Structural basis of DNA N^6-adenine methylation in eukaryotes

Bei Nan[1,2,#], Wentao Yang[3,#], Dantong Wang[4,#], Lanheng Nie[1,2,#], Yuanyuan Wang[1,2,#], Mengyang Xu[4,#], Wenjia Guan[5,6,7], Zhao Chen[5,6,7], Guangchao Zhuang[5,6,7], Shan Gao[1,2,*], Yifan Liu[3,*]

[1]MOE Key Laboratory of Evolution & Marine Biodiversity and Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao 266003, China

[2]Laboratory for Marine Biology and Biotechnology, Qingdao Marine Science and Technology Center, Qingdao 266237, China

[3]Department of Cancer Biology, University of Southern California Keck School of Medicine, Los Angeles, CA 90033, USA

[4]BGI Research, Qingdao 266555, China

[5]Frontiers Science Center for Deep Ocean Multispheres and Earth System (FDOMES), Key Laboratory of Marine Chemistry Theory and Technology, Ministry of Education, Ocean University of China, Qingdao 266100, China.

[6]Laboratory for Marine Ecology and Environmental Science, Qingdao Marine Science and Technology Center, Qingdao 266237, China

[7]College of Chemistry and Chemical Engineering, Academy of Future Ocean, Ocean University of China, Qingdao 266100, China

#These authors contributed equally

*Corresponding author:

Shan Gao (shangao@ouc.edu.cn); Yifan Liu (Yifan.Liu@med.usc.edu)

Abstract

N^6-adenine methylation occurs in both DNA and RNA (referred to as 6mA and m6A, respectively). As an extensively characterized epi-transcriptomic mark found in virtually all eukaryotes, m6A in mRNA is deposited by METTL3-METTL14 complex. As a transcription-associated epigenetic mark abundantly present in many unicellular eukaryotes, 6mA is coordinately maintained by two AMT1 complexes, distinguished by their mutually exclusive subunits, AMT6 and AMT7. These are all members of MT-A70 family methyltransferases (MTases). Despite their functional importance, no structure for holo-complexes with cognate DNA/RNA substrate has been resolved. Here, we employ AlphaFold3 (AF3) and molecular dynamics (MD) simulations for structural modeling of *Tetrahymena* AMT1 complexes, with emphasis on ternary holo-complexes with double-stranded DNA (dsDNA) substrate and cofactor. Key structural features observed in these models are validated by mutagenesis and various other biophysical and biochemical approaches. Our analysis reveals the structural basis for DNA substrate recognition, base flipping, and catalysis in the prototypical eukaryotic DNA 6mA-MTase. It also allows us to delineate the reaction pathway for processive DNA methylation involving translocation of the closed form AMT1 complex along dsDNA. As the active site is highly conserved across MT-A70 family of eukaryotic 6mA/m6A-MTases, the structural insight will facilitate rational design of small molecule inhibitors, especially for METTL3-METTL14, a promising target in cancer therapeutics.

Introduction

As a base modification, N^6-adenine methylation occurs in both DNA and RNA, referred to as 6mA and m6A, respectively (Figure 1A). While long-established in prokaryotes, the widespread presence of 6mA in eukaryotes has only been recognized lately[1-4]. As a transcription-associated epigenetic mark, 6mA —occurring specifically at the ApT duplex— is best characterized in the unicellular eukaryote *Tetrahymena thermophila*[4-8]. MT-A70 family methyltransferases (MTases), classified into several clades with distinct structures and functions, are responsible for N^6-adenine methylation in eukaryotes[9-11]. As founding members of the family, METTL3 and METTL14 form a heterodimer and deposit m6A in mRNA, a well characterized epi-transcriptomic mark[12,13]. *Tetrahymena* MT-A70 MTase AMT1 is required for semiconservative transmission of 6mA[5,7,14,15]. Two AMT1-containing complexes, distinguished by their exclusive subunits, AMT6 and AMT7, coordinate in 6mA transmission[6] (Figure 1B).

Extensive structural studies of prokaryotic DNA MTases, including both 5mC (5-methyl cytosine)[16,17]and 6mA-MTases[18], reveal a conserved MTase fold at the active site, which can bind both the substrate (cytosine or adenine) and cofactor (S-adenosyl methionine (SAM) or S-adenosyl homocysteine (SAH))[19-21]. Flipping of the target base, out of the double-stranded DNA (dsDNA) and into the active site, is a universal mechanism for MTase catalysis[22]. MT-A70 family of eukaryotic 6mA/m6A-MTases are classified as hetero-dimeric beta class amino MTases[20,21]. The structure of METTL3-METTL14 complex reveals division of labor between the cofactor-binding catalytic METTL3 subunit and the non-catalytic METTL14 subunit[23-25]. Similarly, the structure of the hetero-tetrameric AMT1 complex also features AMT1 as the catalytic subunit, AMT7 as the non-catalytic subunit, and two auxiliary subunits, AMTP1 and AMTP2[26,27]. The structure of *Tetrahymena* AMT1 complex associated with a DNA mimic, the Overcoming

Classical Restriction (OCR) protein from bacteriophage T7, has been recently reported, shedding light on the substrate binding[28]. However, no structure for MT-A70 family MTases complexed with their cognate DNA/RNA substrates has been resolved, and the molecular basis for substrate recognition and catalysis still needs to be clarified.

The last few years have witnessed dramatic advances in AI-based protein structure prediction, with milestones like AlphaFold2 (often simply referred to as AlphaFold) and RoseTTAFold[29,30]. The advent of AlphaFold3 (AF3) and RoseTTAFold all-atom extends accurate modeling to other biomolecular interactions, including those with nucleic acids, as well as small-molecule metabolites and inhibitors[31,32]. Here, we employ AF3 and molecular dynamics (MD) simulations for structural modeling of AMT1 complex, with special emphasis on AMT1 ternary holo-complex containing both dsDNA substrate and cofactor (Figure 1B). Structural features observed in these models are further verified by mutagenesis and various other biophysical and biochemical approaches. Our analysis reveals the structural basis for DNA substrate recognition, base flipping, and catalysis in AMT1 complex—the prototypical eukaryotic DNA 6mA-MTase. Moreover, it allows us to delineate the reaction pathway for processive DNA methylation, which involves translocation of the closed form AMT1 complex along dsDNA. As the active site is highly conserved in MT-A70 family of eukaryotic 6mA/m6A-MTases, the structural insight will also facilitate rational design of small molecule inhibitors, especially for METTL3-METTL14, a promising target in cancer therapeutics[33,34].

Results

Structural modeling of AMT1 complex

We used AF3 for structural modeling of *Tetrahymena* AMT1 complex (Figure 1B, C, Figure S1A, B, Table S1), which are mainly involved in 6mA maintenance methylation—converting the ApT duplex from the state of hemi-methylation to full-methylation, but also capable of de novo methylation—converting unmethylated ApT to its hemi-methylated state, albeit with reduced MTase activity[5-7,14]. We modeled AMT1 apo-complex without DNA, binary holo-complex with only DNA, and ternary holo-complex with both DNA and cofactor (Figure S1A). For AMT1 holo-complex, we mainly focused on dsDNA substrates with an identical palindromic sequence for both strands, featuring a central ApT duplex with different methylation states (un-, hemi-, and full-methylation), corresponding to reactants and products in de novo and maintenance methylation (Figure S1B). We also modeled AMT1 complex with AMT6 or AMT7 as the alternative subunit, referred to as AMT6 complex and AMT7 complex, respectively (Figure S1A). For each input setting, 1,000 seeds were used, generating 5,000 models. We examined 24 input settings with various combinations, generating 120,000 models in total (Figure 1C, Figure S1A, Table S1). This comprehensive approach allows us to not only identify highly confident and representative structures of AMT1 complex, but also explore its conformational heterogeneity, providing multiple snapshots along the reaction pathway of AMT1-catalyzed DNA methylation.

We used the molecular dynamics (MD) simulation program CHARMM to optimize critical AF3 models by minimizing local steric clashes[35] (Figure 1C). We focused on SAH-containing AMT1 ternary complex due to parameterization constraints: SAH can be accurately modeled with the CHARMM General Force Field (CGenFF), whereas SAM lacks compatible parameters[36]. Structural features observed in these models were verified

experimentally, including extensive mutagenesis followed by in vitro MTase assays and 6mA detection in genomic DNA of *Tetrahymena* mutants (Figure 1C). This allows us to identify biologically relevant interactions and establish critical structure-function relationships.

Overall structural organization of AMT1 complex

The DNA-binding domain (DBD) of AMTP1, long proposed to interact with the substrate, is not solved in published AMT7 apo-complex structures[26-28]. AF3 modeling of AMT1 apo-complex reveals a high-confidence, well-defined local structure for AMTP1 DBD, which is tethered to the AMT1-AMT6/AMT7 structural core with a long and flexible linker region (Figure S1C). This leads to dispersed global arrangements, with AMTP1 DBD often at substantial distances from the structural core, representing various open forms of AMT1 complex (Figure S1C). However, a stable global arrangement emerges in AMT1 holo-complex, as AMTP1 DBD, confined to the DNA major groove, form direct contacts with the structure core (Figure 1D, Figure S1C, D). In this closed form, the four proteins subunits are linked into a torus with dsDNA substrate passing through the central hole (Figure 1D, Figure S1D).

