The bacteriophage T2 and T4 DNA-[N⁶-adenine] methyltransferase (Dam) sequence specificities are not identical

Samuel L. Schlagman and Stanley Hattman*

Department of Biology, University of Rochester, Rochester, NY 14627, USA

Received August 14, 1989; Revised and Accepted October 17, 1989

ABSTRACT

Bacteriophages T2 and T4 encode DNA-[N⁶-adenine] methyltransferases (Dam) which differ from each other by only three amino acids. The canonical recognition sequence for these enzymes in both cytosine and 5-hydroxymethylcytosine-containing DNA is GATC; at a lower efficiency they also recognize some non-canonical sites in sequences derived from GAY (where Y is cytosine or thymine). We found that T4 Dam fails to methylate certain GATA and GATT sequences which are methylated by T2 Dam. This indicates that T2 Dam and T4 Dam do not have identical sequence specificities. We analyzed DNA sequence data files obtained from GenBank, containing about 30% of the T4 genome, to estimate the overall frequency of occurrence of GATC, as well as non-canonical sites derived from GAY. The observed N⁶methyladenine (m⁶A) content of T4 DNA, methylated exclusively at GATC (by *Escherichia coli* Dam), was found to be in good agreement with this estimate. Although GATC is fully methylated in virion DNA, only a small percentage of the non-canonical sequences are methylated.

INTRODUCTION

Escherichia coli encodes a DNA [N⁶-adenine] methyltransferase (Dam) that methylates the palindromic tetranucleotide sequence GATC (1–4). Bacteriophages T2 and T4 (but not the related phage T6) also specify Dam proteins (5–9). These phage-encoded enzymes methylate the canonical recognition site, GATC, as well as additional (non-canonical) sites (1,10–15); their derived polypeptide sequences differ from each other by only three amino acids (16). Furthermore, T2 Dam and T4 Dam share four regions of amino acid homology with *E. coli* Dam (17). Three of these are also shared by the *DpnII* modification enzyme, which also methylates the sequence GATC (4), suggesting that one or more of the regions may be involved in DNA sequence recognition (17).

In T-even phage DNA, cytosine (C) is replaced by 5-hydroxymethylcytosine (hmC), which is usually glucosylated; mutants lacking glucosylation are designated *gt*[−]. T2 and T4 Dam methylate hmC-containing DNA more readily than does *E. coli* Dam (12,14). In virion DNA, GAThmC appears to be present at similar frequencies in T2 and T4 and may be fully methylated (14). Nevertheless, the m⁶A content of T2*gt*[−]*dam*⁺ virion DNA is almost two-fold higher than in T4*gt*[−]*dam*⁺ virion DNA (8). This suggests that more non-canonical sites are methylated in T2*gt*[−]*dam*⁺ than in T4*gt*[−]*dam*⁺. This is supported by the finding that T2 Dam methylates non-canonical sequences *in vitro* more efficiently than does T4 Dam (1).

The T4 *dam*⁺ and T2 *dam*⁺ genes can mutate to a form designated *dam*^h, which encodes an enzyme capable of methylating non-canonical sequences much more efficiently than

Table 1. GenBank sequence files of bacteriophage T4 used in the analysis

File name ^a	Number of nucleotides ^b	Genes identified ^c
pT4rIIAB.ph	1206	rIIA and rIIB; partial
pT4dexA.ph	1334	dexA; complete
pT4uvsyso.ph	8062	uvsX, 41, 61/58, dam, 69, and soc; complete
pT4bgt.ph	1200	β gt; complete
pT4g55g43.ph	12,977	55, α gt, 47, 46, 44, 62, regA, and 43; complete
pT4denv.ph	569	denV; complete
pT4e.ph	988	e; complete / ipIII; partial
pT4g1trna.ph	5015	1, 57A, 57B, ipl, tRNA-arg, tRNA-gln, tRNA-gly, tRNA-ile, tRNA-leu, tRNA-pro, tRNA-ser, tRNA-thr; complete /3; partial
pT4g67g21.ph	767	67 and 68; complete/ 21; partial
pT4g22g23.ph	2280	23; complete/ 22; partial
pT4uvsy.ph	1001	uvsY; complete
pT4g30.ph	1993	30; complete
pT4pset.ph	2393	pseT; complete
pT4g63alc.ph	1781	63 and alc; complete
pT4frtd1.ph	1482	frd and td; complete
pT4g32.ph	1340	32; complete
pT4g36g37.ph	4418	36 and 37; complete/ 35 and 38; partial

^a The files are arranged by their map location clockwise relative to genes rIIA and rIIB.

^b The nucleotides in each file are those found on the non-coding strand.

