Volume 220, number 1, 167–176 FEB 04983 August 1987

The GATATC-modification enzyme *Eco*RV is closely related to the GATC-recognizing methyltransferases *Dpn*II and *dam* from *E. coli* and phage T_4

Roland Lauster*, Anastasios Kriebardis and Wilhelm Guschlbauer

*Max-Planck-Institut für molekulare Genetik, Ihnestrasse 73, D-1000 Berlin 33, Germany and Service de Biochimie, Département de Biologie, Bât. 142, Centre d'Etudes Nucléaires de Saclay, F-91191 Gif-sur-Yvette Cédex, France

Received 12 June 1987

The amino acid sequence of *Eco*RV DNA methyltransferase which methylates the amino group of the 5'-adenine residue of the target sequence GATATC has been found to be closely related to that of three other adenine methyltransferases, *Dpn*II, *dam* and *dam*T_4, the target sequence of which is GATC. Despite large differences on the DNA level, the four sequences show four blocks of homologies. One of these blocks has the sequence DVYXDPPY and is found with little modification in numerous other DNA methyltransferases. It is speculated that it could be the binding site of the methyl donor, *S*-adenosylmethionine. On the other hand, the identification of a DNA-binding region is more tenuous. As expected, no analogies with (dimeric) repressors and *cro* proteins which have the characteristic helix-turn-helix motif have been observed.

DNA recognition; Secondary structure prediction; Sequence homology; *S*-Adenosylmethionine; Molecular evolution; Enzyme domain

1. INTRODUCTION

While type II restriction endonucleases generally act as dimers, the respective modification methyltransferases appear to prefer the monomeric state [1]. Methyltransferases which methylate the amino group of adenine are particularly interesting: not only do there exist methyltranferases associated with the restriction system, but also certain methyltransferases appear to act alone. The best known is *dam* methylase, which specifically methylates the second adenine of the hemimethylated sequence GATC. This enzyme has been associated with strand discrimination in mismatch repair in *E. coli* [2], as well as in the expression of certain genes [3,4] and in replication [5–7]. Adenine methylation has profound effects on the dynamics of the structure of DNA: the opening and closing rates of the $m^6A \cdot T$ base pairs are greatly reduced, suggesting the existence of a kinetic recognition mechanism in methylated GATC sites [8–10].

The recent interest in methyltransferases has stimulated the cloning and sequencing of their genes [11]. Of the adenine methyltransferases the sequences of which are known, three recognize the same DNA sequence, GATC; these are *dam* methylase from *E. coli* [12], *dam* methylase from bacteriophage T_4 [13] and the *Dpn*II methyltransferase from *D. pneumoniae* [14]. The modification enzyme associated with *Eco*RV nuclease methylates the 5'-adenine of the sequence GATATC [15], while *Eco*RI methyltransferase methylates the central adenine in GAATTC [16,17].

Correspondence address: W. Guschlbauer, Service de Biochimie, Département de Biologie, Bât. 142, Centre d'Etudes Nucléaires de Saclay, F-91191 Gif-sur-Yvette Cédex, France

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

Similarities between the *Dpn*II methyltransferase and *dam* methylase have been recognized [14], as well as with the *dam*T_4 enzyme [18]. We show that the similarities are even more pronounced between *Eco*RV and the three GATC-recognizing methyltransferases than between these three enzymes themselves. On the other hand, only limited sequence similarities with other methyltransferases have been observed.

All DNA methyltransferases use a common methyl donor: *S*-adenosylmethionine (Ado-Met) [19]. Thus, any similarities between enzymes have to take into account both the binding to different DNA target sequences and to Ado-Met.

2. METHODS

The fast dot-matrix program DPSA.A of Marck [20] was used for searching for sequence homologies in the nucleotide and amino acid sequences of the methyltransferases. Diagonals in the matrix plots permitted the localization of homologous blocks. This program also permits determination of the codon usage of the various DNA sequences. The program runs on an Apple IIe microcomputer with the 80-column extended memory card.

Secondary structure predictions performed in Saclay used the GOR method [21], using a scanning window of 17 amino acids, and that of Chou and Fasman [22]. The hydropathy index determined by the method of Kyte and Doolittle [23], using a scanning window of 11 amino acids, was used to search for surface and interior parts of the protein. The three methods were programmed on an Apple IIe microcomputer. The program permits the graphic output of the results of the three methods together. Four different secondary structure prediction methods [21,22,24,25] were applied to the four enzymes in Berlin. The program was run and coordinately printed on a Digital Dec 2020 computer (not shown). The results were essentially similar to those obtained in Saclay.

3. RESULTS

3.1. *Sequence homologies*

Dot-matrix plots of the nucleotide sequences of the genes of the four methyltransferases did not reveal any similarities (not shown). On the other hand, matrix plots which compare the four protein sequences clearly showed diagonals indicating regions of homology (fig.1). The plots between *Eco*RV methyltransferase and the three other enzymes show several homologies, while these are lower between the three GATC-recognizing enzymes. In particular, only limited homologies between *dam*T_4 methyltransferase and the three bacterial enzymes were found.

The dot-matrix program, however, does not allow for deletions or insertions. We have therefore attempted an alignment of sequences with the maximum fits (fig.2), including similarity between amino acids. The total alignment comprises 316 amino acid positions, which is close to the length of the longest sequence, i.e. *Eco*RV methyltransferase with 298 amino acids.

