	ala	*	*	*	ala	*

FIG. 4. Comparison of the amino acid sequences in the high-homology regions of the T4, *E. coli* (ECO), and *S. pneumoniae* (DPN) Dam proteins. The *S. pneumoniae DpnII* sequence is taken from Mannarelli et al. (21), and three of the high *S. pneumoniae-E. coli* homology regions are aligned with three of the high *T4-E. coli* homology regions (I, III, and IV). Asterisks indicate amino acids in T4 or *E. coli* which are identical to the corresponding *S. pneumoniae* residue; boxes indicate that all three are identical; circles indicate that only the T4 and *E. coli* proteins are identical. Superscripts indicate amino acid residue numbers. Dashes in the region I sequence are assumed deletions in the *E. coli* and T4 or an insertion in the *S. pneumoniae* protein.

TABLE 2. Codon usage in the *E. coli* and T4 Dam methylases

Codon	Amino acid	No. of occurrences		Codon	Amino acid	No. of occurrences	
		E. coli Dam	T4 Dam			*E. coli* Dam	T4 Dam
UUU	Phe	10	15	UAU	Tyr	11	12
UUC	Phe	7	1	UAC	Tyr	9	2
UUA	Leu	2	12	UAA		1	1
UUG	Leu	6	4	UAG		0	0
CUU	Leu	3	9	CAU	His	4	4
CUC	Leu	2	2	CAC	His	3	5
CUA	Leu	0	2	CAA	Gln	2	5
CUG	Leu	13	0	CAG	Gln	5	1
AUU	Ile	3	7	AAU	Asn	6	19
AUC	Ile	5	2	AAC	Asn	9	5
AUA	Ile	1	7	AAA	Lys	12	26
AUG	Met	4	3	AAG	Lys	6	5
GUU	Val	6	6	GAU	Asp	10	15
GUC	Val	5	1	GAC	Asp	3	3
GUA	Val	3	7	GAA	Glu	8	14
GUG	Val	5	2	GAG	Glu	11	1
UCU	Ser	3	4	UGU	Cys	3	2
UCC	Ser	2	1	UGC	Cys	2	0
UCA	Ser	1	7	UGA		0	0
UCG	Ser	1	1	UGG	Trp	2	3
CCU	Pro	2	4	CGU	Arg	6	3
CCC	Pro	5	0	CGC	Arg	10	0
CCA	Pro	2	2	CGA	Arg	1	0
CCG	Pro	5	2	CGG	Arg	2	0
ACU	Thr	1	6	AGU	Ser	3	3
ACC	Thr	2	0	AGC	Ser	5	0
ACA	Thr	3	3	AGA	Arg	0	4
ACG	Thr	4	0	AGG	Arg	0	0
GCU	Ala	3	2	GGU	Gly	4	2
GCC	Ala	8	1	GGC	Gly	6	2
GCA	Ala	7	0	GGA	Gly	1	8
GCG	Ala	8	1	GGG	Gly	2	1

pared the primary structure of the two proteins by using the corresponding amino acid homology matrices. This analysis revealed four regions of homology (containing about 45 to 64% identity; Table 1); the relative locations of these regions are displaced by several amino acid residues from their respective amino termini (Fig. 1). The nucleotide and amino acid sequences of each of the four homologous regions are aligned in Fig. 2. It is interesting that homologous region I was encoded by the longer T4 *dam* translation frame only.

Having found patch homology between the T4 and *E. coli* Dam proteins, it was of interest to compare their predicted hydrophobic and hydrophilic domains (Fig. 3). Hydropathicity profiles (18) were derived by continuously determining an average hydropathy for a segment of predetermined length as it advances from the amino to the carboxy terminus; the hydrophobic regions are above the zero line, and the hydrophilic regions are below the zero line. The two patterns are strikingly similar for the first 140 amino acids, after which they diverge.

The codon usage frequencies of the two proteins were determined (Table 2). It should be noted that CUG is the leucine codon most frequently used in *E. coli* (1), and this is also observed for the *E. coli* Dam protein; there are 13 CUGs among the 26 leucine codons. However, phage T4 infection leads to a rapid and specific cleavage of the leucyl-tRNA species that reads this codon (17, 28, 30); consequently, CUG is infrequently found as a T4 leucine codon. The results are in keeping with this pattern; of 29 T4 leucine codons, not one was CUG (Table 2). Thus, despite the evolutionary relationship of the *E. coli* and T4 *dam* genes, all the CUG codons have been changed in T4.

