Published in final edited form as:

Nature. 2007 September 13; 449(7159): 248–251. doi:10.1038/nature06146.

Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation

Da Jia^{1,*}, Renata Z. Jurkowska^{2,*}, Xing Zhang¹, Albert Jeltsch², and Xiaodong Cheng¹

¹Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, Georgia 30322, USA

²Biochemistry Laboratory, School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany

Abstract

Genetic imprinting, found in flowering plants and placental mammals, uses DNA methylation to yield gene expression that is dependent on the parent of origin¹. DNA methyltransferase 3a (Dnmt3a) and its regulatory factor, DNA methyltransferase 3-like protein (Dnmt3L), are both required for the *de novo* DNA methylation of imprinted genes in mammalian germ cells. Dnmt3L interacts specifically with unmethylated lysine 4 of histone H3 through its amino-terminal PHD (plant homeodomain)-like domain². Here we show, with the use of crystallography, that the carboxy-terminal domain of human Dnmt3L interacts with the catalytic domain of Dnmt3a, demonstrating that Dnmt3L has dual functions of binding the unmethylated histone tail and activating DNA methyltransferase. The complexed C-terminal domains of Dnmt3a and Dnmt3L showed further dimerization through Dnmt3a–Dnmt3a interaction, forming a tetrameric complex with two active sites. Substitution of key non-catalytic residues at the Dnmt3a–Dnmt3L interface or the Dnmt3a–Dnmt3a interface eliminated enzymatic activity. Molecular modelling of a DNA–Dnmt3a dimer indicated that the two active sites are separated by about one DNA helical turn. The C-terminal domain of Dnmt3a oligomerizes on DNA to form a nucleoprotein filament. A periodicity in the activity of Dnmt3a on long DNA revealed a correlation of methylated CpG sites at distances of eight to ten base pairs, indicating that oligomerization leads Dnmt3a to methylate DNA in a periodic pattern. A similar periodicity is observed for the frequency of CpG sites in the differentially methylated regions of 12 maternally imprinted mouse genes. These results suggest a basis for the recognition and methylation of differentially methylated regions in imprinted genes, involving the detection of both nucleosome modification and CpG spacing.

In both flowering plants and placental mammals, DNA methylation has a central role in imprinting, but in neither case is it clear how imprinted genes are targeted for methylation. Imprinted genes in mammals are often associated with differentially methylated regions (DMRs)³, which show DNA methylation patterns that depend on the parent of origin. How the imprinting machinery recognizes DMRs is unknown. The Dnmt3 family includes three members: two *de novo* CpG methyltransferases, namely Dnmt3a and Dnmt3b (ref. 4), and an

Correspondence and requests for materials should be addressed to A.J. [E-mail: a.jeltsch@jacobs-university.de] or X.C. [E-mail: xcheng@emory.edu].

*These authors contributed equally to this work.

Author Information The X-ray structure of Dnmt3a–Dnmt3L C-terminal tetramer complex is deposited in the Protein Data Bank under ID code 2QRV. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

enzymatically inactive paralogue, Dnmt3L, that functions as a regulatory factor in germ cells⁵. Inactivating both Dnmt3a and Dnmt3b abolishes *de novo* methylation in mouse embryos⁴. Although Dnmt3b conditional germline knockout animals and their offspring show no apparent phenotype, the phenotype of a corresponding Dnmt3a conditional knockout⁶ is indistinguishable from that of Dnmt3L knockout mice^{5,7} with altered sex-specific *de novo* methylation of DNA sequences in male and female germ cells. These results indicate that Dnmt3a and Dnmt3L are both required for the methylation of most imprinted loci in germ cells.

We undertook structural and biochemical studies of a homogeneous complex of Dnmt3L and Dnmt3a2, generated by a co-expression and co-purification system (Supplementary Fig. 1). Dnmt3a2 is a shorter isoform of Dnmt3a that is the predominant form in embryonic stem cells and embryonal carcinoma cells and can also be detected in testis, ovary, thymus and spleen⁸. For crystallography, we focused on a stable complex of the C-terminal domains from both proteins (Dnmt3a-C and Dnmt3L-C) that retains substantial methyltransferase activity (Supplementary Fig. 1)⁹. The structure of the C-terminal complex was determined to a resolution of 2.9 Å in the presence of cofactor product *S*-adenosyl-L-homocysteine (AdoHcy) (Supplementary Table 1).

