

UC Riverside

UC Riverside Previously Published Works

Title

The structure of DNA methyltransferase DNMT3C reveals an activity-tuning mechanism for DNA methylation.

Permalink

<https://escholarship.org/uc/item/2461750g>

Journal

Journal of Biological Chemistry, 300(9)

Authors

Khudaverdyan, Nelli

Lu, Jiuwei

Chen, Xinyi

et al.

Publication Date

2024-09-01

DOI

10.1016/j.jbc.2024.107633

Peer reviewed



The structure of DNA methyltransferase DNMT3C reveals an activity-tuning mechanism for DNA methylation

Received for publication, March 27, 2024, and in revised form, July 20, 2024. Published, Papers in Press, August 2, 2024.
<https://doi.org/10.1016/j.jbc.2024.107633>

Nelli Khudaverdyan¹ , Jiuwei Lu¹, Xinyi Chen¹ , Genevieve Herle², and Jikui Song^{1,2,*}

From the ¹Department of Biochemistry, and ²Biophysics Program, University of California, Riverside, California, USA

Reviewed by members of the JBC Editorial Board. Edited by Brian D. Strahl

DNA methylation is one of the major epigenetic mechanisms crucial for gene regulation and genome stability. *De novo* DNA methyltransferase DNMT3C is required for silencing evolutionarily young transposons during mice spermatogenesis. Mutation of DNMT3C led to a sterility phenotype that cannot be rescued by its homologs DNMT3A and DNMT3B. However, the structural basis of DNMT3C-mediated DNA methylation remains unknown. Here, we report the structure and mechanism of DNMT3C-mediated DNA methylation. The DNMT3C methyltransferase domain recognizes CpG-containing DNA in a manner similar to that of DNMT3A and DNMT3B, in line with their high sequence similarity. However, two evolutionary covariation sites, C543 and E590, diversify the substrate interaction among DNMT3C, DNMT3A, and DNMT3B, resulting in distinct DNA methylation activity and specificity between DNMT3C, DNMT3A, and DNMT3B *in vitro*. In addition, our combined structural and biochemical analysis reveals that the disease-causing *rahu* mutation of DNMT3C compromises its oligomerization and DNA-binding activities, explaining the loss of DNA methylation activity caused by this mutation. This study provides a mechanistic insight into DNMT3C-mediated DNA methylation that complements DNMT3A- and DNMT3B-mediated DNA methylation in mice, unraveling a regulatory mechanism by which evolutionary conservation and diversification fine-tune the activity of *de novo* DNA methyltransferases.

DNA methylation is an evolutionarily conserved epigenetic mechanism that is important for silencing retrotransposons, which account for ~40% of the mammalian genome (1–3). Retrotransposon silencing is essential for proper mammalian development, such as embryogenesis, gametogenesis, and sexual reproduction (4, 5). Failure to silence retrotransposons leads to genomic instability and consequently affects the fitness of subsequent generations (6). Germline is most vulnerable to this threat as some of the retrotransposons escape from the first wave of DNA methylation in early embryogenesis and must be silenced during the reprogramming of gametogenesis to avoid havoc in germ cells (6, 7).

The establishment of DNA methylation in mammals is mainly orchestrated by three paralogs of DNA

methyltransferases (DNMTs): DNMT3A, DNMT3B, and DNMT3C, and an enzymatically inactive regulator DNMT3L (6, 8–15). Among these, DNMT3C was evolutionarily derived from tandem duplication of DNMT3B but lacks a corresponding N-terminal Proline-Tryptophan-Tryptophan-Proline (PWWP) domain (6). The three members of the *de novo* DNA methyltransferase family share similar but distinctive functionalities during development, which is in part attributed to the diversification selection of their regulatory and catalytic domains (16, 17). It has been demonstrated that DNMT3A is required for methylating major satellite repeats and allele-specific imprinting during gametogenesis, and DNMT3B is important for methylation of minor satellite repeats during embryogenesis; knock out of either DNMT3A or DNMT3B in a mouse model led to prenatal fatality or death shortly after birth (18), indicating that both DNMTs are indispensable. In contrast, DNMT3C forms a methyltransferase complex with DNMT3L to exclusively methylate and silence evolutionarily young yet transcriptionally active transposons, such as L1, EVKs, and IAPs, in the male germline (6, 7). DNMT3C mutation, such as the so-called *rahu* (recombination affected with hypogonadism from under-populated testes) mutation (E693G), which was identified through a forward genetics screen, led to meiotic failure defecting spermatogenesis, thereby affecting male sterility of mice (7).

Previous studies of DNMT3A and DNMT3B in the context of DNMT3A-DNMT3L and DNMT3B-DNMT3L complexes revealed a DNMT3L-DNMT3A/DNMT3B-DNMT3A/DNMT3B-DNMT3L tetrameric assembly, mediated by both the polar, homodimeric interface (RD interface) and the non-polar, heterodimeric interface (FF interface) (16, 19–23). The interaction between DNMT3A or DNMT3B methyltransferase (MTase) domain and DNA substrates is mediated by a loop from the target recognition domain (TRD loop), an α -helix at the RD interface (RD helix), and the catalytic loop (16, 19, 22, 23). Of note, the catalytic loop and the TRD loop coordinately govern the CpG specificity of DNMT3A/DNMT3B. An intramolecular hydrogen bond in the catalytic loop, present between DNMT3B N656 and R661 but not between the corresponding sites (I715 and R720) of DNMT3A, gives rise to a higher methylation activity but a lower CpG specificity for DNMT3B than DNMT3A (16). Introducing DNMT3A-converting mutation to DNMT3B or DNMT3B-converting mutations to DNMT3A on these sites led to a switch of the

* For correspondence: Jikui Song, jikui.song@ucr.edu.

Structure of DNMT3C-DNA complex

activity and substrate specificity between the two enzymes (16). In addition, DNMT3A and DNMT3B show different context-dependent interactions with the +1 CpG-flanking site, resulting in their distinct flanking sequence preference (16). On the other hand, how DNMT3C, the MTase domain of which bears ~92% and ~77% sequence identity with the corresponding regions of DNMT3B and DNMT3A, respectively (Fig. S1), attains its substrate recognition and specificity remains unknown.

To provide the molecular basis for DNMT3C-mediated DNA methylation, we determined the crystal structure of the DNMT3C MTase domain in complex with DNMT3L and CpG-containing DNA. Whereas the substrate-recognition mechanism of DNMT3C resembles that of DNMT3A and DNMT3B, the DNA methylation activity of DNMT3C is restricted by the unique presence of a glutamate (E590) in the DNA-contact sites. Furthermore, our combined sequence, structural, and biochemical analysis reveals that residues C543 and E590 constitute two coevolving sites that impact the catalytic-loop conformation of DNMT3C, DNMT3A, and DNMT3B differentially, which consequently leads to distinct DNA methylation activity and specificity between DNMT3C, DNMT3A and DNMT3B. In addition, our biochemical and enzymatic analysis of the *rahu* mutant of DNMT3C, residing near to the RD interface, reveals a defect in protein oligomerization and DNA-binding activity, thereby explaining the loss of DNA methylation activity caused by this mutation. Together, this study reveals how evolutionary conservation and diversification fine-tune the enzymatic activity and specificity of *de novo* DNA methyltransferases, which may contribute to their distinct functionalities.

Results

Structural overview of the DNMT3C-DNMT3L-CpG DNA complex

The complex of DNMT3C with DNMT3L and CpG DNA was prepared as previously described for the DNMT3A-DNMT3L-DNA and DNMT3B-DNMT3L-DNA complexes (16, 19, 22, 23). For the purpose of crystallization, we focused on the mouse DNMT3C MTase domain and hDNMT3L C-terminal domain for structural determination (Fig. 1A). Note that the C-terminal domains of hDNMT3L or mDNMT3L, which do not engage in DNA interaction (16, 19, 22, 23), share similar oligomerization-interface residues (Fig. S2A). Accordingly, a comparison of the *in vitro* DNA methylation activity of the DNMT3C-mDNMT3L and DNMT3C-hDNMT3L on CpG-, CpA-, and CpT-containing DNAs revealed a similar substrate specificity on CpG- and non-CpG-containing DNAs (Fig. S2, B–D).

The DNA substrate is a 25-mer, self-complementary DNA containing a zebularine followed by guanine (ZpG) site near one end and a CpG site near the other end. Annealing of this DNA results in a duplex with two ZpG/CpG sites in a spacing of 14 base pairs (Fig. 1B). As established previously (16, 19, 23–25), the replacement of the target cytosine with a zebularine permits the formation of a stable DNMT3C-DNA

covalent adduct. The crystal structure of the DNMT3C-DNMT3L-DNA complex bound to *S*-adenosyl-homocysteine (SAH), the by-product of cofactor *S*-adenosyl-methionine (SAM), was solved at 3.2-Å resolution (Fig. 1C and Table S1).

