UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mechanistic Insights into Dynamic Regulation and Substrate Recognition in Mammalian DNA Methylation

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Environmental Toxicology

by

Linfeng Gao

June 2020

Dissertation Committee:
Dr. Jikui Song, Chairperson
Dr. Yinsheng Wang
Dr. Seán O'Leary

The Dissertation of Linfeng Gao is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

The text of this dissertation, in part or in full, is a reprint of the material as it appears in:

1. Linfeng Gao, Xiao-Feng Tan, Shen Zhang, Tianchen Wu, Zhi-Min Zhang, Hui-wang Ai, Jikui Song. An intramolecular interaction of UHRF1 reveals dual control for its histone association. Structure. 2018.

2. Linfeng Gao, Hiwot Anteneh, Jikui Song. Dissect the DNMT3A- and DNMT3B-mediated DNA co-methylation through a covalent complex approach. J. Mol. Biol. 2019.

The co-author Dr. Jikui Song listed in those publications directed and supervised the research which forms the basis for this dissertation.

ABSTRACT OF THE DISSERTATION

Mechanistic Insights into Dynamic Regulation and Substrate Recognition in Mammalian DNA Methylation

by

Linfeng Gao

Doctor of Philosophy, Graduate Program in Environmental Toxicology
University of California, Riverside, June 2020
Dr. Jikui Song, Chairperson

Mammalian DNA methylation serves an important role in epigenetic regulation. The methylation marks are imprinted by *de novo* methyltransferases DNMT3A and DNMT3B during gametogenesis and early embryonic development, and further preserved by UHRF1-mediated maintenance DNA methylation machinery in each cell cycle. Here, we present the crystal structures of an intramolecular complex of UHRF1, as well as DNMT3B-DNMT3L tetramer in complex with CpG/CpA DNA, respectively. Combined with biochemical and functional analysis, we've revealed the molecular basis of the allosteric regulation of UHRF1 and DNMT3B-mediated *de novo* DNA methylation, which comprehensively provides mechanistic insights in mammalian DNA methylation.

Table of Contents

List of Figures

Introduction

Chapter 1

Chapter 2

Chapter 3

Conclusion

List of Tables

Chapter 1

Chapter 2

General introduction into DNA methylation

Epigenetics is recognized as stable heritable phenotypes or traits changes in cells that affect certain genes activation and protein expression but without DNA sequence alteration.[1] Examples of mechanisms that produce such changes include: chromatin modification (DNA methylation and histone modification) [2] and RNA modification [3]. DNA methylation is well studied and it is recognized as one of the major epigenetic mechanisms. It is a biochemical event in which a methyl group is added onto adenine or cytosine base of DNA. [4, 5] Among all methylation modifications of DNA, cytosine methylation occuring at the C5 position, where the methyl group from S-Adenosyl methionine (SAM or AdoMet) is transferred to the fifth carbon of a cytosine residue to form 5mC, is the most widespread form in all organisms [6]. (Figure.1) In prokaryotes, DNA methylation serves in defense mechanisms against extraneous DNA, transcription regulation, replication initiation and in Dam-directed mismatch repair [7]. In eukaryotes, it is associated with silencing of retrotransposons, genomic imprinting and X-chromosome inactivation through silencing the specific genome regions and genes selectively, assembly of heterochromatin in a heritable manner, aging, and carcinogenesis. [8-12] In human, there are around 60-70% of genes have a CpG island in their promoter region [7], and about 60–80% of the CpG sites in the mammalian genome are modified by 5mC in somatic cells [8], while non-CpG methylation occurs in a quarter of total methylation in embryonic stem cells (ESCs) [13], with remaining unmethylated CpG islands near promoter region [14].

Mammalian DNA methylation has two major categories: maintenance DNA methylation and *de novo* DNA methylation, which are catalyzed by different DNA methyltransfereases [15]. Firstly, the DNA methylation pattern is established by *de novo* DNA methyltransferases DNMT3A/3B during germ cell development and early embryogenesis, and maintained by DNA methyltransferase 1 (DNMT1) in each cell cycle [16-18]. (Figure 2) Sequentially, symmetrically methylated CpGs are recognized by the reader proteins like MECP2, MBD1 and MBD2 that contains Methyl-CpG-binding domain (MBD) that mediate recruitment of other factors to block transcription and condense the chromatin structure or DNA repair. [19] Furthermore, DNA methylation marks can be removed through demethylation processes. [20-22]

DNMT1 and maintenance DNA methylation.

DNMT1 plays a crucial role in maintenance DNA methylation. After establishment, DNA methylation patterns needs to be preserved by DNMT1 over cell divisions [24] During S-phase, the newly synthesized DNA strand loses all methylation marks; it forms hemimethylated DNA with its parental DNA strand. DNMT1 is recruited to replication foci and methylates the hemimethylated DNA into full methylated DNA double strands, thus precisely passing the original DNA methylation patterns into the next generation [25]. (Figure 2b) In the absence of DNMT1, the replication machinery would lead to passive demethylation over cell cycles. DNMT1 is a multi-domain protein with around 1600 amino acids. It contains a complicated a long N-terminal regulatory region and a smaller C-terminal catalytic core----methyltransferase domain which are linked by a highly conserved (GK)n dipeptide repeat. The N-terminal regulatory region, which covers around two thirds

of the total sequence can be further separated into five domains: the N-terminal independently folded domain (NTD) which harbors a variety of protein and/or DNA interaction sites, followed by replication foci-targeting sequence (RFTS) domain, a CXXC zinc finger motif and two tandemly connected bromo adjacent homology (BAH1 and BAH2) domains. [26](Figure 3) Until now, several crystal structures of mouse or human DNMT1 in free form or in complex with different DNA substrates have been solved [27-30]. The structure-function studies have illustrated how different domains of this enzyme orchestrate its activity under different activation states in maintenance DNA methylation [26]. Based on them, both mouse and human DNMT1 show very similar compact structures, with the catalytic domain in the middle and the regulatory domains surrounding it, forming an autoinhibitory conformation achieved by either CXXC-Mtase interaction or interaction between the RFTS domain and Mtase domain. The multi-layer intramolecular interactions provide allosteric regulation mechanisms of DNMT1, which ensure its substrate-recognition specificity and localization specificity in maintenance DNA methylation. [26, 31, 32]

UHRF1, a key regulator in DNA methylation

Early studies indicated that the recruitment of DNMT1 to replication foci depends on the interaction with the proliferating cell nuclear antigen (PCNA) component of the replication machinery [33]. However, disruption of the interaction between DNMT1 and PCNA only has a very slight effect on maintenance DNA methylation [34]. It was shown that UHRF1 (Ubiquitin-like PHD and RING finger domain 1), plays an important role in DNMT1-mediated DNA methylation [35, 36]. UHRF1, also known as NP95 in mouse and

ICBP90 in human, is a nuclear protein [37] that belongs to a subfamily of RING-finger type E3 ubiquitin ligases [38]. It has E3 ubiquitin-protein ligase activity by promoting the ubiquitination of target proteins such as histone H3 [39], P53 [40] and promyelocytic leukemia (PML) protein [41]. UHRF1 has recently been identified as a novel oncogene and overexpression of UHRF1 correlates with development of many different cancers, which suggest it to be a novel biomarker and potential therapeutic target [42]. UHRF1 is a multi-domain protein that contains: an N-terminal ubiquitin-like (UBL) domain, then followed by a tandem Tudor domain (TTD), a plant homeodomain (PHD), a SET- and RING-associated (SRA) domain, and a C-terminal RING-finger domain [35, 43]. It plays an important role in DNA methylation maintenance and chromatin modification throughout S phase of the cell cycle [22]. The SET- and RING-associated (SRA) domain of UHRF1 specifically binds to hemimethylated CpG sites of DNA, which also is the physiological substrate of DNMT1 specifically, indicating a model in which UHRF1 recruits DNMT1 to hemimethylated DNA sites for the accomplishment of DNA methylation during the DNA replication [35, 36, 44-47]. Disruption of UHRF1-DNMT1 interaction causes dramatic DNA methylation reduction, which leads to global DNA hypomethylation [48, 49], which highlight its essential role as a key regulator in maintenance DNA methylation. Furthermore, UHRF1 can also interacts with DNMT3A and DNMT3B, which may suggest a role for UHRF1 in *de novo* DNA methylation [50, 51]. Besides to its role in DNA methylation, it also plays a key role in chromatin modification recognition: the tandem Tudor domain (TTD) specifically binds to histone H3 trimethylated at 'Lys-9' (H3K9me3), and the plant homeodomain (PHD) can recognize

H3R2me0 [52-54]. The C-terminal RING finger domain serves as an E3 ubiquitin ligase that mediates the ubiquitination of histone H3 lysine 18 or 23, which in turn recruits DNMT1 to the replication foci [39, 55], and regulate the RFTS domain-mediated DNMT1 autoinhibition [56]. Moreover, UHRF1 has been associated with heterochromatin replication and assembly [57], DNA damage response (DDR) [58], and gene transcription [59].

DNMT3s in *de novo* DNA methylation

In mammals, there are three main members in DNMT3 family: DNMT3A, DNMT3B and DNMT3L. During the early development and differentiation, *de novo* DNA methyltransferases DNMT3A and DNMT3B are responsible for the establishment of DNA methylation patterns [17]. Both of them are regulated by DNMT3-like protein (DNMT3L), which is a catalytically inactive paralog of DNMT3 enzymes, which directly interacts with the catalytic domain of DNMT3A and DNMT3B and stabilizes their binding of AdoMet in germ and embryonic stem (ES) cells [60]. Thus, it is suggested that DNMT3L is responsible for the stimulation of enzymatic activity of DNMT3A and DNMT3B [61]. DNMT3A and DNMT3B are highly homologous, sharing similar domain arrangement: N-terminal domain, followed by a Pro-Trp-Trp-Pro (PWWP) domain, an ADD (Atrx-Dnmt3-Dnmt3l) domain (also known as the plant homeodomain (PHD)), and a methyltransferase domain in the C-terminus all together with 46% seuqence identity. (Figure 4) The cofactor DNMT3L contains an ADD domain and a non-active methyltransferase fold with the absence of a DNA recognition domain [60, 62, 63]. (Figure 4)

The N-Terminal tail is highly divergent between DNMT3A and DNMT3B. The N-terminal sequence (residues 1–211) of DNMT3A can bind to DNA and enhance methylation activity compared to DNMT3A2 which lacks first 219 amino acid residues [64]. Secondly, the N-terminal domain can guide the localization of DNMT3A in cells [64-66]. With its N-terminal sequence, DNMT3A is enriched in heterochromatin regions, while DNMT3A2 was shown to be localized at euchromatin region in the absence of the N-terminal sequence [65]. The PWWP domains of DNMT3A and DNMT3B, which belong to the Royal super-family of domains that recognize histone tails with different modifications [67], can bind both DNA and histone tail H3K36 tri-methylation [68-71], which are essential for targeting de novo DNA methylation to heterochromatin regions in the nucleus [68, 72]. The DNMT3A PWWP domain was also reported to bind NSD1-mediated H3K36me2 at non-coding regions of euchromatin [73, 74]. All DNMT3 family members proteins contain a cysteine rich plant homeodomain (PHD)-like Atrx-Dnmt3-Dnmt3l (ADD) domain [75, 76], which comprise three subdomains: GATA binding protein 1 (GATA1), plant homeodomain (PHD)-type and a C-terminal α-helix [77, 78]. Three subdomains were packed together to form a single globular domain through extensive hydrophobic interactions [77]. It was reported that the ADD domain of all DNMT3 protein can specifically recognize histone unmethylated H3K4 with a conserved recognition mechanism [78-80]. This interaction between ADD domain and H3K4 mark does not only guide DNMT3 to chromatin region in the nucleus [80], but also allosterically activates the enzymatic activity of DNMT3A and DNMT3B through disrupting their intramolecular interactions [81, 82]. Previous study showed that the ADD domain of DNMT3A can

interact with the methyltransferase domain and inhibit its enzymatic activity. Binding of H3 (but not H3K4me3) tail induced a conformational transition of DNMT3A from inhibitory form ('X') shape to active form (butterfly shape). All residues involved in the intramolecular interaction are highly conserved in DNMT3A and DNMT3B, suggesting a conserved allosteric regulation mode of DNTM3 methyltransferases.[82] The C-terminal methyltransferase domain (or catalytic domain) of DNMT3 family enzymes (DNMT3A and DNMT3B) is highly conserved, with over 80% identity [83, 84]. It belongs to a classic AdoMet-dependent DNA 5-mC MTase that contains six methyltransferase motifs (I, IV, VI, VIII, IX, and X) which shares a Rossmann fold core structure. The regulatory cofactor DNMT3L lacks the MTase motifs IX and X and key catalytic residues in its C-terminal domain [84-86]. A crystal structure of DNMT3A-DNMT3L in complex with DNA (PDB: 5YX2) demonstrated that DNMT3A binds to DNA though three parts: DNMT3A catalysis region, target recognition domain (TRD) and homodimeric interface, which together create a continuous DNA-binding surface [84]. The catalysis region contains three conserved motifs: 1) IV motif (catalytic loop) where the active site is located; 2) VI motif (ENV motif) that provide the important glutamate for cytosine protonation ; 3) RXR motif (VIII motif) for flipped cytosine binding and neutralization of the negative charge of the DNA backbone; all together are essential for catalysis of DNA methylation [84, 87-90]. Meanwhile, it has been shown that the C-terminal catalytic domain of DNMT3A and DNMT3B can directly interact with their regulatory cofactor DNMT3L to forms a 'butterfly shape' heterotetramer [61, 63, 91, 92]. In the crystal structure of the DNMT3A-DNMT3L complex, the heterotetramer comprises two DNMT3A molecules in the middle and one DNMT3L

molecule at each edge. The interaction that forming the heterotetramer is mediated by two interfaces: DNMT3A homodimeric interface (RD interface) formed by a hydrogen bonding network mainly between R881 and D872 from both DNMT3A and DNMT3A/DNMT3L heterodimeric interface (FF interface) formed by two pairs of phenylalanine residues (F728 and F768 of DNMT3A, and F261 and F301 of DNMT3L) through hydrophobic stacking interaction[63]. Notably, DNMT3L contains only the heterodimeric interface (FF interface) with homodimeric interface (RD interface) absent where DNMT3A contains both, suggesting that DNMT3A can form homotetramer or to be oligomerized without DNMT3L [63, 93, 94]. Although a structural study of DNMT3B is lacking, high sequence identity suggests that DNMT3A and DNMT3B may share the similar catalysis and oligomerization mechanism in mammalian *de novo* DNA methylation.

Diseases in *de novo* DNA methylation

Failure of appropriate DNA methylation regulation is connected to human diseases, including neurodevelopmental disorders and cancer. [9, 95, 96] For instance, dysregulation of de novo DNA methyltransferase caused by mutations will lead to human pathogenesis of leukemia, and neurodevelopmental disorders of ICF (for immunodeficiency, centromeric instability and facial anomalies).

DNMT3A and Leukemia

Recently, mutations in the gene encoding DNA methyltransferase 3A (DNMT3A) were reported in patients with acute myeloid leukemia (AML), and in patients with various other hematological cancers, suggesting wild-type DNMT3A is a critically important new tumor suppressor and DNMT3A mutations play a prominent role in clonal and malignant

hematopoiesis.[97-99] Particularly, DNMT3A regulates DNA methylation at the edges of large regions of hypomethylation called DNA methylation canyons and also gene body in hematopoietic stem cells (HSCs), highlighted its unique role in HSCs development and self-renewal. [100, 101] Mutations in DNMT3A have now been found in most types of hematological malignancy with varying frequency. DNMT3A is mostly frequently mutated in patients diagnosed with AML in myeloid lineage, with most studies reporting a mutation frequency of 20–25% in DNMT3A [102-104]. Many studies examining all or most of the coding region in AML reported that around 60% of DNMT3A mutations are found at the residue R882 in the methyltransferase domain, highlighting this mutation as a hotspot. In all of these diseases, DNMT3A mutations are typically heterozygous, with biallelic involvement essentially confined to non-R882 mutants. [105-107]

The prevalence of the R882 mutants has caught people's attention. Mutational analysis of R878 in the catalytic domain of the murine protein (equivalent to R882 of the human protein; discussed below) showed abrogated enzymatic activity and reduced DNA binding.[109] The methyltransferase activity of R882H DNMT3A is reduced by around 80% compared to the WT enzyme. Co-expression of DNMT3A wild-type and R882H mutant in cells profoundly inhibits the WT enzyme by disrupting its ability to form active wild-type homo-tetramer, suggesting the possibility of dominant negative effect. AML cells with the R882H mutation have significant enzymatic activity reduction, resulting in focal hypomethylation at specific CpGs throughout AML cell genomes and cause the initiation of carcinogenesis of AML. [105, 110] Meanwhile, DNMT3A R882H mutant was found to have a gain of function effect. Notably, DNMT3A R882H can directly binds to

and potentiates transactivation of stemness genes critical for leukemogenicity including Meis1, Mn1 and Hoxa gene cluster, inducing focal epigenetic alterations, including CpG hypomethylation and concurrent gain of active histone modifications, at cis-regulatory elements such as enhancers to facilitate gene transcription. Cooperation with NRase mutation can transform hematopoietic stem/progenitor cells and induce acute leukemia development. [111] Moreover, DNMT3A R882 mutant can interact with polycomb repressive complex 1 (PRC1), causing transcriptional silencing, then block the differentiation of HSCs and leukemic cells.[112] Other studies also show that R882 is involved in the indirect readout of flanking sequence preferences of DNMT3A and the R882 mutant altered substrate specificity of DNMT3A.[113] Despite previous clinical research on DNMT3A mutation in acute myeloid leukemia, the molecular mechanisms regarding how this mutation causes leukemogenesis *in vivo* are largely unknown. As R882 is located at homodimer interface of DNMT3A, which does not belong to catalytic center. This suggests that an indirect enzymatic inhibition effect in the R882 mutant. [63, 84] Therefore, the molecular basis of R882H-mediated loss of function and gain of function still remains elusive.

DNMT3B and ICF Syndrome

The autosomal recessive Immunodeficiency, Centromeric instability and Facial anomalies (ICF) syndrome is a rare disease and was initially distinguished based on centromeric instability, namely multi-branching of juxta-centromeric regions of chromosomes 1, 9, and 16, from the blood samples of patients with primary immunodeficiency and recurrent infections.[114-116] ICF is a clinically heterogeneous

disease, although it is almost invariantly characterized by a severe impairment to humoral immunity involving hypo- or agamma-globulinemia. B and T cell numbers and V(D)J recombination are normal, suggesting a defect in activation and maturation of lymphocyte cells development at late stages. Failure of proper immune system construction will result in severely raised infection risk which causes death of young patients during early age. [117, 118] ICF is a genetically heterogeneous disease [119, 120], with 50% to 60% of cases are associated with germ line mutations in DNMT3B gene. [118, 121, 122] *De novo* DNA methyltransferase DNMT3B is responsible for CpG island methylation- mediated transcriptional repression in centromeric, pericentromeric, subtelomeric repeats, and also the germline genes. [123-125] Recent studies have proposed three situations demonstrating how DNMT3B mutation affects transcription and thereby contributes to initiation and development of ICF syndrome. Firstly, DNMT3B mutations alter the transcription through disrupting the proper transcription start site (TSS) usage and causing false transcription from intragenic cryptic transcription start site. The aberrant hypomethylation of CpG at alternative or cryptic TSSs caused by DNMT3B dysfunction leads to their illegitimate activation interfering with the transcription and elongation of the appropriate mRNA. Inappropriate translocation of *de novo* methyltransferase DNMT3B will cause hypermethylation of the canonical transcription starting site which will move transcription initiation from alternative sites. (Figure 6A) [126] Secondly, DNMT3B ICF mutants can dysregulate sense-antisense transcription through altering the epigenetic-based self-regulatory circuit. In the general situation, natural antisense RNAs (NATs) can induce a threshold-dependent turn on/off of sense-gene expression, resulting a well-maintained

transcription regulation machinery.[127, 128] However, abnormal activity of DNMT3B mutant would lead to failure of inhibiting the expression of the antisense CD27AS TSS (located at 3′ of the memory B-cell marker CD27 gene locus). Oppositely, it will promote antisense CD27AS transcription. The upregulated antisense transcript recruits the extra *de novo* methyltransferase to the sense TSS, inducing CpG hypermethylation and CD27 gene silencing. (Figure 6B) [126]

Lastly, past studies have been shown that DNA methylation plays an essential role in the alternative splicing of mRNA. [8, 130, 131] In ICF, DNMT3B mutants can affect appropriate alternative exons splicing during the maturation of mRNA. In patient-derived lymphoblastoid cell lines (LCLs), cells, three alternative exons of the trans-membrane protein tyrosine phosphatase (PTPRC) gene are abnormally spliced giving rise to a shorter mature mRNA compared to control cells. Thus, the mutant DNMT3B protein would contribute to the exon skipping acting as an adaptor protein, by interacting with both heterogeneous nuclear ribonucleoproteins (hnRNPs) and the pre-mRNA (Figure 6C).[126] However, due to lack of structural observation of DNMT3B-mediated DNA methylation and transcription regulation, the molecular mechanisms of the function of DNMT3B and DNA methylation in gene expression regulation in association with ICF syndrome still remains poorly understood. In summary, the developmental disorders caused by DNA methyltransferases mutations highlight their importance of DNA methylation in transcription regulation and development. Accumulating observations emphasize that not only the sequence of the human genome, but also the chromosomal modifications, play essential roles in healthy and diseased cells.

Figures

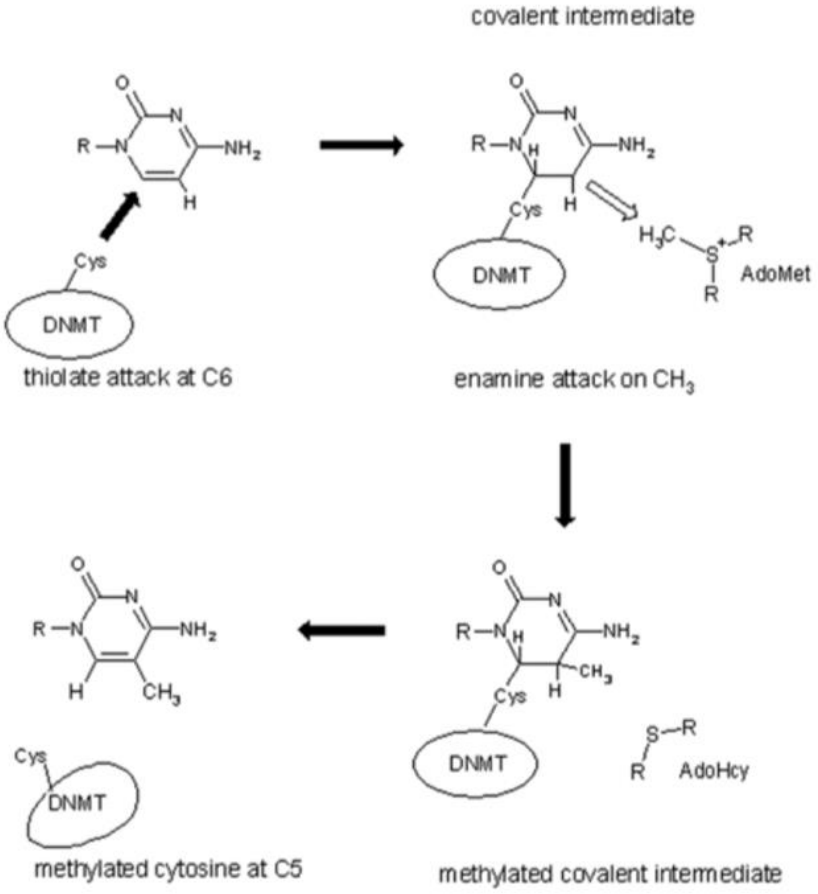


Figure 1. Catalysis of C5-methylation by DNA methyltransferase. [6]

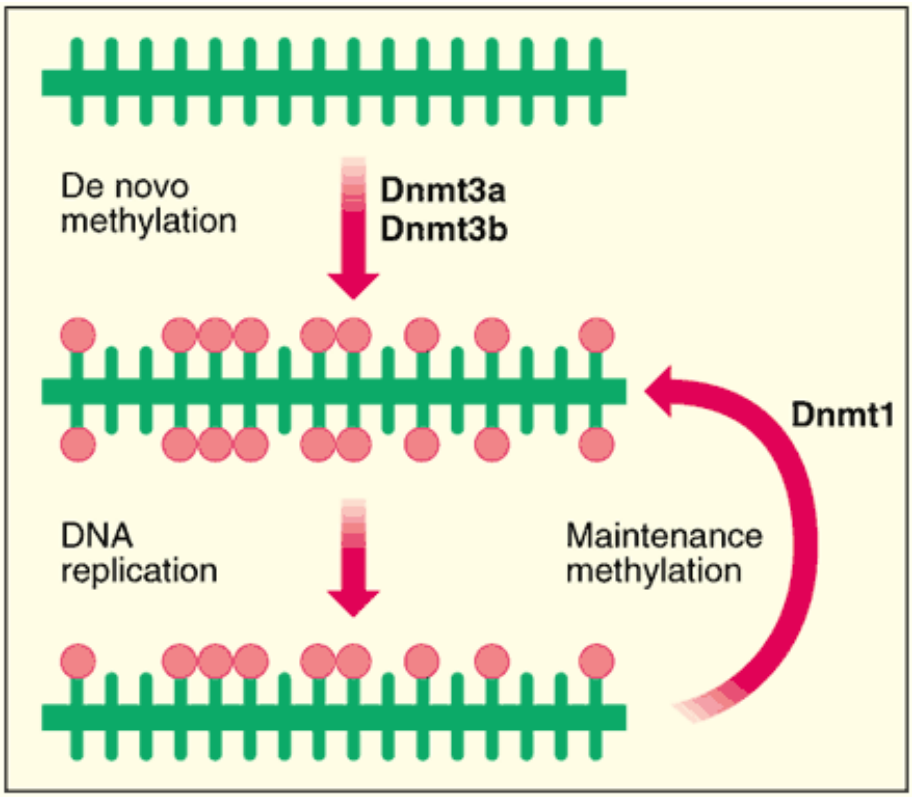


Figure 2. Two categories of DNA methylation. (a) DNMT3A and DNMT3B are the de novo DNMTs and transfer methyl groups (red) onto naked DNA. (b) When DNA undergoes semiconservative replication, the parental DNA stand retains the original DNA methylation pattern (red). DNMT1 associates at the replication foci and precisely replicates the original DNA methylation pattern by adding methyl groups (red) onto the newly formed daughter strand.[15] [23]

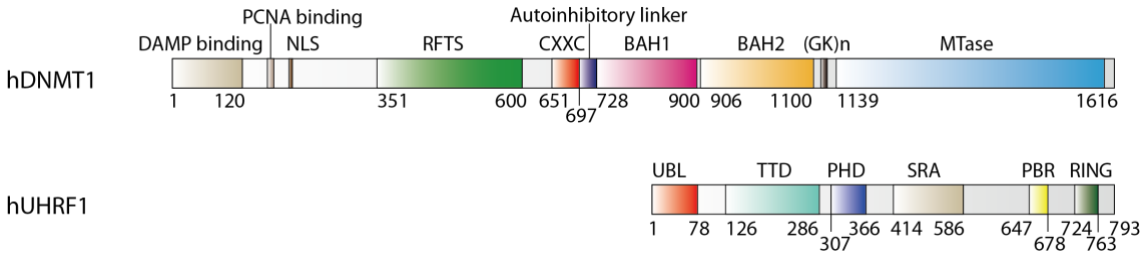


Figure 3. Domain architecture of hDNMT1 and hUHRF1 in maintenance DNA methylation.

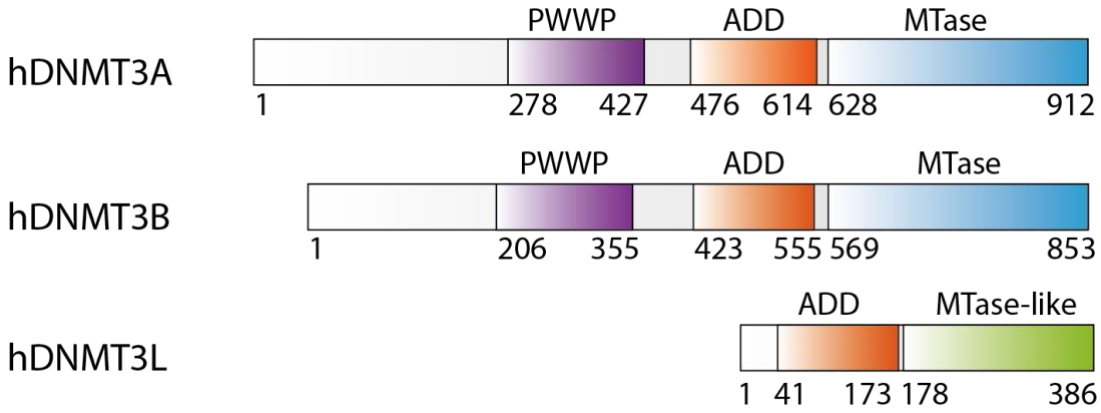


Figure 4. Domain architecture of hDNMT3A, hDNMT3B and hDNMT3L.