Four regions of the amino acid sequences appear to have considerable homologies and are indicated in fig.2:

(I) from alignment positions 19 to 85, containing several extensive homologies between two or three sequences;

(II) a highly conserved short region from positions 102 to 112;

(III) from positions 128 to 158, containing a second highly conserved block around position 150;

(IV) from positions 205 to 214, including the highly conserved sequence from 207 to 213, where all amino acids but one are identical.

The identities between any two sequences are summarized in matrix form in table 1: in the lower left half the total identities over the whole alignment sequence are shown and in the upper right half those in the four blocks of fig.2. In the total aligned sequence the highest score (*Eco*RV vs *dam*) is approx. 27% identity, but 50% within the four blocks, while it is only 20 and 34%, respectively, for the least close pair (*dam* vs *dam*T_4). These results confirm the conclusion from fig.1 that *Eco*RV methyltranferase is most closely related to the three other enzymes, followed by *E. coli dam* methylase.

The entire amino-terminal half (positions 1–158, fig.2) of the four proteins has a high degree of conservation. In 18 positions identical amino acids are found in all four sequences, 30 positions containing related amino acids in equivalent positions. Only three short regions spanning positions 1–18, 86–101 and 115–127 show gross

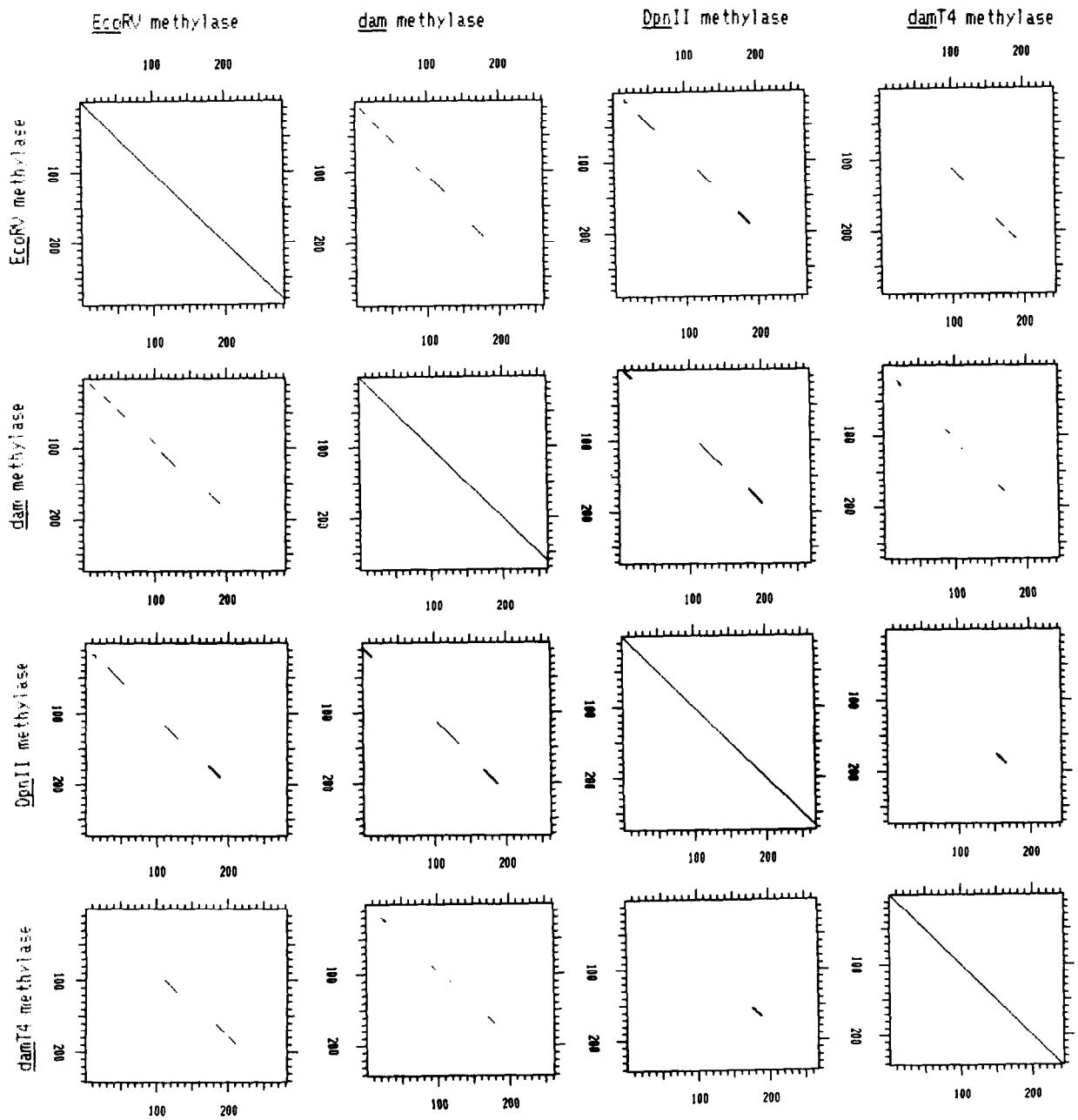

Fig.1. Dot-matrix comparison [20] of the various methyl transferases studied. A window of 20 amino acids was used with 8 or more residues correct.

dissimilarities and frequent large deletions. The previous analysis of Hattman et al. [18] was limited to positions 14–53 and 138–160.

