

Cloning of the *Haemophilus influenzae* Dam methyltransferase and analysis of its relationship to the Dam methyltransferase encoded by the HP1 phage^{*}

Janusz M. Bujnicki¹ ✉, Monika Radlinska², Piotr Zaleski² and Andrzej Piekarowicz² ✉

¹Bioinformatics Laboratory, International Institute of Molecular and Cell Biology, ks. Trojdena 4, 02-109 Warszawa, Poland; ²Institute of Microbiology, Warsaw University, Miecznikowa 1, 02-093 Warszawa, Poland

Received: 24 September, 2001; accepted: 3 December, 2001

Key words: DNA methyltransferase, protein structure, sequence specificity, molecular evolution, bioinformatics

In this paper we report cloning and experimental characterization of the DNA adenine methyltransferase (*dam*) gene from *Haemophilus influenzae* and comparison of its product with the Dam protein from the lysogenic phage of *H. influenzae*, HP1. Molecular modeling of M.*Hin*Dam and M.HP1Dam was carried out, providing a framework for a comparative analysis of these enzymes and their close homologs in the structural context. Both proteins share the common fold and essential cofactor-binding and catalytic residues despite overall divergence. However, subtle but significant differences in the cofactor-binding pocket have been identified. Moreover, while M.*Hin*Dam seems to contact its target DNA sequence using a number of loops, most of them are missing from M.HP1Dam. Analysis of both MTases suggests that their catalytic activity was derived from a common ancestor, but similar sequence specificities arose by convergence.

Adenine-*N*⁶ (m⁶A) MTases are enzymes that transfer a methyl group from *S*-adenosyl-L-methionine (AdoMet) to the N⁶ position of adenine in a specific DNA sequence. Amin- acid

sequence comparisons among the members of this group revealed the presence of nine relatively weakly conserved motifs and a region of essentially higher variability [1, 2]. Based on

^{*}This work was supported by the State Committee for Scientific Research (KBN, Poland) grant 6P04 B00519 and by Warsaw University, grant 1485/12.

✉ Authors, to whom correspondence should be addressed: tel: (48 22) 668 5384; fax: (48 22) 668 5288; e-mail iamb@bioinfo.pl; e-mail: anpiek@plearn.edu.pl

Abbreviations: RM, restriction-modification; MTase, methyltransferase; ENase, endonuclease; TRD, target recognition domain; m⁶A, *N*⁶-methyladenine; m⁴C, *N*⁴-methylcytosine; m⁵C, C⁵-methylcytosine; AdoMet, *S*-adenosyl-L-methionine.

the results of X-ray crystallography of m⁶A MTase *M.TaqI*, motifs IV–VIII were assigned to the active-site subdomain, motifs X and I–III to the AdoMet-binding subdomain, and the variable region was recognized as a separate target-recognizing domain (TRD) [3]. According to the possible linear arrangements of these three functional regions, m⁶A MTases were subdivided into 6 classes: α , β , γ , δ , ε and ζ [1]. To date, nearly all characterized enzymes fall into the α , β and γ classes, with only a few exceptions (reviewed in ref. [4]). Structural characterization of the *M.DpnM* MTase (α) and its comparison with *M.TaqI* (γ) demonstrated that despite absolute conservation of the catalytic domain, the TRDs of m⁶A MTases from different classes are unrelated and do not overlap spatially, when atomic coordinates are superimposed [5].

M.EcoDam, one of three DNA MTases of *E. coli* K-12, is – like *M.DpnM* – a member of the α class, which methylates adenine in both strands of the GATC palindrome. However, *M.EcoDam* is not a part of a regular restriction-modification system. Instead, it regulates numerous molecular events [6]. Immediately after replication GATC sites are transiently hemimethylated, allowing corrections of mismatches by unmethylated strand-directed mismatch repair pathway [7]. Dam methylation of *oriC* is required for efficient initiation of chromosomal replication [8]. The sequestration of *oriC* and *dnaA* ensures single initiation of chromosome replication per cell division [9]. Moreover, methylation of dam sites at numerous promoters regulates their expression [10]. Dam methylation is not essential for viability, although Dam⁻ strains show a variety of phenotypic effects including increased spontaneous mutations, moderate SOS induction, enhancement of duplication segregation, inviability of *dam recA* and *dam recB* mutants, and suppression of that inviability by mutations that eliminate mismatch repair [11]. DNA methylation plays an important role during development of *E. coli* virus T1 [12], which encodes a close homolog

of the HP1 Dam MTase (*M.T1Dam*). The order of subdomains in *M.HP1Dam* and *M.T1Dam* is similar to that of the γ -class of m⁶A MTases, in contrast to the Eubacterial and T-even phage Dam MTases (α -class). *In vivo*, all known coliphage MTases are capable of methylating the adenines in all GATC sequences in the viral DNA. However, it is not clear why these and possibly other viruses afford the luxury of their own MTases, whose specificity overlaps with that of the host enzyme.

It is widely recognized that differences in virulence may be related to the presence or absence of particular genes. Recently it has been shown that *Salmonella typhimurium* mutants lacking Dam activity were totally avirulent and effective as live vaccines against murine typhoid fever [13]. It is not surprising, as the expression of at least 20 *S. typhimurium* genes known to be induced during infection is regulated by dam methylation – hence, it was suggested that Dam inhibitors could have broad antimicrobial action, and Dam-derivatives of many pathogens might serve as live attenuated vaccines [13].

To investigate the function of DNA adenine methylation in *H. influenzae*, an important human respiratory tract pathogen, we attempted to clone and analyze the pair of *dam* genes from the *H. influenzae* Rd chromosome [14] and from its lysogenic bacteriophage HP1 [15]. Products of these genes are homologous to *M.EcoDam* and *M.T1Dam* MTases, respectively. The *H. influenzae* Rd strain is non-pathogenic, however these two types of Dam MTases occur commonly in enteropathogenic bacteria and their phages, respectively. As such, the study of DNA methylation in *H. influenzae* may improve our understanding of pathogenic mechanisms in virulent strains of this bacterium, evolution of enterobacterial phages and their genes, as well as evolution of DNA MTases in general. Cloning and functional characterization of *M.HP1Dam* has been published as a separate paper [16]. Here, we report comparative bioinformatic analysis

of both Dam MTases and experimental characterization of M.HinDam, which was undertaken to confirm the theoretically predicted function of the bacterial gene. The homology models described in this paper are intended to guide future mutagenesis experiments and help in rational design of potential Dam inhibitors.

