CARYOLOGIA Vol. 57, no. 3: 305-311, 2004

The structure of insect DNA methyltransferase 2 (DNMT2) DNA binding domain is responsible for the non-CpG methylation in insect genomes

FEDERICA BORSATTI and MAURO MANDRIOLI*

Dipartimento di Biologia Animale, Università di Modena e Reggio Emilia, Via Campi 213/D, 41100 Modena, Italy. E-mail: mandriol@unimo.it.

Abstract — Alignment of vertebrate and invertebrate DNA methyltransferases 2 (Dnmt2) evidenced an overall evolutionary conservation of these proteins. However, alignment revealed a vertebrate-specific stretch of about forty amino acids located between the catalytic motif VIII and the target recognition domain that is constantly absent from insect homologues. The analysis of the three-dimensional structure of DNA methyltransferase indicated that this vertebrate specific Dnmt2 portion is located at the DNA binding domain whose structure is essential for the discrimination of the proper target sequence. Insect Dnmt2 enzymes are, therefore, devoid of a portion of the DNA binding domain suggesting that this structural change may alter the methylation target of insect Dnmt2 making cytosine methylation not limited to the vertebrate canonical CpG but extended to cytosine residues belonging to other dinucleotides.

Key words: DNA methyltransferase 2, DNA methyltransferase 2 DNA binding domain, insect genome methylation, non-CpG methylation.

INTRODUCTION

It is well known that a variable portion of cytosine residues is methylated in the form of 5-methylcytosine in eukaryotic genomes (BIRD 2002). DNA methylation has been associated with numerous functions depending on the model organism and the experimental context. In general, the presence of DNA methylation, in and around the promoter of genes, is associated with gene silencing (BIRD 2002). On a cellular level, loss of DNA methylation was shown to affect apoptosis in mice (JACKSON-GRUSBY *et al.* 2001) and *Xenopus* (STANCHEVA *et al.* 2001), X-chromosome inactivation and chromosomal stability in mice (PANNING AND JAENISCH 1996; GAUDET *et al.* 2003) and the overall chromosome organization in *Arabidopsis* (SOPPE *et al.* 2002).

In eukaryotes, DNA methylation is carried out by DNA methyltransferases that are grouped into different families (BESTOR 2000; LI 2002). Dnmt1 enzymes preferentially bind to hemi-methylated DNA and are responsible for the maintenance of DNA methylation after each round of replication (BESTOR *et al.* 1988; YODER *et al.* 1997; MARGOT *et al.* 2000). Dnmt2 proteins are similar to the prokaryotic methyltransferases but their function is still partially enigmatic since they seem unable to methylate DNA *in vitro*. Moreover, loss of function mutations of Dnmt2 gene did not showed any effect on mice genomic methylation patterns (OKANO *et al.* 1998) on the contrary of what happen with mutations in Dnmt1 that resulted in developmental defects (LI *et al.* 1992; LEI *et al.* 1996). The last methyltransferase family consists of Dnmt3a and Dnmt3b that are the main players involved in *de novo* methylation (OKANO *et al.* 1998). The third member of this family is Dnmt3L that shares some homologies with Dnmt3a and Dnmt3b and plays a central role in the establishment of maternal genomic imprinting even though it does not have *in vitro* catalytic activity (AAPOLA *et al.* 2001; DEPLUS 2002; HATA *et al.* 2002).

Up to date, the presence of 5-methylcytosine has been reported in several insect species belonging to various orders (FIELD *et al.* 2004). However, its role is still poorly understood and the available data demonstrates varying levels of methylation and different roles suggesting that DNA methylation could not play an evolutionary conserved function.

The presence of a discontinuity in the functional role of methylation from invertebrates to vertebrates

* Corresponding author: phone (+39) 59 2055544; fax (+39) 59 2055548; e-mail: mandriol@unimo.it.