In the carboxyl half of the aligned sequences (fig.2) only the highly conserved region IV between positions 205 and 214 shows any significant homology: six amino acids occupy identical positions in all four proteins: this corresponds to the homologous box 203–219 observed by Hattman et al. [18] when comparing the sequences of *dam*,

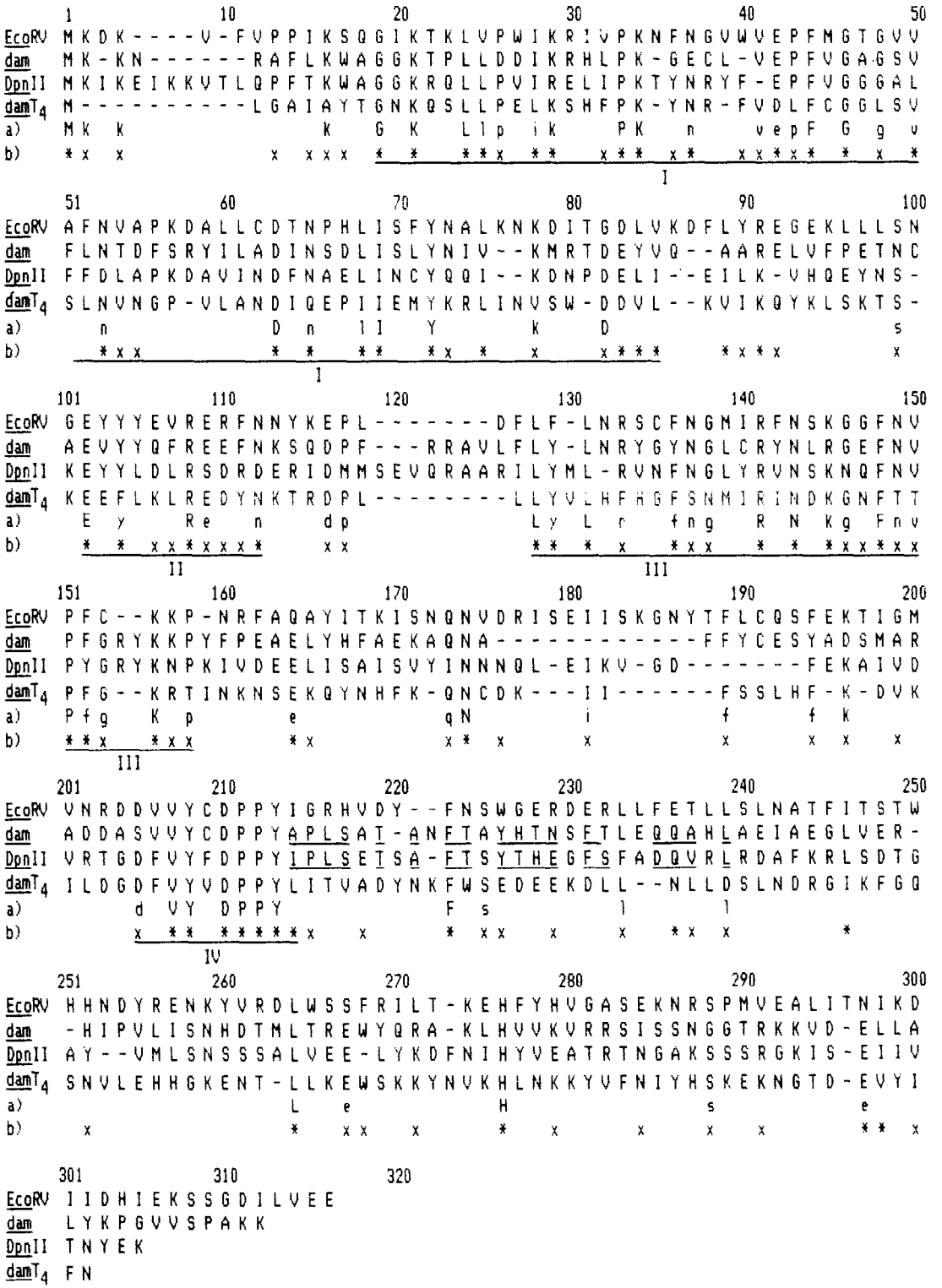
1 10 20 30 40 50
EcoRV MKDK----V-FVPPIKSQGIKTKLVPWIKRIVPKNFNGVWVEPFMGTGVV
dam MK-KN------RAFLKWAGGKTPLLDDIKRHLPK-GECL-VEPFVGAGSV
DpnII MKIKEIKKVTLQPFTKWAGGKRQLLPVIRELIPKTYNRYF-EPFVGGGAL
damT4 M----------LGAIAYTGNKQSLLPELKSHFPK-YNR-FVDLFCGGLSV
a) Mk k K G K L l p i k P K n v e p F G g v
b) *x x x x x x * * * * x ** x ** x * x x * x * x *
I
51 60 70 80 90 100
EcoRV AFNVAPKDALLCDTNPHLISFYNALKNKDITGDLVKDFLYREGEKLLLSN
dam FLNTDFSRYILADINSDLISLYNIV--KMRTDEYVQ--AARELVFPETNC
DpnII FFDLAPKDAVINDFNAELINCYQQI--KDNPDELI--EILK-VHQEYNS-
damT4 SLNVNGP-VLANDIQEPIIEMYKRLINVSW-DDVL--KVIKQYKLSKTS-
a) n D n l I Y k D s
b) *xx * * ** *x * x x**** *x*x x
I
101 110 120 130 140 150
EcoRV GEYYYEVRERFNNYKEPL--------DFLF-LNRSCFNGMIRFNSKGGFNV
dam AEVYYQFREEFNKSQDPF---RRAVLFLY-LNRYGYNGLCRYNLRGEFNV
DpnII KEYYLDLRSDRDERIDMMSEVQRAARILYML-RVNFNGLYRVNSKNQFNV
damT4 KEEFLKLREDYNKTRDPL---------LLYVLHFHGFSNMIRINDKGNFTT
a) E y Re n dp Ly L r fng R N kg Fnv
b) * * xx*xxx* xx ** * x *xx * * *xx*xx
II III
151 160 170 180 190 200
EcoRV PFC--KKP-NRFAQAYITKISNQNVDRISEIISKGNYTFLCQSFEKTIGM
dam PFGRYKKPYFPEAELYHFAEKAQNA----------FFYCESYADSMAR
DpnII PYGRYKNPKIVDEELISAISVYINNNQL-EIKV-GD-------FEKAIVD
damT4 PFG--KRTINKNSEKQYNHFK-QNCDK---II------FSSLHF-K-DVK
a) Pfg K p e qN i f f k
b) **x *xx *x x* x x x x
III
201 210 220 230 240 250
EcoRV VNRDDVVYCDPPYIGRHVDY--FNSWGERDERLLFETLLSLNATFITSTW
dam ADDASVVYCDPPYAPLSAT-ANFTAYHTNSFTLEQQAHLAEIAEGLVER-
DpnII VRTGDFVYFDPPYIPLSETSA-FTSYTHEGFSFADQVRLRDAFKRLSDTG
damT4 ILDGDFVYVDPPYLITVADYNKFWSEDEEKDLL--NLLDSLNDRGIKFGQ
a) d VY DPPY F s l l
b) x ** ***** x x * x x x ** x *
IV
251 260 270 280 290 300
EcoRV HHNDYRENKYVRDLWSSFRILT-KEHFYHVGASEKNRSPMVEALITNIKD
dam -HIPVLISNHDTMLTREWYQRA-KLHVVKVRRSISSNGGTRKKVD-ELLA
DpnII AY--VMLSNSSSALVEE-LYKDFNIHYVEATRTNGAKSSSRGKIS-EIIV
damT4 SNVLEHHGKENT-LLKEWSKKYNVKHLNKKYVFNIYHSKEKNGTD-EVYI
a) L e H s e
b) x * xx x x * x x x x x ** x
301 310 320
EcoRV IIDHIEKSSGDILVEE
dam LYKPGVVSPAKK
DpnII TNYEK
damT4 FN