Both Dnmt3a-C and Dnmt3L-C have the classical fold characteristic for *S*-adenosyl-L-methionine (AdoMet)-dependent methyltransferases¹⁰, but AdoHcy was found only in Dnmt3a-C, not in Dnmt3L-C (Fig. 1a). This is consistent with Dnmt3a-C being the catalytic component of the complex, whereas Dnmt3L is inactive and unable to bind cofactor⁹. The overall complex is elongated (about $160 \times 60 \times 50$ Å³) with a butterfly shape (Fig. 1a, b). The complex contains two monomers of Dnmt3a-C and two of Dnmt3L-C, forming a tetramer (Dnmt3L–Dnmt3a–Dnmt3a–Dnmt3L) with two Dnmt3L–Dnmt3a interfaces (about 906 Å² interface area) and one Dnmt3a–Dnmt3a interface (about 944 Å²). The Dnmt3L–Dnmt3a interface of Dnmt3L also supports a Dnmt3L homodimer (Supplementary Fig. 2a, b). Dnmt3a2 might use the same interface to form a Dnmt3a2 homo-oligomer, as suggested by analytical size exclusion chromatography (a broad peak of about 500 kDa; Supplementary Fig. 2c, d). An F728A mutation of Dnmt3a2, which eliminates a hydrophobic interaction at the Dnmt3a–Dnmt3L interface, disrupted the Dnmt3a2 homo-oligomer to yield a roughly 150-kDa dimer (Supplementary Fig. 2c, d; the calculated mass of a Dnmt3a2 monomer is 78 kDa) and abolished methyltransferase activity (Fig. 1c; compare lanes 1 and 4). The equivalent mutant in Dnmt3L, F261A, lost its ability to form a homodimer (Supplementary Fig. 2a, b) and simultaneously its ability to stimulate wild-type Dnmt3a2 activity (Fig. 1c; compare lanes 1–3). At the Dnmt3a–Dnmt3a interface, an R881A mutation of Dnmt3a that eliminates a network of polar interactions (Fig. 1d) abolished the activity of Dnmt3a-C (ref. 11). These data indicate that both interfaces (Dnmt3a–Dnmt3L and Dnmt3a–Dnmt3a) are essential for catalysis. Dnmt3L might stabilize the conformation of the active-site loop of Dnmt3a (residues 704–725 before helix αD, containing the key nucleophile Cys 706), by means of interactions with the C-terminal portion of the active-site loop (Supplementary Fig. 3). These stabilizing interactions could explain the stimulation of Dnmt3a2 activity by Dnmt3L (Fig. 1c; lanes 1 and 2)^{9,12–14}, as well as the linked loss of the Dnmt3a–Dnmt3L interface and of catalytic activity in Dnmt3a2 F728A (Fig. 1c, lanes 4–6).

Among the known active DNA methyltransferases, Dnmt3a and Dnmt3b have the smallest DNA-binding domain (absent from Dnmt3L; Supplementary Fig. 4). This domain includes about 50 residues in Dnmt3a/Dnmt3b in comparison with, for example, about 85 residues in the bacterial GCGC methyltransferase *M.HhaI* (ref. 15). However, dimerization by means of the Dnmt3a–Dnmt3a interface brings two active sites together and effectively doubles the DNA-binding surface. We superimposed the Dnmt3a structure on that of *M.HhaI* complexed with a short oligonucleotide¹⁵. This yielded a model of a Dnmt3a–DNA complex with a short

DNA duplex bound to each active site (Fig. 2a). The two DNA segments can be connected easily to form a contiguous DNA, such that the two active sites are located in the major groove about 40 Å apart (Fig. 2b). This model indicates that dimeric Dnmt3a could methylate two CpGs separated by one helical turn in one binding event.

Electrophoretic mobility-shift assays revealed cooperative multimerization on DNA of Dnmt3a-C alone or of the Dnmt3a-C–Dnmt3L-C complex, with each monomer of Dnmt3a-C binding to about 12 base pairs (Fig. 2c and Supplementary Fig. 5a). Gel-filtration experiments, using short oligonucleotides of different lengths, confirmed the oligomerization of Dnmt3a-C on DNA with one monomer bound for each roughly nine base pairs (Supplementary Fig. 5b). Oligonucleotides containing a single CpG site are substrates for the Dnmt3a–Dnmt3L C-terminal complex, but at least eight base pairs on each side of the CpG are required for substantial activity, which is consistent with a possible requirement for DNA contact by both Dnmt3a molecules (Supplementary Fig. 1d).