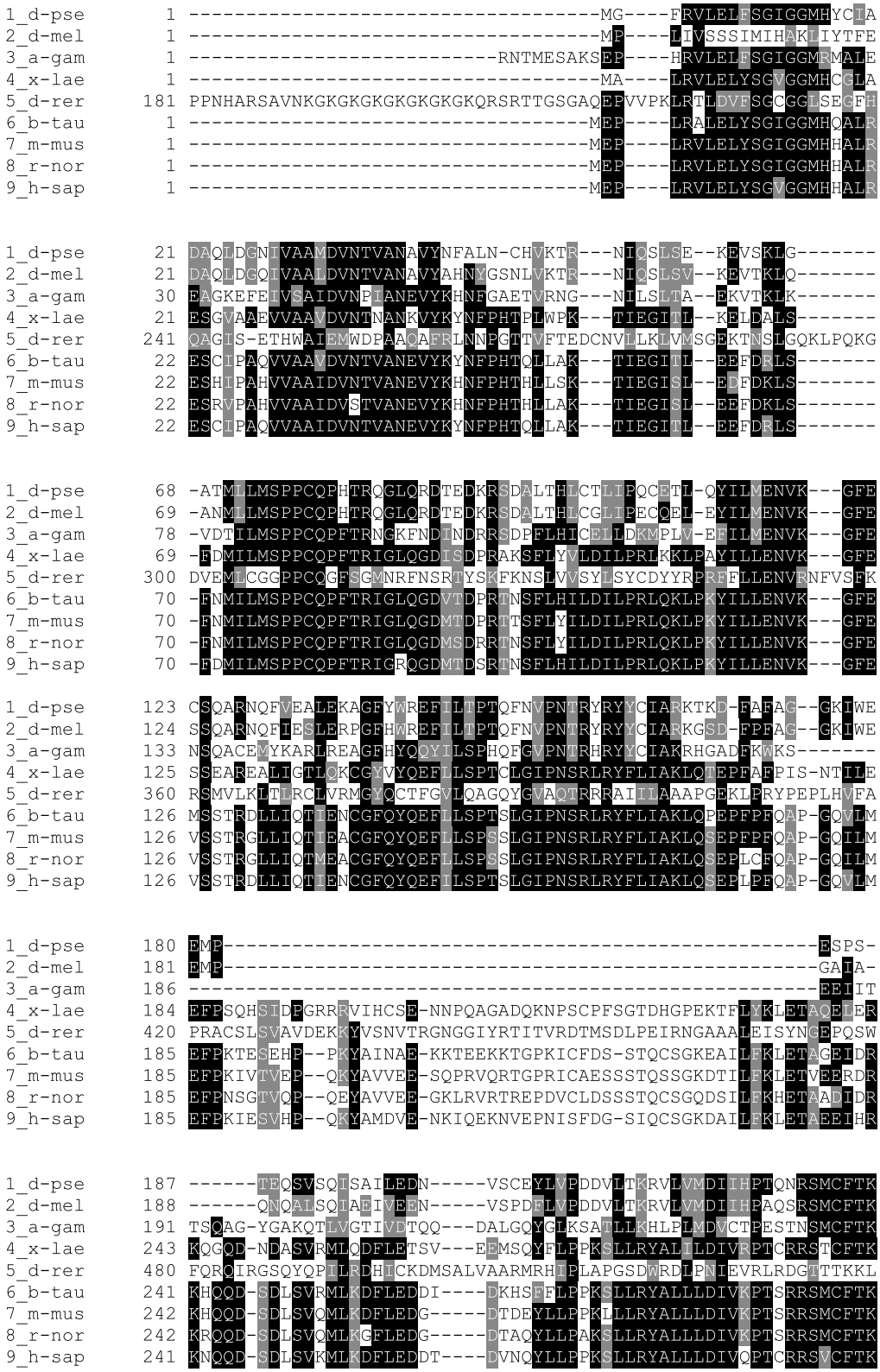
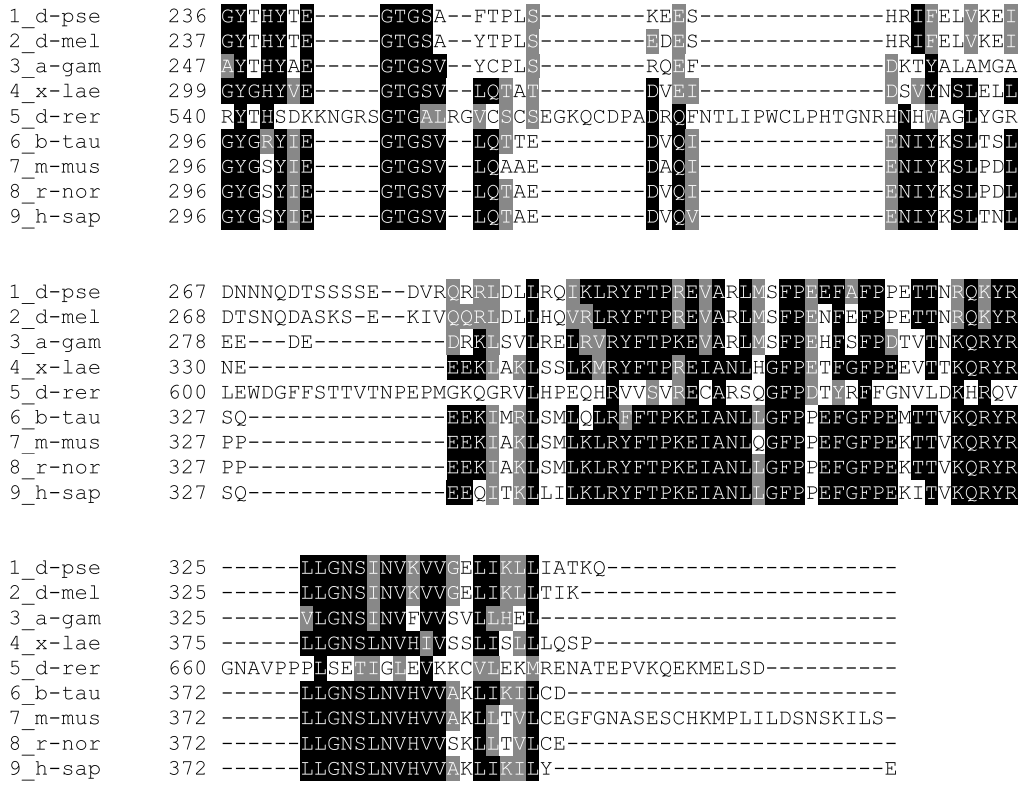


