

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Structure and Regulation of Eukaryotic 5mC and 6mA DNA Methyltransferases

Permalink

<https://escholarship.org/uc/item/2tk61057>

ISBN

9798186178750

Author

Shao, Zengyu

Publication Date

2026-05-27

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Structure and Regulation of Eukaryotic 5mC and 6mA DNA Methyltransferases

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

Doctor of Philosophy

in

Biochemistry and Molecular Biology

by

Zengyu Shao

June 2026

Dissertation Committee:

Dr. Jikui Song, Chairperson

Dr. Charles Hille

Dr. Sean O'Leary

The Dissertation of Zengyu Shao is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to my advisor, Dr. Jikui Song, for his continuous guidance, support, and mentorship throughout my PhD training. His scientific insight, intellectual rigor, and high standards have profoundly shaped my development as a researcher. I am especially grateful to him for emphasizing the importance of strong critical thinking, persistent self-motivation, and independent working capability—qualities that are essential for an academic career. I will continue to keep these principles in mind as I move forward in my professional development.

I am also deeply grateful to my dissertation committee members, Dr. Charles Hille and Dr. Sean O’Leary, as well as Dr. Gregor Blaha, one of my guidance committee members, and Dr. Howard S. Judelson, who served on my oral qualifying examination committee. Thank you all very much for your time, valuable feedback, and encouragement throughout my Ph.D. journey.

I would also like to thank all Song Lab members, especially Jiuwei Lu and Nelli Khudaverdyan for their help with the DIM2 project, and Sol Yoon and Pranav Athavale for purifying mutants for the AMT project. Thanks to Jian Fang for his help in my life and early Ph.D. training.

I also sincerely appreciate the support from the Program of Biochemistry and Molecular Biology at the University of California, Riverside. I am grateful for the prizes, awards, training opportunities, seminars, and academic environment that have enriched my

graduate education. I would also like to acknowledge the BMPN Administration Assistants for their help.

Finally, I would like to express my deepest appreciation to my family for their unconditional love, encouragement, and sacrifice. Their constant support has been the foundation of my perseverance.

This work would not have been possible without all of you.

ABSTRACT OF THE DISSERTATION

Structure and Regulation of Eukaryotic 5mC and 6mA DNA Methyltransferases

by

Zengyu Shao

Doctor of Philosophy, Graduate Program in Biochemistry and Molecular Biology
University of California, Riverside, June 2026
Dr. Jikui Song, Chairperson

DNA methylation is a fundamental epigenetic mechanism that shapes genome function, chromatin organization, and cellular differentiation. Although the regulation of 5-methylcytosine (5mC) in higher eukaryotes has been extensively studied, key questions remain regarding how distinct DNA methyltransferases recognize DNA substrates under different chromatin contexts. This dissertation investigates the structure and regulation of representative eukaryotic DNA methyltransferases that establish either 5mC or N6-methyladenine (6mA). First, biochemical analysis of the *Neurospora crassa* cytosine methyltransferase Defective-In-Methylation-2 (DIM2) shows that its activation is tightly controlled by heterochromatin factors. The highly activated DIM2 requires the cooperativity between HP1 and H3K9me3. Three cryo-EM structures reveal how these multivalent inputs enable region-specific DNA methylation. Second, structural characterization of the *Tetrahymena thermophila* N6-adenine methyltransferase AMT complex captured in an inhibitor-trapped state reveals how a eukaryotic 6mA writer

engages a DNA-mimicking substrate. The structure reveals that the AMT1, AMT7, AMTP1, and AMTP2 form a saddle shape in which the complex undergoes an open-to-closed transition during productive DNA engagement. Together, these studies provide structural and mechanistic insights into how eukaryotic DNA methyltransferases orchestrate protein cofactors in different chromatin contexts to enable DNA recognition for distinct methylation pathways. Together, this work expands the structural framework for understanding the regulation of both 5mC and 6mA methyltransferases across eukaryotes.

TABLE OF CONTENTS

LIST OF FIGURES	X
-----------------------	---

LIST OF TABLES	XII
----------------------	-----

CHAPTER 1	1
-----------------	---

1.1 DNA Methylation as a Conserved yet Diversified Epigenetic System	1
1.2 Two Major Eukaryotic DNA Methylation Marks: 5mC and 6mA.....	2
1.2.1 5mC Methylation	2
1.2.2 6mA Methylation	3
1.3 DNA Methyltransferases as Writers of Epigenetic Information	5
1.3.1 Functional Diversification of Eukaryotic Cytosine Methyltransferases	5
1.3.2 Emerging Enzymatic Systems for Eukaryotic 6mA Methylation.....	8
1.3.3 Shared Principles of Substrate Recognition and Catalytic Activation	10
1.4 Outstanding Mechanistic Questions and Dissertation Objectives	11
1.5 Organization of the Dissertation	12
References.....	17

CHAPTER 2	23
-----------------	----

Abstract.....	23
Introduction.....	24
Materials and Methods	28
Results.....	36
Discussion.....	54
Supplementary Data.....	73
References.....	102

CHAPTER 3.....	107
Abstract.....	107
Introduction.....	108
Materials and Methods	110
Results.....	115
Discussion.....	123
Supplementary Data.....	132
References.....	146
CHAPTER 4.....	149
General Conclusion	149
References.....	152

List of Figures

Chapter 1

Figure 1. Molecular mechanism of DNA methylation in eukaryotes	14
Figure 2. DNA methyltransferases in eukaryotes	15
Figure 3. The reaction pathway of C5 MTases in the presence and in the absence of mechanism-based inhibitors.....	16

Chapter 2

Figure 1. Cryo-EM structure of the DIM2-HP1 complex.....	59
Figure 2. Structural basis for the DIM2-HP1 interaction	61
Figure 3. Cryo-EM structure of apo-DIM2.....	63
Figure 4. Cryo-EM structure of the DIM2-HP1-H3K9me3-DNA complex.....	65
Figure 5. Structural basis of the DIM2-DNA interaction	67
Figure 6. Structural basis for H3K9me3-mediated DIM2 activation	69
Figure 7. Structural comparison of DIM2 and DNMT1	70
Figure 8. Multi-layered regulation of DIM2-mediated DNA methylation	72

Chapter 3

Figure 1. Biochemical analysis of the AMT-OCR complex.....	125
Figure 2. Cryo-EM structure of the OCR-bound AMT complex	126
Figure 3. Structural and mutational analysis of the potential DNA-binding surface of the AMT complex.....	128
Figure 4. Structural comparison of the AMT-DNA model with the enzyme-substrate complexes of bacterial 6mA DNA methyltransferases.....	130

Figure 5. A working model of the AMT complex	131
--	-----

List of Tables

Chapter 2

Table 1. Summary of ITC binding parameters 100

Table 2. Cryo-EM data collection, refinement, and validation statistics 101

Chapter 3

Table 1. Cryo-EM data collection, refinement and validation statistics 145

Chapter 1

Introduction into DNA methylations and DNA methyltransferases

1.1 DNA Methylation as a Conserved yet Diversified Epigenetic System

DNA methylation is a fundamental epigenetic modification that contributes to gene expression, genome stability, and cellular differentiation across diverse organisms¹. It involves the covalent addition of a methyl group to specific nucleotides in DNA, thereby influencing its interaction network with regulatory proteins and altering genome function. By affecting the epigenetic landscapes, DNA methylation plays critical roles in transcription regulation, chromatin integrity, and epigenetic inheritance.

Although DNA methylation is largely conserved, its regulation pattern on gene expression, genome stability, and biological functions varies significantly across organisms. In eukaryotes, 5-methylcytosine (5mC) remains the best-studied methylation mark and is well-established as a major driver of gene silencing and chromatin compaction. In contrast, N6-methyladenine (6mA), a well-known epigenetic mark in prokaryotes that involves in DNA replication, DNA repair, gene regulation, and crucially host defense, has emerged more recently as a DNA modification present in unicellular eukaryotic organisms with much lower abundance. This diversity suggests that DNA methylation is broadly present in various organism, but the function and regulation is subject to species-specific genome and chromatin context.

Because this dissertation focuses on the structural basis of eukaryotic DNA methylation systems, the following content will focus on the two most relevant DNA methylation marks, 5mC and 6mA, as well as the DNA methyltransferases that establish them.

1.2 Two Major Eukaryotic DNA Methylation Marks: 5mC and 6mA

1.2.1 5mC Methylation

5mC is among the most extensively recognized DNA modifications and is ubiquitously distributed across eukaryotes, although its genomic abundance varies significantly between different species. In mammalian somatic tissues, 5mC DNA methylation occurs predominantly at CpG dinucleotides, with approximately 70–80% of CpG sites methylated across the genome^{2,3}. Plants have lower levels of 5mC methylation, and the abundance is dependent on species. In *Arabidopsis thaliana*, for example, approximately 24% of CpG, 6.7% of CHG, and 1.7% of CHH cytosines (H = A, C, or T) are methylated⁴. In other lineages, DNA methylation is minimal or nearly absent. For instance, 5mC is essentially undetectable in *Dictyostelium* and *Caenorhabditis elegans*^{5,6,7}. In fungi, the DNA methylation level is diverse. Though 5mC is absent from species such as *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*, it is easily detectable in *Neurospora crassa*, where approximately 1.5% of total cytosines are methylated^{8,9,10,11,12}.

Despite 5mC shows various genomic abundance, it has emerged as a crucial epigenetic mark with broad regulatory significance. In mammals, DNA methylation is predominantly associated with genome stability, and plays fundamental roles in X-

chromosome inactivation, genomic imprinting, and the stable silencing of repetitive elements¹³. In plants, cytosine methylation is found in CG, CHG, and CHH sequence contexts¹⁴, reflecting a more complex methylation landscape than that observed in mammals. Plant CG methylation is enriched at transposable elements and repetitive regions^{15,4}. Functionally, CG methylation in plants mainly contributes to genome stability by silencing repetitive DNA sequences and transposon elements, gene-body regulation, plant development, and environmental stress response^{16–19}. Non-CG methylation further contributes to transcriptional regulation and genome stability in plants^{20,21}. In fungi such as *Neurospora crassa*, methylation appears to be closely related to repeat-induced point mutation (RIP), which is a homology-based process that converts G:C base pairs to A:T transitions in repetitive DNA sequences and subsequently promotes epigenetic silencing of duplicated regions through DNA methylation^{10,22}. Methylation in *Neurospora crassa* is not biased towards symmetric (e.g. CpG) sites²³, instead every cytosine in the affected region having a >80% chance of being methylated²⁴. Together, these observations indicate that the levels of 5mC methylation are various among species and the biological functions are even more dynamic and highly dependent on genomic contexts and have been adapted across evolution to meet the different demands of distinct organisms.

1.2.2 6mA Methylation

6mA methylation is a well-characterized DNA modification in prokaryotes and has also been detected in several eukaryotic organisms with various abundances. The 6mA was found in unicellular eukaryotes such as *Tetrahymena pyriformis* and *Paramecium aurelia*, as well as in the green alga *Chlamydomonas reinhardtii*^{25,26,27}. In these organisms, 6mA

accounts for approximately 0.5%–2.5% of total adenines^{25,26,27}, which is substantially lower than the abundance of 5mC in mammals and plants²⁸. Later, more and more multicellular eukaryotes have observed 6mA modifications even though with lower abundance, such as *Caenorhabditis elegans*, in which reported levels range from 0.01% to 0.4%, and *Drosophila melanogaster*, in which reported levels range from 0.001% to 0.07%^{29,30}. These observations suggest that the abundance of 6mA levels is relatively different in specific unicellular eukaryotes but generally occurs at much lower levels in multicellular organisms.

Despite its relatively low abundance in many eukaryotic systems, 6mA has emerged as a DNA modification with potentially broad regulatory biological functions. In bacteria, 6mA plays central roles in restriction-modification systems, DNA replication, DNA damage responses, and transcriptional regulation³¹. In eukaryotes, accumulating evidence has linked 6mA to gene transcription, chromatin organization, epigenetic inheritance, DNA repair, and cell-cycle regulation^{32,33,34,35,36}. In mammalian cells, mitochondrial DNA has been reported to contain substantially higher levels of 6mA than nuclear genomic DNA, suggesting that the distribution and function of this modification may be compartment-specific^{37,38}. Elevated 6mA at regions of stress-induced DNA duplex destabilization has also been proposed to modulate gene expression during early embryonic development³⁹. In addition, altered 6mA levels have been reported in multiple human cancers, including gastric, liver, triple-negative breast cancer, and glioblastoma^{40,41,42}. However, the exact function of 6mA in human is still debated. Together, these findings indicate that although 6mA is less prevalent than 5mC in most

eukaryotic genomes, it may nonetheless serve important regulatory functions in genome biology and human disease.

1.3 DNA Methyltransferases as Writers of Epigenetic Information

The installation of DNA methylation marks is catalyzed by DNA methyltransferases, which are enzymes that transfer a methyl group from donor *S*-adenosyl methionine (SAM) to specific DNA nucleotides. Although the catalysis mechanisms are similar between different DNA methylation marks, various DNA methyltransferases display distinct substrate recognition, regulatory modes, and biological functions. In eukaryotes, the most well-known cytosine methyltransferases are DNMT1, DNMT3A, and DNMT3B, which mediate 5mC methylation, whereas the adenine methyltransferase AMT complex has been responsible for 6mA methylation. Therefore, understanding DNA methylation system requires not only exploring the marks themselves, but also defining how these distinct DNA methyltransferases recognize, catalyze, and establish the DNA methylation patterns.

1.3.1 Functional Diversification of Eukaryotic Cytosine Methyltransferases

Cytosine methyltransferases are responsible for transferring a methyl group from donor SAM to the C5 position of cytosine, thereby establishing 5mC DNA methylation mark¹⁴. In general, cytosine methyltransferases are conserved, which share a catalytic core together with a target recognition domain (TRD) that contributes to specific DNA sequence recognition and catalysis⁴³. During methyl transfer, the catalytic core and target recognition domain form a two-lobed architecture that accommodates the DNA duplex

and enables formation of the covalent enzyme–DNA intermediate required for catalysis⁴⁴. Although the catalytic mechanism of cytosine methyltransferases is highly conserved, the enzymes responsible for 5mC deposition in eukaryotes have diversified substantially over evolution.

Cytosine methyltransferases are functionally divided into maintenance and *de novo* enzymes^{45,46}. *De novo* methyltransferases establish methylation patterns at previously unmethylated sites, whereas maintenance methyltransferases preferentially recognize hemimethylated DNA generated during DNA replication and restore the parental methylation pattern on the newly synthesized strand. (Fig. 1a). Based on structural difference and biological function, cytosine methyltransferases in eukaryotes can be divided into five major families: DNMT1, DNMT2, DNMT3/DRM2, Msc1/RID, and DNMT5^{11,47,48} (Fig. 2).

The DNMT1 family is most closely associated with maintenance methylation and includes mammalian DNMT1, plant MET1 and chromomethylases (CMTs), as well as several DNMT1-homologues in fungi, including Msc2 in *Ascombolus*, DIM2 in *Neurospora crassa*^{49,48}. Members of this family share conserved roles in propagating DNA methylation patterns and related domain architectures, such as a bromo-adjacent homology domain (BAH), which contributes to protein-protein interactions and nucleosome binding^{44,50}. For example, CMT3 in plants mediates CHG methylation by recognizing H3K9me2 through its BAH and chromo domains, illustrating how DNA methylation can be coupled to histone modification readout⁵¹ (Fig. 1c). In addition, mammalian DNMT1 requires additional regulatory region replication focus targeting

sequence (RFTS) for targeting to replication fork by recognizing and binding to histone H3 ubiquitinated at K18 and K23 by UHRF1^{52,53,54} (Fig. 1b). RFTS also functions as an autoinhibitory element in DNMT1 and MET1^{55,56}.

DNMT2 represents a second class of methylase-domain-containing proteins that is highly conserved across fungi, plants, and animals⁵⁷. In contrast to DNMT1-family enzymes, DNMT2 primarily functions as an RNA methyltransferase rather than a canonical DNA methyltransferase. It specifically methylates cytosine 38 in the anticodon loop of tRNA Asp, and this activity is conserved across diverse eukaryotic lineages^{57,58}. Although DNMT2 displays little or no conventional DNA methyltransferase activity, it retains the ancestral methyltransferase fold, suggesting functional divergence from a common catalytic framework.

The DNMT3 family functions predominantly in *de novo* DNA methylation. In animals, *de novo* DNA methyltransferases DNMT3A and DNMT3B establish DNA methylation patterns during embryogenesis and development (Fig. 1d), whereas in plants, the *de novo* DNMT3-homologous DRM proteins mediate asymmetric CHH methylation in the RNA-directed DNA methylation pathway involving 24-nucleotide small interfering RNAs (24nt-siRNAs)^{59,60}. DNMT3 homologs are absent from fungal genomes and have also been lost in certain green algae and animal lineages, indicating that alternative methylation systems have evolved in different branches of eukaryotic evolution^{61,62}. In fungi, additional methyltransferase families contribute to cytosine methylation. RID and Msc1 are fungal-specific *de novo* methyltransferases that primarily target repetitive sequences^{63,64}.

DNMT5 as a maintenance 5mC methyltransferase in *Cryptococcus neoformans*, harbors an SNF2 ATPase domain and requires ATP for DNA methylation, which is the only known ATP-dependent DNA 5mC methyltransferase that displays exquisite specificity for hemimethylated DNA⁶⁵. Loss of DNMT5 causes decreased *Cryptococcus neoformans* infectivity in mice, making it a potential target for pharmacological inhibition to combat infection^{66,67}.

Taken together, these observations indicate that eukaryotic cytosine methyltransferases comprise a structurally related but functionally diverse group of enzymes whose evolutionary specialization reflects the distinct regulatory needs of different genomes. This diversity is particularly relevant to the present dissertation, which examines DIM2 as a chromatin-regulated fungal cytosine methyltransferase.

1.3.2 Emerging Enzymatic Systems for Eukaryotic 6mA Methylation

Unlike the well-defined eukaryotic cytosine methyltransferases, the adenine methyltransferases in eukaryotes responsible for installing 6mA DNA methylation mark seem to be much more limited over evolution. It is challenging to identify adenine methyltransferases in higher eukaryotes due to their relative sparsity, even though there are some potential candidates of 6mA DNA methyltransferases^{29,68–71}. However, it has been clearly demonstrated in ciliates that exhibit nuclear dimorphism, such as *Tetrahymena thermophila* and related ciliates²¹. In these organisms, 6mA is found specifically in the transcriptionally active macronucleus and occurs dominantly at ApT dinucleotides, whereas it is absent from the transcriptionally silent micronucleus²¹. During sexual reproduction, when the zygotic nucleus derived from the parental

micronucleus differentiates into the new somatic macronucleus, a genome-wide wave of 6mA methylation is established²¹. This process is catalyzed by two *de novo* adenine methyltransferases 2 (AMT2) and AMT5, whose deletion abolishes 6mA at the majority of methylatable sites and markedly reduces methylation at the remaining sites, resulting in altered gene expression and reduced survival rate of sexual progenies^{72,73}.

In unicellular eukaryote *Tetrahymena thermophila*, maintenance of 6mA epigenetic mark is mediated primarily by adenine methyltransferase AMT complex (Fig. 1e), including a catalytic subunit AMT1, and regulatory subunits AMT6/AMT7, AMTP1, and AMTP2. It has been reported that maintenance methyltransferase AMT1-AMT7-AMTP1-AMTP2 complex (denoted as AMT complex herein) colocalizes with transcriptionally active mark H3K4me3, histone variant H2A.Z, as well as chromatin remodeling factor interactive bromodomain 2 (IBD2) for bulk of region-specific maintenance methylation, coupling DNA methylation and histone modification⁷⁴. In contrast, the AMT1-AMT6-AMTP1-AMTP2 complex is recognized and recruited by Proliferating cell nuclear antigen (PCNA) to initiate maintenance methylation right after DNA replication⁷⁴. These two complexes ensure faithful 6mA methylation at newly replicated loci.

Taken together, current evidence indicates that eukaryotic 6mA methyltransferase systems are best understood in unicellular eukaryotes, particularly ciliates, where *de novo* and maintenance activities can be assigned to distinct adenine methyltransferases. Outside these organisms, the enzymatic basis of 6mA deposition remains incompletely defined. This limited phylogenetic distribution provides the conceptual basis for the

second half of this dissertation, which focuses on the structural basis of substrate recognition by the eukaryotic 6mA methyltransferase AMT complex.

1.3.3 Shared Principles of Substrate Recognition and Catalytic Activation

Despite their evolutionary diversity in structure and functions, DNA methyltransferases share a common catalytic logic. These enzymes use SAM as the universal methyl donor and must gain access to reactive atoms that are normally buried within the DNA duplex. To achieve this, DNA methyltransferases employ a conserved base-flipping mechanism, in which the target base is rotated out of the DNA helix and inserted into the catalytic pocket of the enzyme⁷⁵. This conformational change allows DNA nucleobase to be precisely positioned for methyl transfer and has emerged as a conserved catalytic mechanism in DNA methylation systems.

Cytosine methyltransferases possess a highly conserved catalytic cysteine residue in the enzyme's relatively conserved Pro-Cys motif. It can perform a nucleophilic attack on the C6 position of the target cytosine ring, thereby activating the C5 carbon for methyl transfer from SAM^{75,76}. This mechanism has been elucidated extensively through the use of mechanism-based nucleoside analog inhibitors, including 5-azadeoxycytidine (AzaC)⁷⁷, 5-fluorodeoxycytidine (FdC)⁷⁸, 5,6-dihydro-5-azacytidine⁷⁹, 4'-thio-2'-deoxycytidine⁸⁰, as well as 2-H pyrimidinone-containing oligodeoxynucleotide (Zebularine)⁸¹ (Fig. 3), which have been used in plenty of studies for trapping catalytic intermediates and stabilizing the enzyme-DNA complex. Structural studies using such modified substrates provided direct evidence for the base-flipping mechanism and established the structural basis for target-base extrusion from the DNA duplex.

In contrast, 6mA DNA methyltransferases generally contain a conserved catalytic Asp-Pro-Pro-Trp (DPPW) motif⁸², which is also found in the catalytic domain of MT-A70 family RNA methyltransferases⁸³. The catalytic aspartate residue (D) within motif is responsible for coordinating methyl-donor SAM and facilitating a nucleophilic attack on the target adenine, thereby transferring a methyl group from SAM to the N6 position of adenine. The rest residues of the conserved motif are likely for stabilizing SAM and recognizing DNA substrates⁸⁴.

These biochemical and structural studies support a conserved catalytic mechanism in which SAM and substrate recognition, base flipping, activation of the target base, methyl transfer, and resolution of the covalent intermediate coordinately take place in the catalytic steps. In eukaryotes, these catalytic steps are frequently accompanied by more elaborate regulatory systems involving chromatin recognition, partner proteins, and higher-order conformational control. Defining how these additional layers of regulation are integrated with catalytic chemistry remains a major challenge in the field.

1.4 Outstanding Mechanistic Questions and Dissertation Objectives

Despite substantial progress in understanding DNA methylation across eukaryotes, important mechanistic questions remain regarding how structurally diverse DNA methyltransferases achieve substrate specificity and how distinct DNA methylation pathways are integrated with broader chromatin regulatory systems. Although 5mC is the best-characterized eukaryotic DNA modification, the mechanistic basis by which fungal 5mC methyltransferases are activated and regulated by heterochromatin cues remains not clear. Likewise, although 6mA has emerged as a potentially important eukaryotic DNA

epigenetic mark, the structural basis by which eukaryotic 6mA methyltransferase complexes engage their substrates remains not understood.

This dissertation addresses these questions through exploring the structure and regulation of two types of DNA methyltransferases mediated two distinct DNA methylation systems. The first part focuses on DIM2, a fungal 5mC methyltransferase, and seeks to define how chromatin-associated factors, including HP1 and H3K9me3, regulate substrate engagement and catalytic activation. The second part focuses on the ciliate AMT complex and seeks to define the structural basis by which a eukaryotic 6mA methyltransferase machinery integrates substrate recognition. By examining both a chromatin-regulated 5mC methyltransferase and a multi-subunit 6mA methyltransferase complex, this dissertation aims to identify both the specialized and the shared principles that govern eukaryotic DNA methylation.

More broadly, this work seeks to connect enzyme architecture with epigenetic function. By defining how eukaryotic DNA methyltransferases are regulated, how they recognize DNA or chromatin substrates, and how their activities are coordinated within distinct genomic contexts, this dissertation aims to deepen our understanding of the molecular logic underlying eukaryotic DNA methylation and to clarify how different methylation systems have diversified during evolution.

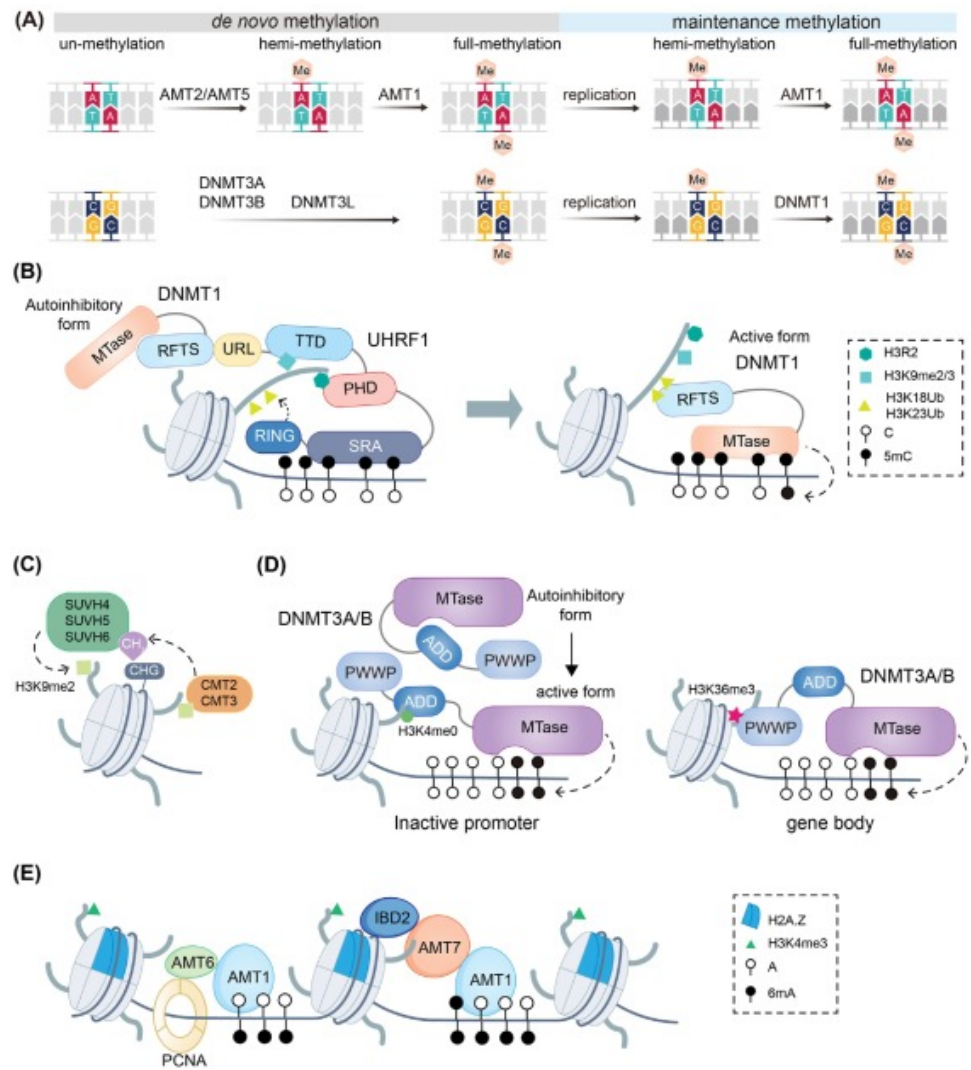
1.5 Organization of the Dissertation

This dissertation is organized into four major chapters, started by a general introduction in Chapter 1. Chapters 2 and 3 are reprints of the material from published papers by Shao et al. in *Nature Communications* (2024)⁸⁵ and *Protein Science* (2025)⁸⁶, respectively.

Chapter 2 focuses on the fungal 5mC methyltransferase DIM2 and examines how heterochromatin-associated factors regulate its activity. Through structural, biochemical, and functional analyses, this chapter defines how HP1 and H3K9me3 cooperate to promote productive substrate engagement and DNA methylation, thereby establishing a mechanistic framework for chromatin-coupled regulation of fungal 5mC deposition.

Chapter 3 focuses on the eukaryotic 6mA methyltransferase AMT complex from *Tetrahymena thermophila*. Using cryo-electron microscopy, mutational analysis, and biochemical characterization, this chapter investigates the structural basis of substrate recognition by the AMT complex and provides insight into how a multi-subunit eukaryotic 6mA methyltransferase machinery engages substrate. Together, Chapters 2 and 3 address complementary aspects of eukaryotic DNA methylation, with one chapter emphasizing chromatin-dependent regulation of a 5mC methyltransferase and the other emphasizing substrate recognition by a 6mA methyltransferase complex.

The dissertation concludes with a general conclusion in Chapter 4 that integrates the findings from both research chapters and discusses their broader implications for the regulation, substrate recognition, and evolutionary diversification of eukaryotic DNA methyltransferase systems.



Trends in Biochemical Sciences

Figure 1. Molecular mechanisms of DNA methylation in eukaryotes⁷².

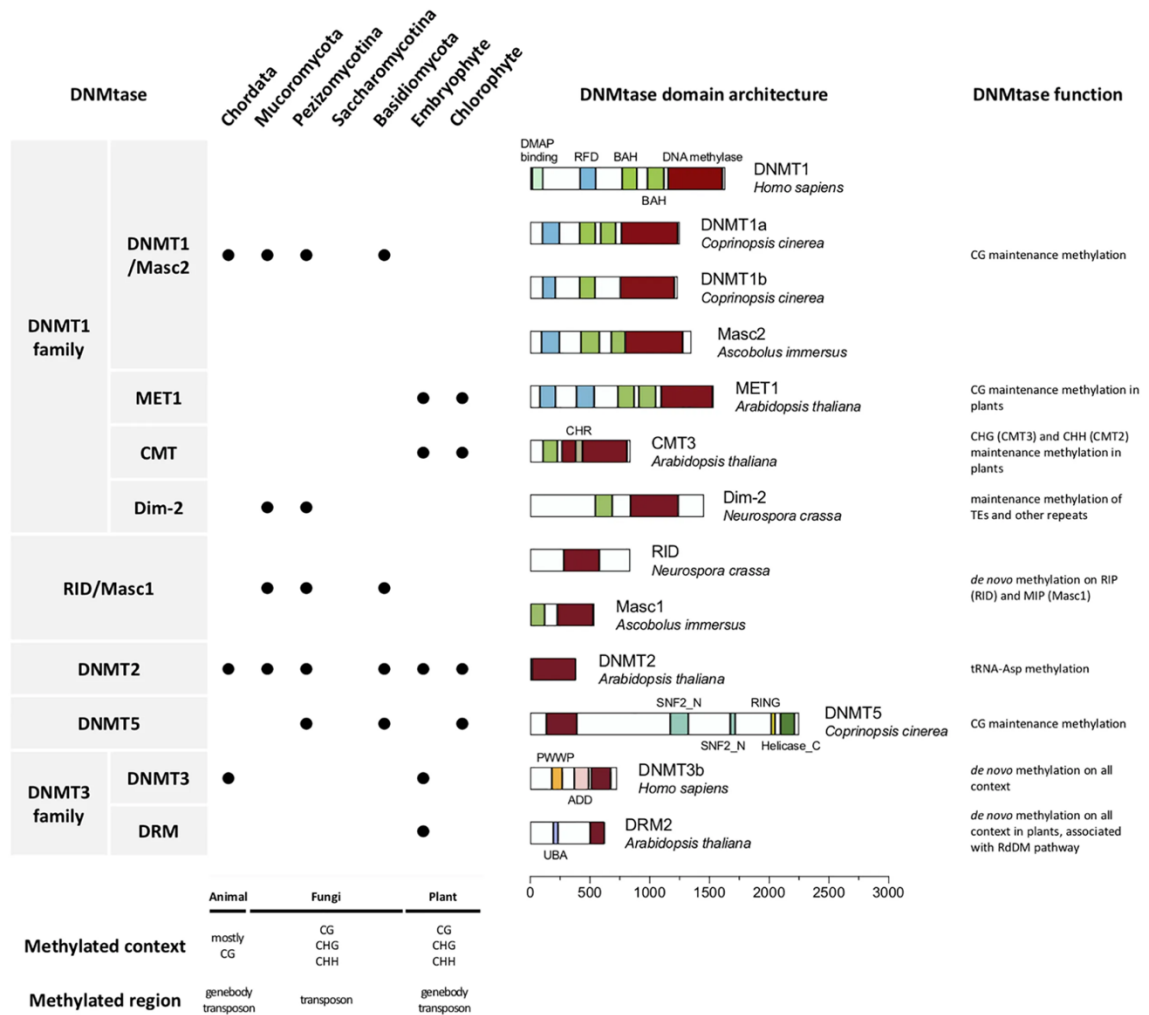


Figure 2. DNA methyltransferases in eukaryotes¹¹.

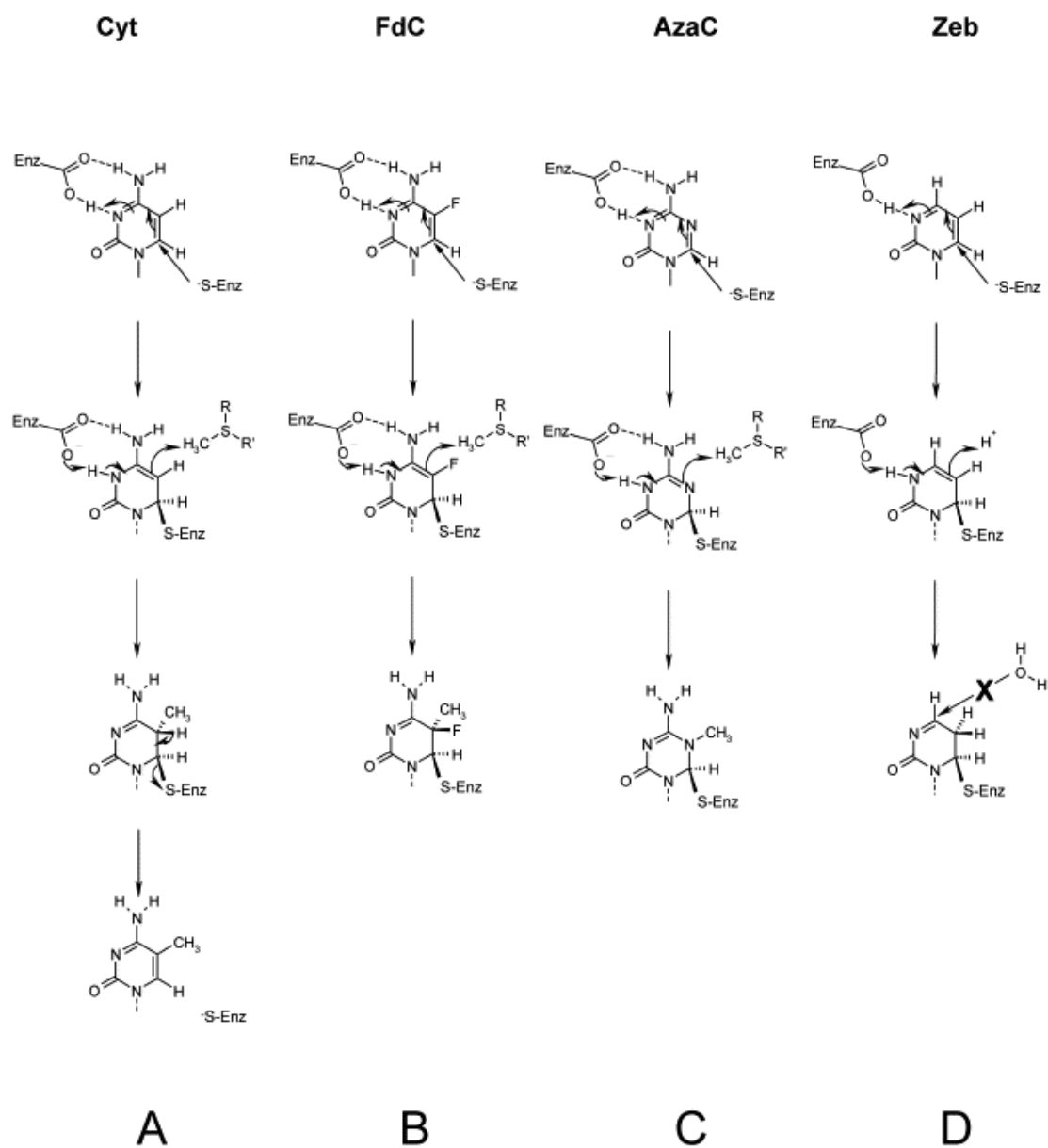


Figure 3. The reaction pathway of C5 MTases in the presence and in the absence of mechanism-based inhibitors⁸¹.

References

1. Mattei, A. L., Bailly, N. & Meissner, A. DNA methylation: a historical perspective. *Trends in Genetics* 38, 676–707 (2022).
2. Li, E. & Zhang, Y. DNA methylation in mammals. *Cold Spring Harb Perspect Biol* 6, a019133 (2014).
3. Jang, H. S., Shin, W. J., Lee, J. E. & Do, J. T. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes (Basel)* 8, 148 (2017).
4. Cokus, S. J. et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215–219 (2008).
5. Smith, S. S. & Ratner, D. I. Lack of 5-methylcytosine in Dictyostelium discoideum DNA. *Biochem J* 277 (Pt 1), 273–275 (1991).
6. Drewell, R. A. et al. The Dictyostelium discoideum genome lacks significant DNA methylation and uncovers palindromic sequences as a source of false positives in bisulfite sequencing. *NAR Genom Bioinform* 5, lqad035 (2023).
7. Hu, C.-W., Chen, J.-L., Hsu, Y.-W., Yen, C.-C. & Chao, M.-R. Trace analysis of methylated and hydroxymethylated cytosines in DNA by isotope-dilution LC-MS/MS: first evidence of DNA methylation in Caenorhabditis elegans. *Biochem J* 465, 39–47 (2015).
8. Bird, A. Molecular biology. Methylation talk between histones and DNA. *Science* 294, 2113–2115 (2001).
9. Capuano, F., Mülleder, M., Kok, R., Blom, H. J. & Ralser, M. Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. *Anal Chem* 86, 3697–3702 (2014).
10. Foss, H. M., Roberts, C. J., Claeyss, K. M. & Selker, E. U. Abnormal chromosome behavior in Neurospora mutants defective in DNA methylation. *Science* 262, 1737–1741 (1993).
11. Nai, Y.-S., Huang, Y.-C., Yen, M.-R. & Chen, P.-Y. Diversity of Fungal DNA Methyltransferases and Their Association With DNA Methylation Patterns. *Front Microbiol* 11, 616922 (2020).
12. Russell, P. J., Rodland, K. D., Rachlin, E. M. & McCloskey, J. A. Differential DNA methylation during the vegetative life cycle of Neurospora crassa. *J Bacteriol* 169, 2902–2905 (1987).
13. Jones, P. A. & Takai, D. The role of DNA methylation in mammalian epigenetics. *Science* 293, 1068–1070 (2001).
14. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11, 204–220 (2010).
15. Zhang, X. et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126, 1189–1201 (2006).

16. Liang, W. et al. Deciphering the synergistic and redundant roles of CG and non-CG DNA methylation in plant development and transposable element silencing. *New Phytol* 233, 722–737 (2022).
17. Muyle, A. M., Seymour, D. K., Lv, Y., Huettel, B. & Gaut, B. S. Gene Body Methylation in Plants: Mechanisms, Functions, and Important Implications for Understanding Evolutionary Processes. *Genome Biol Evol* 14, evac038 (2022).
18. Wang, Y. et al. A modulatory role of CG methylation on gene expression in soybean implicates its potential utility in breeding. *Plant Biotechnol J* 23, 1585–1600 (2025).
19. Wang, Y. et al. A modulatory role of CG methylation on gene expression in soybean implicates its potential utility in breeding. *Plant Biotechnol J* 23, 1585–1600 (2025).
20. Kato, M., Miura, A., Bender, J., Jacobsen, S. E. & Kakutani, T. Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. *Curr Biol* 13, 421–426 (2003).
21. Stroud, H. et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* 21, 64–72 (2014).
22. Galagan, J. E. & Selker, E. U. RIP: the evolutionary cost of genome defense. *Trends Genet* 20, 417–423 (2004).
23. Selker, E. U. & Stevens, J. N. DNA methylation at asymmetric sites is associated with numerous transition mutations. *Proc Natl Acad Sci U S A* 82, 8114–8118 (1985).
24. Selker, E. U., Fritz, D. Y. & Singer, M. J. Dense nonsymmetrical DNA methylation resulting from repeat-induced point mutation in *Neurospora*. *Science* 262, 1724–1728 (1993).
25. Gorovsky, M. A., Hattman, S. & Pleger, G. L. (6 N)methyl adenine in the nuclear DNA of a eucaryote, *Tetrahymena pyriformis*. *J Cell Biol* 56, 697–701 (1973).
26. Cummings, D. J., Tait, A. & Goddard, J. M. Methylated bases in DNA from *Paramecium aurelia*. *Biochim Biophys Acta* 374, 1–11 (1974).
27. Hattman, S., Kenny, C., Berger, L. & Pratt, K. Comparative study of DNA methylation in three unicellular eucaryotes. *J Bacteriol* 135, 1156–1157 (1978).
28. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 9, 465–476 (2008).
29. Greer, E. L. et al. DNA Methylation on N6-Adenine in *C. elegans*. *Cell* 161, 868–878 (2015).
30. Zhang, G. et al. N6-methyladenine DNA modification in *Drosophila*. *Cell* 161, 893–906 (2015).
31. Wion, D. & Casadesús, J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat Rev Microbiol* 4, 183–192 (2006).
32. Boulias, K. & Greer, E. L. Means, mechanisms and consequences of adenine methylation in DNA. *Nat Rev Genet* 23, 411–428 (2022).
33. Li, X. et al. The exploration of N6-deoxyadenosine methylation in mammalian genomes. *Protein Cell* 12, 756–768 (2021).