The structure of the DNMT3C-DNMT3L-DNA complex reveals a heterotetrameric assembly, organized in a DNMT3L-DNMT3C-DNMT3C-DNMT3L fashion reminiscent of those previously observed for the DNMT3A-DNMT3L-DNA and DNMT3B-DNMT3L-DNA complexes (Fig. 1C) (16, 19, 22, 23). As previously observed for the DNMT3A and DNMT3B complexes, the interaction between DNMT3C and DNA is mainly mediated by three distinct regions: TRD loop (residues 660–675) that interacts with DNA major groove, the catalytic loop (residues 535–559) that interacts with the DNA minor groove, and the RD helix that interacts with the backbone of the central DNA segment (Figs. 1D and S3, A and B). At the active site of each DNMT3C subunit, the zebularine (Z5/Z5') on the target strand is flipped out of DNA helix and covalently bound to catalytic cysteine C538 of DNMT3C (Fig. 1E). In addition, the base of Z5/Z5' is stabilized *via* hydrogen-bonding interactions with residues E584, R618, and R620 in the active site (Fig. 1E).

It is worth mentioning that in this study, the DNMT3C-bound DNA contains a guanine (G7) at the +1 ZpG-flanking site (Fig. 1B). Structural comparison of the DNMT3C-DNMT3L-CGG DNA complex with human DNMT3B-DNMT3L-CGG DNA complex (PDB 6KDA) and human DNMT3A-DNMT3L-CGA DNA (PDB 6W8B) reveals high structural similarity, with root-mean-square deviation (RMSD) of 0.49 Å and 0.48 Å over 548 and 545 aligned C α atoms, respectively (Fig. 1, F and G), in line with the high sequence identity between the MTase domain of DNMT3C and that of DNMT3B or DNMT3A (Fig. S1).

Residue E590 restricts the DNA methylation activity of DNMT3C

Toward the DNA minor groove, the DNMT3C MTase domain interacts with both DNA strands *via* the catalytic loop, involving residues C538, S542, C543, V544, and P546 (Figs. 2A and S3B). In addition to the covalent bonding between DNMT3C C538 and Z5', DNMT3C V544 extends its side chain to occupy the DNA cavity vacated by Z5', stabilizing the unpaired Gua20 *via* a hydrogen bond between the backbone carbonyl group of V544 and the N2 atom of the Gua20 (Fig. 2A). Furthermore, the backbone carbonyl of C543 receives a hydrogen bond from the N2 atom of Gua4; the sidechain hydroxyl group of S542 donates a hydrogen bond to the backbone phosphate of Z5'; and the prolyl ring of P546 interacts with the sugar/base moieties of Gua6'/Gua20 *via* van der Waals contacts (Fig. 2A). In addition to the catalytic loop, the α 5-helix located at the DNMT3C-DNMT3L interface further stabilizes the DNMT3C-DNA association *via* a backbone hydrogen bond between DNMT3C V591 and Thy23 (Fig. 2A).

Structural comparison of the DNMT3C-DNMT3L-DNA complex with DNMT3A-DNMT3L-DNA and DNMT3B-DNMT3L-DNA complexes reveal subtle but notable

Structure of DNMT3C-DNA complex

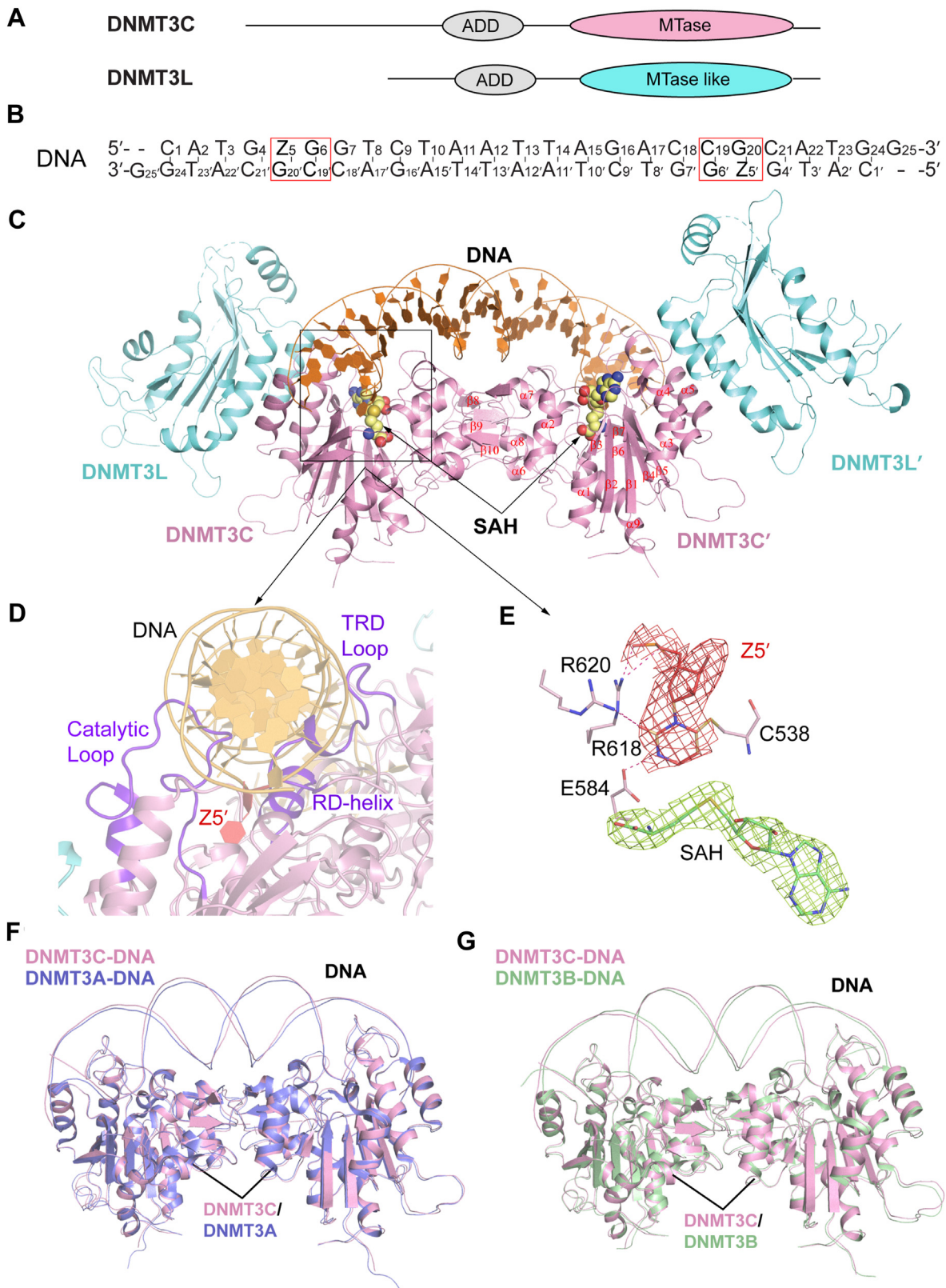


Figure 1. Structure of the DNMT3C-DNMT3L tetramer bound to a CpG DNA. *A*, domain architecture of DNMT3C and DNMT3L. *B*, the DNA sequence used for structural study. Z denotes zebularine, a cytosine analog. *C*, structure of DNMT3C-DNMT3L bound to CpG DNA and SAH (sphere representation). *D*, close-up view of DNA-interacting RD helix, TRD loop, and Catalytic loop. *E*, close-up view of the catalytic site of DNMT3C bound to ZpG/CpG DNA. Fo-Fc omit maps for the flipped zebularine (Z5') and the SAH molecule at 2 σ contour level were shown as salmon and green mesh, respectively. The hydrogen bonds are shown as dashed lines. *F*, structural overlay of the DNMT3C-DNMT3L-DNA and DNMT3A-DNMT3L-DNA complex (PDB 6W8B). *G*, structural overlay of the DNMT3C-DNMT3L-DNA and DNMT3B-DNMT3L-DNA complex (PDB 6KDA).

