doi:10.1006/jmbi.2000.3588 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 297, 293–300

JMB

COMMUNICATION

Structure and Function of the Mouse DNA Methyltransferase Gene: *Dnmt1* shows a Tripartite Structure

Jean B. Margot[1]†, Ana M. Aguirre-Arteta[1]†, B. Viviana Di Giacco[1] Sriharsa Pradhan[2], Richard J. Roberts[2], M. Cristina Cardoso[1] and Heinrich Leonhardt[1]*

[1]*Max Delbrück Center for Molecular Medicine D-13125, Berlin, Germany*

[2]*New England Biolabs, 32 Tozer Road, Beverly MA 01915, USA*

**Corresponding author*

Dnmt1 is the predominant DNA methyltransferase (MTase) in mammals. The C-terminal domain of Dnmt1 clearly shares sequence similarity with many prokaryotic 5mC methyltransferases, and had been proposed to be sufficient for catalytic activity. We show here by deletion analysis that the C-terminal domain alone is not sufficient for methylating activity, but that a large part of the N-terminal domain is required in addition. Since this complex structure of Dnmt1 raises issues about its evolutionary origin, we have compared several eukaryotic MTases and have determined the genomic organization of the mouse *Dnmt1* gene. The 5′ most part of the N-terminal domain is dispensible for enzyme activity, includes the major nuclear import signal and comprises tissue-specific exons. Interestingly, the functional subdivision of Dnmt1 correlates well with the structure of the *Dnmt1* gene in terms of intron/exon size distribution as well as sequence conservation. Our results, based on functional, structural and sequence comparison data, suggest that the gene has evolved from the fusion of at least three genes.

Keywords: DNA methyltransferase; genomic organization; methylase activity; Dnmt1

Introduction

In eukaryotes DNA methylation plays a central role in genomic imprinting (Li *et al*., 1993), X chromosome inactivation (Panning & Jaenisch, 1996) and development (Li *et al*., 1992). Changes in methylation levels have also been correlated with cancer and gene expression patterns (Bakin & Curran, 1999; Boyes & Bird, 1991; Vertino *et al*., 1996). Dnmt1 is the enzyme responsible for the post-replicative addition of a methyl group to cytosine residues at CpG sites of the newly synthesized strand and thereby acting as a maintenance methyltransferase. The protein has been typically divided into a ~1000 aa N-terminal domain linked to a ~500 aa C-terminal domain by a stretch of 12 alternating glycyl and lysyl residues. Several functions have been localized to the N-terminal domain of the mammalian enzyme including several NLS (Cardoso & Leonhardt, 1999; Leonhardt *et al*., 1992), a PCNA binding site (Chuang *et al*., 1997), several sequences targeting the enzyme to replication foci (Leonhardt *et al*., 1992; Liu *et al*., 1998), and a cysteine-rich Zn binding region (Bestor, 1992). The C-terminal domain of Dnmt1 is clearly related to the prokaryotic 5mC MTases, since it harbours a set of ten motifs that are involved in the catalytic reaction and whose sequence is shared with over 100 prokaryotic MTases (Cheng, 1995; Lauster *et al*., 1989; Posfai *et al*., 1989). It is not clear, however, whether the C-terminal domain of the eukaryotic enzyme is sufficient for the catalytic activity of the enzyme or whether the N-terminal domain is also required.

Mapping of Dnmt1 catalytic domain

To determine which parts of Dnmt1 are necessary for enzymatic activity, several deletion mutants

†Both authors contributed equally to this work.

Abbreviations used: aa, amino acid; MTase, methyltransferase.

E-mail address of the corresponding author: leonhardt@fvk-berlin.de

0022-2836/00/020293–8 $35.00/0

of the enzyme were constructed. Two isoforms of Dnmt1 have been characterized. A long form, which is present in embryonic stem cells and somatic tissues, starts at the third in-frame ATG (ATG_3) in exon 1 (Gaudet *et al.*, 1998; Glickman *et al.*, 1997) and a shorter form that is found in oocytes and pre-implantation embryos and starts at the methionine (ATG_4) in exon 4 (Gaudet *et al.*, 1998). As the protein starting at ATG_4 shows similar enzymatic activity *in vitro* (Pradhan *et al.*, 1997) and *in vivo* (Gaudet *et al.*, 1998) to the one starting at ATG_3, we chose the shorter isoform for our analysis. Constructs expressing various deletions of Dnmt1 were expressed in COS-7 cells and partially purified enzymes were assayed for methylation activity (Figure 1). Although previous biochemical analyses on purified Dnmt1 had suggested that the C-terminal domain would, like their prokaryotic