MATERIALS AND METHODS

Bacterial strains, plasmids, growth conditions and reagents, and DNA techniques. All general techniques were used according to protocols described for both host organisms – *E. coli* [17] and *H. influenzae* [18]. The *H. influenzae* Rd30 was used for cloning the *dam* gene. All standard reactions were done according to Sambrook *et al.* [17] or to recommendation of the supplier. The recombinant plasmid containing *dam* gene of *H. influenzae* was constructed as follows: The *dam* coding sequence was amplified by PCR using primers: Mleft: 5'-GCGCGCCATGGTGTTACGTCCGAAAAACAATC-3' and Mright: 5'-GGCCCCGGGTTTACGAGCACC-AAATATCGCAA-3'. Both the PCR product and the vector DNA pMPMT4 Ω [19] were digested with *Nde*I and *Sma*I and then ligated using T4 DNA ligase. The ligation mixture was transformed into *E. coli* GM2163 (*dam*⁻ *dcm*⁻) and the resulting transformants containing the desired insert were tested for the presence of MTase activity. One plasmid, pHindDam was chosen for further study. To obtain inducible synthesis of *H. influenzae* Dam MTase the *E. coli* GM2163(pHindDam) cells were grown in LB broth supplemented with tetracycline (15 μ g/ml) to mid-log phase. Then, to induce expression of the *dam* gene, L-arabinose was added to a final concentration of 1%, and the cultures were incubated for additional 18 h at 37°C with vigorous shaking. The cells extract was obtained as follows: 50 ml of induced cell culture was suspended in 50 mM Tris/HCl, pH 7.5, 1 mM EDTA, 7 mM

2-mercaptoethanol (ME) (buffer A). Cells were disrupted by sonication and cellular debris was removed by centrifugation at 100000 $\times g$ for 1 h. The MTase activity was detected by measurement of transfer of CH₃ groups from labeled AdoMet to different types of substrate DNA. Reaction mixtures (20 μ l) contained 50 mM Tris/HCl, pH 7.5, 10 mM EDTA, 5 mM ME, 2 μ Ci of [³H-methyl]AdoMet, different amounts of DNA and 1 μ l of cell extract. After incubation at 37°C for 30 min, the extent of methylation was assayed as described by Renbaum & Razin [20]

Sequence retrieval, alignment and structure prediction. Dam-related sequences were identified using a variety of BLAST and PSI-BLAST searches [21] of a non-redundant (nr) database and publicly available nucleotide sequences from both complete and unfinished genome projects at NCBI (<http://www.ncbi.nlm.nih.gov>) and IIMCB (<http://blast.bioinfo.pl>) using sequences of M.HinDam and M.HP1Dam as queries. Multiple alignments were extracted from the BLAST output using the BIBVIEW software (<http://bioinfo.pl/bibview.pl>) and corrected manually, taking into account preservation of continuity of predicted secondary structural elements. All structure predictions were carried out *via* the MetaServer/Pcons interface (<http://bioinfo.pl/meta/>; ref. [22]; links to the individual structure prediction servers are provided therein).

Homology model building, refinement and evaluation. The three-dimensional modeling was carried out using the program MODELLER [23], which reconstructs protein structures by satisfaction of spatial restraints. For homology modeling of M.HinDam and M.HP1Dam these restraints were derived from aligned query sequences and template structures of the DNA:m⁶A MTases M.DpnM and M.TaqI obtained from Protein Data Bank (entries 2dpm and 1g38, respectively). The modeling procedure allowed to retain the conformation of backbone and orientations of the sidechains as much as possible in respect to

the type of modeled amino-acid residue, local environment and similarity to either of the template structures, i.e. the choice of the modeling template (*M.DpnM* or *M.TaqI*) was not forced explicitly. Geometrical strain and steric clashes generated by overlapping van der Waals radii were relaxed by limited energy minimization protocol optimizing the objective target function employing conjugate gradients and molecular dynamics with simulated annealing [23]. During the process of modeling the emerging target structures were evaluated by internal self-consistency check, whether the intermediate model satisfied most restraints used to calculate it. Assessment of reliability of *M.HinDam* and *M.HP1Dam* models was carried out using WHATCHECK [24]. The compatibility of residues with the environments provided by neighboring atoms or exposed to the solvent was assessed using VERIFY3D [25].

RESULTS AND DISCUSSION

Cloning of the *H. influenzae* Dam MTase gene

The open reading frame of the MTase gene of *H. influenzae* was amplified by PCR using primers Mleft and Mright and cloned into the vector pMPMT4 Ω in the orientation that allows for a high-level expression and tight regulation based on the promoter and AraC repressor/activator of the *ara* operon resulting in plasmid pHindDam.

Expression and functional activity of *H. influenzae* Dam was tested in *E. coli* strain GM2163, carrying mutations in the *dam* as well as in *dcm* genes. The cell extracts prepared from *E. coli* GM2163(pHindDam) cells induced by the presence of the arabinose showed the presence of MTase activity that was much higher if the substrate DNA (λ DNA) was obtained from *E. coli dam* cells (not shown). The specificity of this MTase was tested by comparative digestion of plasmid

DNA by isoschizomeric restriction enzymes *Sau3AI* and *MboI*. It is known that the cleavage by *Sau3AI* is not affected by adenine methylation in the GATC sequence, whereas *MboI* digestion is prevented by this methylation [26]. Restriction analyses using plasmid DNA isolated from strains GM2163(pMPMT4 Ω) and GM2163(pHind-Dam) confirmed a MTase activity in the latter strain. As shown in Fig. 1, plasmid DNA containing *H. influenzae* MTase gene was susceptible to *Sau3AI* but not to *MboI*, indicating the presence of adenine methylation within the GATC sequence. This results is in perfect agreement with sequence analysis and structure prediction, suggesting that *M.HinDam* shares DNA recognition determinants with other GATC-targeting MTases.