Structure of DNMT3C-DNA complex

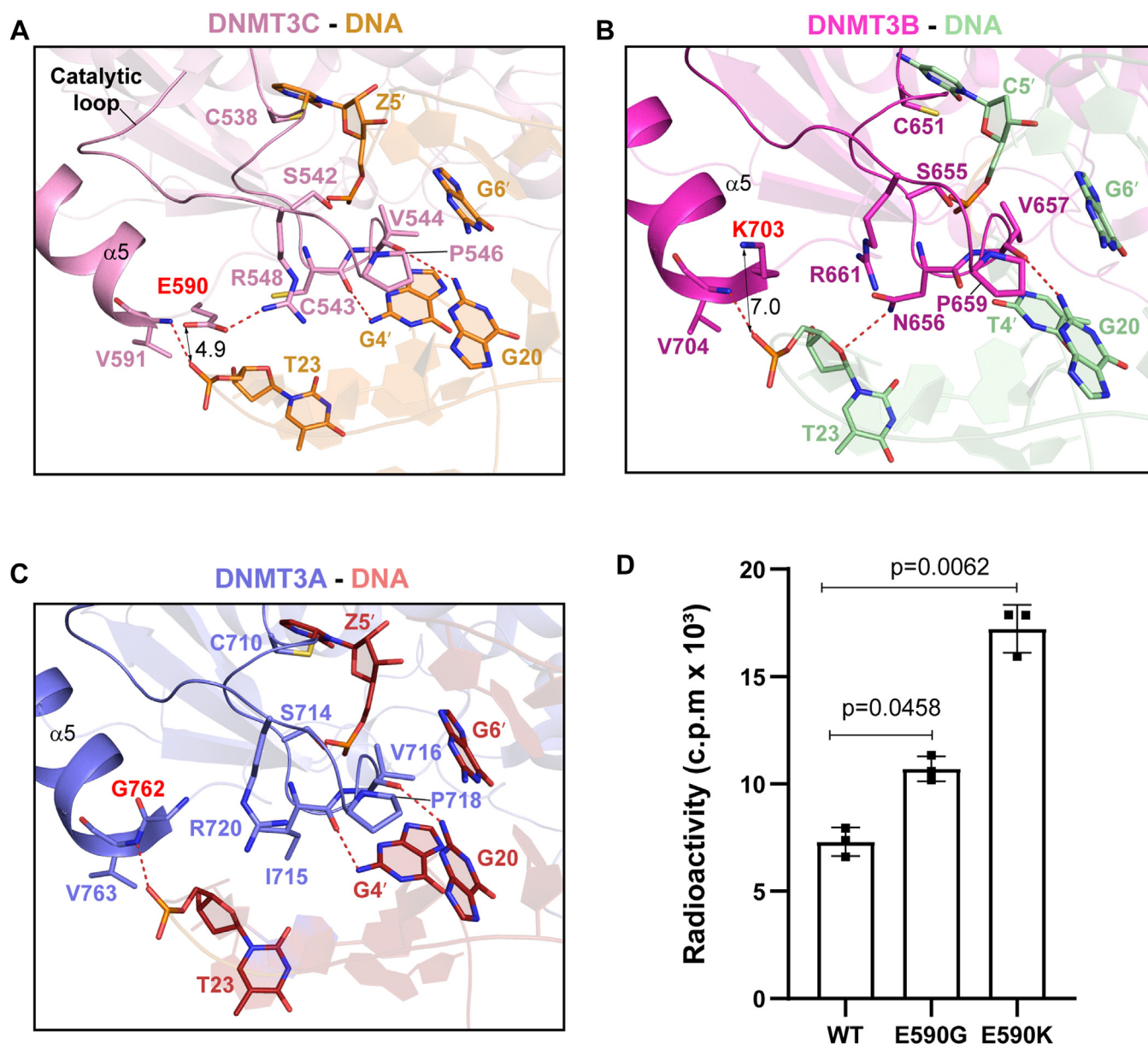


Figure 2. Amino acid variation at the DNMT3C E590-corresponding site leads to distinct DNA interaction between DNMT3C, DNMT3B, and DNMT3A. *A*, close-up view of the interactions between DNMT3C E590, the catalytic loop of DNMT3C, and DNA. *B*, close-up view of the interactions between DNMT3C E590-corresponding site, DNMT3B K703, the catalytic loop of DNMT3B, and DNA. *C*, close-up view of the interactions between DNMT3C E590-corresponding site, DNMT3A G762, the catalytic loop of DNMT3A, and DNA. *D*, *in vitro* DNA methylation activities of DNMT3C, WT or mutant, on CpG DNA. Data are mean \pm s.d. ($n = 3$ biological replicates). Statistical analysis for WT DNMT3C vs mutants used a two-tailed Student's *t* test.

differences in the minor-groove interaction (Fig. 2, A–C). The most striking difference lies in the N-terminus of $\alpha 5$ -helix, where DNMT3C E590 is replaced by a glycine (G762) in DNMT3A but a lysine (K703) in DNMT3B (Fig. 2, A–C). In the DNMT3C-DNMT3L-DNA complex, the side chain of residue E590 forms a salt-bridge with catalytic-loop residue R548, which places E590 in proximity with the backbone of Thy23 for electrostatic repulsion (Fig. 2A). In contrast, the charge reversal for the corresponding site (K703) in DNMT3B leads to an electrostatic attraction with the DNA backbone in the DNMT3B-DNMT3L-DNA complex, thereby contributing to the substrate binding of DNMT3B (Fig. 2B). Different from DNMT3C and DNMT3B, the corresponding residue G762 in

the DNMT3A-DNMT3L-DNA does not engage in appreciable DNA interaction (Fig. 2C). Sequence analysis of DNMT3C across various species indicates that residue E590 is highly conserved throughout evolution (Fig. S4), implying an important role of this residue in DNMT3C function.

To examine the effect of amino acid variation on the DNMT3C E590-corresponding site between DNMT3C, DNMT3A, and DNMT3B, we compared the *in vitro* DNA methylation activities of DNMT3C, DNMT3A, DNMT3B, as well as DNMT3A-converting mutant E590G and DNMT3B-converting mutant E590K of DNMT3C, on a 36-mer CpG-containing DNA duplex. Under the experimental condition, we observed a differential DNA methylation activity for

DNMT3C, DNMT3A, and DNMT3B, with DNMT3B possessing a much higher activity than DNMT3C and DNMT3A (Fig. S5A). On the other hand, introducing the DNMT3A-converting E590G mutation and DNMT3B-converting E590K mutation to DNMT3C both substantially increase its DNA methylation efficiency, with the E590K mutation exhibiting a stronger effect that matches the relatively higher activity of DNMT3B (Fig. 2D). These data support the structural observations that the amino acid variation at the DNMT3C E590-corresponding sites influences the DNA interaction of DNMT3C, DNMT3A, and DNMT3B differentially, which contributes to the different DNA methylation activities among the DNMT3 family of DNA methyltransferases.

Residues C543 and E590 coordinately modulate the DNA methylation activity and specificity of DNMT3C

Sequence comparison of the catalytic loops of DNMT3A, DNMT3B, and DNMT3C reveals three sites that undergo amino acid variations, which correspond to DNMT3C C543, V547, and F552 (Figs. S1 and S4). Among these, the DNMT3C C543-corresponding site is involved in DNA interaction (Fig. 2, A–C). Along the line, previous studies have revealed that the DNA-bound DNMT3A and DNMT3B proteins exhibit a similar but distinct catalytic-loop conformation: the catalytic loop of DNMT3B harbors a hydrogen bond between residues N656 and R661, whereas no such interaction was observed for the corresponding pair (I715 and R720) in DNMT3A (Figs. 3, A and B and S5B) (16, 19, 22, 23). Such a conformational difference is associated with a higher DNA methylation efficiency but a reduced CpG specificity for DNMT3B than DNMT3A, likely caused by differential DNA contacts by the catalytic-loop residues (e.g. DNMT3A V716 and P718-corresponding residues) (Fig. 3, A and B) (16).

Given the fact that the corresponding pair C543/R548 in DNMT3C, as with that in DNMT3A, does not form a hydrogen bond within the pair (Fig. 3, C and D), we set out to determine whether the amino acid variation on the DNMT3C C543-corresponding site influences the DNA methylation activity and substrate specificity of DNMT3C. Toward this, we first compared *in vitro* DNA methylation activities between WT DNMT3C, DNMT3A, and DNMT3B on CpG-, CpA- and CpT-containing DNA substrates. Among the three *de novo* DNA methyltransferases, DNMT3B methylates CpG and non-CpG DNAs both in the highest efficiency under the experimental condition (Fig. S5A). As previously observed (16), whereas DNMT3B shows a substrate preference in the order of CpG > CpA > CpT (Fig. S5A), the relatively high methylation activity of DNMT3B on the non-CpG DNA results in a CpG specificity (relative methylation efficiency of ~1.9-fold for CpG/CpA and ~7.2-fold for CpG/CpT) lower than that of DNMT3A (relative methylation efficiency of 30.0-fold for CpG/CpA and 45.2-fold for CpG/CpT) (Fig. S5A). Interestingly, whereas DNMT3C methylates the CpG DNA in an efficiency similar to that of DNMT3A, it shows a modestly higher methylation efficiency on non-CpG DNAs, especially CpA DNA (Fig. S5A), which gives rise to a CpG specificity

(relative methylation efficiency of ~3.9-fold CpG/CpA and ~14.3-fold CpG/CpT) closer to that of DNMT3B. Next, we mutated DNMT3C C543 into isoleucine (C543I) or asparagine (C543N), mimicking the corresponding sites in DNMT3A and DNMT3B, respectively. Our *in vitro* DNA methylation analysis revealed that introducing the C543I and C543N mutations to DNMT3C each led to a notable decrease in DNA methylation efficiency on CpA- and CpG-containing DNAs (Fig. 3E). This observation suggests that introducing the C543I or C543N mutation alone to DNMT3C is insufficient to convert its DNA methylation activity and specificity toward those of DNMT3A and DNMT3B, respectively.

