Published in final edited form as:

Nat Struct Biol. 2002 March ; 9(3): 217–224. doi:10.1038/nsb759.

The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds

Chen Qiu^{1,2}, Ken Sawada¹, Xing Zhang¹, and Xiaodong Cheng¹

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

²Department of Chemistry, Emory University, 1510 Clifton Road, Atlanta, Georgia 30322, USA

Abstract

The PWWP domain is a weakly conserved sequence motif found in >60 eukaryotic proteins, including the mammalian DNA methyltransferases Dnmt3a and Dnmt3b. These proteins often contain other chromatin-association domains. A 135-residue PWWP domain from mouse Dnmt3b (amino acids 223–357) has been structurally characterized at 1.8 Å resolution. The N-terminal half of this domain resembles a barrel-like five-stranded structure, whereas the C-terminal half contains a five-helix bundle. The two halves are packed against each other to form a single structural module that exhibits a prominent positive electrostatic potential. The PWWP domain alone binds DNA *in vitro*, probably through its basic surface. We also show that recombinant Dnmt3b2 protein (a splice variant of Dnmt3b) and two N-terminal deletion mutants (218 and 369) have approximately equal methyl transfer activity on unmethylated and hemimethylated CpG-containing oligonucleotides. The 218 protein, which includes the PWWP domain, binds DNA more strongly than 369, which lacks the PWWP domain.

Genomic patterns of cytosine methylation at the C5 position play key roles in the organization and control of mammalian chromatin (reviewed in refs 1–3). Mammalian DNA methyltransferases (MTases) all contain a C-terminal catalytic region that is structurally and functionally homologous to bacterial MTases. The eukaryotic DNA MTases have been grouped into three families based on (i) distinctive properties of their N-terminal regions and (ii) intriguing differences in DNA substrate preferences. The first eukaryotic DNA MTase, Dnmt1, contains a large N-terminal regulatory region of ~1,000 amino acids and shows a preference for hemimethylated DNA *in vitro*^{4–6}. Dnmt2, a relatively small protein, contains only the MTase domain, and its function is still unclear⁷. Recombinant Dnmt3 acts on unmethylated and hemimethylated DNA at equal rates *in vitro*^{8,9}.

The task of dissecting the functional roles of the N-terminal region of the different MTase families is just beginning. These regions vary widely in size and are probably responsible for the diverse biological functions of eukaryotic DNA MTases. These functions include the

normal development of animals^{10,11} and plants^{12–14} *via* effects on gene repression¹⁵, X-chromosome inactivation¹⁶, genomic imprinting¹⁷ and replication timing¹⁸.

The Dnmt3 family consists of at least two members, Dnmt3a and Dnmt3b^{8,19}, which are both essential. Inactivation of both Dnmt3a and Dnmt3b abolishes *de novo* methylation in mouse embryos. Dnmt3a knockout mice died four weeks after birth, whereas disrupting Dnmt3b caused embryonic lethality. Dnmt3a and Dnmt3b exhibit nonoverlapping functions in development, with Dnmt3b specifically required for methylation of centromeric minor satellite repeats¹¹. Human ICF (immunodeficiency, centromere instability and facial anomalies) syndrome, which is accompanied by hypomethylation of classical satellite DNA, has been linked to mutations in the C-terminal catalytic domain of DNMT3B²⁰. In addition, DNMT3B plays a role in different forms of cancer^{19,21–23}; its expression level is correlated with the methylation stage of the p15^{INK4B} tumor suppressor gene in acute myelogenous leukemia²⁴. As with Dnmt1, Dnmt3a and Dnmt3b repress transcription in a methylation-independent manner²⁵, primarily through association with the histone deacetylase HDAC1 (ref. 26) using its Cys-rich domain, which contains six CXXC repeats.

No previous structural analysis of the N-terminal regions of any mammalian DNA MTases has been done. Here we present the structure of an amino-proximal PWWP domain of mouse Dnmt3b2 (residues 223–357). Dnmt3b2, an isoform resulting from alternative splicing of the Dnmt3b gene^{8,19}, is active *in vitro*²⁷. Although the importance of individual isoforms is unclear, all of the isoforms (3b1–3b5)²¹ are the same in the 223–357 range — this is, all isoforms have the PWWP domain. We show here that this region of Dnmt3b comprises a globular domain with a five-stranded β -barrel followed by a five-helix bundle. We find that structural orthologs of this 135-amino acid domain are present in >60 eukaryotic proteins, most of which contain known chromatin-association domains. For Dnmt3b2, the PWWP domain is shown to mediate DNA binding *in vitro*.

Dnmt3 contains a PWWP domain

The sequences of Dnmt3 orthologs have been determined from human and mouse. They all contain an N-terminal variable region (~280 amino acids in Dnmt3a and ~220 amino acids in Dnmt3b), followed by three stretches of conserved regions: the PWWP motif, six repeats of a CXXC motif and a set of 10 motifs conserved among DNA MTase catalytic regions (Fig. 1a). The PWWP motif²⁸ was first identified in a gene family related to the hepatoma-derived growth factor (HDGF)²⁹ and WHSC1 genes (Wolf-Hirschhorn Syndrome Candidate)³⁰. The corresponding sequence in Dnmt3 is SWWP (Fig. 1b); only the last two positions of the motif (WP) are invariant among the large number of sequences examined (see below). However, for reasons of consistency, we retained the PWWP nomenclature²⁸.