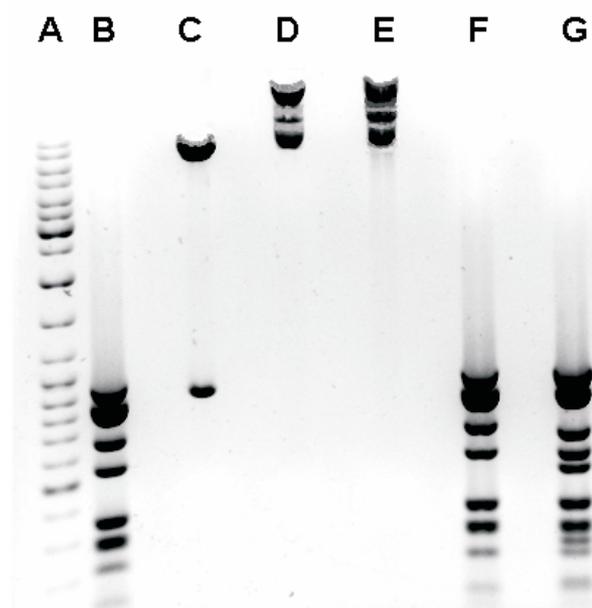


Figure 1. Dam methylation of pHindDam DNA isolated from *E. coli* GM2163 cells induced by the presence of L-Ara.

Aliquotes of 0.4 μ g DNA were digested in 20 μ l reaction volumes with 10u (25-fold excess) of enzymes in buffers recommended by the manufacturers for 8 h at 37°C. Lane A, DNA Ladder Mix (GeneRuler™ - Fermentas); lane B, pMPMT4 Ω digested with *MboI*; lane C, pHindDam digested with *NdeI* and *SmaI*; lane D, pHindDam isolated from induced cells digested with *MboI*; lane E, pHindDam isolated from non-induced cells digested with *MboI*; lane F, pMPMT4 Ω digested with *Sau3AI*; lane G, pHindDam digested with *Sau3AI*.

M.VT2Dam [27] MTases and several putative products of open reading frames identified by genomic sequencing (Fig. 3). Since no similar-

to the DNA:m⁶A MTase *M.TaqI* and γ -class MTases *ErmC'* and *ErmAM* that generate m⁶A in RNA (not shown). However, optimal

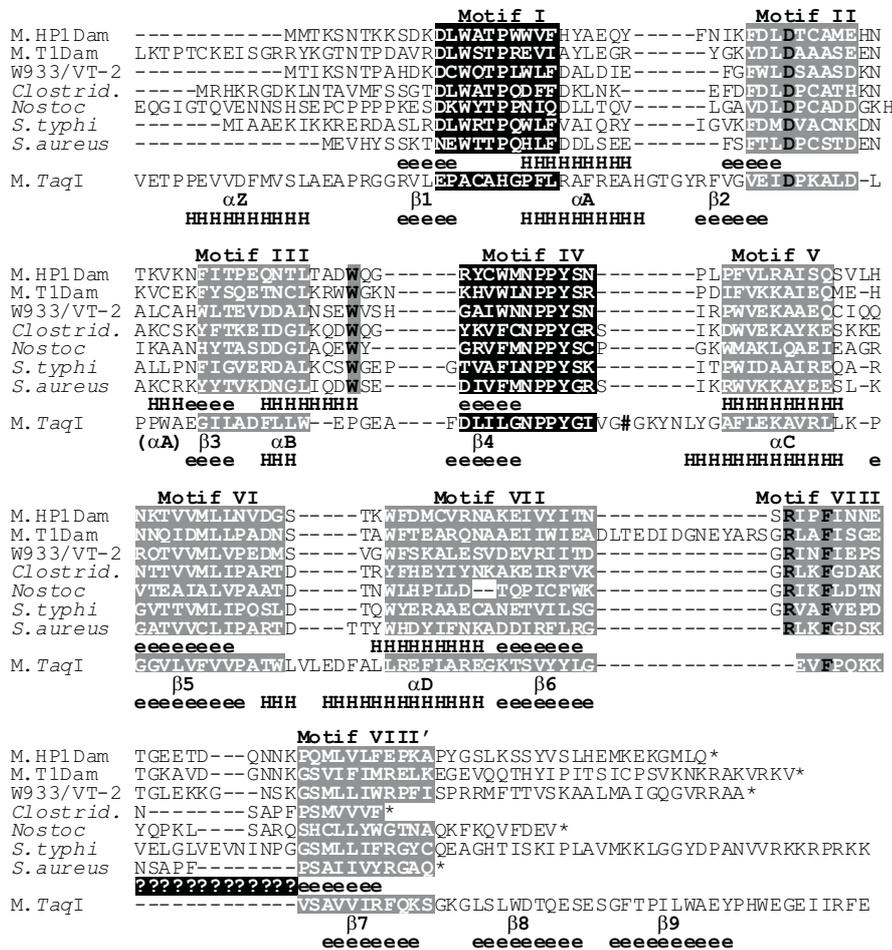


Figure 3. Sequence alignment of M.HP1Dam and its homologs with a structurally characterized γ -m⁶A MTase *M.TaqI*.