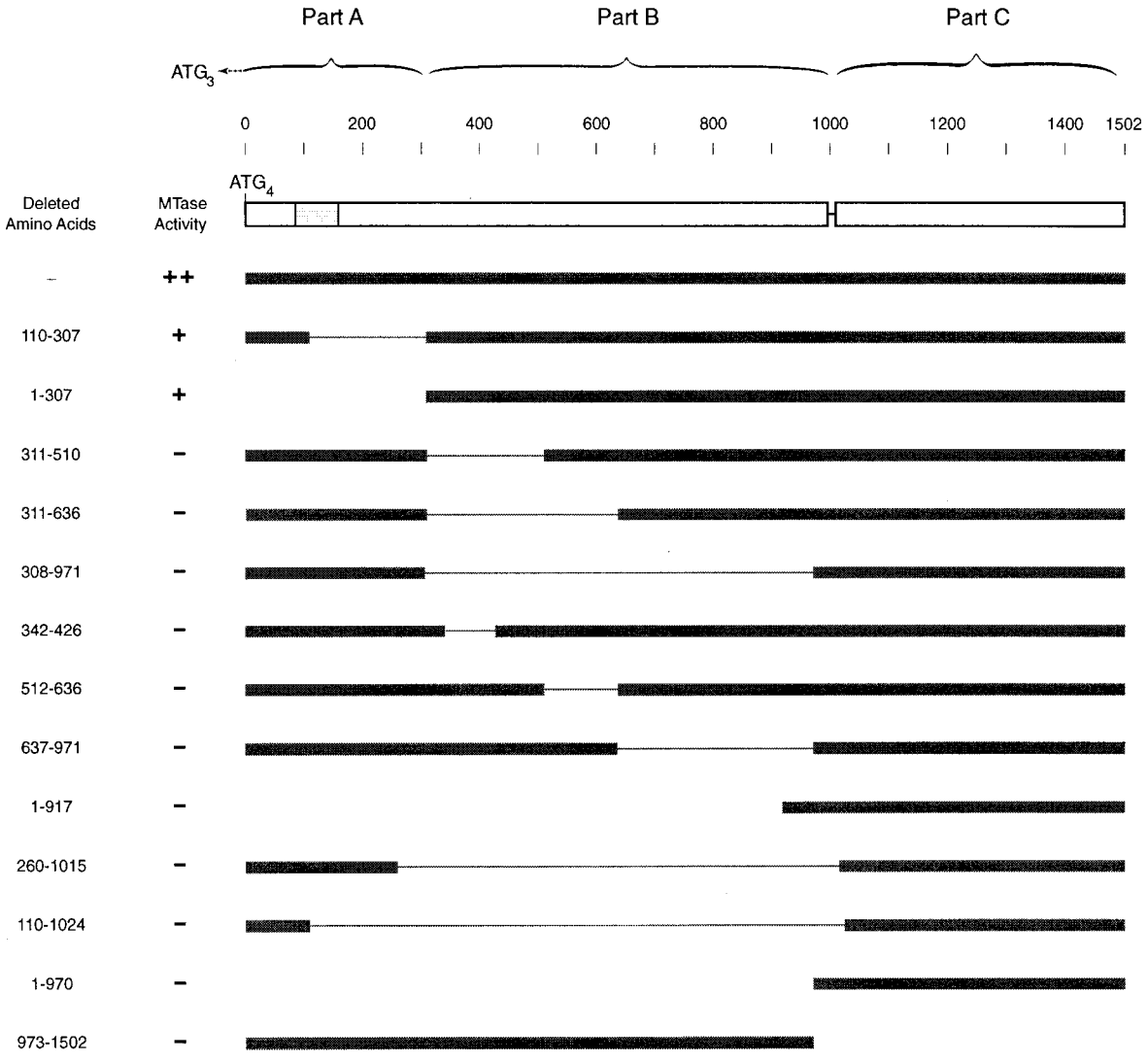

Figure 1. Enzymatic activity of Dnmt1 deletion mutants. Deletion mutants were generated from an initial construct carrying the short MTase isoform cloned into the mammalian expression plasmid pEVRF0 (Czank *et al.*, 1991). Two days after transfection, COS-7 cells were extracted for five minutes on ice in 0.32 M sucrose, 0.3% Triton X-100, 20 mM Tris-HCl (pH 7.4), 0.2 M NaCl, 3 mM $MgCl_2$, 0.5 mM DTT, 0.2 mM PMSF. Cleared supernatants were adjusted to 40 mM NaCl and incubated with pre-equilibrated DEAE Sephacel. After washing with 20 mM Tris-HCl (pH 7.4), 40 mM NaCl, Dnmt1 was eluted with 20 mM Tris-HCl (pH 7.4), 200 mM NaCl. DNA methyltransferase activity was assayed by incubation for 15, 30 or 60 minutes at 37 °C in reaction buffer (20 mM Tris-HCl (pH 7.4), 1 mM EDTA, 20% (v/v) glycerol, 0.5 mM DTT, 0.2 mM PMSF, 10 µg/ml poly(dI-dC), 20 µCi/ml *S*-adenosyl-L-[^{3}H-*methyl*]methionine). After removal of proteins by phenol extraction, the DNA was TCA precipitated and counted. The (+) enzymatic activity corresponds to ~25% of the activity of the intact protein (++) while the (–) activity is indistinguishable from that of the vector control. N and C-terminal domains of the protein are diagrammed as open boxes linked by 12 alternating lysyl glycyl residues (thin line). The corresponding position of the ~90 aa deletion in chicken and *Xenopus* MTases is indicated by a grey box (see the text below). Amino acids remaining in the deletion constructs are indicated with filled bars.