34. Shen, C., Wang, K., Deng, X. & Chen, J. DNA N6-methyldeoxyadenosine in mammals and human disease. *Trends Genet* 38, 454–467 (2022).
35. Feng, X. & He, C. Mammalian DNA N6-methyladenosine: Challenges and new insights. *Mol Cell* 83, 343–351 (2023).
36. Kong, Y. et al. Critical assessment of DNA adenine methylation in eukaryotes using quantitative deconvolution. *Science* 375, 515–522 (2022).
37. Koh, C. W. Q. et al. Single-nucleotide-resolution sequencing of human N6-methyldeoxyadenosine reveals strand-asymmetric clusters associated with SSBP1 on the mitochondrial genome. *Nucleic Acids Res* 46, 11659–11670 (2018).
38. Hao, Z. et al. N6-Deoxyadenosine Methylation in Mammalian Mitochondrial DNA. *Mol Cell* 78, 382–395.e8 (2020).
39. Li, Z. et al. N6-methyladenine in DNA antagonizes SATB1 in early development. *Nature* 583, 625–630 (2020).
40. Xiao, C.-L. et al. N6-Methyladenine DNA Modification in the Human Genome. *Mol Cell* 71, 306–318.e7 (2018).
41. Xie, Q. et al. N6-methyladenine DNA Modification in Glioblastoma. *Cell* 175, 1228–1243.e20 (2018).
42. Sheng, X., Wang, J., Guo, Y., Zhang, J. & Luo, J. DNA N6-Methyladenine (6mA) Modification Regulates Drug Resistance in Triple Negative Breast Cancer. *Front Oncol* 10, 616098 (2020).
43. Martin, J. L. & McMillan, F. M. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr Opin Struct Biol* 12, 783–793 (2002).
44. Yang, N. & Xu, R.-M. Structure and function of the BAH domain in chromatin biology. *Crit Rev Biochem Mol Biol* 48, 211–221 (2013).
45. Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* 14, 9–25 (1975).
46. Chen, T. & Li, E. Structure and function of eukaryotic DNA methyltransferases. *Curr Top Dev Biol* 60, 55–89 (2004).
47. Ponger, L. & Li, W.-H. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol* 22, 1119–1128 (2005).
48. de Mendoza, A., Lister, R. & Bogdanovic, O. Evolution of DNA Methylome Diversity in Eukaryotes. *J Mol Biol* 432, 1687–1705 (2020).
49. Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74, 481–514 (2005).
50. Kim, D. J. The Role of the DNA Methyltransferase Family and the Therapeutic Potential of DNMT Inhibitors in Tumor Treatment. *Curr Oncol* 32, 88 (2025).

51. Du, J. et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* 151, 167–180 (2012).
52. Qin, W. et al. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res* 25, 911–929 (2015).
53. Zhang, J. et al. S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. *Cell Res* 21, 1723–1739 (2011).
54. 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* 68, 350–360.e7 (2017).
55. Kikuchi, A. et al. Structural basis for activation of DNMT1. *Nat Commun* 13, 7130 (2022).
56. Lu, J. et al. Structure and autoinhibitory regulation of MET1 in the maintenance of plant CG methylation. *Plant Cell* 37, koaf246 (2025).
57. Goll, M. G. et al. Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science* 311, 395–398 (2006).
58. Jeltsch, A. et al. Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. *RNA Biol* 14, 1108–1123 (2017).
59. Zhong, X. et al. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell* 157, 1050–1060 (2014).
60. Cao, X. et al. Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr Biol* 13, 2212–2217 (2003).
61. Zemach, A. & Zilberman, D. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr Biol* 20, R780–785 (2010).
62. de Mendoza, A., Lister, R. & Bogdanovic, O. Evolution of DNA Methylome Diversity in Eukaryotes. *J Mol Biol* 432, 1687–1705 (2020).
63. He, Z. et al. RID is required for both repeat-induced point mutation and nucleation of a novel transitional heterochromatic state for euchromatic repeats. *Nucleic Acids Res* 53, gkaf263 (2025).
64. Malagnac, F. et al. A gene essential for de novo methylation and development in *Ascombolus* reveals a novel type of eukaryotic DNA methyltransferase structure. *Cell* 91, 281–290 (1997).
65. Wang, J. et al. Structural insights into DNMT5-mediated ATP-dependent high-fidelity epigenome maintenance. *Mol Cell* 82, 1186–1198.e6 (2022).
66. Liu, O. W. et al. Systematic genetic analysis of virulence in the human fungal pathogen *Cryptococcus neoformans*. *Cell* 135, 174–188 (2008).
67. Huff, J. T. & Zilberman, D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* 156, 1286–1297 (2014).
68. Hsu, K.-W. et al. METTL4-mediated nuclear N6-deoxyadenosine methylation promotes metastasis through activating multiple metastasis-inducing targets. *Genome Biol* 23, 249 (2022).

69. Wong, J. M. & Eirin-Lopez, J. M. Evolution of Methyltransferase-Like (METTL) Proteins in Metazoa: A Complex Gene Family Involved in Epitranscriptomic Regulation and Other Epigenetic Processes. *Mol Biol Evol* 38, 5309–5327 (2021).
70. Ma, C. et al. A novel N6-Deoxyadenine methyltransferase METL-9 modulates *C. elegans* immunity via dichotomous mechanisms. *Cell Res* 33, 628–639 (2023).
71. Wu, T. & He, C. 6mA-METL-9 axis regulates innate immunity in *C. elegans*. *Cell Res* 33, 581–582 (2023).
72. Nan, B. & Gao, S. 5mC and 6mA DNA methylation in eukaryotic genomes. *Trends in Biochemical Sciences* 51, 261–274 (2026).
73. Cheng, T. et al. Identification and characterization of the de novo methyltransferases for eukaryotic N6-methyladenine (6mA). *Sci Adv* 11, eadq4623 (2025).
74. Wang, Y. et al. Dual modes of DNA N6-methyladenine maintenance by distinct methyltransferase complexes. *Proc Natl Acad Sci U S A* 122, e2413037121 (2025).
75. Cheng, X. & Roberts, R. J. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res* 29, 3784–3795 (2001).
76. O’Gara, M., Klimasauskas, S., Roberts, R. J. & Cheng, X. Enzymatic C5-cytosine methylation of DNA: mechanistic implications of new crystal structures for HhaI methyltransferase-DNA-AdoHcy complexes. *J Mol Biol* 261, 634–645 (1996).
77. Jones, P. A. & Taylor, S. M. Cellular differentiation, cytidine analogs and DNA methylation. *Cell* 20, 85–93 (1980).
78. Santi, D. V., Garrett, C. E. & Barr, P. J. On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell* 33, 9–10 (1983).
79. Sheikhnejad, G. et al. Mechanism of inhibition of DNA (cytosine C5)-methyltransferases by oligodeoxyribonucleotides containing 5,6-dihydro-5-azacytosine. *J Mol Biol* 285, 2021–2034 (1999).
80. Kumar, S. et al. DNA containing 4’-thio-2’-deoxycytidine inhibits methylation by HhaI methyltransferase. *Nucleic Acids Res* 25, 2773–2783 (1997).
81. Zhou, L. et al. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J Mol Biol* 321, 591–599 (2002).
82. Beh, L. Y. et al. Identification of a DNA N6-Adenine Methyltransferase Complex and Its Impact on Chromatin Organization. *Cell* 177, 1781–1796.e25 (2019).
83. Oerum, S., Meynier, V., Catala, M. & Tisné, C. A comprehensive review of m6A/m6Am RNA methyltransferase structures. *Nucleic Acids Res* 49, 7239–7255 (2021).
84. Wang, P., Doxtader, K. A. & Nam, Y. Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. *Mol Cell* 63, 306–317 (2016).
85. Shao, Z., Lu, J., Khudaverdyan, N. & Song, J. Multi-layered heterochromatin interaction as a switch for DIM2-mediated DNA methylation. *Nat Commun* 15, 6815 (2024).

- 86. Shao, Z. et al. Structural insight into the substrate binding of the AMT complex via an inhibitor-trapped state. *Protein Sci* 34, e70265 (2025).
- 87. Bird, A. P. & Wolffe, A. P. Methylation-Induced Repression— Belts, Braces, and Chromatin. *Cell* 99, 451–454 (1999).
- 88. Razin, A. & Riggs, A. D. DNA Methylation and Gene Function. *Science* 210, 604–610 (1980).
- 89. Panning, B. & Jaenisch, R. RNA and the Epigenetic Regulation of X Chromosome Inactivation. *Cell* 93, 305–308 (1998).

Chapter 2

Multi-layered heterochromatin interaction as a switch for DIM2-mediated DNA methylation

Abstract

Functional crosstalk between DNA methylation, histone H3 lysine-9 trimethylation (H3K9me3) and heterochromatin protein 1 (HP1) is essential for proper heterochromatin assembly and genome stability. However, how repressive chromatin cues guide DNA methyltransferases for region-specific DNA methylation remains largely unknown. Here, we report structure-function characterizations of DNA methyltransferase Defective-In-Methylation-2 (DIM2) in *Neurospora*. The DNA methylation activity of DIM2 requires the presence of both H3K9me3 and HP1. Our structural study reveals a bipartite DIM2-HP1 interaction, leading to a disorder-to-order transition of the DIM2 target-recognition domain that is essential for substrate binding. Furthermore, the structure of DIM2-HP1-H3K9me3-DNA complex reveals a substrate-binding mechanism distinct from that for its mammalian orthologue DNMT1. In addition, the dual recognition of H3K9me3 peptide by the DIM2 RFTS and BAH1 domains allosterically impacts the DIM2-substrate binding, thereby controlling DIM2-mediated DNA methylation. Together, this study uncovers how multiple heterochromatin factors coordinately orchestrate an activity-switching mechanism for region-specific DNA methylation.

Introduction

In eukaryotes, compartmentalization of chromatin into transcription-active euchromatin and repressive heterochromatin is essential for gene regulation and genome stability¹. Establishment and maintenance of proper heterochromatin assembly depends on an intertwined network of epigenetic mechanisms, such as histone H3 lysine 9 trimethylation (H3K9me3) and DNA methylation at the C-5 position of cytosine^{2, 3}. Specific readout of these epigenetic marks by respective reader proteins, such as heterochromatin protein 1 (HP1) for H3K9me3⁴⁻⁶ and methyl-cytosine binding proteins for DNA methylation⁷, leads to formation of a compact heterochromatin environment underlying gene silencing⁸. Mechanistic understanding of the functional interplay between these epigenetic pathways is crucial for deciphering the mechanisms behind gene regulation and cell fate determination.

C-5 cytosine DNA methylation is widely present in eukaryotic species, ranging from mammals, plants to filamentous fungi^{9, 10}. However, the DNA methylation machinery and its substrate specificity have diversified throughout evolution. In mammals, DNA methylation predominantly occurs within the CG dinucleotide context, accounting for ~70% of total CG sites¹¹. Mammalian DNA methylation is mainly established by *de novo* DNA methyltransferases DNMT3A and DNMT3B, together with a non-catalytic paralog DNMT3L, and maintained by DNA methyltransferase 1 (DNMT1)¹²⁻¹⁵. Plant DNA methylation occurs in all sequence contexts: CG, CHG, and CHH (H = A, C, T)¹⁴, all of which are established by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2)¹⁴. In *Arabidopsis*, DNA METHYLTRANSFERASE 1 (MET1) mediates the

maintenance of CG methylation¹⁶, CHROMOMETHYLASE 3 (CMT3) mediates the maintenance of CHG methylation¹⁷, while CHROMOMETHYLASE 2 (CMT2) and DRM2 mediate CHH methylation at long heterochromatic transposable elements (TEs)¹⁸ and short euchromatic TEs, respectively¹⁹. In *Neurospora*, DNA methylation accounts for 1.5% of total cytosines in all sequence contexts²⁰⁻²², with Defective In Methylation-2 (DIM2) as the sole DNA methyltransferase in vegetative cells²³. It has been established that DNA methylation in *Neurospora* is mostly enriched in the constitutive heterochromatin, closely associated with the AT-rich relics of repeat-induced point mutation (RIP)^{24, 25}. However, due to lack of mechanistic knowledge, how DIM2 orchestrates such unique yet complexed methylome in *Neurospora* is unknown.

Functional crosstalk between DNA methylation and repressive histone modifications is an evolutionarily conserved mechanism that underpins stable maintenance of heterochromatic assembly²⁶. Both mammalian DNMT1 and plant CMT3 harbor reader modules that specifically recognize histone H3K9 methylation, which in turn allosterically regulates DNMT1- and CMT3-mediated DNA methylation²⁷⁻³⁰. Likewise, DIM2-mediated DNA methylation in *Neurospora* was shown to be initiated by Defective-In-Methylation-5 (DIM-5), a histone H3K9me3 writer³¹. The mechanistic link between DIM2-mediated DNA methylation and heterochromatin is further strengthened by the observation of a direct interaction between DIM2 and HP1³². Mutation of either DIM-5 or HP1 led to DNA methylation abolishment in *Neurospora*^{31, 33, 34}, highlighting an essential role of heterochromatin regulation in DIM2 activity. To date, the molecular basis for the functional regulation of DIM2 by HP1 and H3K9me3 remains elusive.

DIM2 belongs to the superfamily of DNMT1, bearing ~16% sequence identity with human DNMT1^{21, 23}. Like DNMT1, it is predicted to contain a C-terminal methyltransferase (MTase) domain preceded by an N-terminal tail (NT), a putative replication-foci-targeting sequence (RFTS) domain and a pair of bromo-adjacent homology (BAH1 and BAH2) domains (Fig. 1a)²³. The corresponding RFTS, BAH1 and MTase domains in DNMT1 have been characterized previously^{15, 29, 35-40}. For instance, the DNMT1 MTase domain strictly methylates CG sites, with marked substrate preference for hemimethylated CGs^{13, 37, 41}; the DNMT1 RFTS domain interacts with the MTase domain to inhibit the DNA binding of DNMT1³⁸⁻⁴⁰, which can be relieved through the interaction between the RFTS domain and histone H3K9me3 and H3 mono-ubiquitylated at lysine 14, 18 and/or 23^{29, 42-45}; and the DNMT1 BAH1 domain serves as a specific reader for histone H4 lysine 20 trimethylation (H4K20me3) to fine-tune the region-specific DNA methylation in L1 transposons³⁵. In contrast, the biochemical roles of the functional domains of DIM2 have yet to be characterized. Considering that *Neurospora* and mammalian genomes are associated with distinct patterns of cytosine methylation (all sequence contexts in *Neurospora* vs CG-specific in mammals)²¹, understanding the structure and mechanism of DIM2-mediated DNA methylation is important for unravelling how DNA methyltransferases have evolved to adapt various genomic complexities.

To delineate the mechanistic basis of DIM2-mediated DNA methylation, we set out to investigate the functional regulation of *Neurospora Crassa* DIM2 by HP1 and H3K9me3. Unlike most other DNA methyltransferases characterized so far that, in the absence of

regulatory proteins, retain a basal DNA methylation activity^{18, 27, 46-50}, DIM2 requires the presence of HP1 and H3K9me3 for a notable DNA activity. Next, we solved the single-particle cryogenic electron microscopy (cryo-EM) structures of DIM2 in complex with HP1, DIM2 in apo form (apo-DIM2), and DIM2 in complex with HP1, H3K9me3 peptides and a DNA duplex. Structural and biochemical analysis of the DIM2-HP1 complex reveals a bipartite interaction between DIM2 and HP1, involving the N-terminal tail (NT) and the target recognition domain (TRD) of DIM2 and the chromoshadow domains (CSDs) of HP1 dimer. In comparison with apo-DIM2, the interaction between DIM2 TRD and HP1 CSDs leads to a disorder-to-order transition of the TRD, providing an explanation for the requirement of HP1 in DIM2-mediated DNA methylation. Furthermore, the structure of DIM2 in complex with HP1, H3K9me3, and DNA reveals that the DIM2 RFTS domain, unlike its counterpart in DNMT1 that functions as a DNA-competitive inhibitor³⁸⁻⁴⁰, forms an additional DNA-binding site to facilitate substrate association. The interaction between DIM2 and H3K9me3 peptides is mediated by both the RFTS and BAH1 domains of DIM2, with the RFTS-H3K9me3 interaction positioning H3 K14 for a direct interaction with substrate DNA, and the BAH1-H3K9me3 interaction repositioning a loop bridging the TRD and the catalytic core of DIM2 for DNA contact. Disruption of the DIM2 residues key for HP1, H3K9me3, or DNA binding severely impairs the DNA methylation activity of DIM2. Together, this study unravels a regulatory model for DIM2, where multiple interactions of heterochromatin factors synergistically control DIM2-mediated DNA methylation, with important implications in the functional crosstalk between DNA methylation and heterochromatin cues.

Materials and Methods

Protein expression and purification.

DNA encoding *Neurospora Crassa* DIM2 (NCBI accession no. XP_959891.1) (residues 1-1242) was derived from a genomic DNA library of *Neurospora Crassa* by PCR amplification. The DNA fragment encoding *Neurospora Crassa* HP1 (residues 1-266; uniprot number. Q1K612) was chemically synthesized by Integrated DNA Technologies, codon optimized for bacterial expression. Both the DNA fragments for DIM2 and HP1 were cloned into an in-house vector preceded by an N-terminal His₆-MBP tag and a TEV cleavage site. The expression plasmids were transformed into BL21 (DE3) RIL cells strain (Agilent Technologies). The transformed cells were first grown at 37 °C until OD₆₀₀ reached 1.0. Afterwards, the temperature was shifted to 16 °C and the cells were induced by addition of 0.13 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and continued to grow overnight. The cells were harvested and lysed in buffer containing 50 mM Tris-HCl (pH 8.0), 1 M NaCl, and 25 mM imidazole. After centrifugation, the His₆-MBP-DIM2 or His₆-MBP-HP1 fusion protein was purified from the supernatant through a Ni-NTA column (Cytiva), followed by ion-exchange chromatography on a HiTrap Heparin HP column (Cytiva), removal of the His₆-MBP tag by TEV-mediated proteolytic cleavage, and a second round of 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 20 mM Tris-HCl (PH 7.5), 100 mM NaCl, 5% glycerol and 5 mM DTT. Purified protein samples were stored at -80 °C in a concentration of ~6 mg/mL for future use.

For biochemical analysis of the DIM2 RFTS domain, the DNA encoding residues 1-389 of DIM2 was inserted into the expression vector described above, and the protein was purified in the same manner as that for residues 1-1242 of DIM2, except that a Superdex 75 16/600 column (GE Healthcare) was used for final protein purification. The DIM2 mutants were generated by site-directed mutagenesis and purified using the same approach as described for WT protein.

In vitro DNA methylation assay.

In vitro DNA methylation assay was performed in a 20- μ L reaction mixture. For the substrate specificity assay, the following DNA duplexes used as substrates: CG DNA (Upper strand: 5'-ATTATTAATCGAAATTTA-3'), hemimethylated CG DNA (Upper strand: 5'-TAAATTTXGATTAATAAT-3', X=5-methylcytosine; Lower strand: 5'-ATTATTAATCGAAATTTA -3'), CHG DNA (Upper strand: 5'-ATTATTAATCTGAATTTA-3'), and CHH DNA (Upper strand: 5'-ATTATTAATCTAAATTTA-3'). The reaction mixture contains 0.25 μ M DIM2, 1.0 μ M DNA duplex, 1.0 μ M synthesized H3K9me3 peptide (residues 1-24 of H3 plus a C-terminal tryptophan), 0.5 μ M HP1, and 0.56 μ M S-adenosyl-L-[methyl- 3 H] methionine with a specific activity of 18 Ci/mmol (PerkinElmer) in buffer containing 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.05% β -mercaptoethanol, 5% glycerol, and 200 μ g/mL BSA. For the rest of the DNA methylation assay, each reaction mixture contains 1.0 μ M DIM2, 1.0 μ M (CAT)₁₂/(ATG)₁₂ DNA duplex and 0.56 μ M S-adenosyl-L-[methyl- 3 H] methionine with a specific activity of 18 Ci/mmol (PerkinElmer) in buffer containing 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.05% β -mercaptoethanol, 5% glycerol, and 200

μg/mL BSA. In addition, 1.0 μM synthesized H3K9me3 or H3K9me0 peptide (residues 1-24 of H3 plus a C-terminal tryptophan) and/or 2 μM HP1 protein was included to evaluate the effect of H3 and HP1 regulation. The reaction mixture was incubated at 30 °C for 30 minutes, unless indicated otherwise, before being quenched by the addition of 5 μL of 10 mM cold SAM. After the reaction, 8 μL of the mixtures was loaded onto Hybond N nylonmembrane (GE Healthcare), which was left to dry out at room temperature. The membrane was subsequently washed with 0.2 M ammonium bicarbonate (pH 8.2) two times (20 minutes per time), deionized water (20 minutes) once, and 95% ethanol (5 minutes) once. After air dried, the membrane carrying each sample was transferred into a vial containing 3 mL scintillation buffer (Fisher). The tritium scintillation was measured and recorded by a Beckman LS6500 counter. Each the reaction was repeated three times.

Analytical size-exclusion chromatography

For size-exclusion chromatography analysis, 400 μL of the sample mixture containing 1.0 μM DIM2 and/or 2.5 μM HP1 was loaded onto Superdex 200 increase 10/300 GL column (GE Healthcare) and eluted at a flow rate of 0.5 mL/min, in the buffer containing 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5% glycerol, and 5 mM DTT. The proteins in each fraction were then visualized by SDS-PAGE.

ITC binding assay

ITC measurements were performed using a MicroCal iTC200 instrument (GE Healthcare). To measure the bindings between proteins and H3K9me3 peptide, 0.04-0.05 mM DIM2, WT or mutant (W261A, W581A, or W261A/W581A), DIM2 RFTS domain,

or HP1, WT or mutant (W98A), was titrated with 0.5 mM histone peptide at 7 °C. Prior to the titration, both peptide and protein samples were subjected to overnight dialysis against buffer containing 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, and 1 mM β -mercaptoethanol. Buffer to buffer titration was performed to ensure no abnormality of baseline. Analyses of all data were performed with MicroCal Origin software, fitted with single-site binding mode.

Thermal shift assay

Thermal shift assays were performed for WT and R1104A DIM2 proteins using a Bio-Rad CFX Connect Real-Time PCR Detection System. Each 20- μ L reaction mixture contained 5 μ M WT or R1104A-mutated DIM2, and/or 10 μ M HP1 protein, dissolved in buffer containing 20 mM Tris-HCl (PH 7.5), 100 mM NaCl, 5% glycerol, 5 mM DTT, and 1 $\times$ GloMelt Dye (Biotium). The sample plates were subject to heat from 4 to 95 °C with a stepwise increment of 5 °C. Fluorescence intensity was recorded with the excitation and emission wavelength of 470 nm and 510 nm, respectively. The experiments were carried out in triplicate.

Electrophoretic mobility shift assay

Each 10- μ L reaction mixture contained 0.1 μ M DNA duplex (Upper strand: 5'-ATTATTAATCTAAATTTA-3') mixed with 0, 0.2, or 0.4 μ M DIM2, WT or mutant (W261A or W581A), and/or 0.8 μ M H3K9me3 peptide in binding buffer containing 20 mM Tris-HCl (PH 8.0), 75 mM NaCl, 10% glycerol, 5 mM DTT, and 2% NP-40. Samples were resolved on a 5% w/v polyacrylamide gel (59:1 for acrylamide:bis-acrylamide) with the running buffer containing 1 $\times$ Tris-Glycine (pH 8.6) at 4 °C for 1

hour. The gel was stained with SYBRTM Gold (Thermo Fisher Scientific) for 10 minutes and visualized by ChemiDoc Imaging System (Bio-Rad).

Cryo-EM data collection

For DIM2-HP1 complex, aliquots of 3 μ L of DIM2-HP1 complex after gel filtration in a buffer containing 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 2.5% Glycerol, and 5 mM DTT were applied to glow-discharged Quantifoil[®] (1.2/1.3) grids. The grids were blotted for 6 s at 95% humidity with an offset of 5 s and plunge frozen into liquid ethane using a Vitrobot Mark IV (Thermo Fisher). Grids were imaged on a 300 keV Titan Krios cryo-electron microscope (Thermo Fisher) equipped with a K3 camera (Gatan) by Pacific Northwest Center for Cryo-EM (PNCC). Movies were collected at a calibrated magnification of $\times 81,000$, corresponding to a 1.056 Å per physical pixel. The dose was set to a total of 67 electrons/Å² over an exposure of 68 frames. Automated data collection was carried out using SerialEM with a nominal defocus range set from -0.8 to -2.5 μ m. A total of 4009 movies were collected over 24 h.

For apo-DIM2, aliquots of 3 μ L of DIM2 in a dialysis buffer containing 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 2.5% Glycerol, and 5 mM DTT were applied to glow-discharged Quantifoil[®] (1.2/1.3) grids. The grids were blotted for 6 s at 95 % humidity and plunge frozen into liquid ethane using a Vitrobot Mark IV (Thermo Fisher). Grids were imaged on a 300 keV Titan Krios cryo-electron microscope (Thermo Fisher) equipped with a K3 camera (Gatan) by NCI cryo-EM Facility (NCEF). Movies were collected at a calibrated magnification of $\times 81,000$, corresponding to a 1.07 Å per physical pixel. The dose was set to a total of 50 electrons/Å² over an exposure of 40 frames.

Automated data collection was carried out using SerialEM with a nominal defocus range set from -0.6 to -1.5 μm . A total of 8327 movies were collected over 48 h.

For DIM2-HP1-H3K9me3-DNA complex, aliquots of 3 μL of DIM2-HP1-H3K9me3-DNA complex after reaction in a buffer containing 50 mM Tris-HCl (pH 8.0), 2.5% Glycerol, and 40 mM DTT were applied to glow-discharged Quantifoil® (1.2/1.3) grids. The grids were blotted for 6 s at 95% humidity and plunge frozen into liquid ethane using a Vitrobot Mark IV (Thermo Fisher). Grids were imaged on a 300 keV Titan Krios cryo-electron microscope (Thermo Fisher) equipped with a K3 camera (Gatan) by Pacific Northwest Center for Cryo-EM (PNCC). Movies were collected at a calibrated magnification of $\times 81,000$, corresponding to a 1.056 $\text{\AA}$ per physical pixel. The dose was set to a total of 70 electrons/ $\text{\AA}^2$ over an exposure of 30 frames. Automated data collection was carried out using SerialEM with a nominal defocus range set from -0.8 to -2.5 μm . A total of 3728 movies were collected over 24 h.

Image processing

The cryo-EM data for the DIM2-HP1 complex, apo-DIM2, and the DIM2-HP1-H3K9me3-DNA complex, containing 4009, 8327 and 3728 movies, respectively, were processed using the CryoSPARC software package (v4.01)⁷². All movies were motion-corrected using patch motion correction, and the resulting micrographs were subject to patch-based contrast transfer functions (CTFs) estimation. Subsequently, peak picking was carried out using the Topaz program⁷³.

For the DIM2-HP1 complex, 1.5 million particles were extracted, followed by several rounds of 2D classifications. One million particles were then selected based on visible

features of secondary and tertiary structures. Next, three *ab initio* models were generated and subsequently heterogeneously refined. The particles associated with one class with well-defined feature of DIM2 and HP1 were selected for non-uniform refinement, resulting in a final resolution of 2.76 Å.

For apo-DIM2, 4.8 million particles were extracted for 2D classifications. Among them, 2.9 million particles, with a down-scaled pixel size of 2.14 Å, were selected after removal of obvious junk and ice. Two rounds of *ab initio* reconstruction and heterogeneous refinement were carried out, with three models in the first round and six models in the second round. For each round, the particles for the classes with well-defined secondary and tertiary structures were selected. After second round, the selected particles were re-extracted with a pixel size of 1.07 Å and used to generate another four *ab initio* models, followed by another round of heterogeneous refinement. These particles associated with class with the highest resolution were then subject to non-uniform refinement, resulting in a final resolution of 2.88 Å.

For the DIM2-HP1-H3K9me3-DNA complex, 1.6 million particles were extracted with a down-scaled pixel size of 2.11 Å. After 2D classifications, 1.1 million particles were selected based on visible feature of secondary and tertiary features. Five *ab initio* models were generated and subsequently heterogeneously refined. The particles associated with the class with defined features of DIM2 were re-extracted with pixel size of 1.06 Å. Next, the particles were subject to 3D classification into three models, followed by heterogeneous refinement. The particles associated with the class with defined feature of DIM2 and DNA molecules were selected for further CTF refinement and non-uniform

refinement, leading to a final resolution of 2.79 Å. The map was further improved using the anisotropic sharpening module in Phenix (v1.20.1-4487)⁷⁴.

Model building and refinement

The structural models of DIM2 and HP1 were predicted using the AlphaFold2 server⁷⁵. The structural model for the DNA and H3K9me3 peptide were derived from that in the DNMT3A-DNMT3L-DNA complex (PDB 6F57)⁵⁶ and that in the DNMT1 RFTS-H3K9me3-ubiquitin complex (PDB 6PZV)²⁹, respectively. Fitting of the structural models of DIM2, HP1 and/or DNA into the individual maps were performed using UCSF Chimera (v1.16)⁷⁶ and ChimeraX (v1.6.1)⁷⁷. The final model was then obtained after iterative model building in Coot (v0.8.9.1)⁷⁸ and real-space refinement in Phenix. The RMSD for pairwise protein structural alignment is calculated using the online TM-align method⁷⁹ (<https://www.rcsb.org/alignment>).

Results

HP1 and H3K9me3 are both required for DIM2-mediated DNA methylation

To elucidate how DIM2-mediated DNA methylation interplays with heterochromatin factors HP1 and H3K9me3, we performed in vitro DNA methylation assay for an N-terminal fragment of DIM2 (residues 1-1242), spanning the NT, RFTS, BAH1, BAH2, and the MTase domains (Fig. 1a), on a 36-mer DNA duplex containing multiple cytosines (Fig. 1b). Surprisingly, under the experimental condition, DIM2 alone showed no detectable activity (Fig. 1b). The presence of histone H3 peptide (residues 1-24) harboring the H3K9me3 modification (H3₁₋₂₄K9me3) failed to activate DIM2 appreciably, while the presence of full-length HP1 led to a slight but notable activity of DIM2 (Fig. 1b). In contrast, the presence of both HP1 and H3K9me3 peptide greatly increased the DNA methylation efficiency of DIM2, higher than that of DIM2-HP1 mixture by ~90-fold (Fig. 1b and Supplementary Fig. 1a), suggesting a synergistic activity-stimulating effect between HP1 and H3K9me3. Note that such an activity-stimulation effect appears specific to heterochromatin factors, as the replacement of the H3K9me3 peptide by the H3 peptide containing unmethylated H3 K9 (H3K9me0) failed to synergize with HP1 to boost the DNA methylation activity of DIM2 (Fig. 1b). The requirement for HP1 protein is unlikely related to its H3K9me3-binding activity, as our Isothermal Titration Calorimetry (ITC) analysis revealed that impairing the H3K9me3-binding activity of HP1 via the H3K9me3-cage mutation W98A⁵¹ (Supplementary Fig. 1b-d and Supplementary Table 1) failed to abolish HP1-mediated regulation; rather, it

increases the activity-stimulation effect of HP1 on DIM2 by ~2-fold (Supplementary Fig. 1e).

Next, we compared the DNA methylation activity of DIM2 on DNA substrates containing one central CG, CHG or CHH target site. As expected, DIM2 methylates both CG and non-CG DNAs (Supplementary Fig. 1f). Furthermore, under the experimental condition, DIM2 shows a modest substrate preference for hemimethylated over unmodified CG DNA. The caveat of this observation is that only one set of CG, CHG and CHH substrates were used. A detailed understanding of how the DNA sequence context impacts the activity of DIM2 awaits further investigation.

Structural overview of the DIM2-HP1 complex

Consistent with a previous report that DIM2 directly interacts with HP1³², our size-exclusion chromatography analysis reveals co-migration of DIM2 (residues 1-1242) fragment and full-length HP1 (Supplementary Fig. 1g), confirming the formation of the DIM2-HP1 complex. Subsequently, we solved the cryo-EM structure of DIM2 (residues 1-1242) in complex with HP1 and cofactor byproduct S-adenosyl homocysteine (SAH) (denoted as DIM2-HP1 complex thereafter) at overall 2.76-Å resolution (Fig. 1c-f and Supplementary Fig. 2 and Supplementary Table 2). Structural analysis of the DIM2-HP1 complex reveals that one DIM2 molecule binds to a homodimer of HP1 CSD domain (HP1-I and HP1-II) (Fig. 1c-f). We were able to trace the majority of DIM2 residues, except for M1-Q125, D436-E546, D562-D578, K592-S602, K699-T702, T933-K936, K1143-D1148 (Fig. 1c-d and Supplementary Fig. 3a-e). On the other hand, although full-length HP1 was used for cryo-EM study, we were only able to trace the density for the

CSD domains (residues K189-I251 of HP1-I CSD and residues C191-I251 of HP1-II CSD) (Supplementary Fig. 3f).

The structure of DIM2 reveals that the N-terminal domains, composed of the extended NT and the RFTS, BAH1 and BAH2 domains, are centered around the C-terminal MTase domain (Fig. 1c-f): the RFTS domain is positioned on one side, and the BAH1 and BAH2 domains are positioned on the other side, reminiscent of what was observed for DNMT1^{39, 40}. The RFTS domain is followed by an extended linker that traverses the catalytic core, forming a pair of α -helices anchored next to the catalytic site (Fig. 1e,f). The MTase domain is further comprised of two subdomains: the catalytic core and the TRD, with the catalytic core harboring the active site and the SAH molecule (Fig. 1e-g and Supplementary Fig. 3g). On the other hand, the structure of HP1 CSD dimer resembles that of mammalian HP1⁵¹, with each CSD containing a three-stranded antiparallel β -sheet that is packed by two C-terminal α -helices (Fig. 1e,f).

Structural basis for the DIM2-HP1 interaction

Association of DIM2 with HP1 CSDs is mediated by a bipartite interaction, involving one interface formed by the canonical peptide-binding groove of the CSD homodimer^{51, 52} and a portion of the NT and TRD of DIM2 (Interface I), and another interface formed by the C-terminal helix-proximate region of HP1-I CSD and DIM2 TRD (Interface II) (Fig. 2a-d).

At the interface I, the N-terminal region (residues Q125-H129) of the DIM2 NT pairs in parallel with the first β -strand of HP1-II CSD via hydrogen-bonding interactions between DIM2 S127 and HP1 H247' (prime symbol denotes residues from the HP1-II) and

between DIM2 H129 and HP1 A204'. Next, the DIM2 NT diverts by 90° at residue H129 and runs in an extended form along the surface groove formed by the C-terminal tails of both HP1 CSDs (Fig. 2a, c), mediated by backbone hydrogen bonds involving DIM2 T131, V132, D133 and L134 and HP1 H247, R249 and I251. In addition, residues I130, V132, L134, and P135 of DIM2 engage in non-polar contacts with HP1 (L213, F243, Y244, and V248 of HP1-I and residues F243', Y244', V248', I250', and I251' of HP1-II), and residue D133 of DIM2 forms a salt bridge with residue R249' of HP1-II (Fig. 2c). Note that the positioning of residues I130, V132, and L134 of DIM2 matches the -2, 0, and +2 sites of the canonical HP1-interacting $P_2XV_0XL_{+2}$ motif^{51, 52}, respectively, indicative of a canonical interaction mechanism. At the exit of the surface groove, DIM2 L139 further engages in a hydrophobic contact with HP1-I Y216, and DIM2 T138 forms a hydrogen bond with HP1-I A204 (Fig. 2c). Subsequently, residues N141-R148 of DIM2 divert by 90° at residue N141 to traverse the β -sheet of the HP1-I CSD, involving a salt-bridge interaction between DIM2 R143 and HP1-I D207, and hydrogen-bonding interactions between DIM2 R143 and Q147 and HP1-I Y216 and K224, respectively (Fig. 2c). Additional intermolecular interactions at the interface I include a salt bridge between DIM2 R1104 and HP1-I D203, and hydrogen-bonding interactions involving DIM2 A1108 and D1185 and HP1-I Q242 and R246 (Fig. 2c, e).

At the interface II, the helical segments (residues 1101-1138) of DIM2 TRD are packed against the β -sheet HP1-I CSD as well as the loop bridging the first two β -strands of HP1-II CSD (Fig. 2d), engaging in intermolecular hydrogen-bonding interactions involving DIM2 H1113, Y1115 and D1134, HP1-I K239, and V248, and HP1-II T210',

and van der Waals contacts involving DIM2 H1113, P1114 and Y1115, and HP1-II H211' and K212' (Fig. 2d). Together, these interactions underpin the association between DIM2 and HP1 CSD homodimer.

To test the structural observations, we selected key HP1-interacting residues (L134, L139 and R1104) (Fig. 2c, e) of DIM2 for mutagenesis, followed by size-exclusion chromatography and in vitro DNA methylation analysis. Introducing the L134A/L139A double mutation to DIM2 led to partial disruption of the DIM2-HP1 complex, as indicated by the presence of three discrete elution peaks corresponding to the DIM2-HP1 complex, DIM2 and HP1, respectively (Fig. 2f and Supplementary Fig.4a-d). Introducing the R1104A mutation did not affect the assembly of the DIM2-HP1 complex appreciably (Supplementary Fig. 4e). However, introducing the L134A/L139A/R1104A triple mutation to DIM2 (DIM2^{LLR}) led to complete disruption of the DIM2-HP1 complex, as indicated by the disappearance of the elution peak for the DIM2-HP1 complex (Fig. 2f and Supplementary Fig. 4f). Consistently, our in vitro DNA methylation assay showed that introducing the L134A/L139A double mutation reduced the DNA methylation efficiency of DIM2 by 14-folds, whereas introducing the DIM2^{LLR} mutation completely abolished the DNA methylation activity of DIM2 (Fig. 2g). These data lend a strong support for the observed DIM2-HP1 interaction and its critical role in DIM2-mediated DNA methylation.

Mechanistic basis for HP1-mediated allosteric regulation of DIM2

To illustrate the mechanism by which HP1 binding controls DIM2-mediated DNA methylation, we further solved the cryo-EM structure of apo-DIM2 at 2.88-Å resolution

(Fig. 3a-d and Supplementary Fig. 5,6 and Supplementary Table 2). We were again able to trace the majority of DIM2 molecule, except for residues M1-Q160, D436-K541, D562-D578, K592-N601, L931-K936, and S1097-G1152 (Fig. 3a-d and Supplementary Fig. 6a-e). As with the DIM2-HP1 complex, the MTase domain of apo-DIM2 harbors a SAH molecule in the active site (Fig. 3e and Supplementary Fig. 6b). Structural alignment of apo-DIM2 with HP1-bound DIM2 gave a root-mean-square deviation (RMSD) of 1.12 Å over 873 aligned Cα atoms (Fig. 3f), indicative of structural similarity. Nevertheless, substantial structural changes were observed for the HP1-interacting regions: residues Q125-S161 of the NT and residues A1096-W1153 of the TRD of DIM2 are well defined in the DIM2-HP1 complex but become disordered in apo-DIM2 (Fig. 3g), suggesting that the HP1 binding induces a disorder-to-order transition of the DIM2 NT and TRD. In addition, less pronounced but notable conformational changes were observed for residues N198-Q204 and A213-K220 of the RFTS domain and residues S1172-G1179 of the TRD (Fig. 3g), reflecting an indirect effect of HP1 interaction.