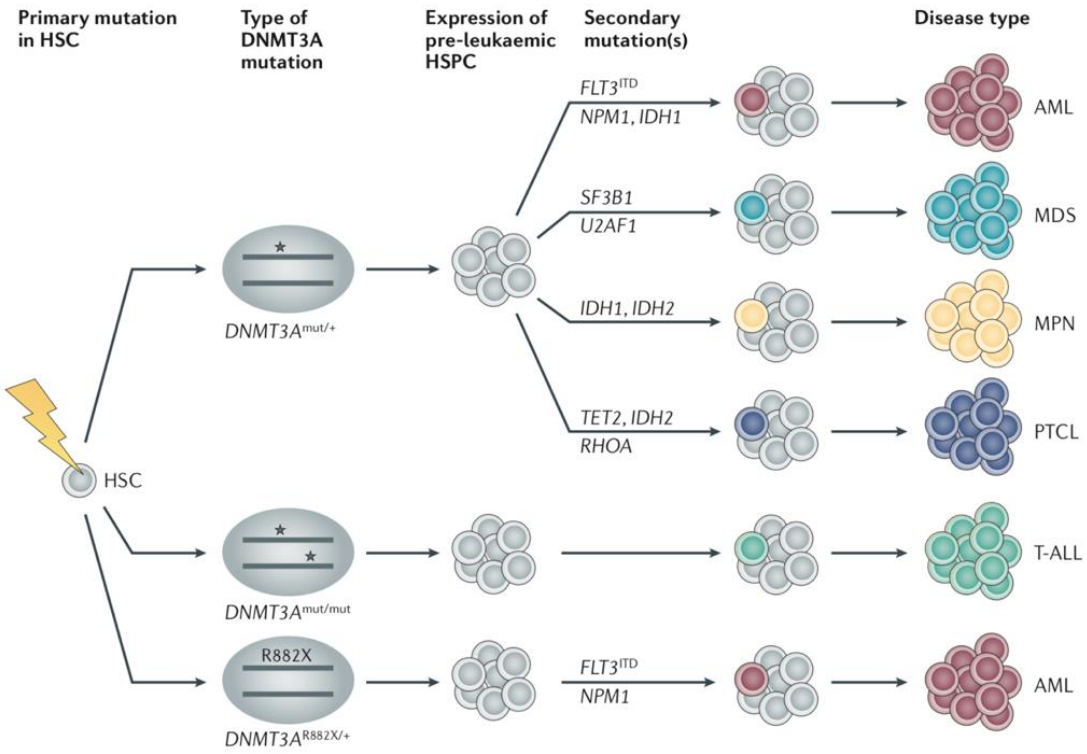


Figure 5. DNMT3A mutation allele and gene dosage, combined with secondary mutations, are likely to dictate the type of haematological disease.[108]

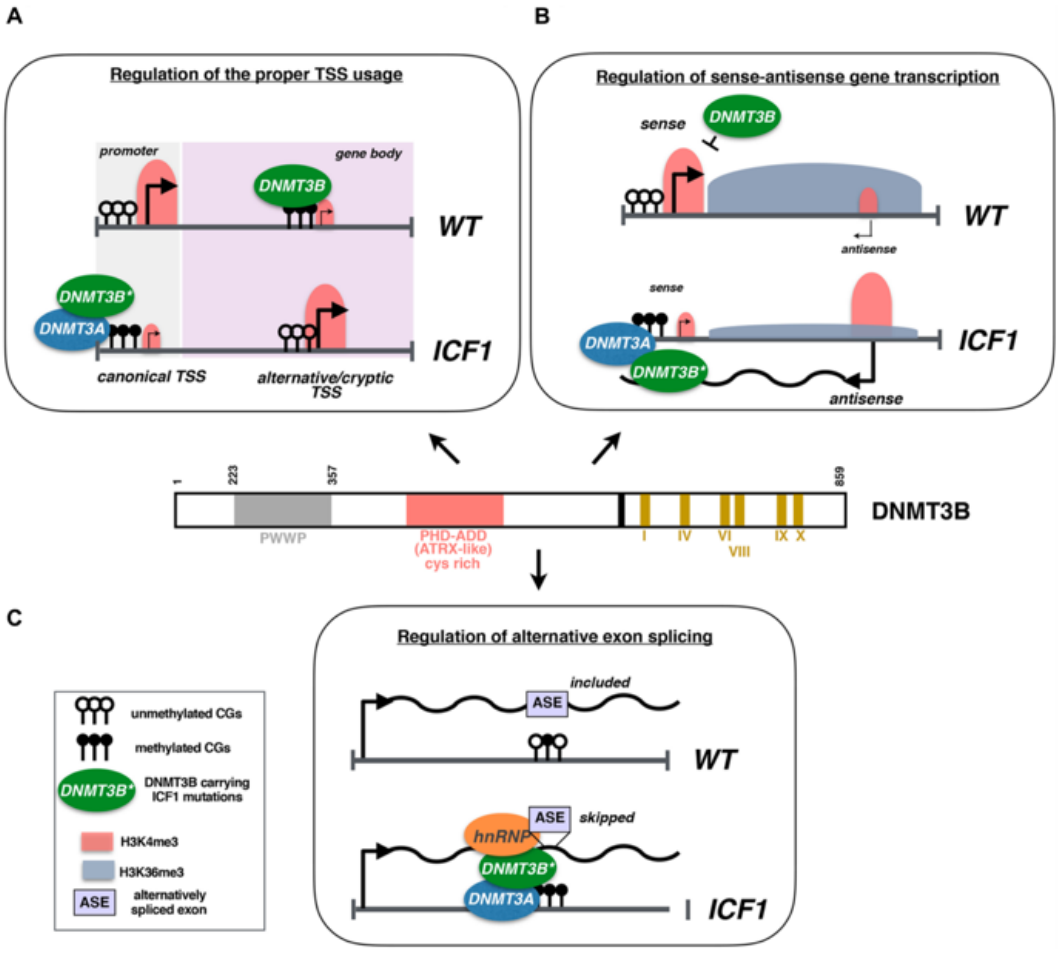


Figure 6. DNMT3B mutation induced dysregulation in ICF syndrome. [129]

Reference

1. Deans, C. and K.A. Maggert, *What do you mean, "epigenetic"?* Genetics, 2015. **199**(4): p. 887-96.
2. Ptashne, M., *On the use of the word 'epigenetic'.* Curr Biol, 2007. **17**(7): p. R233-6.
3. Holoch, D. and D. Moazed, *RNA-mediated epigenetic regulation of gene expression.* Nat Rev Genet, 2015. **16**(2): p. 71-84.
4. Baniushin, B.F., *[Methylation of adenine residues in DNA of eukaryotes].* Mol Biol (Mosk), 2005. **39**(4): p. 557-66.
5. Gromova, E.S. and A.V. Khoroshaev, *[Prokaryotic DNA methyltransferases: the structure and the mechanism of interaction with DNA].* Mol Biol (Mosk), 2003. **37**(2): p. 300-14.
6. Siedlecki, P. and P. Zielenkiewicz, *Mammalian DNA methyltransferases.* Acta Biochim Pol, 2006. **53**(2): p. 245-56.
7. Blow, M.J., et al., *The Epigenomic Landscape of Prokaryotes.* PLoS Genet, 2016. **12**(2): p. e1005854.
8. Anastasiadou, C., et al., *Human epigenome data reveal increased CpG methylation in alternatively spliced sites and putative exonic splicing enhancers.* DNA Cell Biol, 2011. **30**(5): p. 267-75.
9. Baylin, S.B., et al., *Alterations in DNA methylation: a fundamental aspect of neoplasia.* Adv Cancer Res, 1998. **72**: p. 141-96.
10. Kumar, S., V. Chinnusamy, and T. Mohapatra, *Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond.* Front Genet, 2018. **9**: p. 640.
11. Bird, A., *DNA methylation patterns and epigenetic memory.* Genes Dev, 2002. **16**(1): p. 6-21.
12. Klutstein, M., et al., *DNA Methylation in Cancer and Aging.* Cancer Res, 2016. **76**(12): p. 3446-50.
13. Jin, B., Y. Li, and K.D. Robertson, *DNA methylation: superior or subordinate in the epigenetic hierarchy?* Genes Cancer, 2011. **2**(6): p. 607-17.
14. Suzuki, M.M. and A. Bird, *DNA methylation landscapes: provocative insights from epigenomics.* Nat Rev Genet, 2008. **9**(6): p. 465-76.
15. Moore, L.D., T. Le, and G. Fan, *DNA methylation and its basic function.* Neuropsychopharmacology, 2013. **38**(1): p. 23-38.
16. Kato, Y., et al., *Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse.* Hum Mol Genet, 2007. **16**(19): p. 2272-80.
17. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.
18. Goll, M.G. and T.H. Bestor, *Eukaryotic cytosine methyltransferases.* Annu Rev Biochem, 2005. **74**: p. 481-514.
19. Nan, X., R.R. Meehan, and A. Bird, *Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2.* Nucleic Acids Res, 1993. **21**(21): p. 4886-92.
20. Dean, W., et al., *Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos.* Proc Natl Acad Sci U S A, 2001. **98**(24): p. 13734-8.

21. Reik, W., W. Dean, and J. Walter, *Epigenetic reprogramming in mammalian development.* Science, 2001. **293**(5532): p. 1089-93.

22. Oswald, J., et al., *Active demethylation of the paternal genome in the mouse zygote.* Curr Biol, 2000. **10**(8): p. 475-8.

23. Bird, A., *Molecular biology - DNA methylation de novo.* Science, 1999. **286**(5448): p. 2287-2288.

24. Kim, J.K., M. Samaranayake, and S. Pradhan, *Epigenetic mechanisms in mammals.* Cell Mol Life Sci, 2009. **66**(4): p. 596-612.

25. Law, J.A. and S.E. Jacobsen, *Establishing, maintaining and modifying DNA methylation patterns in plants and animals.* Nat Rev Genet, 2010. **11**(3): p. 204-20.

26. Ren, W., L. Gao, and J. Song, *Structural Basis of DNMT1 and DNMT3A-Mediated DNA Methylation.* Genes (Basel), 2018. **9**(12).

27. Song, J., et al., *Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation.* Science, 2011. **331**(6020): p. 1036-40.

28. Takeshita, K., et al., *Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1).* Proc Natl Acad Sci U S A, 2011. **108**(22): p. 9055-9.

29. Song, J., et al., *Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation.* Science, 2012. **335**(6069): p. 709-12.

30. Zhang, Z.M., et al., *Crystal Structure of Human DNA Methyltransferase 1.* J Mol Biol, 2015. **427**(15): p. 2520-2531.

31. Jeltsch, A., *On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme.* Epigenetics, 2006. **1**(2): p. 63-6.

32. Leonhardt, H., et al., *A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei.* Cell, 1992. **71**(5): p. 865-73.

33. Chuang, L.S., et al., *Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1.* Science, 1997. **277**(5334): p. 1996-2000.

34. Spada, F., et al., *DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells.* J Cell Biol, 2007. **176**(5): p. 565-71.

35. Bostick, M., et al., *UHRF1 plays a role in maintaining DNA methylation in mammalian cells.* Science, 2007. **317**(5845): p. 1760-4.

36. Sharif, J., et al., *The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA.* Nature, 2007. **450**(7171): p. 908-12.

37. Tauber, M. and W. Fischle, *Conserved linker regions and their regulation determine multiple chromatin-binding modes of UHRF1.* Nucleus, 2015. **6**(2): p. 123-32.

38. Jenkins, Y., et al., *Critical role of the ubiquitin ligase activity of UHRF1, a nuclear RING finger protein, in tumor cell growth.* Mol Biol Cell, 2005. **16**(12): p. 5621-9.

39. Nishiyama, A., et al., *Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication.* Nature, 2013. **502**(7470): p. 249-53.

40. Ma, J., et al., *Ubiquitin E3 ligase UHRF1 regulates p53 ubiquitination and p53-dependent cell apoptosis in clear cell Renal Cell Carcinoma.* Biochem Biophys Res Commun, 2015. **464**(1): p. 147-53.

41. Guan, D., et al., *The epigenetic regulator UHRF1 promotes ubiquitination-mediated degradation of the tumor-suppressor protein promyelocytic leukemia protein.* Oncogene, 2013. **32**(33): p. 3819-28.

42. Bronner, C., M. Krifa, and M. Mousli, *Increasing role of UHRF1 in the reading and inheritance of the epigenetic code as well as in tumorogenesis.* Biochem Pharmacol, 2013. **86**(12): p. 1643-9.

43. Hashimoto, H., et al., *UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications.* Epigenetics, 2009. **4**(1): p. 8-14.

44. Avvakumov, G.V., et al., *Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1.* Nature, 2008. **455**(7214): p. 822-5.

45. Arita, K., et al., *Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism.* Nature, 2008. **455**(7214): p. 818-21.

46. Hashimoto, H., et al., *The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix.* Nature, 2008. **455**(7214): p. 826-9.

47. Qian, C., et al., *Structure and hemimethylated CpG binding of the SRA domain from human UHRF1.* J Biol Chem, 2008. **283**(50): p. 34490-4.

48. Hervouet, E., et al., *Disruption of Dnmt1/PCNA/UHRF1 interactions promotes tumorigenesis from human and mice glial cells.* PLoS One, 2010. **5**(6): p. e11333.

49. Pacaud, R., et al., *The DNMT1/PCNA/UHRF1 disruption induces tumorigenesis characterized by similar genetic and epigenetic signatures.* Sci Rep, 2014. **4**: p. 4230.

50. Jia, Y., et al., *Negative regulation of DNMT3A de novo DNA methylation by frequently overexpressed UHRF family proteins as a mechanism for widespread DNA hypomethylation in cancer.* Cell Discov, 2016. **2**: p. 16007.

51. Meilinger, D., et al., *Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells.* EMBO Rep, 2009. **10**(11): p. 1259-64.

52. Cheng, J., et al., *Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein.* J Biol Chem, 2013. **288**(2): p. 1329-39.

53. Hu, L., et al., *Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2.* Cell Res, 2011. **21**(9): p. 1374-8.

54. Wang, C., et al., *Structural basis for site-specific reading of unmodified R2 of histone H3 tail by UHRF1 PHD finger.* Cell Res, 2011. **21**(9): p. 1379-82.

55. Qin, W., et al., *DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination.* Cell Res, 2015. **25**(8): p. 911-29.

56. Ishiyama, S., et al., *Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance.* Mol Cell, 2017. **68**(2): p. 350-360 e7.

57. Taylor, E.M., et al., *Depletion of Uhrf1 inhibits chromosomal DNA replication in Xenopus egg extracts.* Nucleic Acids Res, 2013. **41**(16): p. 7725-37.

58. Tian, Y., et al., *UHRF1 contributes to DNA damage repair as a lesion recognition factor and nuclease scaffold.* Cell Rep, 2015. **10**(12): p. 1957-66.

59. Kim, J.K., et al., *UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells.* Nucleic Acids Res, 2009. **37**(2): p. 493-505.

60. Hata, K., et al., *Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice.* Development, 2002. **129**(8): p. 1983-93.

61. Suetake, I., et al., *DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction.* J Biol Chem, 2004. **279**(26): p. 27816-23.

62. Aapola, U., et al., *Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family.* Genomics, 2000. **65**(3): p. 293-8.

63. Jia, D., et al., *Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation.* Nature, 2007. **449**(7159): p. 248-51.

64. Suetake, I., et al., *Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a.* Biochem J, 2011. **437**(1): p. 141-8.

65. Chen, T., et al., *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**(41): p. 38746-54.

66. Kotini, A.G., A. Mpakali, and T. Agalioti, *Dnmt3a1 upregulates transcription of distinct genes and targets chromosomal gene clusters for epigenetic silencing in mouse embryonic stem cells.* Mol Cell Biol, 2011. **31**(7): p. 1577-92.

67. Qin, S. and J. Min, *Structure and function of the nucleosome-binding PWWP domain.* Trends Biochem Sci, 2014. **39**(11): p. 536-47.

68. Chen, T., N. Tsujimoto, and E. Li, *The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin.* Mol Cell Biol, 2004. **24**(20): p. 9048-58.

69. Purdy, M.M., C. Holz-Schietinger, and N.O. Reich, *Identification of a second DNA binding site in human DNA methyltransferase 3A by substrate inhibition and domain deletion.* Arch Biochem Biophys, 2010. **498**(1): p. 13-22.

70. Dhayalan, A., et al., *The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation.* J Biol Chem, 2010. **285**(34): p. 26114-20.

71. Rondelet, G., et al., *Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B.* J Struct Biol, 2016. **194**(3): p. 357-67.

72. Ge, Y.Z., et al., *Chromatin targeting of de novo DNA methyltransferases by the PWWP domain.* J Biol Chem, 2004. **279**(24): p. 25447-54.

73. Weinberg, D.N., et al., *The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape.* Nature, 2019. **573**(7773): p. 281-286.

74. Xu, W., et al., *Correction to: DNMT3A reads and connects histone H3K36me2 to DNA methylation.* Protein Cell, 2020. **11**(3): p. 230.

75. Jurkowska, R.Z., T.P. Jurkowski, and A. Jeltsch, *Structure and function of mammalian DNA methyltransferases.* Chembiochem, 2011. **12**(2): p. 206-22.

76. Jeltsch, A. and R.Z. Jurkowska, *Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm.* Nucleic Acids Res, 2016. **44**(18): p. 8556-8575.

77. Argentaro, A., et al., *Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX.* Proc Natl Acad Sci U S A, 2007. **104**(29): p. 11939-44.

78. Otani, J., et al., *Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain.* EMBO Rep, 2009. **10**(11): p. 1235-41.

79. Ooi, S.K., et al., *DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.* Nature, 2007. **448**(7154): p. 714-7.

80. Zhang, Y., et al., *Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail.* Nucleic Acids Res, 2010. **38**(13): p. 4246-53.

81. Li, B.Z., et al., *Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase.* Cell Res, 2011. **21**(8): p. 1172-81.

82. Guo, X., et al., *Structural insight into autoinhibition and histone H3-induced activation of DNMT3A.* Nature, 2015. **517**(7536): p. 640-4.

83. Barau, J., et al., *The DNA methyltransferase DNMT3C protects male germ cells from transposon activity.* Science, 2016. **354**(6314): p. 909-912.

84. Zhang, Z.M., et al., *Structural basis for DNMT3A-mediated de novo DNA methylation.* Nature, 2018. **554**(7692): p. 387-391.

85. Malone, T., R.M. Blumenthal, and X. Cheng, *Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes.* J Mol Biol, 1995. **253**(4): p. 618-32.

86. Jurkowska, R.Z. and A. Jeltsch, *Mechanisms and Biological Roles of DNA Methyltransferases and DNA Methylation: From Past Achievements to Future Challenges.* Adv Exp Med Biol, 2016. **945**: p. 1-17.

87. Posfai, J., et al., *Predictive motifs derived from cytosine methyltransferases.* Nucleic Acids Res, 1989. **17**(7): p. 2421-35.

88. Kumar, S., et al., *The DNA (cytosine-5) methyltransferases.* Nucleic Acids Res, 1994. **22**(1): p. 1-10.

89. Jeltsch, A., *Molecular enzymology of mammalian DNA methyltransferases.* Curr Top Microbiol Immunol, 2006. **301**: p. 203-25.

90. O'Gara, M., et al., *Enzymatic C5-cytosine methylation of DNA: mechanistic implications of new crystal structures for HhaL methyltransferase-DNA-AdoHcy complexes.* J Mol Biol, 1996. **261**(5): p. 634-45.

91. Chen, Z.X., et al., *Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family.* J Cell Biochem, 2005. **95**(5): p. 902-17.

92. Jurkowska, R.Z., et al., *Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L.* Nucleic Acids Res, 2008. **36**(21): p. 6656-63.

93. Holz-Schietinger, C., et al., *Oligomerization of DNMT3A controls the mechanism of de novo DNA methylation.* J Biol Chem, 2011. **286**(48): p. 41479-88.

94. Holz-Schietinger, C. and N.O. Reich, *De novo DNA methyltransferase DNMT3A: Regulation of oligomeric state and mechanism of action in response to pH changes.* Biochim Biophys Acta, 2015. **1850**(6): p. 1131-9.

95. Jones, P.A. and P.W. Laird, *Cancer epigenetics comes of age.* Nat Genet, 1999. **21**(2): p. 163-7.

96. Tycko, B. and J. Ashkenas, *Epigenetics and its role in disease.* J Clin Invest, 2000. **105**(3): p. 245-6.

97. Kim, M.S., et al., *Mutational analysis of DNMT3A gene in acute leukemias and common solid cancers.* APMIS, 2013. **121**(2): p. 85-94.

98. Kandoth, C., et al., *Mutational landscape and significance across 12 major cancer types.* Nature, 2013. **502**(7471): p. 333-339.

99. Forbes, S.A., et al., *COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer.* Nucleic Acids Res, 2011. **39**(Database issue): p. D945-50.

100. Jeong, M., et al., *Large conserved domains of low DNA methylation maintained by Dnmt3a.* Nat Genet, 2014. **46**(1): p. 17-23.

101. Wu, H., et al., *Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes.* Science, 2010. **329**(5990): p. 444-8.

102. Kataoka, K., et al., *Integrated molecular analysis of adult T cell leukemia/lymphoma.* Nat Genet, 2015. **47**(11): p. 1304-15.

103. Yamashita, Y., et al., *Array-based genomic resequencing of human leukemia.* Oncogene, 2010. **29**(25): p. 3723-31.

104. Ley, T.J., et al., *DNMT3A mutations in acute myeloid leukemia.* N Engl J Med, 2010. **363**(25): p. 2424-33.

105. Russler-Germain, D.A., et al., *The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers.* Cancer Cell, 2014. **25**(4): p. 442-54.

106. Gaidzik, V.I., et al., *Clinical impact of DNMT3A mutations in younger adult patients with acute myeloid leukemia: results of the AML Study Group (AMLSG).* Blood, 2013. **121**(23): p. 4769-77.

107. Thol, F., et al., *Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia.* J Clin Oncol, 2011. **29**(21): p. 2889-96.

108. Yang, L., R. Rau, and M.A. Goodell, *DNMT3A in haematological malignancies.* Nat Rev Cancer, 2015. **15**(3): p. 152-65.

109. Gowher, H., et al., *Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase.* J Mol Biol, 2006. **357**(3): p. 928-41.

110. Spencer, D.H., et al., *CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression.* Cell, 2017. **168**(5): p. 801-816 e13.

111. Lu, R., et al., *Epigenetic Perturbations by Arg882-Mutated DNMT3A Potentiate Aberrant Stem Cell Gene-Expression Program and Acute Leukemia Development.* Cancer Cell, 2016. **30**(1): p. 92-107.

112. Koya, J., et al., *DNMT3A R882 mutants interact with polycomb proteins to block haematopoietic stem and leukaemic cell differentiation.* Nat Commun, 2016. **7**: p. 10924.

113. Emperle, M., et al., *The DNMT3A R882H mutant displays altered flanking sequence preferences.* Nucleic Acids Res, 2018. **46**(6): p. 3130-3139.

114. Maraschio, P., et al., *Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome.* J Med Genet, 1988. **25**(3): p. 173-80.

115. Howard, P.J., et al., *Centromeric instability of chromosomes 1 and 16 with variable immune deficiency: a new syndrome.* Clin Genet, 1985. **27**(5): p. 501-5.

116. Fryns, J.P., et al., *Centromeric instability of chromosomes 1, 9, and 16 associated with combined immunodeficiency.* Hum Genet, 1981. **57**(1): p. 108-10.

117. Tiepolo, L., et al., *Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency.* Hum Genet, 1979. **51**(2): p. 127-37.

118. Ehrlich, M., et al., *DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes.* Hum Mol Genet, 2001. **10**(25): p. 2917-31.

119. Jiang, Y.L., et al., *DNMT3B mutations and DNA methylation defect define two types of ICF syndrome.* Hum Mutat, 2005. **25**(1): p. 56-63.

120. Wijmenga, C., et al., *Genetic variation in ICF syndrome: evidence for genetic heterogeneity.* Hum Mutat, 2000. **16**(6): p. 509-17.

121. Xu, G.L., et al., *Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene.* Nature, 1999. **402**(6758): p. 187-91.

122. Hansen, R.S., et al., *The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome.* Proc Natl Acad Sci U S A, 1999. **96**(25): p. 14412-7.

123. Jeong, S., et al., *Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA.* Mol Cell Biol, 2009. **29**(19): p. 5366-76.

124. Walton, E.L., C. Francastel, and G. Velasco, *Maintenance of DNA methylation: Dnmt3b joins the dance.* Epigenetics, 2011. **6**(11): p. 1373-7.

125. Walton, E.L., C. Francastel, and G. Velasco, *Dnmt3b Prefers Germ Line Genes and Centromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases.* Biology (Basel), 2014. **3**(3): p. 578-605.

126. Gatto, S., et al., *ICF-specific DNMT3B dysfunction interferes with intragenic regulation of mRNA transcription and alternative splicing.* Nucleic Acids Res, 2017. **45**(10): p. 5739-5756.

127. Pelechano, V. and L.M. Steinmetz, *Gene regulation by antisense transcription.* Nat Rev Genet, 2013. **14**(12): p. 880-93.

128. Sorek, R. and P. Cossart, *Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity.* Nat Rev Genet, 2010. **11**(1): p. 9-16.

129. Gagliardi, M., M. Strazzullo, and M.R. Matarazzo, *DNMT3B Functions: Novel Insights From Human Disease.* Front Cell Dev Biol, 2018. **6**: p. 140.

130. Chodavarapu, R.K., et al., *Relationship between nucleosome positioning and DNA methylation.* Nature, 2010. **466**(7304): p. 388-92.

131. Gelfman, S., et al., *DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure.* Genome Res, 2013. **23**(5): p. 789-99.

Chapter 1

An intramolecular interaction of UHRF1 reveals dual control for its histone association

Abstract

UHRF1 (Ubiquitin-like, containing PHD and RING Finger domains, 1) is one of the essential components of mammalian DNA methylation machinery. Chromatin association of UHRF1 is controlled via an interplay between its intramolecular interaction and dual recognition of histone H3 trimethylated at lysine 9 (H3K9me3) and hemimethylated DNA. Here, we report the crystal structure of the N-terminal tandem Tudor domain (TTD) of UHRF1 in complex with the C-terminal polybasic region (PBR). Structural analysis reveals that PBR binding leads to displacement of the TTD – Plant homeodomain (PHD) linker, as well as blockage of the H3K9me3-engaging cage, both of which contribute to a chromatin-occluded UHRF1 conformation. Disruption of the TTD – PBR interaction, which is facilitated by the binding of UHRF1 to hemimethylated DNA or regulatory protein USP7, shifts the UHRF1 conformation toward an open state, allowing for efficient H3K9me3 binding. Together, this study provides structural basis for the allosteric regulation of UHRF1.

Introduction

UHRF1, also known as NP95 or ICBP90, is a multi-domain protein that contains an N-terminal ubiquitin-like (UBL) domain, a tandem Tudor domain (TTD), a plant homeodomain (PHD), a SET- and RING-associated (SRA) domain, and a C-terminal RING-finger domain. In the crystal structure of UHRF1 TTD and PHD in complex with H3K9me3 tail (PDB: 3ASK, 4GY5), TTD and PHD together with their linker in between, form a histone binding interface where TTD recognizes trimethylation at Lysine 9 and PHD recognizes unmodified Alanine 1, Arginine 2 and Lysine 4 [1, 2], generating a coordinated reader cassette for the multivalent silencing marks at the H3 tail. The SRA domain specifically binds to hemimethylated DNA [3-5]. In the crystal structure of UHRF1 SRA domain in complex with hemimethylated CpG DNA (PDB: 2ZO0), two loops from SRA domain are inserted into both major and minor grooves respectively, where the 5-methylcytosine is flipped out into the binding pocket of SRA [5]. Together, recognition of both histone and DNA modification contributes to the association of UHRF1 into the replicating heterochromatin region during the S phase of cell cycle [6] This chromatin association of UHRF1 is reported to be controlled by the conformational change of UHRF1 [7-10]. Unliganded form of UHRF1 adopts a "closed" conformation which is driven by the intramolecular interaction between the TTD domain and a C-terminal polybasic region (PBR), which blocks the binding of TTD to H3K9me3. Intermolecular interaction of the PBR with hemimethylated DNA [7, 9], USP7 [10, 11] or PI5P [8] helps transit UHRF1 into an "open" conformation, thereby promoting the TTD – H3K9me3 binding and consequent association of UHRF1 with heterochromatin [7, 8, 10], as well as UHRF1 E3

ligase activity targeting H3 ubiquitylation [9]. An NMR structure of TTD domain and PBR has provided structural evidence on the interaction between the TTD domain and PBR [7]. However, due to lack of high-resolution structural characterization, the mechanism by which the H3K9me3 epigenetic mark "unlocks" the UHRF1 TTD – PBR interaction was not well explained. Here, we present the crystal structure of the N-terminal tandem Tudor domain (TTD) of UHRF1 in complex with the C-terminal polybasic region (PBR) at 1.68 Ångstrom resolution. Structural and biochemical analysis reveals that PBR binding competes both the H3K9me3-binding cage of TTD domain and the linker between TTD and PHD domain. Consistent with previous study, disruption of this intramolecular interaction by the binding of UHRF1 to hemimethylated DNA or regulatory protein USP7, shifts the UHRF1 conformation toward an open state, allowing for efficient H3K9me3 binding [7, 9-11]. Together, this study provides molecular basis for the allosteric regulation of UHRF1.

Materials and Methods

Protein preparation.

DNA sequences encoding UHRF1 TTD domain from zebrafish (residues 129-280) or human (residues 126-284), human UHRF1 TTD-PHD dual domains (residues 126-366), human UHRF1 PBR (residues 634-665) and the five UBL domains of USP7 (residues 560-1084) were each inserted in a pRSFDuet vector (Novagen), preceded by a hexahistidine and SUMO tag. All the proteins were overexpressed in E. coli BL21 (DE3) RIL strain. The transformed cells were grown at 37 °C in LB medium and induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to 0.1 g/L when the OD600 reached 0.8. After induction, the cells continued to grow at 16 °C overnight. The cells were harvested and lysed in buffer containing 50 mM Tris-HCl (pH 8.0), 1 M NaCl, 25 mM imidazole, and 0.5 mM DTT. After centrifugation, the fusion proteins were purified from the supernatant by a Ni-NTA column, followed by removal of the hexahistidine-SUMO tag by Ubiquitin-like-specific protease 1 (ULP-1) cleavage and subsequent Ni-NTA chromatography. The tag-free proteins were further purified by size exclusion chromatography on a Superdex 200 16/600 column (GE Healthcare) pre-equilibrated with a buffer containing 25 mM Tris (pH 7.5), 100 mM NaCl and 2 mM DTT. Purified samples were stored in -80°C at a concentration of ~ 20 mg/mL for future use.

DNA sequence encoding human full-length UHRF1 was inserted in frame with an N-terminal enhanced cyan fluorescence protein (CFP) sequence and a C-terminal yellow fluorescence protein (YFP, Ypet version) sequence in pBad Vector to generate hexihistidine-CFP-UHRF1-YFP construct. Wild type and D142A/E153A mutant of CFP-

UHRF1-YFP proteins were expressed in E. coli BL21 (DE3) RIL strain and purified sequentially through affinity chromatography using a Ni-NTA column, hydrophobicity interaction chromatography using a Phenly column (GE Healthcare), ion exchange chromatography using a Heparin column (GE Healthcare) and size exclusion chromatography using a Superdex 200 16/600 column (GE Healthcare). The purified protein was dissolved in a buffer containing 25 mM Tris (pH 7.5), 100 mM NaCl and concentrated to 1 mg/ml for Förster resonance energy transfer (FRET) assays.

Expression and purification of UHRF1 PBR peptide (residues 634-665) were as described previously.

Crystallization and structure determination.

The UHRF1 TTD-PBR complex was prepared by mixing of zebrafish UHRF1 TTD domain (residues 129-280) with the human UHRF1 PBR peptide in a molar ratio of 1:1.5 at a total concentration of ~25 mg/mL. The complex was mixed with precipitant containing 0.1 M Succinic acid pH7.0, 15% (v/v) Polyethylene glycol 3350 and crystallized using hanging-drop diffusion method. Before flash freezing in liquid nitrogen, crystals were soaked in cryo-protectant made of crystallization solution supplemented with 20% (v/v) glycerol. The X-ray diffraction data for the UHRF1 TTD-PBR complex were collected on the BL 5.0.3 beamline at the Advanced Light Source, Lawrence Berkeley National Laboratory. The diffraction data were indexed, integrated and scaled using the HKL2000 program35. The structure was solved using the molecular replacement method in PHASER36 with the structure of UHRF1 TTD (PDB ID: 3DB3) as search model. The resulting electron density revealed that there are three molecules of TTD and one PBR

peptide in each asymmetric unit. The structure of TTD-PBR complex was improved by iterative modeling building and refinement with Coot and PHENIX software packages. The same R-free test set was used throughout the refinement.