Fig.2. Aligned sequences of bacterial adenine methylases. (a) First byline: identity in all four sequences (capitals) or in three sequences (lower-case letters). (b) Second byline: similar amino acids in all four sequences (*) or in three sequences (×). Similar amino acids: (R,K), (F, Y, W, H), (D, E, N, Q), (A, L, I, V), (C, M), (T, S), (P, G).

damT_4 and *Dpn*II. The alignment of the four proteins here clearly shows that this sequence box can be reduced to a stretch of nine amino acids.

Even outside the four homology blocks I–IV several similarities between two or three sequences appear. Although some of the alignments may appear tenuous, the presence of certain characteristic, but rare dipeptides in two or more sequences (like DP (EP) at 116–117, QN (NN) at 173–174, SW (WS) or HT (TH) around position 225 and HH around position 251–256) appears to warrant these choices. The absence of cysteines in analogous positions indicates the lack of disulfide bridges. The lack of PC sequences, observed in methyltransferases acting on cytosine and in thymidylate synthetase and implicated in the methyl transfer reaction [26], suggests that methyl transfer does not proceed by the same pathway, i.e. intermediate binding of C6 of the base to cysteine via the S-H group.

3.2. *Codon usage*

The absence of homologies in the nucleotide sequences can be rationalized from the codon usage of the four methyltransferases (not shown). While the *dam* gene has a base composition similar to *E. coli* DNA, i.e. ~50% G-C, the genes of the three other methyltransferases have extremely low G-C contents, ranging from 27.7 to 32.7% G-C. The result is the systematic use of G-C rich codons in *dam* methylase. Thus, the very rarely used Pro codon CCC, the rare Ala codons GCG and GCC, the Leu codon CUG, the Arg codons AGG and CGC, the Ser codon AGC or the Gly codon GGC are preferentially used in *dam* methylase, while they are frequently excluded by the other methyltransferases, and are often little used in other *E. coli* genes [26a]. This strong divergence of codon usage while large parts of the amino acid sequences are preserved indicates strong selective pressure to maintain the four homologous regions which must have an essential role in the structure and function of these enzymes.

Table 1

Identity matrix between four bacterial methyltransferases

	Number of identities			
	*Eco*RV	*dam*	*Dpn*II	*dam*T_4
*Eco*RV		60 (50)	55 (46)	44 (37)
dam	85 (27)		50 (42)	41 (34)
*Dpn*II	77 (25)	90 (29)		42 (35)
damT_4	73 (24)	60 (20)	63 (21)	

Upper right: identical amino acids in conserved regions (119 amino acids, underlined in fig.2). Lower left: identical amino acids in total alignment sequence (fig.2). Percentages in parentheses

3.3. *Secondary structure homologies*

The secondary structure of the four methyltransferases was investigated using four methods [21,22,24,25]. These methods generally predict the content in α-helix, β-sheet, turn and coil with about 65–75% accuracy. The hydropathy method of Kyte and Doolittle [23] attempts to distinguish between regions of amino acid residues which are buried inside the protein or are on the surface.