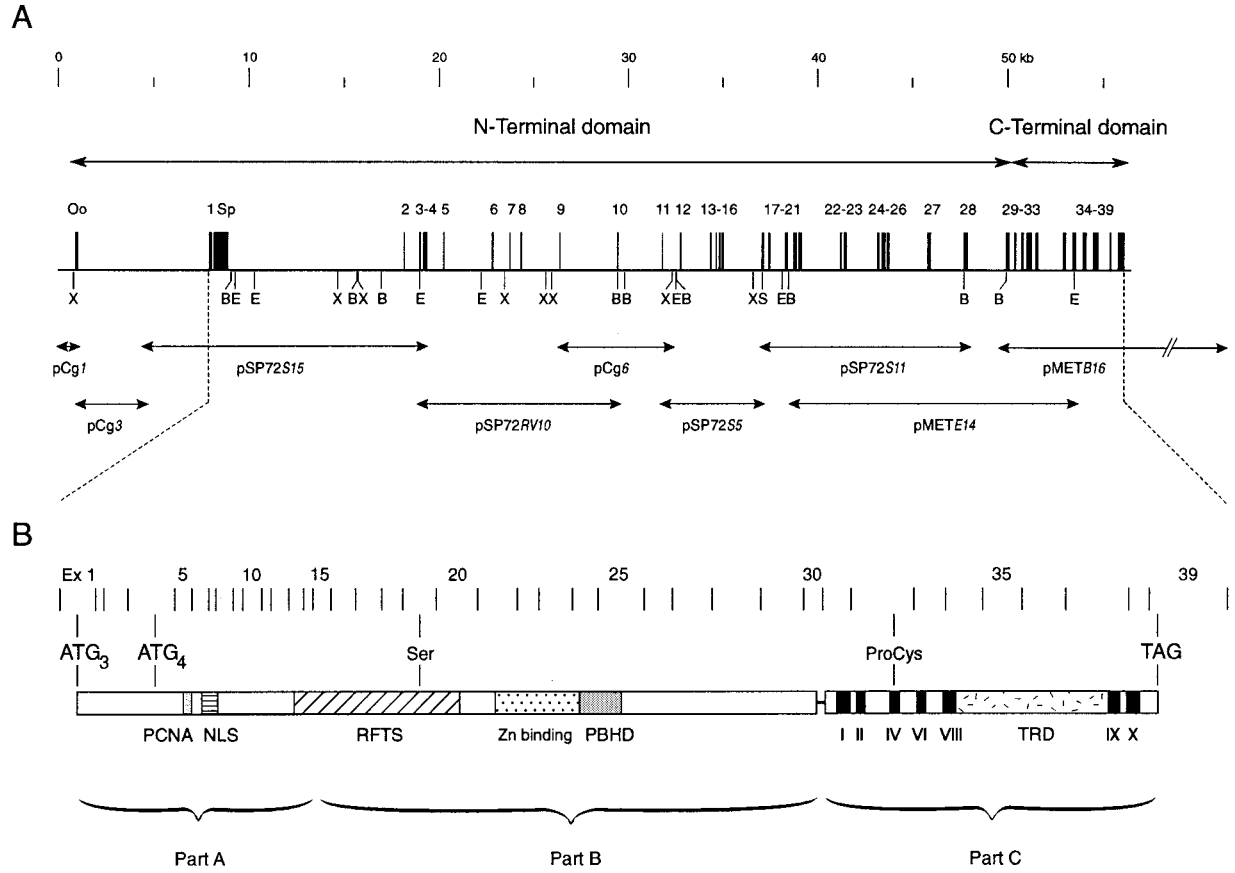

Figure 2. Genomic organization of the mouse DNA MTase gene. (a) Restriction map and location of the cloned genomic DNA fragments within the *Dnmt1* locus. Overlapping phage clones were isolated by screening a mouse 129/Sv genomic library (Li *et al.*, 1992) with the Dnmt1 cDNA. The phage inserts were subdoned into pSP72 (Promega) or pMET (Bluescript KS vector, Stratagene). The region between pSP72*RV10* and pSP72*S5* was obtained by genomic PCR using DNA prepared from frozen 12 day-old mouse embryos strain C57BL6. DNA (100 ng in a reaction volume of 100 µl) were amplified with a mix of *Taq* and *Pwo* DNA polymerases (Expand™ Long Template PCR System, Roche Molecular Biochemicals). Primers as well as annealing and melting temperatures were selected with the Lasergene program (DNASTAR Inc., Madison, WI). The ~6 kb PCR product was directly cloned into PCRII-Topo vector (Invitrogen) and the resulting plasmid designated pCg*6*. The region upstream of pSP2*S15* was obtained by a combination of genomic PCR and genome walking using the GenomeWalker kit (Clontech, Germany). Plasmid insert sizes are indicated in Italics; B: *Bam*HI; E: *Eco*RI; S: *Sal*I; X: *Xba*I. The mouse Dnmt1 cDNA was completely resequenced and the revised sequence submitted to GenBank (accession number AF 162282). (b) Alignment of the exon boundaries with the functional domains of the protein. The boxed N and C-terminal domains are linked by 12 alternating lysyl glycyl residues (thick line between exons 30-31). NLS: nuclear localization sequence (Leonhardt *et al.