Fig. 1 — Alignment of eukaryote Dnmt2s evidenced that these proteins are evolutionary conserved even if some differences are present between insects and vertebrates. In particular, alignment revealed a vertebrate-specific stretch of about forty amino acids located between the catalytic motif VIII and the target recognition domain that is constantly absent from insect homologues.



is straightened by the fact that the Dnmt2 proteins represent the only candidate DNA methyltransferases in *Drosophila melanogaster*, *Drosophila pseudoobscura* and *A. gambiae* as deduced by the absence of other methyltransferase genes in their genome (Lyko 2001; Marhold *et al.* 2004).

Finally, a further difference is due to the fact that insect methylation is not limited to the CpG target: CpA, CpT and methylated doublets were, in fact, also reported in insects (Lyko *et al.* 2000; Kunert *et al.* 2003; Mandrioli and Volpi 2003; Marhold *et al.* 2004) with the peculiarity that, at least in *D. melanogaster*, DNA methylation is concentrated at the non-symmetrical CpA and CpT dinucleotides (Lyko *et al.* 2000).

The present paper analyse eukaryote Dnmt2 sequence and structure in order to verify if insect Dnmt2 possesses peculiarities useful to explain such a different pattern of methylation in insects in respect to vertebrates.

MATERIALS AND METHODS

Sequence retrieval form databases - Dnmt2 sequences were retrieved at NCBI using the ENTREZ software (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) that perform a search across all Entrez databases, whereas *D. pseudoobscura* Dnmt2 homologue was identified using the BLAST tool available at the *D. pseudoobscura* genome website (http://www.hgsc.bcm.tmc.edu/projects/drosophila/).

BLAST - The BLAST 2 software (Basic Local Alignment Search Tool) was used to search the NCBI databases (http://www.ncbi.nlm.nih.gov/BLAST/). In particular BLAST provided a method for rapid searching of Dnmt2 sequence in both nucleotide and protein databases. BLAST algorithm detects in fact local, as well as global, regions of similarity embedded in otherwise unrelated proteins (Altschul *et al.* 1997)

Sequence alignments by CLUSTALW and DNAstar - The CLUSTALW software at the European Bioinformatics Institute (EBI) (ww.ebi.ac.uk/clustalw) was used to look for biologically meaningful sequence alignments of evolutionary conserved DNA and protein sequences. The default alignment parameters were used.