Conserved regions are boxed and numbered according to Malone *et al.* [1]. Motifs I and IV are shown in black, the remaining motifs and the residues invariant for the M.HP1Dam subfamily are colored in gray. The secondary structural elements of M.HP1Dam and *M.TaqI* (predicted and experimentally determined, respectively) are shown, with e (extended) indicating β -strands and H (helical) indicating α - or 3_{10} -helices. Question marks indicate the insertion unique for the M.HP1Dam and its homologs, which might participate in target recognition. Stars indicate protein C-termini. For the clarity of presentation the N-terminal region (MKDFNDIETIDFAETGCSFTREAIASGGYYQA) has been omitted from the M.T1Dam sequence, as well as the insertion indicated by “#” (VFKAVKDLYKKA-FSTWK) and the TRD of *M.TaqI* (aa 248–421) as these regions have no counterpart in sequences of other proteins in the presented alignment.

ity of the M.HP1Dam family to known MTases was apparent from PSI-BLAST searches, we used the sequences of all its members to query the fold recognition MetaServer in order to identify proteins with similar structure. Among MTases with known structure, M.HP1Dam homologs showed similarity only

alignment of the phage MTases with the potential structural templates was problematic, given low level of mutual sequence similarity. It has been demonstrated that incorporating sequence-derived structure predictions into comparative sequence analysis radically improves results of protein fold assignments

[28]. Fortunately, the pattern of secondary structural elements predicted for the M.HP1Dam family was very similar to that experimentally determined for *M.TaqI*. Therefore, the alignment shown in Fig. 3 was based on combined sequence and structure threading followed by manual editing, rather than sequence comparisons alone. The pattern of predicted secondary structures and conserved residues known to be involved in cofactor binding and catalysis in the majority of MTases was used as a guide in cases where threading programs provided several alternatives.

Structural models for M.*Hin*Dam and M.HP1Dam

Homology modeling of M.HP1Dam and M.*Hin*Dam MTases was based on the alignments displayed in Figs. 2 and 3, respectively. The MODELLER program starts from an extended conformation of a polypeptide chain and builds three-dimensional protein models by satisfaction of spatial restraints extracted from the template structures according to given alignments. In the model building the backbone conformation of regions aligned with *M.DpnM* and *M.TaqI* structures was not changed, while the side-chain conformations were calculated based on local similarity to either of the templates. Analysis of the stereochemical parameters (backbone angles, bond lengths and chirality) did not reveal any unusual structural features. No major steric clashes were found in the final models. Although the structure evaluation programs reported low scores for some regions, all the problematic residues were found in loops, which we predict to interact with the cofactor AdoMet and the target DNA (see below). It is widely recognized that structure evaluation programs often misjudge the conformation and 3D-environment of ligand-binding residues as “unusual”, because such amino acids are generally excluded from calculations of parameters of evaluation functions [29]. Nev-

ertheless, the core regions of the M.*Hin*Dam and M.HP1Dam models were evaluated as reasonably folded by VERIFY3D (mean scores 0.463 and 0.291, respectively, with no regions scored < 0). The lower score for M.HP1Dam is quite typical for models based on correct alignments with templates of low sequence identity, and most likely reflects subtle differences in packing of side chains in the hydrophobic core caused by mutual shifts of peripheral secondary structure elements. However, these differences usually do not influence the global fold or the conformation of the essential side chains in the active site. Hence, we are confident that the predicted structures of both M.*Hin*Dam and M.HP1Dam reflect the essential features of the native proteins.

Comparative analysis of M.*Hin*Dam and M.HP1Dam

Figure 4 shows the cartoon representations of structural models of M.HP1Dam and M.*Hin*Dam in similar orientation. Taken together with the sequence alignments (Figs. 2 and 3) it depicts mutual similarities and differences at the level of primary, secondary and tertiary structure. The M.HP1Dam MTase and its homologs lack the separate target recognition domain positioned between motifs II and III, characteristic for M.*Hin*Dam and the α -subfamily of m⁶A MTases in general [5] (compare also Fig. 2 and 3). In addition, several short insertions/deletions (indels) are required to make the structure-based sequence alignment of the two enzymes (Figs. 4, 5). The overall lack of regions of strong sequence similarities, except for the residues of the catalytic motif IV, is striking. Although both isomethylomers appear markedly different at all levels of organization, some correlation between individual conserved positions and common structural features can be easily noted.

The AdoMet-binding pockets of M.HP1Dam and M.*Hin*Dam display pronounced differ-

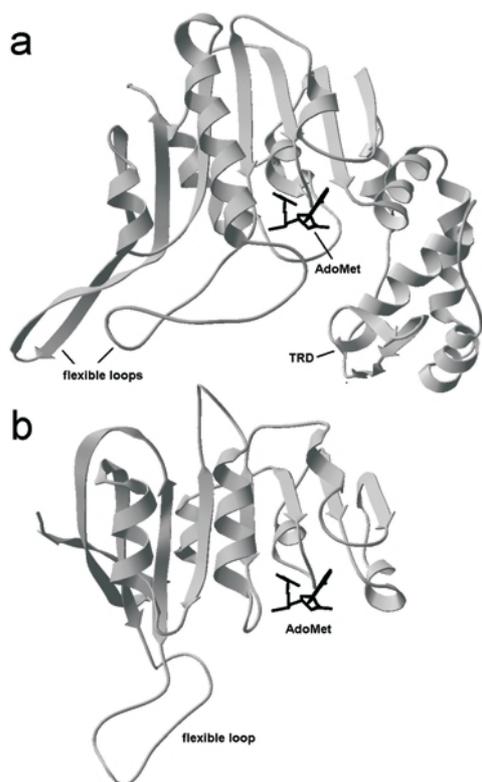


Figure 4. Cartoon representations of molecular models of a) *M.HinDam* and b) *M.HP1Dam* depicting the common structural core and strikingly different peripheral elements presumably involved in recognition of the target DNA sequence.

ences, which is in contrast to the previously reported high degree of conservation in most AdoMet-dependent MTases [2]. In *M.HinDam* the “FVGAG⁵¹” pentapeptide perfectly matches the motif I consensus pattern