In the present case the agreement between the GOR [21] and Chou-Fasman methods [22] is not entirely satisfactory (fig.2), nor with the two other methods [23,24]. It is, however, noteworthy that the highly conserved region IV appears to be part of two β-sheets, connected by a turn, followed by a helical region. This may be significant (see below). Also, this region of intermediate hydrophobicity is not buried in the protein, nor is it completely on the surface. Region III, another area of high homology, also appears to consist of β-sheets.

The hydrophobicity data (fig.2) suggest a recurring pattern. Region I appears to be a β-fold entering into the interior of the protein, while region II

is probably on the surface, as are most of the regions with little or no amino acid homology. In all four methyltransferases the N-terminal and C-terminal appear to be on the surface.

4. DISCUSSION

The remarkable similarities in sequence and structural parameters between the four deoxyadenosine methyltransferases cannot be accidental. Several properties have to be fulfilled by all enzymes: the recognition of the specific sequence GATC or GATATC, respectively, the possible unspecific binding of parts of the enzyme up- or downstream from the recognition site, the binding of the methyl donor, Ado-Met and methyl group transfer.

Unfortunately, no crystal structure exists of an Ado-Met-binding enzyme. Since Ado-Met is not a nucleotide, the numerous investigations on nucleotide-binding proteins [34–36] are not of much help. As expected, the signature sequence GXXXXGK [35,36] characteristic for the phosphate-binding sites is not present in any of the sequences studied. Since, contrary to the nucleotides, Ado-Met is positively charged, only the base and/or sugar part may be relevant. The suggested base-binding sites [36,37] are located between two loops in *ras* P21 protein and elongation factor EF-Tu. Their characteristic sequences, FLNKXD and SAXKXXG respectively, are, however, also absent in the methyltransferases (fig.2).

In this context the crystal structure of RNase T_1 in its complex with 2′-GMP [38] is particularly interesting. In this complex the guanine base of 2′-GMP is sandwiched between the two tyrosine residues of the sequence $HK^{42}YNN^{45}YE$. The two asparagine residues [39] form a very sharp turn in order to permit the two tyrosines to take the guanine base in their middle. Several hydrogen bonds between asparagine and guanine stabilize the complex.

We suggest that the highly conserved sequence DXVYXDPPY (positions 205–213, region IV in fig.2, which is very probably inside the protein in a pocket of intermediate hydrophobicity; fig.3) could serve a similar purpose: the binding of Ado-Met. The two proline residues will certainly cause a sharp turn in the polypeptide chain so that the two tyrosine residues could face each other. Thus, the adenine ring could sandwich between and stack with one or both of the two tyrosines, with possible hydrogen bonds to the aspartate in position 205 (or 203) holding the positive charge of the methyl-S^+ of adenosylmethionine in place. The R (or K) residue upstream (around position 200) could interact with the carboxyl group of the methionine moiety of Ado-Met. A second possible candidate could be region III (position 141–153, fig.2) with the highly conserved sequence RXNXK(R)XFNVPF(Y)G. Sequences similar to those on regions III or IV are found in other procaryotic methyltransferases (table 2), like those of *Eco*RI [16,17], *Pae*R7 [27], *Bsp*RI [28], *Bsu*RI [29], *Hha*I [29a], *Pst*I [30] and *Eco*RII [31], as well as from *Bacillus* phage SPR [32], and the ribosomal methyltransferases *Ksq*A, *erm*D, *erm*C and pAM77 [33]. The existence of these sequences in other methyltransferases with other methylation sites suggests that these sequences are involved in the binding of Ado-Met or in the methylation reaction and are not the binding sites of the enzymes to DNA.

As far as the DNA-binding region is concerned several considerations have to be taken into account. Results on the mode of action of *dam* methylases, which was described as a monomer [11], and the related *Dpn*II methyltransferase which has recently been reported to act as a dimer [40] are contradictory. If the methyltransferases act as monomers, the rules for DNA binding observed for the most studied DNA-binding proteins, i.e. dimeric repressors and *cro* proteins [41–45], will certainly not hold. In these proteins two identical subunits bind symmetrically one helix turn apart. As expected, attempts to align the characteristic helix-turn-helix motif [45] with sequences in region I, II or III failed. Similarly, the helix-turn-β-sheet motif in *Eco*RI endonuclease [46] which binds to the large groove of the GAATTC target sequence in a dyadic manner is not present in the four methyltransferases considered.

A discussion of a DNA-binding mechanism should include biochemical parameters of which only very few are known for the methyltransferases. No mutants of the four enzymes have been described and nothing is yet known on the mechanism of transfer of the methyl group. The