ITC measurements.

Histone H3 peptides containing residues 1-22 (H31-22), followed by a C-terminal tyrosine, were used for ITC assay. Protein or peptide samples were dialyzed against the ITC buffer (25 mM Tris-HCl, pH 7.5, 100 mM NaCl) at 4°C overnight. A MicroCal iTC200 system (GE Healthcare) was used to conduct the ITC measurements. For competitive binding assay, protein samples for UHRF1 domains were first mixed with unmodified H3 or H3K9me3 peptides in a molar ratio of 1:1, which were then titrated by the PBR peptide. A total of 17 injections with a spacing of 180 seconds and a reference power of 5 μcal/s were performed at 25°C. The ITC curves were processed with software ORIGEN (MicroCal) using one-site fitting model.

Peptide pull-down assay.

40 μL streptavidin paramagnetic beads (MagneSphere, Promega) were washed three times with 1x PBS (pH 7.4) buffer and resuspended in 100 μL PD buffer (25 mM Tris, pH 7.5, 300 mM KCl, 20% glycerol, 1 mM DTT and 0.2% v/v Triton X-100). To pull down UHRF1 wild type and R649A/P656G mutant with histone peptides, biotin-labeled $H3_{1-20}$K9me3 or unmodified $H3_{1-20}$, recombinant UHRF1 wild type or R649A/P656A mutant were mixed with the beads to a final concentration of 2.0 μM each and incubated at 4 °C for 1 hr. The beads were washed five times with PD buffer, followed by releasing the proteins into elution buffer (50 mM Tris, pH7.5, 25 % v/v glycerol, 1 mM EDTA, 2% SDS,

10 mM DTT) by boiling for 10 min. All the samples were finally resolved by SDS-PAGE and stained by Sypro Ruby (Bio-rad).

***In vitro* FRET measurements.**

For FRET assay, CFP-UHRF1-YFP wild-type and mutant proteins were adjusted to 0.2 μM, titrated with increasing amount of a hemimethylated DNA duplex (upper strand: 5'-GGCCCXGCAGGCCTG-3'; lower strand: 5'-CAGGCCTGCGGGGCC-3'; X = 5-methylcytosine), USP7 UBL domains or histone peptides. Assays were performed in Take3-plate (#259913 BioTek Instruments) in a buffer containing 25 mM Tris (pH 7.5), 100 mM NaCl, 1 mM Tris(2-carboxyethyl)phosphine (TCEP). Emission intensities were scanned using a SynergyMX microplate reader (#255439 BioTek Instruments) from 470 to 600 nm, with 450 nm as the excitation wavelength. The FRET efficiency ratio was calculated as fluorescence intensity at 525 nm divided by the sum of fluorescence intensities at 480 nm and 525 nm. Emission spectra were normalized to the sum of the fluorescence peak intensities at 480 nm and 525 nm. All assays were performed in duplicate, and the data were processed and normalized using Origin version 7.0 (OriginLab).

Results

Crystal structure of the UHRF1 TTD – PBR complex.

In order to provide molecular basis of TTD-PBR interaction, we determined the crystal structure of the TTD domain of zebrafish UHRF1 (zUHRF1$_{TTD}$) (Fig. 1A) in complex with the PBR peptide from human UHRF1 at 1.68 Å resolution (Fig. 1B, C and Table 1). In each asymmetric unit, there are three TTD domains and one PBR (Fig. S1). Most residues (from K648-R664) of PBR are bound with one TTD molecule predominantly, with its C-terminal tail (from S661-R664) contacting with another TTD molecule, leaving the third TTD unbound due to crystal packing effects (Fig. S1A). The TTD contains two tudor domains, both of which are dominated by a 'Royal' fold with 5-stranded β-barrel. (Fig. 1B). Notably, the linker sequence between the two subdomains joins in antiparallel with the C-terminal tail, together help integrate TTD into structural unit (Fig. 1B). The PBR peptide fills in an acidic groove extending from the subdomain interface, followed by a 90° turn, toward one end of the N-terminal β-barrel (Fig. 1B and Fig. S1B). The interaction between PBR and TTD is formed by a set of hydrogen bonds and hydrophobic interaction that spans into two binding surfaces: one involves PBR K648-G654, while the other involves PBR P656-G660 (Fig. 1B,D,E). In the first interface, PBR K650, S651, A652 and G654 interact with residues R227, G228, Y229 and W230 from the C-terminal subdomain hydrogen bonding interactions, while the side chains of PBR K648 and R649 interact with residues D144, E155 and D182 from the TTD N-terminal subdomain through direct and/or water-mediated hydrogen bonding interactions (Fig. 1D,E). At the second interface, the pyrrolidine of PBR P656 is inserted in the "aromatic cage" formed by

residues F154, Y180 and Y183, and the carbonyl group on P656 backbone interacts with the side chain of TTD N186 (Fig. 2C,D,E). Additional intermolecular interactions involve hydrogen bonds formed between PBR S657 and R658 and a number of TTD residues (D147, N148, Y180 and N186), which further supports the TTD-PBR association (Fig. 1D,E). High residues similarity of interacting sites on TTD across species validates zUHRF1$_{TTD}$ as a model for its human orthologue (Fig. S2).

Mutational analysis of the TTD-PBR interaction.

From our crystal structure of the TTD-PBR complex, we observed that the interaction of TTD-PBR involves a number of residues that contribute the specific interaction. To test our structural results, we introduced mutations to the binding sites both on human TTD and PBR. Then we performed Isothermal Titration Calorimetry (ITC) to quantify the binding affinity of the TTD and PBR interaction (Table S1). The dissociation constant (Kd) for the wild-type TTD and PBR was measured to be 7.4±0.68 μM (Figure 2A,B). In comparison with the wild-type, the binding affinity of TTD D142 (equivalent to D144 in zUHRF1), D145 (equivalent to D147 in zUHRF1), F152 (equivalent to F154 in zUHRF1) and E153 (equivalent to E155 in zUHRF1) mutations to wild-type PBR has decreased by 37-, 3.2-, 12- and 38-fold, respectively (Figure 2A). On other hand, the binding affinity of wild-type TTD to PBR R649A and P656G mutations decreased by 20-, 4- , respectively (Figure 2B). Furthermore, to abolish the interaction, TTD D142A/E153A, PBR R649A/P656G double mutants were constructed and subjected to ITC. The dissociation constant (Kd) between the TTD D142A/E153A double mutant and wild-type PBR was measured to be 534±71.9 μM (Figure 2A), which shows 72-fold reduction compare to the wild-type. Similarly, the

PBR R649A/P656G double mutant reduced the binding affinity by 40-fold, with a dissociation constant (Kd) of 301±25.6 μM (Figure 2B). Together, these data lend a strong support to the structural observation of the TTD-PBR complex. Mutation of either N147 (equivalent to N149 in zUHRF1) or N194 (equivalent to N186 in zUHRF1) into alanine did not lead to appreciable change in the TTD – PBR binding (Fig. S2C, D), suggesting that the intermolecular interactions involving these sites likely arise from the crystal packing effects.

The TTD-PBR interaction perturbs the conformation and H3K9me3 binding of the TTD-PHD dual domains.

Previous research has found that TTD and PHD domains serve as an integrated histone-binding cassette [1, 2]. We want to investigate whether the TTD-PBR interaction affects the conformation of TTD-PHD dual domains and also H3K9me3 binding. We compared our TTD-PBR complex structure with the previously determined structure of UHRF1 TTD-PHD tandem domains in complex with histone H3K9me3 peptide and found that the TTD domains are well aligned (Fig. 3A), with a RMSD of 0.73 Å over 122 Cα atoms, suggesting the structural rigidity of the UHRF1 TTD domain. In addition, we found that the PBR-binding sites significantly overlap with the binding surface for the PHD-TTD linker or the H3K9me3 peptide (Fig. 3A). In particular, the N-terminus of PBR (residues K648-S651) superimposes well with the TTD-PHD linker (residues R295-S298), a region that facilitates the UHRF1-H3K9me3 association through regulation of the TTD-PHD domain orientation [1, 2]. As the backbone and side chain conformations of these two segments are similar (Fig. 3B), they might compete each other for the TTD binding site, suggesting that the

TTD-PBR association would lead to displacement of the TTD-PHD domain linker. Meanwhile, the structural superimposition of the TTD-PBR and TTD-PHD-H3K9me3 complexes shows that the PBR peptide-binding site also partially overlaps with the H3K9me3-binding site, with residue P656 occupying the H3K9me3-engaging aromatic cage of the TTD domain (Fig. 3C,D). In summary, these structural observations have provided some new insights that the TTD-PBR intramolecular interaction may affect the UHRF1-H3K9me3 binding through a two-layered mechanism: First, binding of PBR to the TTD domain dislodges the TTD-PHD linker from the TTD surface, causing a conformational change of the TTD-PHD dual domains that would in turn block the UHRF1-H3K9me3 binding. Second, PBR can directly occupy the H3K9me3-binding pocket with a proline, thus inhibiting the binding of H3K9me3 to TTD. Further, ITC assays consistently validated the structural observations. Binding of the PBR peptide with TTD-PHD is weaker than with TTD alone by ~3 fold (Fig. 3E). In addition, the presence of unmodified H3 and H3K9me3 peptides both further impairs the binding of PBR with TTD-PHD, with a relatively more severe effect associated with the $H3_{1-22}$K9me3 peptide (Fig. 3E). However, the presence of the same excess of P656G PBR has a more mild effect on the TTD-PHD – $H3_{1-22}$K9me3 binding (Fig. S2E).

Competing intramolecular and intermolecular interactions govern the conformational states of UHRF1.

To trace the dynamic transition between multiple conformational stages of UHRF1 governed by intramolecular and intermolecular interactions, we conduct the *in vitro* fluorescence resonance energy transfer (FRET) analysis (Fig. S3). We generated a CFP-

UHRF1-YFP fusion construct, in which full-length human UHRF1 was N- and C-terminally fused to CFP and YFP proteins, respectively. Wild-type CFP-UHRF1-YFP fusion protein presents a high-FRET signal (increased acceptor signal associated with decreased donor signal), with IAcceptor/(IDonor+IAcceptor) ratio of 0.6, indicating that the FRET paired proteins CFP and YFP, separated by the ~800 amino acid long full-length UHRF1 protein, are spatially close enough to generate a high FRET signal. Then we introduced the D142A/E153A double mutation that can disrupt TTD-PBR interaction (Fig. 2B). As we expected, the FRET ration of this mutant protein has decreased to 0.52, suggesting that disruption of the TTD-PBR interactions leads to a more open conformational state for UHRF1 (Fig. 4A). When adding the $H3_{1-22}$K9me3 peptide gradually into the wild-type protein can also open up the conformation that leading to decreased FRET, but has no effect for D142A/E153A-mutated CFP-UHRF1-YFP fusion protein. It is noting that D142A/E153A double mutation not only disrupts TTD-PBR intramolecular interaction, but also reduces the $H3_{1-22}$K9me3-binding affinity of UHRF1 TTD-PHD dual domains by ~4 fold (Fig. S4). However, titration of the unmodified $H3_{1-22}$ peptide did not lead to significant change of FRET signals for both wild-type and D142A/E153A-mutated CFP-UHRF1-YFP fusion protein, suggesting the important role of the H3K9me3-binding cage in the TTD-PBR interaction (Fig. 4A). Besides, Furthermore, titration with hemimethylated DNA resulted a strong FRET signals reduction for both wild-type and D142A/E153A-mutated CFP-UHRF1-YFP (Fig. 4A), suggesting that the interaction of hemimethylated DNA with SRA and PBR synergistically contributes to the conformational “opening” of UHRF1. Finally, we measured the effect of the

intermolecular interaction between UHRF1 PBR and USP7 UBL domains, which were previously shown to allosterically regulate the chromatin association of UHRF1 through interaction with PBR (Fig. S5) [10], on the FRET efficiency of CFP-UHRF1-YPF fusion proteins (Fig. S3). Similar to hemimethylated DNA titration, USP7 UBL domains reduced the FRET efficiency for both wild-type UHRF1 fusion protein and D142A/E153A UHRF1 fusion protein (Fig. 4A,B), confirming the PBR-mediated intermolecular interactions synergize with the UHRF1 TTD-H3K9me3 binding to "open up" the conformation of UHRF1. Next we wanted to ask whether the conformational transition of UHRF1 lead to an alteration of its histone binding affinity. We introduced R649A/P656G mutations located on PBR which can "open" UHRF1 conformation through disrupting its intramolecular interaction with TTD, to perform H3 peptide pull-down assays. Compared with wild-type UHRF1, R649A/P656G-mutated UHRF1 bound stronger to H3K9me3 (Fig. 4B), suggesting that disruption of the TTD-PBR intramolecular interaction facilitates the chromatin association of UHRF1.

Discussion

UHRF1 plays an essential role in DNA methylation maintenance, in which it can regulate the target recruitment of DNMT1 through specific association with replicating heterochromatin enriched with H3K9me3 and hemi-methylated DNA [12-14]. Binding to both H3K9me2/3 and hemi-methylated DNA ensures high fidelity DNA maintenance methylation, thus connecting H3K9 methylation and DNA methylation in maintenance DNA methylation [6]. It was established that the intramolecular interaction between the TTD domain and PBR of UHRF1 leads to a "closed" conformation, which occludes its chromatin association. Intermolecular interactions of the PBR with hemimethylated DNA, DNMT1, USP7 or PI5P "open up" the UHRF1 conformation, promoting the TTD – H3K9me3 mediated heterochromatin association, as well as UHRF1 E3 ligase activity targeting H3 ubiquitylation. [7-10] In addition to the previous NMR structure of TTD-PBR [7], our high-resolution crystal structure provides a detailed information of this intramolecular interaction. Combined with mutagenesis analysis and biochemical characterizations, we demonstrated that the intramolecular TTD-PBR interaction of UHRF1 competes against the intermolecular association between UHRF1 TTD and H3K9me3 through both repositioning of the TTD-PHD linker and blockage of the H3K9m3-binding cage, thereby providing a more comprehensive view on the conformational regulation of UHRF1. Disruption of TTD-PBR interaction by either direct mutagenesis or binding other factors helps expose the H3K9me3-binding pocket and arrangement of TTD-PHD linker, thus promotes H3K9me3 association of UHRF1, which is important for DNMT1-mediated maintenance DNA methylation. (Fig. 4C) Together, our

study provides the molecular mechanism of conformational transition of UHRF1 regulated by PBR mediated intra-/inter- molecular interaction, highlighting its central role for allosteric regulation of UHRF1.

Figures and Table

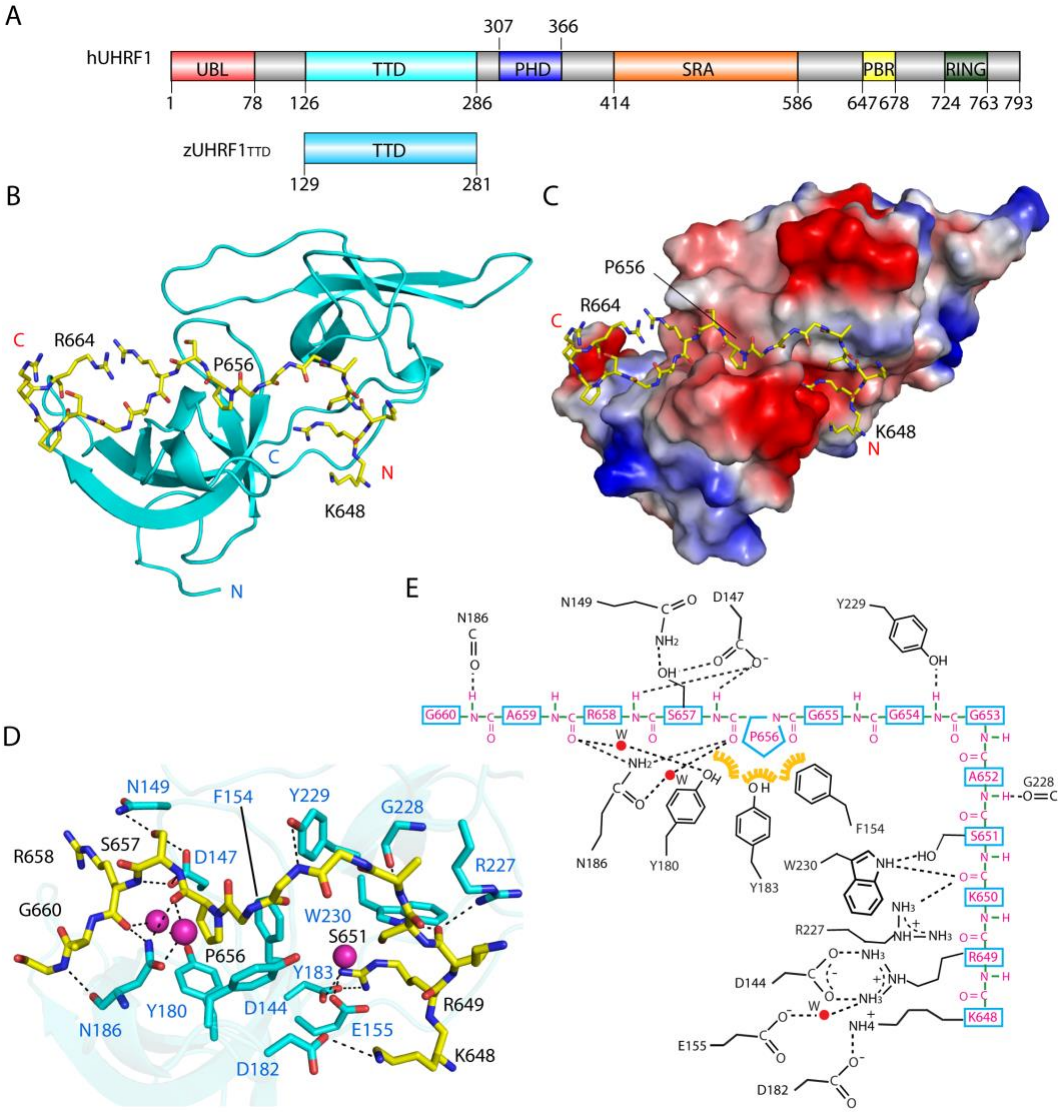


Figure 1. Crystal structure of the UHRF1 TTD – PBR complex. (A) Domain architecture of human UHRF1, with individual domains labeled with residue numbers. The TTD domain of Zebrafish UHRF1 (zUHRF1TTD) used for structure determination is also shown. (B) Ribbon representation of zUHRF1TTD (cyan) bound to the PBR peptide (yellow sticks). (C) Surface representation of the zUHRF1TTD – PBR complex. (D) Close-up view of the intermolecular interaction between zUHRF1TTD and PBR. The water molecules are shown in sphere representation. The hydrogen bonds are shown as dashed lines. (E) Schematic representation of the zUHRF1TTD – PBR interaction. The residues from zUHRF1TTD and PBR are colored in magenta and black, respectively. Yellow: hydrophobic contact. See also Figure S1.

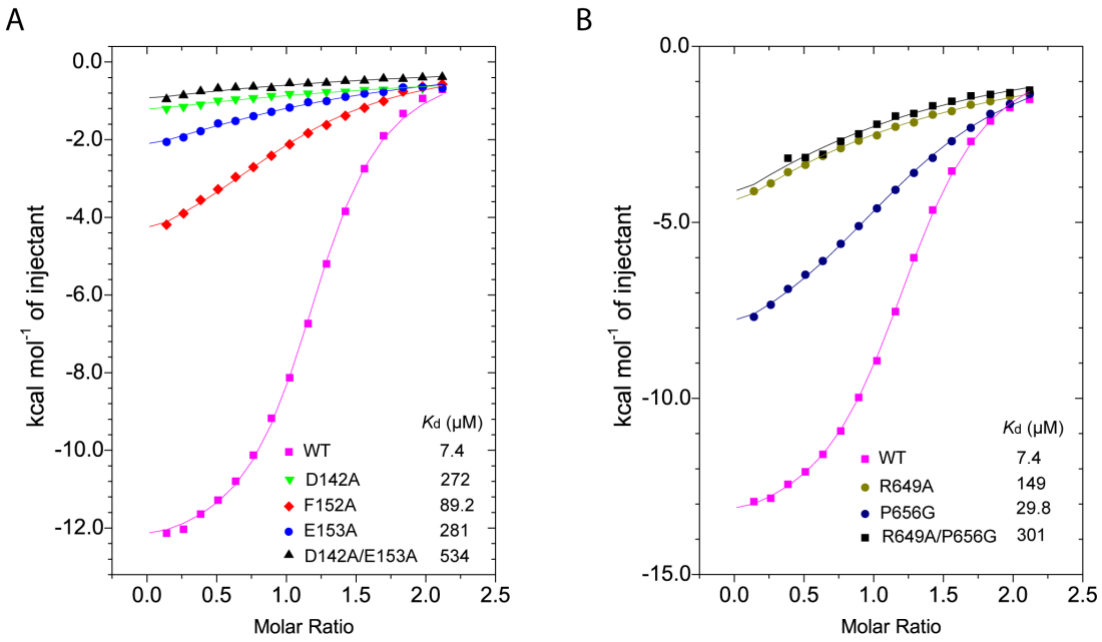
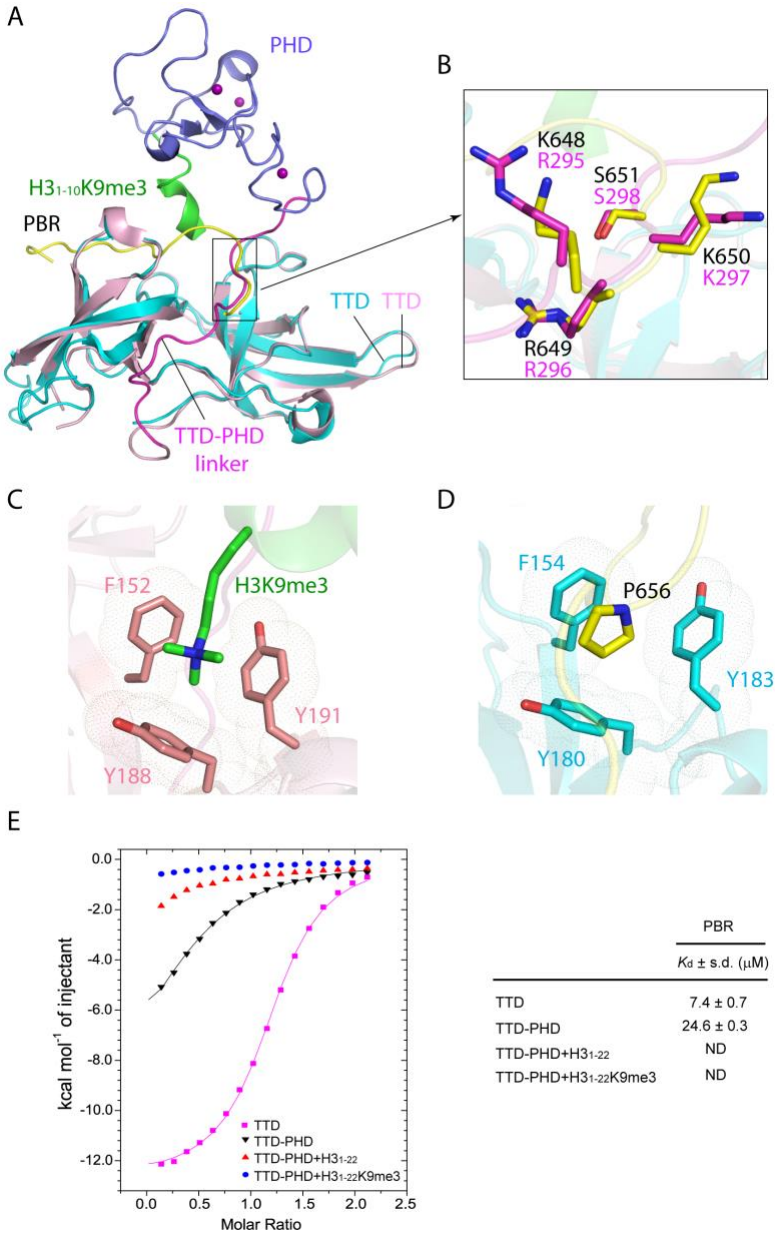


Figure 2. ITC binding assays for human UHRF1 TTD and PBR. (A) Mutational effects of the UHRF1 TTD domain on the TTD – PBR interaction. (B) Mutational effects of the UHRF1 PBR peptide on the TTD – PBR interaction. See also Table S1.



	PBR
	K_d ± s.d. (μM)
TTD	7.4 ± 0.7
TTD-PHD	24.6 ± 0.3
TTD-PHD+H3 1-22	ND
TTD-PHD+H3 1-22 K9me3	ND

Figure 3. UHRF1 PBR and histone H3K9me3 compete on TTD binding. (A) Structural comparison of the UHRF1 TTD – PBR complex with the UHRF1 TTD-PHD – H3K9me3 complex (PDB 4GY5), with individual domains and peptides labeled. The zinc ions are shown as purple spheres. (B) Structural overlay the TTD-bound PBR peptide (yellow) and the TTD-PHD linker (magenta), with the aligned residues shown in stick representation. (C) The side chain of H3K9me3 inserted into the aromatic cage formed by residues F152, Y188 and Y191 of human UHRF1 TTD (PDB 4GY5). (D) The side chain of P656 embedded in the equivalent aromatic cage formed by residues F154, Y180 and Y183 of zUHRF1$_{TTD}$. (E) ITC binding curves of the PBR peptide over the TTD domain or the TTD-PHD dual domains of human UHRF1, in the absence or presence of unmodified H3$_{1-22}$ or H3$_{1-22}$K9me3 peptides. The average and standard deviation of the dissociation constants were derived from two independent measurements. ND, not determinable due to undetectable or non-stoichiometric binding. See also Figure S1, S2 and Table S1.

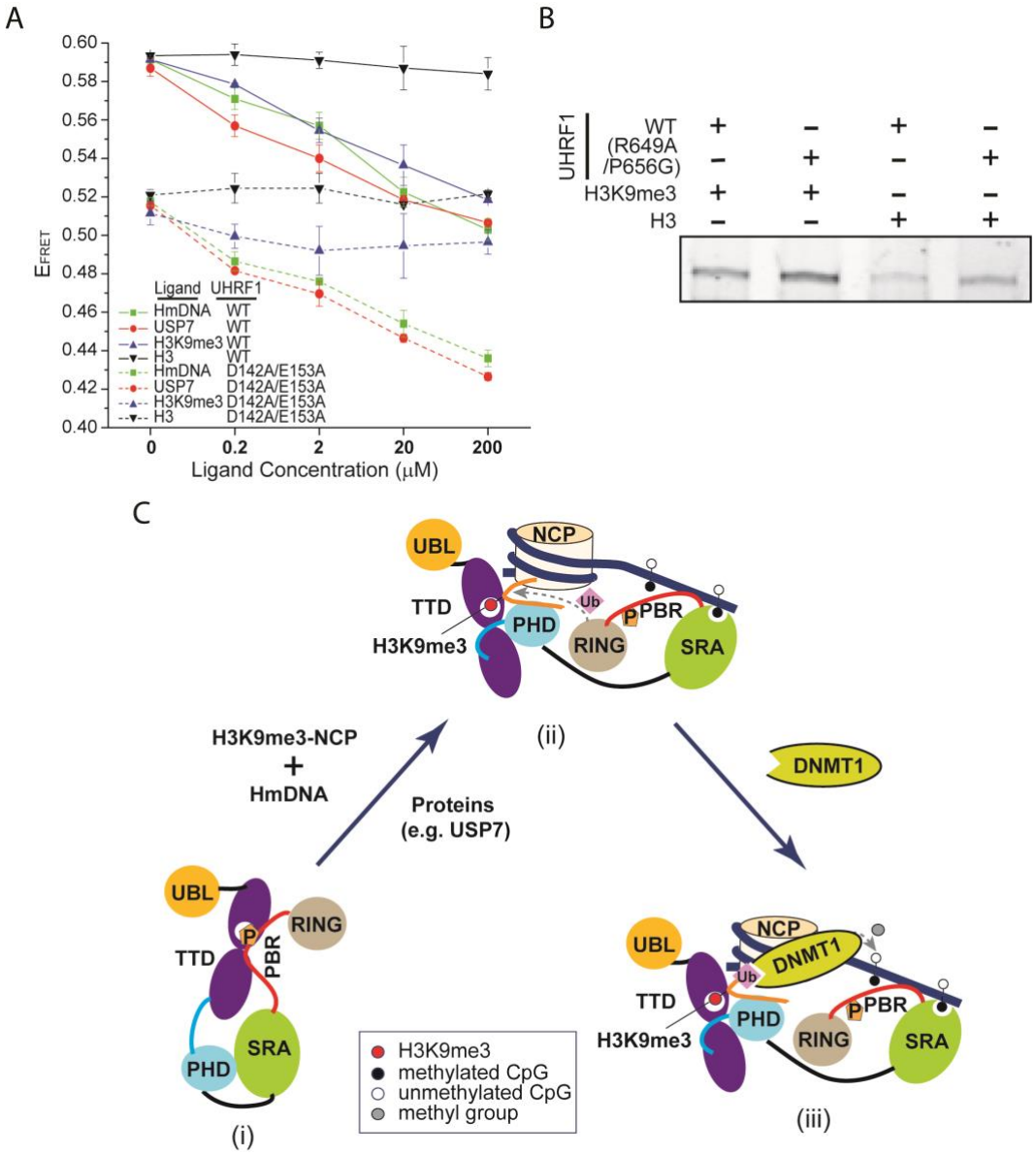


Figure 4. Coupling between the conformational transition of UHRF1 and its chromatin or USP7 binding. (A) FRET ratio, $E_{FRET} = I_{acceptpor}/(I_{acceptpor}+I_{donor})$, of CFP-UHRF1-YFP as function of the concentrations of hemimethylated DNA (HmDNA), USP7 UBL domains (USP7), $H3_{1-22}K9me3$ (H3K9me3) and $H3_{1-22}$ (H3) peptides. (B) Pull-down assays of wild-type (WT) or R649A/P656G-mutated UHRF1 (residues 126-793) with H3K9me3 or unmodified H3 peptides. The gel image cropped from a full gel is boxed. (C) A model for the conformational transition of UHRF1. (i) UHRF1 in free state is dominated by a "closed" conformation. (ii) Association of UHRF1 with hemimethylated DNA (HmDNA) and H3K9me3-modified nucleosome (H3K9me3-NCP) transits UHRF1 into an "open" conformation, permitting strong chromatin association and enhanced H3 ubiquitylation activity. (iii) Ubiquitylated histone H3 subsequently recruits DNMT1 for maintenance DNA methylation. See also Figure S3, S4 and S5.