*, 1992); Ser: phosphorylation site at serine 514 (Glickman *et al.*, 1997); RFTS: replication foci targeting sequence (Leonhardt *et al.*, 1992); PCNA: binding site to PCNA (Chuang *et al.*, 1997); *in vitro* zinc binding (Cys-rich) region (Bestor, 1992); PBHD: polybromo-1 protein homologous domain (Liu *et al.*, 1998); ProCys: dipeptide part of the catalytic site (Cheng *et al.*, 1993; Wu & Santi, 1987); TRD: target recognition domain (Balganesh *et al.*, 1987; Trautner *et al.*, 1988; Wilke *et al.*, 1988); filled boxes: highly conserved sequence motifs (I, II, IV, VI, VIII, IX, X).

relatives, be enzymatically active by itself (Bestor, 1992), our results indicate that constructs carrying only the C-terminal domain failed to show methyltransferase activity. In fact, any tested deletion beyond amino acid 308 completely abolished enzyme activity. These results thus show that, in addition to the C-terminal domain, most of the N-terminal domain (marked as part B) is also required for methyltransferase activity, while the first ~300 aa (part A, exons 1-16) seem to be dispensable. The discrepancy with a previous biochemical study (Bestor, 1992) could be explained either (i) by the fact that both N and C-domains were still present in the proteolysed mix used for enzyme activity or (ii) by a suggested requirement of the N-terminal domain (part B) for the correct three-dimensional folding of the protein (Zimmermann *et al.*, 1997).

Mouse *Dnmt1* gene organization and evolution

The above data led us to look at the genomic organization of mouse *Dnmt1* to determine whether the region necessary for methyltransfer-