The TRD of a DNA methyltransferase is knowingly essential for its substrate binding^{15,53}. The fact that the DIM2 residues A1096-W1153, which account for over one-third of the DIM2 TRD (Fig. 1a), undergo a disorder-to-order transition upon HP1 binding raises a possibility that HP1 binding allosterically activates DIM2 through structural stabilization of the DIM2 TRD. To test this notion, we first performed thermal shift assays to compare the stabilities of wild-type (WT) DIM2 with that harboring a TRD mutation, in which the HP1-interacting residue R1104 was replaced by an alanine (R1104A). WT and R1104A

DIM2 both yielded a melting temperature (T_m) of 39.5 °C (Supplementary Fig. 7a,b), in line with the fact that the TRD is structurally disordered in apo-DIM2. Co-incubation of WT DIM2 with HP1 led to a 5.0 °C increase of T_m (Supplementary Fig. 7a,b), consistent with the formation of DIM2-HP1 complex that further stabilizes the DIM2 TRD. In contrast, co-incubation of R1104A DIM2 with HP1 only increased the T_m by 2.5 °C (Supplementary Fig. 7a,b), supporting the notion that the R1104A, through disruption of the interaction between DIM2 TRD and HP1, reduces the stability of the TRD in the DIM2-HP1 complex. Next, we compared the in vitro DNA methylation activities of WT and R1104A-mutated DIM2. Despite that the DIM2 R1104A mutation alone barely impacts the association of the DIM2-HP1 complex (Supplementary Fig. 4e), the activity-stimulation effect of HP1 on R1104A-mutated DIM2 was reduced by ~8-fold when compared with that on WT DIM2 (Fig. 3h). These observations therefore support the notion that HP1 allosterically regulates the DNA methylation activity of DIM2 via structural stabilization of the DIM2 TRD.

Structural overview of the DIM2-HP1-H3K9me3-DNA complex

To elaborate how HP1 and H3K9me3 cooperate in activating DIM2-mediated DNA methylation, we further solved the cryo-EM structure of DIM2 in complex with HP1, H3₁₋₂₄K9me3 peptides and an 18-mer DNA duplex. The DNA molecule contains a zebularine (Z7) on the target DNA strand (Fig. 4a), which permits the formation of a stable, productive complex between a DNA methyltransferase and DNA, as demonstrated previously⁵⁴⁻⁵⁷. The structure of the DIM2-HP1-H3K9me3-DNA complex was solved at

an overall resolution of 2.79 Å (Fig. 4b-d and Supplementary Fig. 8,9 and Supplementary Table 2).

We were able to trace the entire DNA molecule and most of the DIM2 molecule, except for residues M1-S127, E438-V540, and K592-N601 (Fig. 4b-d and Supplementary Fig. 9a-g). The structure reveals that DIM2 binds to HP1, H3₁₋₂₄K9me3 peptide and DNA molecule in a stoichiometry of 1:2:2:1 (Fig. 4b,d). The chromodomains of HP1 remain untraceable, whereas the homodimeric CSD domains associate with DIM2 in the same manner as that in the DIM2-HP1 complex (Fig. 4d vs. Fig. 1e,f). Notably, the DNA molecule is embedded in the cleft between the catalytic core and TRD (Fig. 4d-f), with the Z7 breaking away from its Watson-Crick pair Gua7' (' denotes the base on the non-target strand) and inserting into the active site for a covalent-linkage with the catalytic cystine C926 of DIM2, where Z7 is further locked in place through interactions with other catalytic residues, such as S924, P925, E966, R1013 and R1015 (Fig. 4e). Each DIM2 molecule binds to two H3K9me3 peptides, with one associated with the RFTS domain (Fig. 4d,g) and the other associated with the BAH1 domain (Fig. 4d,h). We were able to trace residues T6-R17 for the H3K9me3 peptide bound to the RFTS domain (Fig. 4f,g), and residues T3'-G13' (prime symbol denotes residues from the BAH1-bound H3K9me3 peptide) for the one bound to the BAH1 domain (Fig. 4f,h).

Structural overlay of the DIM2-HP1-H3K9me3-DNA complex with the DIM2-HP1 complex reveals high similarity, with an RMSD of 0.65 Å over 809 aligned C α atoms. Nevertheless, DNA and H3K9me3 peptides-induced conformational changes were observed for residues I209-Y221 of the RFTS domain, residues V401-V408 of the RFTS-

BAH1 linker, the catalytic loop (residues S924-V938), and residues R1039-S1053 and G1142-A1149 of the TRD (Supplementary Fig. 10a-c). In addition, a loop in the BAH1 domain (residues R561-H579, denoted as allosteric loop herein) undergoes a disorder-to-order transition upon H3K9me3 binding (Supplementary Fig. 10c).

Structural basis for the DIM2-DNA substrate interaction

The interaction between DIM2 and the 18-mer DNA spans 16 base pairs and discrete regions of the catalytic core, the TRD, and the RFTS domain of DIM2 (Fig. 5a and Supplementary Fig. 11a), involving the loop harboring the catalytic cysteine C926 (catalytic loop: residues S924-V938) that accesses the minor groove for stabilizing the flipped Z7, a loop in the RFTS domain (RFTS loop: residues Y199-K220) that contacts the minor groove of the DNA segment downstream of Z7, a loop bridging the catalytic core with the TRD (bridging loop: Y1038-F1059) that cradles the minor groove of the DNA segment upstream of Z7, a loop of the TRD (TRD loop: S1140-L1180) that accesses the major groove, and residues S1097-Y1102 at the N-terminus of the R1104-residing helix in TRD (TRD helix) that interacts with HP1, lining the DNA backbone along the non-target strand (Fig. 5a).

The catalytic loop migrates into the DNA minor groove to interact with both strands of the DNA (Fig. 5a,b). In addition to the interaction with Z7 by DIM2 residues P925, C926, R1013 and R1015 (Fig. 4e), the side chain of DIM2 L932 occupies the space vacated by base flipping of Z7, engaging in van der Waals contacts with the two neighboring bases (Thy6 and Cyt8) (Fig. 5b). Furthermore, DIM2 S930 forms a hydrogen bond with the backbone phosphate of Z7, while DIM2 Q934 donates a hydrogen bond to

the sugar ring of orphan guanine Gua7' that otherwise pairs with Z7. In addition, DIM2 Q941 from the α -helix C-terminal to the catalytic loop (denoted as catalytic helix herein) donates a hydrogen bond to the backbone phosphate of Thy9, the side chain of DIM2 L931 makes van der Waals contacts with the base rings of Thy6 and Gua5' in the minor groove, and residue R406 from the RFTS-BAH1 linker forms a hydrogen-bonding interaction with Cyt10 (Fig. 5b), thereby reinforcing the minor-groove interaction.

Upstream of the target site Z7, the bridging loop of DIM2 runs across both DNA strands, involving residues Y1038, R1039 and K1041 for hydrogen-bonding and/or electrostatic interactions with the backbone phosphate of Cyt5, residues R1043 and N1044 for electrostatic and/or hydrogen-bonding interactions with the backbone of Gua2', and residue N1042 for van der Waals contacts with the backbone of Thy1' (Fig. 5c). Downstream of the target site Z7, residue Y201 from the RFTS loop approaches the minor groove for van der Waals contacts with the Gua11'-Gua10' step, residues Y199 and A200 engage in hydrogen-bonding interactions with the backbone phosphate of Ade16 and Thy15, respectively (Fig. 5d). In addition, the RFTS loop engages in van der Waals contacts with the minor groove, involving DIM2 Y199, Y201, H202, V216, L217, D219, and K220 and both DNA strands (Gua13'-Gua11' and Cyt14-Ade16) (Fig. 5d).

On the major groove side, TRD loop engages in extensive contacts with the nucleotides surrounding orphan guanine Gua7'. Of note, the guanidinium group of DIM2 R1145 forms hydrogen bonds with the O6 atom of Gua7', the N7 atom of Gua8', and the N7 atom of Ade9', while the sidechain carboxylate of DIM2 D1173 forms a bifurcated

hydrogen bond with the N1 and N2 atoms of Gua7', and another hydrogen bond with the N4 atom of Cyt8 (Fig. 5e). Furthermore, the guanidium group of DIM2 R1175 intercalates into the non-target strand, prying open the Gua7'-Ade6' base step, as well as forming a salt bridge with the backbone phosphate of Gua7' (Fig. 5e and Supplementary Fig. 11b). Such an R1175-mediated DNA intercalation increased the helical rise by 3.9 Å (7.1 Å for Ade6'-Gua7' vs. 3.2 Å for B-form DNA in Supplementary Fig. 11c), while introduced a roll of -24°, reminiscent of what was previously observed for plant *de novo* DNA methyltransferase DRM2, in which residue R595 intercalates between the orphan guanine (G₀') and the +1-flanking nucleotide (A₊₁') on the non-target strand, leading to substantial DNA deformation (Supplementary Fig. 11d)⁵⁸. In addition, DIM2 T1164 and T1167 engage in hydrogen-bonding interactions with the backbone of Thy6 and Cyt8, respectively, DIM2 K1143 engages in electrostatic contacts with the backbone phosphates of Gua8' and Ade9', and DIM2 T1144 is involved in van der Waals contacts with the backbone of Gua7' (Fig. 5e).

Additional DNA-contacting sites of the MTase domain arise from residues S1097-Y1102 at the N-terminus of the TRD helix, which interact with the non-target strand via hydrogen-bonding and van der Waals contacts, and residues R1208-T1210 at the C-terminus of the MTase domain, which interact with the DNA backbone via salt-bridge and van der Waals interactions (Fig. 5a,e).

To test the observed DIM2-DNA interactions, we mutated key DNA-interacting residues on the RFTS, catalytic loop, bridging loop, and TRD of DIM2 and carried out in vitro DNA methylation analysis. Introducing Y201A mutation to the RFTS domain reduced

the DNA methylation efficiency of DIM2 by ~3-fold, introducing the R1039A and R1043A mutations to the bridging loop reduced the DNA methylation efficiency of DIM2 by ~3- and ~8-folds respectively, and introducing the Y1102A, R1145A, and D1173A mutations to the TRD reduced the DNA methylation activity of DIM2 by ~6-, ~8- and ~10-folds, respectively (Fig. 5f). More strikingly, introducing the mutation to the catalytic loop (S930A) or its subsequent helix (Q941A), and Y1100A, T1164A, T1166A/T1167A and R1175A mutations to the TRD each led to abolished DNA methylation activity of DIM2 (Fig. 5f). Together, these data lend a strong support for the observed DIM2-DNA interaction.

Structural basis for RFTS-H3K9me3 and BAH1-H3K9me3 interactions

The RFTS-H3K9me3 interaction is mediated by residues T6-R17 of histone H3K9me3 peptide, which runs through a surface groove formed by the N-terminal and C-terminal lobes of the RFTS domain (Fig. 4d,g). Of note, the side chain of residue H3K9me3 inserts into the aromatic cage at the N-terminal lobe, engaging hydrophobic contacts with DIM2 Y185, W261 and Y273 and an electrostatic contact with DIM2 E271 (Fig. 6a). The RFTS-H3K9me3 association is further supported by an array of backbone hydrogen-bonding interactions, involving DIM2 Y249 and H3 A7 and R8, DIM2 N248 and H3 R8, DIM2 V246 and H3 S10, DIM2 L244 and H3 G12, and a sidechain-main chain hydrogen bond between DIM2 E363 and H3 S10 (Fig. 6a). In addition, the RFTS-H3K9me3 association is supported by van der Waals contacts involving DIM2 V246, N248, Y249, W361, K362, and H3 T6-R17. Such an RFTS-H3K9me3 association in turn positions H3 K14 in proximity with DNA backbone of Cyt17 for a salt-bridge interaction (Fig. 6a).

The fact that the H3K9me3 peptide concurrently interacts with the DIM2 RFTS domain and the DNA substrate explains why the RFTS-H3K9me3 interaction promotes DIM2-mediated DNA methylation.

In addition to the RFTS-H3K9me3 interaction, a second H3K9me3 peptide is anchored to the surface groove of the BAH1 domain, lined by the allosteric loop on one side and the loop connecting β 3- and β 4-strands (l_{34}) of the BAH1 domain on the other side (Fig. 6b). The side chain of H3K9me3 inserts into the hydrophobic cage formed by DIM2 W571, W581, W609, and Y611 for hydrophobic and cation- π interactions, while engaging in an electrostatic interaction with neighboring D615 (Fig. 6b). Furthermore, the DIM2 BAH1 domain interacts with the H3K9me3 peptide via hydrogen-bonding interactions involving DIM2 H634 and H3 S10', DIM2 T559 and H3 R8', and DIM2 N566 and H3 Q5', an electrostatic interaction between DIM2 E649 and H3 R8' (Fig. 6b). In addition, a hydrogen-bonding interaction is formed between DIM2 N1050 of the bridging loop and H3 T11' (Fig. 6b). Structural comparison of the DIM2-HP1-H3K9me3-DNA quaternary complex with the DIM2-HP1 binary complex indicates that in the absence of the H3K9me3 peptide, the bridging loop in the DIM2-HP1 complex would be positioned in steric clash with the DNA substrate (Supplementary Fig. 10c); the H3K9me3 binding induces the structural ordering of the allosteric loop, which in turn moves the bridging loop to a position suitable for DNA binding (Supplementary Fig. 10b,c).

Cooperative regulation of DIM2 by the RFTS and BAH1 domains

The observations for (i) the interaction between DIM2 RFTS-bound H3K9me3 peptide and DNA and (ii) the H3K9me3 binding-induced repositioning of the bridging loop of

the DIM2 MTase domain provide an explanation for the H3K9me3-mediated enzymatic stimulation of DIM2. To test the notion, we mutated the aromatic residues constituting the H3K9me3 cage in the RFTS domain (W261) or BAH1 domain (W581) into alanine and performed ITC binding assays. First, fitting of the ITC binding curve for WT DIM2 with H3K9me3 peptide using one-site binding mode gave an apparent dissociation constant (K_d) of 17.4 μ M (Fig. 6c and Supplementary Fig. 12a). Next, introducing the W581A mutation to the BAH1 domain reduced the peptide-binding affinity of DIM2 by $\sim$ 10-fold (K_d of 182 μ M), to a level that is comparable with the binding affinity of the isolated DIM2 RFTS domain for the H3K9me3 peptide (Supplementary Fig. 12b,c), whereas introducing the W261A mutation to the RFTS domain led to a $\sim$ 2.3-fold enhanced binding affinity (K_d of 7.6 μ M) (Fig. 6c and Supplementary Fig. 12d). Finally, introducing the W261A/W581A double mutation to DIM2 led to complete abolishment of the DIM2-H3K9me3 interaction (Fig. 6c and Supplementary Fig. 12e). These data support the H3K9me3 bindings by the RFTS and BAH1 domains, and suggest that in the absence of DNA, the BAH1-H3K9me3 binding is $\sim$ 24-fold stronger than the RFTS-H3K9me3 binding. However, in the presence of DNA, the RFTS-H3K9me3 binding may be strengthened by the additional interaction between H3 K14 and DNA (Fig. 6a). Presumably owing to such concurrent yet distinct H3K9me3 bindings by the RFTS and BAH1 domains, we found it intractable to fit the ITC binding curve measured for WT DIM2 using two-site binding mode. In support of the role of the RFTS and BAH1 domains in the DIM2-DNA interaction, our electrophoretic mobility shift assay (EMSA)

revealed that both the W261A and W581A mutation modestly reduced the DNA-binding affinity of DIM2 (Supplementary Fig. 12f).

Consistent with the peptide and DNA binding assays, our *in vitro* DNA methylation analysis reveals that introducing the W261A and W581A mutation each led to 4-5-fold decrease in DNA methylation efficiency of DIM2, whereas introducing the W261A/W581A double mutation largely abolished the DNA methylation activity of DIM2 (Fig. 6d). In addition, mutation of some other H3K9me₃-binding sites in the RFTS domain (K362A), the BAH1 domain (E649A) or the bridging loop (N1050A) also led to marked decrease of the DNA methylation activity of DIM2 (Fig. 6d). Together, these data support the notion that separate H3K9me₃ bindings by the DIM2 RFTS and BAH1 domains cooperate in allosteric stimulation of DIM2-mediated DNA methylation.

Structural comparison of DIM2-DNA and DNMT1-DNA complexes

Although DIM2 and DNMT1 share a similar domain architecture, they possess distinct enzymatic specificities, with DIM2 methylating both CG and non-CG DNA (Supplementary Fig. 1f)²³ and DNMT1 strictly methylating CG sites^{13, 41}. To illustrate how such two evolutionarily linked DNA methyltransferases functionally diverge, we performed structural comparison of DIM2 and DNMT1 (Fig. 7a). DIM2 and DNMT1 shares a similar core structure, formed by the BAH1, BAH2 and MTase domains, with an RMSD of 4.4 Å over 497 aligned C α atoms (Fig. 7a), in line with the modest sequence identity between the two orthologues (Supplementary Fig. 13).

The most notable difference between DIM2 and DNMT1 lies in the RFTS domain, which shows drastically different domain conformations: the DIM2 RFTS domain associates

with the side of the MTase domain to support the DNA binding, whereas the DNMT1 RFTS domain occupies the front face of MTase domain to block DNA binding in DNMT1 (Fig. 7a-c), underpinned the distinct RFTS-MTase interdomain interactions between the two proteins (Fig. 7d,e). Accordingly, whereas the DIM2 and DNMT1 RFTS domains share a similar two-lobe fold, with an RMSD of 4.3 Å over 160 aligned C α atoms (Supplementary Fig. 13a-c), they possess distinct surface electrostatic potential for inter-domain interactions (Supplementary Fig. 14d,e). Nevertheless, the DIM2 and DNMT1 RFTS domains interact with the H3K9me3 peptide in a similar mechanism: both of the RFTS domains bind to the H3K9me3 peptide via the cleft between the N- and C-lobes of RFTS, lining the H3K9me3 mark by one (W463 in bovine DNMT1) or three (Y185, W261, and Y273 in DIM2) aromatic residues, reinforced by electrostatic contacts with acidic residues (Supplementary Fig. 14b, c). On the other hand, unlike the DNMT1 RFTS-H3K9me3 interaction that involves residues R2-K23 of H3 as well as H3 ubiquitylation, the DIM2 RFTS domain engages a shorter H3K9me3 peptide (residues T6-R17) (Supplementary Fig. 14a-c) and contains a different sequence for the region corresponding to the H3 ubiquitin-binding sites of DNMT1 RFTS (Supplementary Fig. 13).

Structural comparison of DNMT1 and DIM2 further reveals that, unlike the DIM2 TRD that undergoes an extensive disorder-to-order transition upon HP1 binding, the DNMT1 TRD is well defined even in the DNA-free state of DNMT1 (Supplementary Fig. 14f), providing an explanation for the fact that DNMT1 alone, but not DIM2 alone, shows an appreciable methyltransferase activity⁵⁹ and in line with the fact that DIM2 TRD and

DNMT1 TRD are not conserved in both sequence and structure (Fig. 7a and Supplementary Fig. 13). Along the line, DNMT1 and DIM2 employ a different set of residues for substrate binding (Fig. 7f-i). For instance, DNMT1 inserts three bulky residues (W1512, M1535 and K1537) to occupy the DNA cavity after base flipping of the target cytosine, moving the orphan guanine (G₀') toward the -1-flanking nucleotide on the non-target strand (Fig. 7h), which contributes to the flanking sequence-dependent DNA deformation at the -1 CG-flanking site as observed previously⁶⁰. On the other hand, the DIM2-DNA interaction results in DNA deformation in a different fashion: DIM2 R1175 intercalates between G₀' and -1-flanking nucleotide on the non-target strand, pushing G₀' toward the +1-flanking nucleotide (Fig. 7i), an opposite direction from that in the DNMT1-DNA complex (Fig. 7h). Furthermore, unlike the DNMT1-induced DNA deformation that is compensated by extensive base-specific hydrogen-bonding interactions involving the CG dyad, which gives rise to the strict CG specificity of DNMT1^{37, 41}, the DIM2-mediated DNA deformation is not accompanied by extensive base-specific interactions (Fig. 7h,i). In addition, DNMT1 TRD forms a hydrophobic concave anchoring the 5-methyl group of 5-methylcytosine on the template strand (Fig. 7f,h), underpinning its substrate preference for hemimethylated CG sites; in contrast, the corresponding region of DIM2 contains polar residues (N1170 and R1145) in contact with the +1-flanking site (G₊₁') on the non-target strand (Fig. 7g,i), which may relax the substrate selectivity of DIM2 for hemimethylated CG sites²³. How the DIM2-DNA interaction interplays with substrate specificity of DIM2, if any, awaits further investigation.

It is worth mentioning that the DIM2 RFTS-BAH1 linker contains a pair of α -helices (α A and α B denoted herein) positioned next to the catalytic helix, resembling the corresponding segments in DNMT1 (Supplementary Fig. 14g). Note that previous studies have demonstrated that the corresponding linker segments in DNMT1, through interaction with the catalytic helix, play a critical role in enzyme activation, allowing for a kinked-to-straight conformational adjustment of the catalytic helix of DNMT1 when transiting from an autoinhibitory state to the active state (Supplementary Fig. 14h)^{39, 40}. In contrast, the catalytic helix of DIM2 remains straight in both the DNA-free and -bound states (Supplementary Fig. 14i). In this regard, structural analysis of the DNMT1 reveals that the kinked conformation of the catalytic helix is in part stabilized by an inter-domain contact between residue E572 of the RFTS domain and residue R1238 of the catalytic helix (Supplementary Fig. 14j); in contrast, no corresponding interaction is observed for the apo-DIM2 due to a DNMT1-distinct positioning of the RFTS domain, providing an explanation that why the catalytic helix of DNMT1, but not DIM2, undergoes a conformation readjustment between the apo and DNA-binding states^{36, 37, 43, 60, 61}.

Another structural distinction between DIM2 and DNMT1 lies in the BAH1 domain (Supplementary Fig. 15a-c). Structural analysis of the DIM2 BAH1-H3K9me3 association reveals a binding mode similar to that of the previously reported interaction between the BAH domain of maize CHG DNA methyltransferase ZMET2 and the H3K9me2 peptide (Supplementary Fig. 15d, e), rather than the DNMT1 BAH1-H4K20me3 binding (Supplementary Fig. 15b). In fact, structural alignment of the DIM2 BAH1-H3K9me3 and ZMET2 BAH-H3K9me2 complex gives an RMSD of 2.2 Å over

103 aligned C α atoms (Supplementary Fig. 15e); in contrast, structural alignment of the DIM2 BAH1-H3K9me3 and DNMT1 BAH1-H4K20me3 complex gives an RMSD of 2.4 Å over 79 aligned C α atoms, with the histone peptides anchored in an opposite polarity (Supplementary Fig. 15c). Together, these data reveal an interplay of DIM2 with DNA and repressive histone marks that is distinct from that of DNMT1.

Discussion

DNA methylation, H3K9me3, and HP1 protein together serve as the hallmarks of heterochromatin. How they crosstalk to underpin the heterochromatic assembly and genomic stability remains a fundamental question. This study characterized the structure and regulation of DIM2, the sole DNA methyltransferase in the vegetative tissue of *Neurospora* that specifically mediates DNA methylation at constitutive heterochromatin²⁵. Importantly, through structure delineation of apo-DIM2, DIM2-HP1 complex, and DIM2-HP1-H3K9me3-DNA complex, combined with biochemical and enzymatic analysis, our work provides an unprecedented view on the functional interplay between heterochromatin factors, in which multivalent readouts of H3K9me3 and HP1 act as a switch in controlling the DNA methylation activity of *Neurospora* DIM2, with important implication in understanding heterochromatin-directed DNA methylation establishment and maintenance.

First, this study reveals a mechanism by which DIM2-mediated DNA methylation is strictly controlled by heterochromatin factors. Unlike most other DNA methyltransferases characterized so far that, in the absence of regulatory proteins, exhibit a basal DNA methylation activity, DIM2 alone possesses no appreciable DNA methylation activity;

rather, it requires the presence of both H3K9me3 and HP1 proteins for a substantial DNA methylation activity. Such a tight control of DIM2-mediated DNA methylation is underpinned by three parallel intermolecular interactions: DIM2-HP1, DIM2 RFTS-H3K9me3, and DIM2 BAH1-H3K9me3. First, the DIM2-HP1 interaction leads to two consequences that both contribute to DIM2-mediated DNA methylation in heterochromatin: on one hand, the interaction of HP1 with both the I130-V132-L134 motif-residing NT segment and the TRD of DIM2 leads to a disorder-to-order transition of the TRD, thereby controlling the TRD-mediated DNA interaction; on the other hand, the simultaneous association of HP1 with DIM2 and H3K9me3 helps recruit DIM2 to H3K9me3-marked heterochromatin, thereby contributing to heterochromatin-specific DNA methylation⁶². Second, the interaction of the DIM2 RFTS domain with H3K9me3 helps position the latter for a DNA interaction, thereby supporting the substrate association of DIM2. Third, the interaction of the DIM2 BAH1 domain with H3K9me3 leads to repositioning of the bridging loop of DIM2, which further facilitates the DIM2-substrate association. These multiple epigenetic regulations provide a mechanism allowing heterochromatin factors H3K9me3 and HP1 to tightly control DIM2-mediated DNA methylation, which in turn reinforces genomic stability at the constitutive heterochromatin.

Second, this study reveals an intricate interplay between HP1 and DNA methylation. As a well-characterized reader for the methylated H3K9 (H3K9me)⁴⁻⁶, HP1 knowingly employs its homodimeric assembly to promote phase-separated condensates, thereby contributing to chromatin compaction⁶³⁻⁶⁶. In addition, HP1 proteins reportedly bind to

the DNA methyltransferases (DNMTs) across various species, such as plant CMT3⁶⁷, and mammalian DNMT1^{68, 69}, DNMT3A⁶⁸ and DNMT3B⁷⁰. Whereas these DNMT-HP1 interactions conceivably contribute to the recruitment of respective DNMTs to H3K9me3-enriched heterochromatic regions⁷¹, the molecular basis for the HP1-DNMT axis remains vague. In this regard, this study reveals that the HP1 binding induced an extensive disorder-to-order transition of the DIM2 TRD, thereby coupling the HP1 interaction with DIM2-mediated DNA methylation in heterochromatin. Such an HP1-mediated regulation lends an explanation to a previous observation that disruption of the HP1 interaction of DIM2 led to a loss of DNA methylation in *Neurospora*³², providing a mechanism for heterochromatin-specific DNA methylation by DIM2. Interestingly, the HP1 binding also appears to stimulate the DNA methylation activity of DNMT1⁶⁹. Whether the HP1-mediated regulation of DIM2 is shared with DNMT1 and other DNMTs awaits further investigation.

Third, this study adds another example on the functional crosstalk between DNA methylation and repressive histone modifications, which represents a conserved epigenetic mechanism in reinforcing heterochromatin assembly and gene silencing²⁶. Direct recognition of H3K9me3 and H4K20me3 has been observed for the RFTS and BAH1 domains of DNMT1, respectively, leading to allosteric activation of its DNA methylation activity^{29, 35}. For CMT3, it employs both the Chromodomain and BAH domain to interact with H3K9me2^{27, 28}, with the Chromodomain-H3K9me2 binding mediating the chromatin targeting and the BAH-H3K9me2 interaction leading to allosteric activation^{28, 30}. Our study reveals that both the RFTS and BAH1 domains of

DIM2 recognize H3K9me3, which in turn contribute to the association between DIM2 and DNA substrates through direct DNA contact (for RFTS-bound H3K9me3 peptide) or conformational adjustment of a DNA-binding site (for BAH1-bound H3K9me3 peptide). It remains to be investigated that whether the two DIM2-bound H3K9me3 marks arise from a di-nucleosome, as that demonstrated for CMT3³⁰ (Fig. 8), or a mono-nucleosome. Nevertheless, these observations add an example of the functional crosstalk between DNA methylation and H3K9me3.

Fourth, this study uncovers how closely related DNMT1 and DIM2, with a common scaffold, have evolved into distinct functional specificities. For DNMT1, its RFTS domain interacts with the front face of the MTase domain to inhibit its DNA methylation activity³⁸⁻⁴⁰; the association of H3 two-monoubiquitylation/H3K9me3 with the RFTS domains promotes the repositioning of the latter, thereby relieving the autoinhibitory state^{29, 42, 43}. In contrast, the DIM2 RFTS domain is positioned on the side face of the MTase domain, allowing the DNA binding interface of the latter fully accessible to DNA substrates. The association of H3K9me3 with the RFTS domain did not lead to domain repositioning of the latter, rather, it anchors the H3K9me3 peptide for a direct interaction with the DNA substrates, thereby supporting the DIM2-DNA association. Furthermore, unlike the DNMT1 BAH1 domain that specifically recognizes H4K20me3, this study reveals that the DIM2 BAH1 domain interacts with H3K9me3 peptide in a manner resembling that of ZMET2 BAH domain-H3K9me2 interaction. The DIM2 BAH1-H3K9me3 interaction allosterically activates DIM2 via conformational adjustment of a

DNA-binding loop of the MTase domain, which is again reminiscent of ZMET2 BAH-H3K9me2 interaction²⁸.

In addition, this study reveals that DIM2 interacts with the DNA substrates in a manner different from that of the DNMT1-DNA interaction. The DNMT1-DNA interaction involves a network of base-specific hydrogen-bonding interactions involving the CG site, as well as a hydrophobic concave for anchoring the 5-methyl group of 5-methylcytosine on the hemimethylated CG site^{37, 60}. These base-specific contact sites are not present in DIM2, due to lack of sequence similarity between the TRDs of DIM2 and DNMT1. On the other hand, the DIM2-DNA interaction involves strong DNA intercalation induced by an arginine residue (R1175), resulting in base unstacking between the orphan guanine and the -1-flanking nucleotide. A similar intercalation-induced DNA deformation was previously observed for plant DNA methyltransferase DRM2, which gives rise to a substrate preference of DRM2 for a CHH site containing an A/T nucleotide at the +1 position⁵⁸. Whether the DIM2 R1175-induced DNA deformation confers DIM2, which also targets genomic regions enriched with A/T nucleotides a flanking sequence preference awaits further investigation. Together, these observations provide insight into the evolutionary conservation and divergence among DNMT1 family of DNA methyltransferases.

In summary, this study provides insight into an intricate interplay between heterochromatic factors in regulating DNA methylation, and how DIM2, with a DNMT1-like scaffold, has evolved to orchestrate the unique DNA methylation landscape in *Neurospora*.

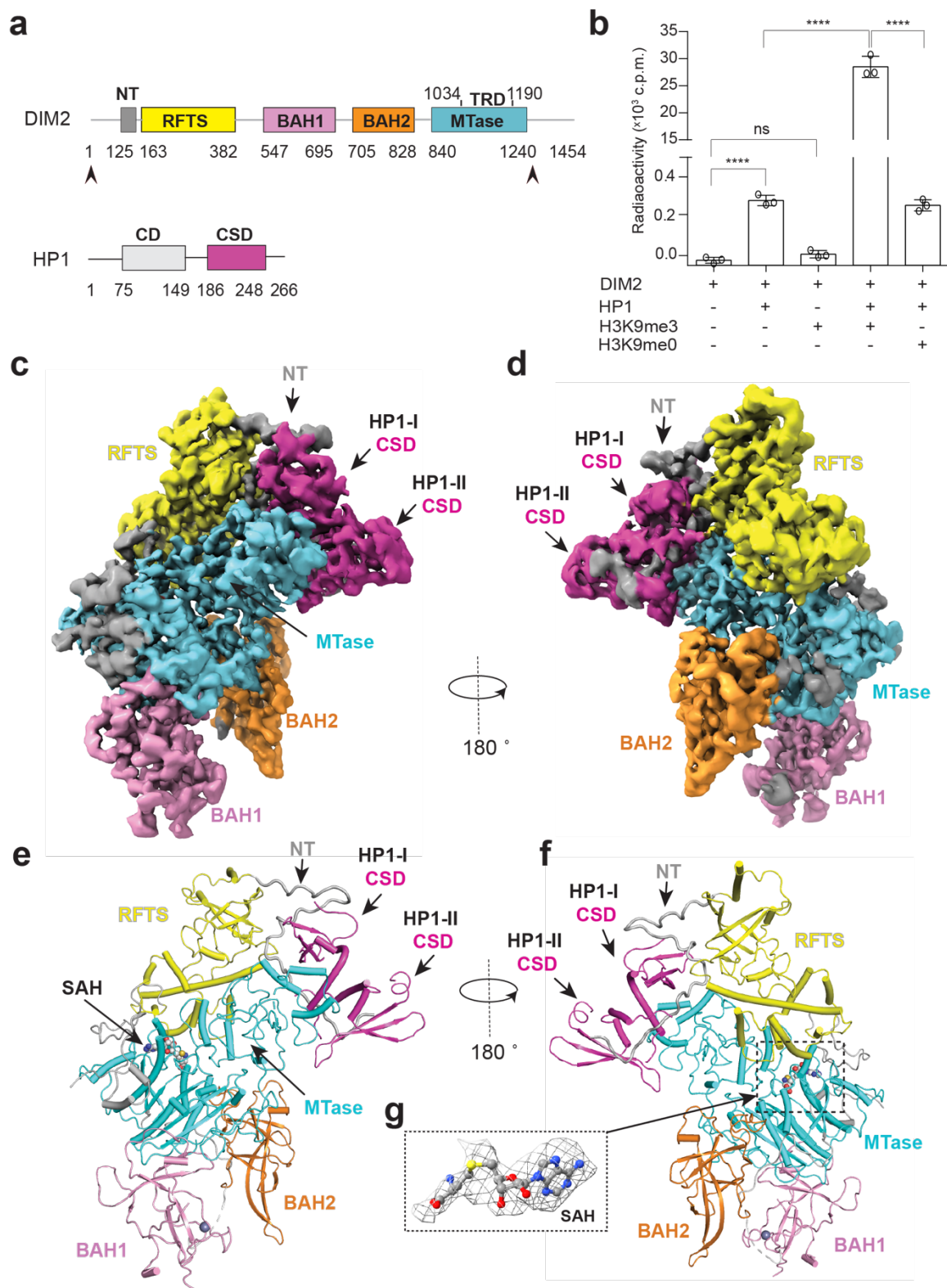


Figure 1. Cryo-EM structure of the DIM2-HP1 complex. (a) Domain architecture of DIM2 and HP1, with individual domains color coded and delimited by amino acid numbers. The DIM2 fragment used for structural and biochemical characterization is indicated by arrows. (b) In vitro DNA methylation assay of DIM2 on a (CTA)₁₂/(TAG)₁₂ 36-mer DNA duplex, in the presence or absence of HP1 and/or histone peptides (H3₁₋₂₄K9me3 or H3₁₋₂₄K9me0). Data are mean $\pm$ s.d. (n = 3 biological replicates). The two-tailed Student's t test statistical analysis was performed to compare DIM2 activity in the absence vs. presence of HP1 and/or histone peptides. ns, not significant; ****p < 0.0001. Source data are provided as a Source Data file. (c,d) Two opposite views of the cryo-EM density map of the DIM2-HP1 complex. (e,f) Atomic model of the DIM2-HP1 complex in two opposite views, with individual domains of DIM2 color coded. The two HP1 CSD subunits are colored in magenta. The NT and domain linkers are colored grey. The SAH molecule and zinc ion are shown in sphere representation. The color scheme for the individual domains of DIM2 is applied to subsequent figures unless otherwise indicated. (g) Cryo-EM density map for the SAH molecule bound to DIM2.

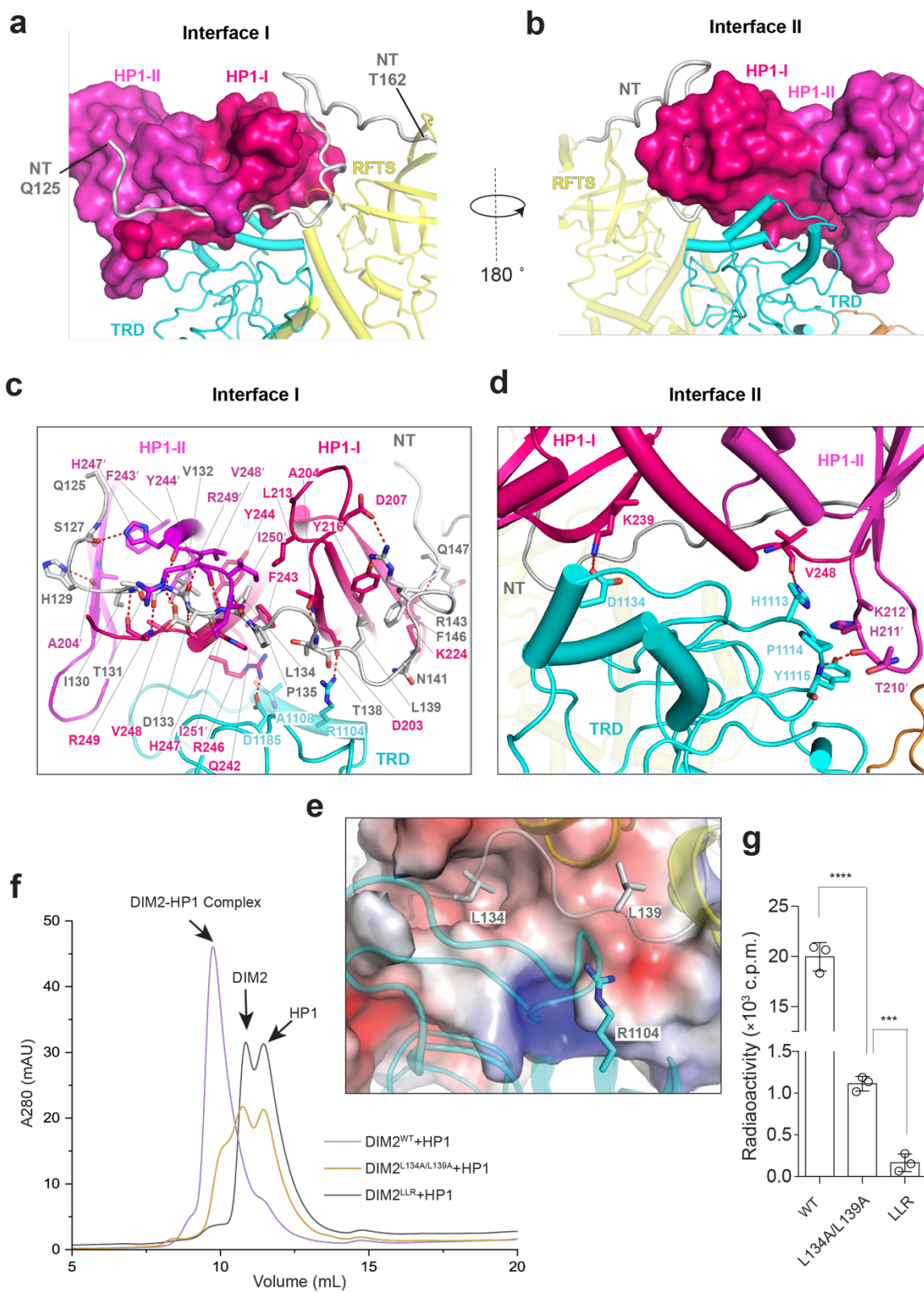


Figure 2. Structural basis for the DIM2-HP1 interaction. (a, b) Close-up view of the interaction between DIM2 and HP1 homodimer, with the two HP1 CSD subunits shown in surface view (HP1-I: hot pink, HP1-II: magenta) and DIM2 shown in ribbon representation. (c) Close-up view of the interaction between HP1 and DIM2 at interface I. (d) Close-up view of the interaction between HP1 and DIM2 at interface II. (e) Close-up view of key HP1-contacting residues of DIM2, with HP1 shown in electrostatic surface representation. (f) Size-exclusion chromatography analysis of DIM2, WT or mutant, mixed with HP1. (g) In vitro DNA methylation assay of DIM2, WT or mutant, on a (CTA)₁₂/(TAG)₁₂ DNA duplex, in the presence of HP1 and histone H3₁₋₂₄K9me3 peptide. LLR, DIM2 L134A/L139A/R1104A triple mutation. Data are mean $\pm$ s.d. (n = 3 biological replicates). Statistical analysis used two-tailed Student's t test. ***p = 0.0003; ****p < 0.0001. Source data are provided as a Source Data file.