Table 1. Crystallographic data collection and refinement statistics of the zUHRF1^{TTD} – PBR complex.

Data collection	
Space group	*C2*
Cell dimensions	
a, *b*, *c* (Å)	81.2, 66.1, 120.5
α, β, γ (°)	90, 105, 90
Wavelength	0.9774
Resolution (Å)	28.76-1.68 (1.74-1.68)[a]
R_{merge}	0.038 (0.47)
$I / \sigma I$	27 (2.0)
Completeness (%)	98.4 (96.4)
Redundancy	3.0 (2.9)
Total reflections	208438
Unique reflections	69241
Refinement	
Resolution (Å)	28.76-1.68 (1.74-1.68)[a]
No. reflections	69239 (6738)
R_{work} / R_{free}	18.3/20.7 (27.0/29.4)
No. atoms	
Protein	3777
Water	535
B factors (Å^2)	
Protein	41.54
Water	47.40
r.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.14
Ramachandran	
Favored (%)	97.4
Allowed (%)	2.6
Outliers (%)	0

[a]Values in parentheses are for highest-resolution shell.

Supplementary Data

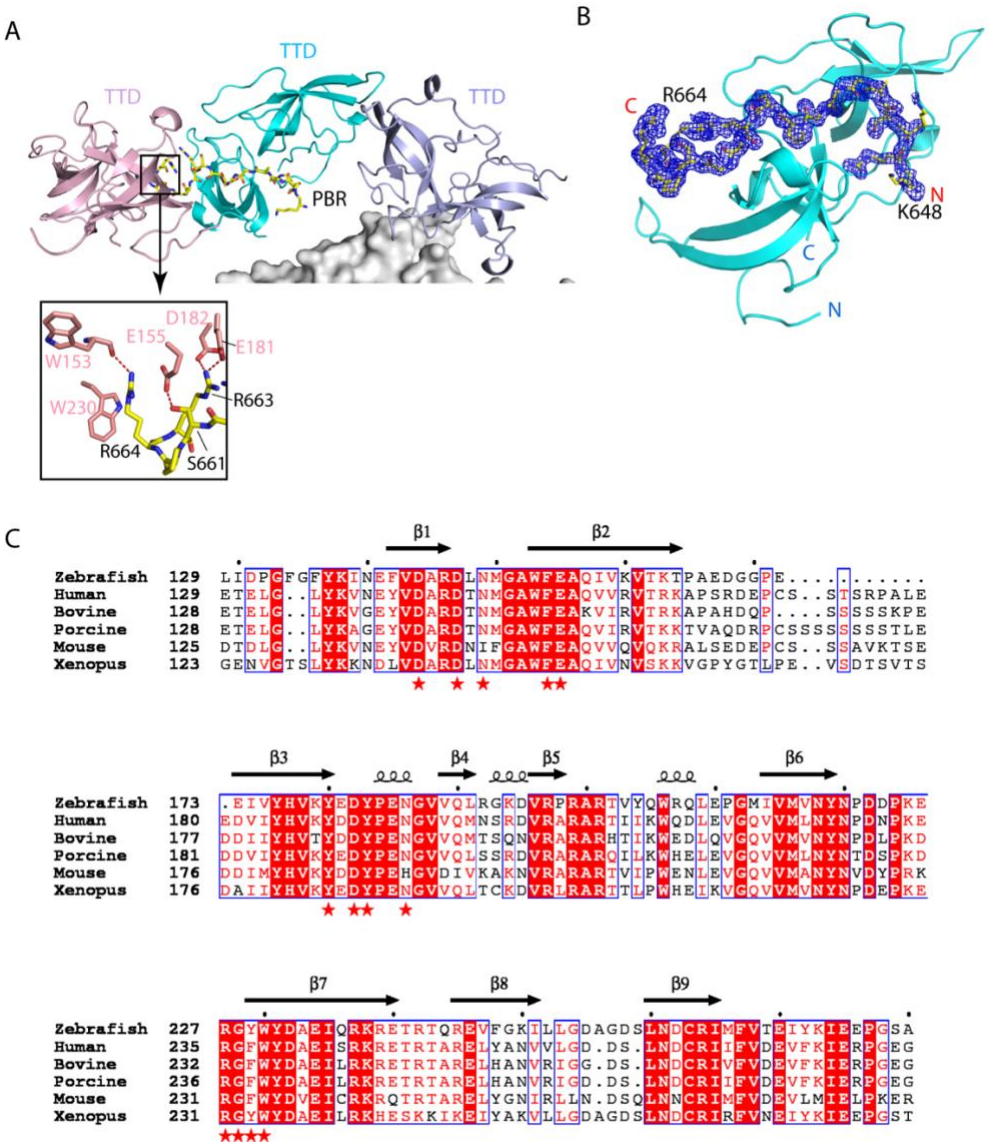


Figure S1. Crystallographic and sequence analyses of the UHRF1 TTD – PBR complex, related to Fig. 1 and Fig. 3. (A) Molecular components in one asymmetric unit (ASU). The three TTD domain molecules are colored in pink, cyan and light blue, respectively. The PBR peptide is shown in yellow sticks. The intermolecular interactions mediating the crystal packing of PBR with one of the TTD molecules are shown in the expanded view. The hydrogen bonds are sown as dashed lines. The crystallographic symmetry-related TTD molecules are shown in surface representation. (B) 2Fo-Fc omit map of PBR peptide bound to zUHRF1$_{TTD}$. The 2Fo-Fc omit map of PBR peptide was shown as blue mesh, contoured at 1σ level. The zUHRF1$_{TTD}$ is colored in cyan. (C) Sequence alignment of UHRF1 TTD domains across different species. The secondary structures of the TTD domain are indicated above the sequence. The PBR binding sites are marked with red asterisks.

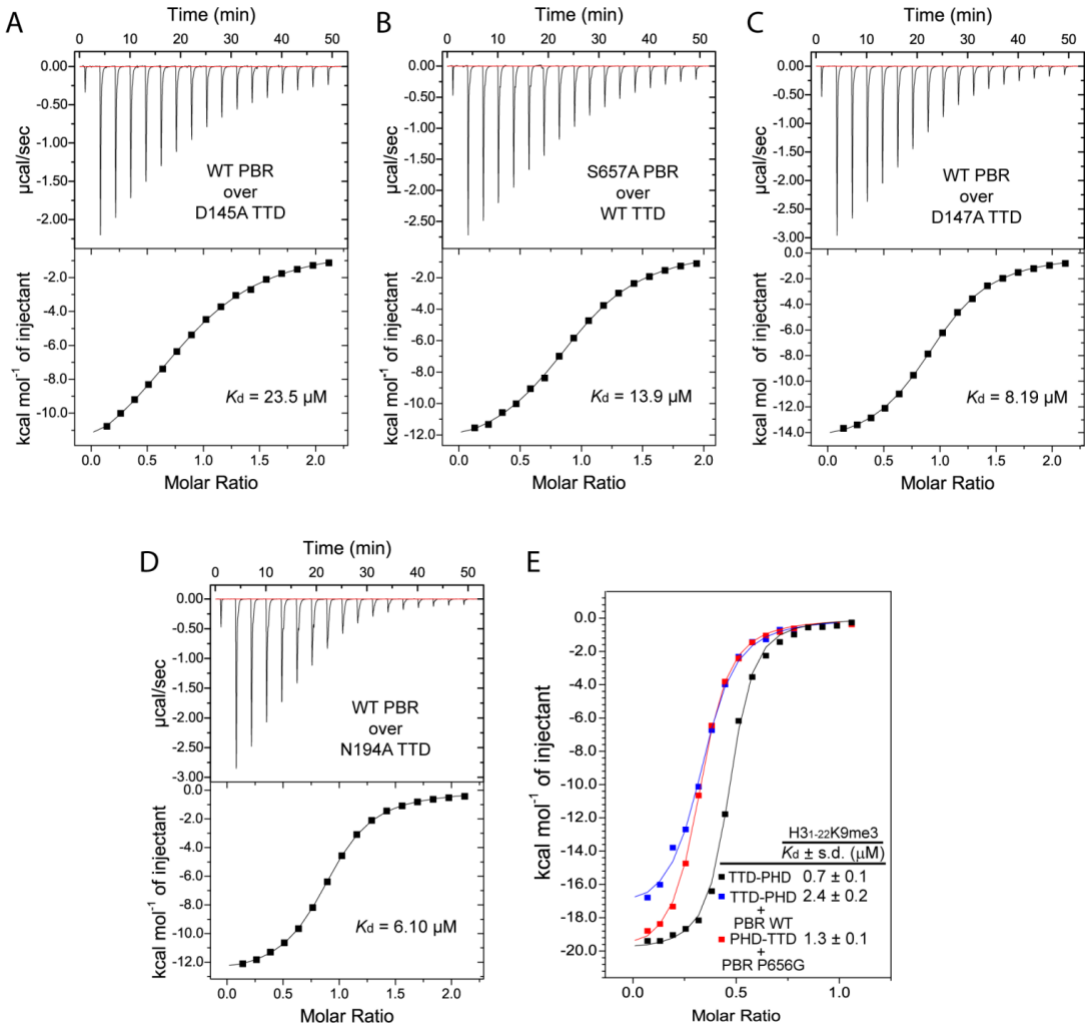


Figure S2. ITC analysis of the UHRF1 TTD – PBR interaction, related to Fig. 1 and Fig. 3. (A) Titration of wild-type (WT) PBR over D145A TTD. (B) Titration of S657A PBR over WT TTD, (C) Titration of WT PBR over D147A TTD. (D) Titration of WT PBR over N194A TTD. Related to Fig. 2. (E) ITC binding curves of the $H3_{1-22}K9me3$ peptide over the UHRF1 TTD-PHD dual domains, in the absence or presence of WT PBR or P656G-mutated PBR. The average and standard deviation of the dissociation constants were derived from two independent measurements.

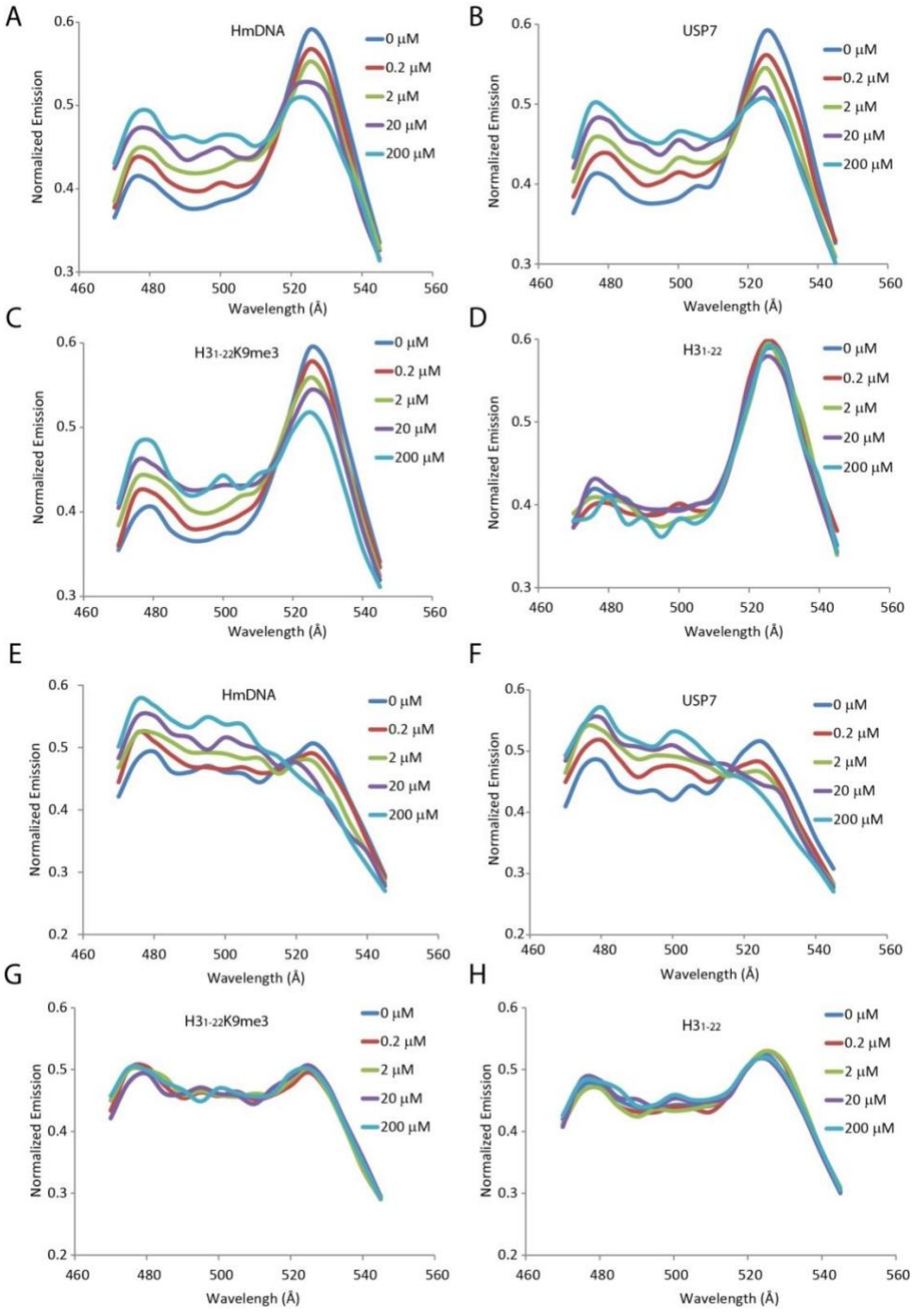


Figure S3. FRET analysis of UHRF1, wild type or D142A/E153A, titrated with different ligands, related to Fig. 4A. (A-D) FRET analysis of wild-type UHRF1 in the presence of increasing amount of hemimethylated DNA (HmDNA) (A), USP7 UBL domains (B), H3$_{1-22}$K9me3 (C) and unmodified H3$_{1-22}$ (D). (E-H) FRET analysis of D142A/E153A UHRF1 in the presence of increasing amount of hemimethylated DNA (HmDNA) (E), USP7 UBL domains (F), H3$_{1-22}$K9me3 (G) and unmodified H3$_{1-22}$ (H).

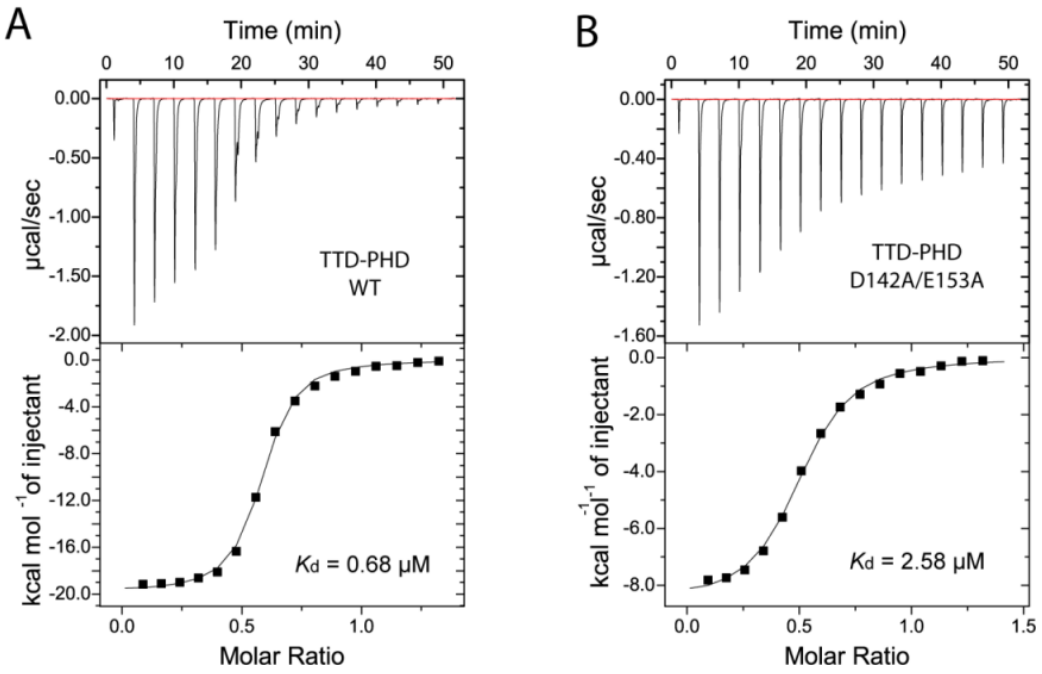


Figure S4. Effect of the D142A/E153A mutation on the UHRF1 TTD-PHD – H3K9me3 interaction, related to Fig. 4A. ITC binding assays of the H3$_{1-22}$K9me3 peptide over (A) wild type (WT) UHRF1 TTD-PHD and (B) D142A/E153A-mutated UHRF1 TTD-PHD.

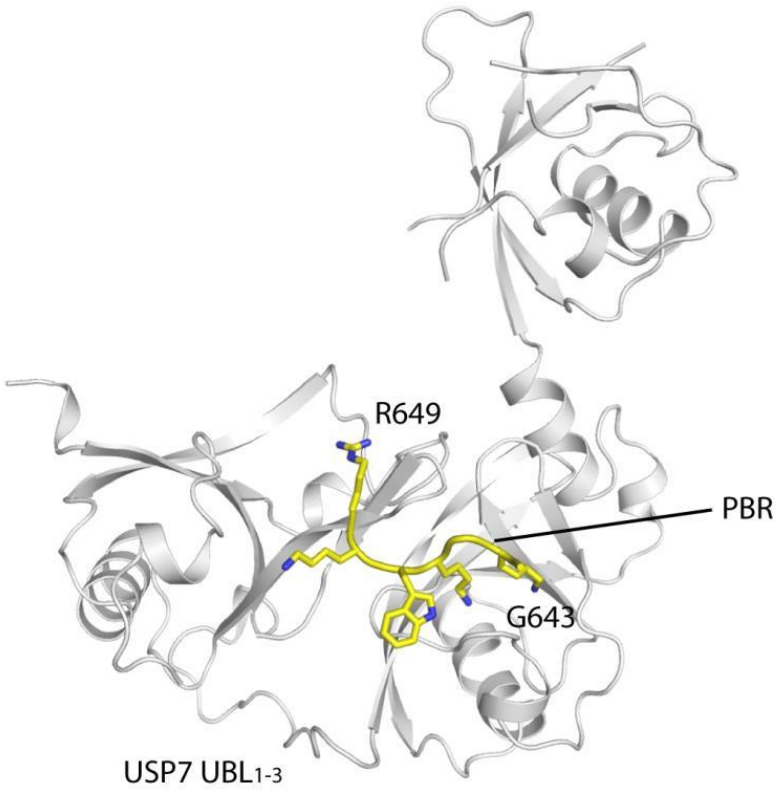


Figure S5. Structure of USP7 UBL domains bound to a PBR peptide, related to Fig. 4A. The first three UBL domains of USP7 (UBL_{1-3}) is colored in grey. The PBR peptide is shown in yellow sticks.

Table S1. Summary of ITC binding parameters, related to Fig. 2 and Fig. 3E.

Syringe / Cell	K_d (μM)	N
WT PBR / WT TTD		
	7.4±0.7	1.0±0.1
WT PBR / TTD mutants		
D142A	272±19	1.0*
D145A	23.5±0.6	0.8±0.1
N147A	8.2±0.0	0.9±0.1
F152A	89.2±13.9	1.3±0.1
E153A	281±11.2	1.0*
D142A/E153A	534±72	1.0*
N194A	6.1±0.2	0.9±0.1
PBR mutants / WT TTD		
R649A	149±1	1.2±0.1
P656G	29.8±4.7	1.3±0.1
S657A	13.9±0.2	0.9±0.1
R649A/P656G	301±26	1.0*
WT PBR / TTD-PHD		
TTD-PHD	24.6±0.3	0.6±0.1
TTD-PHD+$H3_{1-22}$	ND	ND
TTD-PHD+$H3_{1-22}$K9me3	ND	ND
$H3_{1-22}$K9me3 / TTD-PHD		
TTD-PHD	0.7±0.1	0.6±0.1
TTD-PHD+WT PBR	2.4±0.2	0.3±0.1
TTD-PHD+ P656G PBR	1.3±0.1	0.3±0.1
D142A/E153A TTD-PHD	2.6±0.1	0.5±0.1

* Value was set manually. WT, wild type. ND, not determinable due to undetectable or non-stoichiometric binding.

Reference

1. Arita, K., et al., *Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1.* Proc Natl Acad Sci U S A, 2012. **109**(32): p. 12950-5.

2. Cheng, J., et al., *Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein.* J Biol Chem, 2013. **288**(2): p. 1329-39.

3. Arita, K., et al., *Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism.* Nature, 2008. **455**(7214): p. 818-21.

4. Avvakumov, G.V., et al., *Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1.* Nature, 2008. **455**(7214): p. 822-5.

5. Hashimoto, H., et al., *The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix.* Nature, 2008. **455**(7214): p. 826-9.

6. Liu, X., et al., *UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9.* Nat Commun, 2013. **4**: p. 1563.

7. Fang, J., et al., *Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition.* Nat Commun, 2016. **7**: p. 11197.

8. Gelato, K.A., et al., *Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate.* Mol Cell, 2014. **54**(6): p. 905-19.

9. Harrison, J.S., et al., *Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1.* Elife, 2016. **5**.

10. Zhang, Z.M., et al., *An Allosteric Interaction Links USP7 to Deubiquitination and Chromatin Targeting of UHRF1.* Cell Rep, 2015. **12**(9): p. 1400-6.

11. Ma, H., et al., *M phase phosphorylation of the epigenetic regulator UHRF1 regulates its physical association with the deubiquitylase USP7 and stability.* Proc Natl Acad Sci U S A, 2012. **109**(13): p. 4828-33.

12. Bostick, M., et al., *UHRF1 plays a role in maintaining DNA methylation in mammalian cells.* Science, 2007. **317**(5845): p. 1760-4.

13. Miura, M., et al., *Dynamic changes in subnuclear NP95 location during the cell cycle and its spatial relationship with DNA replication foci.* Exp Cell Res, 2001. **263**(2): p. 202-8.

14. Sharif, J., et al., *The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA.* Nature, 2007. **450**(7171): p. 908-12.

Chapter 2

Structural basis of DNMT3B-mediated *de novo* DNA methylation

Abstract

Mammalian DNA methylation patterns are established by two *do novo* DNA methyltransferases, DNMT3A and DNMT3B, which exhibit overlapping, yet distinct DNA methylation activities across different developmental stages and tissues. How their activities diverge remains elusive. Here we reported the crystal structures of DNMT3B-DNMT3L tetramer in complex with CpG and non-CpG DNAs, respectively. Combined with enzymatic analysis, we demonstrated that the hydrogen bonding interaction between N656 and R661 located on the catalytic loop causes a lower CpG specificity than DNMT3A, while the interplay of target recognition domain and homodimeric interface fine-tunes the distinct target selection between the two enzymes, with K777 from DNMT3B TRD loop recognizing +1 flanking site. Together, this study provides a mechanistic understanding of DNMT3B-mediated *de novo* DNA methylation.

Introduction

Establishment of DNA methylation is mainly achieved by *de novo* DNA methyltransferases DNMT3A and DNMT3B during gametogenesis and postimplantation development[1], regulated by DNMT3-like protein (DNMT3L)[2-4]. In addition to CpG methylation, DNMT3A and DNMT3B also catalyze non-CpG (CpH; H = A, C or T) methylation in oocytes, embryonic stem (ES) cells and neural cells[5-7]. Non-CpG methylation is associated with transcriptional repression [8, 9] and the pluripotent epigenetic state[6, 10]. DNMT3A and DNMT3B share the same domain architecture arrangement: PWWP domain, ADD domain and C-terminal methyltransferase domain with high sequence similarity (Fig. S1) , which suggests that DNMT3A and DNMT3B have partial functional redundancy in *de novo* methylation across the genome[1, 11]. On the other hand, more and more studies have demonstrated their non-overlapping functions [1, 12-18]. For example, DNMT3A and DNMT3B show distinct enzymatic properties [17], with DNMT3B associated with a higher non-CpG methylation activity both *in vitro* and *in vivo* [12, 19]. Notably, DNMT3B-mediated CpH methylation primarily occurs in the context of CAG motif in ES cells [6, 20], whereas DNMT3A-mediated CpH methylation shows preference toward the CAC motif in neurons [21, 22]. Dysregulation of DNMT3A or DNMT3B contributes to various human diseases [23]. In particular, mutation of DNMT3B has been linked to ICF (immunodeficiency, centromeric instability, facial anomalies) syndrome, characterized by immunological defects, centromeric heterochromatin instability, and facial anomalies [24], whereas DNMT3A mutations are prevalent in acute myeloid leukemia (AML) and other hematological cancers [25, 26].

Previous studies have indicated subtle mechanistic differences between DNMT3A and DNMT3B [12, 15, 16, 18, 27], including their differential preference toward the flanking sequence of CpG target sites[28-31]. However, due to the limited number of different substrates investigated in these studies, global differences in substrate recognition of DNMT3A and DNMT3B remain elusive.

Our recently reported crystal structure of the DNMT3A-DNMT3L heterotetramer in complex with CpG DNA[32] revealed that the two central DNMT3A subunits bind to the same DNA duplex through a set of interactions mediated by protein motifs from the target recognition domain (TRD), the catalytic core and the DNMT3A-DNMT3A homodimeric interface (also called RD interface below). However, the structural basis of DNMT3B-mediated methylation remains unclear. To gain further understanding of the DNMT3B-mediated *de novo* methylation, we reported the crystal structures of the complexes of the DNMT3B-DNMT3L tetramer with CpG and CpA DNA, respectively. DNMT3B recognizes the target sites through a mechanism distinct from that of DNMT3A, underpinning its relative higher non-CpG preference. Together, this study provides a mechanistic view on DNMT3B-mediated methylation, which impacts the establishment of specific CpG and non-CpG methylation patterns in cells.

Materials and Methods

Protein expression and purification.

MTase domain (residues 562–853) of human DNMT3B (NCBI accession NM_006892) or the MTase domain (residues 628–912) of human DNMT3A (NCBI accession NM_022552) was co-expressed with the C-terminal domain (residues 178–386) of human DNMT3L (NCBI accession NM_175867) on a modified pRSFDuet-1 vector (Novagen), in which the DNMT3B or DNMT3A sequence was preceded by a hexahistidine (His_6) and SUMO tag. The *Escherichia coli* BL21 DE3 (RIL) cell strains transformed with the DNMT3A- or DNMT3B-expression plasmids were first grown at 37 °C, but shifted to 16 °C after induction by IPTG at OD_{600} (optical density at 600 nm) of 0.8. The cells continued to grow overnight. The His_6-SUMO–tagged DNMT3B or DNMT3A fusion proteins in complex DNMT3L were purified using a Ni^{2+}-NTA column. Subsequently, the His6-SUMO tag was removed through Ubiquitin-like-specific protease 1 (ULP1) cleavage, followed by ion exchange chromatography using a Heparin HP column (GE Healthcare) and further purification through size-exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) in a buffer containing 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.1% β-mercaptoethanol, and 5% glycerol. Purified protein samples were stored in -80°C for future use.

To generate the covalent protein–DNA complex, a 25-mer Zebularine-containing DNA (CGA: 5′-GCATGZGTTCTAATTAGAACGCATG-3′ or CGT: 5′-CATGZGATCTAAT TAGATCGCATGG-3′; CAG: 5′-CATGZAGTCTAATTAGACTGCATGG-3′; Z:

zebularine) was self-annealed and incubated with wild-type or mutant DNMT3B-DNMT3L in 20 mM Tris-HCl (pH 8), 20% glycerol, and 40 mM DTT at room temperature. The reaction products were further purified through a HiTrap Q XL column (GE Healthcare), followed by size-exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column. The purified covalent protein-DNA complexes were concentrated to about 0.1–0.2 mM using Ultracel®-10K Centrifugal Filters (Millipore) in a buffer containing 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.1% β-mercaptoethanol, and 5% glycerol.

Crystallization and structure determination.

The crystals of covalent complexes of wild-type or mutant DNMT3B-DNMT3L-DNA complexes were generated by the hanging-drop vapor-diffusion method at 4°C, from drops mixed from 0.5 μl of the protein solution and 0.5 μl of precipitation solution containing 0.1 M Tris-HCl (pH 8.0), 200 mM $MgCl_2$, 8% PEG4000 and 0.2 mM AdoHcy. Crystals for the DNMT3B–DNMT3L tetramer in complex with the CG2 were generated by the hanging-drop vapor-diffusion method at 16°C, from drops mixed from 0.5 μL of DNMT3B–DNMT3L–DNA solution and 0.5 μl of precipitant solution containing 0.1 M Tris-HCl (pH 8.0), 100 mM $MgCl_2$, 7% PEG8000 and 0.2 mM AdoHcy. All crystals were soaked in cryo-protectant made of the precipitation solution supplemented with 25% glycerol, before flash frozen in liquid nitrogen for X-ray data collection. The X-ray diffraction datasets for the DNMT3B–DNMT3L–DNA complexes were collected at selenium peak wavelength on the BL 5.0.1 or BL 5.0.2 beamlines at the Advanced Light Source, Lawrence Berkeley National Laboratory. The diffraction data were indexed, integrated, and scaled using the HKL 2000 program[33]. The structures of the complexes

were solved by molecular replacement method using PHASER[34], with the structure of DNMT3A–DNMT3L-DNA complex (PDB 5YX2) as a search model. Further modelling of the covalent DNMT3B–DNMT3L–DNA complexes was performed using COOT[35] and subjected to refinement using the PHENIX software package[36]. The same R-free test set was used throughout the refinement. The statistics for data collection and structural refinement of the productive covalent DNMT3B–DNMT3L–DNA complexes are summarized in Table S1.

***In vitro* DNA methylation assay**.

Synthesized (GAC)12, (AAC)12 and (TAC)12 DNA duplexes were used as CG-, CA- and CT-containing substrates, respectively. A 20-μL reaction mixture contained 0.75 μM DNA, 0.3 μM DNMT3B-DNMT3L or DNMT3A-DNMT3L tetramer, 2.5 μM S-adenosyl-L-[methyl-^{3}H]methionine (AdoMet) (specific activity 18 Ci/mmol, PerkinElmer) in 50 mM Tris-HCl, pH 8.0, 0.05% β-mercaptoethanol, 5% glycerol and 200 μg/mL BSA. The DNA methylation assays were carried out in triplicate at 37°C for 40 min before being quenched by addition of 5 μL of 10 mM AdoMet. For detection, 12.5 μL of reaction mixture was spot on DEAE Filtermat paper (PerkinElmer) and dried out. The DEAE paper was then washed sequentially with 3 x 5 mL of cold 0.2 M ammonium bicarbonate, 5 mL of Milli Q water, and 5 mL of ethanol. The DEAE paper was air dried and transferred to scintillation vials filled with 5 mL of ScintiVerse (Fisher). The radioactivity of tritium was measured with a Beckman LS6500 counter.