To define the Dnmt3b PWWP-containing domain for crystallography, we performed limited proteolysis (see Methods) on a truncated Dnmt3b2 protein named 218 (consisting of amino acids 219–839). Trypsin digestion of 218 resulted in a stable fragment, which mass spectrometry revealed to comprise amino acids 219–365; this region includes the entirety of the conserved PWWP motif.

Dnmt3b PWWP domain structure

The mouse Dnmt3b2 PWWP domain (residues 219–365) was overexpressed as an N-terminal His₆-fusion protein in *Escherichia coli* grown in the presence of selenomethionine (SeMet) (see Methods). Crystals of the His-tagged, SeMet-labeled Dnmt3b2 PWWP domain formed in the P6₅22 space group. Electron density maps were calculated using three-wavelength selenium data from multiwavelength anomalous diffraction (Table 1). A model consisting of residues 223–357 of Dnmt3b2 was built and refined to 1.8 Å resolution, with a crystallographic R-factor of 0.19 and R_{free} value of 0.23.

The structure of Dnmt3b can be divided into β-barrel and helical-bundle substructures (Fig. 2a). The N-terminal β-barrel structure consists of five antiparallel strands (β1–β5), with two short 3₁₀-helices (αA and αB) between strands β2 and β3, and strands β4 and β5, respectively. The C-terminal helical bundle includes five helices (αC–αG), with a single strand (β6) between helices αE and αF. The SWWP motif is part of the β-barrel but is located at the interface of the two substructures. Ser 244 of the PWWP domain is immediately before strand β2, whereas the Trp doublet and the Pro are part of β2. The loop following strand β6 has a seven-residue insertion in Dnmt3a (Fig. 1b), which is the only major difference within the PWWP domain between Dnmt3a and Dnmt3b.

The PWWP domain contains 19 basic residues (5 Arg and 14 Lys residues) scattered throughout the entire 135 amino acids, with a calculated isoelectric point of 9.6. A surface area of $\sim 45 \times 32 \text{ Å}^2$, enriched with basic residues (Fig. 2b), is surrounded $\sim 270^\circ$ by the positive electrostatic potential within the contours of 2 KT e^{-1} (Fig. 2c). This characteristic surface feature suggests that the positively charged surface of the PWWP domain may interact with negatively charged molecules such as DNA. The longest length ($\sim 45 \text{ Å}$) of the domain and the dimensions of positive potential could span ~ 12 base pairs of DNA. Opposite to this positive surface, there are two localized negatively charged patches — one formed by the negative charges from the β-barrel (mostly from the N-terminus) and the other from the helical bundle (helix αF) (Fig. 2c).

The PWWP domain of Dnmt3b binds to DNA

Based on analysis of the structure, we characterized the functional role of the PWWP domain in Dnmt3b2. We compared DNA binding of two Dnmt3b2 truncations with and without the PWWP domain — 218 (deleting the N-terminal 218 residues) and 369 (deleting the N-terminal 369 residues, including the PWWP domain) — using a nitrocellulose filter-binding assay. The PWWP-containing 218 or the PWWP domain alone binds DNA better than PWWP-lacking 369 (Fig. 3a). To confirm the nitrocellulose binding data, a gel mobility shift assay shows that the PWWP domain alone binds DNA (data not shown). The calculated dissociation constant, K_d , of the protein–DNA complex is ~ 8 and 230 nM for 218 and the PWWP domain alone, respectively (Fig. 3b). As a negative control, the CXXC domain of Dnmt3b2 does not bind DNA at all. This domain shares sequence similarity with the ATRX protein and associates with other proteins, the histone deacetylase HDAC1 and transcriptional repressor RP58 (refs 25,26).

We further investigated methyl transfer activity of various Dnmt3b2 fragments (Fig. 3c). Wild type and two large protein fragments (218 and 369) are equally active on unmethylated (CpG) and hemimethylated (hemi-CpG) oligonucleotide substrates, demonstrating that removing the first 369 amino acids does not affect *in vitro* CpG-methylating activity on naked oligonucleotide DNA. However, the PWWP domain thought to be involved in heterochromatin association²⁵ could conceivably affect activity on chromatin *in vivo*. Indeed, mutations within the corresponding PWWP domain of Dnmt3a abolish its association with heterochromatin (T. Chen and E. Li, pers. comm.). Non-CpG methylation has been observed for Dnmt3a^{9,27,31}. Wild type Dnmt3b2 and 218 have some activities on non-CpG-containing oligonucleotides that contain CpA and CpT sites. However, it is puzzling that the 369 protein loses the ability to methylate non-CpG site because recognition of methylating target is thought to be mediated entirely by the C-terminal MTase domain. As a positive control, Dnmt1 has ~10-fold higher activity on hemimethylated than unmethylated CpG-containing DNA^{5,6,32}. The methyl transfer activity of Dnmt3b2 observed here is equivalent to that of Dnmt1 on unmethylated DNA — that is, ~10% of Dnmt1 activity on hemimethylated DNA — comparable to previous reports⁸. In contrast to Dnmt3a and Dnmt3b, Dnmt1 has no detectable activity on non-CpG DNA.