The catalytic subunit AMT1 directly engages the target strand of dsDNA at the ApT site. Using the directionality of the target strand, we define two faces of the closed form AMT1 holo-complex, with its front facing the 3' end and its back facing the 5' end (Figure 1D, side view). Looking down at the front face, AMTP1 DBD, AMT6/AMT7 MT-A70, AMT1 MT-A70, and AMT1 N-terminal extension modules are arranged clockwise (Figure 1D, front view). AMT1 N-terminal extension module, comprising AMT1 N-terminal helix, AMTP1 helix-turn-helix-like domain, and AMTP2, is structurally well-characterized in AMT7 apo-complex[26,27]. It is stably connected to AMT1 MT-A70 on one side and tethered to AMTP1 DBD by a long linker on the other, but not directly

involved in DNA binding (Figure 1D, front view). AMTP1 DBD, AMT6/AMT7 MT-A70, and AMT1 MT-A70 are inter-connected and wrapped around dsDNA (Figure 1D, front view). Both the adenine-binding and cofactor-binding pockets, solely formed by AMT1, open at the front (Figure 1D, front view). By contrast, sequences divergent in AMT6 and AMT7, including their N-terminal tails and long internal loops, protrude from the back (Figure 1D, side view). These AMT6 and AMT7-specific regions are implicated in regulating and targeting MTase activities, with replication-coupled and transcription-associated methylation preferentially performed by AMT6 and AMT7 complexes, respectively[6]. Indeed, the Proliferating Cell Nuclear Antigen (PCNA)-interacting protein (PIP) box of AMT6 resides in the divergent long loop at the back (Figure 1E). This arrangement positions AMT6 PIP for direct binding in a conserved pocket at the front of the tandemly loaded replication sliding clamp (Figure 1E).

The interface between AMT1 and AMT7 MT-A70 modules is well-defined in the solved structures of AMT1 apo-complex[26,27]. We tracked two distances across this interface in AF3 models of all AMT1 apo- and holo-complexes: their values are tightly clustered in scatter plots, supporting a stable and homogeneous local conformation (Figure S1E). Nonetheless, there is still substantial conformational heterogeneity in AMT1 complex. Some recurring themes include different states of the target adenine (base flipping or pairing), different sites of DNA binding (onsite or offsite binding, in relation to the target ApT duplex), as well as the open and closed forms. Indeed, AMT1 complex dynamics are mostly determined by variable interfaces between the largely invariable structural modules (Figure S1E). Building on this insight, we developed several sets of parameters to monitor variable interfaces in AMT1 complex (Figure S1F, Table S3). Each set of parameters track key distances across an interface, and distribution of these values are mapped for all AF3 models; conformational analysis based on each set is covered in Results

sections elaborating on individual structural features. In Discussion, conformational heterogeneity of AMT1 complex is recapitulated by comprehensive analysis based on Uniform Manifold Approximation and Projection (UMAP) of all relevant parameters.

DNA binding by AMT1

DNA and RNA substrates are not present in previously solved structures of AMT1 complex[26,27] and METTL3-METTL14 complex[23-25]. Therefore, molecular mechanisms for adenine methylation in MT-A70 family of eukaryotic amino-MTases still need to be elucidated. We first analyzed AMT1 ternary complex containing dsDNA substrate with a central full-6mApT and cofactor SAH (AMT7-full-SAH), corresponding to the enzyme-product (EP) complex for maintenance methylation (hemi-6mApT + SAM → full-6mApT + SAH)[5,6]. The top 10 AF3 models, ranked by their ipTM scores, were all in the base flipping mode: 6mA on the target strand is flipped out of dsDNA and into AMT1 active site. We performed MD optimization for these AF3 models, generating structures that are mostly superimposable after pruning a long and flexible internal loop of AMT7 (Figure S2). Here, we highlight three important loops in AMT1: *Major Groove Insertion Loop*, *Minor Groove Insertion Loop*, and *Regulation Loop* (Figure 2A). In the presence of dsDNA and cofactor, these three loops form well-ordered structures gripping the target strand and enclosing the extrahelical adenine (Figure 2A, Figure S3A).

AMT1 and METTL3 are well aligned in primary sequences and 3D structures, with one-to-one correspondence at these three loops (Figure S3B, C). Their conservation and divergence highlight shared and unique features of adenine methylation in DNA and RNA, respectively. AMT1 *Major Groove Insertion Loop*, engaging extrahelical adenine as well as the major groove, corresponds to METTL3 *Gate Loop 1*, but harbors an insertion of seven amino

acids (86-92: SSQPSRG) in the middle (Figure 2B, Figure S3B, C). AMT1 *Minor Groove Insertion Loop* corresponds to METTL3 *Interface Loop*, but is highly divergent in sequence (Figure 2C, Figure S3B, C). It is physically associated with AMT7 *Multiple interaction Loop*, which is partially conserved in AMT6, but divergent in METTL14 (Figure 2C, Figure S3B, D). Note that this interaction is part of the stable hetero-dimerization interface between AMT1 and AMT6/AMT7 MT-A70 modules, as demonstrated by conformational analysis of all AF3 models (Figure 2C, Figure S1E). *Regulation Loop*, controlling the interaction between extrahelical adenine and cofactor, is structurally aligned to METTL3 *Gate Loop 2* (Figure 2D, Figure S3B, C). Furthermore, K150 flanking AMT1 *Minor Groove Insertion Loop* and R198 in AMT1 *Regulation Loop* contact the target strand backbone, while K159 and Q166 in AMT1 *Minor Groove Insertion Loop* contact the non-target strand backbone. These residues are absolutely conserved among AMT1 homologues, but often diverge in METTL3 (Figure S3C, Supplemental File S1[37]).

AMT1 onsite binding at the ApT duplex was found in virtually all AF3 models of AMT1 ternary complex (AMT7-full-SAH). We tracked the distances between 6mA (C3') and H161 (C_α) in AMT1 *Minor Groove Insertion Loop* (Figure 2E, Figure S4A). For the palindromic 6mApT duplex, two values were obtained for each AF3 model. H161 was asymmetrically positioned in the minor groove, much closer to the target strand (5-6 Å) than the non-target strand (14-16 Å). In the 2D distribution, these values were tightly clustered into two groups of reflection symmetry along the diagonal line, as expected for randomly selecting one out of two identical DNA strands as the target strand (Figure 2E). Similar distributions were observed in other AMT1 ternary complexes (Figure 2E, Figure S4A). These AF3 modeling results are consistent with the sequence-specificity of AMT1 complex[5]. They also underscore the importance of AMT1 *Minor Groove Insertion Loop* in DNA

binding.

Base flipping mechanisms

While base flipping is observed in all existing structures of DNA MTases, its angle can vary substantially[22]. In Dam and CcrM, the target base flips ~90°; in M.HhaI and EcoP15I, the target base flips ~180°. In the AMT1 ternary complex (AMT7-SAH-full), the target adenine flips out completely (~180°) and inserts into a deep pocket of AMT1 (Figure 2A). Interactions with extrahelical adenine involve residues conserved in both AMT1 and METTL3 (Figure 2B-D, Figure S3C). The proximal portion of AMT1 *Major Groove Insertion Loop*, containing DPPW (AMT1 79-82; METTL3 395-398) and Y97 (METTL3 Y406), is in close contact with extrahelical adenine (Figure 2B). The structure is stabilized by a hydrophobic core, containing P80 (P396), P81 (P397), W82 (W398), L84 (I400), V93 (L402), I95 (L404), Y97 (Y406), and L100 (L409) from AMT1 *Major Groove Insertion Loop*, as well as I124 (T433) connecting β3 and α3 (Figure 2B). Extrahelical adenine also interacts with AMT1 *Regulation Loop*, containing S202 (S511) and K204 (S513) (Figure 2D). Its Hoogsteen edge is buttressed against a chain of alternating acidic-basic residues within AMT1 *Major Groove Insertion Loop* (D79 (D395)) and *Regulation Loop* (K204 (K513)) and flanking *Minor Groove Insertion Loop* (E170 (E481) and K150 (K459)) (Figure 2B-D). Interactions are extended beyond the base and into the sugar-phosphate moiety: AMT1 K150 (K459) contacts the 5' phosphate, while I124 (T433) and V93 (M402) contact 2' and 4' carbon in the sugar ring, respectively (Figure S4B). The switch from AMT1 I124 to METTL3 T433 can be attributed to the presence of 2' hydroxyl in the RNA substrate, but not DNA.

We also tracked the distances between adenine (N^6) and AMT1 D79 (C_α) in the adenine-binding pocket (Figure S4C-D). In the base flipping mode, the 6mA/A on the target strand inserts into the adenine-binding pocket and

approaches AMT1 D79, while the 6mA/A on the non-target strand keeps its distance (Figure S4C). In the base pairing mode, 6mA/A on both strands stay away from AMT1 D79. For most AF3 models in AMT1 ternary complex, the distance between extrahelical adenine and AMT1 D79 fluctuates within a narrow range of 5-6 Å, consistent with close contact and stable binding (Figure S4D).

In the AMT1 ternary complex (AMT7-SAH-full), AMT1 *Major Groove Insertion Loop* and *Minor Groove Insertion Loop* directly engage base stacks around the 6mApT duplex, stretching DNA along the helical axis (Figure 2A). The distal portion of AMT1 *Major Groove Insertion Loop*, featuring seven amino acids (86-92: SSQPSRG) missing from METTL3 *Gate Loop 1* (Figure S3C), inserts into dsDNA from the major groove, and together with AMTP1 DBD, expands it substantially (Figure 2B, 5A). AMT1 *Minor Groove Insertion Loop* and AMT7 *Multiple interaction Loop*, stabilized by a hydrophobic core (AMT1 F163, Y164 and AMT7 L281 (AMT6 L203), H283 (AMT6 H205)), insert into dsDNA from the minor groove, and expand it substantially (Figure 2C). Notably, the planar imidazole group of AMT1 H161 inserts into the target strand base stack and replaces the flipped-out adenine (Figure 2C). Note that these residues—critical for interactions conducive to base flipping in dsDNA, but not single-stranded RNA—are conserved in AMT1 complex but divergent in METTL3-METTL14 complex (Figure S3C, D, Supplemental File S1, S2).