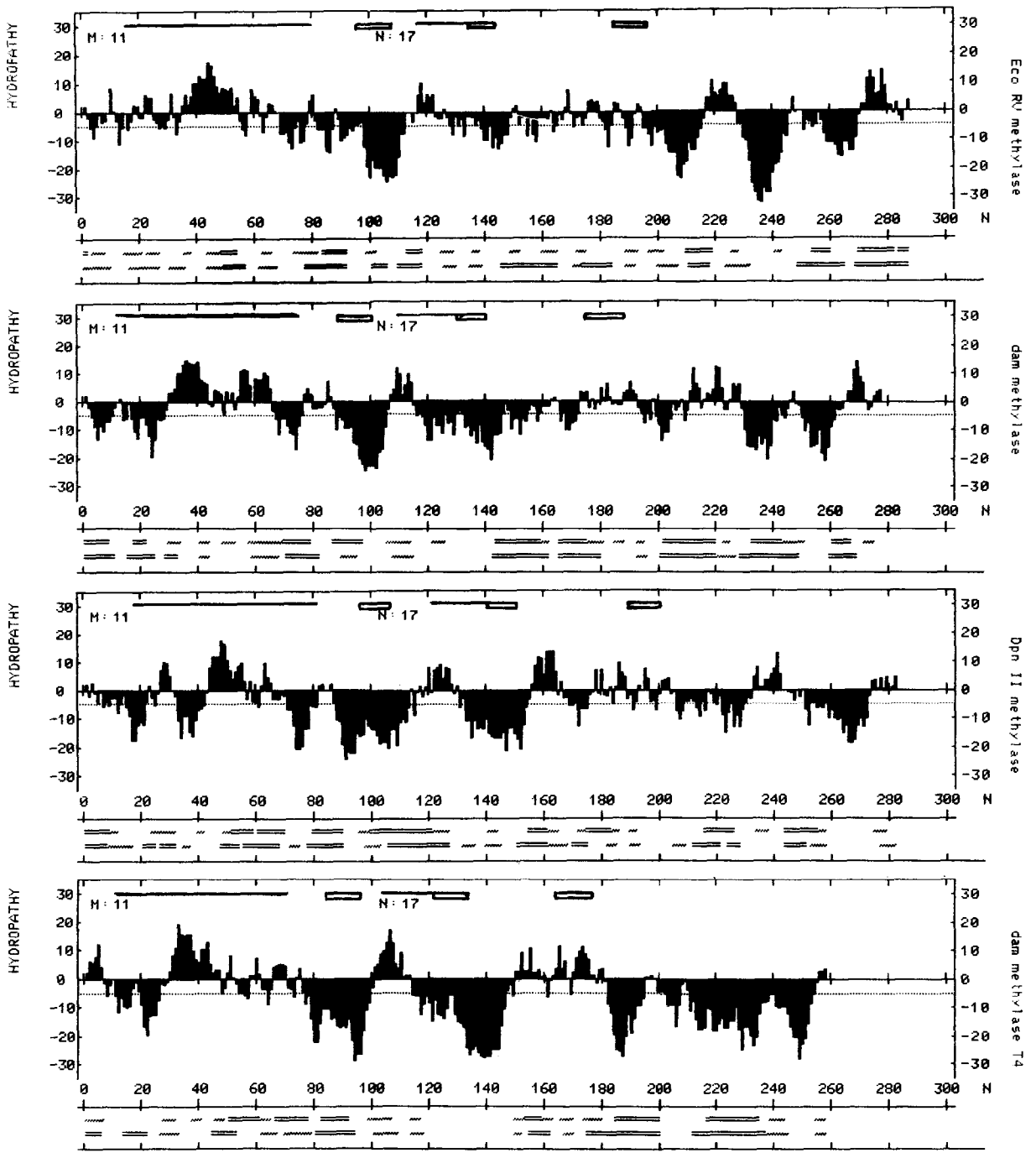

Fig.3. Hydropathy plot [23] of the four bacterial methyltransferases studied using a window of 11 residues. Also plotted are (below) the results of the predictive methods according to Garnier et al. [21] (top lane) and Chou and Fasman [22] (bottom lane). (═══) α-Helix, (~~~~) β-sheet. Sequence homologies are indicated in the top part: (——) medium homology, (═══) high homology.

Table 2

Similarities in amino acid sequences between regions IV and III of *Eco*RV, *dam*, *Dpn*II and *dam*T and other methyltransferases (positions are real sequence positions)

Region		Methylation site	Ref.
Region IV			
*Eco*RV	^{187}D- D- - V V Y C D P P Y I G R H	G^{m}ATATC	15
dam	^{173}D- D A S V V Y C D P P Y A P L S	G^{m}ATC	12
*Dpn*II	^{187}T G D- - F V Y F D P P Y I P L S	G^{m}ATC	14
*dam*T$_4$	^{164}D G D- - F V Y V D P P Y L I T V	G^{m}ATC	13
*Eco*RI	^{133}K S D- - I V V T N P P F S L F R	GA^{m}ATTC	16,17
*Pae*R7	^{113}Q F D- - F V V G N P P Y V R P E L	CTCG^{m}AG	27
*Pst*I	^{145}K Y N- - K A I L N P P Y L K I A	CTGC^{m}AG	30
*Bsp*RI	^{275}I G D- - - L S L D P^{G} P Y F T	GG^{m}CC	28
*Bsu*RI	^{277}I G D- - - L V T D P^{G} P Y F T	GG^{m}CC	29
*Hha*I	^{174}D L N I Q N- F Q F P^{K} P F E L N T F	G^{m}CGC	29a
III			
*Eco*RV	^{128}R F N S K G G F- N V P F C- - K K P	G^{m}ATATC	15
dam	^{124}R Y N L R G E F- N V P F G R Y K K P	G^{m}ATC	12
*Dpn*II	^{134}R V N S K N Q F- N V P Y G R Y K N P	G^{m}ATC	14
*dam*T$_4$	^{116}R I N D K G N F- T T P F G- - K K P	G^{m}ATC	13
*Pae*R7	^{334}R G Q- - G V I - N- P F A E S G G	CTCG^{m}AG	27
*Pst*I	^{186}V A I T P R S F C N G P Y F N E F K K	CTCG^{m}AG	30
*Eco*RII	^{178}D H D V L L A^{G} F P C Q P F S L A G	C^{m}CAGG	31
SPRM	^{101}Y V E T L- K E K Q P K F F V F	G^{m}CNGC	32
	^{345}N G R R F K D D G E P A F T V N		32
*kgs*A	^{110}V F G N L P Y N I S T		33
*erm*D	V V S N I P Y A I T T		33
*erm*A	I Y G N I P Y N I S T		33
pAM177	L V G S T P Y H L S T		33