Structural similarity with other proteins

A DALI³³ search using the PWWP domain revealed no Protein Data Bank entries with structural similarity with the entire domain. However, the five-stranded β -barrel structure of the PWWP domain resembles that of the SAND domain (named after Sp100, AIRE-1, NucP41/75 and DEAF-1)³⁴ and the SMN (survival of motor neuron) Tudor domain³⁵. The SAND domain is present in many proteins linked to chromatin-dependent transcriptional regulation. The SAND domain folds into a five-stranded, twisted antiparallel β -sheet, with four α -helices from three different regions between the strands packing against the β -sheet³⁴ (Fig. 4a). The SAND domains were also demonstrated to bind DNA³⁴. Unlike the SAND domain, the PWWP domain has a five-helix bundle from the contiguous region immediately following the β -sheet that is packed on the opposite side of the β -sheet (Fig. 4a). Despite such a difference, the two domains exhibit similar structure in the β -barrel and are unrelated to any other DNA-binding protein of known structure. Our finding that the PWWP domain alone binds DNA suggests that PWWP domains also define a new family of DNA-binding fold.

The Tudor domain contains only five strands that are remarkably similar to the β -barrel of PWWP domain (Fig. 4b). The β -barrel-like Tudor domain has a negatively charged surface that interacts with the C-terminal Arg-Gly-Gly rich tails³⁵ of the Sm proteins, known substrates of protein arginine methylation. The structure-based sequence alignment revealed noticeable similarity, including the Cys-Ile-Tyr-Pro (CIYP) of SMN1 in the place of SWWP of Dnmt3 (Fig. 4b). At this point, the implication of the apparent structural similarity between Tudor and PWWP domains is unclear.

A widely distributed PWWP domain

The structure of the PWWP domain offers a rationale for the invariant residues in this domain throughout the Dnmt3 family (Fig. 1*b*). The majority of the invariant amino acids are involved in structural packing and intramolecular interactions (Fig. 5). Both the β -barrel and the helical bundle substructures have their own hydrophobic cores (Fig. 5*a,b*). The interface between the two substructures includes interactions mediated by the conserved SWWP motif (Fig. 5*c*). Intramolecular polar or charge interactions probably confer additional stability to the molecule (Fig. 5*d*). The importance of these intramolecular interactions is supported by the observation that most residues involved are invariant in Dnmt3 orthologs (Fig. 1*b*) and conserved among a large number of sequences examined (see below), suggesting that the two substructures behave as a single, folded structural module.

Stec *et al.*²⁸ analyzed 39 proteins and showed that the PWWP domain spans some 70 amino acids, which corresponds only to our β -barrel substructure. However, the extensive interactions at the interface of the two substructures prompted us to reexamine the distribution of PWWP domains. A BLAST search using Dnmt3b2 PWWP domain as the query found homology with 62 eukaryotic proteins from a variety of organisms (Fig. 6*a*). The structure-based multiple sequence alignment revealed three conserved stretches of amino acids, located in β 1–loop L12– β 2, strand β 3–loop L34 and helix α E. These secondary elements are all located at the interface of the two substructures; in particular, the conserved residues in helix α E all face the β -barrel (Fig. 2*a*). Two loops are variable in size among the 62 proteins: the loop between helix α A and strand β 3 (L-A3) and the loop after strand β 5 connecting helix α C (L-5C). In both places, Dnmt3 contains a one-turn 3_{10} -helix (α A or α C) that might become a longer helix in other proteins. Thus, we suggest that the other PWWP domain should include both the β -barrel and the helical bundle substructures as a universal feature.

Only about half of the proteins we identify as carrying a PWWP domain offer any form of characterization, mostly only with a name assigned based on potential functions predicted by sequence similarity (Fig. 6*b*). In addition to the PWWP domain (which is present as two copies in WHSC1)³⁰, one common feature in the proteins is the presence of known chromatin-association domain(s), such as the bromodomain, the chromodomain, the SET domain and the Cys-rich zinc-binding domain (for a recent review, see ref. 36). Similar multiple copies of the CXXC motif are found in Dnmt3a and Dnmt3b and other protein families, such as WHSC1 (ref. 30), BS69 (ref. 37) and BR140 (ref. 38) (Fig. 6*b*, legend).

In addition to the mammalian PWWP-containing proteins, one of the *Arabidopsis thaliana* proteins, AAK01237, is a *Drosophila* trithorax-like protein that contains a Cys-rich region, similar to the one in Dnmt3, and a SET domain, in addition to the PWWP domain (Fig. 6*b*). Notably, the appearance of these domains is in the same order — that is, PWWP, CXXC and MTase/SET domain — in Dnmt3 and the *A. thaliana* protein. The human ortholog of *Drosophila* trithorax (known as HRX, ALL1 or MLL) contains a different zinc-binding region (CXXCXXC) that is closely related to those in DNMT1 and MBD1 (a methyl-CpG binding protein) but not Dnmt3. The combination of the PWWP domain and/or the Cys-rich region, DNA MTase domain, SET domain (found in histone H3 lysine MTases)^{39,40},

bromodomain (binding acetylated lysine)^{41,42} or chromodomain (binding methylated lysine) seems to be a common feature in these fairly large proteins. Similarly, the SAND domain-containing proteins contain several modules, including the PHD-like Cys-rich domain (a protein–protein interaction), domains that mediate transcriptional repression and activation, and a chromatin-association domain (bromodomains)⁴³.