We reconstituted AMT1 complex incorporating mutations affecting these residues, including AMT1 D79A, W82A, D79A+W82A, Y97A, R91A, L84A+I95A (AMT1 *Major Groove Insertion Loop*); AMT1 H161A, F163A+Y164A (AMT1 *Minor Groove Insertion Loop*); AMT1 S202A+K204A (AMT1 *Regulation Loop*) (Figure 2F), and AMT7 L281A+H283A+Q284A, E280A+R282A (AMT7 *Multiple interaction Loop*) (Figure 2G); the in vitro MTase activity was invariably abolished (Figure 2F, G). Some of these

mutations were also introduced into *Tetrahymena*; these mutants showed diminished endogenous 6mA levels comparable to Δ*AMT1* cells (Figure 2H). All this confirms that these loops are critical for DNA methylation.

AMT1 active site

The cofactor-binding pocket is conserved in AMT1 and METTL3[26,27]. Cofactor placement in the top ranking AF3 models for AMT1 ternary complex, all in the base flipping mode, is highly consistent; it is also not significantly altered by MD optimization (Figure 3A, Figure S5A). Indeed, SAH placement in the predicted structures of the AMT1 ternary complex (AMT7-full-SAH) is almost identical to that in the solved structures of AMT1-AMTP1-AMTP2-SAH (7f4n)[26] and METTL3-METTL14-SAH (5il2) [24]. However, SAH conformation in AMT1-AMT7-AMTP1-AMTP2-SAH (7yi8) is substantially different[27] (Figure S5A). This is surprising given that the cofactor-binding pocket is very similar in our MD-optimized AF3 models and all three solved structures (Figure 3A, Figure S5A). We strongly favor our current cofactor placement, as it avoids steric clashes encountered in the alternative placement and supports a S_N2 in-line attack catalytic mechanism.

Both the adenine-binding and cofactor-binding pockets are solely formed by the catalytic subunit AMT1. In AMT1 ternary complex (AMT7-full-SAH), extrahelical 6mA and SAH are separated by a narrow aperture formed by the main/side chains of AMT1 *Major Groove Insertion Loop* and *Regulation Loop* (Figure 3B). Key residues, all conserved in METTL3, include D79 (D395), P80 (P396), P81 (P397), Y97 (Y406), S202 (S511), and K204 (K513) (Figure 3B, Figure S3C). In this way, both the substrate and the cofactor are precisely positioned and orientated, with access between them restricted to exocyclic amino group of extrahelical adenine and sulfide/sulfonium group of SAH/SAM. Very similar arrangements were found in other AMT1 ternary complexes (e.g.,

AMT7-hemi-SAH, AMT6-full-SAH, and AMT6-hemi-SAH) (Figure S5B).

We performed conformational analysis of various AMT6 and AMT7 ternary complexes, corresponding to enzyme-substrate (ES) and enzyme-product (EP) complexes for de novo as well as maintenance methylation (Figure 3C-D, Figure S5C). In most AF3 models, the distance between extrahelical adenine (N^6) and SAH/SAM (Sδ) is less than 5 Å (Figure 3C). For AF3 models with very high scores (ipTM≥0.9), the distance fluctuates around 4.8 Å (Figure S5C). These values are similar to the equivalent distance in DNA adenine- N^6 MTase M.TaqI (4.7 Å) [18]and cytosine-5 MTase M.HhaI (5.0 Å) [17].

We next used MD simulations to closely examine the local conformation of the active site, focusing on SAH-containing EP complex. To mimic the transition state (TS), we assigned an unstable conformer to extrahelical 6mA, with its N^6 in the sp^3 tetrahedral geometry and its N^6-methyl group perpendicular to the planar adenine. EP-TS structures exhibit features consistent with a S_N2 in-line attack mechanism for methyl transfer catalyzed by amino-MTases: ***1)*** the methyl group almost falls on the line between the extrahelical adenine N^6 and SAH S_δ, and ***2)*** the N^6-S_δ distance, at ~4.8 Å, is also optimal (Figure 3D). For the ground state (GS), we assigned a stable conformer to extrahelical 6mA, with its N^6 in the sp^2 trigonal planar geometry and its N^6-methyl group in the stable *syn* conformation[38]. In EP-GS structures, the N^6-S_δ distance decreases substantially to less than 4 Å, as the methyl group moves out of the way (Figure S5D). A strong hydrogen bond between N^6-amine and the backbone carbonyl group of P80 in TS is abolished in GS (Figure S5D). This stabilization of TS and destabilization of GS also contribute to catalysis. Our structural modeling of AMT1 ternary complex thus provides the molecular basis for substrate and cofactor binding, as well as their optimal placement and alignment for methyl transfer. These structural features, recurrent in AMT1 complex and METTL3-METTL14 complex, support a unified catalytic mechanism for MT-A70 family amino-MTases.

Coordination between cofactor binding and base flipping

We monitored base flipping by tracking the N^6-O^4 distance between reverse complement 6mA and thymine in the palindromic (6m)ApT duplex: two values were obtained for each AF3 model of AMT1 holo-complexes. For the AMT1 ternary complex (AMT7-full-SAH), 2D distribution of these values revealed two distinct classes of conformations: ***1)*** base pairing, in which both 6mA remain paired with their complementary thymine (~20%), and ***2)*** base flipping, in which a single 6mA flips out of the DNA double helix, breaking its symmetry (~80%; under rare circumstances (0.1%), both 6mA flip out) (Figure 3E). The distribution exhibited reflection symmetry along the diagonal line, as expected for randomly selecting one out of two identical DNA strands as the target strand (Figure 3E). Indeed, (single) base flipping was predominant in all AMT1 ternary complexes we examined, especially among AF3 models with very high scores (>99% at ipTM≥0.9), regardless of their state in DNA methylation (full-6mApT, hemi-6mA, or ApT), cofactor (SAM/SAH), and subunit (AMT6/AMT7) (Figure 3F, Table S2). However, a substantial drop in base flipping was observed in cofactor-free binary complexes, along with a substantial drop in the AF3 score, suggesting coordination between cofactor-binding and base flipping (Figure 3F, Table S2).

This coordination can be at least partially attributed to the juxtaposition of cofactor-binding and adenine-binding pockets in AMT1. In the base flipping mode AMT1 ternary complex (AMT7-full-SAH), several residues at the aperture separating the two pockets (including AMT1 D79, P81, Y97, and K204) simultaneously engage extrahelical 6mA and SAH (Figure 3B). In the absence of SAH, some contacts, including the favorable edge-to-face aromatic interaction between AMT1 Y97 and extrahelical 6mA, are lost (Figure S5E). This leads to reduced stability for extrahelical adenine placement in its binding

pocket (Figure S5E). Indeed, the distance between the extrahelical 6mA/A and AMT1 D79 is significantly longer and more diffusively distributed in AF3 models of AMT1 binary complexes than ternary complexes (Figure S4D).

Our modeling also reveals that this coordination is further reinforced by interactions between AMT1 active site and AMTP1 N-terminal region. A specific structure, hereafter referred to as AMTP1 *N-terminal Loop*, is formed in AMT1 ternary complexes (Figure 3G). AMTP1 *N-terminal Loop* contacts AMT1 *Major Groove Insertion Loop* and *Regulation Loop*, which are shared by the adenine-binding and cofactor-binding pockets (Figure 3B, G). It also contacts AMT1 at the β7-β8 connecting loop as well as the C-terminal tail, which are part of the cofactor-binding pocket (Figure 3G). Important contacts include AMTP1 S13, K14, and S15 on one side, and AMT1 A96, Y97, D98, Q201, R227, R228, and E241 on the other, with potential to form hydrogen bond networks (Figure 3G).

We tracked the distances between S15 in AMTP1 *N-terminal Loop* and Y97 and E241 in AMT1 active site (Figure 3G, H, Figure S5F). For AMT1 ternary complexes in the base flipping mode, both distances fluctuate narrowly around two small values, supporting the predominance of AMTP1 *N-terminal Loop* docking at AMT1 active site (Figure 3G). This arrangement is substantially reduced in the base pairing mode of AMT1 ternary complexes (Figure 3H). It all but disappears in cofactor-free holo-complexes and apo-complexes (Figure 3H). These observations further link the two binding events at AMT1 active site to the disorder-to-order transition in AMTP1 N-terminal region and docking of AMTP1 *N-terminal Loop*. AMTP1 N-terminal region is mostly conserved (Supplemental File S3). Its functional importance was confirmed by diminished MTase activity in reconstituted AMT7 complex with truncated AMTP1 (Δ1-15) (Figure 3I).