CLUSTALW alignments were edited using BOXSHADE in order to better evidence the presence of conserved domains (http://www.ch.embnet.org/software/BOX_form.html).

Phylogenetic tree was reconstructed on the basis of the CLUSTALW alignments using the tree construction function of the DNAstar software package (DNAstar Inc, Madison, USA).

CD-Search at Conserved Domain Database (CDD) - The search for conserved domain in Dnmt2 was per-

formed using the CD-Search service at the Conserved Domain Database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=cdd) that employs the reverse position-specific BLAST algorithm. The CDD currently contains domains derived from two popular collections, Smart and Pfam, plus contributions from NCBI. The source databases also provide descriptions and links to citations. Since conserved domains correspond to compact structural units, CDs contain links to 3D-structure via Cn3D whenever possible. CD-Search has been run in parallel with protein BLAST searches (MARCHLER-BAUER *et al.* 2003).

RESULTS AND DISCUSSION

The typical eukaryotic DNA methyltransferase is about three times larger than its prokaryotic counterpart (MARGOT *et al.* 2003). By analogy with the prokaryotic enzymes, the C-terminal region has been referred to as the catalytic domain and the N-terminal region as the regulatory domain (BESTOR 2000; MARGOT *et al.* 2003). The N-terminal domain can interact with numerous proteins such as DMAP1, PCNA and Rb and it contains a DNA binding region, a cysteine-rich region, several Zn-binding domains and two regions responsible for the localization to replication foci (LEONHARDT *et al.* 1992; CHUANG *et al.* 1997; ROUNTREE *et al.* 2000; ROBERTSON *et al.* 2000). The lack of extensive homology between the N-terminal domains of maintenance (Dnmt3) and *de novo* methyltransferases (Dnmt1) points towards a possible functional difference of this domain (MARGOT *et al.* 2003).

D.melanogaster, *D. pseudoobscura* and *A.gambiae* genome project revealed that Dnmt2 proteins represent the only candidates DNA methyltransferases suggesting that this enzyme could be the unique responsible for DNA methylation in insect genomes (KUNERT *et al.* 2003; MARHOLD *et al.* 2004). The analysis of the methylation patterns revealed that several insect genomes contain methylated cytosine residues even if they are not concentrated into the CpG doublets, as usually found in vertebrates (FIELD *et al.* 2004). In particular, in *Drosophila* genome methylation resulted concentrated at CpA and CpT targets (LYKO 2001; MARHOLD *et al.* 2004), whereas in the lepidopteran *Mamestra brassicae* methylated cytosines were inserted predominantly into CpC doublets even if methylation was reported also in the CpG, CpA and CpT dinucleotides (MANDRIOLI and VOLPI 2003). This differential target of methylation can reflect the presence of different DNA methyltransferases in insect genomes or the existence of a differential target specificity of the same methylases in insects in respect to vertebrates. In order to answer to this question a comparison of vertebrate and insect Dnmt2 sequences has been performed.

Search in GenBank for Dnmt2 proteins unambiguously retrieved several DNA methyltransferase 2-like sequences in both vertebrates and invertebrates. In particular, homologues were found in the vertebrates *Homo sapiens* (AAC39764), *Mus musculus* (AAC53529), *Rattus norvegicus* (XP_214514),

Table 1 — Similarity and identity values resulting from the alignment of eukaryote Dnmt2s.

Percent Identity

Divergence

	1	2	3	4	5	6	7	8	9		
1		74.8	42.1	37.2	14.4	36.3	37.8	37.5	36.3	1	d-pse
2	27.5		42.4	35.4	13.0	34.8	35.1	35.1	35.4	2	d-mel
3	92.4	95.6		39.8	14.3	38.0	39.5	40.1	38.9	3	a-gam
4	98.7	114.3	101.1		14.4	62.7	61.4	62.1	62.7	4	x-lae
5	289.0	292.0	228.0	241.0		12.3	12.3	13.3	12.8	5	d-rer
6	104.0	109.0	106.6	47.5	236.0		80.1	78.3	85.4	6	b-tau
7	102.1	111.5	100.2	48.2	227.0	20.9		88.5	79.5	7	m-mus
8	102.9	112.4	99.2	48.0	236.0	23.3	12.5		79.3	8	r-nor
9	102.9	110.1	104.4	46.1	215.0	16.3	21.6	21.9		9	h-sap
	1	2	3	4	5	6	7	8	9		