Summary

We have determined the first structure of a PWWP domain. PWWP domains are found in eukaryotic proteins — from yeast to mammals — involved in DNA methylation, DNA repair and regulation of transcription. However, the exact function of PWWP has remained elusive because many proteins containing this domain have never been characterized. We show here that the PWWP domain of Dnmt3b2 has a basic surface that is probably responsible for its DNA binding. We also demonstrated that the PWWP domain is not required for the CpG-methylating activity *in vitro*.

Although our structure and *in vitro* biochemical data suggest that the PWWP domain in Dnmt3b2 is a DNA-binding domain, whether all PWWP-containing proteins would bind DNA is not clear because many of the basic residues in Dnmt3b are not conserved in other members of this family. However, because the majority of PWWP domains (53 out of 62) have a pI value of 8–10, basic residues from locations different than those in Dnmt3b could form a similar basic surface. Consistent with having a universal DNA binding function, PWWP domains have been found, thus far, only in nuclear proteins, some of which bind DNA.

Methods

Expression and purification of Dnmt3b2 protein

The Dnmt3b2 wild type, 218–369, PWWP domain (amino acids 219–365) and the CXXC domain (amino acids 387–561) proteins were expressed in *E. coli* BL21(DE3), with a His₆-tag at the N-terminus. The cells were grown at 22 °C or 37 °C (PWWP) in LB medium containing 50 µg ml^{−1} kanamycin, induced with 0.4 mM isopropyl-β-D-1-thiogalactopyranoside at an OD₆₀₀ of 0.8 for 3 h and harvested by centrifugation. The proteins were purified using chelated-nickel, ion-exchange Mono-S or Mono-Q (CXXC) columns and gel filtration columns (Pharmacia) (Fig. 1a).

Limited proteolysis of Dnmt3b2 218

Purified Dnmt3b2 218 (18 µg) protein was incubated with 10 ng of trypsin (Sigma; 13,600 units mg^{−1}) at room temperature for 40 min. The reaction was stopped with SDS sample buffer or 1 mM phenyl-methyl-sulfonylfluoride. The lower molecular weight sample at ~20 kDa was isolated on a gel filtration column, and its mass was determined by electrospray ionization mass spectrometry to be 17,764.0 Da. The calculated MW of a polypeptide of amino acids 219–365 is 17,761.2 Da.

Oligonucleotides

Oligonucleotides were synthesized by New England Biolabs. Double-stranded oligonucleotides were annealed by mixing equal amounts of complementary strands. In the DNA binding studies, the annealed oligonucleotides were labeled using [γ - 32 P]ATP (Amersham) and polynucleotide kinase (New England Biolabs). The oligonucleotides were CpG: 5'-GAAGCTGGGACTTC-CGGGAGGAGAGTGCAA-3' and 3'-CTTCGACCCTGAAGGCCTCCTCT-CACGTT-5'; Hemi-CpG: 5'-GAAGCTGGGACTTCCGGGAGGAGAGTGCAA-3' and 3'-CTTCGACCCTGAAGGMCCTCCTCTCACGTT-5'; MpG: 5'-GAAGCTGGGACTTCMGGGAGGAGAGTGCAA-3' and 3'-CTTCGACCCTGAAGGMCCTCCTCTCACGTT-5'; and non-CpG: 5'-GAAGCTGGGACTTGGGCCAGGAGAGTGCAA-3' and 3'-CTTCGACCCTGAACCGGGTCCTCTCACGTT-5' (M = 5-methyl-cytosine).

Nitrocellulose filter-binding assay

Different amounts of protein (Fig. 3a,b) were incubated with labeled oligonucleotides (5 nM) in 50 mM Tris-HCl, pH 7.5, and 20 mM NaCl for 30 min at room temperature. The nitrocellulose filter membranes (Protran BA 85, Schleicher & Schuell) was presoaked for 10 min in 0.4 M KOH to reduce retention of free single-stranded DNA and then rinsed in Milli-Q water until the pH returned to neutral. The samples were sucked through the filter membrane and washed four times with 2 ml of 50 mM Tris-HCl, pH 7.5, and 20 mM NaCl, and then washed four times with 2 ml of Milli-Q water. The bound radioactivity was analyzed by liquid scintillation counting. Each point shown was the mean of at least two determinations.

Methyl transfer activity assay

The reaction mixture (20 μ l total volume) contained 0.2 μ M Dnmt3b2 (purified recombinant proteins: wild type, 218, 369, PWWP and CXXC), 0.5 μ M substrate oligonucleotides and 1 μ Ci [methyl- 3 H] AdoMet (78.0 Ci mmol $^{-1}$, NEN) in 50 mM Tris-HCl, pH 7.5, and 1 mM EDTA. The reactions for each protein and each oligonucleotide were in duplicate. Proteins were preincubated with AdoMet at 37 °C for 20 min, and the reactions were carried out at 37 °C for 1 h after adding DNA. Reactions were stopped by transferring the reaction tubes to an ethanol/dry ice bath and processed by spotting the reaction mix on DE81 paper circles (Whatman). The circles were washed sequentially four times with 2 ml of cold 0.2 M NH $_4$ HCO $_3$, four times with 2 ml of deionized water and once with 1 ml ethanol. The dried circles were subjected to liquid scintillation counting using Cytoscint scintillant.