		Motif I		Motif II	
<i>M.HinDam</i>	DE INKAF PNKNC LI	EP FVGAGAV F	LNSN FERY IL	AD IN PDL IN	LF (92) AQS AV
<i>M.HP1Dam</i>	MMTKS NTKK --S DK	DLWAT EWVVF	HYAEQY FN IK	EDLD TCAME	HN ---- TKVK N
		Motif III	Motif IV	Motif V	Motif VI
<i>M.HinDam</i>	FLCC	DFQKTFE	FADKDSV	IYCDPPYA	P (19) QRALADLA
<i>M.HP1Dam</i>	FI -T	PEQNTL	ADWQGRY	CWMNPPYS	N ---- PLPFVLR
		Motif VII		Motif VIII'	
<i>M.HinDam</i>	DTKFTRE	LYNGAKFKRVK	VQRS IS	QN PE	----- KRVRV
<i>M.HP1Dam</i>	VD -GSTK	WFDMCVRNAKE	IVYI TN	SR IP F INNETG	EE TDQNNK
					PEQMLVLFEPKAP

Figure 5. Pairwise alignment of *M.HinDam* and *M.HP1Dam* sequences derived from the structural superposition of molecular models (Fig. 4).

FxGxG, while in *M.HP1Dam* it is substituted by the atypical “WATPW¹⁹” sequence. However, both F⁴⁷ in *M.HinDam* and W¹⁵ in *M.HP1Dam* allow for similar stacking with

the adenine ring, facilitating accommodation of the AdoMet molecule and the sidechains of “atypical” residues in the motif I of *M.HP1Dam* do not seem to interfere with cofactor binding. Nevertheless, we predict that the TP¹⁹ dipeptide conserved in all phage MTases and absent from bacterial enzymes, may cause these two protein families to interact differently with those cofactor analogs, in which the methionine moiety is modified. Other pairs of residues predicted to participate in cofactor binding in a similar manner in bacterial/phage Dam MTases include: E⁴⁵/D¹³, which coordinate the methionine moiety *via* a water molecule, as recently demonstrated in the RrmJ MTase crystal structure [30]; D⁶⁶/D³⁴ and N⁶⁸/D³⁶, which hydrogen bond to the ribose hydroxyl groups; and D¹⁷⁷/E⁵³, which coordinate the adenine N6-amino group.

In the active site of both MTases, the conserved “(D/N)PPY” tetrapeptide of motif IV found in the vast majority of m⁶A MTases (DPPY¹⁹⁷/NPPY⁹² in *M.HinDam*/*M.HP1Dam*, respectively) lies at a strategic position in a bottom of the central cavity. The (D/N)P component is responsible for making two hydrogen bonds to the methylated N6-amino group and the Tyr residue is involved in face-to-face stacking with the adenine ring, as in the *M.TaqI* structure [31]. However, a loop present between motifs IV and V in

M.HinDam as well as in *M.TaqI*, is missing from *M.HP1Dam*. Structural analysis of *M.TaqI* in complex with DNA revealed that this loop participates in target recognition by

making “redundant” contacts with the target sequence *via* the minor groove [31]. This difference suggests that the bacterial and phage MTases differ in respect to the protein-DNA contacts.

The α -class m^6A MTases are believed to use similar elements for sequence recognition; they are the only class of MTases, which seems to show sequence similarity both in the catalytic domain and the TRD [5]. M.*Hin*Dam is predicted to possess the TRD cluster inserted between conserved motifs II and III, essentially identical to that of M.*Dpn*M MTase. It is not surprising that in both MTases recognizing the GATC target, virtually all amino acids of the “recognition loop”, with side chains pointing towards the DNA binding cleft, are perfectly conserved. It is therefore inferred that M.*Hin*Dam and M.*Dpn*M recognize their target identically.

On the other hand, studies on γ - m^6A MTases indicate that the TRD is generally located in the C-terminus and comprises a sizeable (over 100 aa) structural domain, which is not conserved at the level of the amino-acid sequence. It is not known if the TRD of γ - m^6A MTases is conserved at the structural level, since the crystal structure of only one member of this class (M.*Taq*I) is available and to our knowledge no systematic structure prediction study has been carried out to address this issue. Remarkably, the C-terminal extension is absent or not conserved in M.HP1Dam homologs. Besides, its small size (less than 30aa) suggests it forms a variable elaboration of the catalytic domain, as an additional β -hairpin in M.*Taq*I, rather than a separate functional domain (Figs. 2 and 4). But since the phage MTases lack the key elements known to interact with the DNA in M.*Taq*I and M.*Dpn*M, the question arises of how do they accomplish recognition of their target sequence?

Between motifs VII and VIII', which correspond to two antiparallel β -strands, M.HP1-Dam and its homologs possess a common insertion (Figs. 3, 4), which is absent from M.*Taq*I and many other γ - m^6A MTases [1].

α - m^6A MTases possess an insertion in a similar location, although it bears little sequence similarity to the insertion present in phage MTases. Given the high content of Gly, Asp, Asn and Lys in the insert of M.HP1Dam homologs and Ser, Asn, Arg and Lys in the α -MTases, it is quite probable that they both form flexible loops. According to the method of Karplus & Schulz [32] these insertions constitute the most flexible regions in all protein sequences analyzed here (not shown). It is noteworthy, that the most favorable amino acid side chain-DNA base interactions involve Lys, Asn, Asp and Arg [33]. Hence, it is tempting to speculate that this region may contribute to DNA binding in both Dam MTase families.

It is beyond the limits of current modeling methodology to predict with confidence, if these loops adopt similar conformations and make similar contacts with the DNA. Since the primary sequence specificity determinants of α -MTases reside within the TRD, we speculate that this additional loop may contribute to the redundant readout of the target sequence. It is possible that that M.HP1Dam and its homologs use primarily one loop to make only a few essential amino acid-DNA interactions, whereas M.*Hin*Dam and other α -MTases use the “primary” TRD to recognize the GATC sequence and several other loops to precisely tune the sequence context-dependent specificity. This hypothesis can be tested by mutagenesis of the extended loops of M.HP1Dam aimed at fairly easy change in apparently not strongly constrained sequence specificity.