^c Only genes which have been definitively identified within these files are listed. We have indicated whether the file contains the entire gene or not. Some of these files also contain open reading frames for which gene products have not yet been identified.

the wild-type enzyme (1,8,10,11,18,19). Thus, the m⁶A content of *dam*^h virion DNA is two to three times higher than its parental *dam*⁺ form (8); and the m⁶A content of T2gt⁻*dam*^h virion DNA is at least 1.5 fold higher than T4gt⁻*dam*^h (8). Miner et al. (20) have shown that T4 Dam^h and T2 Dam^h have a proline to serine change at amino acid 126. This residue is located within one of the three amino acid homology regions (Region III) mentioned earlier. Finally, *dam*^h phage can mutate to *dam*^h*dam*-x forms which lack enzyme activity and are devoid of m⁶A (19). The double mutants are viable and show no obvious phenotypic differences from *dam*⁺ phage (21). Thus, at the present time the biological function of the T2 and T4 Dam proteins is obscure.

Hattman et al. (11) reported that T2 Dam and T2 Dam^h recognize the sequence GAY, where Y is C (or hmC) or T. However, not all sequences containing GAY are methylated at similar rates, and some may not be methylated at all (1,10). In this communication, we estimate the frequencies at which sequences that can be derived from GAY occur in T2 and T4 DNAs and verify experimentally that GAThmC is found at similar frequencies in both genomes. We show that Dam^h methylates GAThmC sequences *in vivo* as well as the parental Dam form, and we present evidence that neither T4 Dam nor T4 Dam^h are capable of methylating all the sequences recognized by T2 Dam; i.e. , the former appear to be unable to methylate adenine in the sequences GATT and GATA.

MATERIALS AND METHODS

Bacteria, bacteriophage, and plasmids. *Escherichia coli* K704 supII *mcrA mcrB* was from B. de Groot. *E. coli* GM119 supE44 *dam*-3 *dcm*-6 was from J.E. Brooks. *E. coli* GM124

Table 2. Predicted m⁶A content from various methylated GAY sequences in T4 DNA

Sequence ^a	Observed occurrence in the non-coding strand ^b	Average number of methylation sites per strand ^c	Predicted mol % m ⁶ A produced by complete methylation ^d
5' GA*TC	65	65	0.41
5' GA*TA	303	281.5	1.77
5' TATC	260		
5' GA*TG	397	266	1.68
5' CATC	135		
5' GA*TT	367	329	2.07
5' AATC	291		
5' GA*CC	108	119.5	0.75
5' GGTC	131		
5' GA*CT	118	101	0.64
5' AGTC	84		
5' GA*CA	90	88.5	0.56
5' TGTC	87		
5' GA*CG	84	87.5	0.55
5' CGTC	91		

^a Complementary non-palindromic sequences are grouped together. The asterisk indicates the potential target A for methylation.

^b The values represent the total number of occurrences found in the GenBank data base of the files summarized in Table 1.

^c For palindromic sequences this is the number of times the sequence occurs in the data base since it occurs the same number of times in both strands. For non-palindromic sequences only one sequence of each complementary pair is a potential methylation site. Therefore, we add the total number of times a sequence and its complement occur in the data base and divide by two.

^d Calculated from the equation $[(c)/15,861] \times 100$. 15,861 is the average number of A residues per strand.

dam-4 mcrA was from M.G. Marinus. *E. coli* B834 was from J. Wiberg. Bacteriophages T2 $\alpha_{gt}^{-1}dam^{+}$ (designated T2 $gt^{-}dam^{+}$), T2 $\alpha_{gt}^{-1}dam^h$ (designated T2 $gt^{-}dam^h$), T2 $gt^{-1}dam^hdam1$ (designated T2 $gt^{-}dam^hdam1$), T4 $\alpha_{gt}^{-1}\beta_{gt}^{-27}dam^{+}$ (designated T4 $gt^{-}dam^{+}$) and T4 $\alpha_{gt}^{-1}\beta_{gt}^{-27}dam^h$ (designated T4 $gt^{-}dam^h$) were from H.R. Revel. T4 $\alpha_{gt}^{-1}\beta_{gt}^{-27}dam^hdam1$ (designated T4 $gt^{-}dam^hdam1$) was isolated from T4 $gt^{-}dam^h$ by M. Myers in this laboratory. Bacteriophage T4 *amN55X5* (gene 42), *amE51* (gene 56), *nd28* (gene *denA*), *D2a2* (gene *denB*), *alc10*, *dam^{+}* (designated T4cJW819*dam^{+}*) was from J. Wiberg. Bacteriophages T4 *amN55X5* (gene 42), *amE51* X5 (gene 56), *nd28* (gene *denA*), *D2a2X6* (gene *denB*), *alc2*, *dam1* (designated T4cS307*dam^{-}*) and T4 *amN55X5* (gene 42), *amE51* X5 (gene 56), *nd28* (gene *denA*), *D2a2X5* (gene *denB*), *alc2*, *dam^{+}* (designated T4cS334*dam^{+}*) were from K. Carlson. Plasmid pSSH-10, containing the T4 *dam^{+}* gene, has been previously been described (14). Plasmid pTP166, containing the *E. coli dam^{+}* gene under the *tac* promoter, was obtained from M.G. Marinus (22).