Table 1. Intron-exon boundaries in mouse *Dnmt1*

	Exon			Intron			Exon
#	Size	5' Donor site		Size	Phase	3' Acceptor site	
							gccccc..
1	160	..cgc agg CG	g t a g g t..	10.098	2	..a c c c a g	G ctc aaa..
2	37	..gaa AAG	g t a a a c..	0.762	0	..c t a c a g	gag tgt..
3	108	..tct GAG	g t a a g t..	0.115	0	..t t g t a g	gaa ggc..
4	211	..acc ctt T	g t a a g g..	0.818	1	..t t t a a g	CA gtt gaa..
5	79	..acg aag GG	g t g a g t..	2.5	2	..t g a c a g	C ccc act..
6	76	..gac CAG	g t g g g c..	0.876	0	..g c c c a g	gat aag..
7	32	..aca gag AG	g t a a g a..	0.539	2	..t g a c a g	T ggt gct..
8	79	..gaa CAG	g t a a a g..	1.974	0	..t c c t a g	gaa gat..
9	41	..aga gag CT	g t a a g t..	3.2	2	..t g g c a g	A tca ttg..
10	85	..aaa AAG	g c a a a g..	2.1	0	..t g t c a g	gat aaa..
11	41	..aga gat CC	g t a a g t..	0.9	2	..t t g t a g	A gct gcc..
12	82	..gag AGG	g t a a g t..	1.8	0	..g t t t a g	gag gag..
13	65	..cag agc AG	g t g a t g..	0.258	2	..t t c c a g	A tcg gag..
14	43	..ccg AAG	g t a a g t..	0.094	0	..c c a c a g	atc aac..
15	81	..gat GCT	g t g a g t..	0.078	0	..c t t c a g	gtg gat..
16	110	..tcc ttc AG	g t a a t c..	1.9	2	..a t t c a g	T gtg tac..
17	119	..atg gaa G	g t a t g t..	0.228	1	..c t a c a g	GT ggt att..
18	93	..tcc act G	g t g a g t..	0.8	1	..t c c t a g	CA ttt gct..
19	152	..att GAG	g t g c g a..	0.284	0	..t t a a a g	acc act..
20	188	..gga cag AG	g t a a g g..	0.093	2	..c c t c a g	G cga gca..
21	178	..tgt GAG	g t g a g c..	2	0	..t t t c a g	gtc tgt..
22	98	..aag agg AG	g t a g g c..	0.121	2	..c t g c a g	G tgt cct..
23	148	..atg AAG	g t g a g g..	1.5	0	..t t t c a g	att gaa..
24	116	..cta gcc AG	g t a t g c..	0.114	2	..c t g c a g	G gtc aca..
25	209	..atg gag gga G	g t g a g t..	0.077	1	..a g g g a g	GC aca gac..
26	133	..aag cac AA	g t g a g t..	0.435	2	..c t g c a g	G ttc tgc..
27	174	..act ttc AA	g t g a g t..	1.7	2	..c t g c a g	C atc aaa..
28	219	..ttc tac AG	g t g g g c..	1.7	2	..t c a c a g	G cct gag..
29	193	..ctc GAG	g t g g a g..	0.275	0	..c t g c a g	gcc tac..
30	85	..ggg aaa G	g t g t g t..	0.267	1	..g g g a a g	GG aag ggg..
31	129	..caa gca G	g t a a g c..	0.160	1	..t g c c a g	GC atc tcg..
32	283	..ttc ctc AG	g t a a g c..	0.174	2	..c t a c a g	C tac tgt..
33	142	..ctc CAG	g t g g g c..	1.3	0	..g t g c a g	gct gga..
34	167	..ata acg AG	g t c c g g..	0.334	2	..c t g c a g	G ctg agc..
35	178	..tgc AAG	g t a g g t..	0.350	0	..t t c t a g	gac atg..
36	196	..gca gaa G	g t g g g a..	0.337	1	..g a a c a g	GC aag gcc..
37	281	..cgg CAG	g t c a g t..	0.603	0	..c c g c a g	gtg ggt..
38	91	..gca tca G	g t a c g t..	0.309	1	..c c t c a g	CT gca gtt..
39	352	Until poly(A) addition site					
		a g **g t** r a g t				y y n c **a g**	
mMTase consensus		79 84 100 97 95 63 84 45				79 79 74 100 100	
Rodent consensus		56 79 100 100 92 73 82 54				85 85 73 100 100	

The protein starts at the third in-frame ATG (ATG_3) in exon 1 (Gaudet *et al.*, 1998) and stops at the TAG stop codon, which is located 39 bp into exon 39 and is followed by a 311 bp untranslated region. Intron sizes were determined by sequencing (GenBank accession numbers AF 175410-175431 and AF 234317-23418) and/or gel electrophoresis. The size of intron 1 includes the sperm specific exon, while the broken line separating exon 30 and 31 marks the separation between the previously proposed catalytic and regulatory domains.

ase activity is conserved among other mammals and whether functional domains of this MTase correlate with structural features of the gene. The genomic region containing *Dnmt1* has been isolated and the results of the mapping and sequencing data are diagrammed in Figure 2(a). Unless otherwise indicated, we will refer here only to the protein found in somatic cells whose coding sequence starts at ATG_3 (Gaudet *et al*., 1998; Tucker *et al*., 1996). The gene spans over 56 kb of genomic sequence and is divided into 39 exons whose size ranges between 32 bp (exon 7) and 352 bp (exon 39). The TAG stop codon is found 39 bp into exon 39. A polyadenylation signal (AATAAA) was mapped 290 bp downstream of the stop codon. Another potential polyadenylation signal was identified 274 bp further downstream. Intron sizes, with the exception of two large introns after oocyte and spermatocyte specific exons, range between 77 bp and ~3 kb (Table 1). All 5′ and 3′ splice sites agree with the consensus found for rodents (Senapathy *et al*., 1990; Shapiro & Senapathy, 1987) with the exception of the 5′ donor site of intron 10 where the GT consensus is replaced by GC. Interestingly, the human gene, the only other eukaryotic *Dnmt1* MTase gene whose structure is known, has two exceptions to the GT/AG rule, but none correspond to the one in the mouse gene. A comparison between mouse and human DNA sequences reveals a similar genomic organization since the location of 33 out of 38 introns is exactly conserved in both species. The five minor discrepancies are due either to the presence of a small deletion immediately adjacent to the splice site (introns 13 and 36) or to a small shift in the location of the splice site ($\leqslant 4$ nt). Besides these minor differences, the only exception to a perfectly conserved intron/exon organization is the presence of an additional intron in the human gene whose corresponding location would be in mouse exon 37.