Evolutionary relationship of bacterial and phage Dam MTases

It is widely recognized that evolutionary divergence of protein structures occurs much less rapidly than divergence of protein sequences, and that tertiary structural similarities are generally retained at the expense of all else [34]. This indicates that selective con-

straints act to preserve protein structure, including spatial arrangement of both secondary structural elements and amino acid side-chains in functionally important positions, rather than sequence alone, as shown for example for different m⁶A MTase families [5]. The similarities of sequence and structure of both Dam MTases probably reflect the general physicochemical constraints to position the target adenine and the methyl group donor, and to facilitate the packing of hydrophobic interior. Nevertheless, the target recognition determinants of M.HinDam and M.HP1Dam differ to such extent that it makes their close relationship very unlikely. These differences are even higher than differences between M.HinDam and other α -m⁶A MTases, which recognize sequences other than GATC (J.M. Bujnicki & M. Radlinska, manuscript in preparation). Only several key residues and the topological arrangement of secondary structural elements, so far unique for the AdoMet-dependent MTase superfamily, are conserved – these features have been interpreted as evidence of a remote common origin [2, 35].

There is no evidence that M.HinDam and M.HP1Dam retained a similar DNA binding mode in the course of divergent evolution. They share only one candidate “recognition loop” with no mutual sequence similarity. The eroded sequence similarity of M.HP1/T1Dam homologs with all known m⁶A MTases including both α and γ classes suggests that the phage MTase family is very ancient or that it evolved rapidly. Nonetheless, the phage Dam MTases exhibit weak, but detectable similarity to other γ -MTases and no statistically significant similarity to α -MTases, which precludes common origin of the M.HP1/T1Dam subfamily with the α -class Dam MTases in another denotation than a common origin of α and γ classes or of m⁶A MTases in general. The data presented here suggest that M.HinDam and M.HP1Dam share a remote common ancestor, but their similar DNA specificities result from a functional conver-

gence rather than conservation of specificity over vast phylogenetic distances. Moreover, the two remotely homologous isomethylomers of *H. influenzae* might have been brought together only recently, when the HP1 phage (or its ancestor) acquired the “viral” *dam* gene. The role of N6-adenine methylation in the HP1 life cycle is an open question, as neither P2 nor 186 coliphages related to HP1 appear to encode the MTase activity. Given the contrasted dealings of the lytic “killer phage” T1 and lysogenic HP1 with their hosts the mutual similarity of their Dam MTases becomes particularly interesting.

Lysogenic bacteriophages are major vehicles for the transfer of genetic information between bacteria, including pathogenicity and virulence determinants [36]. Defective, incomplete, and cryptic prophages are ubiquitous in all fully sequenced bacterial genomes. Our analysis of the newly discovered M.HP1/T1Dam subfamily showed that these proteins are exclusively present in bacteriophage-related elements, of which the 933W coliphage is a recognized virulence factor [36]. The analysis of sequences adjacent to the M.HP1Dam/M.T1Dam-related open reading frames in all these elements indicated lack of significant mutual similarity (not shown). Given lack of close relationship between Dam-encoding phages, it suggests that the HP1/T1-like *dam* genes have been spread by horizontal gene transfer. A significant role in pathogenesis of *H. somnus* has been suggested for a cryptic prophage related to HP1 [37]. It remains to be determined, whether this prophage carries an M.HP1Dam-related gene.

It has been recognized that there is vigorous and ongoing horizontal gene transfer among the well-studied lambdoid phages of *E. coli* and *Salmonella* [38]. It is known that recombination can create new variants of viruses by either capturing genes or whole DNA segments from cellular nucleic acids or from other viruses [39]. This is well-illustrated in the case of HP1-related phages [40] and the recently discovered cryptic prophage ϕ -flu of

H. influenzae [41]. The data presented here suggest that the HP1/T1-like *dam* genes are intensely exchanged among phage and prophage genomes through the “common gene pool” [41]. The mechanisms ruling the variability of these elements and their recombination with host chromosome play a significant role in bringing together the virulence factors that make up the disease-potential of many bacteria.

Dam MTases as potential targets for the drug-design

Structurally similar AdoMet-utilizing enzymes are crucial not only for bacteria but also for the mammalian metabolism [42]. Hence, despite the availability of AdoMet analogs, like sinefungin, which could inhibit the Dam MTase, their lack of specificity precludes them from clinical use. The alternative strategy, which could overcome dissimilarities of both Dam MTases, would be the use of oligonucleotides, containing the recognition sequence, but incapable of being methylated and blocking the enzyme active site. It has been shown that DNA:m⁶A MTases does not recognize the methylated base *per se*, and that they bind to DNA substrates containing mismatches at the target position within the recognition sequence [43]. Oligonucleotides containing cytosine analogs have been found to have an inhibitory effect on DNA m⁵C MTases [44]. However, in the case of cytosine-C5, methylation proceeds *via* a covalent enzyme-substrate intermediate, whereas adenine-N6 methylation proceeds *via* one-step reaction [45]. Thus, such “suicidal” substrates based on adenine analogs are not available. Gel shift analysis demonstrated that the M.HhaI MTase had a higher affinity for oligonucleotides containing an abasic site instead of the target cytosine compared to the normal substrates [46], what indicates that oligonucleotides lacking target adenines in the symmetrical GATC site might serve as inhibitors for the Dam MTases. All DNA MTases

tested so far, regardless of the target specificity and reaction mechanism, bind the abasic oligonucleotide more strongly than the natural substrates. In principle, regardless of differences in the target DNA binding mode of both Dam families, the same modified substrate might be used against proteins of bacterial and viral origin.

Although the presented models provide a great deal of structural information about cofactor binding, information regarding the DNA binding is necessary for design of more sophisticated Dam inhibitors. This includes definition of specific amino acid-DNA base contacts, the “indirect readout” through phosphate backbone and conformational changes associated with substrate binding. Nevertheless, even in the absence of high-resolution X-ray crystallography or NMR structure determination, our models may serve as a guide for allocation of secondary structural elements and amino-acid residues of importance for further investigations. Importantly, we predict that the cofactor analogs may interact with different side chains in M.HP1Dam and M.HinDam MTases, which suggests that drugs designed against the cofactor-binding pocket of the bacterial enzyme may be ineffective against its phage isomethylomer and *vice versa*.