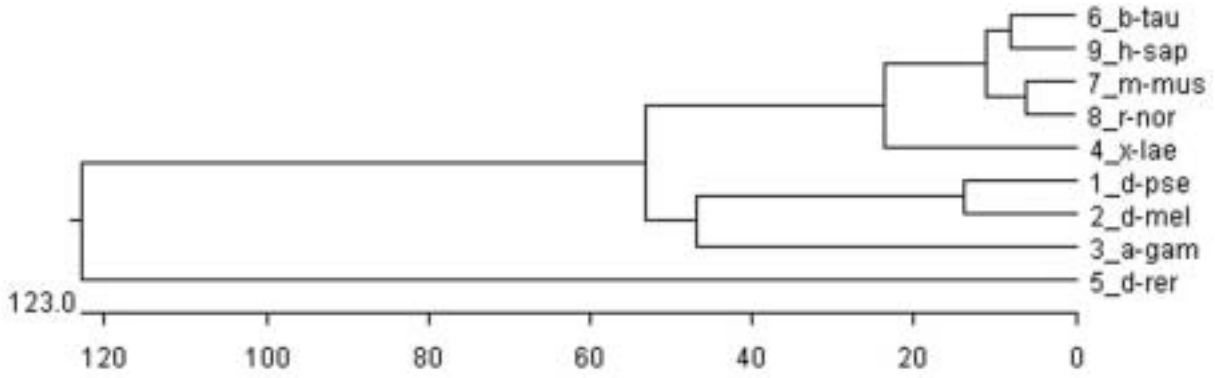


Fig. 2 — Phylogenetic tree reconstructed on the basis of the alignment of eukaryote Dnmt2s that has been used to confirm that Dnmt2 homologue have been really retrieved form sequence databases.