It is interesting to note that the size and spatial arrangement of the exons of the N-terminal domain is quite different from that of the C-terminal exons (Figure 2(a)). For instance, C-terminal exons (29 to 39) are much more uniform in size than N-terminal exons (1 to 29). Furthermore, the average intron size is clearly smaller for C-terminal introns than for N-terminal ones (0.41 *versus* 1.00 kb). Accordingly, N and C-terminal domains have about a threefold difference in exon density (average number of exons per unit length). *Dnmt1* has been proposed to have evolved from the fusion between two different genes, one being a prokaryotic-like DNA cytosine methyltransferase and the other possibly a sequence-specific DNA binding protein (Bestor, 1990). This proposed different evolutionary origin of N and C terminus is also supported by the significant difference in exon density of these domains described above (Figure 2(a)). A closer examination of the structure of the *Dnmt1* gene revealed a further subdivision of the N-terminal domain. The distribution of exon sizes is uneven, since exons from the proximal part of the N-terminal region (ex 1-14) are much smaller than those from the remainder of *Dnmt1* (Figure 2(b)). Quantification of this observation indicates an average exon size of 81($\pm$14) bp (SEM) for the proximal region and 147($\pm$12) bp for the distal part of the N terminus. Statistical analysis shows that these values are significantly different from each other (student's *t*-test, $p < 0.001$). Therefore, on that basis, a bipartite structure can be recognized for the N-terminal domain (labelled as part A and B in all Figures). The human gene shows evidently the same exon distribution as the mouse gene, but the difference in average intron sizes between N and C terminus is even more pronounced due to the presence of two large introns. Additional genomic structures of *Dnmt1* will be necessary to determine whether this pattern is conserved in other species.

A bipartite origin of the N-terminal domain is also supported by comparison of the protein sequences of eukaryotic MTases. The sea urchin, chicken, *Xenopus* and human proteins were aligned against the mouse enzyme and their degree of similarity determined for each mouse exon. The results indicated that, in all cases, the region corresponding to mouse exons 4-16 exhibited a distinct lower similarity than the remainder of the protein (Figure 3). An uneven distribution of similarity between sea urchin and other MTases has been previously reported, although these authors failed to detect the similarity around exon 2 (Aniello *et al*., 1996). The presence of a ~90 aa deletion in *Xenopus* and chicken MTases (corresponding to mouse exons 7-11; Figure 3) lends support to the above finding that part of the N-terminal domain appears dispensable for catalytic activity and localization of the enzyme in these organisms. This could be explained by the presence of redundant nuclear and replication foci targeting signals (Cardoso & Leonhardt, 1999; Liu *et al*., 1998).

The proposal that eukaryotic MTases result from the fusion of an ancestral methyltransferase gene with a second gene was based on the presence of highly conserved sequence motifs in several prokaryotic MTases and in the C-terminal domain of the deduced mouse Dnmt1 open reading frame (Bestor, 1990). Additionally, the size of several prokaryotic enzymes is roughly similar to the size of mouse or human C-terminal domains and a repeating GlyLys dipeptide appears to be critically positioned at the boundary of N and C-terminal domains. Recent data as well as our results have come to question this proposal. First, Dnmt2, a eukaryotic MTase of roughly the size of prokaryotic MTases, does not have methylation activity *in vitro* despite harbouring all the conserved motifs of MTases