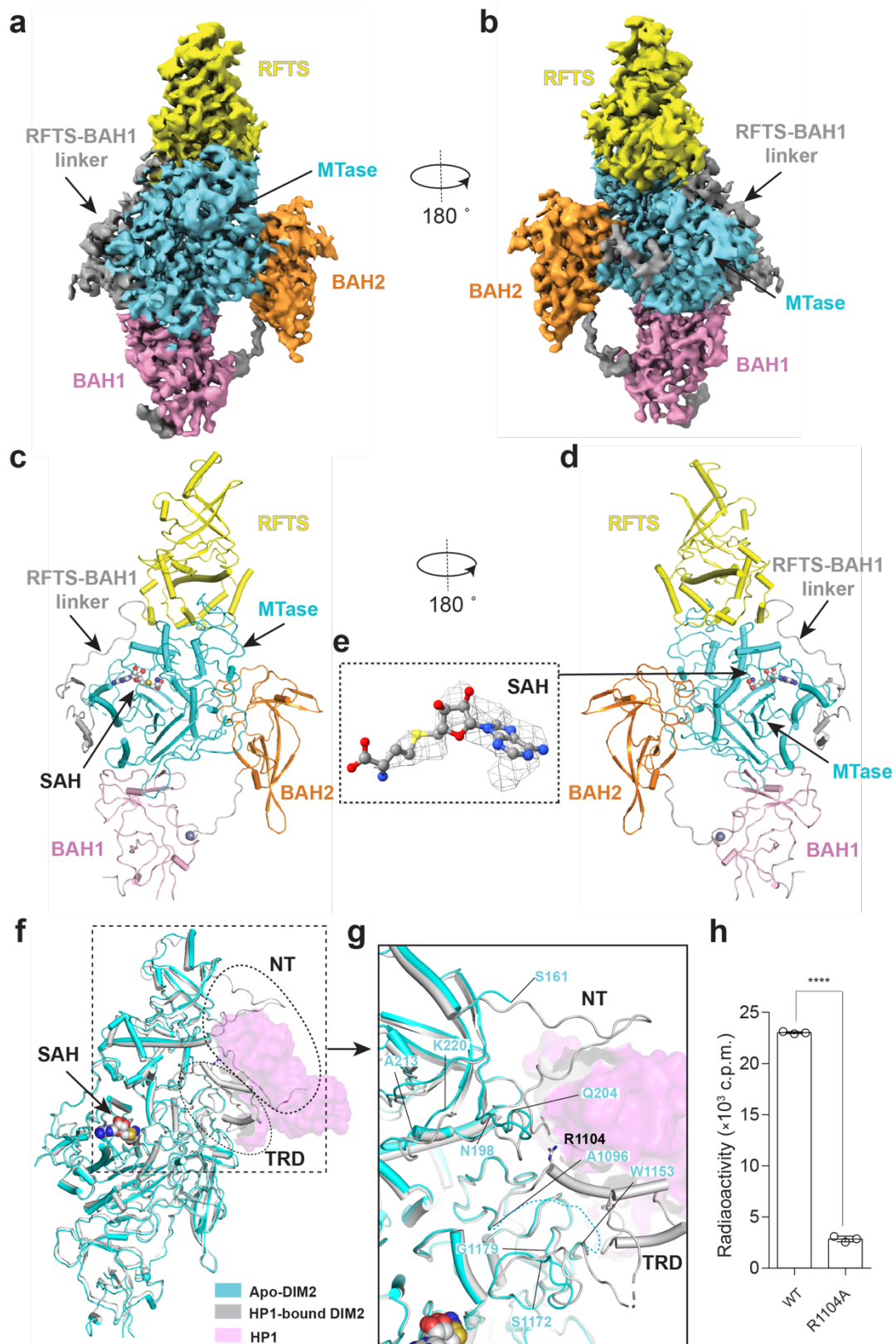


Figure 3. Cryo-EM structure of apo-DIM2. (a, b) Cryo-EM density map for apo-DIM2 in two opposite views. (c, d) Atomic model of apo-DIM2 in two opposite views, with individual domains color coded. The SAH molecule and zinc ion are shown in sphere representation. (e) Cryo-EM density map for the SAH molecule bound to DIM2. (f) Structural alignment between HP1-bound DIM2 (grey) and apo-DIM2 (cyan). The NT and TRD segments that undergo a disorder-to-order transition upon HP1 binding are indicated by dotted circles. The SAH molecules and zinc ions are shown in sphere representation. (g) Close-up view of the NT and TRD, with regions that undergo HP1 binding-induced conformational changes are delimited by residue numbers. (h) In vitro DNA methylation assay of DIM2, WT or mutant, on a (CTA)₁₂/(TAG)₁₂ DNA duplex, in the presence of HP1 and histone H3₁₋₂₄K9me3 peptide. Data are mean $\pm$ s.d. (n = 3 biological replicates). Statistical analysis used two-tailed Student's t test. ****p < 0.0001. Source data are provided as a Source Data file.

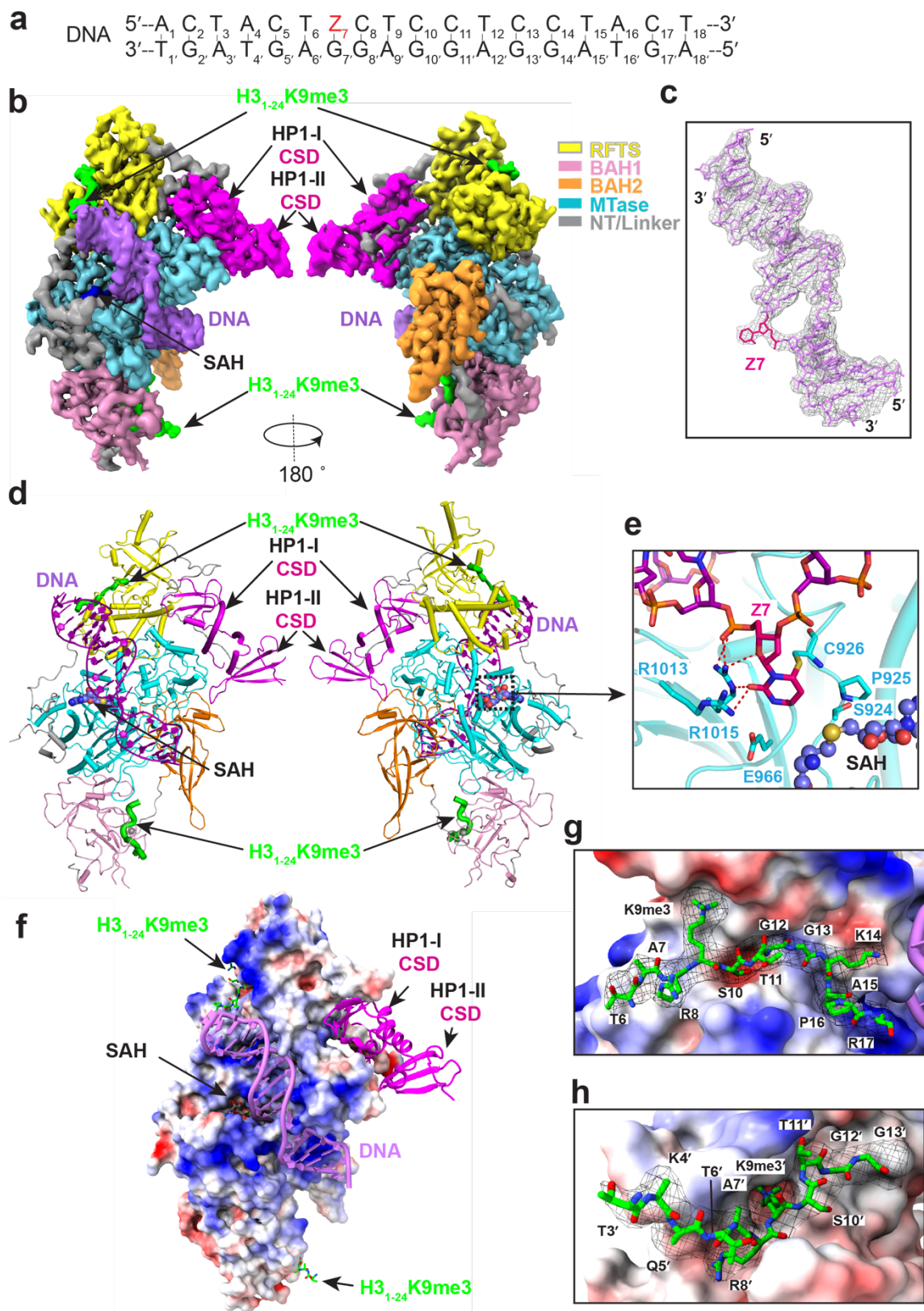


Figure 4. Cryo-EM structure of the DIM2-HP1-H3K9me3-DNA complex. (a) DNA sequence used for structural study. (b) Cryo-EM density map for the DIM2-HP1-H3K9me3-DNA complex in two opposite views. The individual domains of DIM2 are color coded. HP1, H3K9me3, and DNA are colored in magenta, limegreen, and purple, respectively. (c) Atomic model of the DIM2-bound DNA and the corresponding density map (grey). The flipped-out zebularine Z7 is colored hot pink. (d) Atomic model of the DIM2-HP1-H3K9me3-DNA complex, with individual components colored in the same fashion as in (b). The SAH molecule and zinc ion are shown in sphere representation. (e) Close-up view of the flipped-out Z7 bound to DIM2 residues. The hydrogen bonds are depicted as red dashed lines. (f) Electrostatic surface view of DIM2 bound to HP1, H3K9me3 and DNA. (g, h) Close-up view of H3K9me3 peptide bound to DIM2 RFTS (g) or BAH1 domain (h), with the atomic model of the histone peptide covered in the corresponding density map in black mesh and DIM2 shown in electrostatic surface representation.

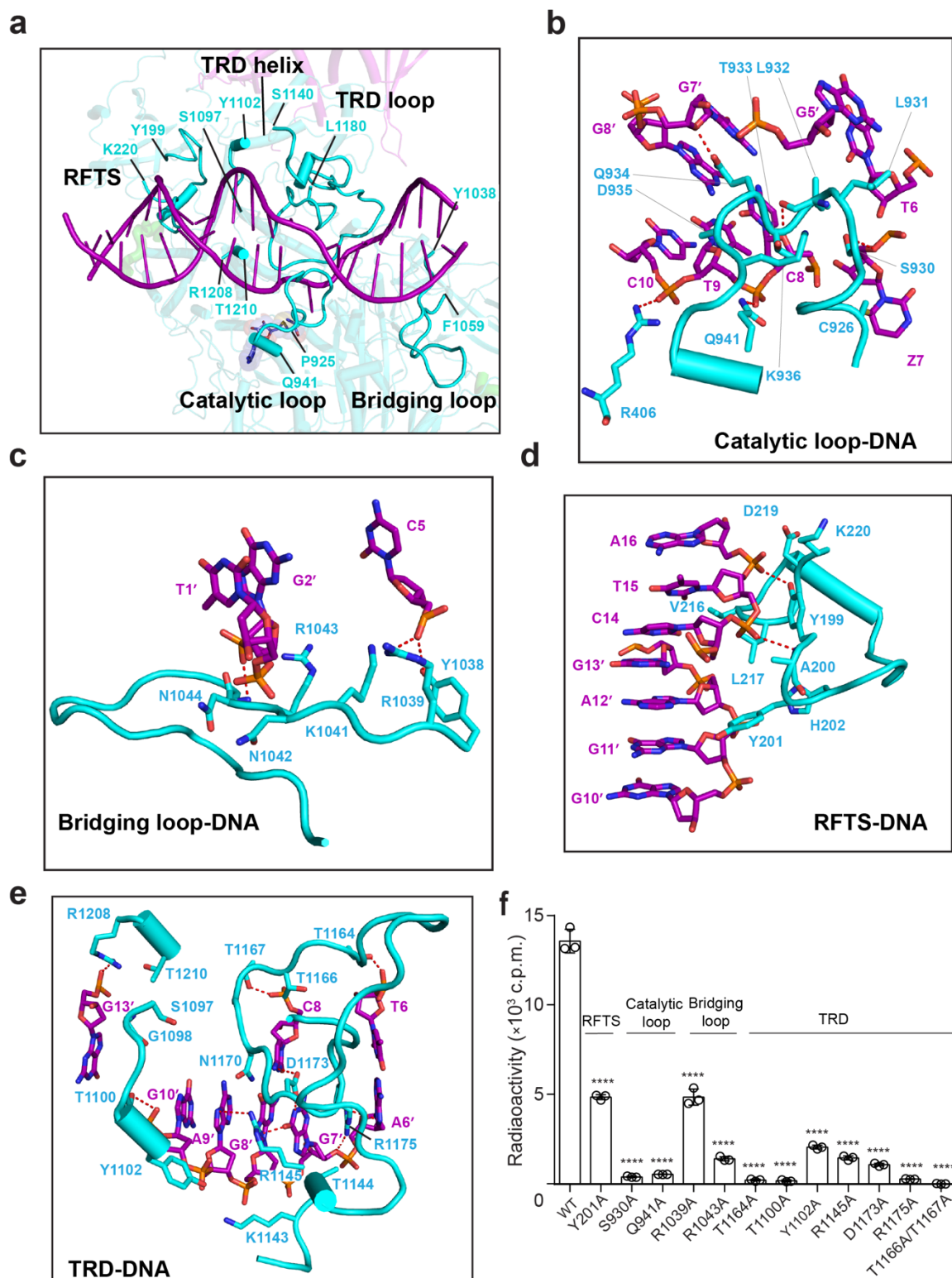


Figure 5. Structural basis of the DIM2-DNA interaction. (a) The overall view of intermolecular interactions between DIM2 (cyan) and DNA (purple), with individual DNA-binding regions of DIM2 delimited by residue numbers. (b-e) Close-up view of the DNA interactions by the catalytic loop (b), bridging loop (c), RFTS (d), and TRD (e). The hydrogen bonds are depicted as red dashed lines. (f) In vitro DNA methylation assay of DIM2, WT or DNA binding-associated mutant, on a (CTA)₁₂/(TAG)₁₂ DNA duplex. The corresponding regions of DIM2 are indicated on top. Data are mean $\pm$ s.d. (n = 3 biological replicates). Statistical analysis used two-tailed Student's t test. ****p < 0.0001. Source data are provided as a Source Data file.

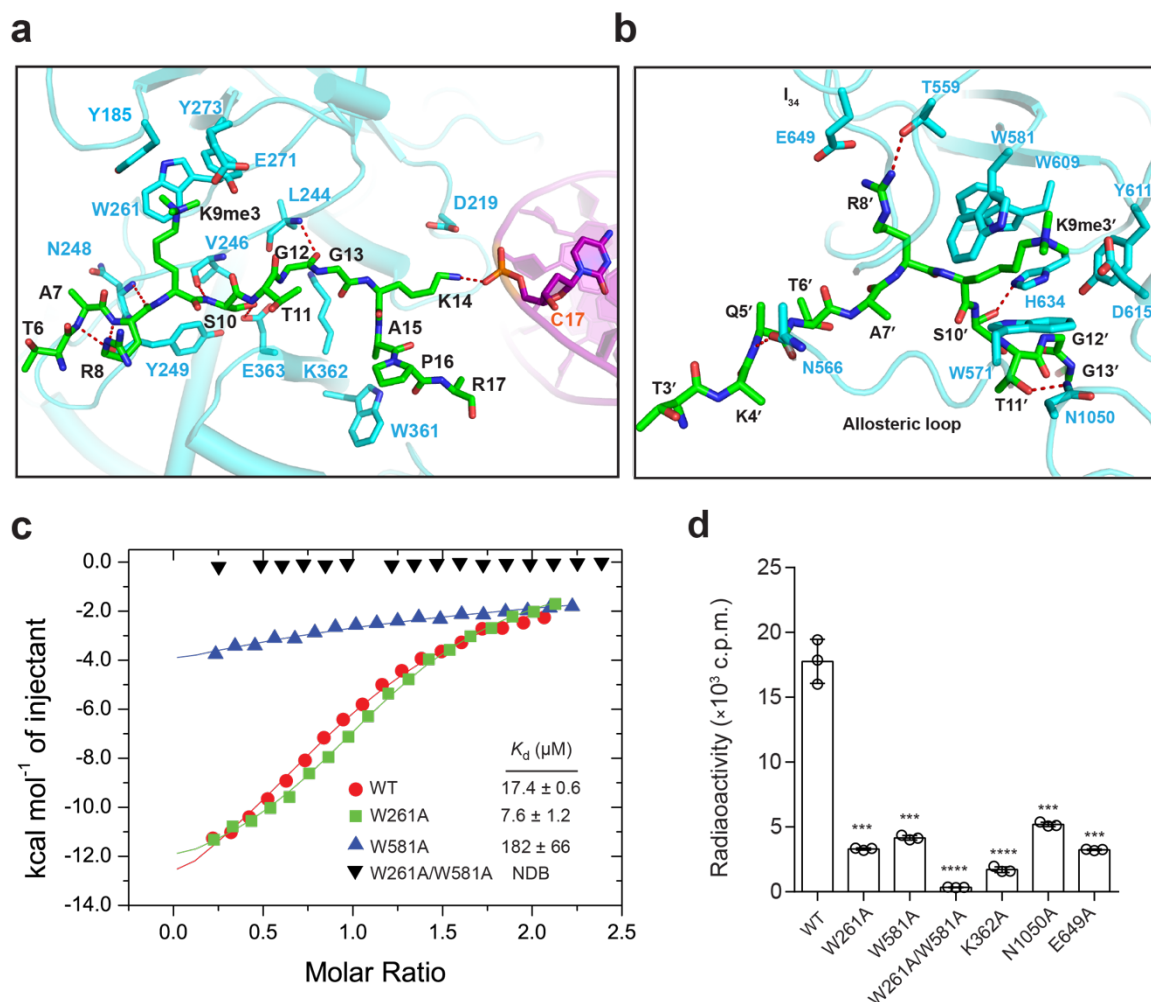


Figure 6. Structural basis for H3K9me3-mediated DIM2 activation. (a) Close-up view of the interaction of the H3₁₋₂₄K9me3 peptide with the DIM2 RFTS domain and DNA. The hydrogen bonds are depicted as dashed lines. (b) Close-up view of the interaction between the H3₁₋₂₄K9me3 peptide and the DIM2 BAH1 domain. (c) ITC binding curves for DIM2, WT or mutant, with the H3₁₋₂₄K9me3 peptide. The average and error estimates of the dissociation constants (K_d) were derived from two independent measurements. NDB, no detectable binding. (d) In vitro DNA methylation assay of DIM2, WT or H3₁₋₂₄K9me3 binding-associated mutant, on a (CTA)₁₂/(TAG)₁₂ DNA duplex. Data are mean $\pm$ s.d. (n = 3 biological replicates). Statistical analysis used two-tailed Student's t test. ***p = 0.0003; ****p < 0.0001. Source data are provided as a Source Data file.

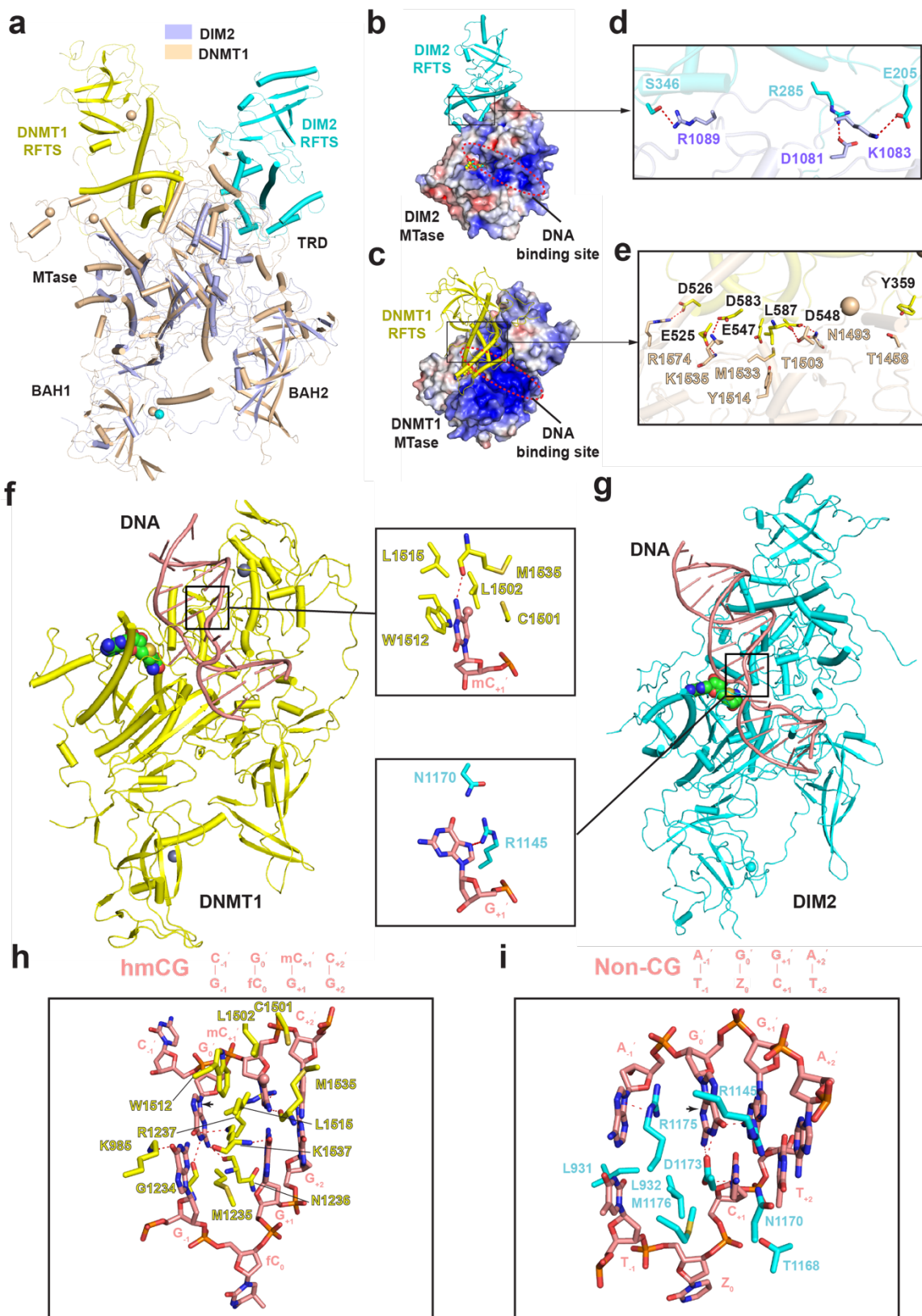


Figure 7. Structural comparison of DIM2 and DNMT1. (a) Structural overlay between the apo forms of DIM2 and DNMT1 (residues 351-1600) (PDB 4WXX). The RFTS domains of DIM2 and DNMT1 are colored in cyan and yellow, respectively. The rest of DIM2 and DNMT1 are colored in slate and wheat, respectively. (b,c) Positioning of the DIM2 (b) or DNMT1 (c) RFTS domain (ribbon representation) on top of the MTase domain (electrostatic surface representation). The DNA-binding site of the MTase domain of each protein is marked by a dotted circle. (d, e) Close-up view of the RFTS-MTase interactions within DIM2 (d) or DNMT1 (e). The hydrogen bonds are depicted as red dashed lines. (f) Atomic model of DNMT1 (yellow)-DNA (salmon) complex (PDB 4DA4), with the residues surrounding the 5-methylcytosine (mC_{+1}') on the template strand shown in the expanded view. The 5-methyl group of mC_{+1}' , zinc ions and SAH molecule are shown in sphere representation. Hydrogen bond is depicted as a dashed line. (g) Atomic model of DIM2 (cyan)-DNA (salmon) complex, with the residues surrounding the +1-flanking nucleotide (G_{+1}') on the complimentary strand shown in the expanded view. (h,i) Close-up view of the DNA interactions by DNMT1 (h) or DIM2 (i) MTase domains, centered on at the target site, as well as the -1 – +2-flanking nucleotides. The black arrows indicate the positional shift of Orphan G_0' from an ideal B-form DNA. The hydrogen bonding interactions are shown as dashed lines.

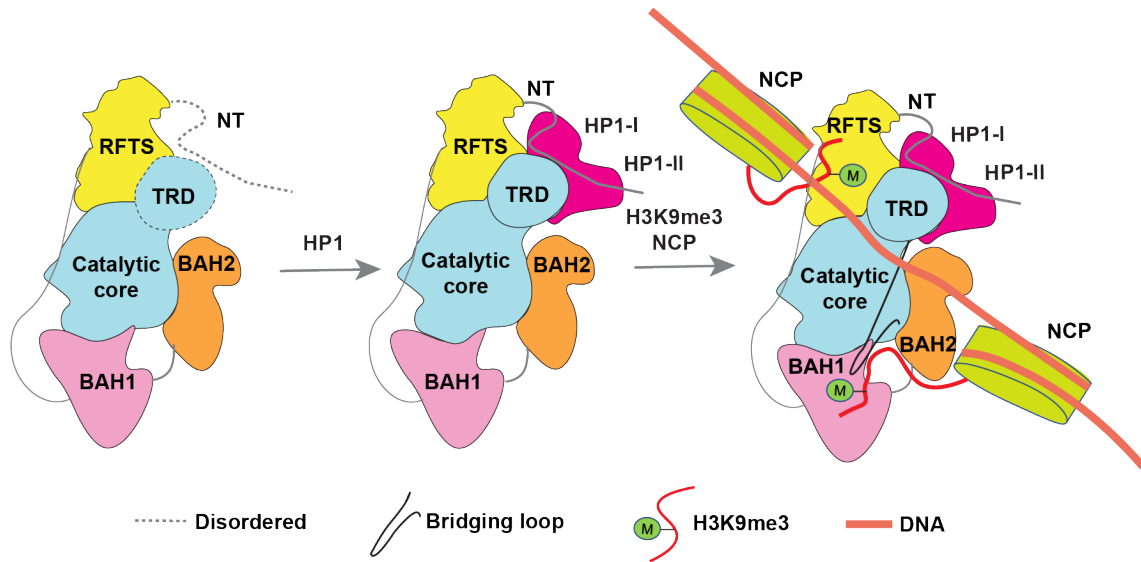
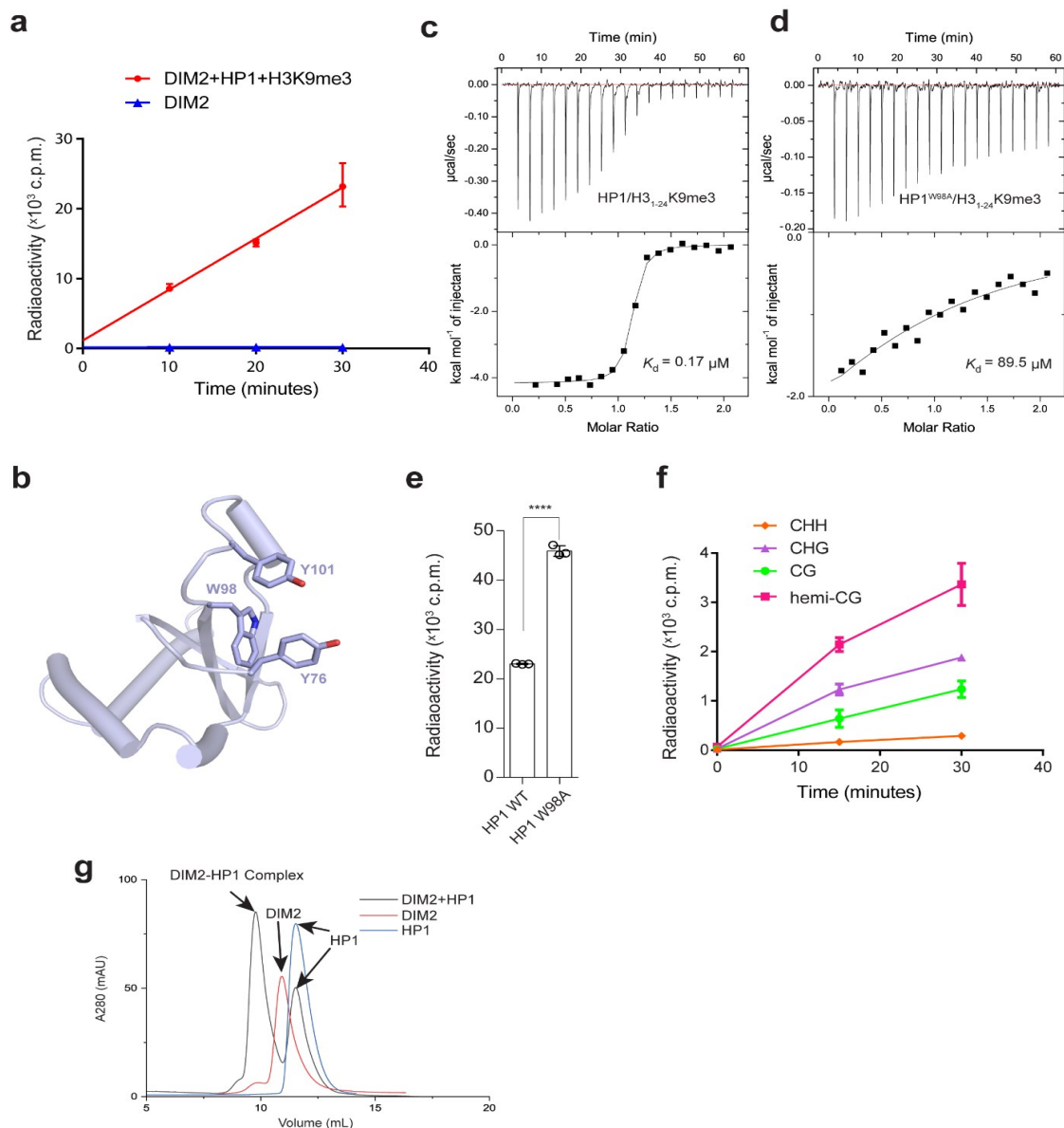


Figure 8. Multi-layered regulation of DIM2-mediated DNA methylation. A working model for the multi-layered regulation of DNA methylation by DIM2. The binding of HP1 induces a disorder-to-order transition of the DIM2 TRD. The dual readout of histone H3K9me3 by the DIM2 RFTS and BAH1 domains allosterically activates DIM2-mediated DNA methylation within the nucleosome environment. NCP, nucleosome core particle. It remains to be investigated whether the two H3K9me3 marks arise from two adjacent nucleosomes or from the same nucleosome.

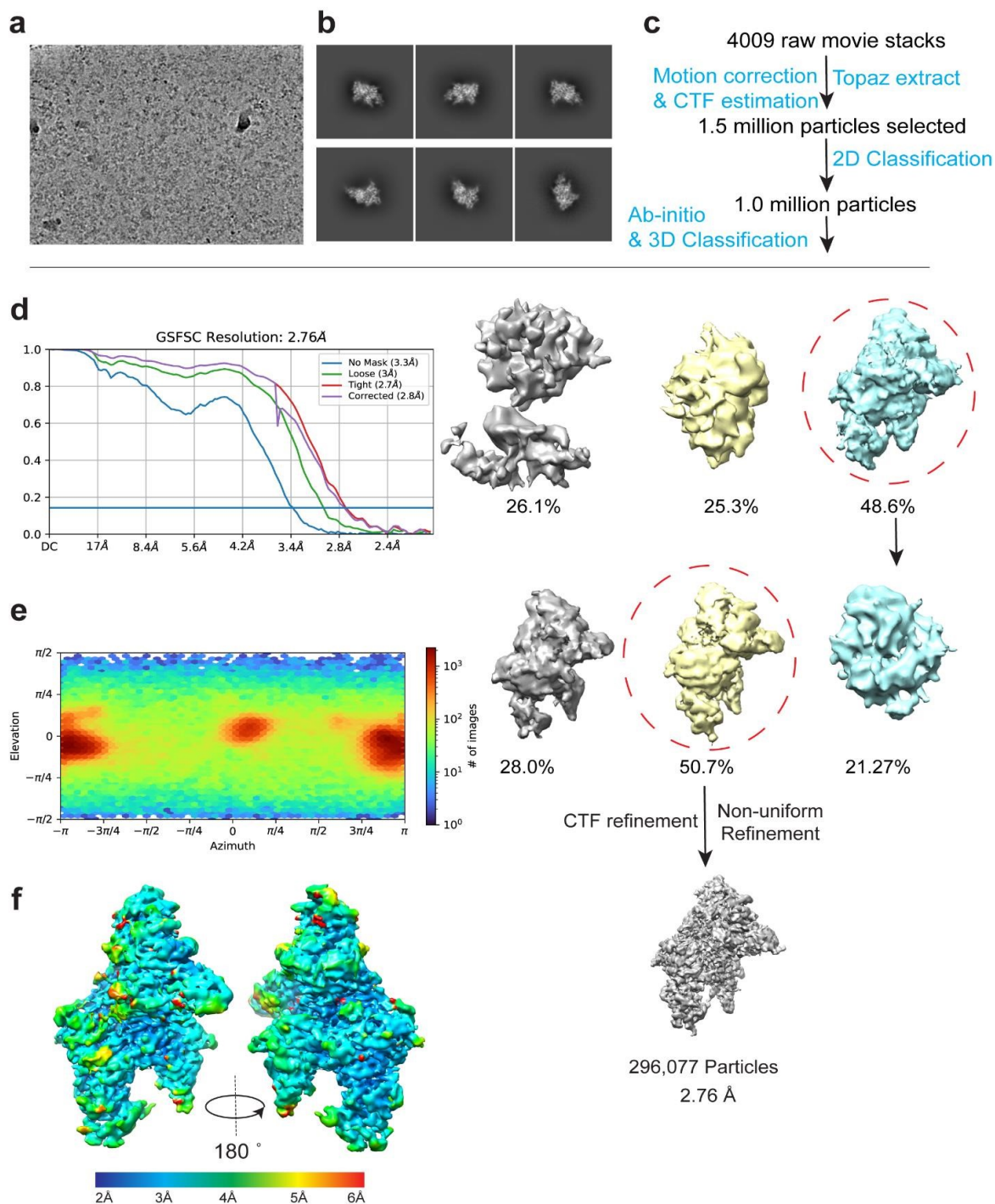
Supplementary Data



Supplementary Fig. 1 | Enzymatic and biochemical analysis of DIM2.

(a) In vitro DNA methylation assay of DIM2 on a (CTA)₁₂/(TAG)₁₂ DNA, in the presence or absence of HP1 and/or histone H3₁₋₂₄K9me3 peptide. Source data are provided as a Source Data file. (b) AlphaFold model of the HP1 chromodomain in *Neurospora Crassa*. The potential protein residues forming H3K9me3-binding pocket are shown in stick representation. (c, d) Representative ITC binding curves for WT (c) or W98A-mutated (d) HP1 with the H3K9me3 peptide. The top panels show the raw thermogram data. The lower panel shows the integrated binding isotherms as a function of the molar ratio of the ligand to the enzyme. (e) In vitro DNA

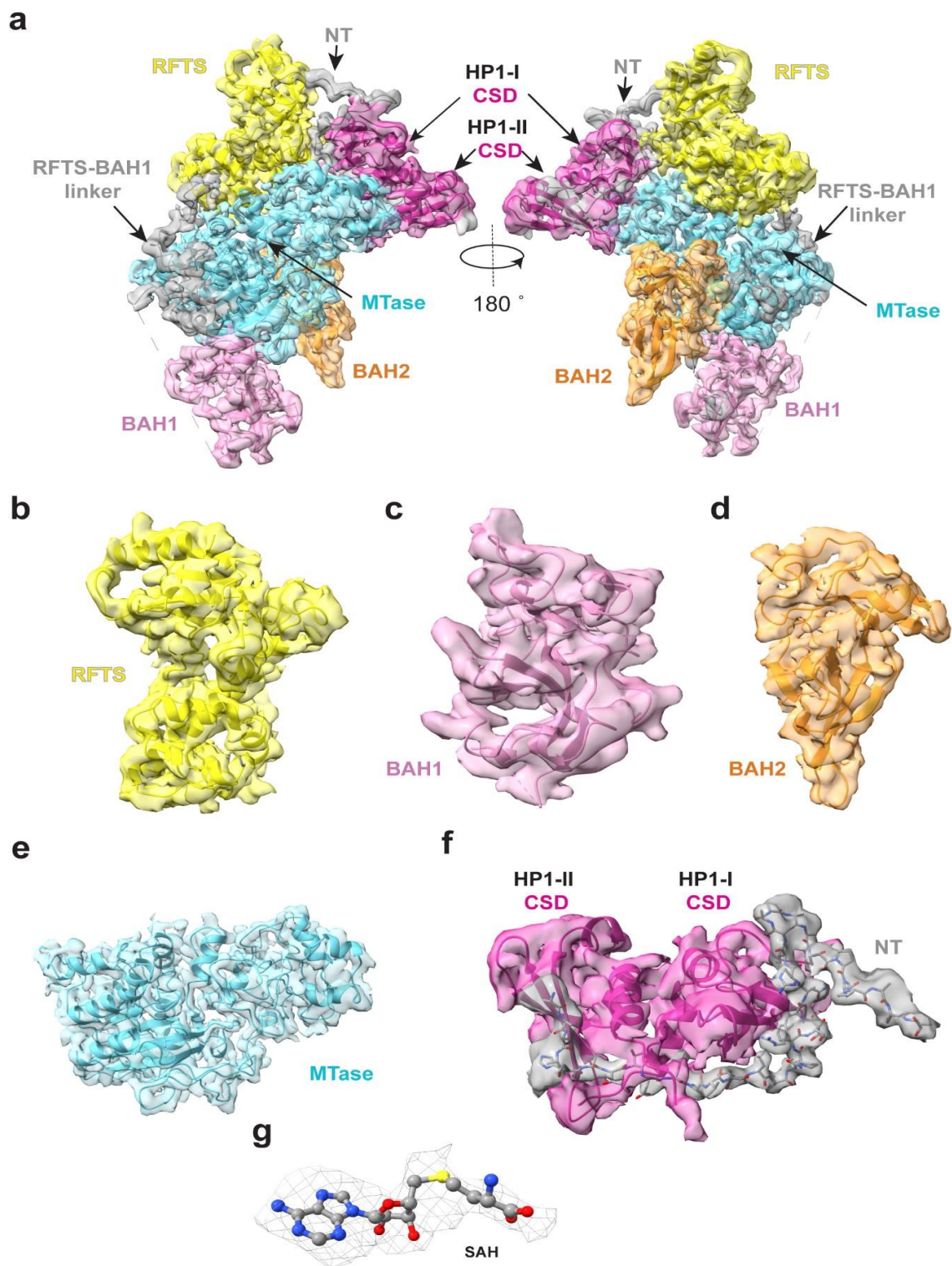
methylation assay of HP1, WT or W98A mutant, on a (CTA)₁₂/(TAG)₁₂ DNA duplex. Data are mean $\pm$ s.d. (n = 3 biological replicates). Statistical analysis used two-tailed Student's t test. ****p < 0.0001. Source data are provided as a Source Data file. (f) In vitro DNA methylation kinetics of DIM2 on DNA substrates containing one central CG, hemimethylated CG (hemi-CG), CHG or CHH target site. Source data are provided as a Source Data file. (g) Size-exclusion chromatography analysis of DIM2, HP1 and DIM2-HP1 complex.



Supplementary Fig. 2 | Data processing workflow for cryo-EM reconstruction of DIM2-HP1 complex.

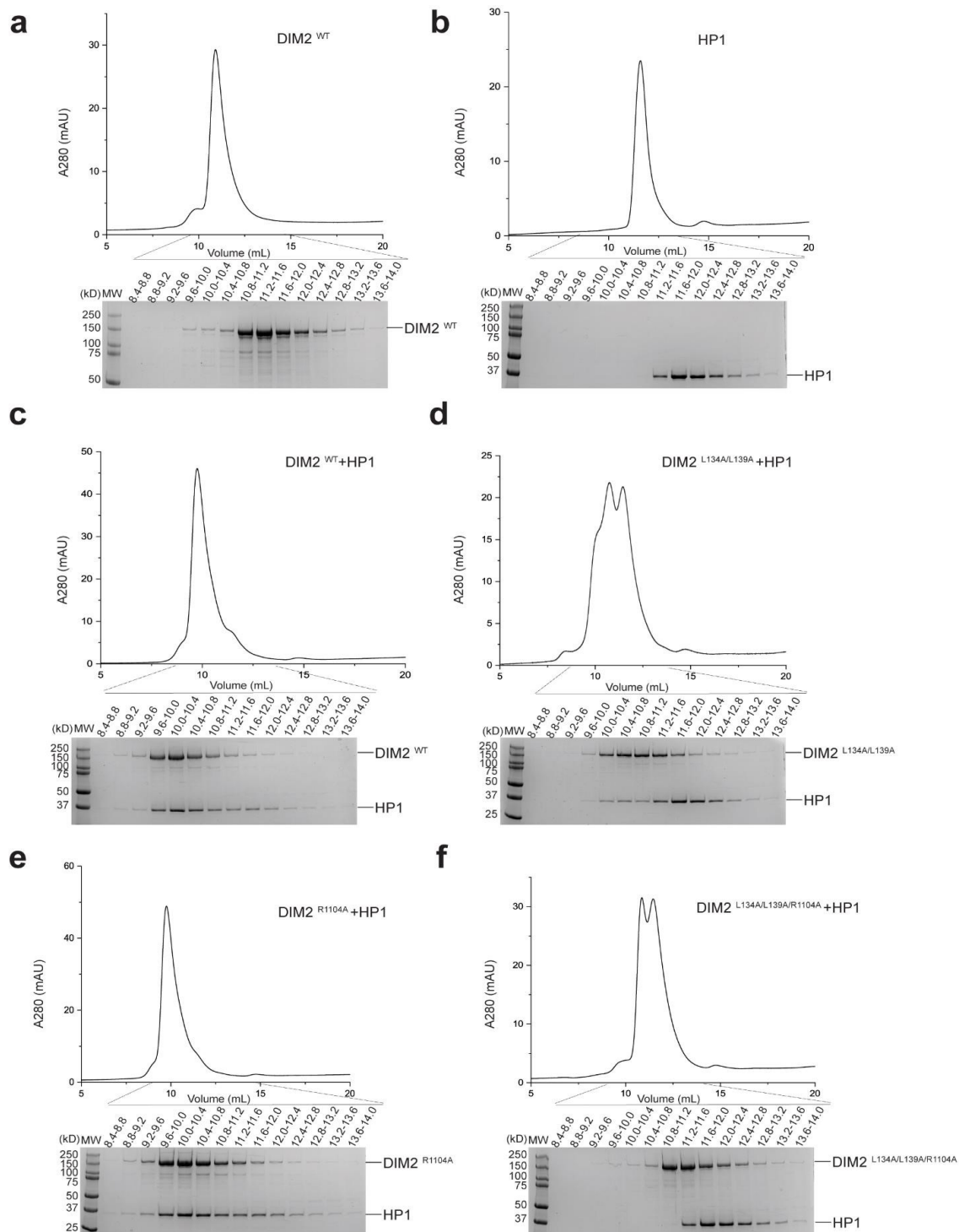
(a) A representative micrograph of the DIM2-HP1 complex by cryo-EM. (b) Representative 2D classes of the DIM2-HP1 complex from CryoSPARC. (c) A simplified flow chart of cryo-EM

data processing. **(d)** Fourier shell correlation (FSC) curve of DIM2-HP1 complex map as a function of resolution using CryoSPARC output. **(e)** Angular distribution calculated in CryoSPARC for particle projections. Heat map shows number of particles for each viewing angle. **(f)** Local resolution map of the DIM2-HP1 complex.