Analysis of T4 DNA sequences. Seventeen GenBank DNA sequence files of bacteriophage

T4 were used for the analysis. They vary in size from 569 to 12,977 nucleotides. They are distributed all over the T4 genome and contain early, middle and late genes, as well as tRNAs. All seventeen files contain the DNA sequence found in the non-coding strand, corresponding to the RNA transcript sequences (summarized in Table I). The University of Wisconsin Genetics Computer Group software package was used to scan these data files in order to determine the number of occurrences (in the non-coding strand) of each sequence of interest.

Chemicals and buffers. Restriction enzymes, *E. coli* Dam, DNA molecular weight markers, S-adenosyl methionine (AdoMet) and BSA were from New England Biolabs (NEB) and Bethesda Research Laboratories (BRL). All enzymes were used under the conditions specified by the manufacturer, unless stated otherwise. The buffer for *Hinf*I digestion of hmC-containing DNA was 10mM Tris-HCl (pH 7.5), 50 mM NaCl, 10mM MgCl₂, 5mM DTT and 1mg/ml BSA. EDTA, TRIS, pancreatic RNase A, DNase I, and BAP (Sigma), phosphodiesterase I (Worthington Biochemical) and agarose (BioRad) were from the manufacturers listed in parenthesis. Agarose gels contained TAE buffer (23).

Growth of bacteriophage, purification and preparation of DNA for analysis by high-performance liquid chromatography. The growth of T4cJW819 and subsequent virion DNA purification has been described previously (12). With the exception of T4cS307*dam*⁻ and T4cS344*dam*⁺, other bacteriophages were grown and the unlabeled virion DNA purified as described by Schlagman et al. (14). Bacteriophage T4cS307*dam*⁻ and T4cS344*dam*⁺ were propagated in *E. coli* GM124 *dam*-4 (phage used to infect *E. coli* GM124 *dam*-4 were grown previously in *E. coli* GM119 *supE*44 *dam*-3 *dcm*-6 to ensure they did not contain m⁶A produced by *E. coli* Dam methylation).

The same DNA was used for restriction enzyme and HPLC analyses. DNA was digested to deoxyribonucleosides (24) and fractionated isocratically at 37°C (14). We found that modifying the buffer pH from 4.1 to 3.58 gave improved separation from a contaminant (sometimes found in T2 DNA preparations) that eluted very near to m⁶dA. If similar results were obtained using both pH conditions, they were averaged to obtain the m⁶A content. If there was a discrepancy, the m⁶A content obtained with pH 3.58 buffer was used.

RESULTS

Estimated occurrence of various GAY sequences in the T4 genome.

Hattman et al. (11) showed that T2 Dam methylates some subset of the sequence, GAY. Therefore, we analyzed the GenBank T4 DNA sequence files (see Table 1) to determine the number of occurrences of each sequence that can be derived from GAY. The DNA sequences contained within these files represented approximately 30% of the T4 genome. From this information, we estimated the frequency at which of each of these potential methylation sites occurs in the T4 genome and the mol % m⁶A that would result from complete methylation at each of these sites. As seen in Table 2, the canonical sequence GATC (GATHmC) is the least abundant of all the potential methylation sequences. It is, therefore, the sequence which would be expected to show the largest error in estimating its occurrence. However, experimental measurement of the GATC frequency in the entire T4 genome would indicate how representative the DNA data files are. This is described below.

Table 3. Mol % m⁶A in virion DNA before and after *in vitro* methylation by M·Eco Dam

DNA ^a	Mol % m ⁶ A prior to <i>in vitro</i> methylation ^b	Mol % m ⁶ A following <i>in vitro</i> methylation ^{b,c}	Mol % m ⁶ A due to methylation of only GATC sequences predicted from computer analysis ^d
T4cS307 <i>dam</i> ⁻	<0.02	0.37	0.41
T4cS334 <i>dam</i> ⁺	0.39	0.38	0.41

^a T4cS307 *dam*⁻ and T4cS334 *dam*⁺ C-containing virion DNAs were isolated from purified phage after propagation in *E. coli* GM124 *dam*⁻.