As described earlier, the DNA-bound DNMT3A, DNMT3B, and DNMT3C also show distinct interactions between the DNMT3C E590- and R548-corresponding sites, with an intramolecular hydrogen bond formed between residues R548 and E590 of DNMT3C but not for the corresponding pairs in DNMT3A or DNMT3B (Fig. 3, A–D). These observations therefore establish a link between DNMT3C C543-, R548- and E590-corresponding sites, prompting us to investigate a potential interplay between the DNMT3C E590-corresponding site and the catalytic loop. Indeed, introducing the DNMT3B-converting C543N/E590K double mutation to DNMT3C greatly increased its methylation efficiencies on CpG, CpG and CpT, resulting in a CpG specificity (relative methylation efficiency of ~1.6-fold for CpG/CpA and ~3.6-fold for CpG/CpT) similar to that of DNMT3B (Figs. 3F vs. S5A). Likewise, introducing the DNMT3A-converting C543I/E590G double mutation to DNMT3C shifted both its DNA methylation efficiencies and CpG specificity (relative methylation efficiency of ~7.0-fold for CpG/CpA and ~21.9-fold for CpG/CpT) toward those of DNMT3A (Figs. 3F vs. S5A). Note that the relative CpG specificity for the C543I/E590G-mutated DNMT3C remains lower than that of DNMT3A, likely owing to the differential DNA contacts by their respective TRD loops (See details below). Next, we performed steady-state kinetic assays to compare the enzymatic kinetics of DNMT3C and its DNMT3A- or DNMT3B-converting mutants on CpG DNA (Fig. S6). In the context of DNMT3C-DNMT3L complex, a k_{cat} of 0.025 min⁻¹ was determined for WT DNMT3C (Fig. S6F). Consistent with the single time-point measurements described above, introducing the single mutations C543N and C543I decreased the k_{cat} of DNMT3C to 0.014 min⁻¹ and 0.011 min⁻¹, respectively, whereas introducing the double mutation C543N/E590K and C543I/E590G increased the k_{cat} to 0.074 min⁻¹ and 0.026 min⁻¹, respectively (Fig. S6F). Together, these observations suggest that the DNMT3C C543- and E590-corresponding sites coordinately modulate the DNA methylation activities and specificities of DNMT3A, DNMT3B, and DNMT3C.

Evolutionary covariation between DNMT3C C543- and E590-corresponding sites

Next, we performed sequence analysis between the subfamilies of DNMT3A, DNMT3B, and DNMT3C, focusing on the DNA-binding sites of the catalytic core. As expected from

Structure of DNMT3C-DNA complex

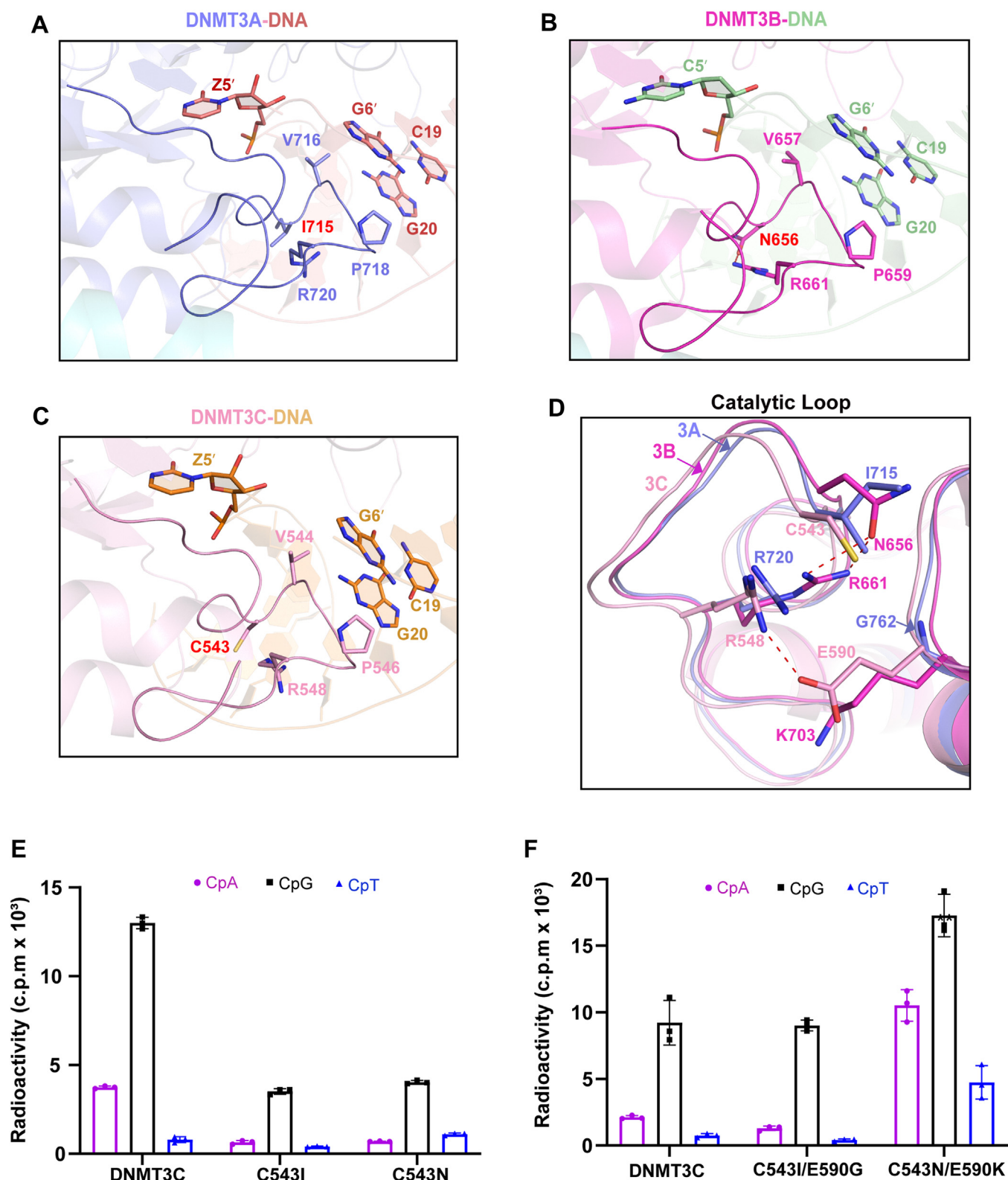


Figure 3. An intramolecular interaction of the catalytic loop controls the CpG specificities of DNMT3A, DNMT3B, and DNMT3C. A–C, structural comparison of the DNMT3C-DNA, DNMT3A-DNA (PDB 6W8B), and DNMT3B-DNA (PDB 6KDA) complexes, highlighting distinct intramolecular interactions within the catalytic loops of DNMT3A (A), DNMT3B (B) and DNMT3C (C). D, close-up view of the overlaid catalytic loops of DNMT3A, DNMT3B, and DNMT3C, the DNMT3C C543-, R548-, and E590-corresponding sites shown in stick representation. The hydrogen-bonding interactions are shown as dashed lines. E and F, *in vitro* DNA methylation activities of DNMT3C, WT or that harboring single (E) or double (F) mutations, on CpA, CpG, and CpT DNAs. Data are mean \pm s.d. (n = 3 biological replicates).

their close evolutionary link, the catalytic loop-surrounding region is highly conserved among DNMT3A, DNMT3B, and DNMT3C (Fig. S4). Nevertheless, several amino acid variations were mapped onto the catalytic loop and its subsequent helix ($\alpha 4$), as well as the spatially proximate $\alpha 5$ -helix (Fig. S4). Among these are the DNMT3C C543- and E590-corresponding sites, which are strictly conserved within each DNMT3A, DNMT3B, or DNMT3C subfamily but distinct between subfamilies (Fig. S4). This observation, together with the fact that the DNMT3C C543- and E590-corresponding sites coordinately modulate the DNA methylation activity and specificity of DNMT3C, DNMT3A, and DNMT3B (Fig. S4), suggests that DNMT3C C543- and E590-corresponding sites have coevolved to confer distinct enzymatic activities and specificities of DNMT3C, DNMT3A and

DNMT3B. How such an activity-tuning mechanism influences the respective DNA methylation by DNMT3A, DNMT3B, and DNMT3C *in vivo* awaits further investigation.