Crystallography

The Dnmt3b2 PWWP domain (amino acids 219–365) was concentrated to 24 mg ml $^{-1}$ in 20 mM HEPES, pH 7.0, 200 mM NaCl, 1 mM EDTA and 5% (v/v) glycerol. Crystals were obtained in a hanging drop in 30% (w/v) PEG 8000, 200 mM sodium acetate and 100 mM sodium cacodylate, pH 6.5, at 16 °C. SeMet-containing derivatives were purified from a methionine auxotrophic *E. coli* strain (B834) grown in the presence of 25 μ g ml $^{-1}$ SeMet. The SeMet protein crystals were obtained under the same conditions as the native crystals.

Complete data sets were collected from a native crystal and a SeMet-labeled crystal (Table 1). The data were processed using the HKL package⁴⁴. There was one molecule per asymmetric unit. Five selenium sites were determined by SOLVE⁴⁵, and RESOLVE⁴⁶ was then used to modify the electron density map. The modified map was of excellent quality to place amino acids 223–357 of Dnmt3b2 into the recognizable densities using O⁴⁷ (Fig. 5e). The electron density is not observed for four N-terminal residues 219–222, eight C-terminal residues 358–365 and Phe 243. The resultant model was refined up to 1.8 Å resolution using X-PLOR⁴⁸ (with 10% reflections excluded for crossvalidation). The final model includes 1,057 protein atoms and 162 water molecules, with r.m.s. deviations of 0.006 Å and 1.3° from ideality for bond lengths and angles, respectively.

Coordinates

Atomic coordinates have been deposited in the Protein Data Bank (accession code 1KHC).

Acknowledgments

We thank E. Li for providing mouse Dnmt3b2 cDNA, S. Pradhan for providing full-length Dnmt1; K.D. Wilkinson for help with analysis of DNA binding titration curves; P. Kearney for constructing overexpression plasmids; and L. Zhou, J.R. Horton, D. Schneider and R.M. Sweet for help with X-ray data collection. We also thank R.M. Blumenthal, T.H. Bestor and P.A. Wade for their critical comments on the manuscript. The study was supported in part by the National Institutes of Health grant to X.C.

References

1. Bestor TH. The DNA methyltransferases of mammals. *Hum. Mol. Genet.* 2000; 9:2395–2402. [PubMed: 11005794]
2. Wade PA. Methyl CpG binding proteins: coupling chromatin architecture to gene regulation. *Oncogene.* 2001; 20:3166–3173. [PubMed: 11420733]
3. Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. *Science.* 2001; 293:1068–1070. [PubMed: 11498573]
4. Yoder JA, Soman NS, Verdine GL, Bestor TH. DNA (cytosine-5) methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J. Mol. Biol.* 1997; 270:385–395. [PubMed: 9237905]
5. Pradhan S, Bacolla A, Wells RD, Roberts RJ. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of *de novo* and maintenance methylation. *J. Biol. Chem.* 1999; 274:33002–33010. [PubMed: 10551868]
6. Fatemi M, Hermann A, Pradhan S, Jeltsch A. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *J. Mol. Biol.* 2001; 309:1189–1199. [PubMed: 11399088]
7. Dong A, et al. Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Res.* 2001; 29:439–448. [PubMed: 11139614]
8. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genet.* 1998; 19:219–220. [PubMed: 9662389]
9. Gowher H, Jeltsch A. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG sites. *J. Mol. Biol.* 2001; 309:1201–1208. [PubMed: 11399089]
10. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell.* 1992; 69:915–926. [PubMed: 1606615]

11. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferase Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell*. 1999; 99:247–257. [PubMed: 10555141]
12. Finnegan EJ, Peacock WJ, Dennis ES. Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proc. Natl. Acad. Sci USA*. 1996; 93:8449–8454. [PubMed: 8710891]
13. Kakutani T, Jeddelloh JA, Flowers SK, Munakata K, Richards EJ. Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proc. Natl. Acad. Sci. USA*. 1996; 93:12406–12411. [PubMed: 8901594]
14. Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL. Demethylation-induced developmental pleiotropy in *Arabidopsis*. *Science*. 1996; 273:654–657. [PubMed: 8662558]
15. Bird AP, Wolffe AP. Methylation-induced repression — belts, braces, and chromatin. *Cell*. 1999; 99:451–454. [PubMed: 10589672]
16. Riggs AD, Pfeifer GP. X-chromosome inactivation and cell memory. *Trends Genet*. 1992; 8:169–174. [PubMed: 1369742]
17. Shemer, R.; Razin, A. Epigenetic mechanisms of gene regulation. Russo, VEA.; Martienssen, RA.; Riggs, AD., editors. New York: Cold Spring Harbor Laboratory Press; 1996. p. 215–229.
18. Siegfried Z, Cedar H. DNA methylation: a molecular lock. *Curr. Biol*. 1997; 7:R305–R307. [PubMed: 9115385]
19. Xie S, Wang Z, Okano M, Li E. Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene*. 1999; 236:87–95. [PubMed: 10433969]
20. Xu G, et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*. 1999; 402:187–191. [PubMed: 10647011]
21. Robertson K, et al. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Res*. 1999; 27:2291–2298. [PubMed: 10325416]
22. Saito Y, et al. Expression of mRNA for DNA methyltransferases and methyl-CpG-binding proteins and DNA methylation status on CpG islands and pericentromeric satellite regions during human hepatocarcinogenesis. *Hepatology*. 2001; 33:561–568. [PubMed: 11230735]
23. Kanai Y, Ushijima S, Kondo Y, Nakanishi Y, Hirohashi S. DNA methyltransferases expression and DNA methylation of CpG islands and pericentromeric satellite regions in human colorectal and stomach cancers. *Int. J. Cancer*. 2001; 91:205–212. [PubMed: 11146446]
24. Mizuno S, et al. Expression of DNA methyltransferases DNMT1 3a, and 3b in normal hematopoiesis and in acute chronic myelogenous leukemia. *Blood*. 2001; 97:1172–1179. [PubMed: 11222358]
25. Bachman KE, Rountree MR, Baylin SB. Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J. Biol. Chem*. 2001; 276:32282–32287. [PubMed: 11427539]
26. Fuks F, Burgers WA, Godin N, Kasai M, Kouzarides T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J*. 2001; 20:2536–2544. [PubMed: 11350943]
27. Aoki A, et al. Enzymatic properties of *de novo* type mouse DNA (cytosine-5) methyltransferases. *Nucleic Acids Res*. 2001; 29:3506–3512. [PubMed: 11522819]
28. Stec I, Nagl S, van Ommen G, Dunnen J. The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? *FEBS Lett*. 2000; 473:1–5. [PubMed: 10802047]
29. Izumoto Y, Kuroda T, Harada H, Kishimoto T, Nakamura H. Hepatoma-derived growth factor belongs to a gene family in mice showing significant homology in the amino terminus. *Biochem. Biophys. Res. Commun*. 1997; 238:26–32. [PubMed: 9299445]
30. Stec I, et al. WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a *Drosophila* dysmorphia gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum. Mol. Genet*. 1998; 7:1071–1082. [PubMed: 9618163]

31. Ramsahoye RH, et al. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. USA*. 2000; 97:5237–5242. [PubMed: 10805783]
32. Flynn J, Glickman JF, Reich NO. Murine DNA cytosine-C5 methyltransferase: pre-steady- and steady-state kinetic analysis with regulatory DNA sequences. *Biochemistry*. 1996; 35:7308–7315. [PubMed: 8652507]
33. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 1993; 233:123–138. [PubMed: 8377180]
34. Bottomley MJ, et al. The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. *Nature Struct. Biol.* 2001; 8:626–633. [PubMed: 11427895]
35. Selenko P, et al. SMN Tudor domain structure and its interaction with the Sm proteins. *Nature Struct. Biol.* 2001; 8:27–31. [PubMed: 11135666]
36. Marmorstein R. Protein modules that manipulate histone tails for chromatin regulation. *Nature Rev. Mol. Cell Biol.* 2001; 2:422–432. [PubMed: 11389466]
37. Hateboer G, et al. BS69, a novel adenovirus E1A-associated protein that inhibits E1A transactivation. *EMBO J.* 1995; 14:3159–3169. [PubMed: 7621829]
38. Thompson KA, et al. BR140, a novel zinc-finger protein with homology to the TAF250 subunit of TFIID. *Biochem. Biophys. Res. Commun.* 1994; 198:1143–1152. [PubMed: 7906940]
39. Rea S, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*. 2000; 406:593–599. [PubMed: 10949293]
40. Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science*. 2001; 292:110–113. [PubMed: 11283354]
41. Jeanmougin F, Wurtz JM, Le Douarin B, Chambon P, Losson R. The bromodomain revisited. *Trends Biochem. Sci.* 1997; 22:151–153. [PubMed: 9175470]
42. Dhalluin C, et al. Structure and ligand of a histone acetyltransferase bromodomain. *Nature*. 1999; 399:491–496. [PubMed: 10365964]
43. Wojciak JM, Clubb RT. Finding the function buried in SAND. *Nature Struct. Biol.* 2001; 8:568–570. [PubMed: 11427878]
44. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 1997; 276:307–326.
45. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallogr. D.* 1999; 55:849–861. [PubMed: 10089316]
46. Terwilliger TC. Maximum likelihood density modification. *Acta Crystallogr. D.* 2000; 56:965–972. [PubMed: 10944333]
47. Jones TA, Kjeldgaard M. Electron-density map interpretation. *Methods Enzymol.* 1997; 277:173–208. [PubMed: 18488310]
48. Brünger, AT. X-PLOR. A system for X-ray crystallography and NMR, version 3.1. New Haven: Yale University; 1992.
49. Carson M. Ribbons. *Methods Enzymol.* 1997; 277:493–505. [PubMed: 18488321]
50. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Protein Struct. Funct. Genet.* 1991; 11:281–296.
51. Palombo F, et al. GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. *Science*. 1995; 268:1912–1914. [PubMed: 7604265]
52. Min J, Landry J, Sternglanz R, Xu R-M. Crystal structure of a SIR2 homolog-NAD complex. *Cell*. 2001; 105:269–279. [PubMed: 11336676]

Fig. 1.