Results

Crystal structure of the DNMT3B-DNMT3L-CpG DNA complexes

We generated the DNMT3B-DNMT3L-CpG DNA complexes using human DNMT3B C-terminal methyltransferase domain, the C-terminal domain of DNMT3L and with two Zebularine containing 25-mer DNA duplexes that harbor two ZGA or ZGT sites as mimics of CGA or CGT motifs with 14 bps spacing (Fig. 1a,b, 2a and S2a). The ZpG-containing DNA is covalently bound to DNMT3B-DNMT3L complex, representing the productive state of DNMT3B-DNMT3L. The structures of the DNMT3B-DNMT3L-CGA and DNMT3B-DNMT3L-CGT DNA complexes, bound to cofactor byproduct *S*-Adenosyl-L-homocysteine (SAH), were solved at 3.0 Å resolution (Table S1). The two structures well-aligned, with a root-mean-square deviation (RMSD) of 0.27 Å over 828 Cα atoms (Fig. S2b). As for the DNMT3A-DNMT3L-DNA complex[32], the structure of the DNMT3B-DNMT3L-DNA complexes reveal a hetero-tetrameric fold arranged in the order of DNMT3L-DNMT3B-DNMT3B-DNMT3L (Fig. 1c, S2b), where the DNA is anchored by two DNMT3B monomers with rotational symmetry. In both structures, DNMT3L does not make any contact with DNA. (Fig. 1c and S2b). We used DNMT3B-CGA complex for detailed analysis.

DNMT3B-CpG DNA interactions

Just like DNMT3A, DNMT3B MTase domain is comprised of a catalytic core with a Rossmann fold and a target recognition domain (TRD), stacking into a two-lobed architecture (Fig. 1c). DNMTB-DNA interaction involves a loop from the catalytic core (Catalytic loop: residues 648-672) that penetrates into the DNA major groove, a loop from

the TRD (TRD loop: residues 772-791) that approaches the DNA minor groove (Fig. 1c), and the homodimeric interface (RD interface) in contact with DNA backbone, creating a continuous basic surface for DNA-binding. (Fig. 1d and Fig. 2a). At each ZpG site, the base of the Zebularine (Z5/ Z20) is flipped into the active-site pocket of each DNMT3B monomer, forming a covalent bond with catalytic cysteine C651 and interacts with other residues from the catalytic pocket through hydrogen-bonding interactions (Fig. 1c). Meanwhile, the side chain of V657 located on the catalytic loop is inserted into the cavity vacated by the base flipping of Zebularine (Fig. 2b). The orphan guanine (G5′/G20) originally paired with Zebularine is further stabilized by a hydrogen bond between the backbone carbonyl group of V657 and the N2 atom of G5′/ G20′ (Fig. 2b), while the ZpG guanine (G6/G19′) is recognized by a hydrogen bond between the side chain of N779 and the O6 atom of G6/G19′, and a water-mediated hydrogen bond between the side chain of T775 and the N7 atom of G6/G19′ (Fig. 2c). Both guanines also make van der Waals contacts with catalytic loop residue P659 (Fig. 2b). Furthermore, residues on the catalytic loop (N652 and S655), the TRD loop (Q772, T773, T776, K782 and K785) and DNMT3B RD interface (R823 and G824) interact with the DNA backbone on both strands through hydrogen bonding interactions or electrostatic interactions (Fig. 2a-d and Fig. S2a,c,d). We next mapped the ICF disease associated DNMT3B mutations onto the structures of the DNMT3B-DNMT3L-DNA complex to understand how these mutations affect DNMT3B activity (Fig. S3). Notably, these mutations are summarized into three categories: the SAM- or DNA-binding sites (e.g. N658S and R823P) (Fig. S3a), “FF” or “RD” interfaces (Fig. S3b), and the regions responsible for the structural integrity of DNMT3B (Fig. S3c),

suggesting that they may impact the function of DNMT3B through different mechanisms. To test these hypotheses arising from our structural observations, we selected some key residues involved in DNA binding and introduced mutation into these sites for *in vitro* methylation analysis using $(GAC)_{12}$ DNA as substrate. Compared with wild-type DNMT3B, all mutations of the catalytic-loop residues (S655A, V657G, N658S and P659A), RD residue (R823P) and the TRD-loop residues (T775A, T776A and K782A) led to decreased enzymatic activity at variable levels (Fig. 2e). Notably, N658S and R823S are physiologically relevant ICF disease associated mutations, which both lead to substantial decrease of the DNMT3B activity (Fig. 2e). Together, these enzymatic data support the structural observations.

Comparative analysis of divergent DNA recognition between DNMT3A and DNMT3B.

The MTase domains of DNMT3A and DNTM3B share about 80% sequence identity (Fig. S1). Consistently, structural superimposition of the two DNMT3B-DNMT3L-DNA complexes with the DNMT3A-DNMT3L-DNA complex reveals high similarity, with an RMSD of 0.62 Å and 0.63 Å over 853 and 863 aligned Cα atoms, respectively (Fig. S4a). Nevertheless, we were able to observe significant discrepancies in those three DNA interacting regions mentioned above: catalytic loop (Fig.3a-c), TRD loop (Fig. 3d-f) and "RD" homodimeric interface (Fig. 3g,h). Notably, a conformational difference is observed for the catalytic loop (Fig. 4a): A side-chain hydrogen bond between DNMT3B N656 and R661 (Fig. 3c), but not between the corresponding DNMT3A residues I715 and R720 (Fig. 3b), results in altered stability and dynamic behavior of catalytic loop, which further leads

to a subtle conformational repositioning of the CpG-contacting residues in DNMT3B, such as V657 and P659, relative to their DNMT3A counterparts (Fig. 3a). To examine whether the conformational divergence between the catalytic loops of DNMT3A and DNMT3B impact their CpG specificities, we measured the enzymatic activities of DNMT3A and DNMT3B, WT and mutants, on a DNA duplex containing either $(GAC)_{12}$, $(AAC)_{12}$ or $(TAC)_{12}$ repeats. (Fig. 4a) WT DNMT3B shows higher enzymatic activity towards all DNA substrates compared to DNMT3A, with higher non-CpG activity/ CpG activity ratio (Fig. 4a), consistent with an earlier report that DNMT3B is more efficient on CpH methylation than DNMT3A [12]. Introduction of the DNMT3A I715N and DNMT3B N656I mutations, which inter-convert the sequences of the catalytic loops of DNMT3A and DNMT3B, led to a consistent shift in DNA methylation efficiencies toward each other (Fig. 4a). It was worth noting that N656-DNMT3B/I715-DNMT3A is the only inconsistent residue on catalytic loop between DNMT3A and DNMT3B (Fig. S1). We also compared TRD loop of DNMT3A and DNMT3B; there are large conformational changes from the side chains of the key residues involved in DNA interaction (Fig. 3d-f). In particular, the O6 atom of CpG guanine (G6) forms a hydrogen bond with R836 in the DNMT3A-DNMT3L-DNA complex, but in DNMT3B-DNMT3L-DNA complex with N779, corresponding to DNMT3A N838. The DNMT3B K777, corresponding to DNMT3A R836, extends its side chain along the DNA major groove without making any direct CpG guanine contact (Fig. 3f). To gain a comprehensive understand of how these residues affect enzymatic behavior of DNMT3A and DNMT3B, we performed mutagenesis and proceeded to *in vitro* methylation assays. Consistent with a previous study on DNMT3A

[32], introduction of DNMT3A R836A led to a modest increase in CpG methylation but a greater change in CpA and CpT methylation, resulting in 2.4- and 2.3-fold decrease in the relative CpG/CpA and CpG/CpT preference (Fig. 4b). Introduction of the DNMT3A N838A mutation, while reduced the overall methylation efficiency, failed to affect the relative CpG/CpA and CpG/CpT preference appreciably (Fig. 4b), consistent with the fact that this residue does not play a role in CpG recognition in DNMT3A. By contrast, introduction of the DNMT3A R836- and N838-corresponding mutations on DNMT3B both led to slight but notable change (1.5- and 0.8-fold for K777A; and 1.3- and 1.4-fold for N779A) in CpG/CpH preference, consistent with the conformational divergence of these sites between DNMT3A and DNMT3B (Fig. 4c). In addition, the conformational divergence on the TRD loop of DNMT3A and DNMT3B is further linked to their RD homodimeric interface: In DNMT3A, a side-chain hydrogen bond is formed between R882 and S837; in contrast, no such hydrogen bond is formed between the equivalent sites in DNMT3B (Fig. 3e,f), which may further contribute to a diverse DNA contacting RD interface: In DNMT3B, only R823 and G824 form hydrogen bonds with DNA (Fig. 3h) While in DNMT3A-DNMT3L-DNA complex S881, R882, L883 and R887 interact with the DNA backbone through hydrogen bonding or van der Waals contacts (Fig. 3g), resulting in larger distance between DNMT3A and DNA in the local region (Fig. S4). Then we introduced G822S, G824L and K828R mutations in DNMT3B respectively which convert into their corresponding residues in DNMT3A. *In vitro* methylation assay performed by our collaborator shows that all mutants resulted in a markedly reduced preference toward the SatII substrate (Fig. 4b), indicating that RD interface contributes to

the substrate sequence recognition preference of DNMT3B and DNMT3A. Together, these data suggest that the conformational divergence on the catalytic loop, the TRD loop and RD interface contributes to a distinct substrate recognition mechanism between DNMT3A and DNMT3B.

Structure of the DNMT3B-DNMT3L-CpA DNA complex.

To understand the structural basis of DNMT3B-mediated non-CpG methylation, we determined the crystal structure of DNMT3B-DNMT3L tetramer covalently bound to ZAG DNA which mimics the CAG motif (Fig. 5a). The structure of the DNMT3B-DNMT3L-CpA DNA complex aligns well with the DNMT3B-DNMT3L-CpG DNA complex (Fig. 5b), with an RMSD of 0.17 Å over 869 aligned Cα atoms (Fig. 5b) Replacement of CpG guanine with CpA adenosine results a reduced interaction contact with P659 (Fig. 5d,e)and loss of hydrogen bonding interaction with N779 (Fig. 5c), which may together explain why DNMT3B has lower methylation activity towards CpA that CpG DNA. Notably, the side chain of K777 interacts with N7 atom of ZAG flanking guanine (G7) through hydrogen bonding interaction, explaining why CAG is most favorable motif among all CpH of DNMT3B [37].

DNMT3B K777 recognizes the +1 flanking nucleotide of the methylation site

We have shown that the side chain of K777 on the TRD loop of DNMT3B adopts distinct conformation from DNMT3A, pointing toward the +1 flanking nucleotide, where the DNMT3A corresponding R836 flip to the opposite side, lying under CpG guanine (G6) with hydrogen bonding interaction (Fig.3 d-f). In fact, structural comparison of the three DNMT3B-DNMT3L in complexs with ZGA/ZGT/ZAG DNA further reveals subtle side

chain conformations towards different +1 flanking sequence (Fig 6a-c, S5a-c). For instance, in DNMT3B-CAG complex, K777 directly interacts with +1 guanine (G7) through hydrogen bonding interaction with N7 atom and electrostatic interaction with the backbone phosphate (Fig. 6a). In DNMT3B-CGA complex, K777 forms a water-mediated hydrogen bond with the backbone of +1 adenosine (A7) (Fig. 6b). In the DNMT3B-CGT complex, in which A7 was replaced by thymine, the side chain of K777 moves away from the DNA backbone (Fig. 6c). Together, structural comparison of those three K777-mediated interactions with DNA reveals a unique K777-meditated +1 flanking nucleotide base reading mechanism of DNMT3B (Fig. 6e).

To further determine the role of DNMT3B K777 in DNA recognition, we generated the complex of K777A-mutated DNMT3B-DNMT3L tetramer with CGT DNA and determined the structure at 2.95 Å resolution (Table 1). The structure of the DNMT3B$_{K777A}$-DNMT3L-DNA complex is well aligned with the wild type DNMT3B-DNMT3L-DNA complex (Fig. S5d), except a subtle change in base stacking along the G6-T7 step (Fig. 6f,g): the T7 base in the K777A DNMT3B-CGT complex undergoes a slide movement toward the major groove (Fig. 5d) . It is worth noting that both the K777A mutant and its corresponding R836A mutant do not lead to conformational change of the neighboring N779 (Fig. S5e,f), suggesting the distinct K777 conformation is not attributed to the sequence difference at the DNMT3B R836 and DNMT3A K777 sites. Together, these structural studies reveal a distinct role of DNMT3B K777 in the shape reading preference of DNMT3B toward the +1 flanking CpG/ non-CpG site.

To validate our structural observations, we conducted *in vitro* methylation assays to analyze how DNMT3B K777 affects the +1 flank site preference of DNMT3B. We first measured WT and K777A DNMT3B-DNMT3L methylation activity on CpG hemimethylated DNAs containing unmethylated (GTC)12 or (GAC)12 on one strand while methylated (GAmC)12 or (GTmC)12 on the complimentary strand, referred to as CGT and CGA DNAs, respectively. WT DNMT3B shows a ~2-fold activity on CGA DNA over CGT DNA (Fig. 7a), whereas the K777A mutant shows higher activity on CGT over CGA (Fig. 6h) Then we compared the methylation efficiencies of WT and K777A DNMT3B on CAA-, CAT-, CAC- and CAG-containing DNA substrates to examine K777-medited +1 flanking site preference on non-CpG methylation. Consistent with previous studies [9, 38], WT DNMT3B shows highest activity on CAG DNA with around 2-, 4, and 5-fold activity over CAA, CAT and CAC, respectively (Fig. 5i). By contrast, the K777A mutant shows an altered substrate preference with greatly reduced methylation activity on CAG DNA, but highly increased methylation efficiency for CAT and CAA DNAs (Fig. 5i), converting to 0.4-, 0.3-, 1-fold activity over CAA, CAT and CAC, respectively (Fig. 5i). Together, our structural and enzymatic characterization reveals the distinct function of DNMT3B K777 on the +1 methylation site flanking sequence shape reading.

Discussion

In mammalian cells, DNMT3B is one of the two major methyltransferases establishing *de novo* methylation across the genome. Despite their high sequence similarity (Fig. S1) and similar domain arrangement (PWWP-ADD-Mtase) suggests functional redundancy, more evidence has revealed that they are distinct from each other [1, 12-18]. Knockout of either DNMT3A and DNMT3B leads to leads to embryonic lethality in mice, indicating that both enzymes are indispensable for development [1]. In addition, DNMT3B exhibits a divergent substrate sequence preference from DNMT3A [9, 37, 38]. Due to lack of further biochemical characterization, the molecular basis of DNMT3B-mediated *de novo* DNA methylation still remained elusive. Here we report crystal structures of DNMT3B-DNMT3L in complex with CpG and non-CpG DNAs. Through structural comparison and enzymatic analysis, we demonstrate that the conformation of DNMT3B differs from DNMT3A in all three DNA-binding regions: the catalytic loop TRD loop and the RD homodimeric interface. Firstly, N656 and R661 on the two ends of the catalytic loop of DNMT3B forms hydrogen bond where not observed in DNMT3A. Presumably the H-bond stabilizes the catalytic loop, resulting an increased non-CpG/CpG methylation efficiency. Secondly, the key residues on TRD loop of DNMT3B, including K777, S778 and N779, adopt distinct side-chain conformations from DNMT3A with a diverse interaction mode. Notably, K777 of DNMT3B undergoes significant side-chain conformational changes, sensing the +1 flanking site through a base readout mechanism in the major groove [39], which determines enzymatic preference at the +1 flanking site on both CpG and CpA DNAs. It is worth mention that in DNMT3B-CAG complex, K777 directly recognizes +1

guanine in CAG motif through hydrogen bonding interaction, resulting in highest efficiency on CAG motif, and supporting previous evidence that CAGs are mostly methylated by dnmt3b in mouse ES cells [37]. Lastly, the RD interface of DNMT3B exhibits less DNA binding than DNMT3A, with DNMT3B R823 exhibiting a different conformation, losing the interaction that connects RD interface and TRD loop from its counterpart in DNMT3A, which leads to a divergent TRD loop conformation. Together, our structural and enzymatic characterization of DNMT3B has revealed a distinct substrate recognition mechanism of DNMT3B-mediated CpG and non-CpG methylation.

Figures and Table

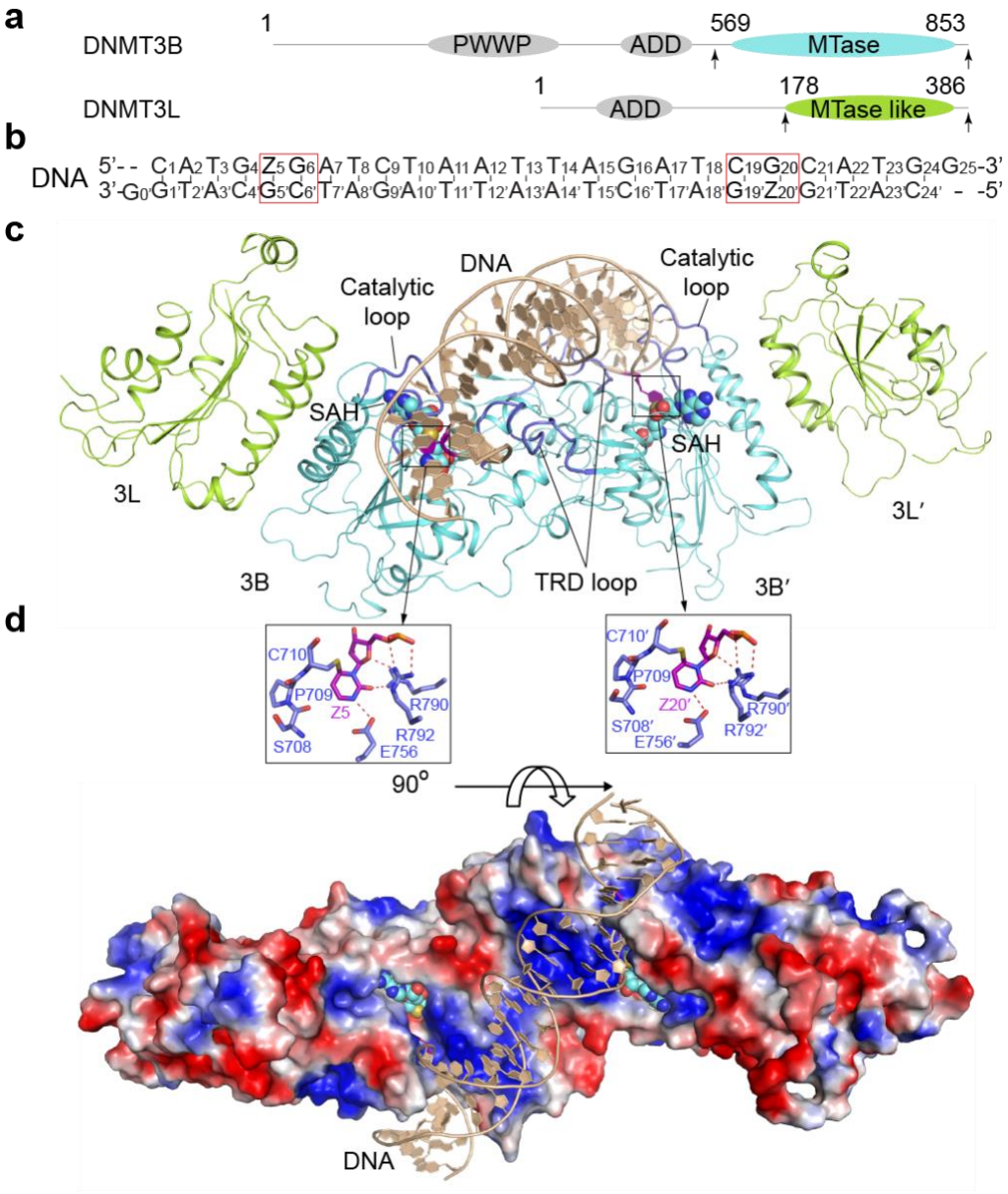


Figure 1. Structure of the DNMT3B–DNMT3L tetramer in complex with a 25-mer CGA DNA containing two CpG sites. (a) Domain architectures of DNMT3B and DNMT3L, with the C-terminal domains marked with arrowheads. **(b)** DNA sequence (CGA) used for structural study. Z, Zebularine. **(c, d)** Ribbon (c) and surface (d) representations of DNMT3B–DNMT3L bound to DNA and SAH. The DNA-contacting TRD loop and catalytic loop are colored in slate. The Zebularines anchored at the two active sites are shown in expanded views, with hydrogen-bonding interactions depicted as dashed lines.

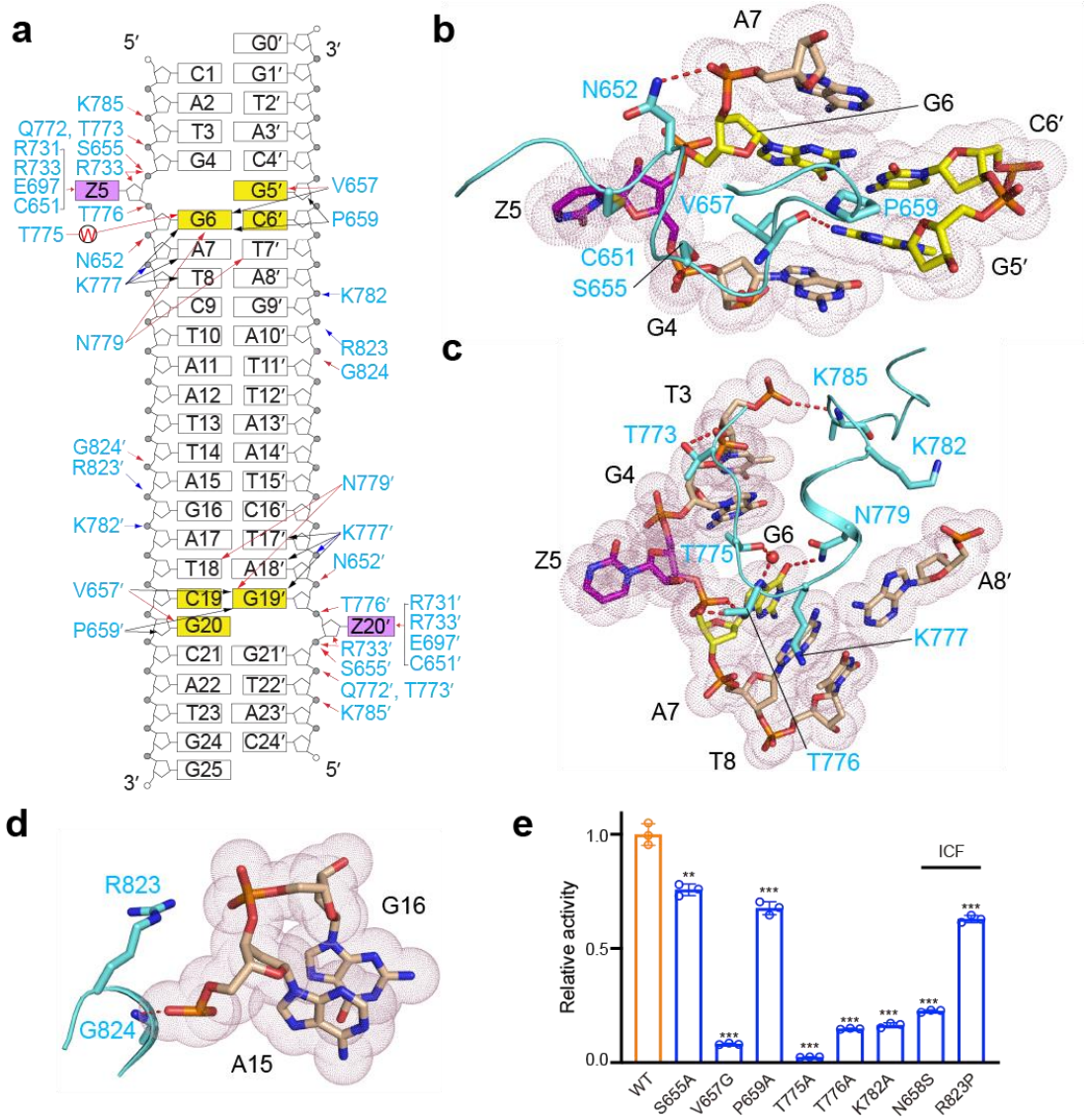


Figure 2. Intermolecular interactions between the DNMT3B–DNMT3L tetramer and DNA. (a) Schematic view of the intermolecular interactions between DNMT3B and CGA DNA. The hydrogen-bonding, electrostatic and van der Waals contacts are represented by red, blue and black arrows, respectively. Water-mediated hydrogen bonds are labeled with letter 'W'. **(b-d)** Close-up views of the DNA interactions of the catalytic loop (b), TRD loop (c) and homodimeric interface (d) of DNMT3B. The DNA is shown in dots-covered stick representation and colored in wheat, except for the ZpG site, in which the Zebularine and the other three nucleotides are colored in purple and yellow, respectively. The hydrogen bonding interactions are shown as dashed lines. **(e)** *In vitro* methylation of wild-type (WT) and mutant DNMT3B in the form of DNMT3B-DNMT3L tetramer. Data are mean ± s.d.. Statistical analysis used two-tailed Student's t-test. **, $p < 0.01$.

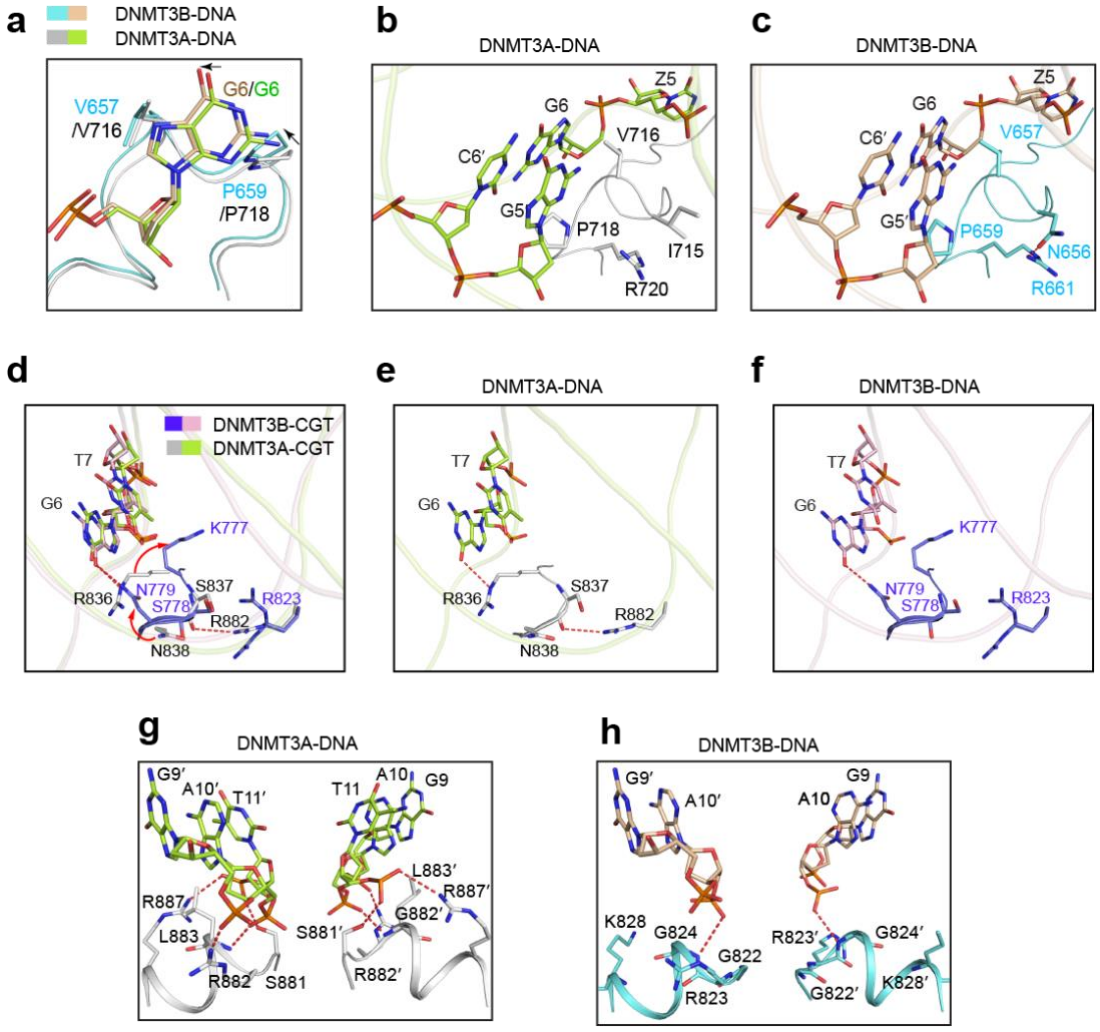


Figure 3. Divergent DNA recognition between DNMT3A and DNMT3B. Structural comparison of the previously reported DNMT3A-DNA complex (PDB 5YX2) and DNMT3B-DNA complexes, highlighting distinct intramolecular interactions between DNMT3A and DNMT3B. **(a-c)** Close-up comparison of the protein-DNA contacts at the catalytic loops of hDNMT3A and hDNMT3B in their respective DNA complexes. **(e-f)** Close-up comparison of the protein-DNA contacts at the TRD loops of hDNMT3A and hDNMT3B in their respective DNA complexes. **(g, h)** Close-up view of the protein-DNA contacts at the RD interfaces of hDNMT3A and hDNMT3B in their respective DNA complexes.