ApT duplex recognition by AMT1

In *Tetrahymena*, 6mA occurs almost exclusively in the ApT dinucleotide; in vitro methylation by reconstituted AMT1 complex also shows a very strong bias for the ApT dinucleotide[5]. AMT1 complex is critical for maintenance methylation, but also capable of de novo methylation; consistently, it can methylate dsDNA with either hemi-6mApT or unmethylated ApT, albeit with much higher activity for the hemi-methylated substrate[5-7,14]. For structural features underpinning substrate recognition, we first closely examined MD-optimized AF3 models of AMT1 ternary complex (AMT7-full-SAH) in the base flipping mode (Figure 4A). The proximal portion of AMT1 *Major Groove Insertion Loop*, part of AMT1 adenine-binding pocket, directly recognizes extrahelical 6mA, while its distal portion inserts deeply into the major groove for sequence-specific interactions with the 6mA-T base pair. The thymine base immediately 3' to the target adenine, particularly its unique 5-methyl group, can pack against the main/side chains of AMT1 L84, P89, R91, and G92, for favorable van der Waals interactions (Figure 4A). N^6-methyl of the non-target strand 6mA, in the *anti*-configuration required for base pairing with the target strand thymine[39], is stabilized by interacting with the aliphatic part of S90 and R91 side chains (Figure S6A). Note that the β-hydroxyl group of AMT1 S90 can rotate away from N^6-methyl to avoid steric clashes (Figure 4B). Additionally, F163 on AMT1 *Minor Groove Insertion Loop* contacts the sugar edge of the A-T base pair; it would have clashed sterically with the exocyclic amino group of a guanine (Figure S6B). The target strand is further stabilized as its sugar-phosphate backbone contacts L84 on the major groove side and G162 on the minor groove side (Figure 4A, Figure S6A, B). All these residues are highly conserved in AMT1 homologues (Supplemental File S1). In AMT7-hemi-SAH, corresponding to the EP complex for de novo methylation (ApT + SAM → <u>hemi-6mApT +SAH</u>), 6mA on the target strand flips out and unmodified A on the non-target strand remains paired (Figure 4B, D). Here, the

β-hydroxyl group of AMT1 S90 can rotate towards the N^6-amino group of adenine to form a stable hydrogen bond between them (Figure 4B). Similar observations were made in the base flipping mode of AMT6 ternary complexes (Figure S6C). By adopting different conformations, AMT1 S90 can interact favorably with both 6mA and unmodified A on the non-target strand (Figure 4B). Indeed, S90A mutation, abolishing the hydrogen bond with unmodified A, led to much reduced de novo methylation activity; by contrast, it only showed a moderate positive effect on maintenance methylation (Figure 4C). Therefore, dual recognition by AMT1 S90 allows AMT1 complexes to perform de novo methylation as well as maintenance methylation.

We tracked the distances between 6mA/A and S90 in AMT1 *Major Groove Insertion Loop* (Figure 4D, Figure S6D, Table S3). In the base flipping mode, the extrahelical 6mA/A on the target strand moves away from AMT1 S90, while the 6mA/A on the non-target strand remains base paired and stays close. In the base pairing mode, 6mA/A on both strands stay close to AMT1 S90. For most AF3 models in AMT7 ternary complexes, especially in the base flipping mode, the distance between the base paired 6mA/A and AMT1 S90 fluctuates around small values (≤5 Å), supporting target site recognition by AMT1 *Major Groove Insertion Loop* (Figure 4D).

DNA binding by AMTP1

Both AMTP1 and AMTP2 contain an N-terminal homeobox-like domain (HD) implicated in DNA binding[40]. Previous structural studies have shown that AMTP2 HD is exapted to stabilize the interaction between AMT1 and AMTP1[26,27]. Our modeling of AMT1 holo-complex supports that AMTP1 N-terminal HD functions as a genuine DNA-binding domain (Figure 1D, E, Figure 5A, Figure S7A). The core of AMTP1 DBD is superimposable to the solved structure of AMTP2 HD (Figure S7B). The structural model is supported

by XL-MS data (Figure S7C)[27]. AMTP1 DBD features a concave interface with DNA (Figure 5B). Interactions with negatively charged DNA are facilitated by many positively charged basic residues at this interface, some of which are conserved in AMTP1 homologues (K44, R68*, K101*, K114, R117, and K118; *: conserved) (Figure 5B, Supplemental File S3). The DNA-recognition helix α3 and its C-terminal loop directly insert into the major groove (Y63, M64, Q67*, and Q73*) (Figure 5C). Underscoring the critical role of AMTP1, AMT1 by itself and even the AMT1-AMT7-AMTP2 sub-complex lack the high DNA binding affinity and MTase activity of AMT1 complexes[14,26]. Furthermore, in the structure of AMT1 complex bound to a DNA mimic, the Overcoming Classical Restriction (OCR) protein from bacteriophage T7, AMTP1 DBD is positioned near OCR[28]. In line with these findings, reconstituted AMT7 complexes incorporating AMTP1 mutations, including Q67A, R68A, and Q73A, showed diminished MTase activities (Figure 5D). Furthermore, *Tetrahymena* Q73A mutant showed diminished endogenous 6mA levels (Figure 5E). All this supports the functional importance of these AMTP1 residues.

We tracked the distances between 6mA/A (N^6) and M64 (C_α) in AMTP1 DNA-recognition helix α3 (Figure 5F, Figure S7D). In the base flipping mode, the extrahelical 6mA/A on the target strand moves away from AMTP1 M64, while the paired 6mA/A on the non-target strand remains close. In the base pairing mode, 6mA/A on both strands stay close to AMTP1 M64. In AMT1 ternary complexes, the distance between the paired 6mA/A on the non-target strand and AMTP1 M64 fluctuates around 10 Å for most AF3 models (Figure 5F, Figure S7D). Here, we focus on the base flipping mode of the AMT1 ternary complex (AMT7-full-SAH). In MD-optimized top scoring AF3 models, AMTP1 DBD co-occupies the major groove with AMT1 *Major Groove Insertion Loop* (Figure 5A), which directly interacts with the ApT duplex, especially on the target strand, as detailed in previous sections. In this cooperative binding mode, insertion of AMTP1 α3 into the major groove is shallow, and its

interaction with the target strand ApT is blocked (Figure 5A). Nonetheless, AMTP1 α3 and its C-terminal loop (Y63, Q67, and Q73) can still contact ApT on the non-target strand (Figure 5C). The aromatic ring of AMTP1 Y63 packs against the 5-methyl group of the orphaned thymine; together with the aliphatic part of AMT1 S90 and R91 side chains, it can also form a hydrophobic cavity that has energetically more favorable interactions with the N^6-methyl group of 6mA than with the amino group of unmodified adenine (Figure S7E), contributing to the substrate preference for hemi-6mApT over unmethylated ApT.

In the cooperative binding mode, there are extensive interactions between the highly conserved tip of AMT1 *Major Groove Insertion Loop* (90-92: SRG) and AMTP1 α3 (Figure 5G). The tip structure is stabilized by the overlaying AMTP1 α3, as well as the underlaying DNA. We tracked the distances between S90 in AMT1 *Major Groove Insertion Loop*, and G60 and Y63 in AMTP1 α3 (Figure 5H, Figure S7F). For most AF3 models in the base flipping mode of AMT1 ternary complexes, both distances fluctuate are small and narrowly distributed, supporting juxtaposition of AMT1 *Major Groove Insertion Loop* and AMTP1 DBD (Figure 5H, Figure S7F). Strikingly, these distances are generally increased and more broadly distributed in the base pairing mode of AMT1 ternary complexes, suggesting reduced cooperativity (Figure 5H, Figure S7F). Cooperative binding all but disappears in cofactor-free binary complexes (Figure 5H, Figure S7F), supporting that it is promoted by cofactor binding, just like base flipping. This allosteric effect is likely also mediated by AMTP1 *N-terminal Loop*, which is directly connected to AMTP1 DBD via a conserved short linker (23-25: PGW). As AMTP1 *N-terminal Loop* docks at AMT1 active site in AMT1 ternary complexes, it brings this linker region closer to the target strand, thereby enhancing the DNA binding of AMTP1 DBD as well as its interaction with the tip of AMT1 *Major Groove Insertion Loop* (Figure S7G).

ApT duplex recognition by AMTP1

We also found AF3 models with significant decreases or increases in the distance between 6mA/A and AMTP1 M64 relative to the cooperative binding mode, corresponding to two alternative DNA binding modes for AMTP1 (Figure 5F). We focus on the former in this section and the latter in the next. Decrease in these distances indicates that AMTP1 α3 insertion into the major groove is deeper than in the cooperative binding mode. In this AMTP1-dominated binding mode, AMTP1 DBD displaces the distal portion of AMT1 *Major Groove Insertion Loop* (Figure 5I), which exhibits structural heterogeneity after losing most of its contacts in the major groove (Figure S7H). The AMTP1-dominated binding mode, while virtually absent in AMT1 ternary complexes, represents a substantial population of AF3 models in the base pairing mode of AMT1 binary complexes (Figure 5F, Figure S7D). We also tracked the distances between S90 in AMT1 *Major Groove Insertion Loop* and G60/Y63 in AMTP1 DBD: in AMT1 ternary complexes, we found mostly narrow distributions around small values, indicative of physical proximity and cooperative binding; in AMT1 binary complexes, we found mostly diffusive distributions of substantially increased values, indicative of physical separation and lack of cooperative binding (Figure 5H). Therefore, AMTP1-dominated binding mode and cooperative binding mode represent distinct classes of conformations for AMT1 complex.

We focus on the AMTP1-dominated binding mode in an AMT1 binary complex (AMT7-hemi) (Figure 5I). Intriguingly, in high-score AF3 models, the 6mA-containing strand is predominantly selected as the non-target strand, while the unmodified A is placed on the target strand. Close examination of MD-optimized AF3 models revealed more direct interactions between AMTP1 α3 and its C-terminal loop (Y63, M64, Q67, and Q73) on the one side, and the

base paired ApT duplex on the other (Figure 5I). There are extensive contacts with 6mApT on the non-target strand, potentially allowing its specific recognition: the 3' thymine contacts M64, Q67, and Q73, while the 5' 6mA contacts Y63, Q67, and Q73, with the aromatic ring of Y63 positioned close to N^6 for favorable interactions with the methyl group in 6mA (Figure 5I). Importantly, with AMT1 *Major Groove Insertion Loop* out of the way, AMTP1 α3 shifts towards the target adenine, with M64 contacting the adenine exocyclic amino group; the presence of the N^6-methyl group would have pushed AMTP1 α3 out of the major groove, thus accounting for the discrimination against 6mA on the target strand (Figure 5I). AMTP1 DBD in this major groove binding mode is structurally similar to the human telomeric protein TRF1/TRF2 complexed with telomere DNA[41,42]. AF3 also predicts that AMTP1 DBD by itself binds DNA sequences with an intrinsically wide major groove, with striking structural similarities to the AMTP1-dominated binding mode of AMT1 complexes (Figure S7I). Our structural analyses therefore distinguish the AMTP1-dominated binding mode and the cooperative binding mode, which likely correspond to two intermediates in the reaction pathway of DNA methylation catalyzed by AMT1 complex.