Dnmt3 MTase family. **a**, Schematic representation of the Dnmt3 family. The amino acid sequences of Dnmt3a and Dnmt3b are highly conserved with respect to one another in the PWWP domain, CXXC domain and C-terminal MTase domain. In contrast, the N-terminal variable domain differs substantially between these two proteins. The alternative splice site in Dnmt3b1 that generates 3b2 is indicated by a triangle. Below the full-length Dnmt3b2 (WT) are the depicted protein fragments (218–369, PWWP and CXXC) and the SDS-PAGE gels (right) of the recombinant proteins used throughout this study. **b**, Sequence alignment of Dnmt3 PWWP domains. The prefix M indicates mouse; H, human; and Zf, zebrafish. The mouse Dnmt3b residue numbering is shown above the sequences, with the secondary structure elements (numbers for strands and letters for α -helices). The highlighted amino acids are either invariant (cyan) or similar (yellow). The SWWP motif is colored in magenta. In zebrafish, there are three partial sequences for Dnmt3 orthologs. Two ESTs (expressed sequence tags) contain only the N-terminal region, including the PWWP domain: AW305577 is more similar to Dnmt3a, whereas BG307263 is more similar to Dnmt3b. The third EST (AAD32631) has a C-terminal portion (CXXC and MTase domains) that shares similarity to the corresponding portion of mammalian Dnmt3 family. However, its N-terminal PWWP domain deviates significantly from the sequences, including eight positively charged (blue) and three negatively charged residues (red), resulting in a pI of 6.3.

Fig. 2.

Structure of Dnmt3b2 PWWP domain. **a**, RIBBON⁴⁹ diagram, with invariant amino acids in cyan, conserved in yellow and varied in gray. The SWWP motif is colored in magenta. In the crystal packing contacts, strand $\beta 6$ from one molecule interacts with strand $\beta 4$, which is the edge strand of the β -sheet (4 3 2 1 5), of a neighboring molecule through backbone–backbone hydrogen bonds in an antiparallel arrangement. **b**, Molecular surface and charge distribution. The view is oriented similarly to that in (a). The surface is colored according to charge: positively charged groups (Arg and Lys) are blue, negatively charged groups (Glu and Asp) are red and uncharged groups are white. **c**, Contours of the electrostatic potential at $\pm 2 \text{ KT e}^{-1}$ calculated and displayed in two orientations as a mesh surface. The positive potential is shown in blue, and the negative potential in red. Panels (b,c) were done using GRASP⁵⁰.

Fig. 3. DNA binding and methyl transfer activity of Dnmt3b2. **a**, Nitrocellulose filter binding experiments using different protein fragments of Dnmt3b2 (300 nM) and labeled oligonucleotides (5 nM), whose sequences contained a single CpG, hemimethylated CpG (hemi-CpG), fully methylated CpG (MpG) or non-CpG (see Methods). **b**, Nitrocellulose filter-binding experiments using different amounts of Δ218 (0, 5, 15, 25, 50, 100, 150, 200 and 300 nM), PWWP (0, 25, 50, 100, 150, 200 and 300 nM) and a labeled CpG-containing oligonucleotide (5 nM). K_d values were estimated from a direct nonlinear regression fit (with the saturation level as one of the variables). **c**, CPM counts. Methyl transfer activities of purified recombinant Dnmt3b2 proteins (expressed in *E. coli*) and full-length Dnmt1 (ref. 5) (expressed in Sf21 insert cells using recombinant baculoviruses).

a *b* **Fig. 4.**

Comparison of PWWP domain with its structural homologs. **a**, Ribbon diagram and stereo view of the superimposition of PWWP (yellow) and Sp100b SAND domain (PDB 1H5P)³³ (orange). For clarity, the C-terminal helix of SAND domain is not shown in the superimposition. The structural similarity between the two domains was detected manually. Below is the structure-based sequence alignment between the β -barrel structure of PWWP and SAND. The line between the two sequences indicates the residues used in superimposition. **b**, Ribbon diagram and stereo view of the superimposition of PWWP (yellow) and the SMN Tudor domain (PDB 1G5V)³⁴ (orange). The r.m.s. deviation is 0.8 Å when 43 C α atoms (out of total 56 residues in Tudor) are structurally aligned between the two domains (DALI Z-score of 6.3). The largest difference lies in the loop between strands β 2 and β 3. In comparison, the five-stranded β -barrel of PWWP is remotely similar to the SH3 structures (DALI Z-score of ~3.3), a small basic folding unit. The structure-based sequence alignment between the β -barrel structure of PWWP and Tudor is shown below, with the line between the two sequences indicating the residues used in superimposition.

Fig. 5.