On longer DNAs, we tested the possibility that the Dnmt3a dimer, in one binding event, methylates two CpG sites separated by a helical turn. Two different DNA fragments were methylated *in vitro* by Dnmt3a-C. Methylation was analysed by bisulphite conversion, followed by cloning and sequencing, of 119 clones in total. At an overall methylation level of 22–26% there was a periodic fluctuation of the relative methylation at the various CpG sites (Fig. 3a and Supplementary Fig. 6). To determine whether there is a correlation between the methylation states of any two CpG sites at a given distance from one another, the autocorrelation of the methylation states was calculated for all pairs of CpGs in each individual clone. We observed a highly significant correlation of methylation status at distances of eight to ten base pairs between two CpG sites (Fig. 3b). These experiments were performed under conditions in which the DNA was saturated with the enzyme. As a result of its large interface with DNA, the enzyme oligomer or polymer cannot move along the DNA, in agreement to the observation that Dnmt3a-C methylates DNA in a non-processive manner¹⁶. The enzyme oligomer on the DNA presents the active sites in a regular spacing, which leads to a correlated methylation of CpG sites. In contrast, CpG sites positioned between the active sites are not readily available for methylation, which causes a correlation of absence of methylation that has the same period (Supplementary Fig. 6c).

The similarity of defects observed in the Dnmt3a conditional germline knockout and the Dnmt3L-null mutants indicates that both Dnmt3a and Dnmt3L are required for the methylation of DMRs in imprinted genes^{5–7}. We studied the distribution of CpG sites among 12 known maternally imprinted DMRs¹⁷ that are methylated in wild-type embryos and responsible for their germline targeting, including the three (*Snrpn*, *Igf2r* and *Peg1*) that were shown experimentally to be unmethylated in affected embryos⁶. The frequencies of the distances between CpG sites peak periodically, with an average interval of 9.5 base pairs (Fig. 4a; the λ DNA fragment and the CpG island used as methylation substrates in Fig. 3 do not contain such pattern). The periodic occurrence of CpG sites 9.5 base pairs apart on average (examples are shown in Fig. 4d) makes these DNA sequences an ideal substrate for the activity of two active sites in Dnmt3a, indicating that the CpGs on maternally imprinted DMRs could be methylated simultaneously by a Dnmt3a–Dnmt3L tetramer or oligomer (as shown in Fig. 2). The periodicity of the distribution of pairwise distances of CpGs was further analysed by calculating an autocorrelation function of the frequencies, which underscored the significance of the observation (Fig. 4b). As controls, ten CpG islands randomly taken from promoter regions of genes on human chromosome 21 (Fig. 4c and Supplementary Fig. 7) did not show any correlation in the positioning of CG sites.

In contrast with the maternally imprinted DMRs, three paternally imprinted DMRs (*H19*, *Dlk1–Gtl2* and *Rasgrf1*; Supplementary Fig. 8a) did not show such periodicity; only *Rasgrf1*

showed a weak periodic pattern similar to maternally imprinted DMRs. The three paternal DMRs showed different methylation levels in impaired spermatogenesis: first, the DMR of *Rasgrf1* is normally methylated in both the Dnmt3a conditional mutant and Dnmt3L^{-/-} as well as in wild-type males⁶, but unmethylated in Dnmt3L knockout animals in a different study¹⁸; second, the *H19* DMR is unmethylated in both mutants⁶ but showed mosaic methylation in two different studies^{7,18}; and third, the DMR at *Dlk1-Gtl2* was methylated in Dnmt3L^{-/-} animals but not in a Dnmt3a conditional mutant. It is possible that additional factors (such as RNA¹⁹) are involved in establishing paternal imprints at specific loci, including paternally imprinted retrotransposons (LINE-1 and IAP; Supplementary Fig. 8b). Dnmt3L and Dnmt3a are at present the only factors known to be required for establishing maternal imprints in germ cells. We conclude that the periodic arrangement of CpGs in maternally imprinted DMRs constitutes an environment that is favourable for methylation by the Dnmt3a–Dnmt3L tetramer, which is consistent with the tetramers having two active sites with similar spacing and might contribute to their preferential methylation in the female germ line. Comprehensive genome-wide studies will be required in the future to determine whether all CpG islands with a periodicity of eight to ten base pairs are maternally imprinted.