DNMT3C TRD loop- and RD helix-mediated DNA interactions

It has been established that the TRD loops of DNMT3A and DNMT3B modulate their respective CpG specificity as well as flanking sequence preference through interacting with the CpG and the +1-flanking sites (16, 19, 22, 23). Notably, DNMT3A and DNMT3B differ in the DNMT3B K777-corresponding site, which is an arginine in DNMT3A (R836) (Fig. 4, A and B). Previous studies from others and we have further demonstrated that DNMT3A and DNMT3B involve different context-dependent CpG interactions: in the context

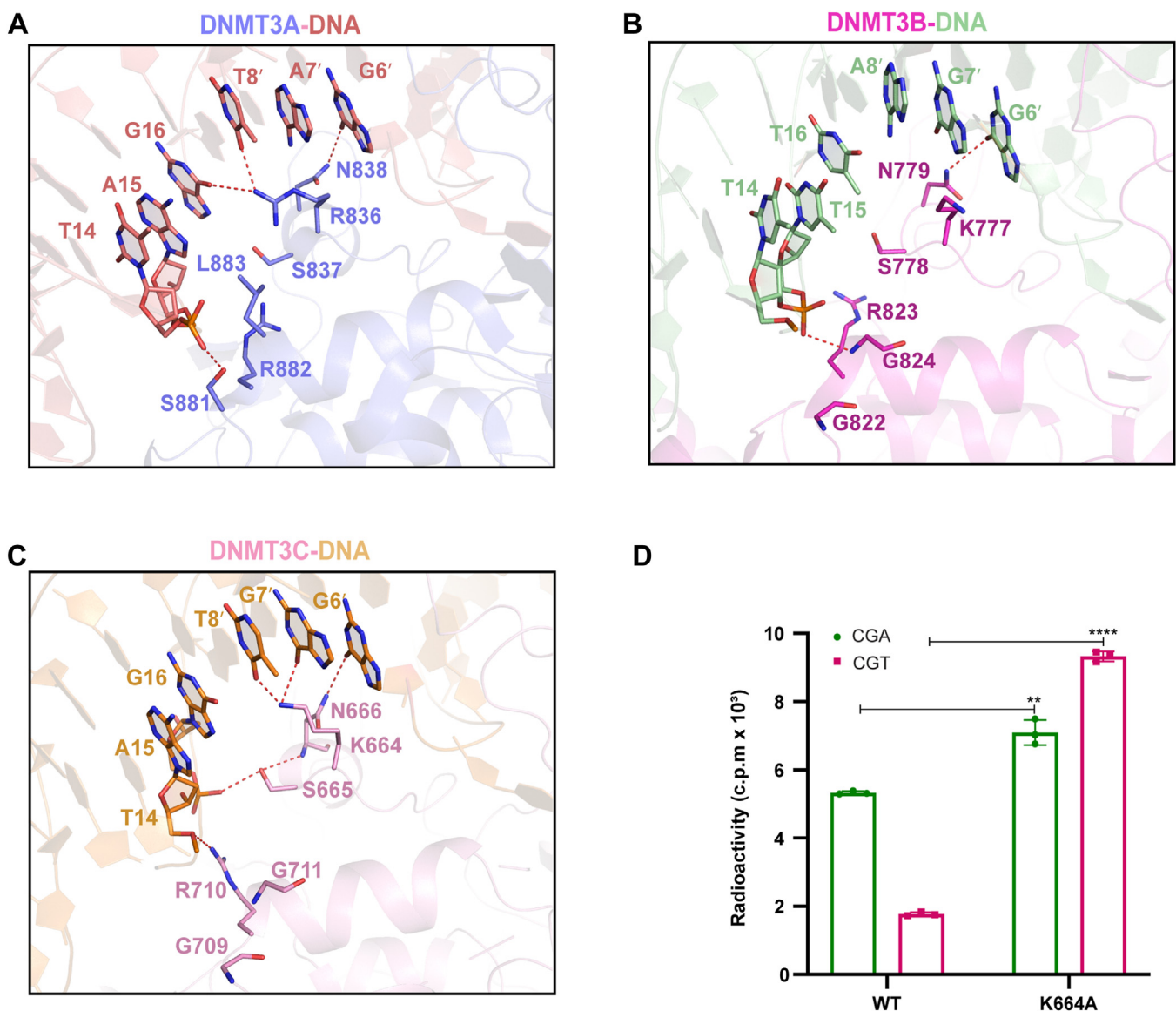


Figure 4. The TRD loop- and RD helix-mediated DNA interactions of DNMT3C resemble those of DNMT3B. Structural overlay of the TRD loop of DNMT3C-DNA, DNMT3A-DNA, and DNMT3B-DNA complexes, with DNA-interacting residues and the corresponding DNAs shown in stick representation. A–C, close-up view of the DNA interaction involving the TRD loop and RD helix of DNMT3A (A), DNMT3B (B), and DNMT3C (C). Hydrogen-bonding interactions are shown as dashed lines. D, *in vitro* DNA methylation activity of DNMT3C, WT or K664A, on CGA and CGT DNAs. Data are mean \pm s.d. (n = 3 biological replicates). The two-tailed Student's *t* test was used for statistical analysis for WT DNMT3C vs K664A. ***p* < 0.01. *****p* < 0.0001.

Structure of DNMT3C-DNA complex

of the CGA or CGG motif, both DNMT3A R836 and DNMT3B K777 interact with the CpG-flanking sites (Fig. 4, A and B) (16, 19, 22, 23); in the context of CGT or CGC motif, DNMT3A R886 engages in a hydrogen-bonding interaction with CpG guanine, while DNMT3B K777 remains engaged with the +1/+2-flanking sites (16, 19, 22, 23). Such a difference in context-dependent CpG recognition between DNMT3A and DNMT3B coincides with the fact that DNMT3A prefers a pyrimidine over purine at the +1-flanking site, while DNMT3B prefers a purine over pyrimidine on the +1-flanking site (16, 26, 27).

DNMT3C contains a TRD loop that shares an identical sequence with the corresponding region of DNMT3B (Fig. S1). Consistently, structural comparison of the DNMT3C-CGG DNA and DNMT3B-CGG DNA interactions reveals a similar conformation for the TRD loop (Fig. 4, B and C). Of note, DNMT3C N666 interacts with CpG guanine (Gua6) *via* a hydrogen-bonding interaction, as with the corresponding DNMT3B N779 in the CGG complex and the corresponding DNMT3A N838 in the CGA complexes (Fig. 4, A–C). Furthermore, DNMT3C K664 engages hydrogen-bonding and/or van der Waals contacts with the +1- and +2-flanking sites in a manner similar to the corresponding DNMT3A R836 in the CGA complex and DNMT3B K777 in the CGG complex (Fig. 4, A–C), reinforcing the notion that the corresponding residue mediates the recognition of CpG-flanking sequences. Nevertheless, in comparison with the DNMT3B K777, DNMT3C K664 appears closer to the +1- and +2-flanking sites for hydrogen-bonding interactions, presumably due to the fact that the DNA molecules in the two complexes used for structural study differ in the +2 site (A8' in the DNMT3B-CGG vs T8' in the DNMT3C-CGG; Fig. 4, B and C). To test the role of residue K664 on the flanking sequence preference of DNMT3C, we performed *in vitro* DNA methylation assays for DNMT3C, WT and K664A mutant, on (CGA)₁₂ and (CGT)₁₂ DNA duplexes, in which the CpG sites on the complementary strands are methylated. WT DNMT3C shows a higher methylation efficiency on CGA DNA than on CGT DNA (Fig. 4D). However, introducing the K664A mutation not only increased the overall DNA methylation efficiency of DNMT3C on both CGT and CGA DNAs but also shifted the relative methylation flanking preference from CGA to CGT (Fig. 4D). These observations recapitulate what was previously observed for the corresponding DNMT3B K777A mutation (16), confirming that the TRD loops of DNMT3C and DNMT3B share conserved CpG-recognition mechanisms. A detailed comparison of the flanking sequence preference between DNMT3C and DNMT3B awaits further investigation.

It is worth noting that the RD helix of DNMT3C also shares a similar DNA interaction mode with that of DNMT3B, both involving an arginine (R710 in DNMT3C and R823 in DNMT3B) for an interaction with the DNA backbone (Fig. 4, B and C). In contrast, the corresponding region of DNMT3A engages in more extensive interactions, involving residues S881, R882, L883, and R887, for the DNA contact (Fig. 4A). The functional consequence of the amino acid variation on the RD helix between DNMT3s remains to be determined.