Views of defined elements in the Dnmt3b2 PWWP structure. **a**, Hydrophobic core of the β -barrel centered on residue Trp 270 (β 3), packing against Gly 238 (β 1), Ile 240 (after strand β 1), Ala 248 (β 2), Val 268 (β 3) and Ile 279 (β 4). A hydrogen bond is formed between Trp 270 and Ser 277 (β 4). These residues are from strands β 1 to β 4; therefore, these interactions appear to anchor the β -subdomain. **b**, Hydrophobic core of the helical bundle. The interactions involve helices α D (Phe 299), α E (Tyr 310, Met 309 and Leu 313), α F (Met 339, Trp 342 and Ala 343), α G (Leu 355), β 6 (Phe 324) and two of their connecting loops — the loop prior helix α D (Phe 294) and the loop between helices α F and α G (Phe 347). **c**, The interactions involving the SWWP motif in the interface between the β -barrel and helical bundle substructures. The Pro 247 at the end of the SWWP motif is completely buried within the interface, whereas the other three residues are exposed on the surface. Trp 245 and Trp 246 point in opposite directions but have a similar network of interactions — the indole rings are flanked by a ring structure (His 293 imidazole ring and Tyr 305 aromatic ring) and by the aliphatic portion of a Lys side chain (Lys 239 of strand β 1 and Lys 241 of loop L-12). One difference between the Trp residues is that a water molecule is held in position by three hydrogen bonds with Trp 246, Ser 244 and Lys 241, but there is no equivalent water associated with Trp 245. **d**, Network of intramolecular polar interactions (centered on residue Arg 319). The two terminal amino groups of the positive charged Arg 319 side chain (helix α E) interact with main chain carbonyl oxygen atoms of Gly 233 (N-terminal loop) and Lys 356 (C-terminal loop). The plane of the Arg 319 guanidino group makes van der Waals contacts with the proline ring of Pro 357. The negatively charged Asp 234 side chain (next to Gly 233) interacts with the main chain amino group of Gly 231. **e**,

Stereo view showing an omit electron density map contoured at 4σ , superimposed with refined coordinates of SWWP motif.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 6.

Sequence alignment of PWWP-containing proteins. **a**, The residues ‘WP’ (magenta) are conserved among all, except for the BR140 family in which a Tyr in the place of Trp. Other conserved residues are highlighted with a yellow background (hydrophobic), blue (basic) and red (acidic). The accession numbers are colored based on the range of calculated pI value: blue (8–10) and magenta (6–7). The prefix M indicates mouse; H, human; At, *A. thaliana*; and Sp, *Schizosaccharomyces pombe*. In addition to the sequences shown, the other PWWP-containing proteins used in alignment are from human (AAD53063, AAD32631, CAC28351, AAF25871, CAB75688, BAA86439, BAA86600, NP_055392, CAB66669, AAG44637); mouse (AAD33084, AF064553, AAB88445, BAA90478, BAA22896, AAF65469, BAA22895); *A. thaliana* (AAF64539, AAF23297, CAB86049, BAB11589, AAF29390, AAD24842, CAB93724, CAB87752, AAF18621, BAA97320, CAB71104); zebrafish (BG307263, AW305577, AAC40177, AI588565), *Drosophila* (AAF56762, AAF59276, AAF57004, AAF55003, AAF57177, AAF58494), *Xenopus* (AW636253); *Saccharomyces cerevisiae* (AAB64719, NP_013758); *S. pombe* (CAA22121, AB027814); bovine (CAA40348); *Onchocerca volvulus* (BE132518); *Ceanorhabditis elegans* (AAB52437); and *Triticum aestivum* (BE500209). **b**, The schematic representation of PWWP-containing proteins. Abbreviations are NLS (nuclear localization signal), HMG (high mobility group), PCNA (PCNA binding motif), SET (Suvar39, Enhancer of zeste, Trithorax), bromo (bromodomain) and chromo (chromodomain). WHSC1 is a 136-kDa protein expressed ubiquitously in early development and preferentially in rapidly growing

embryonic tissues³⁰. BS69 is a 69-kDa nuclear protein that specifically suppresses adenovirus E1A-activated transcription³⁷. BR140 shows similarity to a group of transcriptional coactivators, including the TAF250 subunit of TFIID³⁸. MSH6 is mammalian MutS homolog essential for DNA mismatch repair⁵¹, which contains a PWWP domain at its extreme N-terminus. However, msh6 orthologs in *S. cerevisiae*, *C. elegans* and *A. thaliana* do not contain this domain. The fully sequenced *C. elegans* contains only one PWWP protein, AAB52437; this protein contains an N-terminal PWWP domain and a C-terminal SIR2-like domain. Although the known eukaryotic members of the NAD-dependent protein deacetylase SIR2 family often contains the N- and/or C-terminal extensions outside of a conserved core region, none of these contains a PWWP domain. A CXXC zinc-binding motif is found in the structure of an *Archaeoglobus fulgidus* SIR2 homolog, which contains a CXXC-X₁₅₋₂₀-CXXC zinc-binding sequence motif⁵².

Table 1

Summary of X-ray diffraction data collection

	SeMet			Native
Cell dimension (Å)				
a = b	53.2			53.1
c	186.8			186.8
Wavelength (Å)	0.97881	0.97858	0.95	1.54
Resolution limit (Å) ^I	20–2.1			20–1.8
	(2.14–2.1)			(1.83–1.8)
Completeness (%) ^I	96	96.3	98.5	99.2
	(65.4)	(68.0)	(81.7)	(94.3)
R _{linear} (%) ^I	6.9	7.3	5.7	3.8
	(8.5)	(9.3)	(6.9)	(14.8)
Number of Unique reflections	9,572	9,559	9,800	15,308
Observations	97,945	103,761	105,781	145,433
I / σ (I)	43.4	40.3	37.4	23.6
Overall figure of merit	0.72			

^I Number in parentheses is for the highest resolution shell.