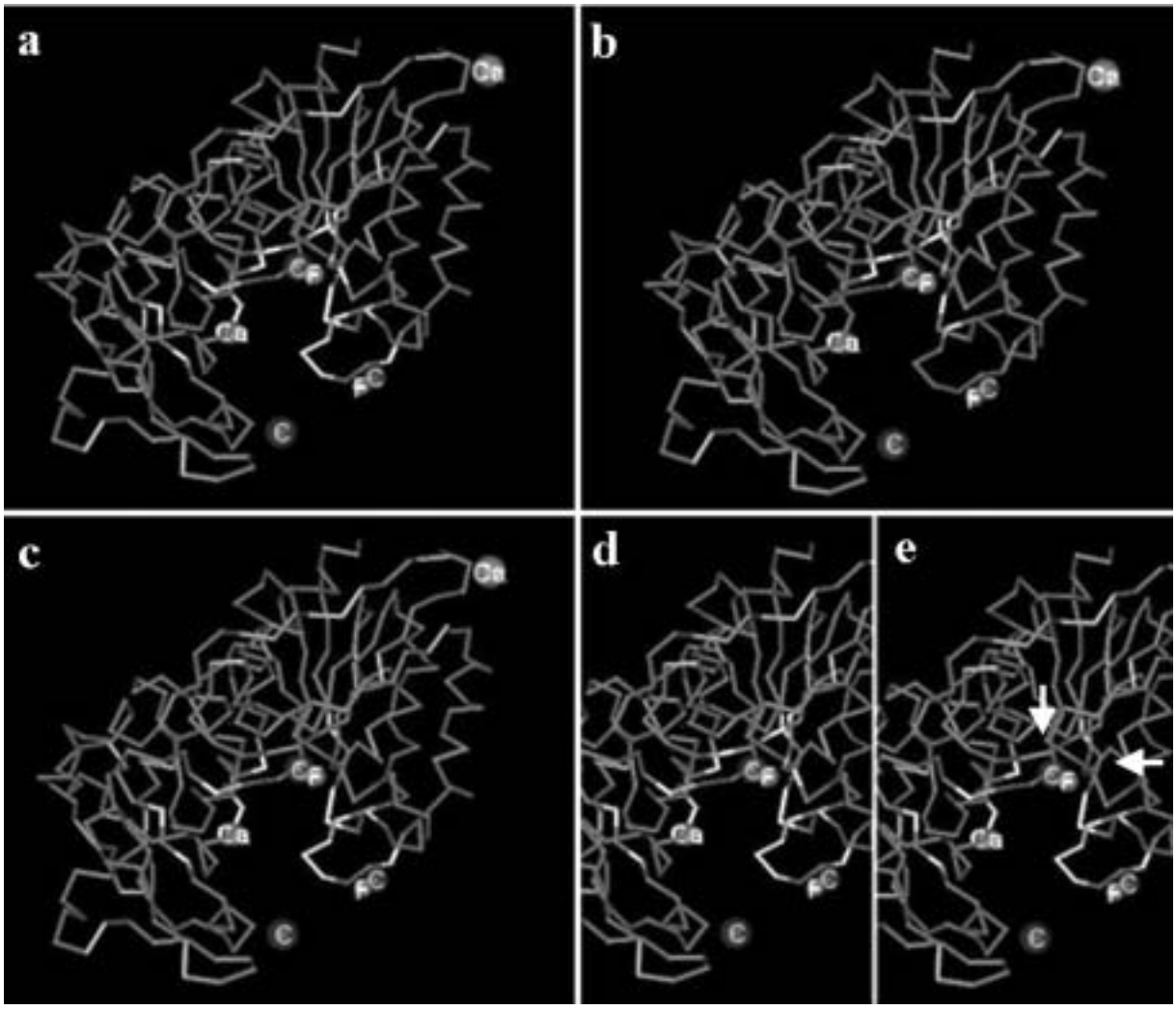


Fig. 3 — Three-dimensional structure of DNA methyltransferase with evidenced in yellow the DNA binding site (a, d) and the substrate interaction site (b). Alignment of Dnmt2 sequences revealed that insect Dnmt2 enzymes lack of a portion of these domains (c, e) suggesting that this differential structure could change the methylation target of insect Dnmt2 making cytosine methylation not limited to the vertebrate canonical CpG.

Xenopus laevis (AAH46854), *Bos taurus* (NP_861528) and *Danio rerio* (AAC69603) and in the invertebrates *D. melanogaster* (AAF03835) and *Anopheles gambiae* (XP_312975). Finally, a Dnmt2 homologue was retrieved in *D. pseudoobscura* in the sequence named Contig1859_Contig703.

Successively, a BLAST analysis has been performed in order to be sure that Dnmt2 homologues were really recovered. Finally, Dnmt2 similarity has been evaluated through the alignment of the retrieved sequences, which showed that Dnmt2 proteins were overall conserved and that they contain conserved catalytic motifs typical for (cytosine-5) DNA methyltransferases (KUMAR *et al.* 1994) (Figure 1). In particular, the highest observed similarities have been observed in Dnmt2 proteins from vertebrates with the exception of the putative *D. rerio* Dnmt2 that resulted poorly conserved suggesting that this sequence could not represent a real DNA methyltransferase 2 (Table 1). This hypothesis is confirmed by the phylogenetic tree reconstructed on the basis of the alignment since *D. rerio* Dnmt2 resulted as an independent branch in respect to the other vertebrate Dnmt2s (Figure 2).

A further analysis of Dnmt2 alignment revealed that all vertebrate Dnmt2s contain a stretch of about forty amino acids between the catalytic motif VIII and the target recognition domain that is absent in insect homologues. Analysis of literature data showed that this stretch of Dnmt2 is also absent in the DNA methyltransferase 2 of *Drosophila virilis*, *D. hydei* and *D. simulans* (MARHOLD *et al.* 2004) indicating that this portion of the Dnmt2 is peculiar of vertebrate and constantly absent in insect DNA methyltransferases 2. Considering that drosophilids and *A. gambiae* diverged about 250 million years ago, the results of alignment indicate that the structure of insect Dnmt2 is highly conserved suggesting that the reported difference in the structure of methyltransferase reflect a peculiar functionality of Dnmt2 in insects in respect to vertebrates. At this regards, we verified the location and the function of the vertebrate specific Dnmt2 stretch in order to identify a possible effect of its absence in insect homologues.