Phage genes can be beneficial to the host by enhancing its survival in certain limiting environments, in some cases providing virulence factors. It is widely accepted that bacterial genomes would ultimately incorporate all genes that potentially can function there, and give the organism selectional advantage. When the bacterial gene is lost or the enzymatic activity of its product is destroyed, the viral paralogous gene might become essential. The *S. typhimurium* bacteria lacking Dam activity were totally avirulent [13]. It is tempting to speculate, and remains to be tested, if Dam mutants of some pathogenic Proteobacteria would lose their virulence due to the disturbance of expression of Dam-regulated genes. To date, it has been shown to not to be

the case in the Dam mutants of *Neisseria meningitidis* [47]. Still, there is a serious threat that the *dam* strains of *Salmonella* and *Escherichia* planned to be used as vaccines may become lysogenized with Dam-expressing phages. The high activity of phage Dam MTases may serve to increase the expression of "unwanted" genes and decrease the value of the vaccine and even make it potentially dangerous. Besides, the virulent strains lysogenized with Dam-expressing phages may be resistant to the drugs targeting solely the bacterial-type Dam MTase. Hence, the inhibition of a bacterial Dam MTase should be accompanied by inhibition of its phage counterpart. We hope that our study will help to better understand the differences between these enzymes and guide the development of inhibitors against them.

CONCLUSIONS

In this paper we addressed the questions regarding common ancestry of two *dam* gene families by comparison of sequence and structures of their products. We also cloned and characterized the function of *M.HinDam*, which corroborated the sequence-based prediction. The lack of 3D information for Dam MTases prompted us to build homology models, which provide insight into the protein sequence-structure-function relationships in advance of experimental structure determination. The predicted structural similarity and the conservation of the functionally important residues suggest a distant evolutionary relationship between Dam MTases of *H. influenzae* and its phage and similar Dam MTase pairs found in other enteropathogenic bacteria. The models presented here helped to visualize significant differences between the host and the viral Dam MTases, which we find crucial in designing future experiments aimed at elucidation of the *M.HP1Dam* precise role in development of HP1 phage and possibly the virulence of *H. influenzae*. This, as well as veri-

fication of the importance of the analyzed regions and individual residues by mutagenesis of both MTases, will be addressed in a subsequent study.

The presented data also considerably extend our knowledge on m⁶A MTases and the problem of evolution of target DNA recognition. To date, results of comparative analyses were published only for closely related MTases, whose cognate specificity resulted from a limited divergence and not evolutionary selection for a particular function. We hope that our study will help to better understand the mechanisms of target recognition employed by DNA m⁶A MTase and the role the bacterial and phage Dam MTases play in the evolution of bacterial virulence. We also hope that our prediction of enzyme-ligand interactions both common and unique to the families of *M.HP1Dam* and *M.HinDam* homologs will aid the design of potential Dam inhibitors.

REFERENCES

1. Malone, T., Blumenthal, R.M. & Cheng, X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **253**, 618-632.
2. Fauman, E.B., Blumenthal, R.M. & Cheng, X. (1999) Structure and evolution of AdoMet-dependent Mtases; in: *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions* (Cheng, X., *et al.*, eds.) pp. 1-38, World Scientific Inc., Singapore.
3. Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I. & Saenger, W. (1994) Three-dimensional structure of the adenine-specific DNA methyltransferase *M.TaqI* in complex with the cofactor *S*-adenosylmethionine. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10957-10961.
4. Bujnicki, J.M. (2001) Understanding the evolution of restriction-modification systems:

- clues from sequence and structure comparisons. *Acta Biochim. Polon.* **48**, 935–967.
5. Tran, P.H., Korszun, Z.R., Cerritelli, S., Springhorn, S.S. & Lacks, S.A. (1998) Crystal structure of the *DpnM* DNA adenine methyltransferase from the *DpnII* restriction system of *Streptococcus pneumoniae* bound to *S*-adenosylmethionine. *Structure* **6**, 1563–1575.
 6. Modrich, P. (1991) Mechanisms and biological effects of mismatch repair. *Annu. Rev. Genet.* **25**, 229–253.
 7. Modrich, P. (1989) Methyl-directed DNA mismatch correction. *J. Biol. Chem.* **264**, 6597–6600.
 8. Oka, A., Sugimoto, K., Takanami, M. & Hirota, Y. (1980) Replication origin of the *Escherichia coli* K-12 chromosome: The size and structure of the minimum DNA segment carrying the information for autonomous replication. *Mol. Gen. Genet.* **178**, 9–20.
 9. Crooke, E. (1995) DNA synthesis initiated at *oriC*: *In vitro* replication reactions. *Methods Enzymol.* **262**, 500–506.
 10. Bogan, J.A. & Helmstetter, C.E. (1997) DNA sequestration and transcription in the *oriC* region of *Escherichia coli*. *Mol. Microbiol.* **26**, 889–896.
 11. Palmer, B.R. & Marinus, M.G. (1994) The *dam* and *dcm* strains of *Escherichia coli* – a review. *Gene* **143**, 1–12.
 12. Scherzer, E., Auer, B. & Schweiger, M. (1987) Identification, purification, and characterization of *Escherichia coli* virus T1 DNA methyltransferase. *J. Biol. Chem.* **262**, 15225–15231.
 13. Heithoff, D.M., Sinsheimer, R.L., Low, D.A. & Mahan, M.J. (1999) An essential role for DNA adenine methylation in bacterial virulence. *Science* **284**, 967–970.
 14. Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. & Venter, J.C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.
 15. Stuy, J.H. (1974) Origin and direction of *Haemophilus* bacteriophage HP1 DNA replication. *J. Virol.* **13**, 757–759.
 16. Piekarowicz, A. & Bujnicki, J.M. (1999) Cloning of the *Dam* methyltransferase gene from *Haemophilus influenzae* bacteriophage HP1. *Acta Microbiol. Pol.* **48**, 123–129.
 17. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*; 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 18. Barcak, G.J., Chandler, M.S., Redfield, R.J. & Tomb, J.F. (1991) Genetic systems in *Haemophilus influenzae*. *Methods Enzymol.* **204**, 321–342.
 19. Mayer, M.P. (1995) A new set of useful cloning and expression vectors derived from pBlueScript. *Gene* **163**, 41–46.
 20. Renbaum, P. & Razin, A. (1995) Interaction of *M.SssI* and *M.HhaI* with single-base mismatched oligodeoxynucleotide duplexes. *Gene* **157**, 177–179.
 21. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
 22. Bujnicki, J.M., Elofsson, A., Fischer, D. & Rychlewski, L. (2001) Structure prediction Meta Server. *Bioinformatics* **17**, 750–751.