Supplementary Fig. 3 | Cryo-EM density maps for the DIM2-HP1 complex.

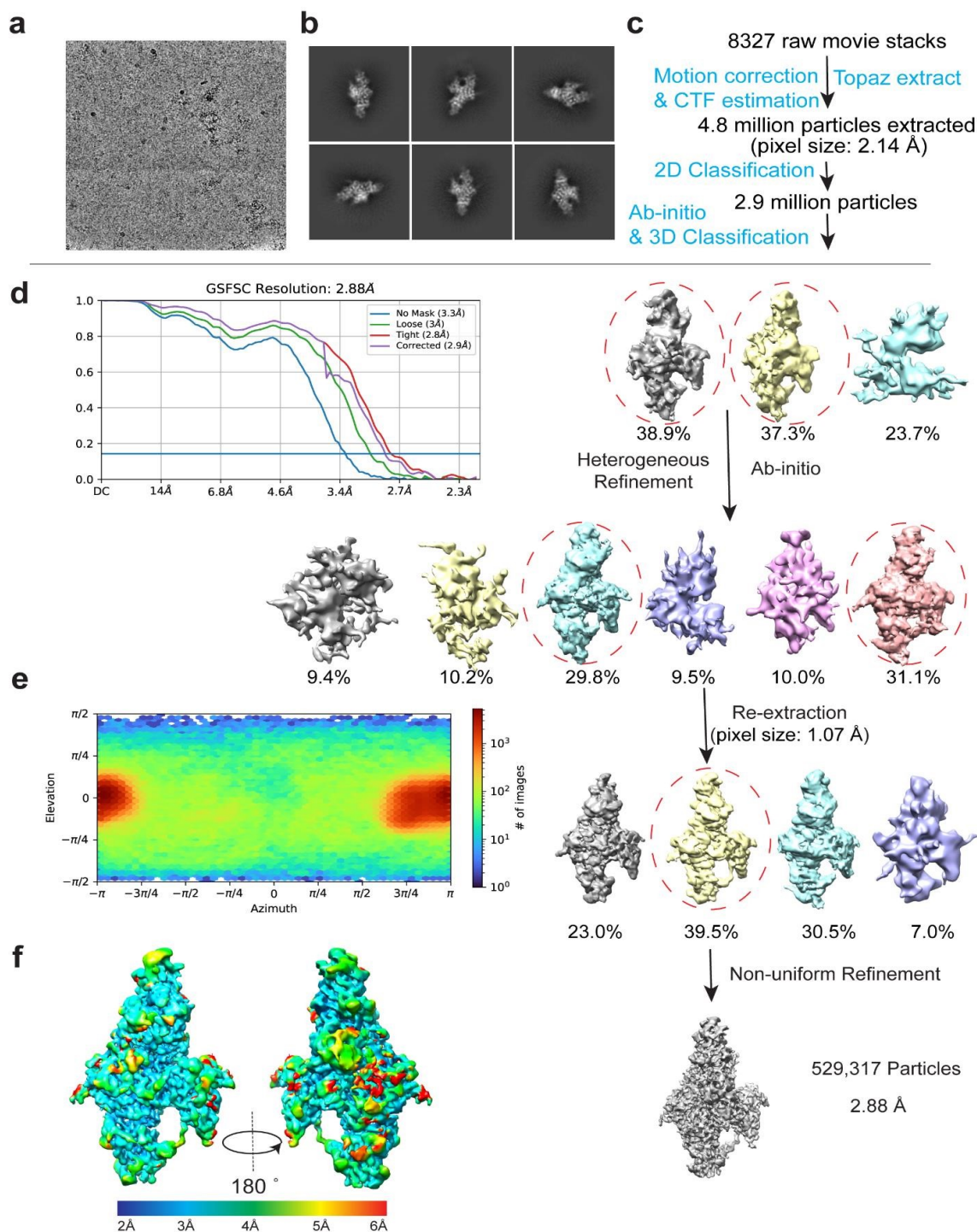
(a-g) Cryo-EM density and built in models for the final map for the DIM2-HP1 complex (a), DIM2 RFTS domain (b), DIM2 BAH1 domain (c), DIM2 BAH2 domain (d), DIM2 MTase domain (e), HP1 dimer and bound DIM2 N-terminal tail (f), and SAH molecule in a different view from that of Fig. 1g (g).



Supplementary Fig. 4 | Size-exclusion chromatography analyses of the DIM2-HP1 interaction.

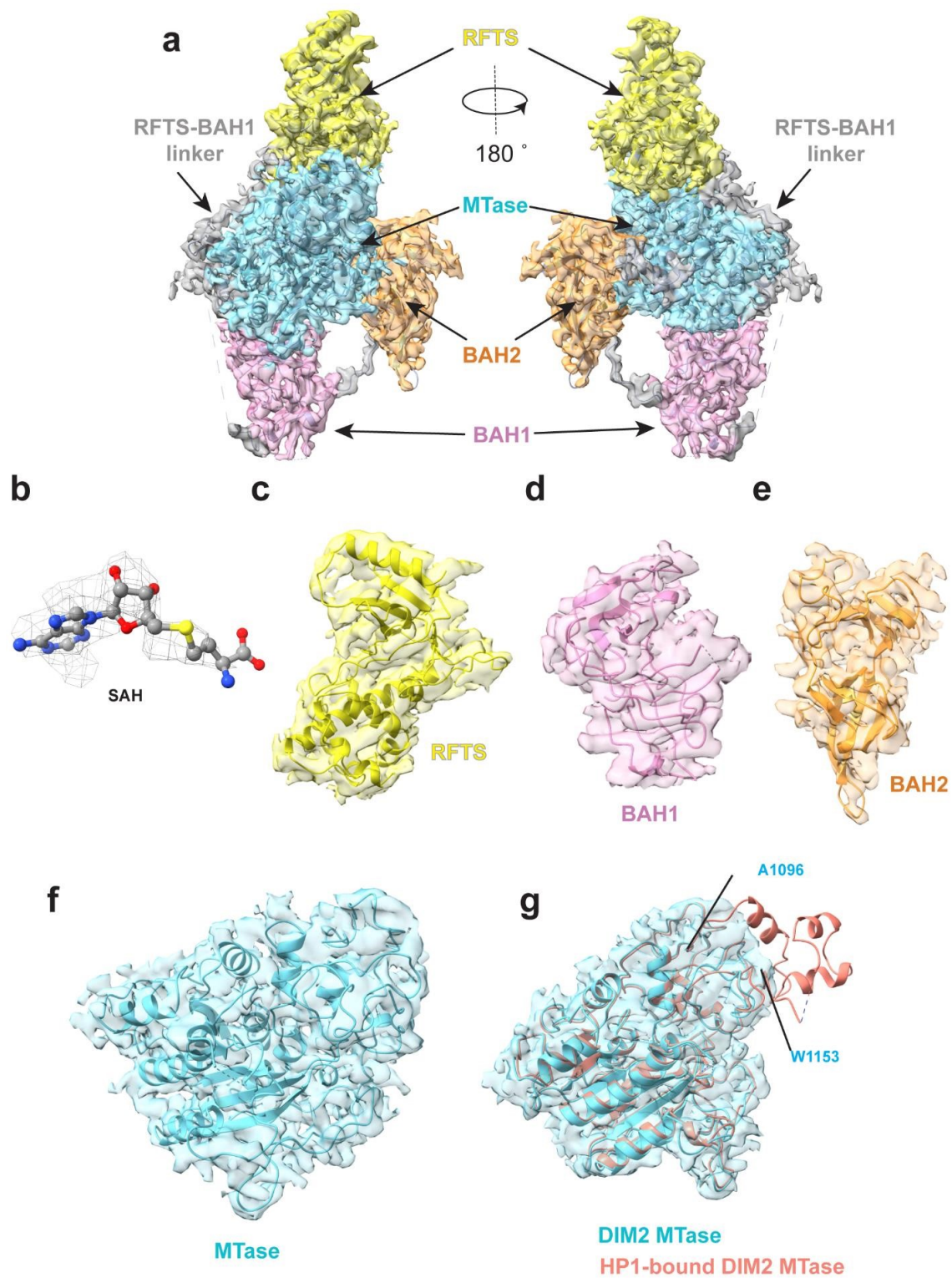
(a, b) Size-exclusion chromatography profile of WT DIM2 (DIM2^{WT}) alone (a) or HP1 alone (b). (c-f) Size-exclusion chromatograph profile of the DIM2-HP1 mixture harboring DIM2^{WT} (c),

L134A/L139A-mutated DIM2 (DIM2^{L134A/L139A}) (d), R1104A-mutated DIM2 (DIM2^{R1104A}) (e), and L134A/L139A/R1104A-mutated DIM2 (DIM2^{L134A/L139A/R1104A}) (f). The SDS-PAGE images for selected fractions are shown below.



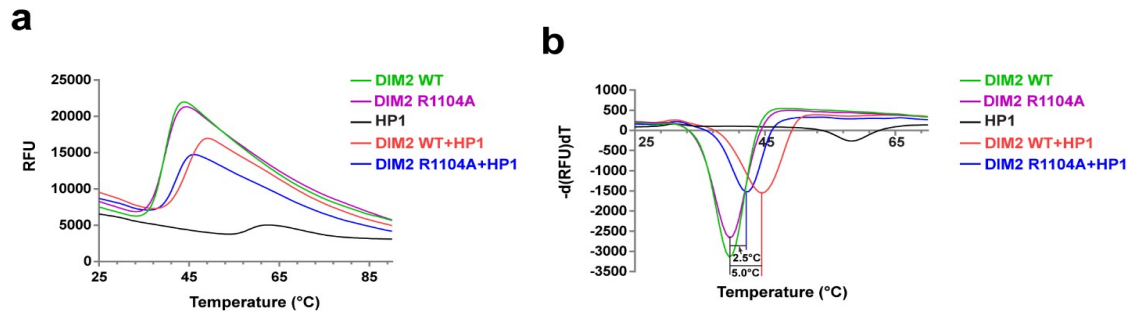
Supplementary Fig. 5 | Data processing workflow for cryo-EM reconstruction of apo-DIM2. (a) A representative micrograph of apo-DIM2 by cryo-EM. (b) Representative 2D classes of apo-DIM2 from CryoSPARC. (c) A simplified flow chart of cryo-EM data processing. (d) Fourier

shell correlation (FSC) curve of apo-DIM2 map as a function of resolution using CryoSPARC output. (e) Angular distribution calculated in CryoSPARC for particle projections. Heat map shows number of particles for each viewing angle. (f) Local resolution map of apo-DIM2.



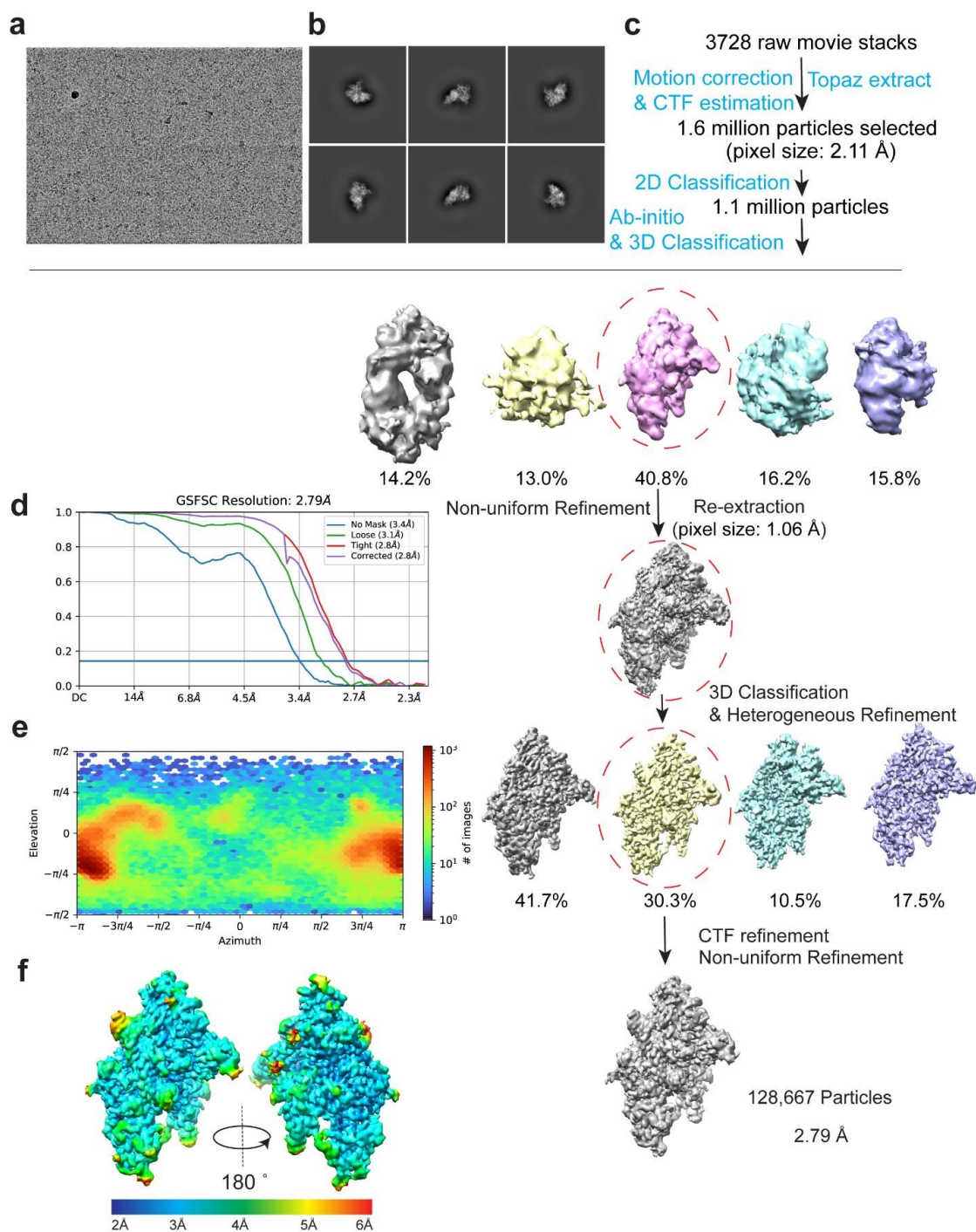
Supplementary Fig. 6 | Cryo-EM density maps for apo-DIM2.

(a-f) Cryo-EM density and built in models for the final map for apo-DIM2 (a), SAH molecule in a different view from Fig. 3e (b), DIM2 RFTS domain (c), DIM2 BAH1 domain (d), DIM2 BAH2 domain (e), and DIM2 MTase domain (f). (g) Structural alignment between HP1-bound DIM2 MTase (salmon) and the MTase domain of apo-DIM2 (cyan).



Supplementary Fig. 7 | Thermal shift assay for wild type (WT) and R1104A-mutated DIM2, in the presence or absence of HP1.

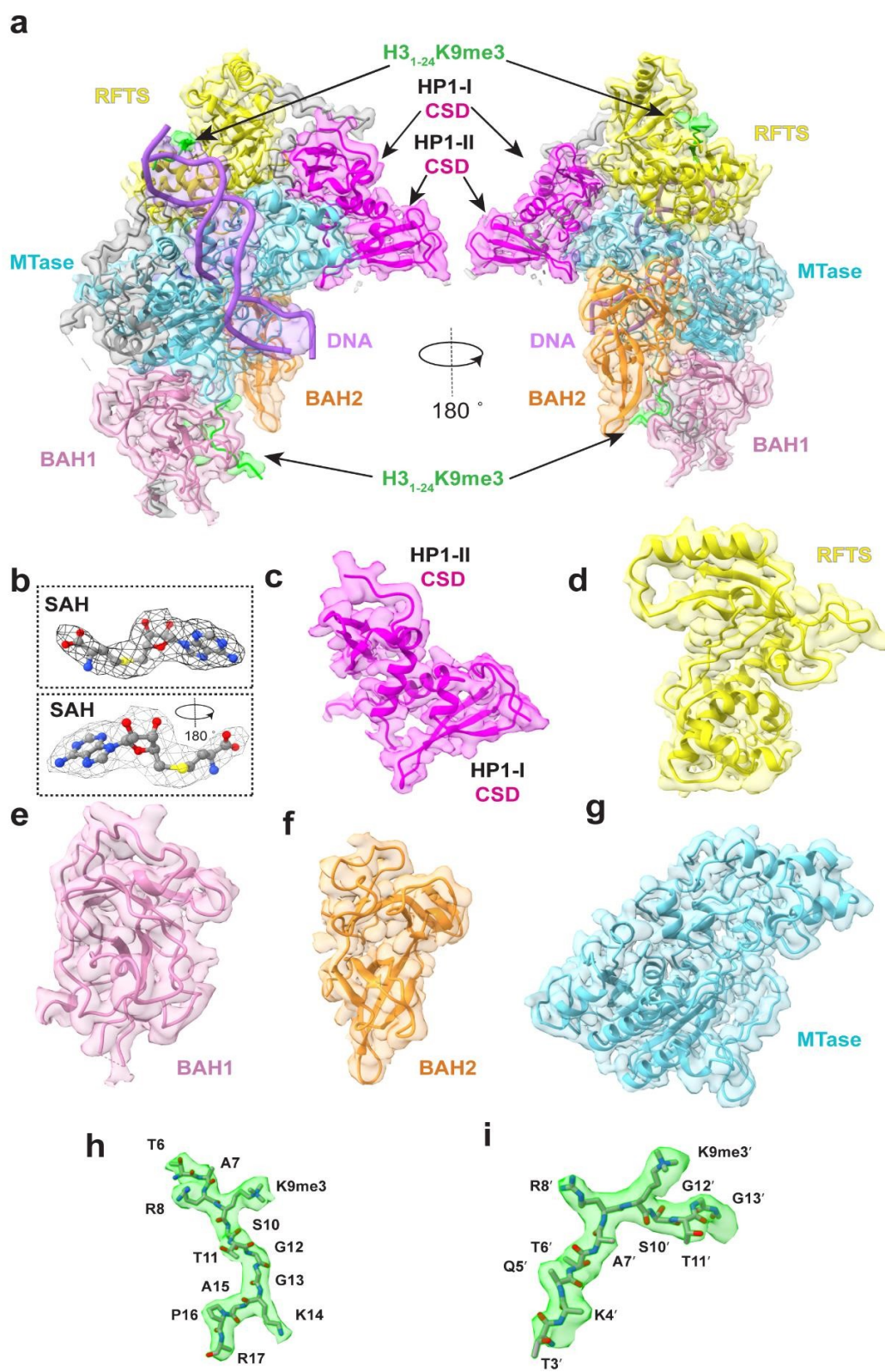
(a, b) Raw fluorescence data (a) and first derivative of the raw data (b). The difference in melting temperature (T_m) between free and complexed WT or R1104A DIM2 is indicated, respectively. Note that HP1 alone (black) in two-fold molar excess than that of DIM2 (green and purple) yielded a much-reduced fluorescence signal, with a T_m of 58.5 °C, which is distinct from the major fluorescence peaks measured for the free or HP1-bound state of DIM2.



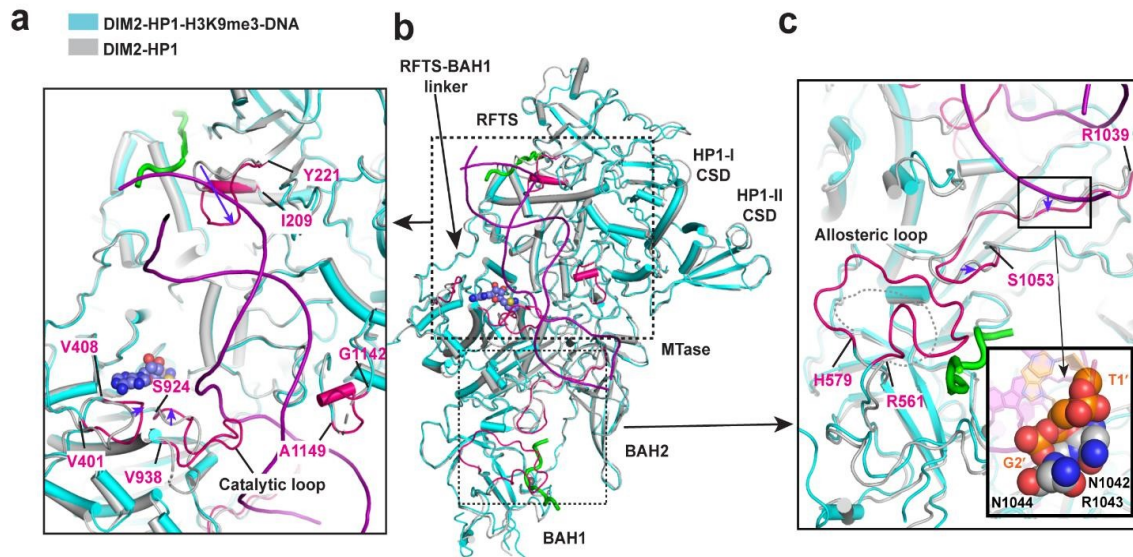
Supplementary Fig. 8 | Data processing workflow for cryo-EM reconstruction of DIM2-HP1H3K9me3-DNA complex.

(a) A representative micrograph of the DIM2-HP1-H3K9me3-DNA by cryo-EM. (b) Representative 2D classes of the DIM2-HP1-H3K9me3-DNA from CryoSPARC. (c) A simplified flow chart of cryo-EM data processing. (d) Fourier shell correlation (FSC) curve of the

DIM2-HP1-H3K9me3-DNA map as a function of resolution using CryoSPARC output. **(e)** Angular distribution calculated in CryoSPARC for particle projections. Heat map shows number of particles for each viewing angle. **(f)** Local resolution map of DIM2-HP1-H3K9me3-DNA complex.

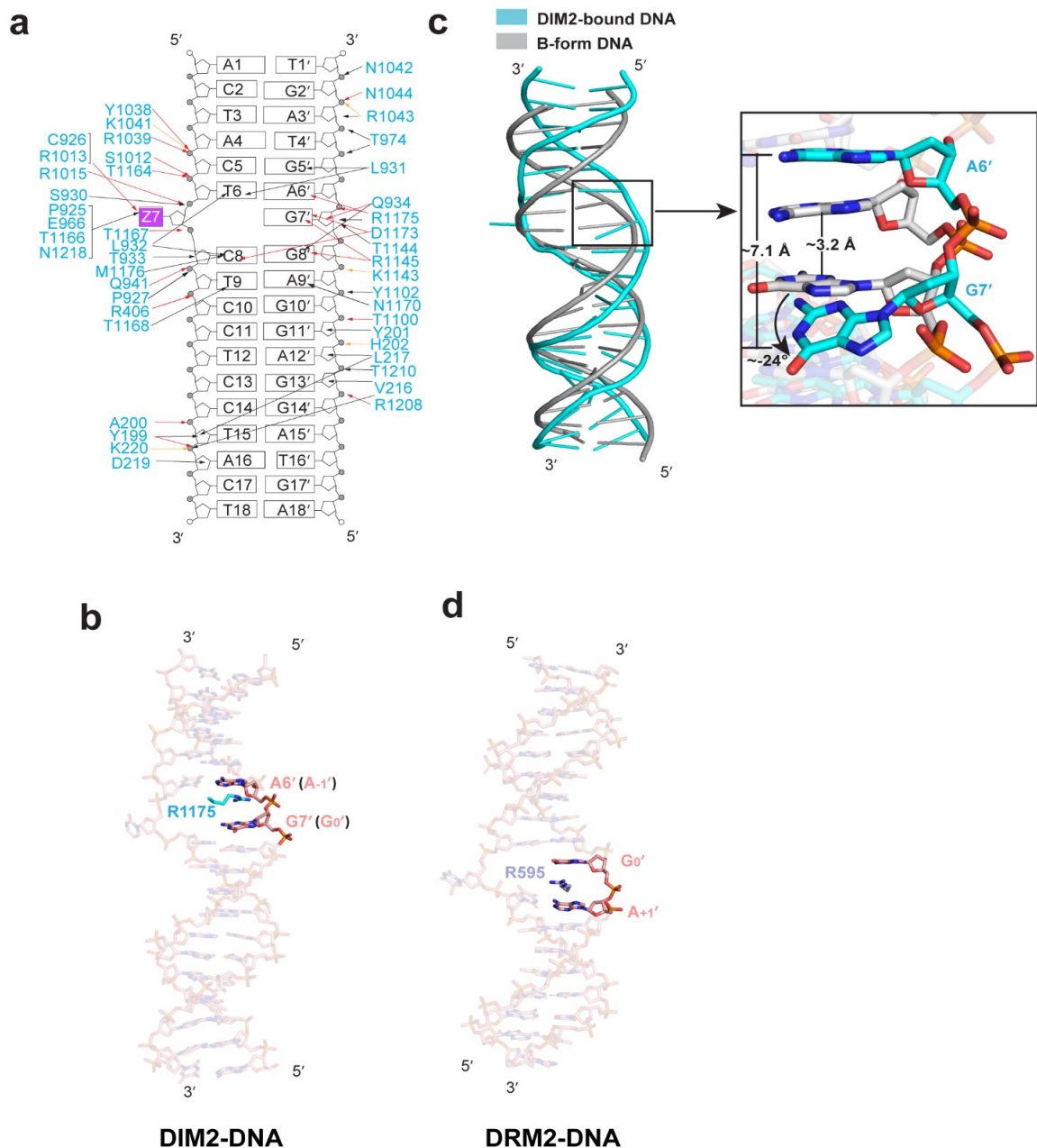


Supplementary Fig. 9 | Cryo-EM density maps for the DIM2-HP1-H3K9me3-DNA complex.
(a-i) Cryo-EM density and built in models for the final map for the DIM2-HP1-H3K9me3-DNA complex (a), SAH molecule in two opposite views (b), HP1 dimer (c), DIM2 RFTS domain (d), DIM2 BAH1 domain (e), DIM2 BAH2 domain (f), DIM2 MTase domain (g), DIM2 RFTS-bound H3K9me3 (h), and DIM2 BAH1-bound H3K9me3 (i).



Supplementary Fig. 10 | Structural comparison of DIM2-HP1-H3K9me3-DNA complex and DIM2-HP1 complex.

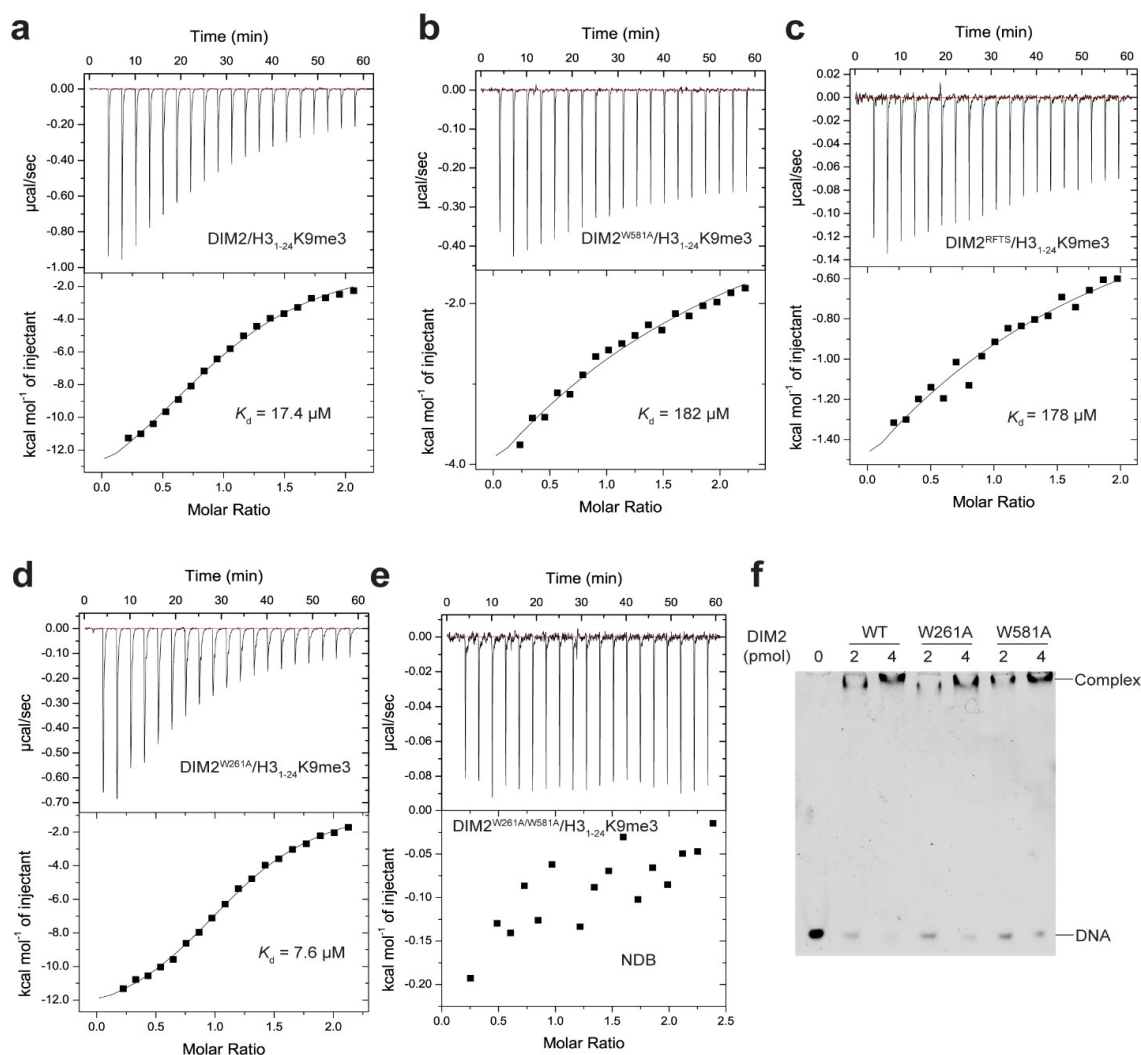
(a-c) Structural overlay of the DIM2-HP1 complex (grey) and the DIM2-HP1-H3K9me3-DNA complex (cyan). The DIM2 regions that exhibit the largest conformational changes are colored in hot pink in the DIM2-HP1-H3K9me3-DNA complex and highlighted in expanded views (a) and (c). The potential steric clashes between the bridging loop of DIM2 and DNA are indicated in sphere representation in (c).



Supplementary Fig. 11 | Structural details for the DIM2-DNA interaction.

(a) Schematic view of the DIM2-DNA interactions. Hydrogen-bonding, electrostatic, and van der Waals contacts are indicated by red, wheat, and black arrows, respectively. (b) Structural overview of the DNA intercalation by DIM2 R1175. Note that DIM2 R1175 intercalates between the orphan guanine G7' (a.k.a. G₀') and its preceding nucleotide A6' (a.k.a. A₋₁'). (c) Structural alignment between DIM2-bound DNA (cyan) and B-form DNA (grey), with the expanded view showing that DIM2 binding increased the rise of the A6'-G7' step from 3.2 Å in B-form DNA to 7.1 Å in DIM2-bound DNA and introduced a roll of ~24°. (d) Structural overview of the

structural intercalation by DRM2 R595 (PDB 7L4C). Note that DRM2 R595 intercalates between the orphan guanine G₀' and its subsequent nucleotide A₊₁'.



Supplementary Fig. 12 | Biochemical analysis of DIM2, WT or mutant.

(a-e) Representative ITC binding curves for the H3K9me3 peptides to the DIM2 (a), W581A-mutated DIM2 (DIM2^{W581A}) (b), DIM2 RFTS domain (DIM2^{RFTS}) (c), W261A-mutated DIM2 (DIM2^{W261A}) (d), and W261A/W581A-mutated DIM2 (DIM2^{W261A/W581A}) (e). The top panels show the raw heat change data. The lower panel shows the integrated binding isotherms as a function of the molar ratio of the ligand to the enzyme. (f) EMSA analysis of the interaction between DIM2 and an 18-bp DNA in the presence of the H3K9me3 peptide. Each sample contains 1 pmol of DNA, 0 (for the sample lack of DIM2) or 8 (for the sample containing DIM2) pmol of H3K9me3 peptide, and various amounts of DIM2 protein. The bands corresponding to free DNA and the protein-DNA complex are indicated, respectively. The experiment was repeated three times with consistent results. Source data are provided as a Source Data file.

DIM2 1 MDSPPRRSGCMFIDVPAPTMGFQPEYLDMFASVLSQGLAEGDYAHHQPLBAGK.EECLPEIAYTTTITPSPDDPOLQLQ
DNMT1 219 AEPERRASG.....TRTEKEERDEKKEKRLRSQTKE.....PTEKQALKKE..EDREARAGVQADEDEDGDEK

DIM2 80 LLEQQPQTESGLNGVDPAAPAESETEADLPDGFSDSPDDFVVQESKHITVDLP.....VSTLINPRS.....T
DNMT1 282 DEKKHRSQPKD...LAAKRRPEEKEPEKVNQISDEKDEDEKEEKRRK.TTPKEPTEKKMARAKTVMNSKTHPPKCIQC

DIM2 146 FORIDEND.....NTVPPQS.TPPEVAVEDLLKAAKAGKNKYIEFEHDFNFVNYAYHPQEMRPTIQ..LVAT
DNMT1 357 GOYLDPPDLKYGQHPDAVDEPOMITNEKLSIFD...ANESGFESYALPQHKITCESVYCKKHG...LCPIIDTGLIEK

DIM2 215 KVLHDKYTFDGVLR..YGNTKHYVTGMQVLEL.PVGNVYASLHVSVKQIIVVRSKHNAKREIY.....YLLKKPAFET
DNMT1 430 NI...ELFESGSAKPIYDDDDPSLEGVNGKNLGPINEW.....WITGFDGGEKALIGFSTSAFAYELMDSEPEY

DIM2 284 QRYT...QPELWIAADLGKHVVDPYCTRMVERKREVTLGCFKSDFIQWASKAHGKSAFQNWRAQHPSDDFRSTVAANIGYI
DNMT1 496 APIELGMQEKLYIS...KIVVEF.....LOSNSDSTYEDL.....

DIM2 361 WKEINQVAGAKRAACDCL..FRELMLVKKPGQVFRQVPPGPVTEGDRVAATIITPYLKECFGHMILGKVLRLAGEDAE
DNMT1 528 ...LNKLETTVPSPGLNLRTEEDSLRLHACQVVEOVE...SYDEAGSDSQPIFLTRCMRD.....LKLKLAG....

DIM2 439 KEKPKVKAARLKIENKNATKADTKDDMKNDTATESLPTPTPSLSPVQVLEATPESDIVSIVSSDLPPSEN.....
DNMT1 590VTLCGRRAOARQTIHSTRE..KDRGPTKATTTKLVYQIFDTFFAQOIEKD.....DREDKENAFKRRRCGVC

DIM2 509 ...NPPPL.....TNGSVKPKAKA.....NPKPKPSTQPLHAAHVRYLSQELVN....
DNMT1 657 EVCCQPECGKCKACKDMVKFGSGSRKQACQERRCPNMAKEADDDEEVDDNIPEMPSPKKMHQCKKQKNKNRISVWGE

DIM2 550KIKVGDVISTPRDDSSNTDTKWKPTDTHHRWFGLVQRVHTAKTKSSGRGLNKSFSFVIVWF
DNMT1 737 AVKTDGKKSYYKKVCIDAETLEVGDCVSVIPDDSS.....KPL.....YLARV.TALWEDSS...NQGMFAHWF

DIM2 611 YRPDTPCCAMKYKWRNELFLSNHCTCQEGHARVKGNEVLAHPVDW.....
DNMT1 798 CAGTDTVLGATSDPL..ELFLVDECDMQLSYLHKKVVLKAPSENWAMEGGMDPESLLEGDDGKTYFYQLWYDQDYAR

DIM2 659 FGTPE.....PSNKEFFVVRQVYESEQRRWITLQK.....DHLTCYHNQ.....PPK.....
DNMT1 876 FESDPKTPQTEENKPKFCVSCARLAEMR.....OKIIPRVLEQLEDLDSRVLYYSATKNGILYRVGDGVYLDPEAFTFNI

DIM2 700PPTAPYKPGDVLATLSPSD..KFS.....PY...EVVEYF...TQGE.KETAF.VRLKLLRRKVD
DNMT1 951 KLSPPVKRPRK..EYVDEDDLYEHYRKYSIDYIKGSLNDAPEPYRIGRIKETIFCPKKSNCRPNETDIKIRVNFYRPENT.