The search for conserved functional domain at the Conserved Domain Database (CDD) indicated the presence of a C-5 cytosine-specific DNA methylase domain in the Dnmt2 sequences. In particular, the amino acidic stretch identified using alignment data resulted involved both in the DNA binding site and in the substrate interaction site of the methyltransferase (Figure 3).

The absence of a portion of the DNA binding domain in insect Dnmt2 is very intriguing since this domain is essential for the discrimination of the proper methylation target sequence. These data, as a whole, suggest that this differential structure could change the methylation target of insect Dnmt2 making cytosine methylation not limited to the vertebrate canonical CpG but extended to cytosine residues belonging to other dinucleotides. This hypothesis is supported by the methylation mechanism originally proposed by SANTI *et al.* (1983) and modified by CHEN *et al.* (1991) and ERLANSON *et al.* (1993) indicating that the target recognition domain makes specific contacts with base edges in the major groove of DNA and is responsible for sequence discrimination (SANTI *et al.* 1983; CHEN *et al.* 1991; ERLANSON *et al.* 1993).

Finally, our proposal could explain the experimental data reported in *D. melanogaster* where it has been showed the presence of methylation at CpA and CpT dinucleotides despite the presence of a unique putative CpG methyltransferase (KUNERT *et al.* 2003).

REFERENCES

AAPOLA U., LYLE R., KROHN K., ANTONARAKIS S.E. and PETERSON, P., 2001 - *Isolation and initial characterization of the mouse Dnmt3l gene*. Cytogenet. Cell Genet. 92, 122-126.

ALTSCHUL S.F., MADDEN T.L., SCHÄFFER A.A., ZHANG J., ZHANG Z., MILLER W. and LIPMAN D.J., 1997 — *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res. 25, 3389-3402.

BESTOR T., LAUDANO A., MATTALIANO R. and INGRAM V., 1988. — *Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases.* J. Mol. Biol. 203, 971-983.

BESTOR T.H., 2000 — *The DNA methyltransferases of mammals*. Hum. Mol. Genet. 9, 2395-2402.

BIRD A., 2002 — *DNA methylation patterns and epigenetic memory*. Genes Dev. 16, 6-21.

CHEN L., MACMILLAN A.M., CHANG W., EZAZ-NIKPAY K., LANE W.S. and VERDINE G.L., 1991 — *Direct identification of the active site nucleophile in a DNA (cytosine-5)-methyltransferase.* Biochemistry 30, 11018–11025.

CHUANG, L.S., IAN, H.I., KOH, T.W., NG, H.H., XU, G. and LI, B.F., 1997 — *Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1.* Science 277, 1996-2000.

DEPLUS, R., BRENNER, C., BURGERS, W.A., PUTMANS, P., KOUZARIDES, T., DE LAUNOIT, Y. and FUKS, F., 2002 — *Dnmt3L is a transcriptional repressor that recruits histone deacetylase.* Nucleic Acids Res. 30, 3831-3838.

ERLANSON, D., DHEN, L. and VERDINE, G.L., 1993 — *Enzymatic DNA methylation through a locally un-*

paired intermediate. J. Am. Chem. Soc. 115, 12583–12584.

FIELD L.M., LYKO F., MANDRIOLI M. and PRANTERA G., 2004 — *DNA methylation in insects.* Insect Mol. Biol. 13, 109-217

GAUDET F., HODGSON J., EDEN A., JACKSON-GRUSBY L., DAUSMAN J., GRAY J.W., LEONHARDT H. and JAENISCH R., 2003 — *Induction of tumors in mice by genomic hypomethylation.* Science 300, 489-492.

HATA, K., OKANO, M., LEI, H. and LI, E., 2002 — *Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice*. Development 129, 1983-1993.

JACKSON-GRUSBY L., BEARD C., POSSEMATO R., TUDOR M., FAMBROUGH D., CSANKOVSZKI G., DAUSMAN J., LEE P., WILSON C., LANDER E. and JAENISCH, R., 2001 — *Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation.* Nat. Genet. 27, 31-39.

KUMAR S., CHENG X., KLIMASAUSKAS S., MI S., POSFAI J., ROBERTS R.J. and WILSON G.G., 1994 — *The DNA (cytosine-5) methyltransferases.* Nucleic Acids Res. 22, 1-10.