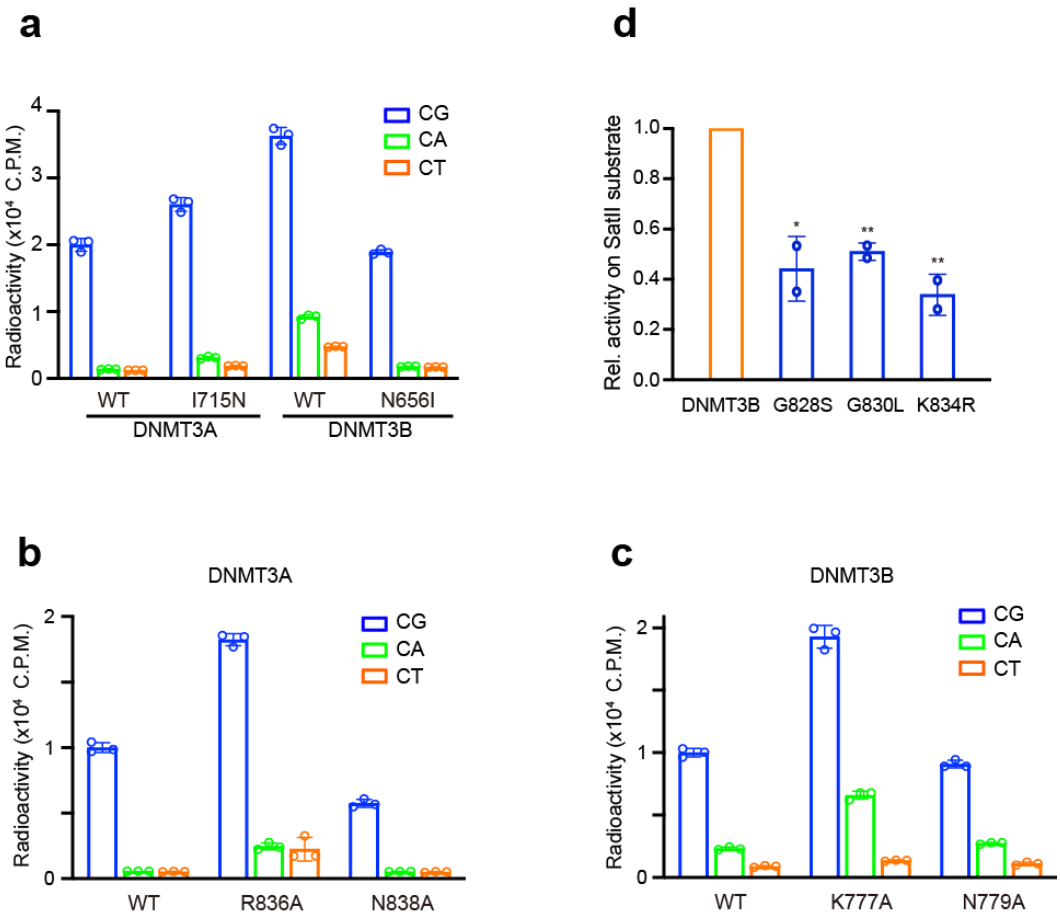


Figure 4. Methylation assay analysis of DNMT3A and DNMT3B. (a) *In vitro* methylation of WT and catalytic loop-mutated hDNMT3A and hDNMT3B on CG-, CA- and CT-containing DNA. The data and error estimate were derived from three independent measurements. **(b, c)** *In vitro* methylation of WT and TRD loop-mutated hDNMT3A (c) and hDNMT3B (d) on CG-, CA- and CT-containing DNA. The data and error estimate were derived from three independent measurements. **(d)** Methylation assay of WT mDNMT3B and RD mutants for the SatII substrate. mDNMT3B residues G828, G830 and K834 correspond to hDNMT3B G822, G824 and K828. Statistical analysis used two-tailed Student's t-test for the difference from WT: ***, $p < 0.001$. Data was generated by collaborator.

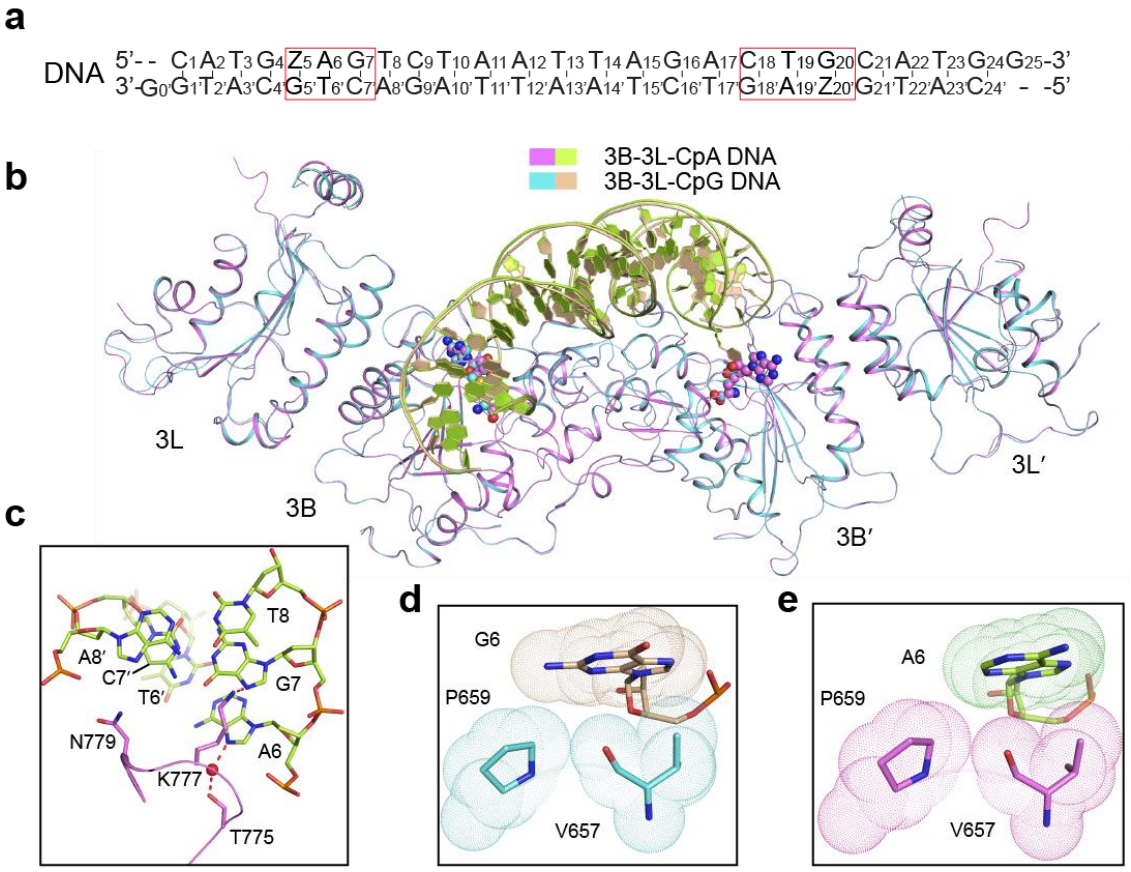


Figure 5. Structure of the DNMT3B–DNMT3L tetramer in complex with a 25-mer DNA containing two CpA sites. (a) DNA sequence used for the structural study. **(b)** Structural overlay of the DNMT3B-CAG DNA and DNMT3B-CGA DNA complexes. **(c)** Close-up view of the interactions between the TRD loop of DNMT3B and the CAG DNA. **(d,e)** Close-up view of the van der Waals contacts between the catalytic loop of DNMT3B and G6 in the ZpG site of the CGA complex (d) or A6 in the ZpA site of the CAG complex (e).

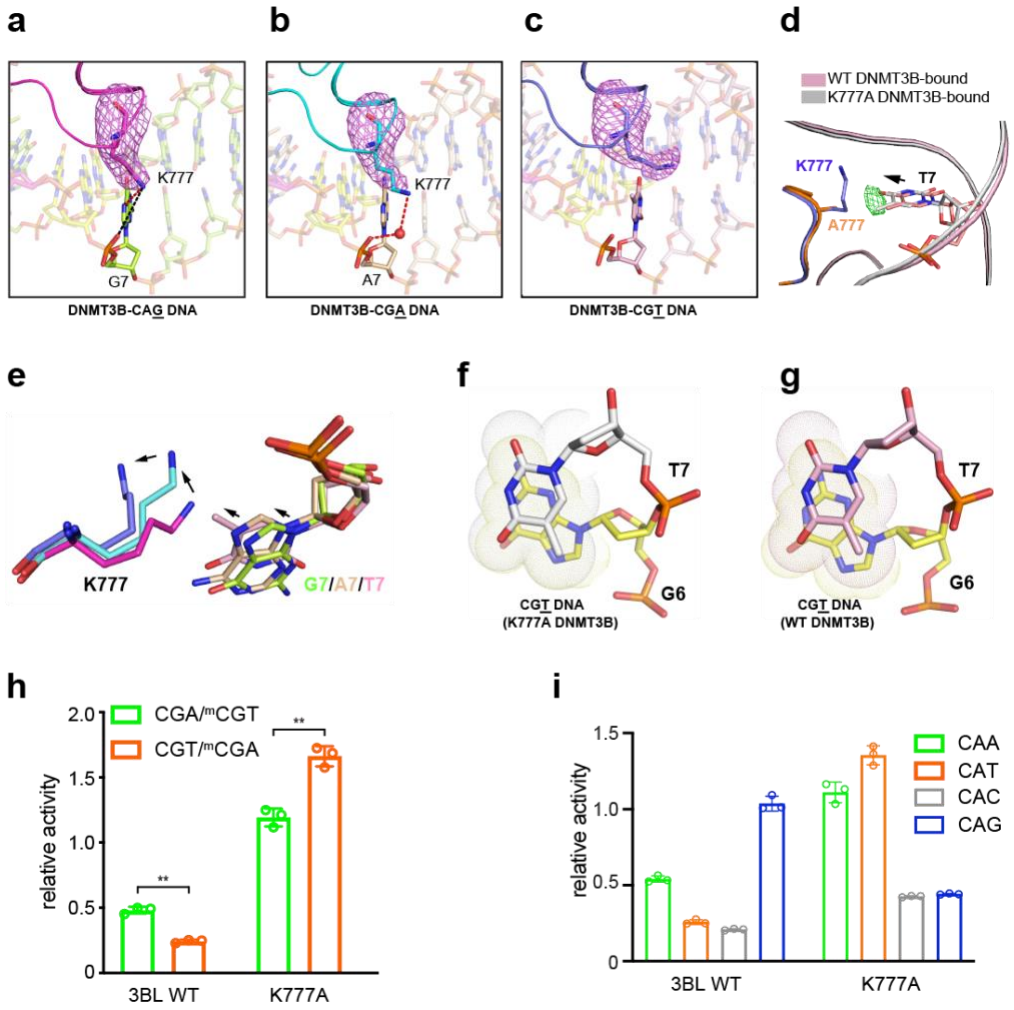


Figure 6. Recognition of the DNA shape of the +1 flanking site by DNMT3B K777. (a-c) Stick representations of the CAG (a), CGA (b) and CGT (c) DNAs bound to the DNMT3B-DNMT3L tetramer. Fo-Fc omit map of the K777 (light magenta) is contoured at 2.0 σ level. The hydrogen bonding and electrostatic interactions are shown as red and black dashed lines, respectively. The water molecule is shown as red sphere. (**d**) Structural comparison of the +1 nucleotide (T7) at the CpG site of the DNA bound to WT or K777A-mutated DNMT3B. The T7 nucleotides associated with WT and K777A-mutated DNMT3B-CGT complexes are colored in light pink and silver, respectively. The conformational shift of the T7 nucleotide between the two complexes is indicated by arrow as well as the Fo-Fo difference map of T7 (green; contoured at 2.0 σ level) derived from the electron density for the WT and K777A DNMT3B complexes. (**e**) Structural superposition of DNMT3B K777 and interacting +1 nucleotides in CAG, CGA and CGT DNA complexes. The color schemes are the same as in (a), (b) and (c)The relative positioning of the +1 bases and DNMT3B K777 are indicated by arrows. **(f, g)** Close-up view of the stacking interaction between CpG guanine (G6) and +1 base (T7) of CGT DNA bound to K777A-mutated (**f**) or WT (**g**) DNMT3B-DNMT3L. **(h)** *In vitro* methylation of WT and K777A hDNMT3B-DNMT3L on hemimethylated CGT and CGA DNA. The data and error estimate were derived from three independent measurements. Statistical analysis used two-tailed Student's t-test. *, $p < 0.05$, **, $p < 0.01$. **(i)** *In vitro* methylation of WT and K777A hDNMT3B-DNMT3L on DNA containing CAA, CAT, CAC and CAG, respectively.

Table 1. X-ray data collection and refinement statistics.

	DNMT3B-DNMT3L-CGA DNA (PDB: 6U8P)	DNMT3B-DNMT3L-CGT DNA (PDB: 6U8V)	DNMT3B-DNMT3L-CAG DNA (PDB: 6U8X)	DNMT3B$_{K777A}$-DNMT3L-CGT DNA (PDB: 6U8W)
Data collection				
Space group	P 31	P 31	P 31	P 31
Cell dimensions				
a, *b*, *c* (Å)	193.3, 193.3, 49.7	194.2, 194.2, 49.7	193.8, 193.8, 49.9	193.5 193.5 49.9
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00 120.00
Resolution (Å)	48.33-3.04 (3.15-3.04)	48.55-3.00(3.10-3.00)	42.88-2.95 (3.06-2.95)	48.38-2.95 (3.05-2.95)
R_{merge}	0.130 (0.833)	0.107 (0.682)	0.110 (0.994)	0.097 (0.733)
$I/\sigma(I)$	8.5 (1.3)	11.8 (2.3)	8.8 (1.2)	7.91 (1.34)
$CC_{1/2}$	0.996 (0.843)	0.997 (0.924)	0.998 (0.763)	0.997 (0.782)
Completeness (%)	99.8 (99.4)	99.3 (94.2)	99.64 (99.20)	99.7 (99.3)
Redundancy	4.3 (4.0)	5.0 (4.9)	5.2 (4.7)	3.4 (3.3)
Refinement				
No. reflections	39736	41814	43905	43844
R_{work} / R_{free}	0.204/0.237	0.204/0.240	0.227/0.252	0.216/0.236
No. atoms				
Protein and DNA	8398	8527	8521	8380
Ligand	52	52	52	52
Water	25	11	19	11
B factors (Å$_2$)				
DNMT3B	69.8	68.5	66.2	70.8
DNMT3L	120.7	121.7	123.6	127.2
DNA	142.7	124.3	124.9	155.2
Ligand	54.4	47.2	50.5	78.4
Water	61.6	52.1	60.6	65.6
r.m.s deviations				
Bond lengths (Å)	0.006	0.007	0.003	0.004
Bond angles (°)	1.11	1.16	0.62	0.68

[a]Values in parentheses are for highest-resolution shell. Each structure was determined using the dataset collected from a single crystal.

Supplementary data

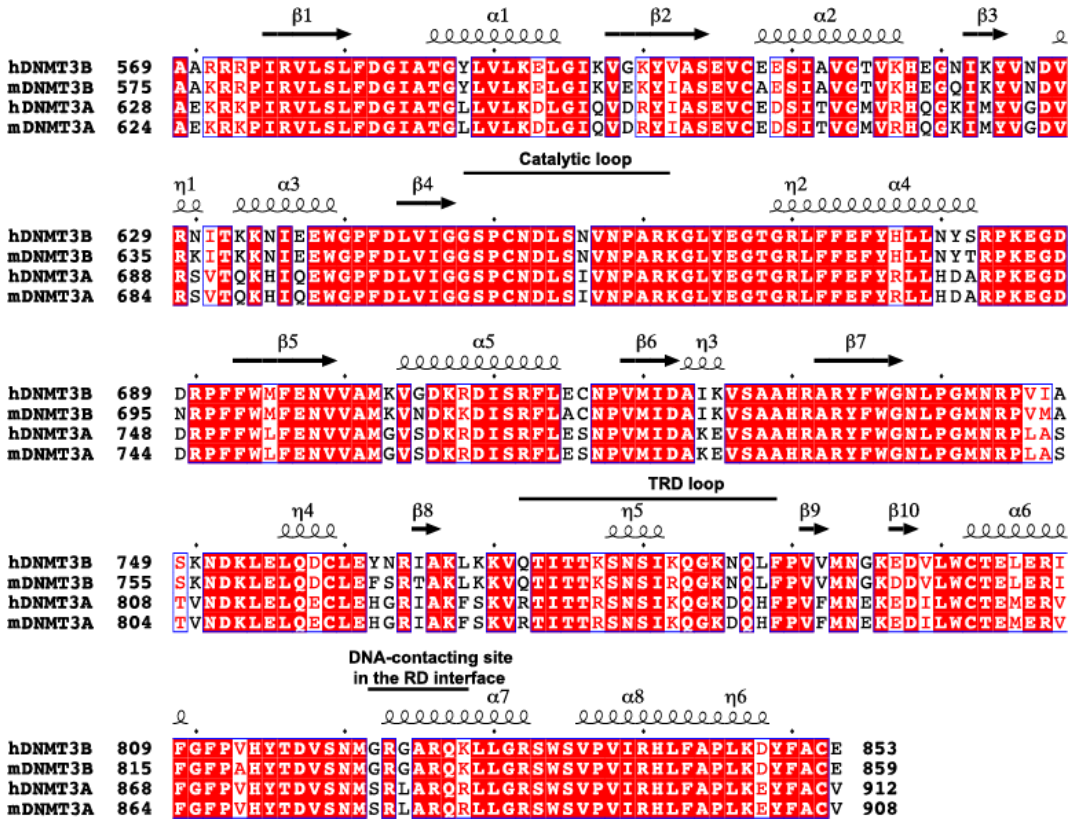


Figure S1. Sequence alignment of DNMT3 proteins from human (hDNMT3B, hDNMT3A) and mouse (mDNMT3B and mDNMT3A). Secondary structures are shown above the aligned sequences. Fully conserved residues are colored in white and highlighted in red shade. Partially conserved residues are colored in red. The residues involved in DNA binding are marked on top of the aligned sequences.

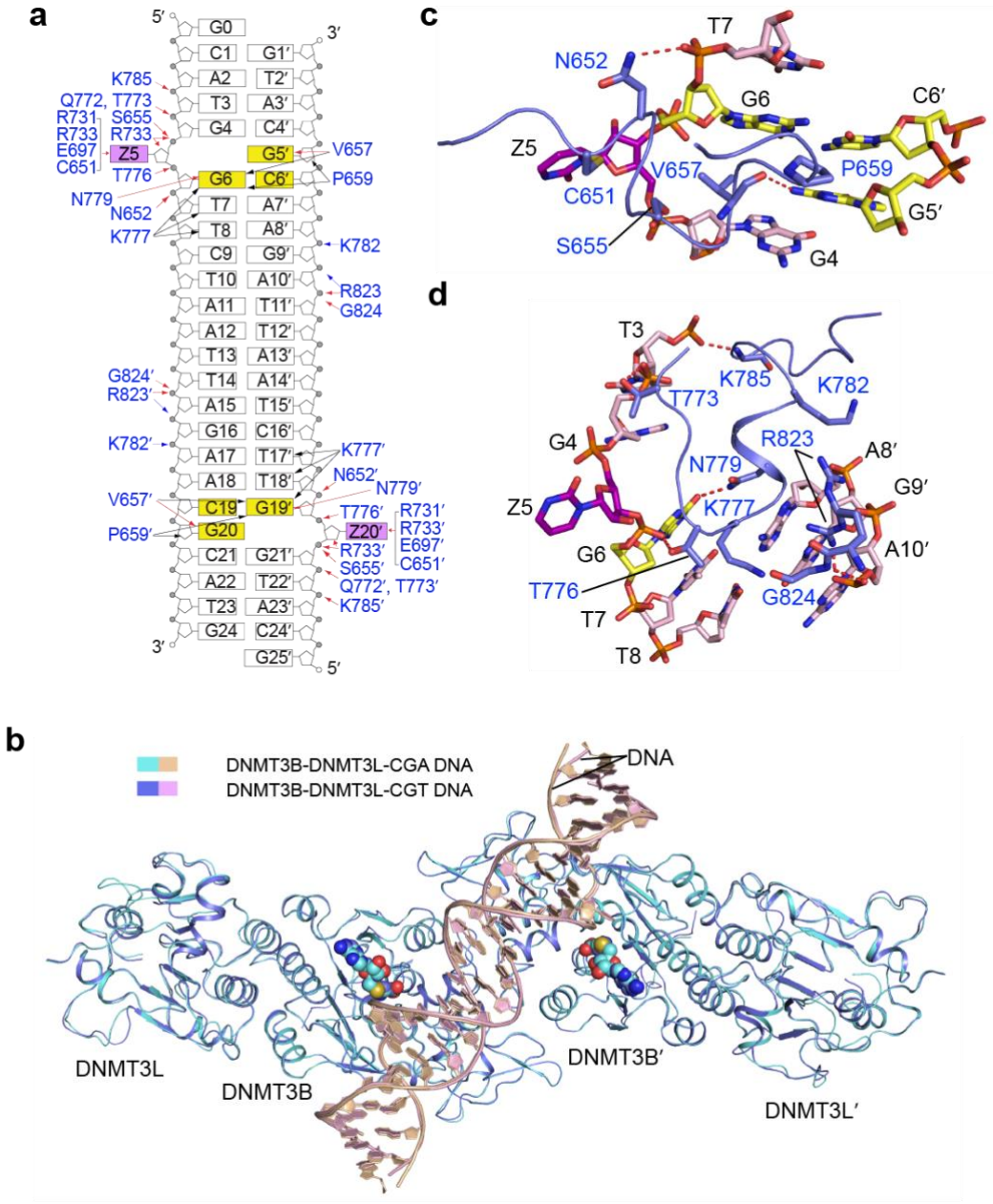


Figure S2. Structural analysis of the DNMT3B-DNMT3L-CGT DNA complex. a, Structural overlay of the DNMT3B-DNMT3L-CGA and DNMT3B-DNMT3L-CGT complexes. **b,** Schematic view of the intermolecular interactions between DNMT3A and CGT DNA. The hydrogen-bonding, electrostatic and van der Waals contacts are represented by red, blue and black arrows, respectively. Water-mediated hydrogen bonds are labeled with letter 'W'. **c, d,** Close-up views of the DNA interactions of the catalytic loop (**c**), and TRD loop and homodimeric interface (**d**) of DNMT3

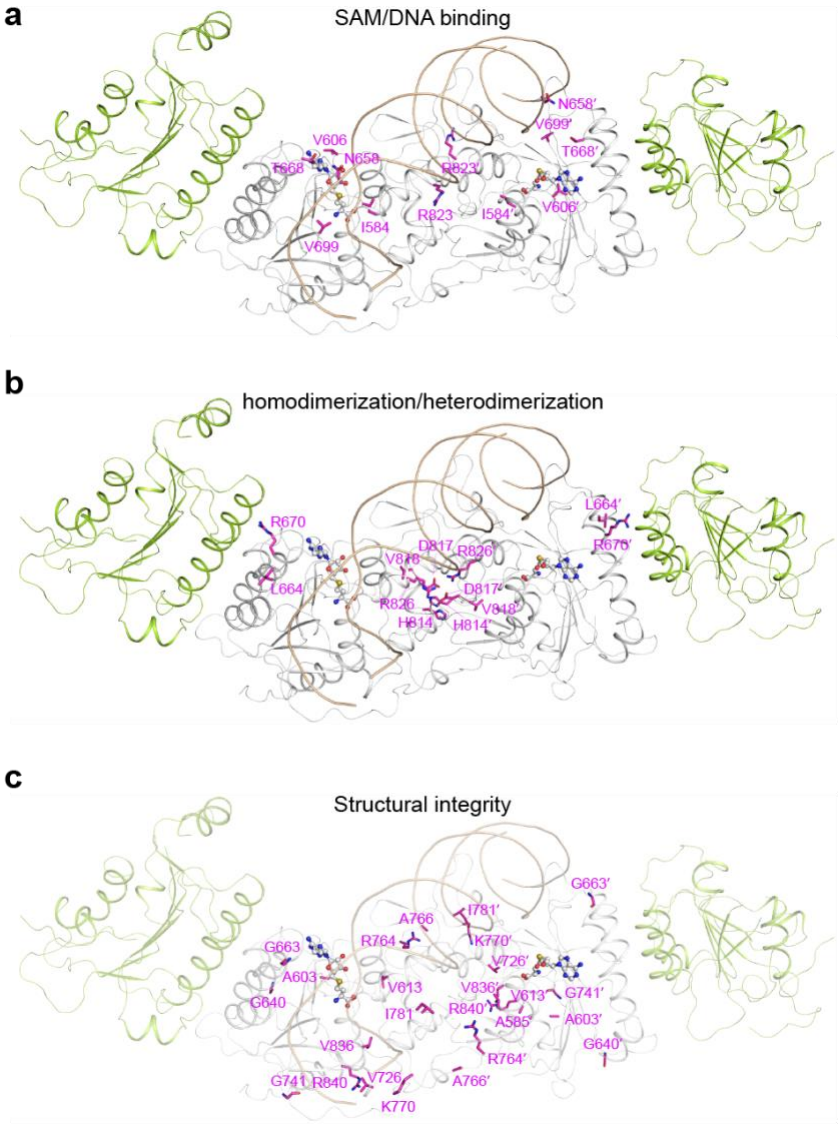
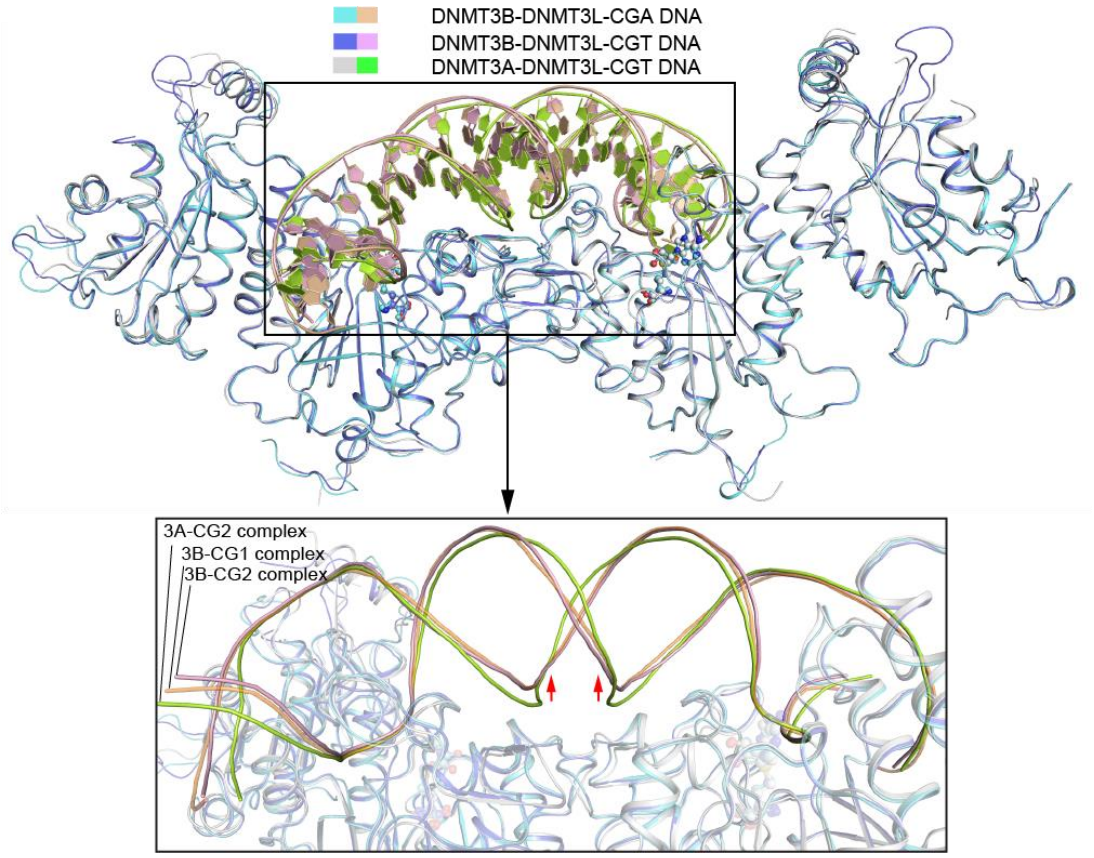


Figure S3. Mapping the ICF-associated mutations on the structure of DNMT3B. (a-c) Ribbon diagram of the structure of the DNMT3B-DNMT3L-CGA DNA, with ICF mutation states show as sticks and colored in light magenta.(**a)** ICF mutations located on the cofactor SAM or DNA binding sites.**(b)** ICF mutations located on the "RD" or "FF" interfaces of DNMT3B. **(c)** ICF mutations located on the regions of DNMT3B responsible for structural integrity.



S4. Structural comparison of the DNMT3B-DNMT3L-DNA and DNMT3A-DNMT3L-DNA complexes. Structural superposition of the DNMT3B-DNMT3L-CGA DNA, DNMT3B-DNMT3L-CGT DNA and DNMT3A-DNMT3L-CGT DNA. The lower part shows a close-up view of the DNAs bound to DNMT3B or DNMT3A.

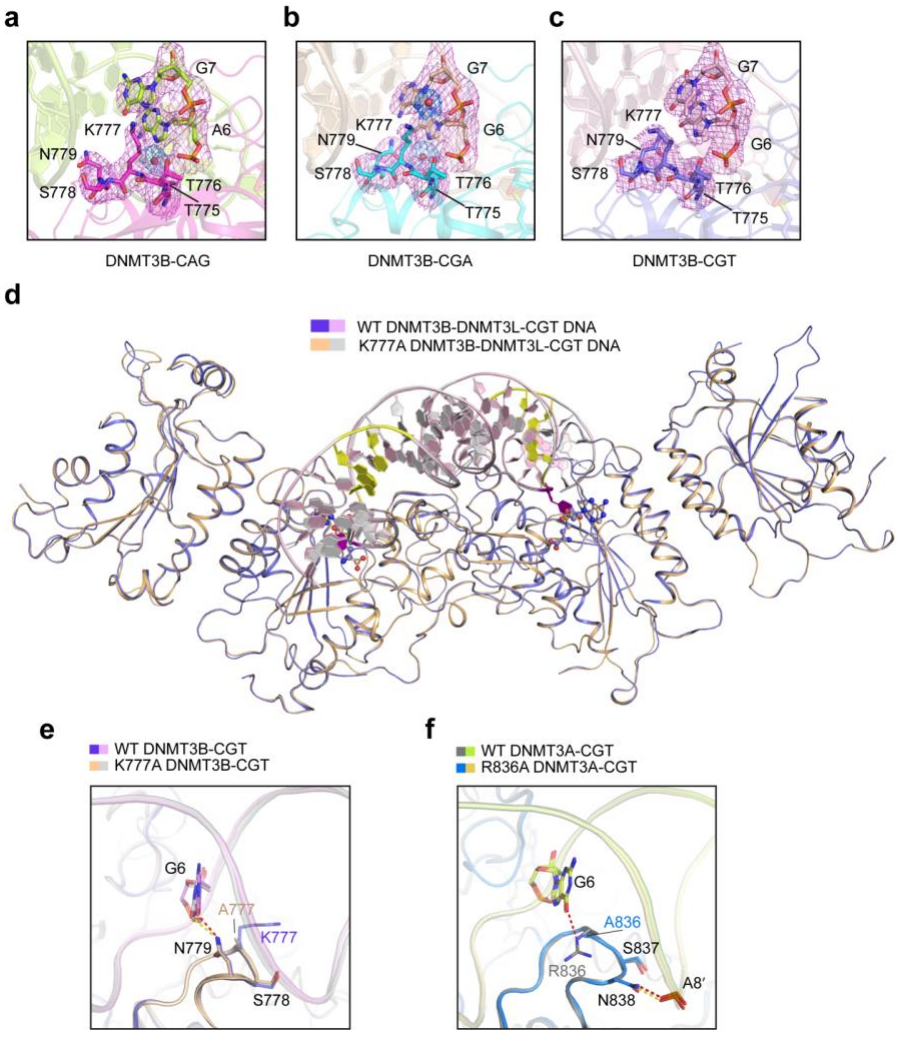


Figure S5. Structural analysis of the TRD loop of DNMT3B in DNA-bound form. (a-c) Fo-Fc omit map (violet) for TRD residues T775-N779 of DNMT3B and interacting DNA nucleotides in the DNMT3B-CAG (a), DNMT3B-CGA (b) and DNMT3B-CGT (c) complexes. The Fo-Fc omit map for water molecules (red sphere) are colored blue in (a) and (b). All the omit maps are contoured at 2.0 σ level. **(d)** Structural superposition of the DNMT3B-DNMT3L-CGT DNA and K777A-mutated DNMT3B-DNMT3L-CGT DNA complexes. The CpG/ZpG sites are colored in purple (Zebularine) or yellow. The SAH molecules are shown in sphere representation. **(e)** Close-up view of the aligned WT DNMT3B-CGT and K777A-mutated DNMT3B-CGT complexes, with the hydrogen bonds in the WT and K777A complexes shown as red and wheat dashed lines, respectively. **(f)** Close-up view of the aligned WT DNMT3A-CGT and R836A-mutated DNMT3A-CGT complexes, with the hydrogen bonds in the WT and R836A complexes shown as red and wheat dashed lines, respectively.