Offsite DNA binding

AMTP1 DBD is positioned at the major groove and near the ApT duplex in both cooperative and AMTP1-dominated binding modes, representing onsite binding. AMTP1 DBD can be positioned further away from the ApT duplex, as indicated by increased distances between 6mA/A and AMTP1 M64, relative to the cooperative binding mode (Figure 5F). This AMTP1 offsite binding mode represents a continuum of structures, in which AMTP1 DBD progressively shifts away from the ApT site. AMTP1 DBD initially moves along the major groove but ultimately reaches the minor groove. AMTP1 DBD offsite binding is

mediated by sequence non-specific interactions, including with the sugar-phosphate backbone (Figure S8A). Conformational heterogeneity of the AMTP1 offsite binding mode is further compounded by weak interactions. Indeed, AMTP1 DBD by itself and even the full length AMTP1 had no affinity for dsDNA (Figure S8B). For AMT1 ternary complexes, AMTP1 offsite binding represents only a small minority of AF3 models in the base flipping mode, but substantially more in the base pairing mode; AMTP1 offsite binding is common in AMT1 binary complexes, especially in the base pairing mode (Figure 5F, Figure S7D). AMTP1 offsite binding mode, just like AMTP1-dominated binding mode, is characterized by physical separation of AMTP1 DBD and AMT1 *Major Groove Insertion Loop*, evident from increased distances between AMT1 S90 and AMTP1 G60/Y63 (Figure 5H).

AMTP1 offsite binding generally occurs when AMT1 *Minor Groove Insertion Loop* still engages the ApT site (Figure S8A). Indeed, this is the most stable protein-DNA interface in AMT1 holo-complex (Figure 2E, Figure S4A), in strong contrast to dynamic DNA binding by AMT1 *Major Groove Insertion Loop* and AMTP1 DBD. This provides a simple reaction pathway: AMT1 is initially anchored by its *Minor Groove Insertion Loop* engaging the ApT site from the minor groove; it subsequently pivots around the target strand and contacts the ApT site from the major groove with its *Major Groove Insertion Loop*. This is coordinated with the motion of AMTP1 DBD, which cooperatively secures the base flipping mode critical for DNA methylation. Intriguingly, AMT1 *Minor Groove Insertion Loop* can also engage DNA away from the ApT site, representing AMT1 offsite binding mode that is readily observed only in AMT1 binary complex, especially with unmethylated ApT (Figure 5J, Figure S8A, Figure S8C). AMT1 offsite binding mostly tracks the minor groove and distributes rather evenly among non-ApT sites along the dsDNA substrate (Figure 5J, Figure S8A, Figure S8C). This strongly suggests a scanning mechanism for processive methylation, by which translocation of AMT1

complex along dsDNA allows efficient one-dimensional search. This requires continuous DNA association facilitated by the closed form AMT1 complex.

Interactions securing the closed form

In the closed form AMT1 holo-complex, AMTP1 DBD occupies the major groove, wraps around the non-target strand, and reaches AMT6/AMT7 MT-A70, forming an interface absent in apo-complex, hereafter referred to as the clasp. In other words, the clasp, with underlaying DNA, connects AMTP1 and AMT6/AMT7 in AMT1 holo-complex (Figure 6A). In AMT1 ternary complex (AMT7-full-SAH), most contacts in the clasp are between the N-terminal portion of AMTP1 *Clasp Loop* and the C-terminal portion of AMT7 *Multiple interaction Loop*. The long side chains of AMT7 E280, R282, and R285 can reach across the interface and form hydrogen bonds/salt bridges with the main/side chains of AMTP1 N107, G109, D110, and N111, as well as the DNA backbone (Figure 6A). The same AMT7 region also inserts into the minor groove and interacts with AMT1 *Minor Groove Insertion Loop* (Figure 2C, Figure S3D). Indeed, this part of AMT7 *Multiple interaction Loop* (280-285: ELRHQR) is absolutely conserved among AMT6 as well as AMT7 homologues (File S2). Its interaction partners on AMTP1 *Clasp Loop* (107-111: NTGDN) are also mostly conserved (File S3). Hydrogen bonds/salt bridges can also be formed between the side chains of D110 on AMTP1 *Clasp Loop* and K320 on AMT7 *Hydrophobic Loop* (Figure 6A). AMT7 K320 is also conserved in AMT6 (K242) (Supplemental File S3). Similar interactions were observed in other AMT1 ternary complexes (Figure S9A), supporting the clasp as a general feature securing the closed form.

We tracked the distances between AMTP1 D110 and AMT6/AMT7 K242/320 and R204/282 (Figure 6B, C). In AMT1 ternary complexes, these values mostly cluster around distances optimal for interactions, supporting

clasp formation (Figure 6B, C). There is a small population of AF3 models with substantial increases in both distances, which are incompatible with clasp formation, thus representing the open form (Figure 6B, C). Such AF3 models increase dramatically in AMT1 binary holo-complex and become predominant in apo-complex (Figure 6B, C). The dynamic nature of the clasp is required for loading and unloading of the dsDNA substrate. The hooking and unhooking of the clasp dictate the switch between the open form predominant in apo-complex and the closed form predominant in holo-complex. By securing the close form, the clasp can potentially increase DNA affinity and methylation processivity of AMT1 complex. Functional importance of the clasp was confirmed by reduced in vitro MTase activity of AMT1 complex incorporating the mutation AMTP1 D110K (Figure 6D). *Tetrahymena* D110K mutant also showed diminished endogenous 6mA levels (Figure 6E)

Structural basis for differential activities of AMT6 and AMT7 complexes

There are two distinct AMT1 complexes in *Tetrahymena*: AMT7 complex, with higher DNA binding and catalytic activities, is important for processive methylation, while AMT6 complex, with lower activities, is much less processive[6]. We next investigated the structural basis for their differential activities. Most regions involved in DNA binding are conserved across AMT6 and AMT7, except for the N-terminal portion of the *Multiple Interaction Loop* (Figure 6F). This region, as part of the clasp, is represented by divergent sequences in AMT6 (195-200: F·NKDA**Q**) and AMT7 (272-278: IGDKNQ**K**) (Figure 6F). Critically, the positively charged K278 in AMT7 is replaced by the neutral Q200 in AMT6. In support of differential interactions, both AF3 modeling and MD simulation showed that AMT7 K278 was placed closer to the DNA backbone and less diffusively distributed than AMT6 Q200, especially in binary complexes (Figure 6F, Figure S9B). The differential stability of this

variance region is propagated to the conserved C-terminal portion of *Multiple Interaction Loop*. Indeed, *Multiple interaction Loop* as a whole was structurally better aligned for MD-optimized AF3 models of AMT7 binary complex, relative to its AMT6 counterpart (AMT7-hemi vs. AMT6-hemi) (Figure 6G, H, Figure S9C). This effect was further amplified by the clasp, leading to a much more stable placement of AMTP1 DBD in AMT7 complex (Figure 6G, H, Figure S9D).

Conformational analysis monitoring clasp formation also revealed substantial differences between AMT6 and AMT7 holo-complexes (Figure 6B, C). For most AF3 models of AMT6 and AMT7 ternary complexes, the distance between AMTP1 *Clasp Loop* (D110) and AMT6/AMT7 *Hydrophobic Loop* (K242/K320) was small (≤10 Å), consistent with well-formed clasp (Figure 6B, C). However, in cofactor-free binary complexes, there were more AF3 models for which this distance was large (>10 Å), indicating clasp formation issues (Figure 6I). Consistent with experimental observations[6], such AF3 models were present in much higher proportions in AMT6 than AMT7 binary complexes (Figure 6I). Indeed, clasp formation was the only conformational analysis in which AMT6 and AMT7 complexes show substantial differences.

Discussion

AMT1-catalyzed DNA methylation is an intrinsically dynamic process, as it entails loading and unloading of DNA substrate and cofactor, as well as target adenine flipping in and out of the active site. This dynamic behavior is attested by conformational heterogeneity in AF3 models of AMT1 apo- and holo-complexes. Intriguingly, distribution of conformations is readily perturbed by AF3 input, particularly ***1)*** presence or absence of DNA substrate and cofactor, ***2)*** DNA substrate with different methylation states, and ***3)*** alternative protein subunit. By extensive AF3 modeling with a full array of input settings, we have surveyed the conformational space of AMT1 complex. Here, we provide a comprehensive analysis based on Uniform Manifold Approximation and Projection (UMAP) of key distances tracking dynamic protein-protein and protein-DNA interfaces in AMT1 complex, the prototypical eukaryotic DNA 6mA-MTase (Figure 7A-D, Figure S10-13). This allows us to home in on conformations critical for catalysis and delineate the reaction pathway for DNA adenine methylation (Figure 7E).