DIM2 754 RODAPAN.....ELVYTED...LVDVRAERTVKKCIIMRCFRPDERVPSPEYDRGCTGNMFFETHRDHGRCVPLDTPPTL
DNMT1 1028 HKSTPASVHADINLWYWSDEFAVVDFA..VOGRCTVEYGEDLPECVQVYSMGGPNRFYFL..EAYNAKSKSFEDEPNHA

DIM2 826 RQGFNPPLGNLGR.....PKLRGMDLYCGGNGFGRGLEEGGVVEMRWANDIWDKATHTYMA
DNMT1 1104 R...SPGNKKGKGGKGGKPKSQACEPSEPEIEIKL PKLRTLDVSGGCGLSRGFHQAGLSTLWAIEMWDPAQAERL

DIM2 881 NTPDPNKTNPPLGSVDDLLRLALEGKFSDN...VPRPGVDFTIAGSPCPGFSLLTQ.DKRVLNQVKNOSLVASFASFV
DNMT1 1180 NNPGL...STVETEDCNILLKLVMAGETTNSRGQRLEQKGEVEMLCGPPCQGFSGMNRFNSTRYSKFKN.SLVVSFSLYC

DIM2 956 DRYRKYGVLENVSGIVQTFVNRQD.VLSOLFCAVGMGYQAQLIGDAWAHCAPOERRERVFYFAAPGLPLPPLPS
DNMT1 1256 DYYRPRFLEN...VRNFVSFKRSMVILKTLRLCLVRMGYQCTFGVLOAGQYGAQRRRAITLLAAPGKPLPFPPL

DIM2 1035 HSHYRVKNRNTIGFLCNGESYVQR...SFIPTAFKRFVSAEGTADLPKIG.....DGKPDACVR.....
DNMT1 1332 HV..TAPRACQLSVVDDKKFVSNITRLSSGP..FRITLTVRDTMSDLPEVRNGASALEISYNGEPQSWFQRLRGAQYQPI

DIM2 1090 FPDHRLASGTTPTYTRAQYACTPETHPYGMNFIAWNN.....GNCVMSKSDRLFPSEGKTRPSDASV.....GW
DNMT1 1409 LRDE..ICKDMALVAAARMRHTEILAP.GSD...WRDLPNIEVRLSDCTMAKRLRYTHHDRKNRSGSGLRGVCSCEVAG

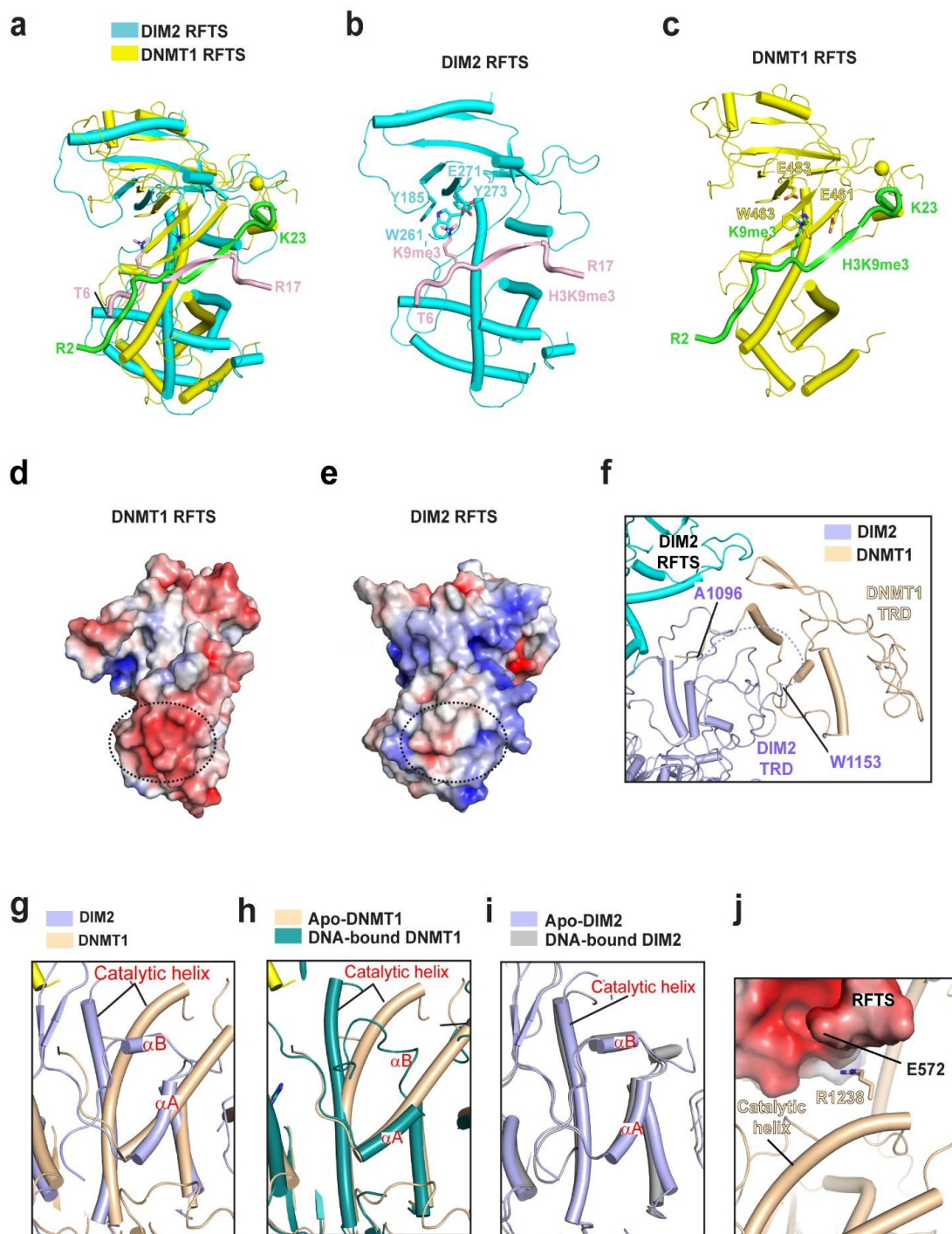
DIM2 1154 KRLNPL...KTLFPTV.....TTSNPDS.ARMGPGLHWEDRPTVQEMRRAQGYLDE
DNMT1 1483 KACDPAARQFNTELEFWCLPHTGNRRHHWAGLYGRLEWDGFFS.TTVNPEPMGKQGRVLHPEQHRRVSVRECARSGGFPT

DIM2 1203 EVLVGRITDQWKLVGNSVSRHMAAIGL...KREAWLGTLYDESAVVATATATATATAAAGVTVPMEEFGIGT
DNMT1 1563 YRLGNILDKHRQVGNVPPPLAKAIGLEIKLCMLAKAREASAKLKEBAAKD.....

DIM2 1275 TESSRPSRSPVHTAVDLDDSKSERSRSTTPATVLS'SSAAGDGSANAAGLEDDDDNDMMEMMEVTRKRSSPAVDEAGMRPS
DNMT1

Supplementary Fig. 13 | Sequence alignment between DIM2 and human DNMT1.

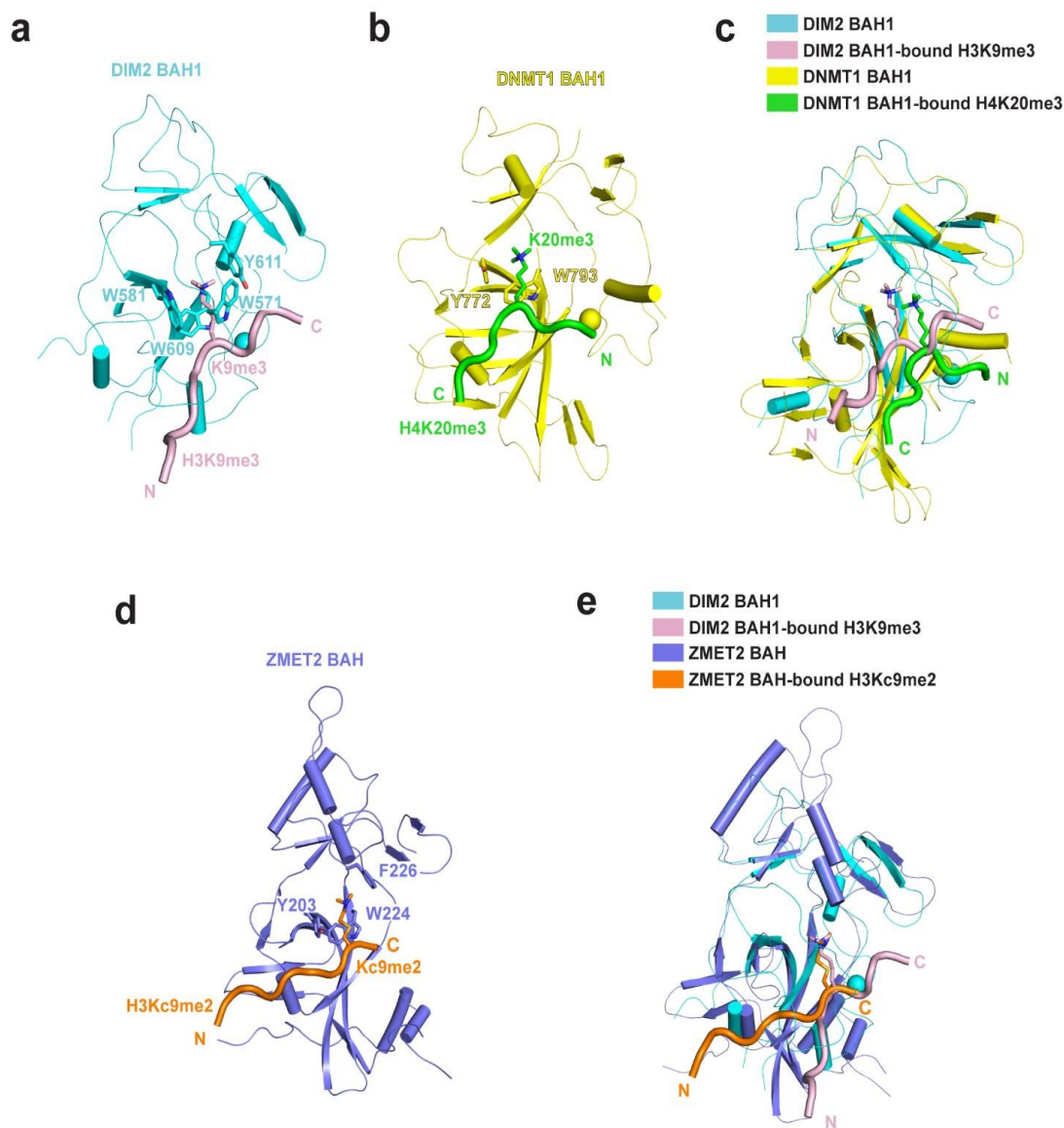
Identical or similar residues are boxed and colored in red. Completely conserved residues are shaded in red. The DIM2 domains are labeled above the corresponding sequences, while the DNMT1 domains are labeled below the corresponding sequences. The DIM2 residues responsible for the H3K9me3 and HP1 interactions are marked by green and magenta arrows, respectively. The DNMT1 residues responsible for the interaction with H3K9me3 and ubiquitin are marked by blue and dark arrows, respectively.



Supplementary Fig. 14 | Structural comparison between the RFTS domains of DIM2 and DNMT1.

(a) Structural alignment of H3K9me3 (light pink)-bound DIM2 RFTS domain (cyan) and the H3K9me3 (green)-bound DNMT1 RFTS domain (yellow) (PDB 6PZV). The zinc ion bound to the DNMT1 RFTS domain is shown in sphere representation. (b) Ribbon representation of

H3K9me3 (light pink)-bound DIM2 RFTS domain (cyan). The H3K9me3 pocket residues of DIM2 RFTS are shown in stick representation. (c) Ribbon representation of H3K9me3 (green)-bound DNMT1 RFTS domain (yellow) (PDB 6PZV). The H3K9me3 pocket residues are shown in stick representation. (d) Electrostatic surface of the DNMT1 RFTS domain (PDB 4WXX). The region involved in the binding to the MTase domain is circled by dotted line. (e) Electrostatic surface of the DIM2 RFTS domain. The region corresponding to the MTase domain-binding site of the DNMT1 RFTS domain is circled by dotted line. (f) Close-up view of the overlaid TRD subdomain between apo-form DIM2 and DNMT1 (PDB 4WXX). (g) Structural overlay between the catalytic helix and RFTS-BAH1 linker helices α A and α B of apo-DIM2 and the corresponding regions of DNA-free DNMT1 (PDB 4WXX). (h) Structural overlay between apo-form (PDB 4WXX) and DNAbound (PDB 7XI9) DNMT1, highlight the region corresponding to the catalytic helix and RFTSBAH1 linker helices α A and α B of DIM2. (i) Structural overlay between apo-DIM2 and DNA-bound DIM2, highlight the catalytic helix and RFTS-BAH1 linker helices α A and α B. (j) Electrostatic contact between residue E572 in the RFTS domain (Electrostatic surface) and residue R1238 (stick representation) in the catalytic helix of DNMT1 (ribbon representation).



Supplementary Fig. 15 | Structural comparison of the histone-binding BAH domains between DIM2, DNMT1, and ZMET2.

(a) Ribbon representation of H3K9me3 (light pink)-bound DIM2 BAH1 domain (cyan), with the residues forming the H3K9me3-binding cage shown in stick representation. The zinc ion is shown in sphere representation. The N- and C-termini of the peptide are labeled “N” and “C”, respectively. (b) Ribbon representation of H4K20me3 (green)-bound DNMT1 BAH1 domain (yellow) (PDB 7LMK), with the residues forming the H4K20me3-binding cage shown in stick representation. The zinc ion is shown in sphere representation. The N- and C-termini of the peptide are labeled “N” and “C”, respectively. (c) Structural overlay of H3K9me3-bound BAH1 (cyan) of DIM2 with H4K20me3-bound DNMT1 BAH1 (yellow). The H3K9me3 and H4K20me3 are shown in pink and green, respectively. (d) Ribbon representation of H3K9me3 (orange) -bound ZMET2 BAH domain (slate) (PDB 7UBU), with the residues forming the

H3K9me2-binding cage shown in stick representation. (e) Structural overlay of H3K9me3 (pink)-bound BAH1 (cyan) of DIM2 with H3K9me2 (orange)-bound ZMET2 BAH domain (slate).

Supplementary Table 1. Summary of ITC binding parameters.

Protein	Peptide	K_d (μM)	N value
DIM2, WT	H3 ₁₋₂₄ K9me3	$17.4 \pm 0.6^{\#}$	0.95 ± 0.08
DIM2, W261A	H3 ₁₋₂₄ K9me3	$7.6 \pm 1.2^{\#}$	1.07 ± 0.11
DIM2, W581A	H3 ₁₋₂₄ K9me3	$182 \pm 67^{\#}$	1.0*
DIM2 RFTS domain	H3 ₁₋₂₄ K9me3	$178 \pm 24^{**}$	1.0*
DIM2, W261A/W581A	H3 ₁₋₂₄ K9me3	NDB	
HP1, WT	H3 ₁₋₂₄ K9me3	$0.17 \pm 0.03^{\#}$	0.97 ± 0.12
HP1, W98A	H3 ₁₋₂₄ K9me3	$89.5 \pm 7.6^{\#}$	1.03 ± 0.03

NDB, no detectable binding. * The N value was set manually. [#] The mean value and S.D. were derived from two-independent measurements. ** The mean value and S.D. were derived from curve fitting of a single measurement.

Supplementary Table 2. Cryo-EM data collection, refinement, and validation statistics.

States	DIM2-HP1 (EMD-44415, PDB 9BAZ)	Apo-DIM2 (EMD-44110, PDB 9BAP)	DIM2-HP1H3K9me3- DNA (EMD-44111, PDB 9BAQ)
Codes			
Data collection and processing			
Microscope	Titan Krios	Titan Krios	Titan Krios
Camera	K3 BioQuantum	K3 BioQuantum	K3 BioQuantum
Magnification	81,000	81,000	81,000
Voltage (kV)	300	300	300
Defocus range (μm)	-0.8 ~ -2.5	-0.6 ~ -1.5	-0.8 ~ -2.5
Exposure time (s)	4.1	2.4	2.7
Dose rate($e^-/\text{\AA}^2/\text{s}$)	16.3	20.8	25.9
Number of frames	68	40	30
Pixel size ($\text{\AA}$)	1.056	1.07	1.056
Micrographs (no.)	4009	8327	3728
Symmetry imposed	<i>C1</i>	<i>C1</i>	<i>C1</i>
Initial particles (no.)	1,476,534	4,792,885	1,550,688
Final particles (no.)	296,077	529,317	128,667
Map resolution ($\text{\AA}$)	2.76	2.88	2.79
FSC threshold	0.143	0.143	0.143
Refinement			
Initial model used	DIM2 (AlphaFold), HP1 (AlphaFold)	DIM2 (AlphaFold)	DIM2 (AlphaFold), HP1 (AlphaFold), DNA (PDB 6F57)
Model resolution ($\text{\AA}$)	3.2	3.2	3.0
FSC threshold	0.5	0.5	0.5
Map sharpening <i>B</i> factor ($\text{\AA}^2$)	-85.8	-107.3	n.a
CC (mask)	0.71	0.67	0.81
Model composition			
Non-hydrogen atoms	8283	6535	9733
Protein residues	1090	887	1154
DNA	-	-	36
Ion (zinc)	1	1	1
SAH	1	1	1
<i>B</i> factors ($\text{\AA}^2$)			
Protein	128.17	98.02	127.86
DNA	-	-	196.65
Ion (zinc)	188.74	153.67	196.94
SAH	114.25	85.7	111.6
R.m.s. deviations			
Bond lengths ($\text{\AA}$)	0.004	0.003	0.003
Bond angles ($^\circ$)	0.690	0.658	0.619
Validation			
MolProbity score	1.79	1.71	1.62
Clashscore	12.04	9.48	8.74
Poor rotamers (%)	0.5	0.84	0.66
Ramachandran plot			
Favored (%)	96.83	96.68	97.18
Allowed (%)	3.17	3.32	2.82
Disallowed (%)	0.00	0.00	0.00

n.a: Anisotropic sharpened by Phenix

References

1. Janssen, A., Colmenares, S.U. & Karpen, G.H. Heterochromatin: Guardian of the Genome. *Annual review of cell and developmental biology* 34, 265-288 (2018).
2. Allshire, R.C. & Madhani, H.D. Ten principles of heterochromatin formation and function. *Nature reviews. Molecular cell biology* 19, 229-244 (2018).
3. Richards, E.J. & Elgin, S.C. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* 108, 489-500 (2002).
4. Bannister, A.J. et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410, 120-124 (2001).
5. Jacobs, S.A. et al. Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J* 20, 5232-5241 (2001).
6. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410, 116-120 (2001).
7. Du, Q., Luu, P.L., Stirzaker, C. & Clark, S.J. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* 7, 1051-1073 (2015).
8. Grewal, S.I. & Moazed, D. Heterochromatin and epigenetic control of gene expression. *Science* 301, 798-802 (2003).
9. Greenberg, M.V.C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology* 20, 590-607 (2019).
10. Martienssen, R.A. & Colot, V. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* 293, 1070-1074 (2001).
11. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315-322 (2009).
12. Cheng, X. & Blumenthal, R.M. Mammalian DNA methyltransferases: a structural perspective. *Structure* 16, 341-350 (2008).
13. Goll, M.G. & Bestor, T.H. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74, 481-514 (2005).
14. Law, J.A. & Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11, 204-220 (2010).
15. Ren, W., Gao, L. & Song, J. Structural Basis of DNMT1 and DNMT3A-Mediated DNA Methylation. *Genes* 9 (2018).
16. Finnegan, E.J. & Dennis, E.S. Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Res* 21, 2383-2388 (1993).
17. Lindroth, A.M. et al. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* 292, 2077-2080 (2001).

18. Stroud, H. et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* 21, 64-72 (2014).
19. Henderson, I.R. & Jacobsen, S.E. Epigenetic inheritance in plants. *Nature* 447, 418-424 (2007).
20. Foss, H.M., Roberts, C.J., Claeys, K.M. & Selker, E.U. Abnormal chromosome behavior in *Neurospora* mutants defective in DNA methylation. *Science* 262, 1737-1741 (1993).
21. Nai, Y.S., Huang, Y.C., Yen, M.R. & Chen, P.Y. Diversity of Fungal DNA Methyltransferases and Their Association With DNA Methylation Patterns. *Front Microbiol* 11, 616922 (2020).
22. Russell, P.J., Rodland, K.D., Rachlin, E.M. & McCloskey, J.A. Differential DNA methylation during the vegetative life cycle of *Neurospora crassa*. *Journal of bacteriology* 169, 2902-2905 (1987).
23. Kouzminova, E. & Selker, E.U. *dim-2* encodes a DNA methyltransferase responsible for all known cytosine methylation in *Neurospora*. *EMBO J* 20, 4309-4323 (2001).
24. Lewis, Z.A. et al. Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*. *Genome research* 19, 427-437 (2009).
25. Rountree, M.R. & Selker, E.U. DNA methylation and the formation of heterochromatin in *Neurospora crassa*. *Heredity* 105, 38-44 (2010).
26. Du, J., Johnson, L.M., Jacobsen, S.E. & Patel, D.J. DNA methylation pathways and their crosstalk with histone methylation. *Nature reviews. Molecular cell biology* 16, 519-532 (2015).
27. Du, J. et al. Dual binding of chromomethylase domains to H3K9me₂-containing nucleosomes directs DNA methylation in plants. *Cell* 151, 167-180 (2012).
28. Fang, J. et al. Mechanistic basis for maintenance of CHG DNA methylation in plants. *Nature communications* 13, 3877 (2022).
29. Ren, W. et al. Direct readout of heterochromatic H3K9me₃ regulates DNMT1-mediated maintenance DNA methylation. *Proc Natl Acad Sci U S A* 117, 18439-18447 (2020).
30. Stoddard, C.I. et al. A Nucleosome Bridging Mechanism for Activation of a Maintenance DNA Methyltransferase. *Mol Cell* 73, 73-83 e76 (2019).
31. Tamaru, H. & Selker, E.U. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* 414, 277-283 (2001).
32. Honda, S. & Selker, E.U. Direct interaction between DNA methyltransferase DIM-2 and HP1 is required for DNA methylation in *Neurospora crassa*. *Mol Cell Biol* 28, 6044-6055 (2008).
33. Freitag, M., Hickey, P.C., Khlafallah, T.K., Read, N.D. & Selker, E.U. HP1 is essential for DNA methylation in *neurospora*. *Mol Cell* 13, 427-434 (2004).
34. Tamaru, H. et al. Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in *Neurospora crassa*. *Nat Genet* 34, 75-79 (2003).
35. Ren, W. et al. DNMT1 reads heterochromatic H4K20me₃ to reinforce LINE-1 DNA methylation.

Nature Communications. In press (2021).

36. Song, J., Rechkoblit, O., Bestor, T.H. & Patel, D.J. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* 331, 1036-1040 (2011).
37. Song, J., Teplova, M., Ishibe-Murakami, S. & Patel, D.J. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* 335, 709-712 (2012).
38. Syeda, F. et al. The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *J Biol Chem* 286, 15344-15351 (2011).
39. Takeshita, K. et al. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proc Natl Acad Sci U S A* 108, 9055-9059 (2011).
40. Zhang, Z.M. et al. Crystal Structure of Human DNA Methyltransferase 1. *J Mol Biol* 427, 2520-2531 (2015).
41. Jeltsch, A. On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics* 1, 63-66 (2006).
42. 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* 68, 350-360 e357 (2017).
43. Kikuchi, A. et al. Structural basis for activation of DNMT1. *Nature communications* 13, 7130 (2022).
44. Nishiyama, A. et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* 502, 249-253 (2013).
45. Qin, W. et al. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell research* 25, 911-929 (2015).
46. Bestor, T.H. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J* 11, 2611-2617 (1992).
47. Cao, X. et al. Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. *Proc Natl Acad Sci U S A* 97, 4979-4984 (2000).
48. Dumesic, P.A., Stoddard, C.I., Catania, S., Narlikar, G.J. & Madhani, H.D. ATP Hydrolysis by the SNF2 Domain of Dnmt5 Is Coupled to Both Specific Recognition and Modification of Hemimethylated DNA. *Mol Cell* 79, 127-139 e124 (2020).
49. Klimasauskas, S., Kumar, S., Roberts, R.J. & Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 76, 357-369 (1994).
50. Reinisch, K.M., Chen, L., Verdine, G.L. & Lipscomb, W.N. The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* 82, 143-153 (1995).
51. Thiru, A. et al. Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. *EMBO J* 23, 489-499 (2004).
52. Liu, Y. et al. Peptide recognition by heterochromatin protein 1 (HP1) chromoshadow domains

- revisited: Plasticity in the pseudosymmetric histone binding site of human HP1. *J Biol Chem* 292, 5655-5664 (2017).
53. Cheng, X. Structure and function of DNA methyltransferases. *Annual review of biophysics and biomolecular structure* 24, 293-318 (1995).
 54. Anteneh, H., Fang, J. & Song, J. Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. *Nature communications* 11, 2294 (2020).
 55. Gao, L. et al. Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. *Nature communications* 11, 3355 (2020).
 56. Zhang, Z.M. et al. Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature* 554, 387-391 (2018).
 57. Zhou, L. et al. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J Mol Biol* 321, 591-599 (2002).
 58. Fang, J. et al. Substrate deformation regulates DRM2-mediated DNA methylation in plants. *Science advances* 7 (2021).
 59. Pradhan, S., Bacolla, A., Wells, R.D. & Roberts, R.J. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J Biol Chem* 274, 33002-33010 (1999).
 60. Adam, S. et al. DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. *Nature communications* 11, 3723 (2020).
 61. Ye, F. et al. Biochemical Studies and Molecular Dynamic Simulations Reveal the Molecular Basis of Conformational Changes in DNA Methyltransferase-1. *ACS chemical biology* 13, 772-781 (2018).
 62. Aramayo, R. & Selker, E.U. *Neurospora crassa*, a model system for epigenetics research. *Cold Spring Harbor perspectives in biology* 5, a017921 (2013).
 63. Larson, A.G. et al. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* 547, 236-240 (2017).
 64. Canzio, D. et al. Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol Cell* 41, 67-81 (2011).
 65. Canzio, D. et al. A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. *Nature* 496, 377-381 (2013).
 66. Machida, S. et al. Structural Basis of Heterochromatin Formation by Human HP1. *Mol Cell* 69, 385-397 e388 (2018).
 67. Jackson, J.P., Lindroth, A.M., Cao, X. & Jacobsen, S.E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* 416, 556-560 (2002).
 68. Fuks, F., Hurd, P.J., Deplus, R. & Kouzarides, T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res* 31, 2305-2312 (2003).

69. Smallwood, A., Esteve, P.O., Pradhan, S. & Carey, M. Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev* 21, 1169-1178 (2007).
70. Geiman, T.M. et al. DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. *Biochemical and biophysical research communications* 318, 544-555 (2004).
71. Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D. & Grewal, S.I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* 292, 110-113 (2001).
72. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature methods* 14, 290-296 (2017).
73. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods* 16, 1153-1160 (2019).
74. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66, 213-221 (2010).
75. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583-589 (2021).
76. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25, 1605-1612 (2004).
77. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132 (2004).
78. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33, 2302-2309 (2005).

Chapter 3

Structural insight into the substrate binding of the *Tetrahymena* AMT complex via an inhibitor-trapped state

Abstract

N6-adenine (6mA) DNA methylation plays an important role in gene regulation and genome stability. The 6mA methylation in *Tetrahymena thermophila* is mainly mediated by the AMT complex, comprised of the AMT1, AMT7, AMTP1 and AMTP2 subunits. To date, how this complex assembles on the DNA substrate remains elusive. Here we report the structure of the AMT complex bound to the OCR protein from bacteriophage T7, mimicking the AMT-DNA encounter complex. The AMT1-AMT7 heterodimer approaches OCR from one side, while the AMTP1 N-terminal domain, assuming a homeodomain fold, binds to OCR from the other side, resulting in a saddle-shaped architecture reminiscent of what was observed for prokaryotic 6mA writers. Mutation of the AMT1, AMT7 and AMTP1 residues on the OCR-contact points led to impaired DNA methylation activity to various extents, supporting a role for these residues in DNA binding. Furthermore, structural comparison of the AMT1-AMT7 subunits with the evolutionarily related METTL3-METTL14 and AMT1-AMT6 complexes reveal sequence conservation and divergence on the region corresponding to the OCR-binding site, shedding light onto the substrate binding of the latter two complexes. Together, this study supports a model in which the AMT complex undergoes a substrate binding-

induced open-to-closed conformational transition, with implications in its substrate binding and processive 6mA methylation.

Introduction

DNA methylation is an epigenetic mechanism critically influencing chromatin structure and function. The most well-known DNA methylation occurs at the C-5 position of cytosines, which functions to regulate gene expression^{87,88}, X-chromosome inactivation⁸⁹, transposon silencing^{4,5}, and genome stability⁶. In addition, DNA methylation at N6-adenine (6mA) plays a role in modulating genome function in both prokaryotes and eukaryotes^{7,8}. Recent studies have further linked 6mA DNA methylation in protists, green algae and basal fungi, occurring predominantly in the context of ApT dinucleotides, to epigenetic regulation due to its association with transcriptionally active domains⁹⁻¹² and epigenetic marks H2A.Z and H3K4me3^{13,14}. However, in contrast to the writers of C-5 DNA methylation that have been well characterized¹⁵⁻²⁰, the structure and mechanism of the writers for 6mA DNA methylation in eukaryotes remains elusive.

In unicellular eukaryote *Tetrahymena thermophila*, 6mA DNA methylation is mainly maintained by the highly processive AMT complex comprised of AMT1 methyltransferase (MTase), and regulatory subunits AMT7, AMTP1, and AMTP2 (a.k.a. MTA1, AMT9, p1 and p2, respectively, in the same organism)^{11,12}. In addition, a variant of the AMT complex, in which AMT7 is replaced by a shorter homologue AMT6, plays a role in maintenance of 6mA methylation in a replication-coupled manner¹⁴. Previous structural and biochemical characterizations of the AMT complex reveal that AMT1,

AMT7, AMTP1, and AMTP2 form a heterotetramer in a stoichiometry of 1:1:1:1^{21,22}. Among these, the catalytic AMT1 subunit harbors a classic DPPW catalytic motif, stacked with the regulatory AMT7 subunit in a manner resembling the heterodimeric RNA adenine methyltransferase METTL3-METTL14 complex^{23–25}. Furthermore, a long N-terminal helix of AMT1 bridges the AMT1-AMT7 heterodimer with the AMTP1 and AMTP2 subunits through interactions with AMTP2 and C-terminal helix (CH) of AMTP1. It has been established that the AMT complex-mediated 6mA methylation requires the presence of all four subunits, with the homeodomain-like DNA-binding proteins AMTP1 and AMTP2 critically regulating substrate binding^{21,22}. However, due to lack of structural information on substrate recognition, the molecular mechanism by which the four subunits of AMT complex orchestrate 6mA methylation remains elusive. To elucidate the mechanism of the substrate binding by the AMT complex, we resorted to the Overcoming Classical Restriction (OCR) protein from bacteriophage T7, a mimic of B-form DNA²⁶, to stabilize the potential DNA-binding surfaces of AMT1, AMT7, AMTP1, and AMTP2. The OCR protein binds to the AMT complex tightly, leading to inhibition of the latter's DNA methylation activity. On this basis, we solved the single-particle cryo-EM structure of AMT1-AMT7-AMTP1-AMTP2-OCR complex. The cryo-EM structure of the AMT complex reveals saddle-shaped architecture, in which the AMT1-AMT7 heterodimer is bridged with the AMTP1 and AMTP2 subunits via the AMT1 N-terminal helix (NH). Strikingly, the AMT1-AMT7 subunits are joined by the N-terminal domain of AMTP1, which is otherwise structurally disordered in the apo form of the AMT complex, to flank the conformationally dynamic OCR, resulting in a closed

conformation reminiscent of what was observed for many other 6mA writers in prokaryotes. Our combined structural and biochemical analysis confirmed the OCR-binding surface as potential DNA-binding sites and shed light onto the substrate-binding mechanism of the evolutionarily related METTL3-METTL14 complex. Together, this study supports the notion that the AMT complex undergoes an open-to-closed conformational transition upon substrate binding, with important implications in the mechanism of the AMT complex-mediated 6mA DNA methylation.

Materials and Methods

Plasmid

The genes for full-length *Tetrahymena thermophila* AMT1 and AMT7 were inserted in tandem into an in-house bacterial expression vector, in which the AMT1 was preceded by an N-terminal His₆-MBP tag and a TEV cleavage site. Both DNA fragments encoding full-length *Tetrahymena thermophila* AMTP1 and AMTP2 were cloned into a modified pRSF vector preceded by an N-terminal His₆-SUMO tag and ULP1 (ubiquitin-like protease 1) cleavage site. The DNA fragment encoding *Escherichia* phage T7 OCR (residues 1-117; uniport number. P03775) was chemically synthesized by Integrated DNA Technologies and cloned into the pRSF vector.

Protein expression and purification

The expression plasmids for AMT1, AMT7, and AMTP2 were co-transformed into BL21 (DE3) RIL cells strain (Agilent Technologies). The cells were initially grown at 37 °C and were shifted to 16 °C when OD₆₀₀ reached 1.0. The cells were induced by the addition of 0.13 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and continued to

grow overnight. The cells were lysed in buffer containing 50 mM Tris-HCl (pH 8.0), 1 M NaCl, 25 mM imidazole, 10% glycerol, 10 µg/mL DNase I, and 1 mM PMSF. The fusion proteins were purified using a Ni-NTA affinity column. The AMT1-AMT7-AMTP2 complex was then treated with ULP1 protease to remove the His₆-SUMO tag and TEV protease to remove the His₆-MBP tag, followed by ion-exchange chromatography on a HiTrap Heparin HP column (Cytiva) and size-exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) in 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5% glycerol, and 5 mM DTT. Purified protein samples were stored at -80 °C for further use.

The AMTP1 and OCR proteins were purified in a similar manner to the above, except that the cells were shifted to 16 °C when OD₆₀₀ reached 0.8. After cell lysis, the AMTP1 or OCR protein was purified using a Ni-NTA affinity column, followed by incubation with ULP1 protease. To remove the His₆-SUMO tag, the AMTP1 protein was further purified by ion-exchange chromatography on a HiTrap Heparin HP column (Cytiva), and the OCR protein was purified using a HiTrap QHP Sepharose column (Cytiva). The tag-free AMTP1 or OCR protein was finally purified by size exclusion chromatography on a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) in 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5% glycerol and 5 mM DTT. Purified protein samples were stored at -80 °C before use. The mutants of AMT1, AMT7 and AMTP1 were generated by site-directed mutagenesis and purified in the same manner as wild-type (WT) proteins.

Radioactivity-based DNA methylation assay

To assess the OCR-mediated inhibition of the AMT complex, a synthesized 12-mer DNA duplex containing a single hemimethylated ApT site (Upper strand: 5'-GCAAGXTCAACG-3', X=6-methyladenine; Lower strand: 5'-CGTTGATCTTGC -3') was used as DNA substrate. The 20- μ L reaction mixture contained 1 μ M AMT1-AMT7-AMTP1-AMTP2 complex, 1 μ M DNA and 0.56 μ M S-adenosyl-L-[methyl- 3 H]methionine with a specific activity of 18 Ci/mmol (PerkinElmer) in buffer containing 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 200 μ g/mL BSA, 0.05% β -mercaptoethanol, and 5% glycerol in the presence or absence of 2 μ M OCR. The reaction was performed at 30 $^{\circ}$ C for 30 minutes before being quenched by 5 μ L of 10 mM cold SAM. Eight μ L of the mixtures were loaded onto Hybond N nylon membrane (GE Healthcare) and dried out at room temperature, followed by washing with 0.2 M ammonium bicarbonate (pH 8.2) twice, deionized water once, and 95% ethanol once. The dry membrane was then transferred into a vial containing 3-mL scintillation buffer (Fisher) and read by a Beckman LS6500 counter. The assays were performed in triplicate.

Luminescence-based DNA methylation assay

For mutational analysis of the AMT complex, the MTase-GloTM Methyltransferase Assay kit was used to detect activity (Promega). Specifically, 1 μ M WT or mutant AMT complex was mixed with 10 μ M SAM and the aforementioned hemimethylated 12-mer DNA duplex or unmethylated, self-complementary 24-mer DNA duplex (5'-CGCGATCGATGATCATCGATCGCG-3') in a buffer containing 20 mM Tris-HCl (pH 8.0), 50 mM NaCl, 1 mM EDTA, 3 mM MgCl₂, 100 μ g/mL BSA, and 1 mM DTT at 30 $^{\circ}$ C. The reaction lasted for 30 minutes before being quenched by 0.1% trifluoroacetic

acid. Subsequently, the reaction mixture was first mixed with 2 μ L 6 $\times$ MTase-GloTM reagent at room temperature for 30 minutes and then with 12 μ L MTase-GloTM detection solution for 30 minutes. The luminescence of the samples was measured in an opaque white 384-well microwell plate (Revvity) by CLARIOstar Plus (BMG LABTECH) at 25 °C.

Analytical size-exclusion chromatography

A 400- μ L sample containing 18 μ M AMT1-AMT7-AMTP1-AMTP2 complex, 36 μ M OCR, or the AMT complex-OCR mixture was loaded onto Superdex 200 increase 10/300 GL column (GE Healthcare) pre-equilibrated with 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5% glycerol, and 5 mM DTT. The proteins in each of the elution fractions were visualized by the SDS-PAGE imaging.

Cryo-EM sample preparation and data collection

The concentration of the AMT1-AMT7-AMTP1-AMTP2-OCR complex was adjusted to 1.1 mg/mL in a buffer containing 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM DTT, and 2.5% Glycerol. Three μ L of the complex sample was applied to Quantifoil® (1.2/1.3) grids, which were glow discharged at 15 mA for 90 s. The grids were blotted for 7 s at 95% humidity with blot force 2 and plunge frozen using a Vitrobot Mark IV (Thermo Fisher).

The cryo-EM grids were imaged on a 300 keV Titan Krios cryo-electron microscope equipped with a K3 camera (Gatan) by the National Center for CryoEM Access and Training (NCCAT). Micrographs were collected at a magnification of $\times 81,000$ with a

super-resolution pixel size of 0.53 Å. The exposures were set with a total dose of $\sim 54 \text{ e}^-/\text{\AA}^2$ over 50 frames and a defocus range of -0.4 μm to -3.1 μm .

Cryo-EM data processing

In total 5137 movies were processed using the CryoSPARC software package (v4.01)²⁷. All movies are subject to 2x2 binning, patch motion correction and patch-based contrast transfer function estimation. Subsequently, 2.6 million particles were picked using the Topaz program²⁸, and extracted with a downsampled pixel size of 2.12 Å per pixel. After several rounds of 2D classification, 0.2 million particles were selected for 3D ab initio reconstruction and subsequent heterogeneous refinement. The particles with visible features of secondary and tertiary structures were re-extracted with pixel size of 1.06 Å per pixel. The final set of particles were subject to removal of duplicates and non-uniform refinement, resulting in a final resolution of 3.59 Å, as given by the Fourier shell correlation criterion (FSC 0.143).

Model building and refinement

The structural model for AMT1-AMT7-AMTP2 complex was derived from that of the reported apo-state AMT complex (PDB 7YI8)²². The structural models for AMTP1 and OCR were generated by AlphaFold3 method²⁹. Fitting the models into the density map was carried out using UCSF Chimera (v1.16)³⁰ and ChimeraX (v1.6.1)³¹. The final structural model was obtained after iterative model building in Coot (v0.8.9.1)³², followed by real-space refinement in Phenix³³. In addition, the structural model for the AMT-bound B-form DNA was generated and aligned with the density corresponding to the OCR protein in Coot.

Results

The OCR protein inhibits the AMT complex-mediated 6mA DNA methylation

A previous structural study of the AMT complex, despite the presence of DNA substrate, failed to observe the density of the DNA²², suggesting that the AMT-DNA association is highly dynamic. To overcome this challenge, we resorted to the OCR protein from T7 bacteriophage for stabilizing the substrate-binding conformation of the AMT complex (Fig. 1a-c). OCR reportedly exists in a homodimeric form mimicking a slightly bent 20-base pair B-form DNA and has been shown to bind and inhibit DNA-templated enzymes^{34,35}. To test the interaction between the AMT complex and OCR, we performed size-exclusion chromatography analysis for the AMT complex and OCR, alone or in mixture. OCR co-migrated with the AMT complex on a Superdex 200 increase 10/300 GL column, indicative of stable association (Fig. 1d, e and Supplementary Fig. 1a, b). To evaluate whether the presence of OCR affects the AMT complex-mediated 6mA DNA methylation, we performed in vitro DNA methylation assay for the AMT complex on a 12-mer DNA duplex containing a single hemimethylated ApT site in the presence or absence of OCR. Consistent with previous observations^{11–13,21,22}, the AMT complex is highly active on the hemimethylated ApT DNA (Fig. 1f). However, the presence of OCR in a 2:1 molar excess reduced the DNA methylation efficiency of the AMT complex by ~100-fold (Fig. 1f), indicative of strong inhibition. These data establish OCR as a DNA-mimicking inhibitor for the AMT complex-mediated 6mA DNA methylation.

Structural overview of the OCR-bound AMT complex

Next, we solved the cryo-EM structure of the OCR-bound AMT complex at an overall 3.59-Å resolution (Fig. 2a-d, Supplementary Fig. 2 and Supplementary Table 1). Structural analysis of the OCR-bound AMT complex reveals that the AMT complex is comprised of one copy of AMT1, AMT7, AMTP1, and AMTP2 each (Fig. 2a-d and Supplementary Fig. 3a-i), consistent with previous observations^{21,22}. We were able to trace the density for AMT1 residues 128-216 and 227-372, AMT7 residues 99-114, 118-156, 162-228, 319-450, AMTP1 residues 16-128 and 154-184, and AMTP2 residues 11-132 (Fig. 2c, d and Supplementary Fig. 3a-i). The AMT1-bound S-adenosyl-homocysteine (SAH), by-product of the cofactor S-adenosyl-methionine (SAM), was also defined (Supplementary Fig. 3j).

The density for the OCR molecules bound to the AMT complex is rather weak and failed to be improved even through local refinement or focused classification, presumably due to conformational flexibility, which prevents us from accurate modeling (Fig. 2a, b and Supplementary Fig. 3e). Nevertheless, it is apparent that the OCR molecules are sandwiched by the N-terminal domain (ND) of AMTP1 from one side and the AMT1 MTase and AMT7 MTase-like domains on the other side (Fig. 2a-d). As observed previously^{21,22}, the MTase domain of AMT1 associates with the MTase-like domain of AMT7 in antiparallel, resulting in a butterfly-like fold (Fig. 2a-c). The AMTP1 protein is comprised of an N-terminal domain (ND) dominated by a helical fold and a structurally independent C-terminal helix (CH) (Fig. 2a-c). The AMTP2 protein is comprised of an ND in a helix-loop-helix conformation and a C-terminal domain (CD) dominated by a four-helix bundle (Fig. 2a-c). Notably, the AMT1 N-terminal helix (NH) engages in

helical contacts with both the AMTP1 CH and the AMTP2 CD, thereby bridging the AMT1-AMT7 heterodimer with AMTP1 and AMTP2 (Fig. 2a-d). Meanwhile, the ND of AMTP2 is docked onto the surface region formed by the beginning α -helix and β -strand of the AMT1 MTase domain as well as the preceding loop (Fig. 2a-c).

Structural superposition of the OCR-bound AMT complex and the previously reported apo-form AMT complex (PDB 7YI8)²² gave a root-mean-square deviation (RMSD) of 1.2 Å over 531 aligned C α atoms, indicative of high structural similarity (Fig. 2e and Supplementary Fig. 3k). The most notable difference between the two complexes lies in the AMTP1 ND domain, which is disordered in the apo-form AMT complex but is well defined in the OCR-bound form (Fig. 2e and Supplementary Fig. 3k), suggesting an OCR binding-induced structural stabilization effect. The conformational difference between the apo and OCR-bound AMT complex was further propagated to the AMTP1 CH (residues E154-Q184) (Fig. 2e and Supplementary Fig. 3k): upon the OCR binding, the N- and C-termini of the AMTP1 CH shifted toward the AMTP1 ND by ~ 4 and ~ 8 Å, respectively (Fig. 2e and Supplementary Fig. 3k).