^b The mol % m⁶A was determined by HPLC analysis before and after *in vitro* methylation with purified M·Eco Dam and unlabeled AdoMet. The results are the averages from at least two independent digests and HPLC analysis, which generally varied by less than 10% from the average.

^c Thirty (30) µg of DNA in 300µl of buffer containing 25 mM Tris-HCl (pH 7.5), 5mM EDTA (pH 8.0), 1mM DTT, 0.03 mg of BSA and 150 U of M·Eco Dam was incubated at 37°C for 4 hr. The DNA was precipitated with ethanol, pelleted by centrifugation, air dried and resuspended in 120µl of water. M·Eco Dam was inactivated by heating at 65°C for twenty min. Five (5) µg of DNA was removed for restriction enzyme analysis (see Figure 1) and the remainder was digested for HPLC analysis.

^d For T4 only 30 % of the genome sequence was known. The prediction is based on the complete methylation of the A in the sequence 5' GATC.

Phage T4c contains mutations which result in the production of C-containing DNA instead of hmC-containing DNA. T4cS307*dam*⁻ is a T4c phage which encodes a defective Dam, whereas T4cS334*dam*⁺ is an isogenic form which encodes an active enzyme. As seen in Table 3, when T4cS307*dam*⁻ is propagated in *E. coli dam*⁻ cells, the phage DNA is devoid of m⁶A; when T4cS334*dam*⁺ is propagated in the same host, virion DNA contains 0.39 mol % m⁶A. This is the same value (0.40 mol%) obtained by a different method for T4cJW819*dam*⁺ propagated in the same *dam*⁻ host (14). *E. coli* Dam methylates the sequence GATC only (1–3); as seen in Table 3, the mol % m⁶A produced by *E. coli* Dam methylation *in vitro* of T4cS307*dam*⁻ DNA is identical to that for T4cS334*dam*⁺ DNA isolated from phage propagated in *E. coli dam*⁻ cells. Furthermore, T4cS334*dam*⁺ DNA has the same m⁶A content prior to and following *in vitro* methylation by *E. coli* Dam (Table 3). These results indicate that viral GATC's are fully methylated *in vivo* by T4 Dam, and that methylation is almost exclusively in this sequence. Furthermore, the three values of mol % m⁶A in T4c DNA are similar (within 10%) to the one calculated from the DNA sequence files assuming complete methylation of GATC (see Table 2). This indicates that the occurrence of GATC in the DNA data base files is representative of the entire T4 genome and that estimations of other sequences (Table 2) are probably not off by more than 10% from actual values.

In order to confirm that GATC sequences in T4cS334*dam*⁺ phage DNA are fully methylated *in vivo* by T4 Dam, we carried out a restriction enzyme analysis using *Mbo*I, *Dpn*I, and *Sau*3AI. *Mbo*I cuts at GATC only when the A is unmethylated on both strands. *Dpn*I cuts at GATC only when the A is methylated on both strands. *Sau*3AI cuts at GATC sequences independent of A methylation. T4cS334*dam*⁺ DNA was refractory to *Mbo*I but sensitive to *Dpn*I (not shown). Furthermore, the *Dpn*I and *Sau*3AI digestion patterns

Table 4. Mol % m⁶A in various T2 and T4 virion DNAs following propagation in cells harboring a T4 *dam*⁺ or an *E. coli dam*⁺ plasmid^{a,b}

Phage	Plasmid	Methyltransferase produced by the plasmid	Mol % m ⁶ A
T2 <i>gt</i> ⁻ <i>dam</i> ^h <i>dam</i> 1	none	none	<0.02
	pTP166	<i>Eco</i> Dam	0.34
	pSSH-10	T4 Dam	0.43
T4 <i>gt</i> ⁻ <i>dam</i> ^h <i>dam</i> 1	none	none	<0.05
	pTP166	<i>Eco</i> Dam	0.28
	pSSH-10	T4 Dam	0.40
T4cS307 <i>dam</i> ⁻	none	none	<0.02
	pSSH-10	T4 Dam	0.42

^a The mol % m⁶A was determined by HPLC.

^b T2*gt*⁻*dam*^h*dam*1 and T4*gt*⁻*dam*^h*dam*1 (hmC-containing DNA phages) were grown in K704. T4cS307*dam*⁻ (C-containing DNA) was grown in GM124*dam*-4.

were identical (this can occur only if the DNA is fully methylated at all GATC sequences). As expected, the digestion pattern for T4cS334*dam*⁺ DNA was the same prior to and following *in vitro* methylation by *E. coli* Dam (not shown). These results confirm that T4c GATC sequences are fully methylated *in vivo*.