Catalytically competent conformation of AMT1 complex

We first focus on AMT1 ternary complex (AMT7-full-SAH), corresponding to the EP complex of maintenance methylation. Approximately 30% of its AF3 models are tightly clustered in UMAP (Figure 7B: middle). These AF3 models are highly uniform with regard to the following structural features tracked by individual sets of parameters: ***1)*** single 6mA flipping out of dsDNA, ***2)*** single 6mA flipping into AMT1 active site, ***3)*** onsite binding by AMT1 *Minor Groove Insertion Loop*, *Major Groove Insertion Loop*, and *Regulation Loop* at the target site, ***4)*** onsite binding by AMTP1 DBD, ***5)*** cooperative binding by AMT1 *Major Groove Insertion Loop* and AMTP1 DBD, ***6)*** AMTP1 *N-terminal Loop* docking at AMT1 active site, and ***7)*** clasp formation (Figure S11, S12). They

also strongly overlap with high-score AF3 models (Figure 7B, Figure S12). These structures are characterized by severe distortions in dsDNA, including widening of major and minor grooves, underwinding and bending of DNA backbone, as well as base flipping. dsDNA is squeezed by AMT1-AMT7 heterodimer and AMTP1 DBD, which ultimately drives the AMT1 intercalating residues into base stacks, leading to target adenine flipping. As a result of protein-protein and protein-DNA interactions, adenine and cofactor placement in AMT1 active site is rigidly defined and optimized for methyl transfer. We therefore identify these structures as the catalytically competent conformation (CCC) of AMT1 ternary complex (C1) (Figure 7A, B). CCC, especially AMT1 active site at its core, is essentially the same in the EP complex of de novo methylation (AMT7-hemi-SAH; C4), as well as in their AMT6 counterparts (AMT6-full-SAH and AMT6-hemi-SAH) (Figure 7A, C, Figure S10, S13). Many critical features in the active site of AMT1 complex are conserved in METTL3-METTL14, thus providing a unified catalytic mechanism for MT-A70 family of eukaryotic amino MTases. The structural insight can promote rational design of small molecule inhibitors for METTL3-METTL14, which have been extensively explored for their potential in cancer therapeutics[33,34].

Reaction pathway for DNA adenine methylation

CCC of AMT1 complex provides a critical snapshot in the reaction pathway for DNA adenine methylation. Consistent with induced fit enzymatic catalysis, CCC assembly depends on both substrate and cofactor binding. Other distinct conformations are readily observed, especially in AF3 models of cofactor-free AMT1 binary holo-complex and DNA-free apo-complex (Figure 7D, Figure S10D). Even minority classes of conformations may provide important insight into the reaction pathway for AMT1-dependent DNA methylation. By systematically characterizing conformational heterogeneity in AMT1 complex,

we reveal a full array of pivotal intermediates, envision the transition between them, and delineate the reaction pathway.

CCC assembly involves long-range movements of several protein regions, especially AMT1 *Major Groove Insertion Loop* and AMTP1 DBD with its N-terminal tail (Figure 7B-D, Figure S10B-D). Conformational heterogeneity of AMT1 *Major Groove Insertion Loop*, especially its distal portion in binary holo-complex and apo-complex, fits its role as a dynamic "lid", which controls the target adenine flipping, both out of and back into the dsDNA substrate, to initiate and conclude the methylation reaction. In CCC, the distal portion of AMT1 *Major Groove Insertion Loop* is caught in between underlaying DNA and overlaying AMTP1 DBD. AMTP1 DBD is in turn anchored by its long C-terminal linker to AMTP1 helix-turn-helix-like domain and AMT1 N-terminal extension module, by its short N-terminal linker to AMTP1 *N-terminal Loop* and AMT1 active site, and by the clasp to AMT6/AMT7 MT-A70 module. The movement of AMTP1 DBD dictates the transition between the open and closed forms of AMT1 complex. Furthermore, AMTP1 *N-terminal Loop* fills the space between AMT1 active site and DNA major groove, functioning like a linchpin by further limiting their relative motion. This double-lock mechanism, by coordinating multiple interactions for synergistic assembly of CCC, plays a critical role in stabilizing the transition state for methyl transfer.

Under both in vivo and in vitro conditions, AMT1-dependent DNA adenine methylation is ApT-specific[5]. Our structural modeling reveals that the ApT duplex is recognized by not only CCC in the base flipping mode (C1 and C4), but also a distinct conformation of binary holo-complex in the base pairing mode (C9) (Figure 7B-D, Figure S10B-D). ApT recognition mostly occurs in the major groove, mediated by AMT1 *Major Groove Insertion Loop* and AMTP1 DBD. In the cooperative binding mode (C1 and C4), AMT1 *Major Groove Insertion Loop* mainly engages the target strand, while also interacting with the

non-target strand; in the AMTP1-dominated mode (C9), AMTP1 DBD mainly engages the non-target strand, but it can also discriminate against the target 6mA. In both modes, there are structural features that interact more favorably with 6mA on the non-target strand, while also accommodating unmodified adenine. Dual recognition by AMT1 complex accounts for its strong maintenance methylation activity and the weaker, but still substantial, de novo methylation activity. We envision a reaction pathway in which the ApT duplex in the base pairing mode is rapidly located by AMTP1 DBD in AMT1 binary complex, before cofactor loading, base flipping, and target verification by AMT1 *Major Groove Insertion Loop* in AMT1 ternary complex. This double-check mechanism allows for both efficient and accurate target sequence recognition.

Due to the semiconservative nature of 6mA transmission in *Tetrahymena*[5], the target strand of DNA methylation is identical to the newly synthesized strand in DNA replication, while the non-target strand corresponds to the template strand. Processive methylation of the target/newly synthesized strand occurs in an AMT7-dependent manner in *Tetrahymena*[6]. Based on our structural modeling, we posit that this may be underpinned by a scanning process, in which AMT1 binary complex performs one-dimensional search for target ApT sites by translocating along the dsDNA substrate. In this context, the offsite DNA binding modes of AMT1 and AMTP1 (C11 and C12) represent key intermediates during the scan, as AMT1 complex moves from one target site to another (Figure 7A, Figure S10A). Scanning is promoted by the close form AMT1 complex, which is secured by the clasp. Importantly, reduced clasp stability in AMT6 complex, especially its binary holo-complex, can account for its reduced DNA affinity and methylation processivity, relative to AMT7 complex[6].

We conclude by positing an integrated reaction pathway for processive maintenance methylation catalyzed by AMT7 complex (Figure 7E). First, AMT7 apo-complex, in its open form, is loaded onto dsDNA substrate. The binary holo-complex is then converted into the closed form by clasp formation. Through the scanning process, AMT7 complex is localized to a target site, mediated by AMTP1 DBD recognition of the hemi-6mApT duplex in the base pairing mode. Subsequent base flipping, cofactor loading, and long-range movement of protein regions (including AMT1 *Major Groove Insertion Loop* and AMTP1 DBD) lead to CCC assembly, recognition of hemi-6mApT in the base flipping mode, and catalysis of methyl transfer. CCC disassembly and cofactor unloading converts the ternary complex back into the binary complex, ready to translocate along dsDNA to another target site for processive methylation. Only when the clasp unravels can the open form AMT7 complex unload from dsDNA.

Materials and Methods

AMT6/AMT7 complexes structure prediction using AlphaFold3

To model the structures of AMT6/7 complexes, we employed AlphaFold3.0.0[31] running within a Singularity containerized environment on a high-performance GPU cluster. Structure prediction was executed in five batches, each consisting of 200 independent prediction seeds to enhance conformational sampling. Each batch was submitted via SLURM with resource specifications of 12 CPU cores, 2 NVIDIA A100 GPUs, and 100 GB RAM per node, under a maximum runtime of 48 hours. The pipeline utilized a customized JSON file for each input batch, specifying the molecular system and sequence configuration. The structure prediction was executed using run_alphafold.py with appropriate paths to the model directory and database resources. All structure predictions were carried out with default AlphaFold3 inference settings.

Reaction coordinates distance profiling and dimensionality reduction

To characterize the conformational landscape across a series of protein complexes, we computed interatomic distances between predefined atom pairs representing key structural features. Distance matrices derived from eight distinct structural descriptors were integrated into a unified matrix, where rows correspond to individual AMT6/AMT7 complex structures predicted by AlphaFold3 and columns represent the eight structural features. Unsupervised dimensionality reduction and visualization were performed on this integrated distance matrix using Uniform Manifold Approximation and Projection (UMAP) with default parameters (R umap package[43]).

Classify AF3 models by reaction coordinates

By combining the corresponding distances of both binary and ternary complexes of AMT6 and AMT7 models, we then used UMAP algorithm to reduce the dimension for further visualization. UMAP results were divided into 24 groups, and representative structures were extracted by projecting the corresponding groups back to the 2-dimensional scatter plots.

MD simulations

MD simulations were performed to refine the local side-chain conformations of AMT6 and AMT7 holo-complexes bound to full-methylation dsDNA and SAH co-substrates. Three distinct target systems were simulated to capture key functional states of the flipped 6mA base: (1) Single-base flipping transition state: Systems included A7-full-SAH, A6-full-SAH, A7-hemi-SAH, and A6-hemi-SAH, with 6mA flipped out and stabilized in the sp^3-hybridized conformer state (Conformer 3, as defined for 6mA in PubChem CID 67955). (2) Single-base flipping ground state: Identical systems as (1), but with 6mA in the Conformer 1 state, as defined in PubChem CID 67955. (3) Double-base paired state: A7-hemi-SAH and A6-hemi-SAH with 6mA proximal/distal to SAH, adopting the sp^2-hybridized Conformer 2 state in PubChem CID 67955.

For each target system, the initial structures were derived from AF3 predictions. The top ten highest-confidence models were selected per target based on per-residue pLDDT scores, ensuring backbone reliability.

MD simulations employed CHARMM[35] version c49b1 with the all-atom additive CHARMM36 force field[44] for proteins and nucleic acids version 36. The SAH cofactor, absent in standard CHARMM parameters, was processed as follows: missing hydrogens in SAH were added using Avogadro[45] v1.2.0 to ensure correct protonation states; SAH topology and parameters were generated via the CGenFF[36] server, leveraging the Merck Molecular Force

Field (MMFF94s) for charge assignment and bonded terms. Penalty scores for SAH parameters were <10, indicating robust compatibility. The complexes (protein-DNA-SAH) were then assembled and preprocessed using CHARMM-GUI[46]. Pre-simulation hybridization states of 6mA conformers were validated and assigned using automated ChimeraX[47] v1.8 scripts. Systems were solvated with CHARMM-modified TIP3P water, optimized for protein-water dispersion interactions, and neutralized with stochastically distributed potassium or chloride ions in the solvent volume to prevent biased electrostatic artifacts.