The effect of rahu mutation on DNMT3C activity

A previous study has demonstrated that DNMT3C E693G, the so-called *rahu* mutation, is responsible for the male sterility phenotype in mice (7). Close inspection of the DNMT3C-DNMT3L-DNA complex reveals that residue E693 interacts with the RD-interface residues, such as H701 and R713, *via* hydrogen-bonding interactions. Conceivably, the hydrogen-bonding interaction between DNMT3C E693 and H701 may help stabilize the homodimerization of DNMT3C (Fig. 5A). In addition, the hydrogen-bonding interaction between residues E693 and R713 may stabilize the DNA-contacting helix at the RD interface (where residue R710 is located) (Fig. 5A), thereby contributing to the DNA-binding activity of DNMT3C. In support of these notions, our size-exclusion chromatography analysis reveals that the *rahu* mutation shifts the elution of DNMT3C-DNMT3L from a volume that corresponds to the heterotetrameric assembly to a dimeric assembly (Fig. 5B), suggesting disruption of the tetrameric assembly of DNMT3C-DNMT3L by the *rahu* mutation. Next, we performed electrophoretic mobility shift assay (EMSA) for DNMT3C, WT and *rahu* mutant. Increasing concentration of WT DNMT3C gradually decreased the intensity of the free DNA band, indicative of the complex formation (Figs. 5C and S7). In contrast, titration of DNA with the *rahu* mutant of DNMT3C did not lead to an appreciable shift of the DNA band (Figs. 5C and S7). These data therefore support that the *rahu* mutation led to disruption of homodimerization as well as reduced DNA-binding activity of DNMT3C. Finally, our *in vitro* DNA methylation analysis reveals that the *rahu* mutation led to largely abolished DNA methylation activity of DNMT3C (Fig. 5D), consistent with a previous observation that the *rahu* mutation caused DNA hypomethylation in cells (7). Together, these data establish that the *rahu* mutation compromised DNMT3C-mediated DNA methylation *via* impairing its oligomerization and DNA-binding activity.

Discussion

DNA methylation plays a critical role in transposon silencing and chromatin compartmentalization during development. Dysregulation of DNA methylation leads to various developmental defects and diseases. The recent identification of DNMT3C, which specifically methylates promoters of young transposons in mouse sperm (6, 7), provides a mechanism for silencing the evolutionarily young transposons (*e.g.* EVRKs) that are enriched in the mouse genome (3), thereby shedding light on how mammalian DNA methyltransferases adapt to the genome complexities of rodents through functional partitioning. DNMT3A, DNMT3B and DNMT3C coexist in mouse male germ cells (6, 28). Compared with closely related DNMT3B, the lack of an N-terminal histone reader PWWP domain in DNMT3C may contribute to its distinct chromatin targeting activity. On the other hand, how the DNA methylation activity of DNMT3C is regulated to maintain a proper balance of DNA methylation across the genome remains unclear. This study, through combined structural and biochemical analysis, provides a mechanistic

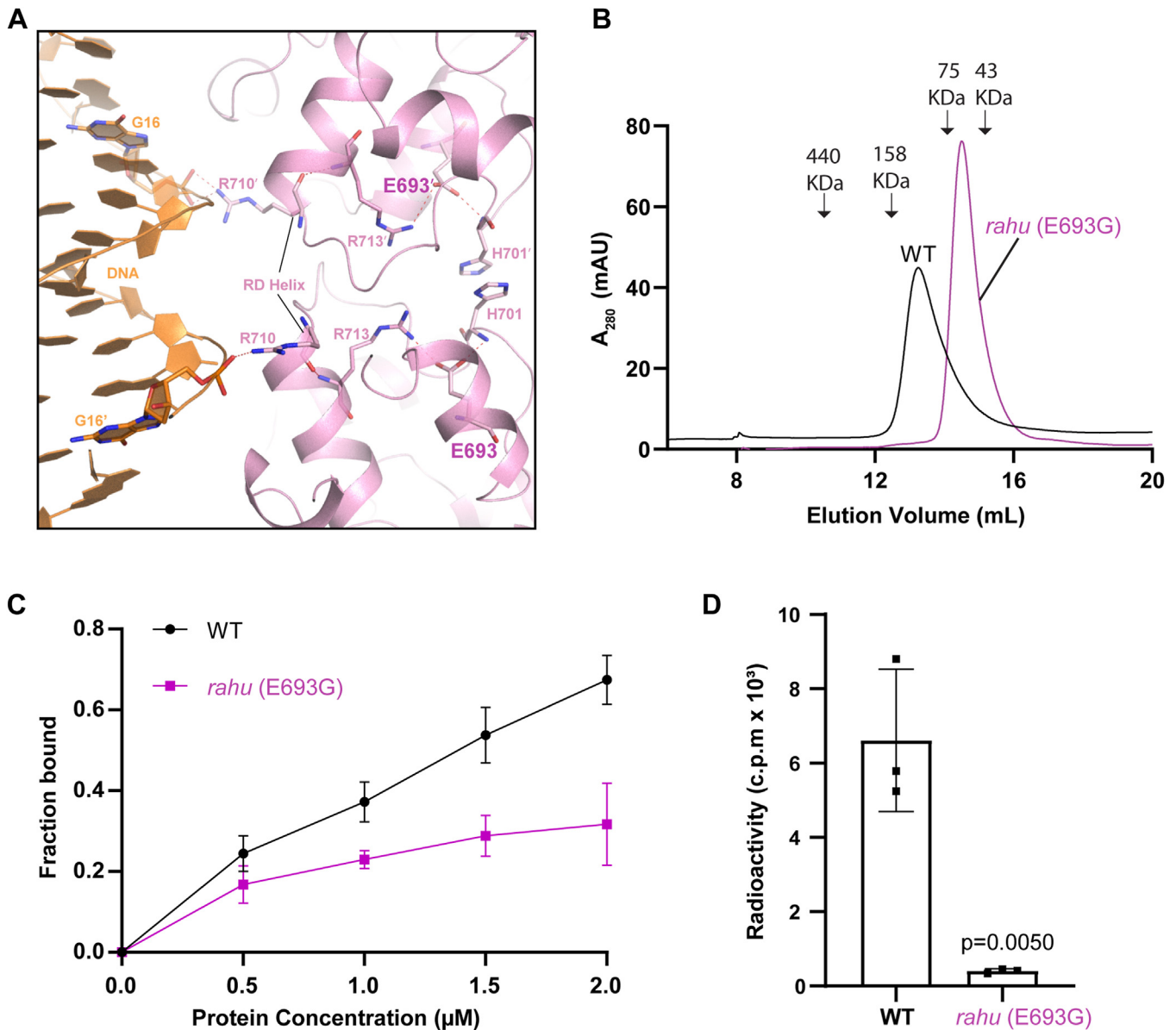


Figure 5. Structural and biochemical analyses of DNMT3C *rahu* mutant. *A*, close-up view of the DNMT3C E693-engaged interactions at the RD interface of DNMT3C-DNA complex. Hydrogen-bonding interactions are shown as dashed lines. *B*, size-exclusion chromatography analysis of WT and *rahu* mutant DNMT3C. The elution volumes for the select protein markers are indicated by arrows. Note that the molecular weight for tetrameric and dimeric DNMT3C-DNMT3L are ~ 134 kD and ~ 67 kD, respectively. *C*, the ratio of DNA-bound WT or mutant DNMT3C-DNMT3L, determined by the EMSA analysis, as a function of protein concentration. Data are mean \pm s.d. ($n = 3$ biological replicates). *D*, *in vitro* DNA methylation analysis of DNMT3C, WT or *rahu* mutant, on CpG DNA. Data are mean \pm s.d. ($n = 3$ biological replicates). Statistical analysis for WT DNMT3C vs *rahu* used two-tailed Student's *t* test.

insight into DNMT3C-mediated *de novo* DNA methylation and the functional impact of its disease mutation.

First, this study reveals that DNMT3C interacts with substrate DNA in a manner similar to but distinct from that of DNMT3A and DNMT3B. On the one hand, the interaction between the TRD loop of DNMT3C and DNA resembles the DNMT3B-DNA interaction, both recognizing the CpG guanine *via* an asparagine (DNMT3C N666 or DNMT3B N779) and the CpG-flanking sites *via* a lysine (DNMT3C K664 or DNMT3B K777). On the other hand, substantial diversification among the three *de novo* DNA methyltransferases was observed for the interaction between their Rossmann-fold catalytic core and DNA. Of note, DNMT3C contains a

glutamate at site 590, rather than a lysine (K703) in DNMT3B or a glycine (G762) in DNMT3A. The presence of E590 results in electrostatic repulsion between DNMT3C and DNA substrate, thereby restricting the DNA methylation activity of DNMT3C that may help avoid genome-wide DNA hypermethylation.