Mapping of the potential DNA-binding interface of the AMT1-AMT7 heterodimer

Analysis of surface electrostatic potential of the OCR-bound AMT complex reveals that the OCR molecules are walled by the positively charged surfaces of the AMT1 MTase domain, the AMT7 MTase-like domain, and the AMTP1 ND domain (Fig. 3a-c). On one side, a cluster of residues at the AMT1-AMT7 interface, including AMT1 K280, K289 and H291 and AMT7 R379 and R382, form a contact surface with the OCR (Fig. 3b). Proximal to this surface are the residues (Y227, R357 and R358) around the SAH-

binding site of AMT1, formed by two loop segments, namely Gate loops 1 and 2^{21,22} (Fig. 3b). On the other side, AMTP1 residues K44, K46, K57, R68, K101, K114 and K118 form another contact surface for the OCR (Fig. 3c).

The fact that the OCR protein serves as a mimic for the B-form DNA suggests that the OCR-contact sites of the AMT complex constitute its DNA-binding interface. To test this possibility, we selected the OCR-contact residues of AMT1, AMT7, and AMTP1 for mutagenesis and performed in vitro DNA methylation assays on substrates containing either a favorable hemimethylated ApT site or unfavorable, unmethylated ApT sites (Fig. 3d-g). For the hemimethylated substrate, introducing the AMT1 K289 mutation led to ~3-fold reduction of the DNA methylation activity, while introducing the AMT1 Y227A, K280A, H291A, R357A or R358A, or the AMT7 R379 or R382 mutation, largely abolished the AMT activity (Fig. 3d). Likewise, introducing the AMT1 and AMT7 mutations also led to substantial reduction of the AMT activity on the unmethylated substrate (Fig. 3e). For AMTP1, introducing the K57A and K101A mutations led to 2-4-fold lower methylation efficiency of the AMT complex on the hemimethylated substrate, and 3-6-fold lower methylation efficiency on the unmethylated substrate, suggesting an important role for these two sites in DNA binding (Fig. 3f, g). On the other hand, a less pronounced effect was observed for the AMTP1 K44A, K46A, R68A, K114A or K118A mutation, which led to slight or no activity change of the AMT complex on the hemimethylated substrate (Fig. 3f). Nevertheless, it is apparent that these mutations significantly impair the AMT-mediated methylation on the less favorable, unmethylated substrates, supporting a role for these residues in fine-tuning the activity of the AMT

complex (Fig. 3g). Sequence analysis of AMT1, AMT7 and AMTP1 across various species further demonstrated that many of these potential DNA-binding sites, such as AMT1 Y227, K280, K289, H291, R357 and R358, AMT7 R379 and R382, and AMTP1 K44, K46, K57, R68 and K101, are evolutionarily conserved (Supplementary Fig. 4 and Supplementary Fig. 5). Together, these data support the notion that the OCR-contact residues provide a DNA-binding interface underpinning the AMT-mediated 6mA DNA methylation.

Comparative structural analysis of AMTP1 in AMT-mediated 6mA methylation

AMTP1 belongs to a broad family of homeodomains, which often interact with cognate DNAs to regulate cell- and tissue-specific gene expression³⁶. Through structural homology search using the DALI program³⁷, we identified a few structural homologues for AMTP1, including Double homeobox protein 4 (DUX4) (PDB 6E8C)³⁸, Myb-containing domain (MCD3) of Double-strand telomeric DNA-binding proteins 2 (TEBP-2) (PDB 8U8L)³⁹, Antennapedia homeodomain-DNA complex (PDB 9ANT)⁴⁰, and TEA domain family transcription factor 4 (TEAD4) (PDB 5GZB)⁴¹ (Supplementary Fig. 6a). Notably, these proteins all harbor a homeodomain, employing a helix-turn-helix (HTH) motif to access the major groove of the DNA molecules (Supplementary Fig. 6b). It's worth noting that several OCR-contact sites for AMTP1 are also located on the corresponding region, spanning from α 2-helix to α 3-helix of the ND, such as K44, K46, K57 and R68, in addition to those from the neighboring β 1-strand (K101) and α 5-helix (K114 and K118) (Supplementary Fig. 5 and Supplementary Fig. 6c). These data provide further insight into the DNA-binding sites of AMTP1.

Structural comparison of the AMT complex with prokaryotic 6mA writers

To illustrate how the AMT complex-substrate binding is linked to prokaryotic adenine DNA methyltransferases, we generated a structural model of the AMT-DNA complex, based on the density corresponding to the OCR protein in the AMT-OCR complex. Due to the weak density of OCR, the DNA molecule was positioned in a random orientation. Nevertheless, the proximity between the DNA backbone and the AMTP1, AMT1 and AMT7 subunits agrees well with the AMT-OCR interactions (Fig. 4a, b). We then performed structural comparison of the AMT-DNA complex with several prokaryotic 6mA writers, *E. coli* Type III restriction-modification enzyme EcoP15I-DNA complex (PDB 4ZCF)⁴², *Caulobacter crescentus* cell cycle-regulated DNA methyltransferase (CcrM)-DNA complex (PDB 6PBD)⁴³, and *Thermus aquaticus* DNA adenine methyltransferase M.TaqI-DNA complex (PDB 1G38)⁴⁴ (Fig. 4c-h). Similar to the AMT complex, the EcoP15I complex employs multiple subunits for DNA binding, with the catalytic ModB subunit binding to the DNA from the minor groove and the target recognition domain (TRD) of ModA subunit (TRD_A) binding to the major groove for sequence-specific recognition (Fig. 4c, d). Likewise, the CcrM complex involves two copies of CcrM molecules for DNA binding, with one molecule approaching the minor groove for catalysis and the TRD from the other CcrM subunit (TRD') approaching the major groove (Fig. 4e, f). In contrast, the M.TaqI complex relies on two subdomains of M.TaqI: the catalytic core and the TRD, to bind DNA from minor groove and major groove, respectively (Fig. 4g, h). Despite that these complexes employ distinct assembly mechanisms, they all form a deep cleft for processive DNA methylation (Fig. 4a-h),

highlighting similar yet distinct mechanisms of substrate DNA binding among the evolutionarily related 6mA DNA methyltransferases.

Structural comparison of the AMT complex with the prokaryotic 6mA writers further indicated that AMTP1 ND structurally corresponds to the major groove-binding components of the prokaryotic methyltransferase complexes, such as the TRD_A of the ModA subunit of EcoP15I, the TRD of the major groove-engaging CcrM molecule, and the TRD of *M.TaqI* (Supplementary Fig. 6d). Whether AMTP1 recognizes the target DNA in a similar manner to that of the TRD of the prokaryotic complexes warrants further investigation.

Structural insights into the substrate-binding site of the METTL3-METTL14 complex

The AMT1-AMT7 heterodimer and the RNA adenine methyltransferase METTL3-METTL14 heterodimer belong to the same MT-A70 methyltransferase family^{11,12}, with high sequence and structural similarity^{21,22,24,25}. Structural overlay of the OCR-bound AMT complex with the METTL3-METTL14 complex (PDB 5IL1)²⁴ reveals high structural similarity between the two complexes, with an RMSD of 1.13 Å over 161 aligned C α atoms (Supplementary Fig. 7a). Notably, the SAM- and target adenosine-binding sites of METTL3-METTL14 complex, formed by Gate loop 1 (METTL3 residues 396-410) and Gate loop 2 (METTL3 residues 507-515), are well aligned with the corresponding regions (residues 210-232 and 328-336, respectively) of AMT1 (Supplementary Fig. 7b). Furthermore, a number of DNA-binding sites of AMT1, such as residues Y227, K280 and R357, are conserved in METTL3 (Supplementary Fig. 7c),

suggesting that AMT1 and METTL3 may employ a similar surface platform for substrate binding. On the other hand, the AMT7 residues R379/R382-corresponding sites in METTL14 are occluded from solvent access by two α -helices (Supplementary Fig. 7a), suggesting that they are unlikely involved in substrate binding in the METTL3-METTL14 complex. These observations shed light onto the functional conservation and divergence between two evolutionarily related methyltransferases.

Structural modeling analysis of the substrate binding by the AMT1-AMT6 complex

The AMT1-AMT6 complex, in which the AMT7 subunit in AMT1-AMT7 complex is replaced by AMT6, is an alternative methyltransferase in maintaining the 6mA patterns shortly after DNA replication¹⁴. With reduced methylation activity than the AMT1-AMT7 complex in vitro, the AMT1-AMT6 complex maintains the 6mA DNA methylation in cells in a proceeding cell nuclear antigen (PCNA)-dependent manner¹⁴. To illustrate the substrate binding by the AMT1-AMT6 complex, we generated a structural model of the AMT1-AMT6 complex based on the structure of AMT1-AMT7-AMTP1-AMTP2 complex and AlphaFold3 prediction²⁹. The AMT1-AMT7 and AMT1-AMT6 complexes are structurally similar, with an RMSD of 1.77 Å over 1279 aligned Ca atoms (Supplementary Fig. 8a). Notably, the DNA-binding sites of AMT7, including R379 and R382, are also conserved in AMT6 (Supplementary Fig. 8b), suggesting a conserved DNA-binding mechanism between the two homologues. Nevertheless, AMT6 differs from AMT7 in the sequence of the disordered N-terminus (Supplementary Fig. 8b). In addition, the surface electrostatic potential of AMT6 appears less positively charged than that of AMT7, due to amino acid variation in the region near to the DNA-binding site

(Supplementary Fig. 8b-d). Whether such a difference in the DNA-binding site contributes to the differential methylation activity between the AMT1-AMT7 and AMT1-AMT6 complexes awaits further investigation.

Discussion

6mA DNA methylation in eukaryotes has gained increasing attention due to its functional implication in epigenetic regulation and genomic assembly⁴⁵. This study, through structure determination of an inhibitor-trapped state of the AMT complex from ciliated *Tetrahymena thermophila*, combined with enzymatic analysis, provides a framework for understanding how the AMT complex engages DNA substrate for 6mA maintenance methylation, with important implications in functional regulation of 6mA DNA methylation in eukaryotes.

First of all, the structure of the AMT complex with the DNA-competitive inhibitor OCR supports a model that the AMT complex undergoes an open-to-closed conformational transition upon the substrate binding (Fig. 5). Previous structural characterization of the AMT complex reveals that, in the absence of DNA, the AMT1, AMT7, AMTP2 and the C-terminal helix of AMTP1 adopts an open conformation, while the ND domain of AMTP1 is disordered^{21,22}. While these studies provide a framework for the functional assembly of the AMT complex, the lack of structural details for the ND domain of AMTP1 raises a question on its functional role in AMT-mediated 6mA DNA methylation. To overcome the challenge of the dynamic AMT-DNA interaction, we took advantage of a DNA-competitive inhibitor, the OCR protein, to stabilize the DNA-binding components of the AMT complex. As a result, we were able to capture a saddle-

shaped architecture formed by AMT1, AMT7, AMTP1, and AMTP2, in which AMTP1 and the AMT1-AMT7 heterodimer coordinately flank the OCR protein. Consistently, our in vitro enzymatic analysis reveals that mutations of the putative DNA-binding sites of AMT1, AMT7 and AMTP1 impact the AMT activity to various extents. Our comparative structural analysis further demonstrated that the ND domain of AMTP1 shares a similar functional site with other members of the homeodomain family, implying a role for this domain to target the major groove of DNA substrate. Whether AMTP1 contributes to the DNA sequence specificity of the AMT complex, as what was observed for the TRD domain of a C-5 DNA methyltransferase^{15,18}, has yet to be elucidated.

It has been well known that DNA-templated enzymes often adopt a ring or saddle-shaped structure to process DNA⁴⁶. For instance, encircling of the DNA by eukaryotic DNA polymerase processivity factor PCNA⁴⁷ and bacteriophage T4 gp45⁴⁸ allows the polymerases to move rapidly without dissociation from the template. Similarly, the restriction-modification enzyme *M.TaqI* from *Thermus aquaticus*⁴⁴, the DNA restriction enzyme *BamHI*⁴⁸, and the TATA-box binding protein (TBP)⁴⁹ all adopt a saddle conformation to slide along DNA for processive enzymatic or scanning actions. In this context, the substrate binding-induced conformational closure of the AMT complex may underpin its processive methylation along DNA^{11,12}. A detailed understanding of the enzymatic kinetics of the AMT complex awaits further investigation.

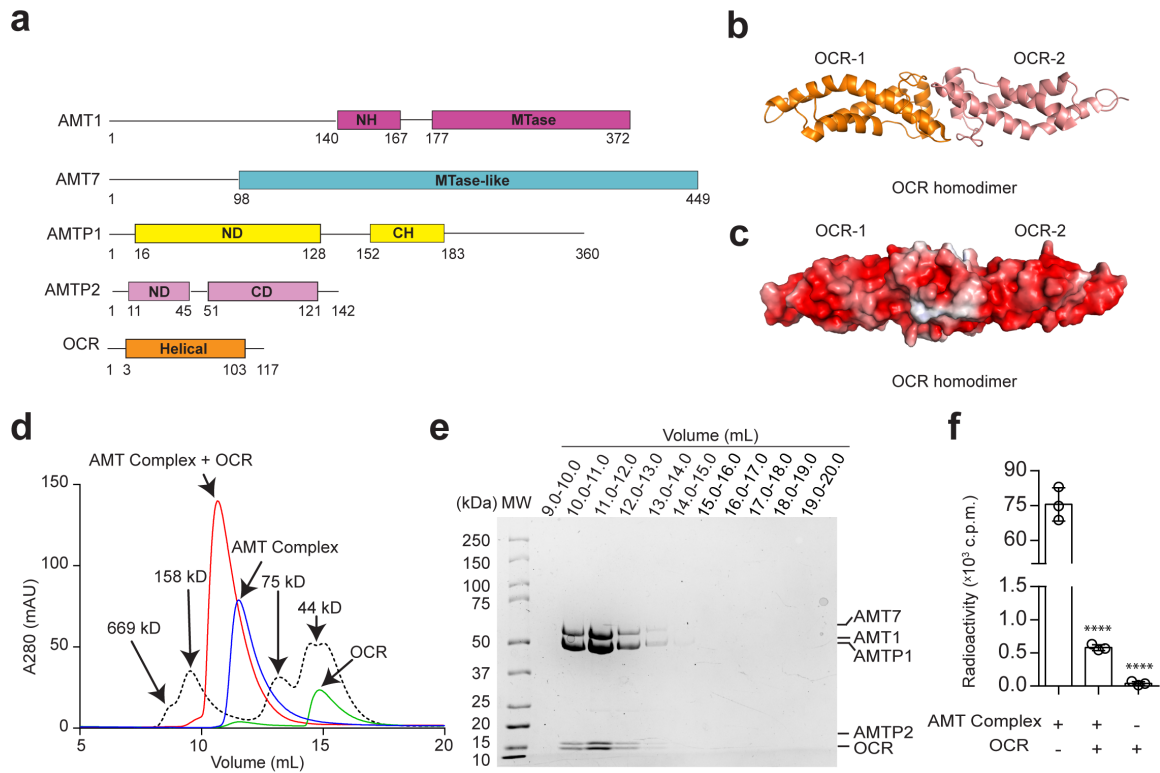


Figure 1. Biochemical analysis of the AMT-OCR complex.

(a) Domain architecture of AMT1, AMT7, AMTP1, AMTP2, and OCR. NH, N-terminal helix; ND, N-terminal domain; CH, C-terminal helix; CD, C-terminal domain. (b, c) Ribbon (b) and surface electrostatic (c) representation of OCR homodimer (PDB 8ZEK). (d) Size-exclusion chromatography analysis of the AMT complex (blue), OCR (green), or AMT-OCR mixture (red). The elution profile for the molecular weight markers: Thyroglobulin (669 kDa), Aldolase (158 kDa), Conalbumin (75 kDa), and Ovalbumin (44 kDa) is shown as dashed line, with individual molecular weight indicated on top. (e) SDS-PAGE analysis of select fractions of the AMT-OCR mixture from size-exclusion chromatography in (d). (f) In vitro DNA methylation assay of the AMT complex on a hemimethylated 12-mer DNA duplex, in the presence or absence of OCR. Data are mean $\pm$ s.d. ($n = 3$ biological replicates). The two-tailed Student's *t* test statistical analysis was performed to compare AMT complex activity in the absence vs. presence of OCR. **** $p < 0.0001$.

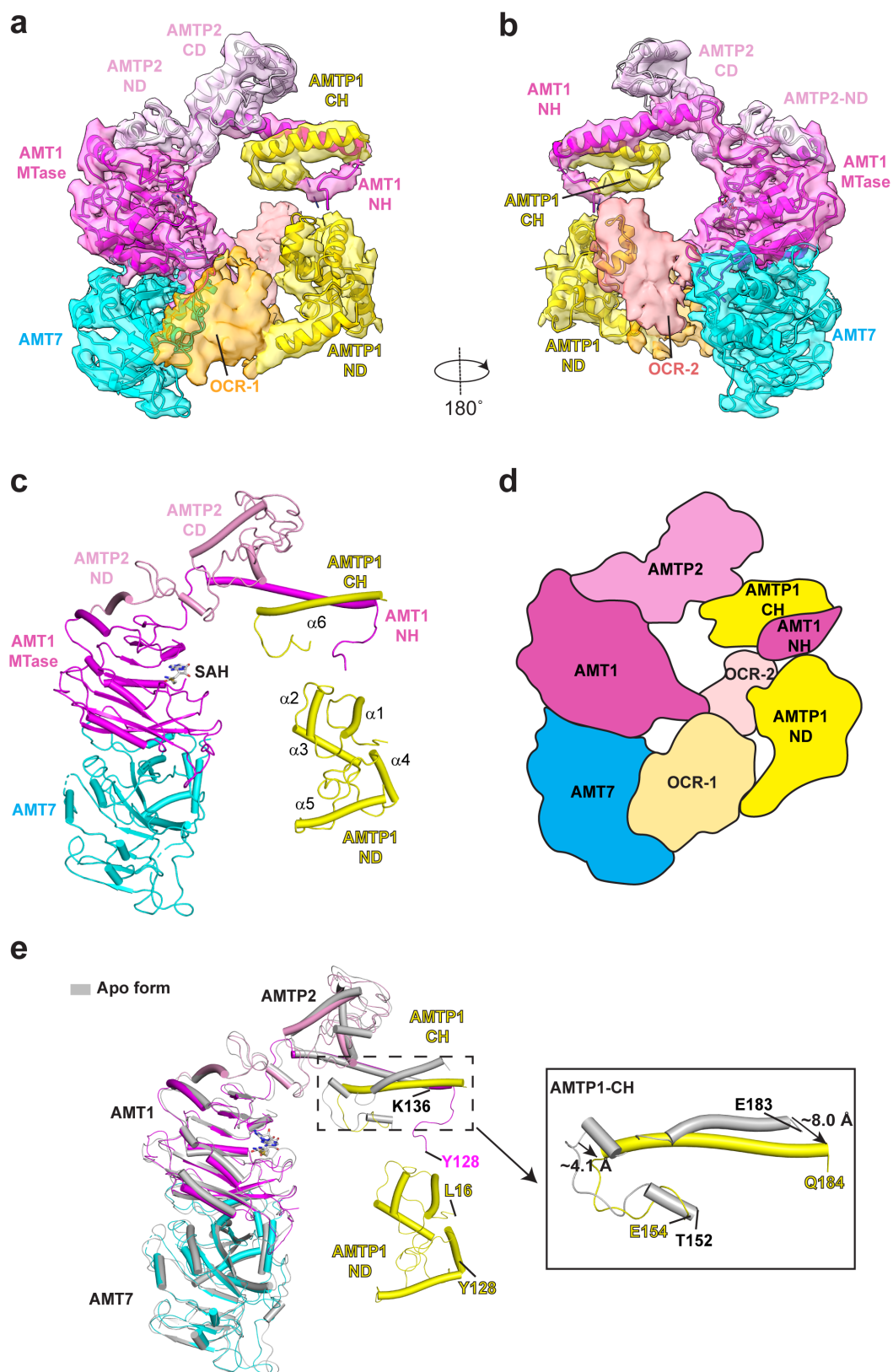


Figure 2. Cryo-EM structure of the OCR-bound AMT complex.

(a, b) Two opposite views of the cryo-EM density map and atomic model of the AMT-OCR complex, with individual subunits and domains labeled. AMT1, AMT7, AMTP1, and AMTP2 are colored in magenta, cyan, yellow, and pink, respectively. The color scheme applies to subsequent figures unless otherwise indicated. The density corresponding to the OCR subunits (OCR-1 and OCR-2) are colored in orange and deep salmon, respectively. The atomic model for OCR dimer was not included in the final model due to its density being weak. (c) Atomic model of the AMT-OCR complex, with the cofactor by-product SAH molecule is shown in stick representation. (d) Schematic view of the assembly of the AMT-OCR complex. (e) Structural overlay of apo-form (PDB 7YI8) and OCR-bound AMT complex, with the conformational shift for the AMTP1 CH highlighted in the expanded view. The N- and C-termini of the AMTP1 ND and the AMTP1 CH, and the N-terminus of AMT1 are indicated.

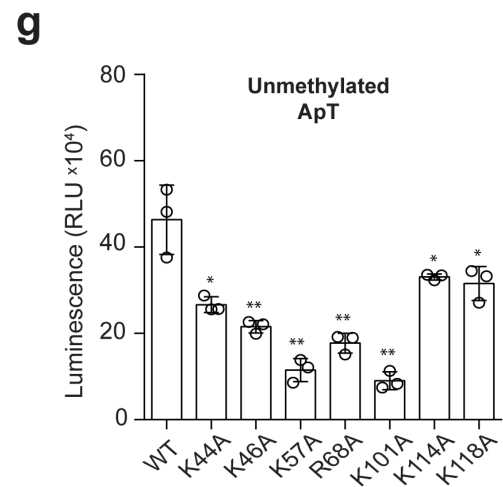
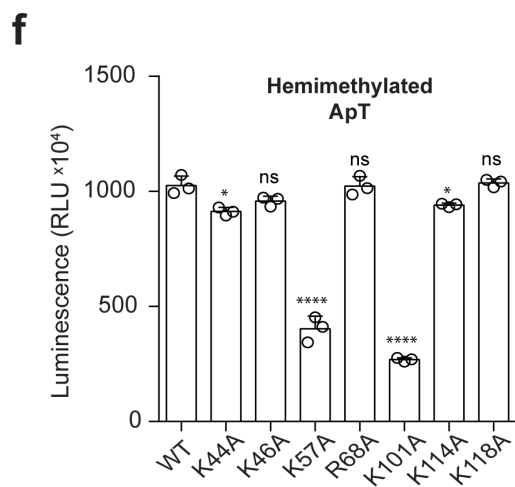
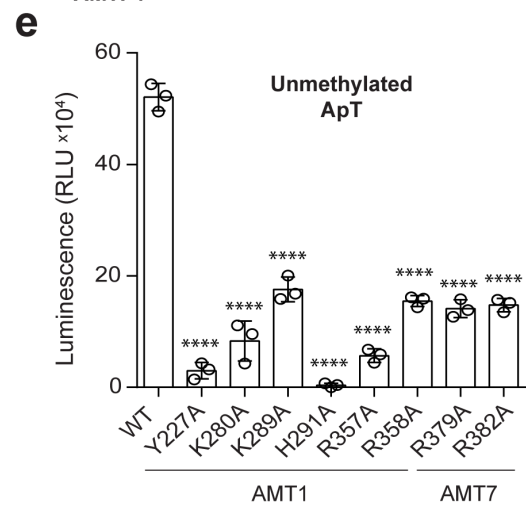
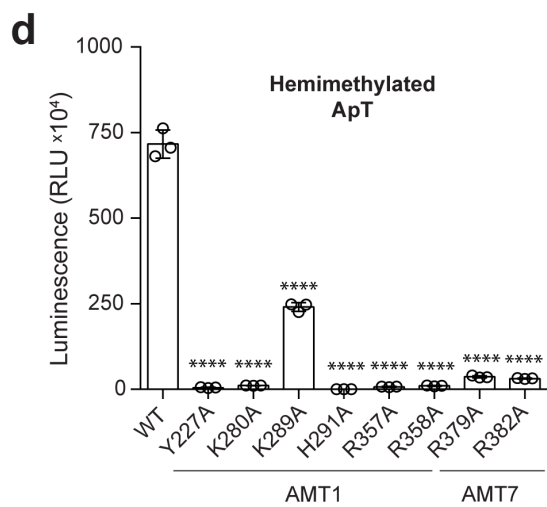
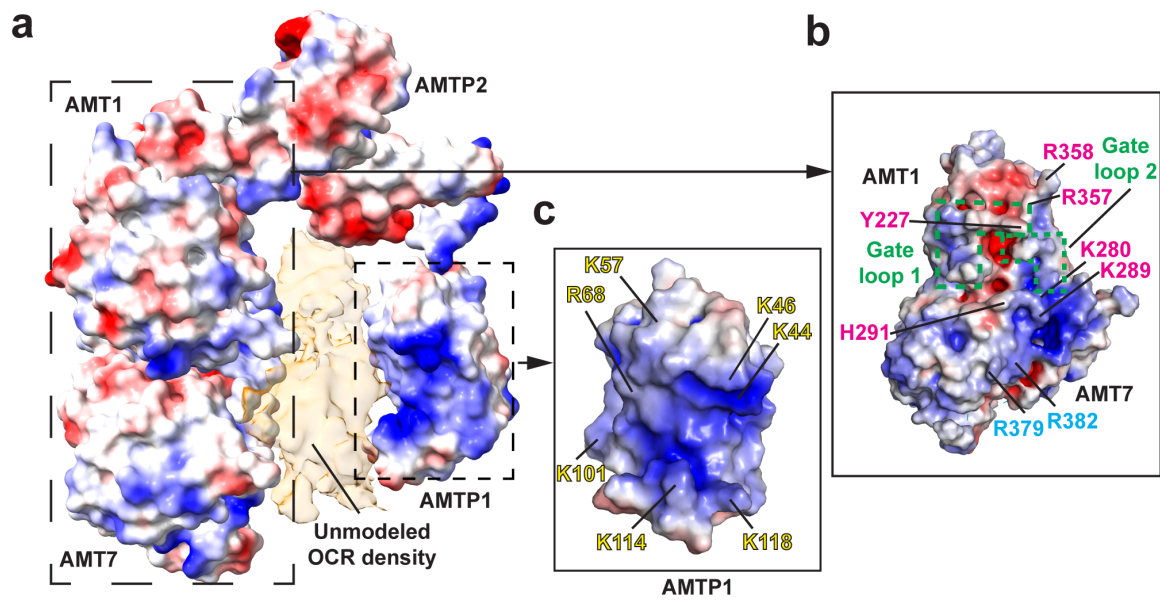


Figure 3. Structural and mutational analysis of the potential DNA-binding surface of the AMT complex.

(a-c) Electrostatic surface of the AMT complex, with the entire surface shown in (a), and the potential substrate-binding sites of the AMT1-AMT7 dimer (b) and AMTP1 (c) shown in the expanded views. Gate loops 1 and 2 in AMT1 are highlighted by green dashed lines. (d, e) In vitro DNA methylation assay of the AMT1- or AMT7-mutated AMT complex on a hemimethylated 12-mer DNA duplex (d) or unmethylated 24-mer DNA duplex (e). Data are mean $\pm$ s.d. (n = 3 biological replicates). (f, g) In vitro DNA methylation assay of the AMTP1-mutated AMT complex on a hemimethylated 12-mer DNA duplex (f) or unmethylated 24-mer DNA duplex (g). Data are mean $\pm$ s.d. (n = 3 biological replicates). The two-tailed Student's t test statistical analysis was performed for pairwise comparison of the activity of WT vs. mutant AMT complex. ns, not significant; *p<0.05, **p<0.01; ****p < 0.0001.

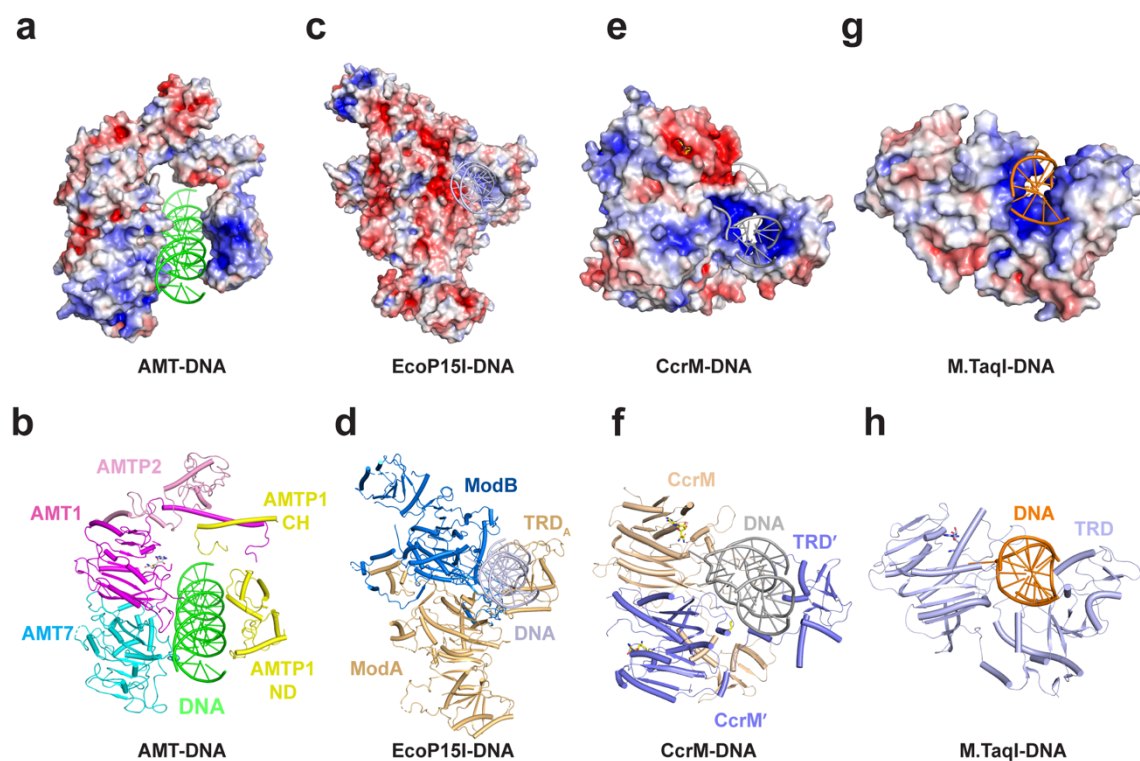


Figure 4. Structural comparison of the AMT-DNA model with the enzyme-substrate complexes of bacteria 6mA DNA methyltransferases. (a, b) Electrostatic surface (a) and ribbon representation (b) of the AMT-DNA model. (c, d) Electrostatic surface (c) and ribbon representation (d) of the EcoP15I Mod subunit-DNA (PDB 4ZCF). (e, f) Electrostatic surface (e) and ribbon representation (f) of the CcrM-DNA complex (PDB 6PBD). (g, h) Electrostatic surface (g) and ribbon representation (h) of the M.TaqI-DNA complex (PDB 1G38).

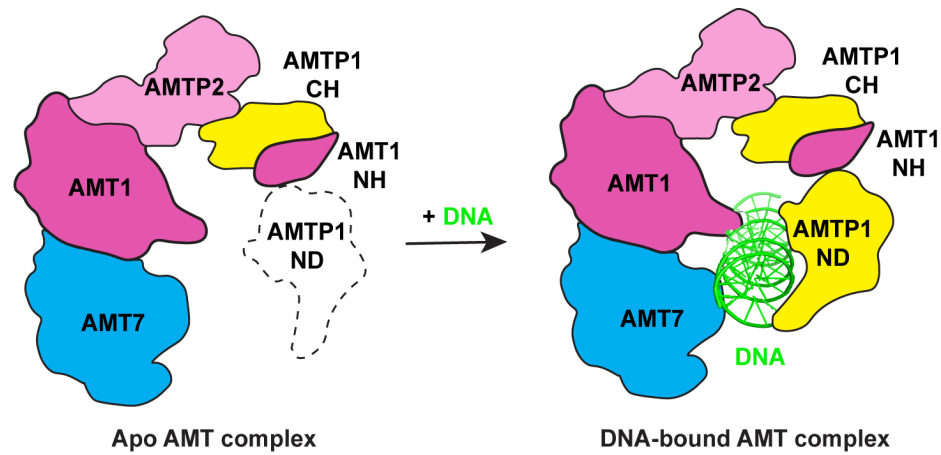
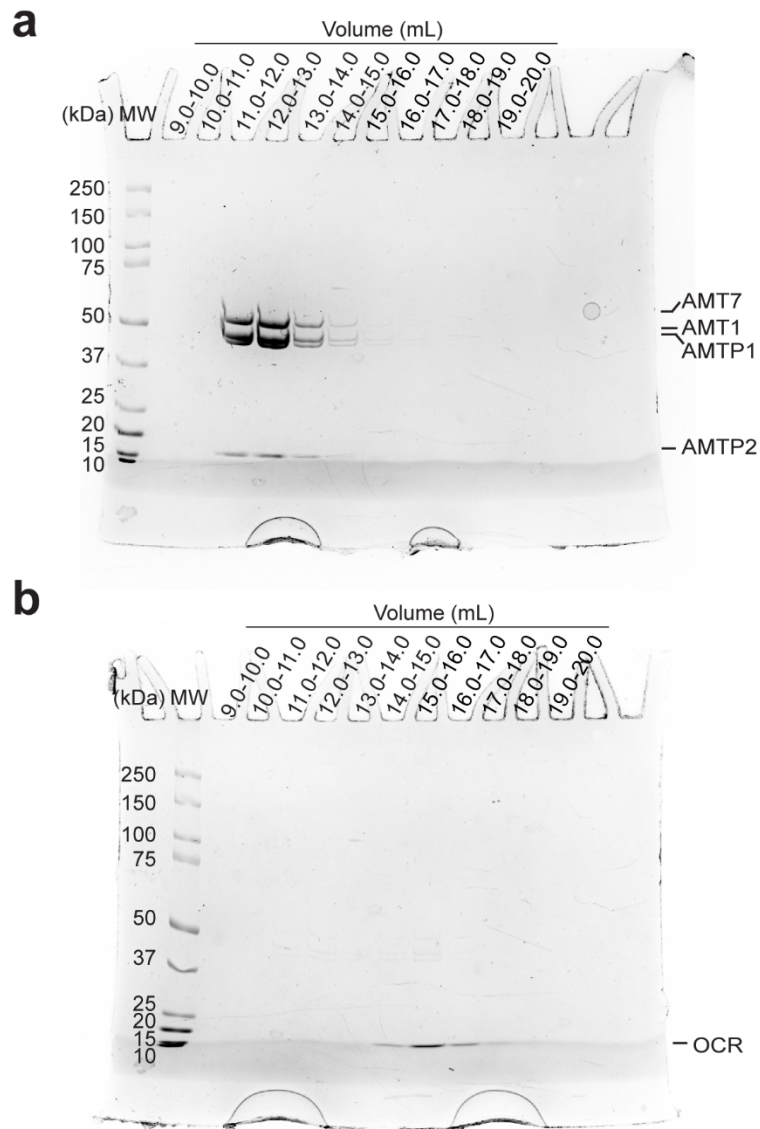


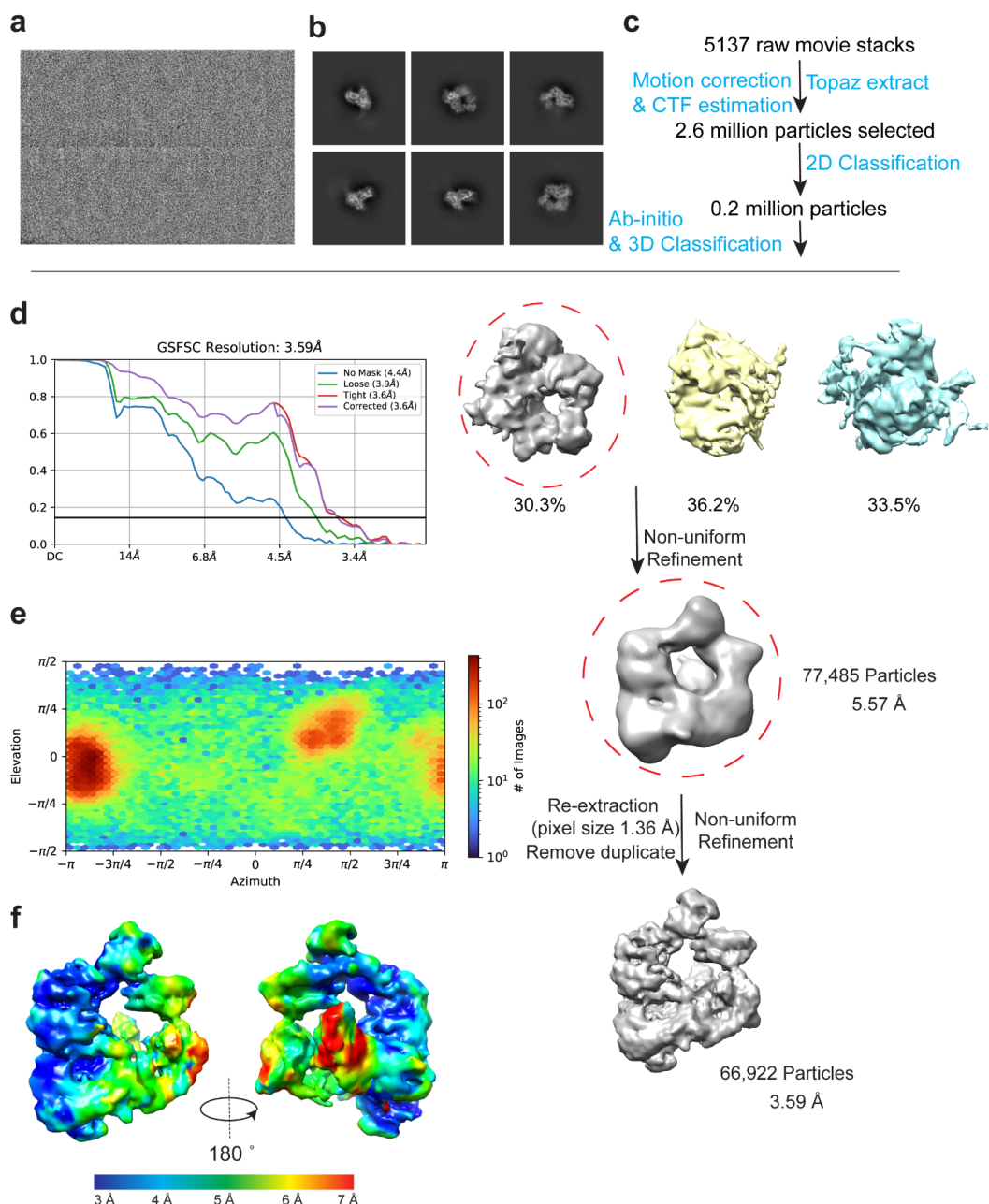
Figure 5. A working model of the AMT complex. In the absence of DNA substrate, the AMTP1 N-terminal domain is dynamically disordered, resulting in an open conformation. In the presence of DNA substrate, the AMT1-AMT7 dimer and the AMTP1 ND close up on DNA for methylation.