T4cS307*dam*⁻ DNA, isolated from phage grown in *E. coli dam*⁻ cells, was analyzed in a similar fashion. Prior to *in vitro* methylation by *E. coli* Dam, this DNA was sensitive to *Mbo*I, refractory to *Dpn*I, and had identical *Mbo*I and *Sau*3AI digestion patterns (data not shown); this occurs only when all the GATC sites in the DNA are devoid of m⁶A. Following *in vitro* methylation by *E. coli* Dam, this DNA was refractory to *Mbo*I, sensitive to *Dpn*I and had identical *Dpn*I and *Sau*3AI digestion patterns (data not shown). Thus, consistent with the m⁶A contents (Table 3) GATC sequences are fully methylated *in vitro* by *E. coli* Dam.

The frequency of GATC sequences is similar in T2 and T4.

Phages T2 and T4 show DNA sequence homology of greater than 85% (25). This suggests that the various T2 Dam sequences (GAY) are likely to occur at similar frequencies in T2 and T4. Consistent with this notion is the finding by Schlagman et al. (14) that T4*gt*⁻*dam*^h*dam*1 and T2*gt*⁻*dam*^h*dam*1 (normally devoid of m⁶A) have similar m⁶A contents (0.42 mol % to 0.45 mol %) following growth in cells harboring a cloned T4 *dam*⁺ gene. We estimate that complete methylation of GATC (GAThmC) sequences would produce approximately 0.4 mol % m⁶A (Tables 2 and 3); it appears, therefore, that the GAThmC content of phages T2 and T4 are similar and that these sequences are fully methylated in T4*gt*⁻*dam*^h*dam*1 and T2*gt*⁻*dam*^h*dam*1 DNAs after growth in cells harboring a T4 *dam*⁺ clone. We verified this by restriction enzyme analysis using *Mbo*I and *Dpn*I (we did not use *Sau*3AI since this enzyme is unable to digest hmC-containing DNA). Phage T2*gt*⁻*dam*^h*dam*1 DNA (propagated in *E. coli* K704) was sensitive to *Mbo*I, resistant to *Dpn*I and devoid of m⁶A (Table 4; restriction digests not shown). However, after growth in *E. coli* K704 [pSSH-10], phage T2*gt*⁻*dam*^h*dam*1 was sensitive to *Dpn*I, refractory to digestion by *Mbo*I and contained 0.43 mol % m⁶A (Table 4; restriction

Table 5. Methylation of non-canonical GAY sequences in various T2 and T4 phage DNAs

Phage ^a	Mol % m ⁶ A in virion DNA ^b	% of total m ⁶ A in GATC sequences ^c	% of total m ⁶ A in GAY non- canonical sequences ^d	% of GAY non-canonical sequences methylated ^e
T2 <i>gt⁻dam⁺</i>	0.85	48.2	51.8	5.49
T2 <i>gt⁻dam^h</i>	2.44	16.8	83.2	25.3
T4 <i>gt⁻dam⁺</i>	0.58	70.7	29.3	2.12
T4cJW819 <i>dam⁺</i>	0.43	95.3	4.7	0.25
T4 <i>gt⁻dam^h</i>	1.43	28.7	71.3	12.7

^a T4cJW819*dam⁺* (C-containing DNA) was grown in B834; the other phages (hmC-containing DNA) were grown in K704.

^b See Table 4 footnote a.

^c Calculated from the equation $[0.41/(\text{column b value})] \times 100$. 0.41 is the estimated mol % m⁶A from resulting from complete methylation of GATC (Table 2).

^d $d = 100 - (c)$

^e Calculated from the equation $[(b - 0.41)/8.02] \times 100$. 8.02 is the estimated mol % m⁶A resulting from complete methylation of all non-canonical GAY sequences (Table 2).

digests not shown). Furthermore, these *Mbo*I and *Dpn*I digestion patterns are identical to those of T2*gt⁻dam⁺* and T2*gt⁻dam^h* DNAs (propagated in *E. coli* K704; data not shown). This indicates that most, if not all, of the GAThmC sites in T2*gt⁻dam^hdam1*, in T2*gt⁻dam⁺* and in T2*gt⁻dam^h* virion DNAs were fully methylated *in vivo* by T4 Dam, T2 Dam and T2 Dam^h, respectively. Finally, after propagation in *E. coli* K704 [pTP166], which overexpresses *E. coli* Dam, phage T2*gt⁻dam^hdam1* DNA contains only 0.34 mol % m⁶A (Table 4); as expected, this DNA was not fully methylated at GAThmC sites since it was partially sensitive to digestion with *Mbo*I and partially resistant to digestion with *Dpn*I (data not shown). Similar results were also obtained for T4 virion DNAs (Table 4; restriction digests not shown) indicating that most if not all GAThmC (GATC) sequences are fully methylated *in vivo* by T4 Dam and T4 Dam^h. Since full methylation of T2 GAThmC sites must produce between 0.34 and 0.43 mol % m⁶A (Table 4), we conclude that T2 and T4 DNAs contain similar frequencies of GAThmC and that non-canonical sites (derived from GAY) probably occur at similar frequencies, as well. Hence, the calculations made for T4 DNA (Table 2) are also valid for T2.