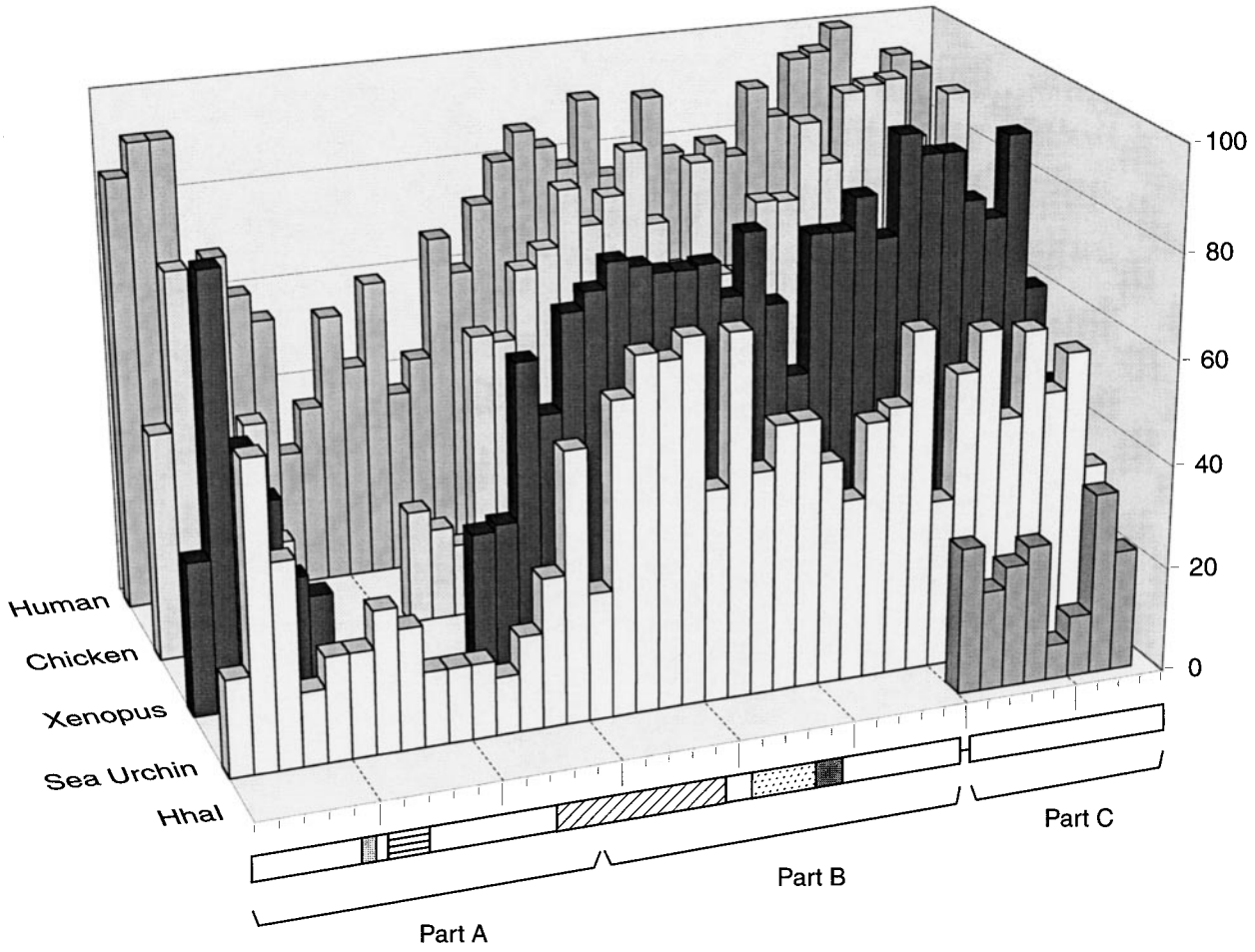

Figure 3. Comparison of prokaryotic and eukaryotic MTases. The amino acid sequences of mouse (GenBank accession number AF 162282), human (X63692), sea urchin (Z50183), chicken (D43920), *Xenopus* (D78638), and *Hha*I (J02677) MTases were first aligned using the Clustal method, then the mouse sequence was aligned pairwise to each of the other sequences using the Pearson-Lipman algorithm (Pearson & Lipman, 1988) of the MegAlign program (Lasergene). Identical amino acid residues are plotted per each mouse exon starting at ATG_3. Amino acids overlapping an exon boundary were attributed to the upstream exon. Deletions/insertions were not weighted in the similarity determinations. The location of mouse MTase functional domains is diagrammed below the graph as in Figure 2.

(Okano *et al.*, 1998b); second, functional rescue of Dnmt1 mutant embryonic stem cells was only achieved by a construct carrying the first four exons of Dnmt1 (Tucker *et al.*, 1996), indicating that the most upstream sequence of Dnmt1 is crucial for *in vivo* activity; and third, the newly isolated Dnmt3a and 3b have *de novo* methylation activity but do not have a GlyLys dipeptide repeat and their size is at least 850 aa (Okano *et al.*, 1998a; Xie *et al.*, 1999). Therefore, we propose instead that metazoan MTases arose *via* the fusion of three genes. Our proposal is based on functional data (deletion mapping of the MTase activity, Figure 1), on structural data (intron/exon size and exon density of the mouse and human genes, Figure 2 and Table 1), on sequence comparisons between different MTases (invertebrates, amphibians, avians and mammals, Figure 3), and on the reports mentioned above. The timing and order of the fusion event(s) leading to the tripartite structure of Dnmt1 MTases is difficult to ascertain. Since all known metazoan Dnmt1 proteins have roughly the same size it seems likely that part A and B were joined before the fusion to part C. The well conserved area of similarity around exon 2 points to an important role for exon 2 but could also indicate that the first few exons were initially linked to part B and were subsequently separated by an insertion event. As the presence of this highly conserved region does not have a drastic effect on the enzymatic properties *in vitro* (Pradhan

et al., 1997), specific mutations have to be further tested in transgenic animals to elucidate the role of this region in mammalian development.

Acknowledgements

We are grateful to Drs En Li and Rudolf Jaenisch for providing genomic clones and to Ingrid Grunewald for excellent technical assistance. This work was supported by grants from the Council for Tobacco Research and the Deutsche Forschungsgemeinschaft to H.L. and an NIH grant (GM 46127) to R.J.R.

References

Aniello, F., Locascio, A., Fucci, L., Geraci, G. & Branno, M. (1996). Isolation of cDNA clones encoding DNA methyltransferase of sea urchin *P. lividus*: expression during embryonic development. *Gene,* **178**, 57-61.