23. Sali, A. & Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
24. Hooft, R.W., Vriend, G., Sander, C. & Abola, E.E. (1996) Errors in protein structures. *Nature* **381**, 272.
25. Eisenberg, D., Luthy, R. & Bowie, J.U. (1997) VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* **277**, 396–404.
26. McClelland, M., Nelson, M. & Raschke, E. (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res.* **22**, 3640–3659.
27. Radlinska, M. & Bujnicki, J.M. (2001) Cloning of enterohemorrhagic *Escherichia coli* phage VT-2 Dam methyltransferase. *Acta Microbiol. Pol.* **50**, 157–163.
28. Fischer, D. & Eisenberg, D. (1996) Protein fold recognition using sequence-derived predictions. *Protein Sci.* **5**, 947–955.
29. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
30. Bugl, H., Fauman, E.B., Staker, B.L., Zheng, F., Kushner, S.R., Saper, M.A., Bardwell, J.C. & Jakob, U. (2000) RNA methylation under heat shock control. *Mol. Cell* **6**, 349–360.
31. Goedecke, K., Pignot, M., Goody, R.S., Scheidig, A.J. & Weinhold, E. (2001) Structure of the N6-adenine DNA methyltransferase M. *TaqI* in complex with DNA and a co-factor analog. *Nat. Struct. Biol.* **8**, 121–125.
32. Karplus, P.A. & Schulz, G.E. (1985) Prediction of chain flexibility in proteins. *Naturwissenschaften* **72**, 212–213.
33. Lustig, B. & Jernigan, R.L. (1995) Consistencies of individual DNA base-amino acid interactions in structures and sequences. *Nucleic Acids Res.* **23**, 4707–4711.
34. Chothia, C. & Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823–826.
35. Bujnicki, J.M. (1999) Comparison of protein structures reveals monophyletic origin of the AdoMet-dependent methyltransferase family and mechanistic convergence rather than recent differentiation of N4-cytosine and N6-adenine DNA methylation. *In Silico Biol.* **1**, 1–8, <http://www.bioinfo.de/isb/1999-01/0016/>.
36. Plunkett, G., Rose, D.J., Durfee, T.J. & Blattner, F.R. (1999) Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J. Bacteriol.* **181**, 1767–1778.
37. Pontarollo, R.A., Rioux, C.R. & Potter, A.A. (1997) Cloning and characterization of bacteriophage-like DNA from *Haemophilus somnus* homologous to phages P2 and HP1. *J. Bacteriol.* **179**, 1872–1879.
38. Campbell, A. (1994) Comparative molecular biology of lambdoid phages. *Annu. Rev. Microbiol.* **48**, 193–222.
39. Holland, J. & Domingo, E. (1998) Origin and evolution of viruses. *Virus Genes* **16**, 13–21.
40. Piekarowicz, A., Brzezinski, R., Smorawska, M., Kauc, L., Skowronek, K., Lenarczyk, M., Golembiewska, M. & Siwinska, M. (1986) Major spontaneous genomic rearrangements in *Haemophilus influenzae* S2 and HP1c1 bacteriophages. *Gene* **49**, 111–118.
41. Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E. & Hatfull, G.F. (1999) Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2192–2197.
42. Fujioka, M. (1992) Mammalian small molecule methyltransferases: Their structural and func-

- tional features. *Int. J. Biochem.* **24**, 1917–1924.
- 43.** Jeltsch, A., Christ, F., Fatemi, M. & Roth, M. (1999) On the substrate specificity of DNA methyltransferases. Adenine-N6 DNA methyltransferases also modify cytosine residues at position N4. *J. Biol. Chem.* **274**, 19538–19544.
- 44.** Sheikhnejad, G., Brank, A., Christman, J.K., Goddard, A., Alvarez, E., Ford, H.J., Marquez, V.E., Marasco, C.J., Sufrin, J.R., O’Gara, M. & Cheng, X. (1999) Mechanism of inhibition of DNA (cytosine C5)-methyltransferases by oligodeoxyribonucleotides containing 5,6-dihydro-5-azacytosine. *J. Mol. Biol.* **285**, 2021–2034.
- 45.** Ho, D.K., Wu, J.C., Santi, D.V. & Floss, H.G. (1991) Stereochemical studies of the C-methylation of deoxycytidine catalyzed by *HhaI* methylase and the N-methylation of deoxyadenosine catalyzed by *EcoRI* methylase. *Arch. Biochem. Biophys.* **284**, 264–269.
- 46.** Yang, A.S., Shen, J.C., Zingg, J.M., Mi, S. & Jones, P.A. (1995) *HhaI* and *HpaII* DNA methyltransferases bind DNA mismatches, methylate uracil and block DNA repair. *Nucleic Acids Res.* **23**, 1380–1387.
- 47.** Richardson, A.R. & Stojiljkovic, I. (2001) Mismatch repair and the regulation of phase variation in *Neisseria meningitidis*. *Mol. Microbiol.* **40**, 645–655.