large degree of homology in the amino-terminal half of the aligned sequences (fig.2) and the recurrent patterns of hydrophobicities (fig.3) lead, however, to the logical assumption that one of these homologous regions should be involved in unspecific DNA binding. On the other hand, one expects that amino acid sequences conserved in *dam* and *Dpn*II methyltransferases (both of which recognize GATC), but not in *Eco*RV methyltransferase (which recognizes GATATC) could be those involved in specific target recognition. Only one region meets these criteria: in the region 215–239 the *dam* and *Dpn*II enzymes have 10 identical and three similar amino acids in common, plus the characteristic inversion HT vs TH (underlined residues in fig.2), but only three in common with *Eco*RV. This region is predicted to be α-helical in all four methyltransferases by all prediction methods applied and very probably on the surface of the protein (fig.3). This further supports speculation on the role of this region as the recognition domain for the DNA target sequences. Its immediate vicinity with the suggested Ado-Met-binding region IV (see above) makes this an attractive possibility of approaching and close interaction of the two substrates.

The methyltransferase *dam*T_4 has been omitted from the above discussion. It is difficult to evaluate the effect of 5-hydroxymethylcytosine residues within the GATC sequence. Certain similarities exist between *Eco*RV and *dam*T_4 methyltransferases, and to a much lesser extent with *Dpn*II and *dam* proteins. In this context it is interesting that *dam*T_4 methylase can replace that from *E. coli* in the methylation of GATC, but is unable to substitute for the *E. coli* enzyme in the methylation-instructed mismatch repair [47].

4.1. *Phylogenetic relationships*

The close relationship of the four enzymes considered can be further accentuated, if the analysis is based on the four highly conserved regions of the amino acid sequences (fig.2). This may be justified by the assumption that these protein sequences are involved in common general functions, like specific and/or unspecific DNA binding, Ado-Met binding or methyl group transfer. The degree of homology of amino acids which amounts to nearly 40% of the total alignment (table 1) varies between 43 and 52% when the three bacterial enzymes are compared, while *dam*T_4 methyltransferase shows only some 35% conserved amino acids in comparison (table 1). This may reflect strong selective pressure on the phage genome, since this protein is also the smallest of the four (deletions in the middle and at both termini).

It has been suggested by Mannarelli et al. [14] that a common ancestral restriction-modification system existed in *E. coli*, of which only *dam* methyltransferase subsisted. The inclusion of the methyltransferases *Eco*RV and *dam*T_4 in this group raises some interesting phylogenetic considerations. Since the *Dpn*II and *Eco*RV enzymes are part of the respective restriction-modification systems the question arises as to whether the restriction enzymes are also related. If they were, one could envisage the existence of a progenic restriction-modification complex. Interestingly, no significant degree of homology can be detected between the sequences of the restriction endonucleases *Eco*RV and *Dpn*II or *Dpn*I [48]. Since only the methyltransferases are related, this strongly indicates that no progenic restriction-modification unit had existed and that both the endonucleases and the methyltransferases have evolved independently. The *dam* and *dam*T_4 gene seem to be adapted forms involved in different biological functions.

ACKNOWLEDGEMENTS

We thank Drs T.A. Trautner and T. Balganesh (Berlin) for helpful discussions. Dr C. Marck (Saclay) is thanked for his help with the DPSA.A programme. A.K. gratefully acknowledges the financial support from Euratom, l'Association pour la Recherche sur le Cancer and the European Community.

REFERENCES

[1] Modrich, P. (1979) CRC Crit. Rev. Biochem. 13, 287–323.
[2] Radman, M. and Wagner, R. (1986) Annu. Rev. Genet. 20, 523–538.
[3] Plasterk, R.H.A., Vrieling and Van de Putte, P. (1983) Nature 301, 344–347.
[4] Brown, R.E. and Wright, A. (1986) Mol. Gen. Genet. 202, 246–250.
[5] Smith, D.W., Garland, A.M., Herman, G., Enns, R.E., Baker, T.A. and Zyskind, J.W. (1985) EMBO J. 4, 1319–1326.
[6] Messer, W., Bellekes, U. and Lother, H. (1985) EMBO J. 4, 1327–1332.
[7] Längle-Rouault, F., Maehaut-Michel, G. and Radman, M. (1986) EMBO J. 5, 2009–2013.
[8] Fazakerley, G.V., Téoule, R., Guy, A., Fritzsche, H. and Guschlbauer, W. (1985) Biochemistry 24, 4540–4548.
[9] Quignard, E., Téoule, R., Guy, A. and Fazakerley, G.V. (1985) Nucleic Acids Res. 13, 7829–7836.
[10] Quignard, E., Fazakerley, G.V., Téoule, R., Guy, A. and Guschlbauer, W. (1985) Eur. J. Biochem. 152, 99–105; (1987) Eur. J. Biochem., in press.
[11] Geier, G.E. and Modrich, P. (1979) J. Biol. Chem. 254, 1408–1413.
[12] Brooks, J.E., Blumenthal, R.M. and Gingeras, T.R. (1983) Nucleic Acids Res. 11, 837–851.
[13] MacDonald, P.M. and Mosig, G. (1984) EMBO J. 3, 2863–2871.
[14] Mannarelli, B.M., Balganesh, T.S., Greenberg, B., Springhorn, S.S. and Lacks, S.A. (1985) Proc. Natl. Acad. Sci. USA 82, 4468–4472.
[15] Bourgueleret, L., Schwarzstein, M., Tsugita, A. and Zabeau, M. (1984) Nucleic Acids Res. 12, 3659–3676.
[16] Newman, A.K., Rubin, R.A., Kim, S.-H. and Modrich, P. (1981) J. Biol. Chem. 256, 2131–2139.