A two-stage protocol of energy minimization was applied. The steepest descent (5,000 steps) was used to eliminate severe clashes. Adopted basis Newton–Raphson (ABNR) minimization was then applied until convergence (energy gradient <10^{-} kcal·mol^{-1}·$Å^{-1}$). Energy minimization ensured systems resided at local potential-energy minima. This refinement corrected AF3-predicted deviations in side-chain rotamers of catalytic residues, and base-stacking geometries in dsDNA, particularly near flipped 6mA. Final structures after energy relaxation were used for the downstream analysis and visualization. This protocol establishes a reproducible framework for refining AF3-predicted enzyme-DNA complexes.

Generation of *Tetrahymena* strains

The *AMT1, AMT7* and *AMTP1* mutations were respectively introduced to *AMT1*-NHA, *AMT7*-CHA and *AMTP1*-CHA construct by fusion PCR with primers containing the designed mutations[6]. These constructs were individually transformed into wild type (WT) *Tetrahymena* cells (SB210) and selected in increasing concentration of paromomycin to generate a somatic gene mutation strain[48,49].

UHPLC-QQQ-MS/MS analysis

Tetrahymena genomic DNA was purified and treated following established protocols[6]. Samples were analyzed by ultra-high-performance liquid chromatography-tandem mass spectrometry (UHPLC-QQQ-MS/MS) on an Acquity BEH C18 column (75 mm × 2.1 mm, 1.7 μm, Waters, MA, USA), using a Xevo TQ-XS triple quadrupole mass spectrometer (Waters, Milford, MA, USA) or a Hypersil GOLD column (100 x 2.1 mm, 1.9 μm, Thermo scientific) using a TSQ Quantiva mass spectrometer (Thermo scientific). The mass spectrometer was set to multiple reaction monitoring (MRM) in positive electrospray mode. The selective MRM transitions were detected under m/z 266/150 for 6mA and m/z 252/136 for dA. The ratio of 6mA/A was quantified by the calibration curves of nucleoside standards running at the same time. UHPLCQQQ-MS/MS was performed at the Key Laboratory of Marine Medicine at Ocean University of China (OUC).

Recombinant protein purification

Full length AMT1, AMT6, AMT7, AMTP1, AMTP2 ORFs were codon-optimized for bacterial expression as previously mentioned[6]. AMT1 was cloned into a pRSF-Duet vector. AMTP1 and AMTP2 were cloned into a pET-Duet vector. AMT7 was cloned into a pET-28b vector. Mutations were generated by fusion PCR. WT or mutant AMT1 subcomplex components (AMT1, AMTP1 and AMTP2) were co-expressed in *E. coli* BL21 (DE3) cells with a 6 × His-SUMO tag attached to AMT1 and AMTP2 N-termini. Wild type or mutant AMT7 possessing a 6 × His-SUMO tag was expressed in *E. coli* strain BL21 (DE3).

E. coli was cultured to an optical density (OD600) of 0.8 at 37°C and protein expression was induced by 0.25 mM IPTG at 16°C overnight. Induced

cells expressing subcomplex or AMT7 protein were mixed and harvested. The cells were then resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 25 mM imidazole, 10% glycerol) and homogenized with an ultra-high-pressure cell disrupter at 4°C. Lysates were cleared by centrifugation at 13,000 g for 45 min at 4°C. Proteins were first purified by Ni-NTA (Novagen, USA) affinity chromatography and eluted through on-column tag cleavage by the ULP1 protease. The elute was further purified by passage through a Heparin ion-ex-change column (GE Healthcare, USA) with buffer A (20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 5% glycerol) and buffer B (20 mM Tris-HCl, pH 7.5, 1 M NaCl, 5% glycerol), before being loaded onto a Superdex 200 10/300 Increase column (GE Healthcare, USA) with buffer C (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 4 mM DTT).

Methyltransferase assay

For radioactive methyltransferase assay, 0.1 μM protein was mixed with 1 μM ^{3}H-labeled S-adenosyl-L-methionine ([^{3}H]SAM, Perkin Elmer) and 1μM dsDNA in 15 μLreaction buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.05% β-Me). Samples were incubated at 30°C for 30 min and subsequently spotted onto Hybond-XL membrane (GE Healthcare). Membranes were then washed three times with 0.2 M ammonium bicarbonate, once with distilled water, twice with 100% ethanol, and finally air-dried for 1 h. Each membrane was immersed in Ultima Gold (PerkinElmer) and used for scintillation counting on a PerkinElmer MicroBeta2 (PerkinElmer).

Electrophoretic mobility shift assay (EMSA)

EMSA was performed using dsDNA and proteins purified as above. Briefly, dsDNA (4 μM) was incubated with a gradient dose of 2, 4, and 8 μM full length

or truncated AMTP1 protein in buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1mM EDTA, 0.05% β-Me for 30 min at 4°C. 2 μM AMT7 complex proteins was used as positive control. Reactions were run on a 6% native polyacrylamide gel (acrylamide:bis-acrylamide=37.5:1) in 1 x TBE buffer at 100 V for 30 min. The gel was stained with ethidium bromide.

In vitro cross-linking of AMT1 complex and DNA

Purified AMT6 and AMT7 complexes were incubated with synthetic oligonucleotides and SAH at 4°C for 30min, and further purified by size exclusion chromatography on a Superdex 200 increase 10/300 column (GE Healthcare). The purified protein-DNA complexes were subjected to cross-linking experiments as per the modified procedures adapted from previously established protocols[6]. Cross-linking mass spectrometry was performed at the Technology Center for Protein Sciences (Proteomics Facility), Tsinghua University.

Acknowledgements

S.G. was supported by the National Natural Science Foundation of China (32125006), Laoshan Laboratory (LSKJ202203203), and the Natural Science Foundation of Shandong Province (ZR2024ZD40). Y.L. was supported by the NSF award (MCB-2435178). D.W. and M.X. were supported by the National Key R&D Program of China (2025YFC2817103). High-performance computing resources for data processing were provided by the Center for Advanced Research Computing (CARC) at the University of Southern California, as well as the Institute of Evolution and Marine Biodiversity IEMB1, the High-Performance Biological Supercomputing Center, and Marine Big Data Center of Institute for Advanced Ocean Study at OUC. Our special thanks are

given to Prof. Weibo Song (OUC) for his helpful suggestions during drafting the manuscript.

Author contribution

Y.L. and S.G. conceived the study and designed the experiments; B.N. and L.N. performed most of the experiments; W.Y. performed AlphaFold 3 prediction; Y.W. collected and organized the data. M.X. performed molecular dynamics simulation; D.W. and W.Y. performed conformational heterogeneity analysis; W.G., Z.C. and G.Z performed the radioactive methylation assay; B.N., Y.L. and S.G. wrote the paper, with inputs from all authors. All authors read and approved the final manuscript.

The authors declare no conflict of interest.

References

1. Bochtler, M., and Fernandes, H. (2021). DNA adenine methylation in eukaryotes: Enzymatic mark or a form of DNA damage? BioEssays *43*, 2000243. https://doi.org/10.1002/bies.202000243.

2. Fu, Y., Luo, G.-Z., Chen, K., Deng, X., Yu, M., Han, D., Hao, Z., Liu, J., Lu, X., Doré, Louis C., et al. (2015). N^6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. Cell *161*, 879-892. https://doi.org/10.1016/j.cell.2015.04.010.

3. Mondo, S.J., Dannebaum, R.O., Kuo, R.C., Louie, K.B., Bewick, A.J., LaButti, K., Haridas, S., Kuo, A., Salamov, A., Ahrendt, S.R., et al. (2017). Widespread adenine N^6-methylation of active genes in fungi. Nature Genetics *49*, 964-968. https://doi.org/10.1038/ng.3859.

4. Wang, Y., Chen, X., Sheng, Y., Liu, Y., and Gao, S. (2017). N^6-adenine DNA methylation is associated with the linker DNA of H2A.Z-containing well-positioned nucleosomes in Pol II-transcribed genes in *Tetrahymena*. *Nucleic Acids Research 45*, 11594-11606. https://doi.org/10.1093/nar/gkx883.

5. Sheng, Y., Wang, Y., Yang, W., Wang, X.Q., Lu, J., Pan, B., Nan, B., Liu, Y., Ye, F., Li, C., et al. (2024). Semiconservative transmission of DNA N^6-adenine methylation in a unicellular eukaryote. Genome Research *34*, 740-756. https://doi.org/10.1101/gr.277843.123

6. Wang, Y., Nan, B., Ye, F., Zhang, Z., Yang, W., Pan, B., Wei, F., Duan, L., Li, H., Niu, J., et al. (2025). Dual modes of DNA N^6-methyladenine maintenance by distinct methyltransferase complexes. Proceedings of the National Academy of Sciences *122*, e2413037121. https://doi.org/10.1073/pnas.2413037121.

7. Bai, L., Yang, G., Qin, Z., Lyu, J., Wang, Y., Feng, J., Liu, M., Gong, T., Li, X., Li, Z., et al. (2021). Proteome-Wide Profiling of Readers for DNA Modification. Advanced Science *8*, 2101426. https://doi.org/10.1002/advs.202101426.

8. Ye, F., Chen, X., Li, Y., Ju, A., Sheng, Y., Duan, L., Zhang, J., Zhang, Z., Al-Rasheid, K.A.S., Stover, N.A., and Gao, S. (2025). Comprehensive genome annotation of the model ciliate *Tetrahymena thermophila* by in-depth epigenetic and transcriptomic profiling. Nucleic Acids Research *53*, gkae1177. https://doi.org/10.1093/nar/gkae1177.

9. Iyer, L.M., Abhiman, S., and Aravind, L. (2011). Chapter 2 - natural history of eukaryotic DNA methylation systems. In Progress in Molecular Biology and Translational Science, X. Cheng, and R.M.