Evolutionary covariation of two functional sites is recurrently observed for enzymes, which leads to fine-tuning of the enzymatic activity, thereby adapting them to distinct environments encountered by various species (29, 30). This study shows that the replacement of DNMT3C C543 by the corresponding residues in DNMT3A and DNMT3B led to a significant decrease in the DNA methylation activity of

Structure of DNMT3C-DNA complex

DNMT3C, suggesting a unique role for this residue in maintaining a proper DNA methylation activity of DNMT3C. Furthermore, this study reveals distinct intramolecular interactions involving the DNMT3C C543-, R548- and E590-corresponding sites: the DNMT3C R548- and E590-corresponding sites form a hydrogen bond in DNMT3C but not in DNMT3A and DNMT3B; the DNMT3C C543- and R548-corresponding sites form a hydrogen bond in DNMT3B but not in DNMT3A and DNMT3C (Fig. 3D). Consistently, our DNA methylation analyses revealed that introducing the DNMT3A-converting C543I/E590G and DNMT3B-converting C543N/E590K double mutations, rather than the C543I and C543N single mutations, to DNMT3C substantially shift the DNA methylation activity and specificity of DNMT3C to those of DNMT3A and DNMT3B, respectively. These observations therefore establish that the DNMT3C C543- and E590-corresponding sites coordinately modulate the activity of DNMT3A, DNMT3B and DNMT3C. Intriguingly, our sequence analysis of the DNMT3 family of *de novo* DNA methyltransferases revealed that DNMT3C C543- and E590-corresponding sites are evolutionarily conserved within each DNMT3A, DNMT3B or DNMT3C subfamily but vary between different DNMT3 subfamilies. Together, these findings support a notion that evolutionary covariation of the DNMT3C C543- and E590-corresponding sites may serve to fine-tune the DNA methylation activity and specificity of DNMT3s, opening a new avenue for designing a novel DNA methyltransferase with unique activity and specificity. How the amino acid variations between the DNMT3 family of DNA methyltransferases contribute to their distinct functionalities *in vivo* awaits further investigation.

As with DNMT3B (16), DNMT3C shows a marked methylation activity toward CpA DNA. CpA methylation is the most abundant non-CpG DNA methylation in mammals, occurring in gene bodies, enhancers, promoters, as well as transposons of germ cells, embryonic stem cells, and neural cells (31). An increasing number of studies have linked non-CpG methylation to gene regulation (31–33). In this context, DNMT3C-mediated non-CpG methylation may provide another mechanism to defend against evolutionarily young transposons. It is also worth mentioning that the ADD domains of DNMT3A and DNMT3B were shown to autoinhibit the activity of DNMT3A/DNMT3B, which can be relieved *via* the interaction between the ADD domain and unmodified histone H3 tail (20, 34, 35). It is conceivable that DNMT3C may adopt a similar regulatory mechanism. However, since the ADD-mediated regulation does not involve a sequence-specific DNA interaction, it may not impact the substrate specificity of DNMT3C substantially at the chromatin level.

Finally, this study provides a molecular basis for the disease-causing *rahu* mutation of DNMT3C, which causes DNA hypomethylation in IAP and L1 transposons, particularly at the 5'-end. A *rahu*-corresponding mutation has not been reported for DNMT3A and DNMT3B. Nevertheless, a three-amino-acid insertion immediately downstream of the DNMT3C E693-corresponding site in DNMT3B has been shown to be

associated with the ICF syndrome and perturb protein localization in cells (14, 36). Our structural analysis reveals that DNMT3C E693 is located near to the RD interface, involving interactions stabilizing the oligomerization and the DNA-contact site. Introducing the *rahu* mutation led to impaired oligomerization and reduced DNA-binding activity of the DNMT3C-DNMT3L complex, which consequently abolished the DNA methylation activity of the DNMT3C-DNMT3L complex. Together, this study provides a mechanistic explanation of the functional consequence of the *rahu* mutation, as well as its related mutation in human ICF syndrome.

Experimental procedures

Protein expression and purification

A synthetic DNA fragment encoding mouse DNMT3C MTase domain (residues 456–740) and the cDNA encoding the C-terminal domain (residues 178–386) of human DNMT3L (hDNMT3L, NCBI accession NM_175867) or the C-terminal domain (residues 220–415) of mouse DNMT3L (mDNMT3L, NCBI accession NM_1081695) were inserted in tandem in a modified pRSFDuet-1 vector (Novagen Inc), in which the DNA sequence for DNMT3C was preceded by an N-terminal His₆-SUMO tag and ULP1 (ubiquitin-like protease) cleavage site and the hDNMT3L or mDNMT3L gene was flanked by the NdeI and XhoI cleavage sites. The expression plasmid was transformed in BL21 (DE3) RIL cell strain (Agilent Technologies). The transformed cells were cultured in LB media at 37 °C until the optical density at 600 nm (OD₆₀₀) reached ~1.0. The temperature was then lowered to 16 °C, and the cells were induced by 80 μM IPTG (isopropyl β-D-galactoside). The cells continued to grow overnight. Subsequently, the cells were harvested and lysed in a buffer (50 mM Tris-HCl, pH8.0, 1 M NaCl, 10% glycerol, 18 mM Imidazole, 10 μg/ml DNase I, and 1 mM PMSF). The His₆-SUMO-DNMT3C fusion protein and hDNMT3L/mDNMT3L were co-purified using a Ni-NTA affinity column (GE Healthcare), followed by ion-exchange chromatography on a Heparin HP column (GE Healthcare). After treatment with ULP1 for 2 to 3 h, the His₆-SUMO tag was removed *via* Ni-NTA affinity chromatography. Finally, the DNMT3C-hDNMT3L or DNMT3C-mDNMT3L complex was purified *via* size-exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) in a buffer containing 25 mM Tris-HCl (pH 7.5), 5% glycerol, 250 mM NaCl, and 5 mM DTT.

For crystallization of the DNMT3C-DNMT3L-DNA complex, the DNA duplex was prepared by annealing of a zebrularine-containing palindromic DNA fragment (5'-CATGZGGTCTTAATTAGACCGCATGG-3'; Z: zebrularine). Subsequently, the DNA was mixed with the DNMT3C-hDNMT3L tetramer in a 1:2 molar ratio in a buffer containing 50 mM Tris-HCl (pH 8.0), 35% glycerol, and 40 mM DTT at 25 °C. The covalent complex of DNMT3C-hDNMT3L-DNA was first purified using a HiTrap Q HP column (GE Healthcare) pre-equilibrated with 20 mM Tris-HCl, 5%

glycerol, 5 mM DTT, and 50 mM NaCl and run in a salt gradient from 50 mM to 1 M NaCl. The protein-DNA complex was further purified on a HiLoad 16/600 Superdex 200 pg column in a buffer containing 20 mM Tris-HCl (pH 8.0), 5% glycerol, 250 mM NaCl and 5 mM DTT. Protein purity was confirmed using SDS-PAGE and the sample was stored at -80°C before use.

X-ray crystallography and structure determination

The DNMT3C-mDNMT3L complex shows a reduced solubility than the DNMT3C-hDNMT3L complex. We, therefore, focused on the DNMT3C-hDNMT3L-DNA complex for crystallographic study. The crystallization condition for ~ 0.1 mM DNMT3C-hDNMT3L-DNA complex mixed with *S*-adenosyl-homocysteine (SAH) in 1:5 molar ratio was first identified using sparse-matrix screening (Hampton Research Inc). Subsequently, crystals were reproduced and optimized by the hanging drop vapor diffusion method at 4°C . For optimal crystallization, 1 μl of DNMT3C-hDNMT3L-DNA sample was mixed with 1 μl of the precipitant solution containing 2% v/v Tacsimate TM (pH 8.0), 0.1 M Tris-HCl (pH 8.5), and 16% w/v Polyethylene glycol 3350. To harvest the crystals, the crystals were soaked in mother liquor with an additional 25% ethylene glycol, followed by flash frozen in liquid nitrogen.

The X-ray diffraction data were collected on the Beamline 8.2.2 at the Advanced Light Source, Lawrence Berkeley National Laboratory. The diffraction data were indexed, integrated, and scaled using the HKL3000 program (37), followed by molecular replacement using the PHASER program (38) embedded in the PHENIX software (39), using the structure of human DNMT3B-DNMT3L-DNA complex (PDB 6KDA) as a search model. The structure was refined using COOT (40) and PHENIX software iteratively. The same R-free test set was used throughout the refinement. Structural statistics is listed in Table S1. The structural figures were generated using the Pymol software (<https://www.pymol.org/pymol>).