Only a fraction of non-canonical sites are methylated in phage DNA.

The frequency of occurrence of GAThmC, as well as non-canonical sequences derived from GAY, is probably similar in both T4 and T2 DNAs. To estimate the amount of methylation that occurs in non-canonical sequences, we subtract 0.41 mol% m⁶A (resulting from full methylation of GAThmC) from the observed total mol % m⁶A. It follows that less than 5% of the total DNA methylation in T4cJW819*dam⁺* occurs at non-canonical sequences (Table 5), whereas 29% of the total methylation in T4*gt⁻dam⁺* occurs at these sequences. Thus, *in vivo*, T4 Dam methylates more non-canonical sequences on T4 hmC-containing DNA than on T4 C-containing DNA. Furthermore, more than 50% of the methylation in T2*gt⁻dam⁺* occurs at non-canonical sequences; this methylation level is intermediate between those found in T4*gt⁻dam⁺* and T4*gt⁻dam^h* (Table 5). Finally, although non-canonical sequences are methylated to the greatest extent in T2*gt⁻dam^h* (83.2% of the total DNA methylation), this represents methylation of only 25.3% of the available non-canonical sequences.

Differential recognition of the sequences GATA and GATT.

The *EcoRV* restriction enzyme recognizes GATATC (GATAThmC) and cuts within this sequence if the 5' A is not methylated in at least one strand (26–29). Since this site also belongs to the non-canonical sequence GATA, we analyzed the susceptibility of various $T2gt^-$ and $T4gt^-$ DNAs to cleavage by *EcoRV*. As seen in Figure 1 (lanes c and d), the extent of digestion of unmethylated $T2gt^-dam^hdam1$ DNA was a function of the enzyme concentration used. We will discuss only the digestion patterns obtained at the higher enzyme concentration.

Compared to unmethylated $T2gt^-dam^hdam1$, $T2gt^-dam^+$ DNA showed evidence of partial protection against *EcoRV*; this is readily observed by the presence of additional bands between the 6.6 kb and 23 kb molecular weight markers (Figure 1, lanes d and f), as well as the presence of an additional band located between 2.3 kb and 4.36 kb. Furthermore, the number and intensity of these additional bands increase for $T2gt^-dam^h$ compared to $T2gt^-dam^+$ (Figure 1, lanes f and h). These additional bands do not appear to be the result of incomplete digestion, since they are not present in the $T2gt^-dam^hdam1$ digest (lane d). Thus, although T2 Dam methylates some GATAThmC sequences, T2 Dam^h methylates them to a greater degree. Nevertheless, only a small percentage of such sequences are methylated by either enzyme.

In contrast, $T4gt^-$ DNA showed no protection against *EcoRV* digestion (data not shown). This indicates that T4 Dam and T4 Dam^h do not methylate GATAThmC sequences *in vivo*.

The *HinfI* restriction enzyme recognizes the sequence GANTC (GANThmC) (where N is any base), and cuts DNA within this sequence if the A is not methylated in at least one strand (27,28). The sequence GANTC is composed of two non-palindromic complementary pairs of sequences, namely GATTC/GAATC and GACTC/GAGTC. Each of these pairs of non-palindromic sequences is a subset of a different non-canonical sequence. Thus, we could use *HinfI* to investigate the degree of methylation occurring in a subset of the non-canonical sequences GATT and/or GACT. Computer analysis of the T4 DNA sequence data files summarized in Table 1 suggests that the non-palindromic complementary pair GATTC/GAATC comprises 78% of the *HinfI* sequences found in the T4 genome (and presumably also in the T2 genome) and that the non-palindromic complementary pair GACTC/GAGTC comprises the remaining 22% of these sequences. Thus, the vast majority of *HinfI* sites in the phage DNA are a subset of the non-canonical sequence GATT.