Reference

1. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.

2. Bourc'his, D., et al., *Dnmt3L and the establishment of maternal genomic imprints.* Science, 2001. **294**(5551): p. 2536-9.

3. Chedin, F., M.R. Lieber, and C.L. Hsieh, *The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16916-21.

4. Hata, K., et al., *Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice.* Development, 2002. **129**(8): p. 1983-93.

5. He, Y. and J.R. Ecker, *Non-CG Methylation in the Human Genome.* Annu Rev Genomics Hum Genet, 2015. **16**: p. 55-77.

6. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-22.

7. Ramsahoye, B.H., et al., *Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a.* Proc Natl Acad Sci U S A, 2000. **97**(10): p. 5237-42.

8. Barres, R., et al., *Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density.* Cell Metab, 2009. **10**(3): p. 189-98.

9. Guo, J.U., et al., *Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain.* Nat Neurosci, 2014. **17**(2): p. 215-22.

10. Ma, H., et al., *Abnormalities in human pluripotent cells due to reprogramming mechanisms.* Nature, 2014. **511**(7508): p. 177-83.

11. Chen, T., et al., *Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b.* Mol Cell Biol, 2003. **23**(16): p. 5594-605.

12. Aoki, A., et al., *Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases.* Nucleic Acids Res, 2001. **29**(17): p. 3506-12.

13. Baubec, T., et al., *Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation.* Nature, 2015. **520**(7546): p. 243-7.

14. Challen, G.A., et al., *Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells.* Cell Stem Cell, 2014. **15**(3): p. 350-364.

15. Gowher, H. and A. Jeltsch, *Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases.* J Biol Chem, 2002. **277**(23): p. 20409-14.

16. Hsieh, C.L., *In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b.* Mol Cell Biol, 1999. **19**(12): p. 8211-8.

17. Norvil, A.B., et al., *Dnmt3b Methylates DNA by a Noncooperative Mechanism, and Its Activity Is Unaffected by Manipulations at the Predicted Dimer Interface.* Biochemistry, 2018. **57**(29): p. 4312-4324.

18. Suetake, I., et al., *Distinct enzymatic properties of recombinant mouse DNA methyltransferases Dnmt3a and Dnmt3b.* J Biochem, 2003. **133**(6): p. 737-44.

19. Liao, J., et al., *Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells.* Nat Genet, 2015. **47**(5): p. 469-78.

20. Laurent, L., et al., *Dynamic changes in the human methylome during differentiation.* Genome Res, 2010. **20**(3): p. 320-31.

21. Lister, R., et al., *Global epigenomic reconfiguration during mammalian brain development.* Science, 2013. **341**(6146): p. 1237905.

22. Xie, W., et al., *Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome.* Cell, 2012. **148**(4): p. 816-31.

23. Hamidi, T., A.K. Singh, and T. Chen, *Genetic alterations of DNA methylation machinery in human diseases.* Epigenomics, 2015. **7**(2): p. 247-65.

24. Smeets, D.F., et al., *ICF syndrome: a new case and review of the literature.* Hum Genet, 1994. **94**(3): p. 240-6.

25. Ley, T.J., et al., *DNMT3A mutations in acute myeloid leukemia.* N Engl J Med, 2010. **363**(25): p. 2424-33.

26. Yang, L., R. Rau, and M.A. Goodell, *DNMT3A in haematological malignancies.* Nat Rev Cancer, 2015. **15**(3): p. 152-65.

27. Gowher, H. and A. Jeltsch, *Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites.* J Mol Biol, 2001. **309**(5): p. 1201-8.

28. Handa, V. and A. Jeltsch, *Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome.* J Mol Biol, 2005. **348**(5): p. 1103-12.

29. Jurkowska, R.Z., et al., *Approaches to enzyme and substrate design of the murine Dnmt3a DNA methyltransferase.* Chembiochem, 2011. **12**(10): p. 1589-94.

30. Lin, I.G., et al., *Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro.* Mol Cell Biol, 2002. **22**(3): p. 704-23.

31. Wienholz, B.L., et al., *DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo.* PLoS Genet, 2010. **6**(9): p. e1001106.

32. Zhang, Z.M., et al., *Structural basis for DNMT3A-mediated de novo DNA methylation.* Nature, 2018. **554**(7692): p. 387-391.

33. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Macromolecular Crystallography, Pt A, 1997. **276**: p. 307-326.

34. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.

35. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

36. Adams, P.D., et al., *PHENIX: building new software for automated crystallographic structure determination.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 11): p. 1948-54.

37. Lee, J.H., S.J. Park, and K. Nakai, *Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s.* Sci Rep, 2017. **7**(1): p. 11295.

38. Lister, R., et al., *Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells.* Nature, 2011. **471**(7336): p. 68-73.

39. Rohs, R., et al., *Origins of specificity in protein-DNA recognition.* Annu Rev Biochem, 2010. **79**: p. 233-69.

Chapter 3

Dissect the DNMT3A- and DNMT3B-mediated DNA co-methylation through a covalent complex approach

Abstract

DNA methylation plays a critical role in regulating gene expression, genomic stability and cell fate commitment. Mammalian DNA methylation, which mostly occurs in the context of CpG dinucleotide, is installed by two *do novo* DNA methyltransferases, DNMT3A and DNMT3B. Oligomerization of DNMT3A and DNMT3B permits both enzymes to co-methylate two CpG sites located on the same DNA substrates. However, how DNMT3A- and DNMT3B-mediated co-methylation contributes to the DNA methylation patterns remain unclear. Here we generated covalent enzyme-substrate complexes of DNMT3A and DNMT3B, and performed bisulfite sequencing-based single turnover methylation analysis on both complexes. Our results showed that both DNMT3A- and DNMT3B-mediated co-methylation preferentially gives rises to a methylation spacing of 14 base pairs, consistent with the previous structural observation for DNMT3A in complex with regulatory protein DNMT3L and CpG DNA. This study provides a novel method for mechanistic investigation of DNMT3A- and DNMT3B-mediated DNA co-methylation.

Introduction

In mammals, the DNA methylation pattern is mainly established by *de novo* DNA methyltransferases DNMT3A and DNMT3B during gametogenesis and postimplantation development [1]. Both of them are regulated by DNMT3-like protein (DNMT3L) which is a catalytically inactive paralogue of DNMT3 enzymes [2-4]. In the crystal structure of DNMT3A methyltransferase (MTase) domain in complex with the C-terminal domain of DNMT3L, DNMT3A directly interacts with DNMT3L to forms a 'butterfly shape' heterotetramer in the order of DNMT3L-DNMT3A-DNMT3A-DNMT3L, where two DNMT3A molecules in the middle and one DNMT3L molecule at each edge [5]. Two active sites from each DNMT3A unit are separated by a distance of ~40 Å, has lent support to a DNA co-methylation model, in which DNMT3A can co-methylate two CpG sites across the DNA double strands separated by 8-10 base pairs (bps)[5]. Followed by the structural modelling, a subsequent study has developed a hairpin bisulfite analysis-based approach to evaluate the DNMT3A-mediated CpG co-methylation, where the most correlated methylation events occur at the CpG sites on the two opposite strands, with spacing of 8-10 bp[6]. These observations therefore led to the proposition that the DNMT3A-mediated CpG co-methylation may contribute to the periodicity of CpG methylation in mammalian genomes[5, 6]. Interestingly, genome-wide methylation analysis has identified a modest enrichment of 8-10 base pair (bp) periodicity among the methylation sites of both mammalian and plant genomes[5, 7-9]. However, this approach is challenged by the fact that it could not distinguish whether the two correlated methylation events arise from co-methylation or sequential methylation events (Fig. 1A).

In fact, it has been argued that the 10-bp periodic distribution of the WW (W= A, T) and SS (SS= G, C) dinucleotides in nucleosomal DNA, which serves to accommodate the DNA wrapping around histone octamers [10], may dictate the periodicity of DNA methylation[7]. A more recent crystal structure of DNMT3A-DNMT3L in complex with DNA has revealed that the distance between two active sites of DNMT3A-DNMT3L heterotetramer is 14 base pairs, where the protein complex is well aligned with DNA-binding free state with minimal conformational change, which contradicts with the 10-bp model [11]. (Fig. 1B) To gain further understanding of the DNMT3A- and DNMT3B-mediated *de novo* methylation, we developed a "single-turnover" methylation assay, in which DNMT3A or DNMT3B is covalently linked to one target site of the substrates, thereby restricting the dissociation of DNMT3A or DNMT3B from the substrate before the next enzymatic attack. Our result unequivocally clarified 14 bp as the co-methylation spacing of DNMT3A and DNMT3B, providing a basis for mechanistic understanding of DNMT3A- and DNMT3B-mediated CpG co-methylation.

Materials and Methods

Protein expression and purification.

The MTase domains of human DNMT3A or DNMT3B (DNMT3A MTase: residues 562–853; DNMT3B MTase: residues 628–912) or was co-expressed with residues 178–386 of human DNMT3L on a modified pRSFDuet-1 vector (Novagen), which contains a hexahistidine (His_6) and SUMO tag preceding the DNMT3B or DNMT3A sequence. The expression and purification of the DNMT3A-DNMT3L or DNMT3B-DNMT3L complex was as previously described[11]. In essence, the protein complexes were expressed using the *Escherichia coli* BL21 DE3 (RIL) cell strains and purified using a Ni^{2+}-NTA column, followed by removal of the His_6-SUMO tag through Ubiquitin-like-specific protease 1 (ULP1) cleavage, ion exchange chromatography on a Heparin HP column (GE Healthcare) and size-exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare). The final protein samples were concentrated in a buffer containing 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.1% β-mercaptoethanol and 5% glycerol, and stored in -80°C before use.

Preparation of covalent complexes.

To prepare the covalent complex of the DNMT3A-DNMT3L or DNMT3B-DNMT3L tetramer with ZpG-containing DNA for methylation assay, the DNMT3A-DNMT3L or DNMT3B-DNMT3L tetramer was mixed with a 37-mer/68-mer hybrid DNA duplex ($DNA_{spacing1}$) (Upper strand: 5′-GCATGZGTTCTAACGCGCGCGCGTGAAGGAAGG AAGG-3′; Lower strand: 5′- GGTGGTGGTGGTGATCCTTCCTTCCTTCACGCG CGC GCGTTAGAACGCATGCGTGGAGATGGAGGAGG-3′; Z: Zebularine) or a 38-mer/69-

mer hybrid DNA duplex ($DNA_{spacing2}$) (Upper strand: 5′- GCATGXGTTCTAA GCGCGCGCGCGTGAAGGAAGGAAGG-3′; Lower strand: 5′- GGTGGTGGTG GTGATCCTTCCTTCCTTCACGCGCGCGCGCTTAGAACGCATGCGTGGAG ATG GAGGAGG-3′; Z: Zebularine) in a 2:1 molar ration and incubated under the condition of 20 mM Tris-HCl (pH 8), 20% glycerol, and 40 mM DTT at room temperature. The 5′-overhang sequences in both DNA substrates serve as templates for strand-specific PCR amplification as well as internal control for bisulfite conversion. Subsequently, the DNMT3A/3B-DNMT3L-$DNA_{spacing1}$ or DNMT3A/3B-DNMT3L-$DNA_{spacing2}$ covalent complexes were purified sequentially through ion exchange chromatography on a HiTrap Q XL column (GE Healthcare) and size-exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column.

Single-turnover methylation assay.

For the enzymatic assays, the DNMT3A/3B-DNMT3L-$DNA_{spacing1}$ or DNMT3A/3B-DNMT3L-$DNA_{spacing2}$ covalent complexes were adjusted to 1 μM and reacted in a buffer containing 32 μM AdoMet, 50 mM Tris-HCl, pH 8.0, 0.05% β-mercaptoethanol, 5% glycerol and 200 μg/mL BSA. The reaction was performed at 37°C for 1h. The reaction products were digested with protease K at 42°C for 2h, purified by agarose gel electrophoresis using GeneJET gel extraction kit (Thermo Scientific). 100 ng of purified DNA was subjected to bisulfite conversion using an EZ DNA Methylation Gold Kit (Zymo Research), followed by PCR amplification using primers: 5'-CCACCACCACCACTA-3' and 5'-GGAGGAGGTAGAGGTG-3'. The PCR products were cloned into pCR4-TOPO vector (Invitrogen) for sequencing analysis.

Results

In our previous study we've reported the crystal structure of DNMT3A-DNMT3L in complex with DNA in which Zebularine was used to replace the cytosine in the CpG site, forming a covalent bond with the active site of DNMT3A [11]. Then we superimposed two crystal structures of the DNMT3A-DNMT3L-CpG DNA complexes: one involves the DNMT3A-DNMT3L tetramer bound to two 10 bp-long DNA duplexes each containing one single central CpG/ZpG site (PDB: 6F57), while the other involves the DNMT3A-DNMT3L tetramer bound to a 25-mer self-complimentary DNA substrate containing two CpG/ZpG sites, with the target Zebularines separated by 14 bps (PDB: 5YX2). Though the space groups of those two structures are different, the protein-DNA interaction interface are well aligned, indicating that those the interaction surface is not affected by crystal packing effect. (Fig. 1B) It is noting that the shorter DNA is tilted by a small angle, mimicking the bending conformation of the longer DNA, which suggests that the 14-bp spacing between the two target sites in the longer DNA may be natural co-methylation spacing of DNMT3A under the physiological conditions. (Fig. 1B) To explore how DNMT3B and DNMT3A may mediate DNA co-methylation in solution, we designed a single turnover methylation assay (Fig. 2A,B), based on the stable DNMT3A-DNMT3L-DNA or DNMT3B-DNMT3L-DNA complex formed prior to the methylation reaction (Fig. 2A). In essence, we first generated a covalent complex between the DNMT3A-DNMT3L or DNMT3B-DNMT3L tetramer and an unmethylated DNA substrate containing one single ZpG site as well as multiple CpG sites, with the Zebularine and cytosines separated by 8-, 10-, 12-, 14- and 16-bp, respectively. Next, the DNMT3A-DNMT3L-DNA or

DNMT3B-DNMT3L-DNA complex was incubated with cofactor S-adenosyl-L-methionine (SAM), which triggers the methylation reaction with DNMT3A or DNMT3B staying bound to the DNA. Subsequently, DNMT3A or DNMT3B was removed from the DNA substrate through protease digestion, the target strand of the DNA product was amplified by PCR reaction, and the distances between the methylcytosine and the ZpG site was measured by bisulfite sequencing analysis (Fig. 2B and Tables S1-S4). In this study, we particularly focus on the DNA strand that is complementary to the ZpG-containing strand, given the fact that structural analysis of the DNMT3A-DNMT3L-DNA complex indicates an opposite directionality between the co-methylated CpG sites (Fig. 1B).

To ensure the methylated cytosine we count is not from the failure of bisulfite conversion reaction, we first analyzed the bisulfite conversion rate for the unmethylated cytosine in both external and internal control. The sequencing date of the external control, which is the unmethyltated CpG DNA that was not subjected to methylation reaction, shows a ~94% conversion rate (Fig. S1). The internal control, which is collection of all the non-CpG sites from the DNA that was under methylation reaction, reveals a >96% conversion rate (Fig. S1), which further validates the efficiency of bisulfite conversion. We are able to obtain 157 sequences from DNMT3A-DNMT3L methylation product and 150 sequences of DNMT3B-DNMT3L methylation product for even number substrate. Among them, 50 and 59 single-methylated sequences are identified respectively for DNMT3A and DNMT3B with both over 90% methylation preference on a spacing of 14bps (Fig. 2C,D and Figure S2 and Tables S2-S3). In addition, we performed the single-turnover methylation assay on the DNA substrate in which the Zebularine within the sole ZpG site

and the cytosines within the CpG sites are separated by 9-, 11-, 13-, 15- and 17-bp, respectively. Under the same reaction condition, we've obtained 12 out of 36 analyzed sequences with a single methylation event which give no significant spacing preference (Fig. 2E, Fig. S2, Table S1 and S4). Together, our data suggest that two target sites with 14 bp spacing most likely represent the most favorable co-methylation substrate for both DNMT3A and DNMT3B.

Discussion

The previously proposed DNA co-methylation spacing model of the DNMT3A-DNMT3L tetramer highlights the importance of DNMT3A oligomerization in defining DNA methylation patterns [5, 6]. This model fits with genome-wide methylation analysis, which has indicated an 8-10 bp of methylation periodicity of mammalian genomes [7, 12, 13]. However, due to lacking direct structural evidence, how such oligomerization of DNMT3A and DNMT3B contribute to the correlation co-methylation of two neighboring sites remains controversial [10]. This study, through single-turnover methylation assay, combined with our previous structural observation of DNMT3A-DNMT3L-DNA complexes, demonstrates 14 bp as the most favorable co-methylation spacing for DNMT3A and DNMT3B.

The apparent discrepancy between the 14-bp co-methylation spacing determined here and the previous observation of 8-10 bp spacing likely arises from different enzymatic catalysis scenarios. The 8-10 bp methylation spacing may arise from an energetically distinct scenario (Fig. 3B): it is likely that the DNMT3A or DNMT3B dimer presents one of the monomers to methylate one CpG site first, with the other monomer binding to the DNA in a non-productive manner; upon the completion of the first methylation event, the DNMT3A or DNMT3B dimer presents the second monomer to attack another CpG site that is in close proximity. During this transition, DNMT3A/3B and DNA substrate may undergo a conformational adjustment without being fully dissociated. The 14-bp co-methylation spacing arises from a scenario in which the two active sites of the DNMT3A or DNMT3B dimer undergo the productive catalysis simultaneously (Fig. 3A). Structurally,

the co-methylation event observed in this study is accompanied by significant bending of DNA substrates, which might not be energetically favorable in solution when compared with the sequential methylation events along the substrate with the standard B-form conformation. A detailed characterization of these enzymatic behaviors awaits further investigation. And how the 14-bp spacing mechanism described here may contribute to the DNA methylation patterns in cells, in particular in the context of the chromatin environment, remains to be elucidated.

Figures

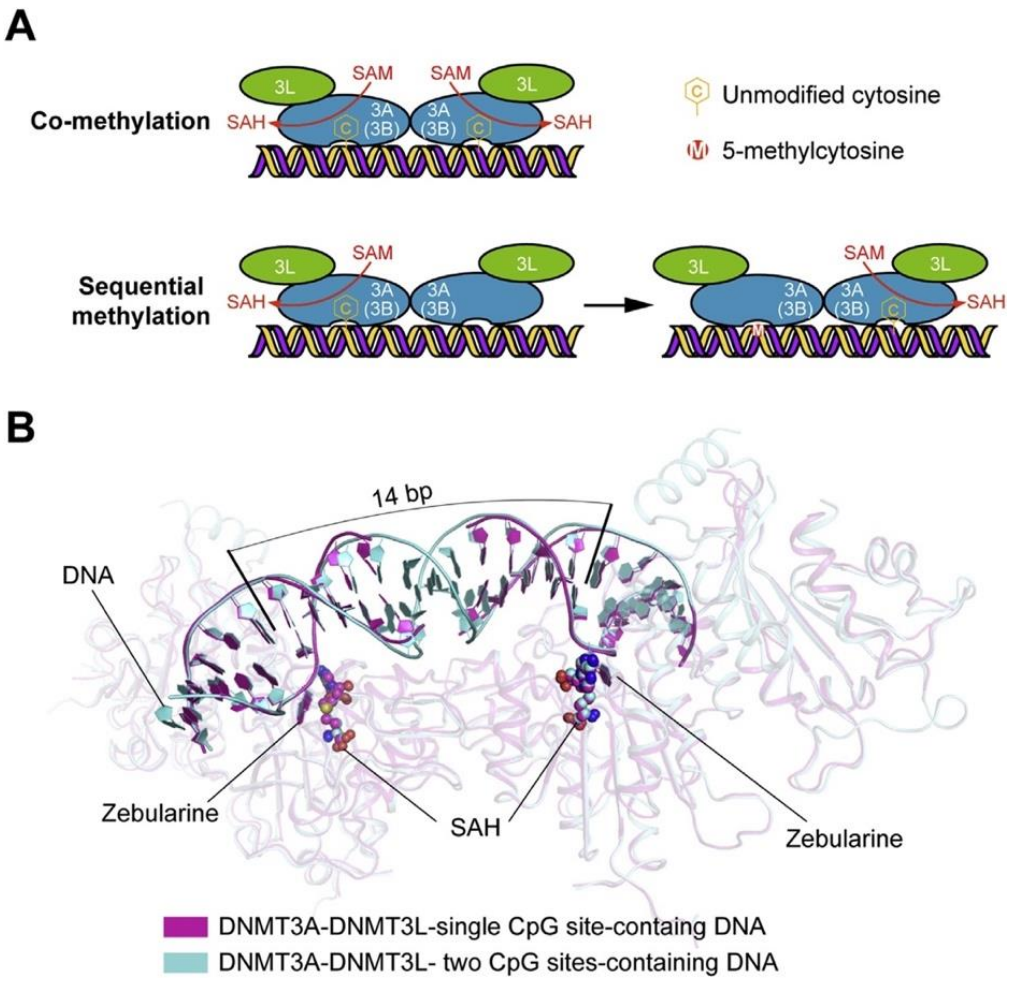


Figure 1. Co-methylation model of DNMT3A- and DNMT3B. (A) Schematic model of co-methylation vs sequential methylation. (B) Structural overlay of the DNMT3A-DNMT3L tetramer bound to two short DNA duplexes containing single CpG site (PDB 6F57) and the DNMT3A-DNMT3L tetramer bound to one long DNA substrate containing two CpG sites (PDB 5YX2). Both structures suggest co-methylation spacing of 14-bp.

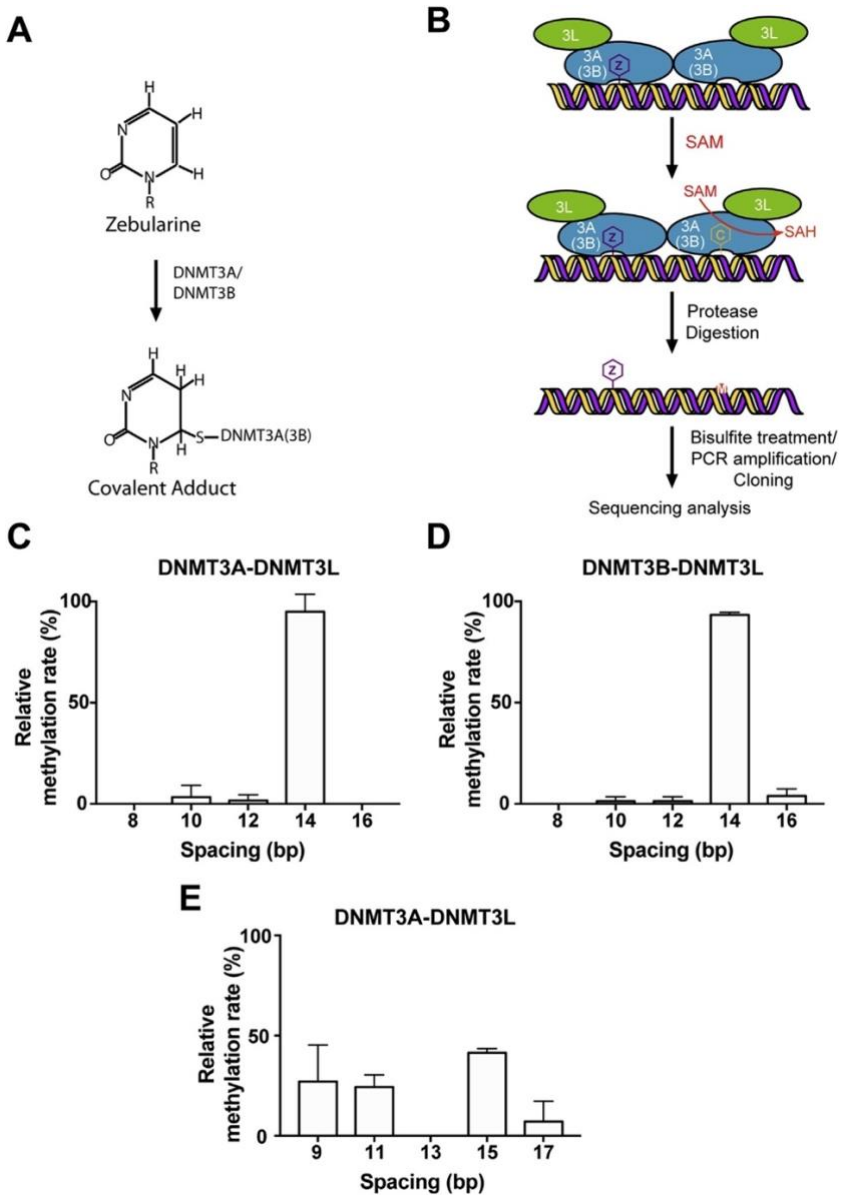


Figure 2. Schematic view of single-turnover methylation assay for measurement of co-methylation spacing. (A) Schematics for generation of the covalent complex of DNMT3A and DNMT3B with Zebularine-containing DNA. (B) Workflow for the single-turnover methylation assay. The Zebularine and unmodified cytosine are denoted as "Z" and "C", respectively. The target strand contains a 5′ overhang for the purpose of strand-specific PCR amplification. (C) The relative abundance of various CpG co-methylation spacing mediated by the DNMT3A-DNMT3L complex. The average and standard deviation were derived from analysis of three independent batches of sequencing data (See Table S2). (D) The relative abundance of various CpG co-methylation spacing mediated by the DNMT3B-DNMT3L complex. The average and standard deviation were derived from analysis of three independent batches of sequencing data (See Table S3). (E) The relative abundance of various CpG co-methylation spacing mediated by the DNMT3A-DNMT3L complex. The average and standard deviation were derived from analysis of three independent batches of sequencing data (See Table S4).

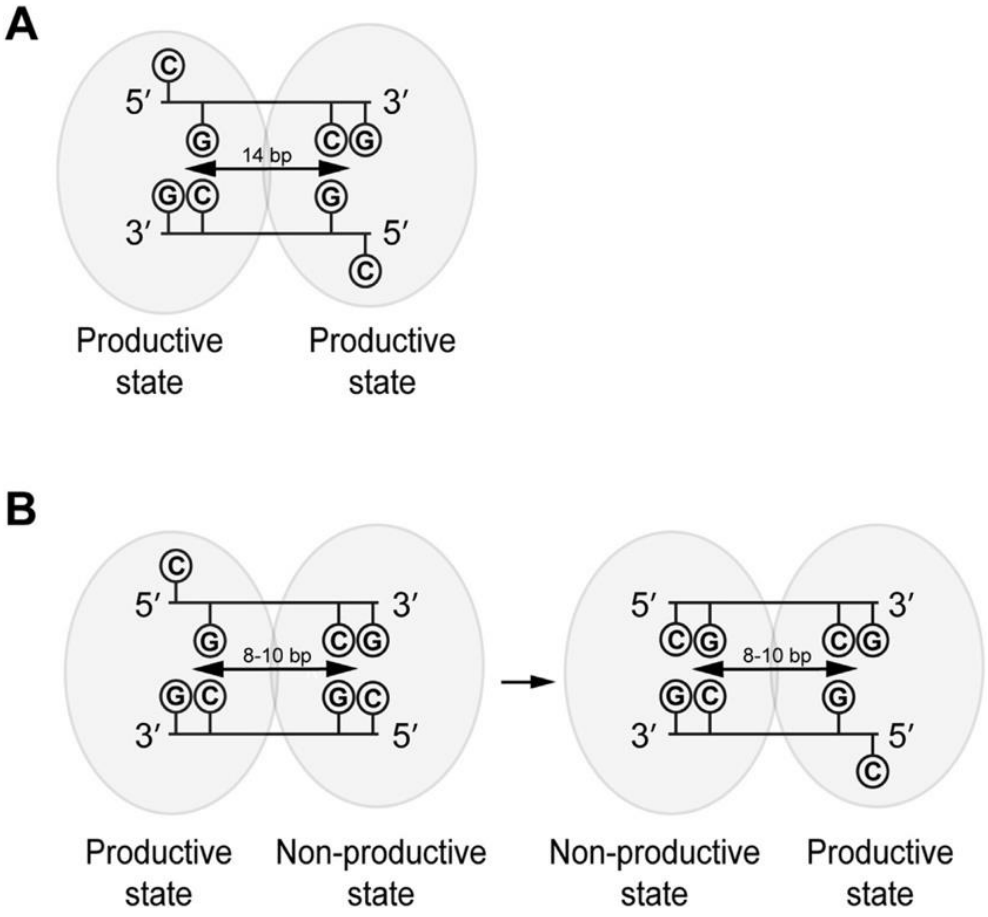


Figure 3. Model for DNMT3A- and DNMT3B-mediated DNA co-methylation. (A) The 14-bp co-methylation spacing arises from the two concurrent methylation events of DNMT3A and DNMT3B. The CpG sites are denoted by letters 'C' and 'G'. The DNMT3A or DNMT3B monomers are represented by oval spheres. The flipped 'C' represents the insertion of target cytosine into the active site of each monomer, corresponding to a productive methylation state. (B) DNMT3A/3B may catalyze the methylation of one CpG site first, and then transit into the catalytic state for methylating the next CpG site. At each catalytic stage, only one monomer undergoes productive methylation, while the other monomer adopts a non-productive state, which does not involve base flipping of target cytosine. During the transition of the methylation events, the DNMT3A/3B dimer may remain associated with the DNA substrate.