Finally, histone methylation has a function in epigenetic signalling in addition to DNA methylation. There have been reports of an inverse relationship between methylation of histone H3 lysine 4 (H3K4) and allele-specific DNA methylation at DMRs; that is, a lack of H3K4 methylation at the methylated allele and the presence of H3K4 methylation at the unmethylated allele^{20–23}. The Dnmt3L–Dnmt3a complex structure presented here indicates a novel mechanism by which an absence of H3K4 methylation is recognized by the PHD (plant homeodomain)-like domain of Dnmt3L (ref. 2), whereas its C-terminal methyltransferase-like domain brings in the active DNA methyltransferase Dnmt3a to establish a heritable DNA methylation pattern. Hence, H3K4 methylation could protect unmethylated DMRs from DNA methylation by the Dnmt3a–Dnmt3L complex.

METHODS SUMMARY

Co-expression and co-purification of Dnmt3 proteins

Co-expression of full-length Dnmt3a2 (residues 220–908 of Dnmt3a; National Center for Biotechnology Information (NCBI) accession number o88508) and Dnmt3L (NCBI accession number AAH83147) was achieved by engineering two expression cassettes in one plasmid. Co-expression of Dnmt3a-C (residues 623–908) and Dnmt3L-C (residues 160–386) was achieved by the sequential transformation of two plasmids (pXC528 and pXC510) into *Escherichia coli* strain BL21 (DE3). Dnmt3a2 or Dnmt3a-C contained an N-terminal His₆ tag, and Dnmt3L or Dnmt3L-C was a glutathione S-transferase (GST) fusion protein. The protein complex was purified with the use of three-column chromatography (GSTrap HP column, Ni²⁺-chelating column and Superdex 75; Amersham-Pharmacia). The GST tag was cleaved by thrombin.

Crystallography

By combining the Se-anomalous diffraction data (Supplementary Table 1) and the molecular replacement with the use of the C-terminal domain of Dnmt3L homodimer as the initial search model, the structure of Dnmt3a-C–Dnmt3L-C complex was solved.

Sequence analyses

The analyses of the periodicity of CpG positioning and in Dnmt3a activity were performed with two in-house programs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. R. Horton for assistance with X-ray diffraction data collection; Z. Yang for assistance in solving the structure; H. Sasaki and R. Hirasawa for providing DMR sequences; S. Devine and R. Mills for help with DMR sequence analysis; A. Pingoud for providing an R.EcoRV expression clone used for calibration of the EMSA experiments; R. M. Blumenthal for critical editing of the manuscript; and E. Bernstein and R. E. Collins for comments on the manuscript. This work was supported by grants from the National Institutes of Health to X.C. and grants from the Deutsche Forschungsgemeinschaft and BMBF (Biofuture programme) to A.J.

References

1. Feil R, Berger F. Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet* 2007;23:192–199. [PubMed: 17316885]
2. Ooi SKT, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* 2007;448:714–717. [PubMed: 17687327]
3. Ferguson-Smith AC, Surani MA. Imprinting and the epigenetic asymmetry between parental genomes. *Science* 2001;293:1086–1089. [PubMed: 11498578]
4. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* 1999;99:247–257. [PubMed: 10555141]
5. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. *Science* 2001;294:2536–2539. [PubMed: 11719692]
6. Kaneda M, et al. Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 2004;429:900–903. [PubMed: 15215868]
7. Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 2004;431:96–99. [PubMed: 15318244]
8. Chen T, Ueda Y, Xie S, Li E. A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active *de novo* methylation. *J Biol Chem* 2002;277:38746–38754. [PubMed: 12138111]
9. Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *J Biol Chem* 2005;280:13341–13348. [PubMed: 15671018]
10. Schubert HL, Blumenthal RM, Cheng X. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci* 2003;28:329–335. [PubMed: 12826405]
11. Gowher H, et al. Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. *J Mol Biol* 2006;357:928–941. [PubMed: 16472822]
12. Chedin F, Lieber MR, Hsieh CL. The DNA methyltransferase-like protein DNMT3L stimulates *de novo* methylation by Dnmt3a. *Proc Natl Acad Sci USA* 2002;99:16916–16921. [PubMed: 12481029]
13. Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem* 2004;279:27816–27823. [PubMed: 15105426]
14. Karet MS, Botello ZM, Ennis JJ, Chou C, Chedin F. Reconstitution and mechanism of the stimulation of *de novo* methylation by human DNMT3L. *J Biol Chem* 2006;281:25893–25902. [PubMed: 16829525]
15. Klimasauskas S, Kumar S, Roberts RJ, Cheng X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 1994;76:357–369. [PubMed: 8293469]
16. Gowher H, Jeltsch A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. *J Biol Chem* 2002;277:20409–20414. [PubMed: 11919202]
17. Kobayashi H, et al. Bisulfite sequencing and dinucleotide content analysis of 15 imprinted mouse differentially methylated regions (DMRs): paternally methylated DMRs contain less CpGs than maternally methylated DMRs. *Cytogenet. Genome Res* 2006;113:130–137.