As seen in Figure 2, although the *HinfI* digestion pattern of $T2gt^-dam^+$ is virtually identical to that of $T2gt^-dam^hdam1$, the *HinfI* digestion pattern of $T2gt^-dam^h$ shows a shift to higher molecular weight fragments. We interpret these results to mean that T2 Dam methylates only a small percentage of the *HinfI* sites in the $T2gt^-dam^+$ [it should be noted that methylation of 25% of the GACTC/GAGTC sequences would result in methylation of only 5.5% ($.219 \times .25 \times 100$) of all the *HinfI* sites and would probably not be detected as a shift in the digestion pattern. Thus, a considerable number of methylated GACTC/GAGTC sequences could be present in the $T2gt^-dam^+$ DNA and still not be

Figure 1. The susceptibility of various $T2gt^-$ phage DNAs to cleavage by the restriction enzyme *EcoRV*. After growth in *E. coli* K704 virion DNAs were purified and digested with *EcoRV*. Samples were loaded on an 0.8% agarose gel containing TAE buffer (23). Lanes c, e, and g contain 1.5 μ g of $T2gt^-dam^hdam1$, $T2gt^-dam^+$, and $T2gt^-dam^h$ DNAs, respectively digested with 30 U of *EcoRV* for 20 h; lanes d, f, and h contain the same DNAs digested with 60 U of *EcoRV* for 20 h. ϕ X174RF *Hae*III (lanes a and i) and λ -*Hind*III (lanes b and j) marker DNAs.

Figure 2. The susceptibility of various $T2gt^-$ phage DNAs to cleavage by the restriction enzyme *HinfI*. Two μg of various $T2gt^-$ DNAs (propagated in *E. coli* K704) were digested for 15 h with 50 U of *HinfI* and loaded in lanes a–c of a 1% agarose gel containing TAE buffer (23). $T2gt^-dam^hdam1$ DNA (lane a); $T2gt^-dam^+$ DNA (lane b); and $T2gt^-dam^h$ DNA (lane c). $\phi\text{X174RF} \cdot \text{HaeIII}$ (lane d) and $\lambda \cdot \text{HindIII}$ DNA (lane e) marker DNAs.

detected by this method]. However, it is evident that T2 Dam^h methylation increases protection against *HinfI*.

$T4gt^-dam^hdam1$, $T4gt^-dam^+$ and $T4gt^-dam^h$ all have virtually identical *HinfI*

digestion patterns (data not shown). This shows that T4 Dam and T4 Dam^h methylated very few, if any, *Hinf*I sites. Thus, T4 Dam and T4 Dam^h appear to be unable to methylate the non-canonical sequence GATT.

DISCUSSION

We have shown that the occurrence of GATC in the DNA sequence data base files listed in Table 1 is representative of the entire T4 and T2 genomes (Tables 3 and 4). This suggests that these files can be used to provide reliable estimates for the abundance of other sequences in T2 and T4, which are found at frequencies equal to or greater than that of GATC. We have used these files to estimate the occurrence of GAY non-canonical sequences (Table 2).

McClelland used a 2nd order Markov chain analysis on a smaller DNA sequence data base to determine the expected occurrence of GATC in the T4 genome; he concluded that this sequence was significantly under-represented (30). We analyzed the DNA sequence data base files (Table 1) in the same manner and have confirmed McClelland's original findings (unpublished). This suggests that there may be selection against GATC sequences in T4 (and in the other T-even phages). If this is true, then this sequence should occur in the greatest abundance in those regions in which it serves a biological function. In this regard, in *E. coli*, GATC is more abundant in open reading frames (31), in the *oriC* region (where it is implicated in DNA replication) (32), and in regions affecting transcription-regulation (33,34).

Although T4cS307dam⁻ and T4gt⁻dam^hdam-1 were methylated to the same extent when grown in cells harboring the T4 dam⁺ clone pSSH-10 (Table 3), T4cS344dam⁺ and T4cJW819dam⁺ have fewer non-canonical sequences methylated than T4gt⁻dam⁺ (Tables 3 and 5). Schlagman et al. (14) found that the ability of T4 Dam to methylate non-canonical sequences *in vivo* depends on the enzyme concentration in the cell; thus, it would appear that cells infected with T4c produce less T4 Dam than cells infected with T4. In addition, the methylation of non-canonical sites may be different for hmC-vs C-containing DNA.

The increase in the number of non-canonical sequences methylated in T2 compared to T4 may also be due, in part, to production of more Dam protein in T2-infected cells. Doolittle and Sirotkin found that T2 Dam is more efficient than T4 Dam at methylating non-canonical sequences *in vitro* (1). We have found that *in vivo* T4 Dam is not able to methylate all of the non-canonical sequences that are methylated by T2 Dam; viz, certain GATT and GATA sequences. It is not clear to what extent differences in enzyme level and sequence specificity contribute to the differences in T2 and T4 methylation levels.

One (or more) of the three amino acid differences between T2 Dam and T4 Dam must be responsible for their slightly different sequence specificities. In this regard it is interesting to note that two of these amino acid changes occur in a region of homology (Region I) shared by T2 Dam, T4 Dam, *E. coli* Dam, and *Dpn*I (17) all of which methylate the sequence GATC (1-4,10-15). Furthermore, these residues are located in a region different than the one responsible for the Dam^h phenotype (Region III) (20). Thus, two domains may be involved in DNA sequence recognition.