The A+T content of the *E. coli dam* gene (50%) is quite different from that of the T4 *dam* gene (72%); these values are similar to the overall base composition of the respective genomes. As might be expected, the large difference in A+T content is also reflected in codon usage. For example, the total number of encoded lysine-plus-arginine residues is almost identical for the two proteins, 37 in *E. coli dam* and 38 in T4 Dam. However, the *E. coli* gene encodes 18 lysine and 19 arginine residues, whereas T4 *dam* encodes 31 lysine and 7 arginine residues. The lysine codons are AAA and AAG, whereas the possible arginine codons are CGU, CGC, CGG, CGA, AGA, and AGG. Thus, on the basis of its high A+T content, T4 should have more lysine than arginine codons, and certain arginine codons (the CGX codons) should be found infrequently. This is precisely what was observed.

Examples for other amino acids were also consistent with this pattern.

It should be recalled that the T4 Dam enzyme methylates the sequence GAThm^{5}C, but this sequence is poorly methylated by the host *E. coli* Dam enzyme. Some of the divergence between the two enzymatic forms may represent evolutionary changes in the phage enzyme that allow it to recognize GAThm^{5}C more efficiently.

The *Streptococcus* (*Diplococcus*) *pneumoniae* DNA (modification) methylase *Dpn*II also methylates the palindromic sequence GATC to Gm^{6}ATC (19), and this DNA methylase gene has also been compared with the *E. coli dam* gene (21). Examination of the coding sequences revealed 30% overall identity between the two amino acid sequences. Moreover, there were four regions of up to 75% amino acid sequence homology, and three of these overlapped, the T4-*E. coli* homologous regions I, III, and IV (Fig. 4). It is evident that many sites are identical in all three proteins; in addition, some sites showed identity between the proteins produced by *S. pneumoniae* and *E. coli*, *S. pneumoniae* and T4, or T4 and *E. coli*. The conservation of these regions suggests that they are important in substrate recognition or enzyme activity; taken together, these results strongly favor a common evolutionary origin for all three *dam* genes, as opposed to a convergent evolution.

A common evolutionary origin of the *dam* genes in *S. pneumoniae* and *E. coli* genes would have to predate the divergence of gram-positive and gram-negative bacteria. This is contrary to the recent proposal (3) that Dam methylation in the *E. coli* lineage appeared late in bacterial evolution, after divergence from the genus *Pseudomonas*. In that report the authors ignored the presence of Dam methylation in *Streptococcus* species as well as in the gram-negative *Moraxella bovis*. It is possible that active gram-positive restriction and modification genes have been transmitted to *Moraxella*, but only the methylase remains functional in the *Enterobacteriaceae*, *Parvobacteriaceae*, and *Vibronaceae*. Alternatively, the methylase gene (or parts of it) might have been transmitted by phages or plasmids after gram-positive and gram-negative bacteria diverged. Finally, by analogy to T4 and *E. coli*, the failure of the *E. coli dam* gene to hybridize to *Anabaena variabilis* DNA (4) does not necessarily preclude a common ancestor for *E. coli* and the cyanobacteria.

We thank James Pustell for supplying the excellent package of DNA analysis programs and William McAllister for supplying a package of protein analysis programs which included a microcomputer implementation of hydropathicity analysis. We also thank Sanford Lacks for sharing and discussing data prior to its publication.

This work was supported by Public Health Service grants GM 29227 (to S.H.) and GM 13221 (to G.M.) from the National Institutes of Health. P.M.M. was supported by Cellular and Molecular Biology Training Grant T-32 GM 07319 from the National Institutes of Health. J.W. was a prebaccalaureate student.