Bakin, A. V. & Curran, T. (1999). Role of DNA 5-methylcytosine transferase in cell transformation by fos. *Science,* **283**, 387-390.

Balganesh, T. S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K. & Trautner, T. A. (1987). Construction and use of chimeric SPR/phi 3T DNA methyltransferases in the definition of sequence recognizing enzyme regions. *EMBO J.* **6**, 3543-3549.

Bestor, T. H. (1990). DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Phil. Trans Roy Soc. ser. B, Biol. Sci.* **326**, 179-187.

Bestor, T. H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J.* **11**, 2611-2617.

Boyes, J. & Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CG binding protein. *Cell,* **64**, 1123-l134.

Cardoso, M. C. & Leonhardt, H. (1999). DNA methyltransferase is actively retained in the cytoplasm during early development. *J. Cell. Biol.* **147**, 25-32.

Cheng, X. (1995). Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 293-318.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W. & Roberts, R. J. (1993). Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell,* **74**, 299-307.

Chuang, L. S., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G. & Li, B. F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science,* **277**, 1996-2000.

Czank, A., Hauselmann, R., Page, A. W., Leonhardt, H., Bestor, T. H., Schaffner, W. & Hergersberg, M. (1991). Expression in mammalian cells of a cloned gene encoding murine DNA methyltransferase. *Gene,* **109**, 259-263.

Gaudet, F., Talbot, D., Leonhardt, H. & Jaenisch, R. (1998). A short DNA methyltransferase isoform restores methylation in vivo. *J. Biol. Chem.* **273**, 32725-32729.

Glickman, J. F., Pavlovich, J. G. & Reich, N. O. (1997). Peptide mapping of the murine DNA methyltransferase reveals a major phosphorylation site and the start of translation. *J. Biol. Chem.* **272**, 17851-17867.

Lauster, R., Trautner, T. A. & Noyer-Weidner, M. (1989). Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.* **206**, 305-312.

Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell,* **71**, 865-873.

Li, E., Bestor, T. H. & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell,* **69**, 915-926.

Li, E., Beard, C. & Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature,* **366**, 362-365.

Liu, Y., Oakeley, B. J., Sun, L. & Jost, J. P. (1998). Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. *Nucl. Acids Res.* **26**, 1038-1045.

Okano, M., Xie, S. & Li, E. (1998a). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases [letter]. *Nature Genet.* **19**, 219-220.

Okano, M., Xie, S. & Li, E. (1998b). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucl. Acids Res.* **26**, 2536-2540.

Panning, B. & Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes Dev.* **10**, 1991-2002.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA,* **85**, 2444-2448.

Postfai, J., Bhagwat, A. S., Posfai, G. & Roberts, R. J. (1989). Predictive motifs derived from cytosine methyltransferases. *Nucl. Acids Res.* **17**, 2421-2435.

Pradhan, S., Talbot, D., Sha, M., Benner, J., Hornstra, L., Li, E., Jaenisch, R. & Roberts, R. J. (1997). Baculovirus-mediated expression and characterization of the full-length murine DNA methyltransferase. *Nucl. Acids Res.* **25**, 4666-4673.

Senapathy, P., Shapiro, M. B. & Harris, N. L. (1990). Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol.* **183**, 252-278.

Shapiro, M. B. & Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* **15**, 7155-7174.

Trautner, T. A., Balganesh, T. S. & Pawlek, B. (1988). Chimeric multispecific DNA methyltransferases with novel combinations of target recognition. *Nucl. Acids Res.* **16**, 6649-6658.

Tucker, K. L., Talbot, D., Lee, M. A., Leonhardt, H. & Jaenisch, R. (1996). Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene. *Proc. Natl Acad. Sci. USA,* **93**, 12920-12925.

Vertino, P. M., Yen, R. W., Gao, J. & Baylin, S. B. (1996). De novo methylation of CpG island sequences in human fibroblasts overexpressing DNA (cytosine-5-)-methyltransferase. *Mol. Cell. Biol.* **16**, 4555-4565.

Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B. & Trautner, T. A. (1988).

Sequential order of target-recognizing domains in multispecific DNA-methyltransferases. *EMBO J.* **7**, 2601-2609.

Wu, J. C. & Santi, D. V. (1987). Kinetic and catalytic mechanism of HhaI methyltransferase. *J. Biol. Chem.* **262**, 4778-4786.

Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W. W., Okumura, K. & Li, E. (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene*, **236**, 87-95.

Zimmermann, C., Guhl, E. & Graessmann, A. (1997). Mouse DNA methyltransferase (MTase) deletion mutants that retain the catalytic domain display neither de novo nor maintenance methylation activity in vivo. *Biol. Chem.* **378**, 393-405.

Edited by J. Karn

(Received 6 September 1999; received in revised form 3 February 2000; accepted 9 February 2000)