ACKNOWLEDGEMENTS

We wish to thank Dr. Karin Carlson for providing us with bacteriophages T4cS307dam⁻ and T4cS334dam⁺. This work was supported by PHS grant no. GM29227.

*To whom correspondence should be addressed

REFERENCES

1. Doolittle, M.M. and Sirotkin, K (1988) *Biochim. Biophys. Acta* **949**, 240–246.
2. Geier, G.E., and Modrich, P. (1979) *J. Biol. Chem.* **254**, 1408–1413.
3. Hattman, S., Brooks, J. E., and Masurekar, M. (1978) *J. Mol. Biol.* **126**, 367–380.
4. Lacks, S., and Greenberg, B. (1977) *J. Mol. Biol.* **114**, 153–168.
5. Fujimoto, D., Srinivasan, P. R., and Borek, E. (1965) *Biochemistry* **4**, 2849–2855.
6. Gefter, M., Hausmann, R., Gold, M. and Hurwitz, J. (1966) *J. Biol. Chem.* **241**, 1995–2006.
7. Gold, M., Hausmann, R., Maitra, U. and Hurwitz, J. (1964) *Proc. Natl. Acad. Sci. USA* **52**, 292–297.
8. Hattman, S. (1970) *Virology* **42**, 359–370.
9. Hausmann, R., and Gold, M. (1966) *J. Biol. Chem.* **241**, 1985–1994.
10. Brooks, J.B. and Hattman, S. (1978) *J. Mol. Biol.* **126**, 381–394.
11. Hattman, S., Van Ormondt, H. and deWaard, A. (1978) *J. Mol. Biol.* **119**, 361–376.
12. Schlagman, S.L. and Hattman, S. (1983) *Gene* **22**, 139–156.
13. Schlagman, S.L., Hattman, S. and Marinus, M.G. (1986) *J. Bacteriol.* **165**, 896–900.
14. Schlagman, S.L., Miner, Z., Fehér, Z. and Hattman, S. (1988) *Gene* **73**, 517–530.
15. Van Ormondt, H., Gorter, J., Havelaar, K. J. and deWaard, A. (1975) *Nucleic Acids Res.* **2**, 1391–1400.
16. Miner, Z. and Hattman, S. (1988) *J. Bacteriol.* **170**, 5177–5184.
17. Hattman, S., Wilkinson, J., Swinton, D., Schlagman, S., Macdonald, P. M. and Mosig, G (1985) *J. Bacteriol.* **164**, 932–937.
18. Hehlmann, R. and Hattman S. (1972) *J. Mol. Biol.* **67**, 351–360.
19. Revel, H.R. and Hattman, S. (1971) *Virology* **45**, 484–495.
20. Miner, Z., Schlagman, S. and Hattman, S. (1988) *Gene* **74**, 275–276.
21. Georgopoulos, C.P. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge MA, 1969.
22. Marinus, M.G., Poteete, A. and Arraj, J.A. (1984) *Gene* **28**, 123–125.
23. Maniatis, T., Fritsch, E.F. and Sambrook, J. *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982 p.454
24. Scarbrough, K., Hattman, S. and Nur, U. (1984) *Mol. Cell Biol.* **4**, 599–603.
25. Kim, J.-S. and Davidson, N. (1974) *Virology* **57**, 93–111.
26. Kholmina, G.V., Rebentish, B.A., Skoblov, Yu.S., Mironov, A.A., Yankovskii, N.K., Kozlov, Yu.I., Glatman, L.I., Moroz, A.F., Debabov, V.G. (1980) *Dokl. Acad. Nauk.* **253**, 495–497.
27. McClelland, M. and Nelson, M. (1988) *Gene* **74**, 291–304.
28. Nelson, M., Christ, C. and Schildkraut, I. (1984) *Nucleic Acids Res.* **12**, 5165–5173.
29. Schildkraut, I., Banner, C.D.B., Rhodes, C.S. and Parekh, S. (1984) *Gene* **27**, 327–329.
30. McClelland, M. (1985) *J. Mol. Evol.* **21**, 317–322.
31. Barras, F. and Marinus, M.G. (1988) *Nucleic Acids Res.* **16**, 9821–9838.
32. Zyskind, J.W. and Smith, D.W. (1986) *Cell* **46**, 219–235.
33. Marinus, M.G. (1987) *Ann Rev. Genet.* **21**, 113–131.
34. Sternberg, N. (1985) *J. Bacteriol.* **164**, 490–493.

<p>This article, submitted on disc, has been automatically converted into this typeset format by the publisher.</p>