Supplementary Data



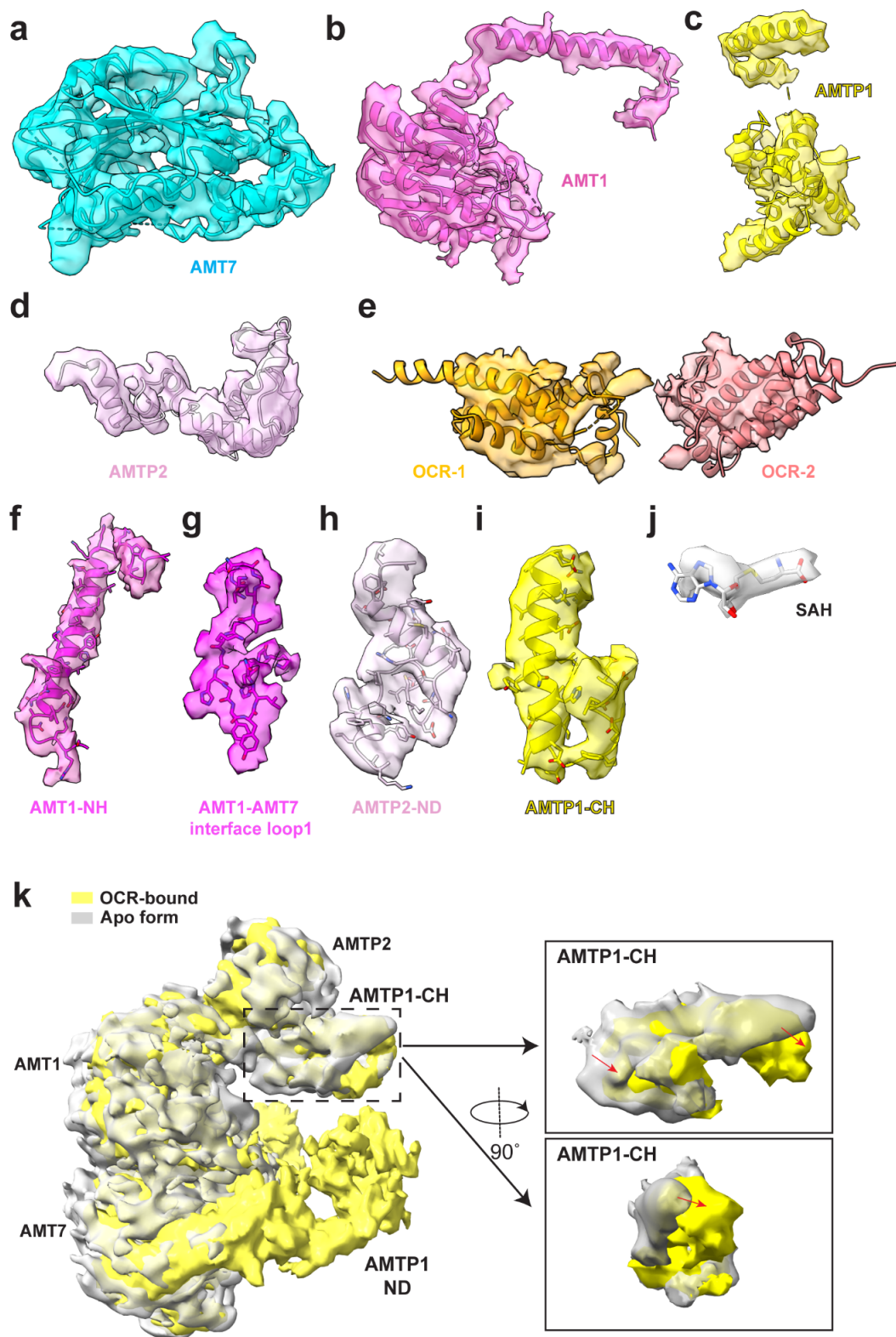
Supplementary Fig. 1 | The SDS-PAGE analysis of the AMT complex and OCR.

(a, b) SDS-PAGE images of the AMT complex (a) and OCR (b) following size-exclusion chromatography, with the elution volume of individual fractions indicated on top.



Supplementary Fig. 2 | Data processing workflow for cryo-EM reconstruction of the AMT-OCR complex.

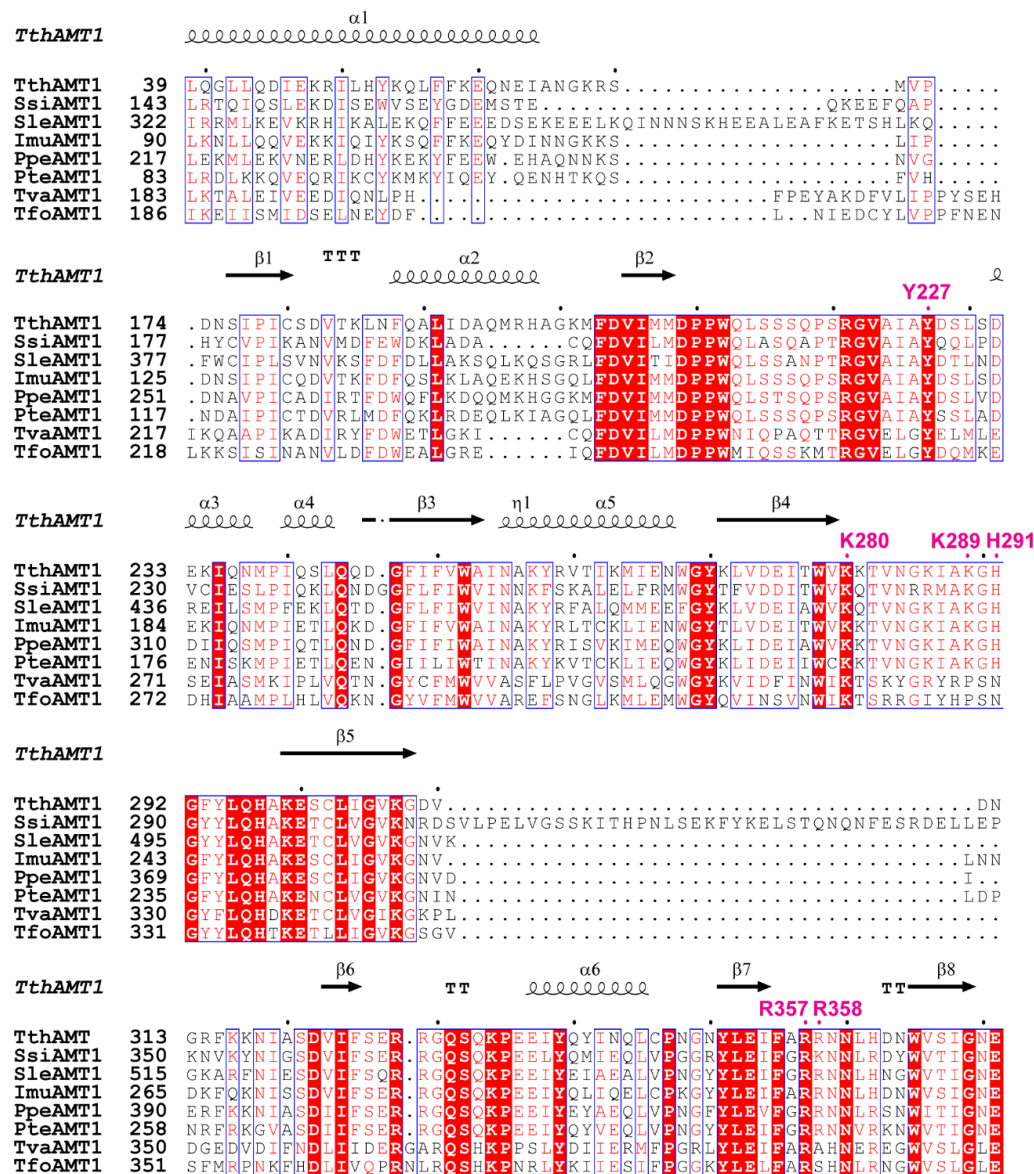
(a) A representative micrograph of the AMT-OCR complex. (b) Representative 2D classes of the AMT-OCR complex from CryoSPARC. (c) Flow chart of the cryo-EM data processing. (d) Fourier shell correlation (FSC) curve of the AMT-OCR complex map as a function of resolution using CryoSPARC output. (e) Angular distribution calculated in CryoSPARC for particle projections. Heat map shows number of particles for each viewing angle. (f) Local resolution map of the AMT-OCR complex.



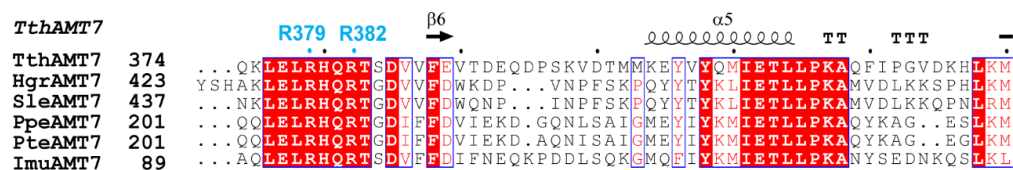
Supplementary Fig. 3 | Cryo-EM density maps for the AMT-OCR complex.

(a-j) Density map and atomic model of the AMT7 (a), AMT1 (b), AMTP1 (c), AMTP2 (d), OCR dimer (e), the AMT1 N-terminal helix (AMT1-NH) (f), the AMT1-AMT7 interface loop (residues 282-299) (g), the AMTP2 N-terminal domain (AMTP2-ND) (h), the AMTP1 C-terminal helix (AMTP1-CH) (i), and SAH molecule (j). (k) Comparison of density map between apo-form (EMD-33853) (grey) and OCR-bound AMT complex (yellow), with the conformational change of the AMTP1 CH shown in the expanded views. A low-pass filter of 5 Å was applied to the density of the apo-form for the purpose of comparison.

a

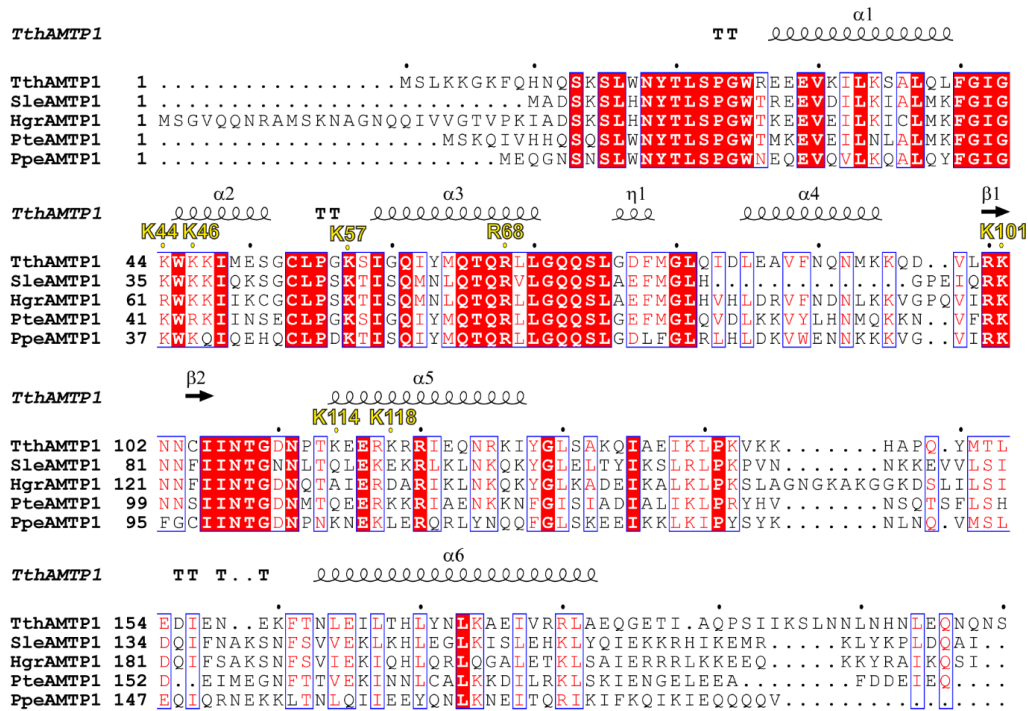


b



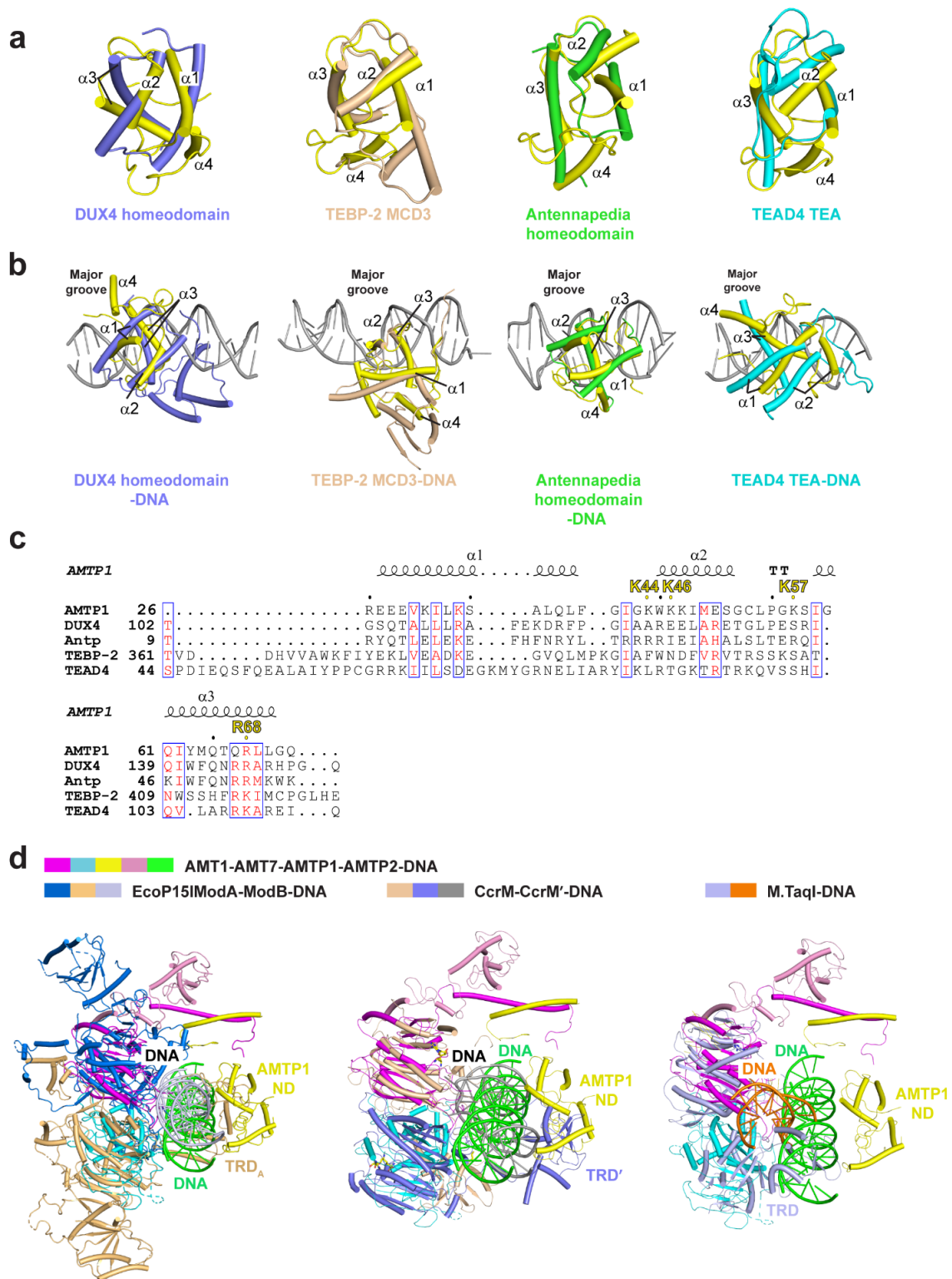
Supplementary Fig. 4 | Sequence alignment of AMT1 and AMT7 from various species.

(a) Sequence alignment of AMT1 proteins. TthAMT1: *Tetrahymena thermophila* AMT1 (XP_001032074), SsiAMT1: *Smittium simulii* AMT1 (PVU88469), SleAMT1: *Stylonychia lemnae* AMT1 (CDW73011), ImuAMT1: *Ichthyophthirius multifiliis* AMT1 (XP_004035673), PpeAMT1: *Pseudocohnilembus persalinus* AMT1 (KRX04523), PteAMT1: *Paramecium tetraurelia* AMT1 (XP_001457964), TvaAMT1: *Trichomonas vaginalis* G3 AMT1 (EAX86932), TfoAMT1: *Tritrichomonas foetus* AMT1 (OHT14409). Identical or similar residues are boxed and colored red. The secondary structures corresponding to TthAMT1 are indicated on top. The OCR-contact sites of AMT1 are indicated in magenta. **(b)** Sequence alignment of AMT7 proteins. TthAMT7: *Tetrahymena thermophila* AMT7 (7YI8_A), HgrAMT7: *Halteria grandinella* AMT7 (TNV81034), SleAMT7: *Stylonychia lemnae* AMT7 (CDW82373), PpeAMT7: *Paramecium pentaurelia* AMT7 (CAD8164721), PteAMT7: *Paramecium tetraurelia* AMT7 (XP_001442122), ImuAMT7: *Ichthyophthirius multifiliis* AMT7 (XP_004030651), TthAMT6: *Tetrahymena thermophila* AMT6 (XP_001010335). Identical or similar residues are boxed and colored red. Completely conserved residues are shaded in red. The secondary structures corresponding to TthAMT7 are indicated on top. The OCR-contact sites of AMT7 are indicated in cyan.



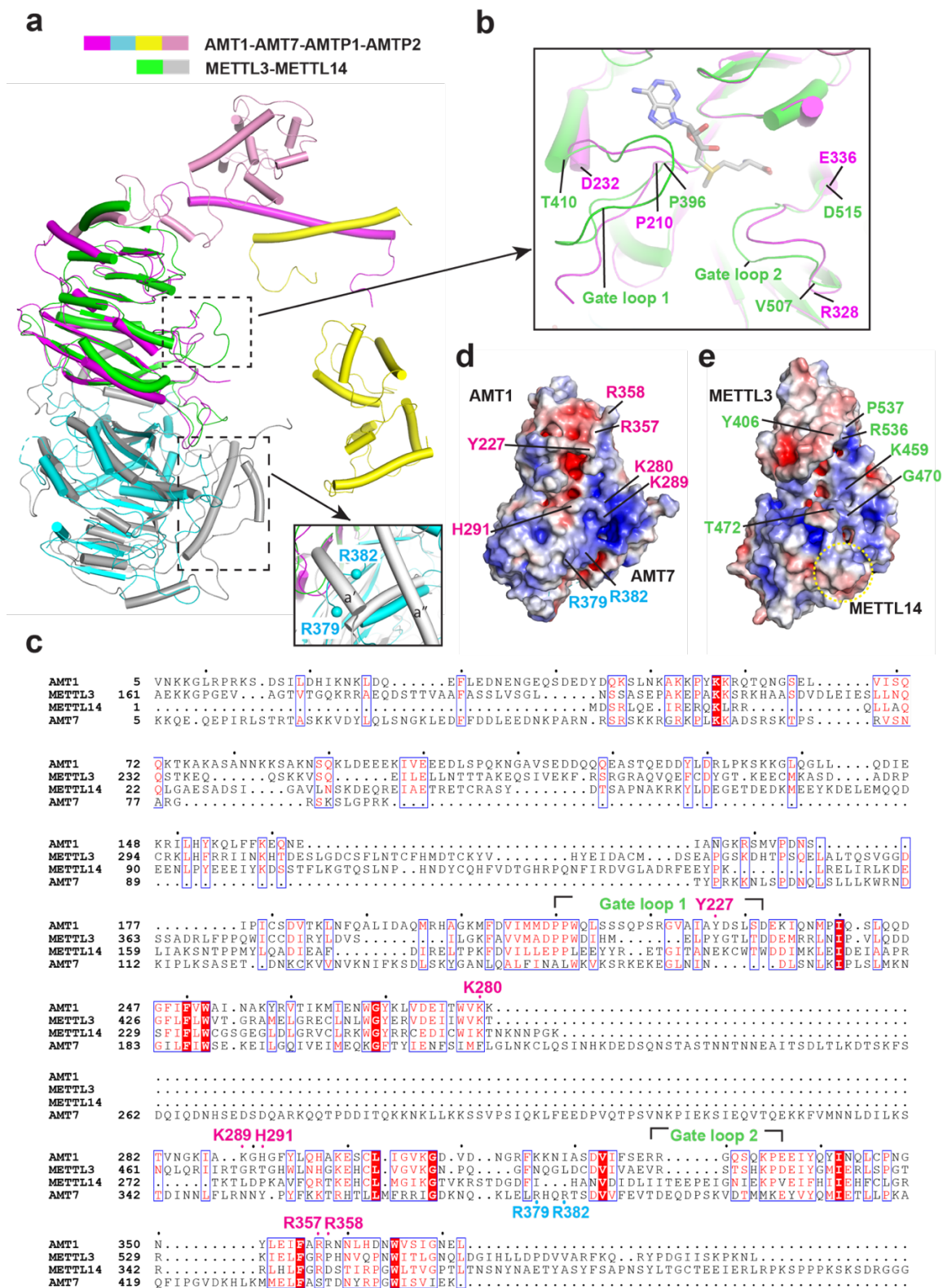
Supplementary Fig. 5 | Sequence alignment of AMTP1 from various species.

TthAMTP1: *Tetrahymena thermophila* AMTP1 (XP_001009903), SleAMTP1: *Stylonychia lemnae* AMTP1 (CDW84997), HgrAMTP1: *Halteria grandinella* AMTP1 (TNV87513), PteAMTP1: *Paramecium tetraurelia* AMTP1 (XP_001443306), PpeAMTP1: *Pseudocohnilembus persalinus* AMTP1 (KRX10274). The OCR-contact residues of AMTP1 are labeled yellow.



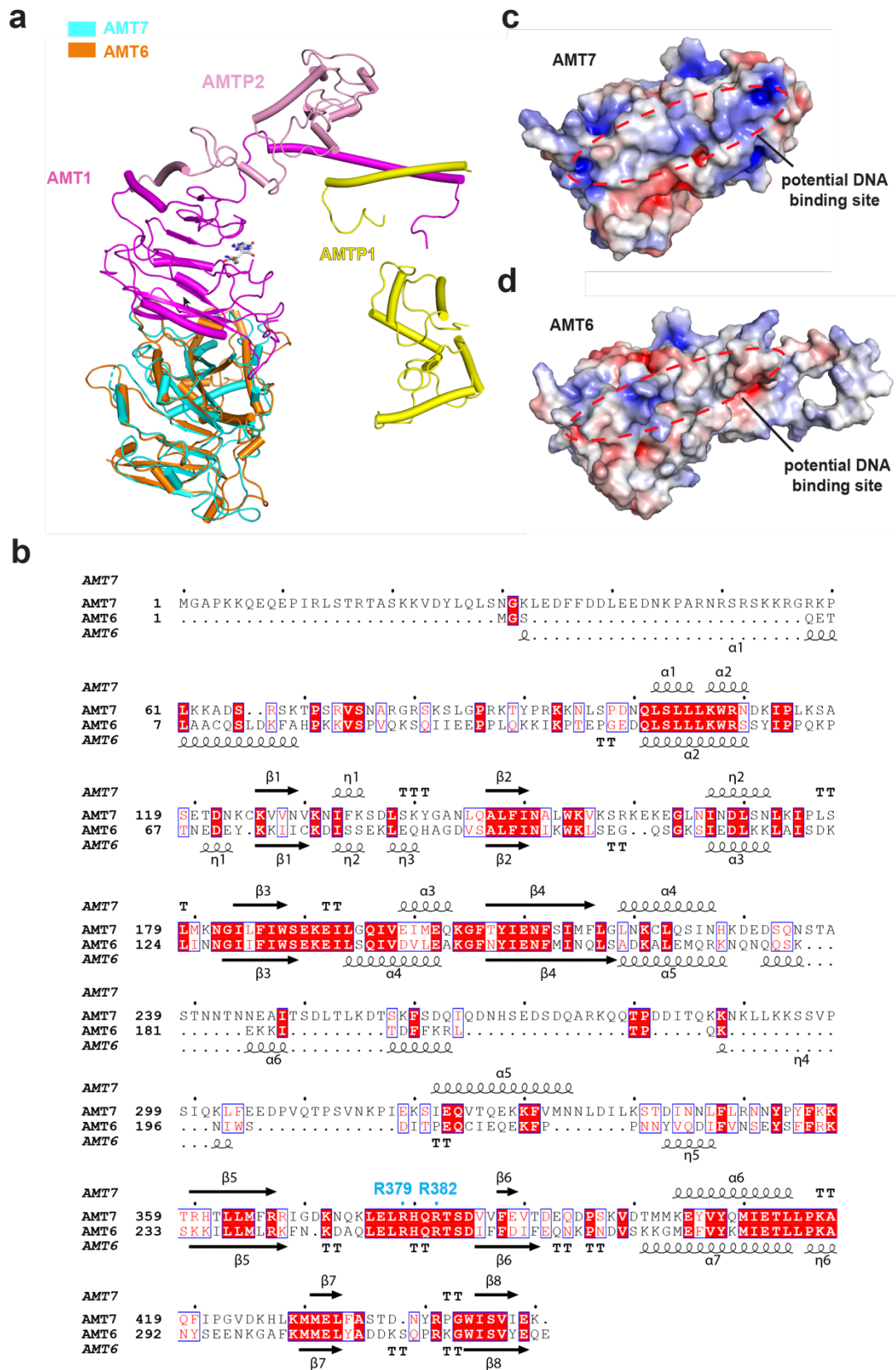
Supplementary Fig. 6 | Evolutionary analysis of AMTP1.

(a) Structural alignment of AMTP1 homeodomain (yellow) with DNA-bound homeodomain-containing proteins: DUX4 homeodomain (PDB 6E8C, slate), TEBP-2 MCD3 (PDB 8U8L, wheat), *Antennapedia* homeodomain (PDB 9ANT, green), and TEAD4 TEA (PDB 5GZB, cyan), with individual helices labeled. The long C-terminus of $\alpha 3$ in TEAD4 TEA was removed for clarity. (b) Structural alignment of AMTP1 (yellow) with the DNA complex of homeodomain-containing proteins: DUX4 homeodomain (slate)-DNA (grey) (PDB 6E8C), TEBP-2 MCD3 (wheat)-DNA (grey) (PDB 8U8L), *Antennapedia* homeodomain (green)-DNA (grey) (PDB 9ANT), and TEAD4 TEA (cyan)-DNA (grey) (PDB 5GZB). (c) Sequence alignment of AMTP1 homeodomain with that of DUX4, *Antennapedia* (Antp), TEBP-2, and TEAD4. The corresponding secondary structure and OCR-contact sites of AMTP1 are indicated on top. (d) Structural overlay of the AMT-DNA model with prokaryotic 6mA writers EcoP15I Mod subunit-DNA (PDB 4ZCF), CcrM-DNA complex (PDB 6PBD), and *M.TaqI*-DNA complex (PDB 1G38). The AMTP1 and the corresponding TRD of the prokaryotic complexes, as well as the bound DNA molecules, are labeled.



Supplementary Fig. 7 | Structural comparison of OCR-bound AMT complex with METTL3-METTL14 complex.

(a) Structural alignment of OCR-bound AMT complex with METTL3-METTL14 complex (PDB 5IL1), with the cofactor-binding site boxed and the AMT7 R379- and R382-corresponding region highlighted in the expanded view. Note that the corresponding region in METTL3-METTL14 is shielded by a pair of helices (α' and α''). (b) Close-up view of the cofactor-binding Gate loops, with residue numbers delimited. The SAM molecule bound to METTL3-METTL14 complex is shown in stick representation. (c) Sequence alignment of Tetrahymena AMT1, human METTL3, human METTL14 and Tetrahymena AMT7. Identical or similar residues are boxed and colored red. Completely conserved residues are shaded in red. The potential DNA-binding sites for AMT1 and AMT7 are marked above and below the alignment respectively. (d, e) Electrostatic surfaces of the AMT1-AMT7 (d) and METTL3-METTL14 (e) complexes. The OCR-contact sites of AMT1 and AMT7 are labeled in magenta and cyan, respectively. The corresponding sites in METTL3 are labeled in green. The METTL14 region corresponding to AMT7 R379 and R382, which is occluded from solvent access, is marked by dotted circle.



Supplementary Fig. 8 | Structural and sequence analysis of AMT7 and AMT6.

(a) Structural overlay of AMT7 (cyan) and AMT6 (orange) in the context of the AMT complex. The structural model for AMT6 was predicted by AlphaFold3. (b) Sequence comparison of AMT6 and AMT7, with the AMT7 residues involved in OCR binding labeled in cyan on top. (c) Electrostatic surface view of AMT7, with the potential DNA-binding surface circled by red dashed line. d Electrostatic surface view of AMT6, with the potential DNA-binding site circled by red dashed line.

Supplementary Table 1. Cryo-EM data collection, refinement and validation statistics.

State	AMT1-AMT7-AMTP1-AMTP2
Code	(EMD-70174, PDB 9O6K)
Data collection and processing	
Microscope	Titan Krios
Camera	Gatan K3
Magnification	81,000
Voltage (kV)	300
Defocus range (μm)	-0.4 ~ -3.1
Exposure time (s)	2
Dose rate($e^-/\text{\AA}^2/\text{s}$)	26.77
Number of frames	50
Pixel size ($\text{\AA}$)	0.529
Micrographs (no.)	5137
Symmetry imposed	<i>C1</i>
Initial particles (no.)	2,645,958
Final particles (no.)	66,922
Map resolution ($\text{\AA}$)	3.59
FSC threshold	0.143
Refinement	
Initial model used	AMT1, AMT7, AMTP2 (7YI8) AMTP1 (AlphaFold)
Model resolution ($\text{\AA}$)	3.8
FSC threshold	0.143
Map sharpening <i>B</i> factor ($\text{\AA}^2$)	-100
CC (mask)	0.62
Model composition	
Non-hydrogen atoms	5847
Protein residues	756
SAH	1
<i>B</i> factors ($\text{\AA}^2$)	
Protein	230.25
SAH	146.36
R.m.s. deviations	
Bond lengths ($\text{\AA}$)	0.004
Bond angles ($^\circ$)	0.775
Validation	
MolProbity score	1.95
Clashscore	9.69
Poor rotamers (%)	0.34
Ramachandran plot	
Favored (%)	93.22
Allowed (%)	6.78
Disallowed (%)	0.00

References

1. Bird, A. P. & Wolffe, A. P. Methylation-Induced Repression— Belts, Braces, and Chromatin. *Cell* 99, 451–454 (1999).
2. Razin, A. & Riggs, A. D. DNA Methylation and Gene Function. *Science* 210, 604–610 (1980).
3. Panning, B. & Jaenisch, R. RNA and the Epigenetic Regulation of X Chromosome Inactivation. *Cell* 93, 305–308 (1998).
4. Bourc’his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431, 96–99 (2004).
5. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20, 116–117 (1998).
6. Jones, P. A. & Gonzalgo, M. L. Altered DNA methylation and genome instability: A new pathway to cancer? *Proc. Natl. Acad. Sci. U.S.A.* 94, 2103–2105 (1997).
7. Luo, G.-Z., Blanco, M. A., Greer, E. L., He, C. & Shi, Y. DNA N6-methyladenine: a new epigenetic mark in eukaryotes? *Nat Rev Mol Cell Biol* 16, 705–710 (2015).
8. Boulias, K. & Greer, E. L. Means, mechanisms and consequences of adenine methylation in DNA. *Nat Rev Genet* 23, 411–428 (2022).
9. Fu, Y. et al. N6-Methyldeoxyadenosine Marks Active Transcription Start Sites in *Chlamydomonas*. *Cell* 161, 879–892 (2015).
10. Mondo, S. J. et al. Widespread adenine N6-methylation of active genes in fungi. *Nat Genet* 49, 964–968 (2017).
11. Wang, Y. et al. A distinct class of eukaryotic MT-A70 methyltransferases maintain symmetric DNA N6-adenine methylation at the ApT dinucleotides as an epigenetic mark associated with transcription. *Nucleic Acids Res* 47, 11771–11789 (2019).
12. Beh, L. Y. et al. Identification of a DNA N6-Adenine Methyltransferase Complex and Its Impact on Chromatin Organization. *Cell* 177, 1781-1796.e25 (2019).
13. Sheng, Y. et al. Semiconservative transmission of DNA N6 -adenine methylation in a unicellular eukaryote. *Genome Res.* 34, 740–756 (2024).
14. Wang, Y. et al. Dual modes of DNA N6 -methyladenine maintenance by distinct methyltransferase complexes. *Proc. Natl. Acad. Sci. U.S.A.* 122, e2413037121 (2025).
15. Cheng, X. DNA modification by methyltransferases. *Current Opinion in Structural Biology* 5, 4–10 (1995).
16. Du, Q., Luu, P.-L., Stirzaker, C. & Clark, S. J. Methyl-CpG-Binding Domain Proteins: Readers of the Epigenome. *Epigenomics* 7, 1051–1073 (2015).
17. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11, 204–220 (2010).

18. Ren, W., Gao, L. & Song, J. Structural Basis of DNMT1 and DNMT3A-Mediated DNA Methylation. *Genes* 9, 620 (2018).
19. Song, J. & Pfeifer, G. P. Are there specific readers of oxidized 5-methylcytosine bases? *BioEssays* 38, 1038–1047 (2016).
20. Zhang, X., Zhang, Y., Wang, C. & Wang, X. TET (Ten-eleven translocation) family proteins: structure, biological functions and applications. *Sig Transduct Target Ther* 8, 297 (2023).
21. Chen, J. et al. Structural basis for MTA1c-mediated DNA N6-adenine methylation. *Nat Commun* 13, 3257 (2022).
22. Yan, J. et al. Structural insights into DNA N6-adenine methylation by the MTA1 complex. *Cell Discov* 9, 8 (2023).
23. Śledź, P. & Jinek, M. Structural insights into the molecular mechanism of the m6A writer complex. *eLife* 5, e18434 (2016).
24. Wang, X. et al. Structural basis of N6-adenosine methylation by the METTL3–METTL14 complex. *Nature* 534, 575–578 (2016).
25. Wang, P., Doxtader, K. A. & Nam, Y. Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. *Molecular Cell* 63, 306–317 (2016).
26. Putnam, C. D. & Tainer, J. A. Protein mimicry of DNA and pathway regulation. *DNA Repair* 4, 1410–1420 (2005).
27. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 14, 290–296 (2017).
28. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* 16, 1153–1160 (2019).
29. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500 (2024).
30. Pettersen, E. F. et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605–1612 (2004).
31. Meng, E. C. et al. UCSF ChimeraX : Tools for structure building and analysis. *Protein Science* 32, e4792 (2023).
32. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486–501 (2010).
33. Adams, P. D. et al. PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66, 213–221 (2010).
34. Li, S. et al. Ocr-mediated suppression of BrxX unveils a phage counter-defense mechanism. *Nucleic Acids Research* 52, 8580–8594 (2024).
35. Walkinshaw, M. D. et al. Structure of Ocr from Bacteriophage T7, a Protein that Mimics B-Form

- DNA. *Molecular Cell* 9, 187–194 (2002).
36. Bürglin, T. R. & Affolter, M. Homeodomain proteins: an update. *Chromosoma* 125, 497–521 (2016).
 37. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Research* 38, W545–W549 (2010).
 38. Lee, J. K. et al. Crystal Structure of the Double Homeodomain of DUX4 in Complex with DNA. *Cell Reports* 25, 2955–2962.e3 (2018).
 39. Padmanaban, S. et al. *Caenorhabditis elegans* telomere-binding proteins TEBP-1 and TEBP-2 adapt the Myb module to dimerize and bind telomeric DNA. *Proc. Natl. Acad. Sci. U.S.A.* 121, e2316651121 (2024).
 40. Fraenkel, E. & Pabo, C. O. Comparison of X-ray and NMR structures for the Antennapedia homeodomain–DNA complex. *Nat Struct Mol Biol* 5, 692–697 (1998).
 41. Shi, Z. et al. DNA-binding mechanism of the Hippo pathway transcription factor TEAD4. *Oncogene* 36, 4362–4369 (2017).
 42. Gupta, Y. K., Chan, S.-H., Xu, S. & Aggarwal, A. K. Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat Commun* 6, 7363 (2015).
 43. Horton, J. R. et al. The cell cycle-regulated DNA adenine methyltransferase CcrM opens a bubble at its DNA recognition site. *Nat Commun* 10, 4600 (2019).
 44. Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J. & Weinhold, E. [No title found]. *Nat. Struct Biol.* 8, 121–125 (2001).
 45. Feng, X. & He, C. Mammalian DNA N6-methyladenosine: Challenges and new insights. *Molecular Cell* 83, 343–351 (2023).
 46. Breyer, W. A. & Matthews, B. W. A structural basis for processivity. *Protein Science* 10, 1699–1711 (2001).
 47. Krishna, T. S. R., Kong, X.-P., Gary, S., Burgers, P. M. & Kuriyan, J. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* 79, 1233–1243 (1994).
 48. Viadiu, H. & Aggarwal, A. K. Structure of BamHI Bound to Nonspecific DNA. *Molecular Cell* 5, 889–895 (2000).
 49. Chasman, D. I., Flaherty, K. M., Sharp, P. A. & Kornberg, R. D. Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proc. Natl. Acad. Sci. U.S.A.* 90, 8174–8178 (1993).

Chapter 4

General Conclusion

DNA methylation serves as one of the fundamental epigenetic marks that shapes genome function across diverse eukaryotic systems¹. Although cytosine and adenine methylation have long been studied in distinct biological contexts, the molecular principles governing how their methyltransferases are activated and regulated, as well as how they engage their substrates, have remained elusive. This dissertation addresses these questions through structural and biochemical studies of two representative eukaryotic DNA methyltransferase systems: the fungal 5mC methyltransferase DIM2 and the ciliate 6mA methyltransferase AMT complex. Together, these studies provide new insight into how distinct methylation machineries have evolved specialized mechanisms for substrate recognition, activation, and epigenetic regulation.

The first part of this dissertation focuses on the *Neurospora crassa* 5mC DNA methyltransferase DIM2, a key enzyme homologous to mammalian cytosine methyltransferase DNMT1 and responsible for cytosine methylation in fungal heterochromatin². Through structural, biochemical, and functional analyses, this work reveals that DIM2 is not simply a catalytic enzyme acting on naked DNA, but rather a highly regulated methyltransferase whose activity is coordinated by multilayered interactions with heterochromatin components. In particular, the studies presented here demonstrate how DIM2 integrates inputs from other repressive marks, including histone H3 lysine 9 trimethylation and HP1, to achieve region-specific DNA methylation. These findings support a model in which heterochromatin-associated interactions act as a

molecular switch that promotes productive substrate engagement and DNA methylation. More broadly, this work highlights how chromatin context can directly regulate the activity performance of a eukaryotic DNA methyltransferase, thereby ensuring that methylation is deposited with both spatial precision and epigenetic fidelity.

The second part of this dissertation examines the AMT complex from *Tetrahymena thermophila*, the first identified eukaryotic 6mA DNA methyltransferase complex. By capturing the AMT complex in an inhibitor-trapped state using the DNA mimic protein OCR, this work provides structural insight into how a eukaryotic adenine methyltransferase machinery assembles for substrate recognition. The cryo-EM structure of AMT complex reveals a saddle-shaped architecture in which multiple subunits cooperate to form a positively charged cleft in the middle suitable for DNA engagement. In addition, the structural and mutational analyses identify key protein surfaces involved in substrate binding and support a model, in which 6mA methyltransferase AMT complex adopting an open-to-closed conformational change is coupled to processive DNA methylation. These findings establish an architectural framework for understanding eukaryotic 6mA methylation and suggest that substrate capture by the AMT complex depends on coordinated assembly rather than the action of a single catalytic subunit in isolation.

Taken together, the studies in this dissertation reveal the conserved and divergent regulation of two types of DNA methyltransferases in eukaryotes. First, DNA methyltransferases are extensively regulated by molecular context. Their activities are not determined solely by the catalytic domain, but instead coordinated with chromatin marks,

partner proteins, and higher-order chromatin environment. Second, substrate recognition for DNA methylation in eukaryotes relies on elaborate regulation rather than simple sequence readout. Both DIM2 and the AMT complex use cooperative structural features to achieve specificity: DIM2 through heterochromatin-coupled regulation, and the AMT complex through multi-subunit assembly and enclosure of the substrate. These principles underscore the idea that eukaryotic DNA methyltransferases function as integrated regulatory machineries rather than isolated enzymes.

Beyond their specific mechanistic contributions, the findings in this study also have broader implications for epigenetics and genome biology. The DIM2 study provides a framework for understanding how fungal DNA methylation is restricted to heterochromatic regions and how histone and DNA methylation pathways are functionally coupled. The AMT complex study expands the current view of eukaryotic DNA methylation by providing a structural basis for adenine methylation in a eukaryotic genome. Together, these studies broaden the conceptual landscape of DNA methylation by showing that distinct methylation marks, although chemically different, are governed by common principles of regulated substrate access, cooperative assembly, and chromatin-dependent control.

Several important questions remain for future investigation. For DIM2, it will be important to define how repressive chromatin states recruit and control DIM2-mediated DNA methylation, and how additional heterochromatin factors contribute to the establishment and maintenance of fungal DNA methylation. For the AMT complex, future work should aim to capture the state of AMT complex bound to nucleosome,

clarify the conformational transitions associated with various chromatin context, and determine how 6mA is propagated and interpreted in vivo. More broadly, continued studies of structurally and evolutionarily diverse methyltransferase systems will be essential for understanding how DNA methylation has adapted to different genomic and cellular environments.

In conclusion, this dissertation provides structural and mechanistic insight into the regulation and substrate recognition of eukaryotic DNA methyltransferases responsible for 5mC and 6mA deposition, respectively. By linking enzyme architecture to epigenetic function, this work advances our understanding of how DNA methylation is established in chromatin and highlights the remarkable diversity of strategies that eukaryotic organisms have adapted to control their genomes. These findings not only deepen our knowledge of epigenetic regulation in fungi and ciliates but also provide a foundation for future studies aimed at understanding the evolution, function, and biological significance of DNA methylation pathways across eukaryotes.

References

1. Bestor, T. H. DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philos Trans R Soc Lond B Biol Sci* 326, 179–187 (1990).
2. Rountree, M. R. & Selker, E. U. DNA methylation and the formation of heterochromatin in *Neurospora crassa*. *Heredity (Edinb)* 105, 38–44 (2010).