[17] Greene, P.J., Gupta, M., Boyer, H.W., Brown, W.E. and Rosenberg, J.M. (1981) J. Biol. Chem. 256, 2143–2153.
[18] Hattman, S., Wilkinson, J., Swinton, D., Schlagman, S., MacDonald, P.M. and Mosig, G. (1985) J. Bact. 164, 932–937.
[19] Cantoni, G.L. (1985) Prog. Clin. Biol. Res. 198, 47–65.
[20] Marck, C. (1986) Nucleic Acids Res. 14, 583–590.
[21] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[22] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.
[23] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[24] Burgess, A.W., Ponnuswamy, P.K. and Scheraga, H.A. (1974) Isr. J. Chem. 12, 239–286.
[25] Nagano, K. (1977) J. Mol. Biol. 109, 252–274.
[26] Wu, J.C. and Santi, D.V. (1987) J. Biol. Chem. 262, 4778–4786.
[26a] Lloubes, R.P., Chartier, M.J., Journet, A.M., Varenne, S.G. and Lazdunski, C.J. (1984) Eur. J. Biochem. 144, 73–78.
[27] Thériault, G., Roy, P.H., Howard, K.A., Benner, J.S., Brooks, J.E., Waters, A.F. and Gingeras, T.R. (1985) Nucleic Acids Res. 13, 8440–8461.
[28] Posfai, G., Kiss, A., Erdei, S., Posfai, J. and Venetianer, P. (1983) J. Mol. Biol. 170, 597–610.
[29] Posfai, G., Baldauf, F., Erdei, S., Posfai, J., Venetianer, P. and Kiss, A. (1984) Nucleic Acids Res. 12, 9039–9049.
[29a] Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G.G. and Wells, R.D. (1987) J. Biol. Chem. 262, 4770–4777.
[30] Walder, R.Y., Walder, J.A. and Donelson, J.E. (1984) J. Biol. Chem. 259, 8015–8026.
[31] Som, S., Bhagwat, A.S. and Friedman, S. (1987) Nucleic Acids Res. 15, 313–331.
[32] Buhk, I., Behrens, B., Tailor, R., Wilke, K., Prada, J.J., Günthert, U., Noyer-Weidner, M., Jentsch, S. and Trautner, T.A. (1984) Gene 29, 51–61.
[33] Van Buul, C.P.J.J. and Van Knippenberg, P.H. (1985) Gene 38, 65–72.
[34] Argos, P. and Leberman, R. (1985) Eur. J. Biochem. 152, 651–656.
[35] Möller, W. and Amons, R. (1985) FEBS Lett. 186, 1–7.
[36] Masters, S.B., Stroud, R.M. and Bourne, H.R. (1986) Protein Eng. 1, 47–54.
[37] McCormick, F., Clark, B.F.C., La Cour, T.F.M., Kjelgaard, M., Norskov-Lauritsen, L. and Nyborg, J. (1985) Science 230, 78–82.
[38] Heinemann, U. and Saenger, W. (1982) Nature 299, 27–31.
[39] Heinemann, U. and Saenger, W. (1983) J. Biomol. Struct. Dyn. 1, 523–538.
[40] Lacks, S.A., Mannarelli, B.M., Springhorn, S.S., Greenberg, B. and De la Campa, A.G. (1986) Abstr., Wind River Conference on Genetic Exchange, Estes Park, CO, June 9–12.
[41] Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M. and Pabo, C.O. (1982) Nature 298, 447–451.
[42] Weber, I.T., McKay, D.B. and Steitz, T.A. (1982) Nucleic Acids Res. 10, 5085–5102.
[43] Matthews, B.W., Ohlendorf, D.H., Anderson, W.F. and Takeda, Y. (1982) Proc. Natl. Acad. Sci. USA 79, 1428–1432.
[44] Ohlendorf, D.H. and Matthews, B.W. (1983) Annu. Rev. Biophys. Bioeng. 12, 259–284.
[45] Pabo, C.O. and Sauer, R.T. (1984) Annu. Rev. Biochem. 53, 293–321.
[46] McClarin, J.A., Frederick, C.A., Wang, B.-C., Greene, P.J., Boyer, H.W., Grable, J. and Rosenberg, J.M. (1986) Science 234, 1526–1541.
[47] Schlagman, S.L., Hattman, S. and Marinus, M.G. (1986) J. Bacteriol. 165, 896–900.
[48] Lacks, S.A., Mannarelli, B.M., Springhorn, S.S. and Greenberg, B. (1986) Cell 46, 993–1000.