In vitro DNA methylation assay

In vitro DNA methylation assays were performed following a previously reported protocol (16). In essence, DNA duplexes (GAC)₁₂, (AAC)₁₂, and (TAC)₁₂ were used for CpG, CpA, and CpT DNA substrates, respectively. For the assays under a single substrate concentration, the experiments were carried out in 20 μl -reaction solution containing 0.75 μM DNA, 0.6 μM DNMT3C-DNMT3L, DNMT3B-DNMT3L or DNMT3A-DNMT3L tetramer, 2.5 μM *S*-adenosyl-L-[methyl-³H]methionine (specific activity 82.3 Ci/mmol, PerkinElmer) in 50 mM Tris-HCl (pH 7.5), 0.05% β -mercaptoethanol, 5% glycerol and 200 $\mu\text{g}/\text{ml}$ BSA. The reactions were carried out in triplicate at 37°C for 10 min unless indicated otherwise, before being quenched by 0.6 μl of 32 mM nonradioactive *S*-adenosyl-methionine (SAM). For the steady-state kinetics assay, the reactions were conducted in a 20- μl solution containing 0.3 μM DNMT3C-DNMT3L C-terminal domains, WT or mutant, various concentrations of CpG DNA (0, 0.0125, 0.025, 0.05, 0.1, 0.2, or 0.35 μM), and 2.5 μM *S*-adenosyl-L-[methyl-³H]

methionine (specific activity 82.3 Ci/mmol, PerkinElmer) in 50 mM Tris-HCl (pH 7.5), 0.05% β -mercaptoethanol, 5% glycerol and 200 $\mu\text{g}/\text{ml}$ BSA. For measurement of the methylation rate, the reactions were carried out for approximately 0, 6, and 12 min before being quenched by 0.6 μl of 32 mM nonradioactive SAM. The reactions were carried out in triplicate.

Subsequently, 8 μl of each reaction mixture was spot on the Amersham Hybond-XL membrane (GE Healthcare). The membrane was washed sequentially with cold 0.2 M ammonium bicarbonate (pH 8.2) twice, deionized water once, and ethanol once, 15 min for each step. The membrane was air-dried before being transferred to scintillation vials each containing 3 ml of ScintiVerse (Fisher). The ³H radioactivity was measured using a Beckman LS6500 counter.

Electrophoretic mobility shift assay (EMSA)

For the DNMT3C-DNA binding assay, a 12- μL sample contained 1.0 μM (GAC)₁₂ DNA duplex mixed with various concentrations (0, 0.5, 1.0, 1.5, and 2 μM) of WT or mutant DNMT3C MTase-mDNMT3L C-terminal domain. The binding mixture was incubated for 30 min on ice in a buffer containing 50 mM Tris-HCl (pH 8.5), 36% glycerol, 8% glucose, 2 mM DTT, 0.1 mg/ml BSA, 100 mM NaCl, and 0.08% Tween 20. Gel electrophoresis was run on an 8% polyacrylamide gel at 110 V using 0.5 \times TBE (pH 8.3) buffer at 4°C for 1 h. The gel was stained using SYBR gold stain (Thermo Fisher) for visualization. The band intensity of free DNA in each sample was quantified using ImageJ software (41) and subsequently used for estimation of the ratio of DNMT3C-bound DNA.

Size-exclusion chromatography

Size-exclusion chromatography analysis was performed on a Superdex 200 increase 10/300 gl column (GE Healthcare). WT and *rahu* DNMT3C protein samples (500 μl of 0.4 mg/ml protein) were injected separately into the column and eluted at a flow rate of 0.5 ml/min in a buffer containing 25 mM Tris-HCl (pH 7.5), 5% Glycerol, 250 mM NaCl, and 5 mM DTT buffer. The fractions were collected in 500 μl each. The protein identity was confirmed using SDS-PAGE.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Coordinates and structure factors for the DNMT3C-DNMT3L-CGG complex have been deposited in the Protein Data Bank under accession code 8TCI.

Supporting information—This article contains supporting information.

Acknowledgments—We thank staff members at the Advanced Light Source (DE-AC02-05CH11231), Lawrence Berkeley National Laboratory for access to X-ray beamlines. This work was supported by

Structure of DNMT3C-DNA complex

NIH grant (R35GM119721) to J. S. G. H. was supported by an NRSA T32 training grant (T32 ES018827).

Author contributions—N. K. writing—original draft; N. K., J. L., J. S., X. C., and G. H. investigation; N. K. and J. S. formal analysis; N. K. and J. S. data curation; J. S. writing—review & editing, J. S. funding acquisition; J. S. conceptualization.

Conflict of interests—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviation—The abbreviations used are: DNMT, DNA methyltransferases.

References

- Goodier, J. L. (2016) Restricting retrotransposons: a review. *Mob DNA* **7**, 16
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921
- Mouse Genome Sequencing, C., Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., et al. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562
- Cordaux, R., and Batzer, M. A. (2009) The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **10**, 691–703
- Savage, A. L., Schumann, G. G., Breen, G., Bubb, V. J., Al-Chalabi, A., and Quinn, J. P. (2019) Retrotransposons in the development and progression of amyotrophic lateral sclerosis. *J. Neurol. Neurosurg. Psychiatry* **90**, 284–293
- Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Herault, Y., et al. (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354**, 909–912
- Jain, D., Meydan, C., Lange, J., Claeys Bouuaert, C., Lailler, N., Mason, C. E., et al. (2017) Rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. *PLoS Genet.* **13**, e1006964
- Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., and Bestor, T. H. (2001) Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536–2539
- Chedin, F., Lieber, M. R., and Hsieh, C. L. (2002) The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16916–16921
- Cheng, X., and Blumenthal, R. M. (2008) Mammalian DNA methyltransferases: a structural perspective. *Structure* **16**, 341–350
- Goll, M. G., and Bestor, T. H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514
- Jeltsch, A., and Jurkowska, R. Z. (2016) Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. *Nucleic Acids Res.* **44**, 8556–8575
- Law, J. A., and Jacobsen, S. E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220
- Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257
- Ren, W., Gao, L., and Song, J. (2018) Structural basis of DNMT1 and DNMT3A-mediated DNA methylation. *Genes* **9**, 620
- Gao, L., Emperle, M., Guo, Y., Grimm, S. A., Ren, W., Adam, S., et al. (2020) Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. *Nat. Commun.* **11**, 3355
- Molaro, A., Malik, H. S., and Bourc'his, D. (2020) Dynamic evolution of de novo DNA methyltransferases in rodent and primate genomes. *Mol. Biol. Evol.* **37**, 1882–1892
- Okano, M., Xie, S., and Li, E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* **19**, 219–220
- Anteneh, H., Fang, J., and Song, J. (2020) Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. *Nat. Commun.* **11**, 2294
- Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., et al. (2015) Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* **517**, 640–644
- Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449**, 248–251
- Lin, C. C., Chen, Y. P., Yang, W. Z., Shen, J. C. K., and Yuan, H. S. (2020) Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. *Nucleic Acids Res.* **48**, 3949–3961
- Zhang, Z. M., Lu, R., Wang, P., Yu, Y., Chen, D., Gao, L., et al. (2018) Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature* **554**, 387–391
- Gao, L., Anteneh, H., and Song, J. (2020) Dissect the DNMT3A- and DNMT3B-mediated DNA Co-methylation through a covalent complex approach. *J. Mol. Biol.* **432**, 569–575
- Zhou, L., Cheng, X., Connolly, B. A., Dickman, M. J., Hurd, P. J., and Hornby, D. P. (2002) Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J. Mol. Biol.* **321**, 591–599
- Lin, I. G., Han, L., Taghva, A., O'Brien, L. E., and Hsieh, C. L. (2002) Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. *Mol. Cell Biol.* **22**, 704–723
- Wienholz, B. L., Karet, M. S., Moarefi, A. H., Gordon, C. A., Ginno, P. A., and Chedin, F. (2010) DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. *PLoS Genet.* **6**, e1001106
- Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., et al. (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900–903
- Laskowski, R. A., Gerick, F., and Thornton, J. M. (2009) The structural basis of allosteric regulation in proteins. *FEBS Lett.* **583**, 1692–1698
- Yang, J. S., Seo, S. W., Jang, S., Jung, G. Y., and Kim, S. (2012) Rational engineering of enzyme allosteric regulation through sequence evolution analysis. *PLoS Comput. Biol.* **8**, e1002612
- He, Y., and Ecker, J. R. (2015) Non-CG methylation in the human genome. *Annu. Rev. Genomics Hum. Genet.* **16**, 55–77
- Jang, H. S., Shin, W. J., Lee, J. E., and Do, J. T. (2017) CpG and non-CpG methylation in epigenetic gene regulation and brain function. *Genes* **8**, 148
- Patil, V., Ward, R. L., and Hesson, L. B. (2014) The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics* **9**, 823–828
- Hu, J. L., Zhou, B. O., Zhang, R. R., Zhang, K. L., Zhou, J. Q., and Xu, G. L. (2009) The N-terminus of histone H3 is required for de novo DNA methylation in chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 22187–22192
- Lu, J., Fang, J., Zhu, H., Liang, K. L., Khudaverdyan, N., and Song, J. (2023) Structural basis for the allosteric regulation and dynamic assembly of DNMT3B. *Nucleic Acids Res.* **51**, 12476–12491
- Ueda, Y., Okano, M., Williams, C., Chen, T., Georgopoulos, K., and Li, E. (2006) Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. *Development* **133**, 1183–1192
- Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
- Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675