LITERATURE CITED

1. **Alff-Steinberger, C.** 1984. Evidence for a coding pattern on the non-coding strand of the *E. coli* genome. Nucleic Acids Res. **12**:2235–2241.
2. **Auer, B., and M. Schweiger.** 1984. Evidence that *Escherichia coli* virus T1 induces a DNA methyltransferase. J. Virol. **49**:588–590.
3. **Barbeyron, T., K. Kean, and P. Forterre.** 1984. DNA adenine methylation of GATC sequences appeared recently in the *Escherichia coli* lineage. J. Bacteriol. **160**:586–590.
4. **Brooks, J. E., R. M. Blumenthal, and T. R. Gingeras.** 1983. The isolation and characterization of the *Escherichia coli* DNA adenine methylase (*dam*) gene. Nucleic Acids Res. **11**:837–851.
5. **Brooks, J. E., and S. Hattman.** 1978. In vitro methylation of bacteriophage λ DNA by wild-type (*dam*$^+$) and mutant (*dam*^{h}) forms of phage T2 DNA adenine methylase. J. Mol. Biol. **126**:381–394.
6. **Cowie, D. B., R. J. Avery, and S. P. Champe.** 1971. DNA homology among the T-even bacteriophages. Virology **45**:30–37.
7. **Dunn, D. B., and J. D. Smith.** 1958. The occurrence of 6-methylaminopurine in deoxyribonucleic acids. Nature (London) **68**:627–636.
8. **Fujimoto, D., P. R. Srinivasan, and E. Borek.** 1965. On the nature of the deoxyribonucleic acid methylases. Biological evidence for the multiple nature of the enzyme. Biochemistry **4**:2849–2855.
9. **Gefter, M., R. Hausmann, M. Gold, and J. Hurwitz.** 1966. The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. X. Bacteriophage T3-induced *S*-adenosyl-methionine cleavage. J. Biol. Chem. **241**:1995–2006.
10. **Gold, M., R. Hausmann, U. Maitra, and J. Hurwitz.** 1964. The enzymatic methylation of RNA and DNA. VII. Effects of bacteriophage infection on the activity of the methylating enzymes. Proc. Natl. Acad. Sci. USA **52**:292–297.
11. **Hattman, S.** 1970. DNA methylation of T-even bacteriophages and of their non-glucosylated mutants: its role in P1-directed restriction. Virology **42**:359–370.
12. **Hattman, S.** 1983. DNA methylation, p. 514–518. *In* The enzymes, vol. 14. Academic Press, Inc., New York.
13. **Hattman, S., J. E. Brooks, and M. Masurekar.** 1978. Sequence specificity of the P1-modification methylase (M · *Eco* P1) and the DNA methylase (M · *Eco dam*) controlled by the *E. coli dam* gene. J. Mol. Biol. **126**:367–380.
14. **Hattman, S., H. Van Ormondt, and A. DeWaard.** 1978. Sequence specificity of the wild-type (*dam*$^+$) and mutant (*dam*^{h}) forms of bacteriophage T2 DNA adenine methylase. J. Mol. Biol. **119**:361–376.
15. **Hausmann, R., and M. Gold.** 1966. The enzymatic methylation of ribonucleic and deoxyribonucleic acid. IX. Deoxyribonucleic acid methylase in bacteriophage-infected *Escherichia coli*. J. Biol. Chem. **241**:1985–1994.
16. **Hehlmann, R., and S. Hattman.** 1972. Mutants of bacteriophage T2 *gt* with altered DNA methylase activity. J. Mol. Biol. **67**:351–360.
17. **Kano-Sueoka, T., and N. Sueoka.** 1969. Leucine tRNA and cessation of *Escherichia coli* protein synthesis upon phage T2 infection. Proc. Natl. Acad. Sci. USA **62**:1229–1236.
18. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
19. **Lacks, S., and B. Greenberg.** 1977. Complementary specificity of restriction endonuclease of *Diplococcus pneumoniae* with respect to DNA methylation. J. Mol. Biol. **114**:153–168.
20. **Macdonald, P. M., and G. Mosig.** 1984. Regulation of a new bacteriophage gene, 69, that spans an origin of DNA replication. EMBO J. **3**:2863–2871.
21. **Mannarelli, B. M., T. S. Balganesh, B. Greenberg, S. S. Springhorn, and S. A. Lacks.** 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.
22. **Marinus, M. G., and N. R. Morris.** 1973. Isolation of DNA methylase mutants from *Escherichia coli* K-12. J. Bacteriol. **114**:1143–1150.
23. **Pustell, J., and F. C. Kafatos.** 1982. A convenient and adaptable package of DNA sequence analysis programs for microcomputers. Nucleic Acids Res. **10**:4765–4782.
24. **Revel, H. R., and S. Hattman.** 1971. Mutants of T2 *gt* with altered DNA methylase activity: relation to restriction by prophage P1. Virology **45**:484–495.
25. **Revel, H. R., S. Hattman, and S. E. Luria.** 1965. Mutants of bacteriophages T2 and T6 defective in α-glucosyl transferase.

Biochem. Biophys. Res. Commun. **18**:545–550.
26. **Schildkraut, C. L., K. L. Wierzchowski, J. Marmur, D. M. Green, and P. Doty.** 1962. A study of base sequence homology among the T series of bacteriophages. Virology **18**:43–55.
27. **Schlagman, S. L., and S. Hattman.** 1983. Molecular cloning of a functional *dam*$^+$ gene coding for phage T4 DNA adenine methylase. Gene **22**:139–156.
28. **Sueoka, N., and T. Kano-Sueoka.** 1964. A specific modification of leucyl-sRNA of *Escherichia coli* after phage T2 infection. Proc. Natl. Acad. Sci. USA **52**:1535–1540.
29. **Van Ormondt, H., J. Gorter, K. J. Havelaar, and A. DeWaard.** 1975. Specificity of a deoxyribonucleic acid transmethylase induced by bacteriophage T2. I. Nucleotide sequences isolated from *Micrococcus luteus* DNA methylated in vitro. Nucleic Acids Res. **2**:1391–1400.
30. **Yudelevich, A.** 1971. Specific cleavage of an *Escherichia coli* leucine transfer RNA following bacteriophage T4 infection. J. Mol. Biol. **60**:21–29.