18. Webster KE, et al. Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc Natl Acad Sci USA* 2005;102:4068–4073. [PubMed: 15753313]
19. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* 2007;316:744–747. [PubMed: 17446352]
20. Delaval K, et al. Differential histone modifications mark mouse imprinting control regions during spermatogenesis. *EMBO J* 2007;26:720–729. [PubMed: 17255950]
21. Fournier C, et al. Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. *EMBO J* 2002;21:6560–6570. [PubMed: 12456662]
22. Yamasaki Y, et al. Neuron-specific relaxation of Igf2r imprinting is associated with neuron-specific histone modifications and lack of its antisense transcript Air. *Hum Mol Genet* 2005;14:2511–2520. [PubMed: 16037066]
23. Vu TH, Li T, Hoffman AR. Promoter-restricted histone code, not the differentially methylated DNA regions or antisense transcripts, marks the imprinting status of IGF2R in human and mouse. *Hum Mol Genet* 2004;13:2233–2245. [PubMed: 15294879]

Figure 1. Structure of the Dnmt3a-Dnmt3L (3a-3L) C-terminal domain complex

a, Side view of the tetramer Dnmt3L (blue)-Dnmt3a (green)-Dnmt3a (green)-Dnmt3L (blue), with the bound AdoHcy shown as a stick model. The inset shows the two pairs of phenylalanine residues (F728 and F768 of Dnmt3a, and F261 and F301 of Dnmt3L) in the centre of the Dnmt3a-Dnmt3L interface. **b**, Top view showing the Dnmt3a-Dnmt3a interface with two pairs of salt bridges formed between R881 and D872 (enlarged). **c**, Activities of Dnmt3a2 and its point mutant F728A in the presence and absence of Dnmt3L. Error bars represent s.d. calculated from two independent experiments. **d**, A network of polar interactions between two Dnmt3a molecules (coloured blue and purple) involves R881-D872, D872-H869, H869-E851 and E851-R881.

Figure 2. Model of Dnmt3a-Dnmt3L tetramer with DNA

a, Surface representation of the Dnmt3a-Dnmt3L tetramer, with two short DNA molecules adopted by superimposition of the *HhaI*-DNA complex structure¹⁵ onto individual Dnmt3a-C. **b**, Two views of the Dnmt3a-Dnmt3L tetramer with one contiguous curved DNA molecule, approximately 25 base pairs in length, covering two active sites. **c**, Cooperative formation of large protein-DNA complexes (red arrow) after incubation of a 146-base-pair DNA with increasing concentrations of Dnmt3a-C or Dnmt3a-C-Dnmt3L-C complex. The lower band (black arrow) corresponds to one Dnmt3L-C bound to the DNA as a result of the presence of some free Dnmt3L-C in the mixture (Supplementary Fig. 5a).

Figure 3. Periodicity of Dnmt3a-C activity on long DNA substrates

a, Two DNA fragments (from λ phage (top; 520 base pairs, 40 CpGs) and the CpG island upstream of the human *SUHWI* gene (bottom; 420 base pairs, 56 CpGs)) were used for methylation. A total of 119 clones (64 from λ phage, 55 from the human CpG island) were sequenced to determine their methylation pattern. **b**, Autocorrelation of CpG methylation. By averaging two substrates, the P values for the correlation factors observed for distances of eight to ten base pairs (arrow) are in the range $(3-7) \times 10^{-4}$. This periodicity is detectable for correlation of methylation of individual CpG sites as well as for correlation of the absence of methylation at two sites (Supplementary Fig. 6).

Figure 4. Periodicity of CpG sites in maternally imprinted DMRs¹⁷

a, Frequencies of CpG pairs at a given distance with respect to all pairs, for 11 maternally imprinted DMRs. The slope of the graph in the inset is 9.5. **b**, The autocorrelation of peaks for 11 maternally imprinted DMRs. One sequence (*Impact*) was excluded from the plot because it contains repeats (see **d**), with each CpG site separated from the second subsequent one by ten base pairs; thus, it follows the same pattern seen by the other 11 DMRs, bringing the total number to 12 DMRs. **c**, Autocorrelation of peaks for ten CpG islands randomly selected from human chromosome 21. **d**, A region of DMR sequences of the *U2af1-rs1* and *Impact* genes¹⁷.