Supplementary Data

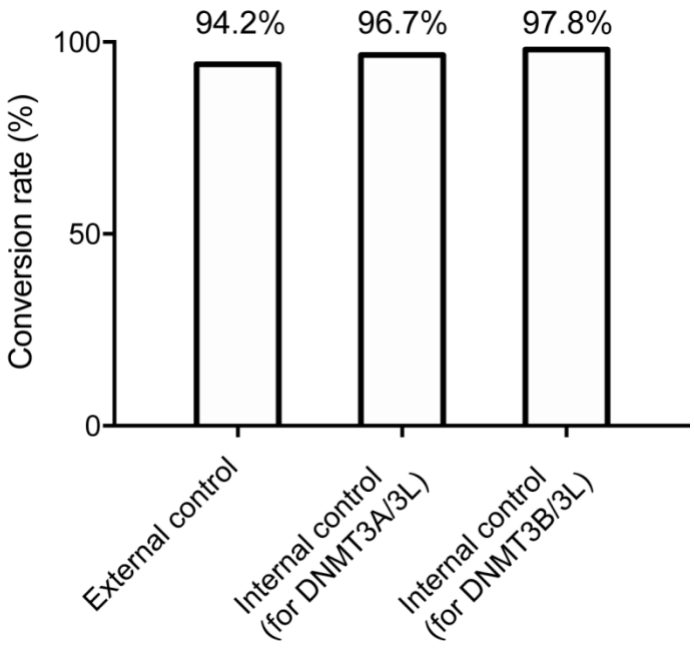


Figure S1. Bisulfite conversion rates of the CpG site complementary to the ZpG site (internal control) and the unmethylated CpG DNA that is not subjected to methylation reaction (external control).

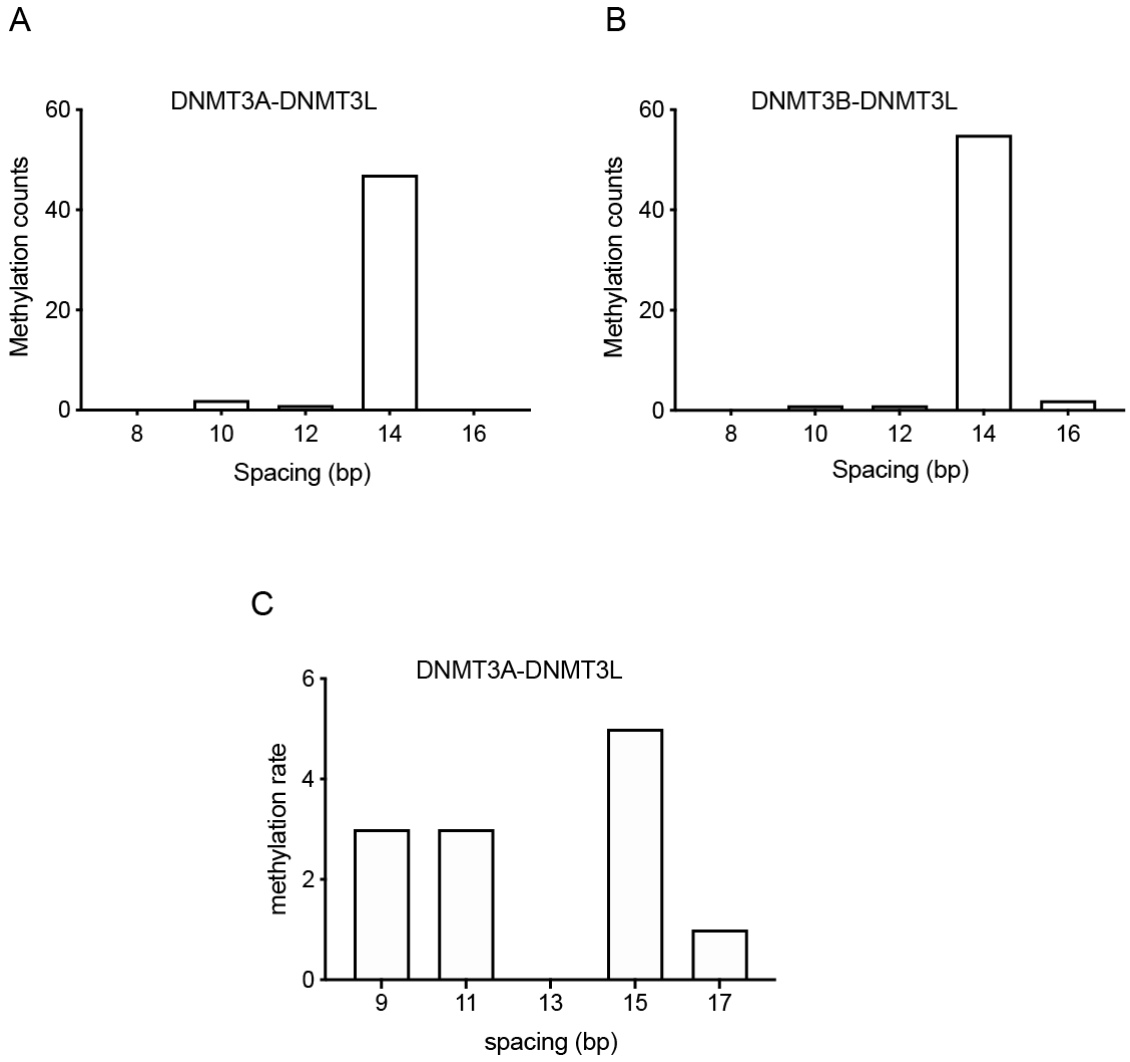


Figure S2. (A) Counts of the DNMT3A-DNMT3L tetramer-mediated methylation events on the CpG sites with indicated distances to the Zebularine site. (B) Counts of the DNMT3A-DNMT3L tetramer-mediated methylation events on the CpG sites with indicated distances to the Zebularine site. (C) Counts of the DNMT3A-DNMT3L tetramer-mediated methylation events on the CpG sites with indicated distances to the Zebularine site.

Table S1. Categorical analysis of the DNA sequencing results.

Substrate	Enzyme	Total sequences	Single-methylation	Un-methylated	Multiple-methylation
Even number CpG spacing	DNMT3A-DNMT3L	157	50	1	106
Even number CpG spacing	DNMT3B-DNMT3L	150	59	3	88
Odd number CpG spacing	DNMT3A-DNMT3L	36	12	5	19

Table S2. Sequencing data for DNA substrates with 8-, 10-, 12-, 14- and 16-bp CpG spacing, methylated by DNMT3A-DNMT3L. These data provide complementary sequence information for the target strand. In comparison with the reference sequence, an A-to-G conversion indicates the presence of methylcytosine, colored in red in sequence. The cytosine located on the Zebularine site is labeled in bold.

Reference sequence	ACATA**C**ATTCTAACACACACATAAAAAAAAAAAAAA (Complementary to bisulfite-converted, unmethylated DNA substrate)	
Group1	ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAGA ACATA**C**ATTCTAACACACACGCATAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAGAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAGAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA	12 sequences
Group 2	ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAGAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAGA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAGAAAAAAAAAA	18 sequences
Group 3	ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAA ACATA**C**ATTCTAACACGCACACATAAAAAAAGAAAGAA ACATA**C**ATTCTAACACACACGCATAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAGA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA CATA**C**ATTCTAACACGCACACATAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAGA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAGAA	20 sequences

	ACATA**C**ATTCTAACACACACGCATGAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATGAAAAAAAAAAGA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACGCACATAAAAAAAAAAAAG ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAGA	

Table S3. Sequencing data for DNA substrates with 8-, 10-, 12-, 14- and 16-bp CpG spacing, methylated by DNMT3B-DNMT3L. These data provide complementary sequence information for the target strand. In comparison with the reference sequence, an A-to-G conversion indicates the presence of methylcytosine, colored in red in sequence. The cytosine located on the Zebularine site is labeled in bold.

Reference sequence	ACATA**C**ATTCTAACACACACACATAAAAAAAAAAAAAA (Complementary to bisulfite-converted, unmethylated DNA substrate)	
Group1	ACATA**C**ATTCTAACACACACGTAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAGA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAGAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA CATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAGA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAAA	16 sequences
Group 2	ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAGAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAGA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA CATA**C**ATTCTAACACACACACGTAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA	18 sequences
Group 3	ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAGAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAGAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA	25 sequences

	ACATA**C**ATTCTAACACACGCACATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAGA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA GCATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAGA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAGA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACGCACACATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATGAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAAA ACATA**C**ATTCTAACACACACGCATAAAAAAAAAAAAAA	

Table S4. Sequencing data for DNA substrates with 9-, 11-, 13-, 15- and 17-bp CpG spacing, methylated by DNMT3A-DNMT3L or DNMT3B-DNMT3L. These data provide complementary sequence information for the target strand. In comparison with the reference sequence, an A-to-G conversion indicates the presence of methylcytosine, colored in red in sequence. The cytosine located on the Zebularine site is labeled in bold.

Reference sequence	ACATA**C**ATTCTAAACACACACACATAAAAAAAAAAAAAA (Complementary to bisulfite-converted, unmethylated DNA substrate)	
Group 1	ACATA**C**ATTCTAAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAAACACGCACACATAAAAAAACAAAAG -CACA**C**ATTCTAAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAAACGCACACACA-AAAA ACATA**C**ATTCTAAACACACACACGTAAAAAAAAAAAAA ACATA**C**ATTCTAAACACACACGCATAAAAAAAAAAAAA ACATA**C**ATTCTAAACACGCACACATAAAAAAAAAAAAAA	7 sequences
Group 2	ACATA**C**ATTCTAAACGCACACACATAAAAAAAAAAAAAA ACATA**C**ATTC-AAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAAACACACACGCATAAAAAAAAAAAAAA ACATA**C**ATTCTAAACGCACACACATAAAAGAAAAAAAAA ACATA**C**ATTCTAAACACGCACACATAAAAAAAAAAAAA	5 sequences

Reference

1. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.

2. Bourc'his, D., et al., *Dnmt3L and the establishment of maternal genomic imprints.* Science, 2001. **294**(5551): p. 2536-9.

3. Chedin, F., M.R. Lieber, and C.L. Hsieh, *The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16916-21.

4. Hata, K., et al., *Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice.* Development, 2002. **129**(8): p. 1983-93.

5. Jia, D., et al., *Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation.* Nature, 2007. **449**(7159): p. 248-51.

6. Jurkowska, R.Z., et al., *Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L.* Nucleic Acids Res, 2008. **36**(21): p. 6656-63.

7. Chodavarapu, R.K., et al., *Relationship between nucleosome positioning and DNA methylation.* Nature, 2010. **466**(7304): p. 388-92.

8. Cokus, S.J., et al., *Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning.* Nature, 2008. **452**(7184): p. 215-9.

9. Glass, J.L., et al., *CG dinucleotide periodicities recognized by the Dnmt3a-Dnmt3L complex are distinctive at retroelements and imprinted domains.* Mamm Genome, 2009. **20**(9-10): p. 633-43.

10. Widom, J., *Role of DNA sequence in nucleosome stability and dynamics.* Q Rev Biophys, 2001. **34**(3): p. 269-324.

11. Zhang, Z.M., et al., *Structural basis for DNMT3A-mediated de novo DNA methylation.* Nature, 2018. **554**(7692): p. 387-391.

12. Zhang, X., et al., *Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis.* Cell, 2006. **126**(6): p. 1189-201.

13. Zhang, Y., et al., *DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution.* PLoS Genet, 2009. **5**(3): p. e1000438.

General conclusion

Since our research interest is focusing on mammalian DNA methylation in epigenetic regulation, which is further divided into two main categories: maintenance DNA methylation and *de novo* DNA methylation. We first investigated the key regulator in maintenance DNA methylation: UHRF1. Identified as one of the essential components of the maintenance DNA methylation machinery, UHRF1 regulates the recruitment of DNMT1 to methylation target sites through its specific association with replicating heterochromatin enriched with H3K9me3 and hemimethylated DNA [1-4]. The synergistic bindings of UHRF1 to H3K9me3 and hemimethylated DNA presumably provide a redundant mechanism for spatial-temporal control of UHRF1 activity within the nuclei, which in turn contributes to the high-fidelity maintenance of DNA methylation. In this regard, recent evidence has indicated that disruption of the UHRF1 – H3K9me3 interaction leads to a modest reduction of genomic DNA methylation, highlighting a role of the UHRF1 – H3K9me3 recognition in promoting DNA methylation maintenance [5]. Previous studies from others and us have revealed that the intramolecular interaction between TTD and PBR of UHRF1 results in a "closed" conformation with reduced chromatin binding activity [6-9], which can be "opened up" through bindings of UHRF1 to H3K9me3 and hemimethylated DNA [6, 8], chromatin modifiers such as USP7 [9] or DNMT1 [6], or nuclear ligand PI5P [7]. Recent studies further demonstrated that conformational opening of UHRF1 elevates both of its chromatin binding and E3 ubiquitin ligase activities [6, 8], therefore establishing a link between the conformational transition

of UHRF1 and its cell cycle-dependent chromatin localization [10] and ubiquitylation activity [8, 11, 12]. Through high-resolution structure determination, mutagenesis and biochemical characterizations, this study confirmed the previous NMR study [6] that the intramolecular TTD – PBR interaction of UHRF1 leads to repositioning of the TTD-PHD linker, which consequently interferes with the UHRF1 – H3K9me3 interaction.

Extending from the NMR study, the crystal structure presented herein further demonstrated that the interaction of PBR with TTD also leads to the blockage of the H3K9m3-binding cage, thereby providing a more comprehensive view on the conformational regulation of UHRF1 (Fig.1). Disruption of the UHRF1 TTD – PBR interaction would presumably lead to simultaneous exposure of the H3K9m3-binding cage and re-alignment of the TTD-PHD linker, resulting in enhanced UHRF1 – H3K9me3 association. This H3K9me3 occlusion-regulated conformational transition may provide a mechanism for the H3K9me3-specific localization of UHRF1, which is important for DNMT1-mediated maintenance DNA methylation. The PBR segment of UHRF1 appears to serve as a central platform for functional regulation of UHRF1. Our previous study indicated that introduction of K644E/K646E mutations in the PBR sequence led to reduced chromatin binding of UHRF1, suggesting an important role of this region in nuclear localization of UHRF1 [9]. In fact, the intramolecular interaction between PBR and the TTD domain has been shown to be modulated by its intermolecular interaction with DNA [6, 8], USP7 [9], DNMT1 [6] and PI5P [7], resulting in a conformational transition of UHRF1 for differential chromatin accessibility. (Fig. 1) It is conceivable that interactions of the PBR sequence with different nuclear factors are subject to a dynamic regulation

throughout the cell cycle, which might be important for its function in maintenance DNA methylation and other cellular activities. For instance, the PBR-mediated interaction between UHRF1 and DNMT1 might facilitate the recruitment of DNMT1 to its target loci, while opening up the conformation of UHRF1 for efficient histone H3 ubiquitylation [8]. Subsequently, association of DNMT1 with ubiquitylated H3 [11, 12] likely leads to release of PBR to engage DNA binding in adjacent regions. The mechanism underlying the dynamic transition between different PBR engagements awaits further investigation.

In the Chapter 2 and Chapter 3, we investigated DNMT3B-mediated DNA methylation, to gain more mechanistic insight into the interplay between DNMT3s and substrate in *de novo* methylation. As one of the two major de novo DNA methyltransferases, DNMT3B plays critical roles in establishing DNA methylation during the early stage of embryonic development, and shares partial functional redundancy with DNMT3A in *de novo* methylation across the genome [13, 14]. On the other hand, increasing evidence has demonstrated their non-overlapping functions [14-21]. For instance, mouse embryos with DNMT3A or DNMT3B knockout died in utero or immediately after birth [14], indicating that both enzymes are indispensable for development. It has been demonstrated that DNMT3A is critical for establishing methylation in major satellite repeats and allelic-specific imprinting during gametogenesis [22, 23], whereas DNMT3B plays a dominant role in early embryonic development [24], giving rise to methylation in centromeric minor satellite repeats [14] and actively transcribed gene bodies [16]. In addition, DNMT3A and DNMT3B show distinct enzymatic properties [20], with DNMT3B associated with a more pronounced activity toward non-CpG methylation both *in vitro* and *in vivo* [15, 25].

Intriguingly, DNMT3B-mediated CpH methylation primarily occurs in the context of CAG motif in ES cells [26, 27], whereas DNMT3A-mediated CpH methylation shows preference toward the CAC motif in neurons [28, 29]. Recent studies have suggested that DNMT3A and DNMT3B give rise to different landscapes of non-CpG methylation across tissues, with DNMT3B producing more methylation at CAG motif in ES cells and DNMT3A producing more methylation at CAC motif in neuronal cells [30-32]. Thus, the molecular basis governing DNMT3B-mediated methylation, and how the activities of DNMT3A and DNMT3B are related, remain elusive.

This study, through systematic structural and biochemical characterizations of DNMT3B in CpG and non-CpG methylation, provides a mechanistic view on DNMT3B-mediated *de novo* methylation, as well as a framework for characterizing disease-associated DNMT3B mutations. Importantly, DNMT3B and DNMT3A, with a similar oligomeric fold and DNA co-methylation behavior, employ distinct CpG and non-CpG recognition mechanisms, which underlie their differential enzymatic preference for CpG and non-CpG sites. (Fig. 2) These findings highlight the structural and functional divergence between the two *do novo* DNA methylation machineries. Through structure determination of DNMT3B in complex with CpG and CpA DNAs, this study reveals a multi-layered substrate recognition mechanism. First, the formation of a specific H-bond in the catalytic loop of DNMT3B gives rises to a reduced CpG/CpH specificity in DNMT3B, likely due to a conformational stabilization effect. Second, the different protein-DNA interactions at the RD interface of DNMT3B and DNMT3A diversify their intramolecular interaction between the RD interface and TRD, which triggers changes in

the side-chain conformations of TRD loop residues, including DNMT3B K777, S778 and N779, leading to a distinct TRD-DNA interaction of DNMT3B. Third, this study reveals a role for the TRD loop in fine-tuning substrate readout of DNMT3B. Strikingly, K777 undergoes significant side-chain conformational changes in response to changes in the major groove environments caused by different +1 bases, providing a mechanism for readout of +1 nucleotides by DNMT3B. This base shape-directed readout of the +1 flanking site by DNMT3B K777 on one hand may reduce the requirement for the N779-CpG contact when the CpG substrate is in a "favorable" flanking sequence context, on the other hand, shifts the preference of DNMT3B toward G at the +1 site on non-CpG substrates. This study therefore adds a new example to the growing family of DNA-interacting proteins that recognizes the DNA shape as a readout mechanism [33].

The previously proposed DNA co-methylation spacing model of the DNMT3A-DNMT3L tetramer highlights the importance of DNMT3A oligomerization in defining DNA methylation patterns [34, 35]. This model fits with genome-wide methylation analysis, which has indicated a 10-bp of methylation periodicity [36-38]. However, how such oligomerization of DNMT3A or DNMT3B contributes to the correlation of two neighboring methylation sites has not been precisely determined, partly due to the challenge in discerning DNMT3A or DNMT3B-mediated co-methylation over their distributive action in solution. Furthermore, it has been long argued that the 10-bp periodic distribution of the WW (W= A, T) and SS (SS= G, C) dinucleotides in nucleosomal DNA, which serves to accommodate the DNA wrapping around histone octamers[39], may dictate the 10-bp periodicity of DNA methylation[36]. Through single-turnover

methylation assay, we demonstrate that the most favorable co-methylation spacing of DNMT3A and DNMT3B is 14 bp, consist with the structural studies of both DNMT3A-DNMT3L-DNA and DNMT3B-DNMT3L-DNA complexes. How the 14-bp spacing mechanism described here may contribute to the DNA methylation patterns in cells, in particular in the context of the chromatin environment, remains to be elucidated. Nevertheless, it is apparent that the unique architecture of DNMT3A oligomer provides a basis for correlative methylation of two cytosines within a relatively short distance in general, through either co-methylation or a sequential enzymatic action. It is important to point out that the unique architecture of DNMT3A and DNMT3B oligomer provides a basis for correlative methylation of two cytosines within a relatively short distance in general, through either co-methylation or a distributive enzymatic action.

Together, through comprehensive structural, biochemical and enzymatic characterizations of DNMT3B and DNMT3A in CpG and non-CpG methylation, this study provides an unprecedented insight on DNMT3B- and DNMT3A-mediated de novo methylation. It is also worth mentioning that the genomic targeting of DNMT3A and DNMT3B in cells are further regulated by additional factors, including their N-terminal domains and DNMT3L. For instance, a previous study has demonstrated that DNMT3L modulates cellular de novo methylation activities through focusing the DNA methylation machineries on well-chromatinized DNA templates [40]. In addition, despite that structural analysis of both DNMT3B-DNA and DNMT3A-DNA complexes reveals that DNMT3L does not directly engage in the DNA interaction, the role of DNMT3L in stimulating the enzymatic activity and/or stability of DNMT3A and DNMT3B has been well

established[22, 23, 34, 41, 42], which may attenuate the impact of the intrinsic sequence specificities of these enzymes on the landscape of DNA methylation[40]. How the intrinsic preferences of DNMT3A and DNMT3B interplay with other cellular factors in regulating genomic DNA methylation awaits further investigation.

Figures

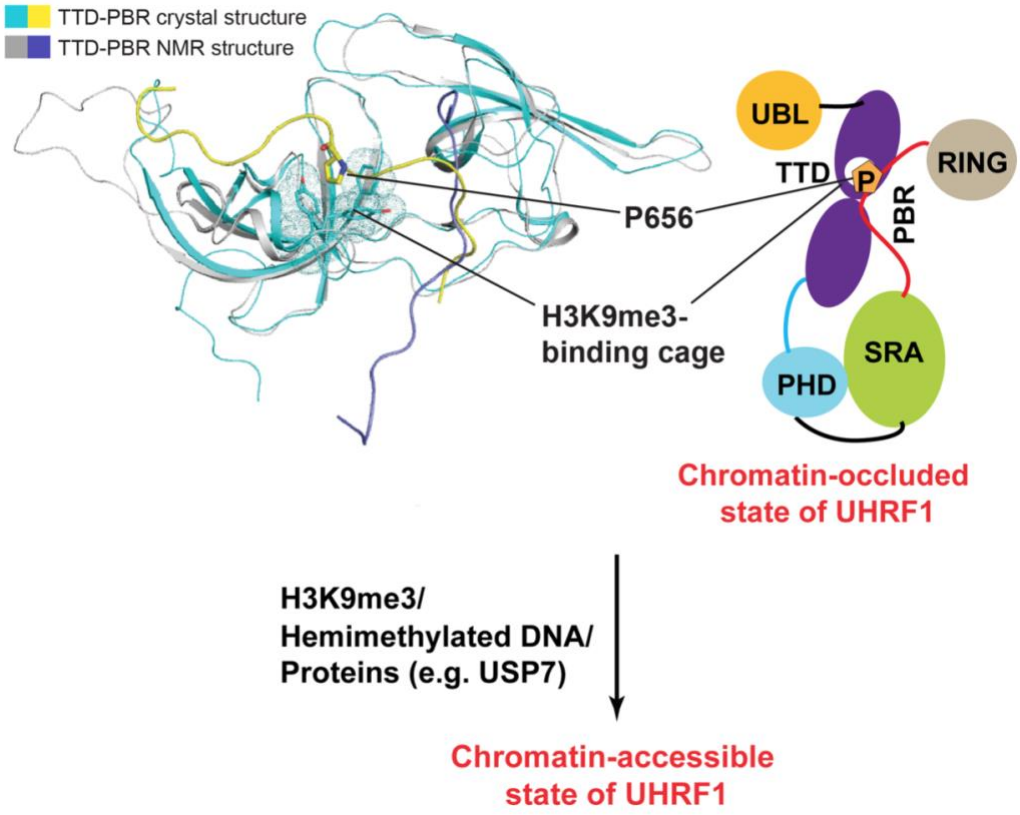


Figure 1. Structure comparison between UHRF1 TTD-PBR crystal structure and NMR structure and graphic demonstration of PBR-mediated UHRF1 allosteric regulation.

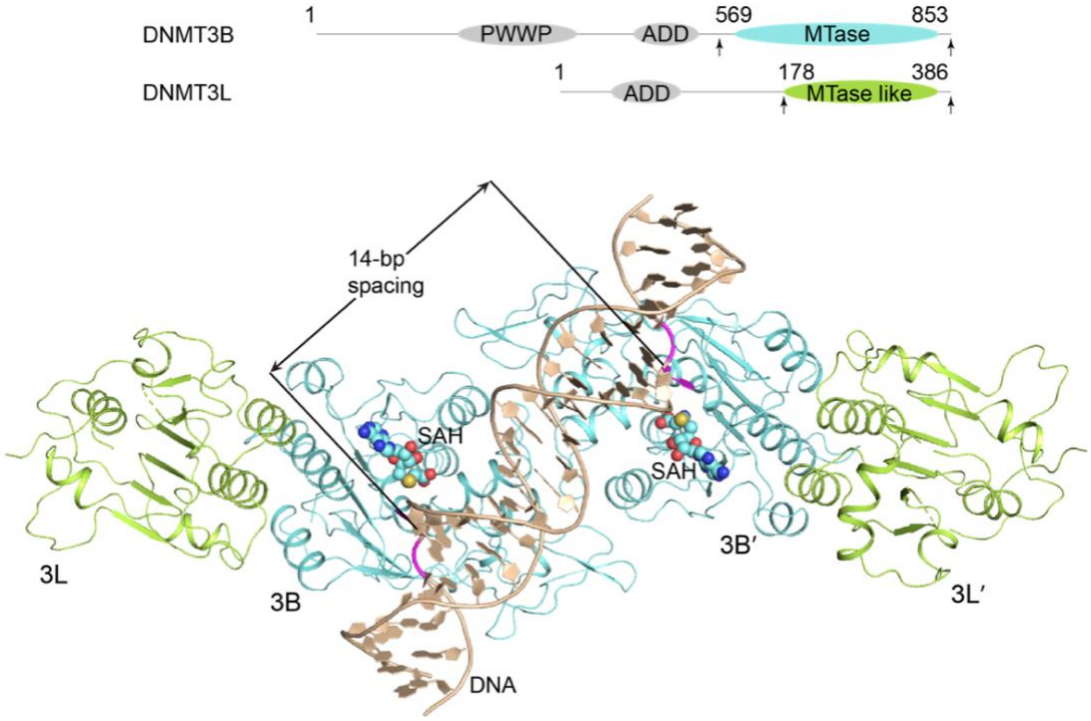


Figure 2. Structure view of DNMT3B-DNMT3L tetramer in complex with 25-mer DNA containing 2 CpG sites with 14-bp spacing.

Reference

1. Bostick, M., et al., *UHRF1 plays a role in maintaining DNA methylation in mammalian cells.* Science, 2007. **317**(5845): p. 1760-4.

2. Liu, X., et al., *UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9.* Nat Commun, 2013. **4**: p. 1563.

3. Miura, M., et al., *Dynamic changes in subnuclear NP95 location during the cell cycle and its spatial relationship with DNA replication foci.* Exp Cell Res, 2001. **263**(2): p. 202-8.

4. Sharif, J., et al., *The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA.* Nature, 2007. **450**(7171): p. 908-12.

5. Zhao, Q., et al., *Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals.* Nat Commun, 2016. **7**: p. 12464.

6. Fang, J., et al., *Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition.* Nat Commun, 2016. **7**: p. 11197.

7. Gelato, K.A., et al., *Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate.* Mol Cell, 2014. **54**(6): p. 905-19.

8. Harrison, J.S., et al., *Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1.* Elife, 2016. **5**.

9. Zhang, Z.M., et al., *An Allosteric Interaction Links USP7 to Deubiquitination and Chromatin Targeting of UHRF1.* Cell Rep, 2015. **12**(9): p. 1400-6.

10. Papait, R., et al., *Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing.* Mol Biol Cell, 2007. **18**(3): p. 1098-106.

11. Nishiyama, A., et al., *Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication.* Nature, 2013. **502**(7470): p. 249-53.

12. Qin, W., et al., *DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination.* Cell Res, 2015.

13. Chen, T., et al., *Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b.* Mol Cell Biol, 2003. **23**(16): p. 5594-605.

14. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.

15. Aoki, A., et al., *Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases.* Nucleic Acids Res, 2001. **29**(17): p. 3506-12.

16. Baubec, T., et al., *Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation.* Nature, 2015. **520**(7546): p. 243-7.

17. Challen, G.A., et al., *Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells.* Cell Stem Cell, 2014. **15**(3): p. 350-364.

18. Gowher, H. and A. Jeltsch, *Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases.* J Biol Chem, 2002. **277**(23): p. 20409-14.

19. Hsieh, C.L., *In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b.* Mol Cell Biol, 1999. **19**(12): p. 8211-8.

20. Norvil, A.B., et al., *Dnmt3b Methylates DNA by a Noncooperative Mechanism, and Its Activity Is Unaffected by Manipulations at the Predicted Dimer Interface.* Biochemistry, 2018. **57**(29): p. 4312-4324.

21. Suetake, I., et al., *Distinct enzymatic properties of recombinant mouse DNA methyltransferases Dnmt3a and Dnmt3b.* J Biochem, 2003. **133**(6): p. 737-44.

22. Bourc'his, D., et al., *Dnmt3L and the establishment of maternal genomic imprints.* Science, 2001. **294**(5551): p. 2536-9.

23. Hata, K., et al., *Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice.* Development, 2002. **129**(8): p. 1983-93.

24. Chedin, F., *The DNMT3 family of mammalian de novo DNA methyltransferases.* Prog Mol Biol Transl Sci, 2011. **101**: p. 255-85.

25. Liao, J., et al., *Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells.* Nat Genet, 2015. **47**(5): p. 469-78.

26. Laurent, L., et al., *Dynamic changes in the human methylome during differentiation.* Genome Res, 2010. **20**(3): p. 320-31.

27. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-22.

28. Lister, R., et al., *Global epigenomic reconfiguration during mammalian brain development.* Science, 2013. **341**(6146): p. 1237905.

29. Xie, W., et al., *Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome.* Cell, 2012. **148**(4): p. 816-31.

30. Guo, J.U., et al., *Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain.* Nat Neurosci, 2014. **17**(2): p. 215-22.

31. Lee, J.H., S.J. Park, and K. Nakai, *Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s.* Sci Rep, 2017. **7**(1): p. 11295.

32. Lister, R., et al., *Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells.* Nature, 2011. **471**(7336): p. 68-73.

33. Rohs, R., et al., *Origins of specificity in protein-DNA recognition.* Annu Rev Biochem, 2010. **79**: p. 233-69.

34. Jia, D., et al., *Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation.* Nature, 2007. **449**(7159): p. 248-51.

35. Jurkowska, R.Z., et al., *Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L.* Nucleic Acids Res, 2008. **36**(21): p. 6656-63.

36. Chodavarapu, R.K., et al., *Relationship between nucleosome positioning and DNA methylation.* Nature, 2010. **466**(7304): p. 388-92.

37. Zhang, X., et al., *Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis.* Cell, 2006. **126**(6): p. 1189-201.

38. Zhang, Y., et al., *DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution.* PLoS Genet, 2009. **5**(3): p. e1000438.

39. Widom, J., *Role of DNA sequence in nucleosome stability and dynamics.* Q Rev Biophys, 2001. **34**(3): p. 269-324.

40. Wienholz, B.L., et al., *DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo.* PLoS Genet, 2010. **6**(9): p. e1001106.

41. Chedin, F., M.R. Lieber, and C.L. Hsieh, *The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16916-21.

42. Veland, N., et al., *DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells.* Nucleic Acids Res, 2018.