KUNERT N., MARHOLD J., STANKE J., STACH D. and LYKO F., 2003 — *A Dnmt2-like protein mediates DNA methylation in Drosophila*. Development 130, 5083-5090.

LEI H., OH S.P., OKANO M., JUTTERMANN R., GOSS K.A., JAENISCH R. and LI E., 1996 — *De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells*. Development 122, 3195-3205.

LEONHARDT H., PAGE A.W., WEIER H.U. and BESTOR T.H., 1992 — *A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei*. Cell 71, 865-873.

LI E., 2002 — *Chromatin modification and epigenetic reprogramming in mammalian development*. Nat. Rev. Genet. 3, 662-673.

LI E., BESTOR T.H. and JAENISCH R., 1992 — *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* Cell 69, 915-926.

LYKO F., 2001 — *DNA methylation learns to fly*. Trends Genet. 17, 169-172.

LYKO, F., RAMSAHOYE, B.H. and JAENISCH, R., 2000 - *DNA methylation in Drosophila melanogaster*. Nature 408, 538-540.

MANDRIOLI M. and VOLPI N., 2003 — *The genome of the lepidopteran Mamestra brassicae has a vertebrate-like content of methyl-cytosine*. Genetica 119, 187-191.

MARCHLER-BAUER A., ANDERSON J.B., DEWEESE-SCOTT C., FEDOROVA N.D., GEER L.Y., HE S., HURWITZ D.I., JACKSON J.D., JACOBS A.R., LANCZYCKI C.J., LIEBERT C.A., LIU C., MADEJ T., MARCHLER G.H., MAZUMDER R., NIKOLSKAYA A.N., PANCHENKO A.R., RAO B.S., SHOEMAKER B.A., SIMONYAN V., SONG J.S., THIESSEN P.A., VASUDEVAN S., WANG Y., YAMASHITA R.A., YIN J.J. and BRYANT S.H., 2003 — *CDD: a curated Entrez database of conserved domain alignments*. Nucleic Acids Res. 31, 383-387.

MARHOLD, J., ROTHE, N., PAULI, A., MUND, C., KUEHLE, K., BRUECKNER, B. and LYKO, F. (2004). — *Conservation of DNA methylation in dipteran insects*. Insect Mol. Biol. 13,117-123.

MARGOT J.B., AGUIRRE-ARTETA A.M., DI GIACCO B.V., PRADHAN S., ROBERTS R.J., CARDOSO M.C. AND LEONHARDT H., 2000 — *Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure*. J. Mol. Biol. 297, 293-300.

OKANO M., XIE S. and LI E., 1998 — *Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells*. Nucleic Acids Res. 26, 2536-2540.

PANNING B. and JAENISCH R., 1996 — *DNA hypomethylation can activate Xist expression and silence X-linked genes*. Genes Dev. 10, 1991-2002.

ROBERTSON K.D., AIT-SI-ALI S., YOKOCHI T., WADE P.A., JONES P.L. and WOLFFE A.P., 2000 — *DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters.* Nat. Genet. 25, 338-342.

ROUNTREE M.R., BACHMAN K.E. and BAYLIN S.B., 2000 — *DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci*. Nat. Genet. 25, 269-277.

SANTI D.V., GARRETT C.E. and BARR P.J., 1983 — *On the mechanism of inhibition of DNA-cytosine methyltransferase by cytosine analogs.* Cell 33, 9–10.

SOPPE W.J., JASENCAKOVA Z., HOUBEN A., KAKUTANI T., MEISTER A., HUANG M.S., JACOBSEN S.E., SCHUBERT I. and FRANZ P.F., 2002 — *DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis*. EMBO J. 21, 6549-6559.

STANCHEVA I., HENSEY C. and MEEHAN R.R., 2001 — *Loss of the maintenance methyltransferase, xDnmt1, induces apoptosis in Xenopus embryos*. EMBO J. 20, 1963-1973.

YODER J., WALSH C.P. and BESTOR T.H., 1997 — *Cytosine methylation and the ecology of intragenomic parasites*. Trends Genet. 13, 335-340.

Received July 7, 2004; accepted October 28, 2004