Blumenthal, eds. (Academic Press), pp. 25-104. https://doi.org/10.1016/B978-0-12-387685-0.00002-0.

10. Iyer, L.M., Zhang, D., and Aravind, L. (2016). Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. Bioessays *38*, 27-40. https://doi.org/10.1002/bies.201500104.

11. Cheng, T., Zhang, J., Li, H., Diao, J., Zhang, W., Niu, J., Kawaguchi, T., Nakayama, J.-i., Kataoka, K., and Gao, S. (2025). Identification and characterization of the de novo methyltransferases for eukaryotic N^6-methyladenine (6mA). Science Advances *11*, eadq4623. https://doi.org/10.1126/sciadv.adq4623.

12. He, P.C., and He, C. (2021). m^6A RNA methylation: from mechanisms to therapeutic potential. The EMBO Journal *40*, e105977. https://doi.org/10.15252/embj.2020105977.

13. Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3–METTL14 complex mediates mammalian nuclear RNA N^6-adenosine methylation. Nature Chemical Biology *10*, 93-95. https://doi.org/10.1038/nchembio.1432.

14. Beh, L.Y., Debelouchina, G.T., Clay, D.M., Thompson, R.E., Lindblad, K.A., Hutton, E.R., Bracht, J.R., Sebra, R.P., Muir, T.W., and Landweber, L.F. (2019). Identification of a DNA N^6-adenine methyltransferase complex and its impact on chromatin organization. *Cell 177*, 1781-1796.e1725. https://doi.org/10.1016/j.cell.2019.04.028.

15. Duan, L., Li, H., Ju, A., Zhang, Z., Niu, J., Zhang, Y., Diao, J., Liu, Y., Song, N., Ma, H., et al. (2025). Methyl-dependent auto-regulation of the DNA N^6-adenine methyltransferase AMT1 in the unicellular eukaryote *Tetrahymena thermophila*. Nucleic Acids Research *53*, gkaf022. https://doi.org/10.1093/nar/gkaf022.

16. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., and Roberts, R.J. (1993). Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-l-methionine. Cell *74*, 299-307. https://doi.org/10.1016/0092-8674(93)90421-L.

17. Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. Cell *76*, 357-369. https://doi.org/10.1016/0092-8674(94)90342-5.

18. Goedecke, K., Pignot, M., Goody, R.S., Scheidig, A.J., and Weinhold, E. (2001). Structure of the N^6-adenine DNA methyltransferase M•TaqI in

complex with DNA and a cofactor analog. Nature Structural Biology *8*, 121-125. https://doi.org/10.1038/8410410.1038/84104.

19. Falnes, P.Ø. (2024). Closing in on human methylation—the versatile family of seven-β-strand (METTL) methyltransferases. Nucleic Acids Research *52*, 11423-11441. https://doi.org/10.1093/nar/gkae816.

20. Malone, T., Blumenthal, R.M., and Cheng, X. (1995). Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyl-transferases, and suggests a catalytic mechanism for these enzymes. Journal of Molecular Biology *253*, 618-632. https://doi.org/10.1006/jmbi.1995.0577.

21. Woodcock, C.B., Horton, J.R., Zhang, X., Blumenthal, R.M., and Cheng, X. (2020). Beta class amino methyltransferases from bacteria to humans: evolution and structural consequences. Nucleic Acids Research *48*, 10034-10044. https://doi.org/10.1093/nar/gkaa446.

22. Ren, R., Horton, J.R., Hong, S., and Cheng, X. (2022). Recent advances on DNA base flipping: a general mechanism for writing, reading, and erasing DNA modifications. In DNA Methyltransferases - Role and Function, A. Jeltsch, and R.Z. Jurkowska, eds. (Springer International Publishing), pp. 295-315. https://doi.org/10.1007/978-3-031-11454-0_12.

23. Śledź, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m^6A writer complex. eLife *5*, e18434. https://doi.org/10.7554/eLife.18434.

24. Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., Gong, Z., Wang, Q., Huang, J., Tang, C., et al. (2016). Structural basis of N^6-adenosine methylation by the METTL3–METTL14 complex. Nature *534*, 575-578. https://doi.org/10.1038/nature18298.

25. Wang, P., Doxtader, Katelyn A., and Nam, Y. (2016). Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases. Molecular Cell *63*, 306-317. https://doi.org/10.1016/j.molcel.2016.05.041.

26. Chen, J., Hu, R., Chen, Y., Lin, X., Xiang, W., Chen, H., Yao, C., and Liu, L. (2022). Structural basis for MTA1c-mediated DNA N^6-adenine methylation. Nature Communications *13*, 3257. https://doi.org/10.1038/s41467-022-31060-6.

27. Yan, J., Liu, F., Guan, Z., Yan, X., Jin, X., Wang, Q., Wang, Z., Yan, J., Zhang, D., Liu, Z., et al. (2023). Structural insights into DNA N^6-adenine

methylation by the MTA1 complex. Cell Discovery *9*, 8. https://doi.org/10.1038/s41421-022-00516-w.

28. Shao, Z., Yoon, S., Lu, J., Athavale, P., Liu, Y., and Song, J. (2025). Structural insight into the substrate binding of the AMT complex via an inhibitor-trapped state. bioRxiv, 2025.2004.2024.650485. http://biorxiv.org/content/early/2025/04/24/2025.04.24.650485.abstract.

29. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science *373*, 871-876. https://doi.org/10.1126/science.abj8754.

30. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589. https://doi.org/10.1038/s41586-021-03819-2.

31. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493-500. https://doi.org/10.1038/s41586-024-07487-w.

32. Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G.R., Morey-Burrows, F.S., Anishchenko, I., Humphreys, I.R., et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. Science *384*, eadl2528. https://doi.org/10.1126/science.adl2528.

33. Fiorentino, F., Menna, M., Rotili, D., Valente, S., and Mai, A. (2023). METTL3 from target validation to the first small-molecule inhibitors: a medicinal chemistry journey. Journal of Medicinal Chemistry *66*, 1654-1677. https://doi.org/10.1021/acs.jmedchem.2c01601.

34. Yankova, E., Blackaby, W., Albertella, M., Rak, J., De Braekeleer, E., Tsagkogeorga, G., Pilka, E.S., Aspris, D., Leggate, D., Hendrick, A.G., et al. (2021). Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. Nature *593*, 597-601. https://doi.org/10.1038/s41586-021-03536-w.

35. Hwang, W., Austin, S.L., Blondel, A., Boittier, E.D., Boresch, S., Buck, M., Buckner, J., Caflisch, A., Chang, H.-T., Cheng, X., et al. (2024). CHARMM at 45: enhancements in accessibility, functionality, and speed. The Journal of Physical Chemistry B *128*, 9976-10042. https://doi.org/10.1021/acs.jpcb.4c04100.

36. Vanommeslaeghe, K., Raman, E.P., and MacKerell, A.D., Jr. (2012). Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. Journal of Chemical Information and Modeling *52*, 3155-3168. https://doi.org/10.1021/ci3003649.

37. Lyu, L., Zhang, X., Gao, Y., Zhang, T., Fu, J., Stover, N.A., and Gao, F. (2024). From germline genome to highly fragmented somatic genome: genome-wide DNA rearrangement during the sexual process in ciliated protists. Marine Life Science & Technology *6*, 31-49. https://doi.org/10.1007/s42995-023-00213-x.

38. Sternglanz, H., and Bugg, C.E. (1973). Conformation of N^6-methyladenine, a base involved in DNA modification: restriction processes. Science *182*, 833-834. https://doi.org/10.1126/science.182.4114.833.

39. Roost, C., Lynch, S.R., Batista, P.J., Qu, K., Chang, H.Y., and Kool, E.T. (2015). Structure and thermodynamics of N^6-methyladenosine in RNA: a spring-loaded base modification. Journal of the American Chemical Society *137*, 2107-2115. https://doi.org/10.1021/ja513080v.

40. Bürglin, T.R., and Affolter, M. (2016). Homeodomain proteins: an update. Chromosoma *125*, 497-521. https://doi.org/10.1007/s00412-015-0543-8.

41. Court, R., Chapman, L., Fairall, L., and Rhodes, D. (2005). How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. EMBO reports *6*, 39-45-45. https://doi.org/10.1038/sj.embor.7400314.

42. Hanaoka, S., Nagadoi, A., and Nishimura, Y. (2005). Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities. Protein Science *14*, 119-130. https://doi.org/10.1110/ps.04983705.

43. Konopka, T., and Maintainer Tomasz Konopka (2018). R-package: umap.

44. Brooks, B.R., Brooks Iii, C.L., Mackerell Jr, A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al. (2009). CHARMM: the biomolecular simulation program. Journal of Computational Chemistry *30*, 1545-1614. https://doi.org/10.1002/jcc.21287.

45. Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., and Hutchison, G.R. (2012). Avogadro: an advanced semantic chemical

editor, visualization, and analysis platform. Journal of Cheminformatics *4*, 17. https://doi.org/10.1186/1758-2946-4-17.

46. Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. Journal of Computational Chemistry *29*, 1859-1865. https://doi.org/10.1002/jcc.20945.

47. Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H., and Ferrin, T.E. (2023). UCSF ChimeraX: tools for structure building and analysis. Protein Science *32*, e4792. https://doi.org/10.1002/pro.4792.

48. Mochizuki, K. (2008). High efficiency transformation of *Tetrahymena* using a codon-optimized neomycin resistance gene. Gene *425*, 79-83. https://doi.org/10.1016/j.gene.2008.08.007.

49. Hao, H., Lian, Y., Ren, C., Yang, S., Zhao, M., Bo, T., Xu, J., and Wang, W. (2024). RebL1 is required for macronuclear structure stability and gametogenesis in *Tetrahymena thermophila*. Marine Life Science & Technology *6*, 183-197. https://doi.org/10.1007/s42995-024-00219-z.

Figures

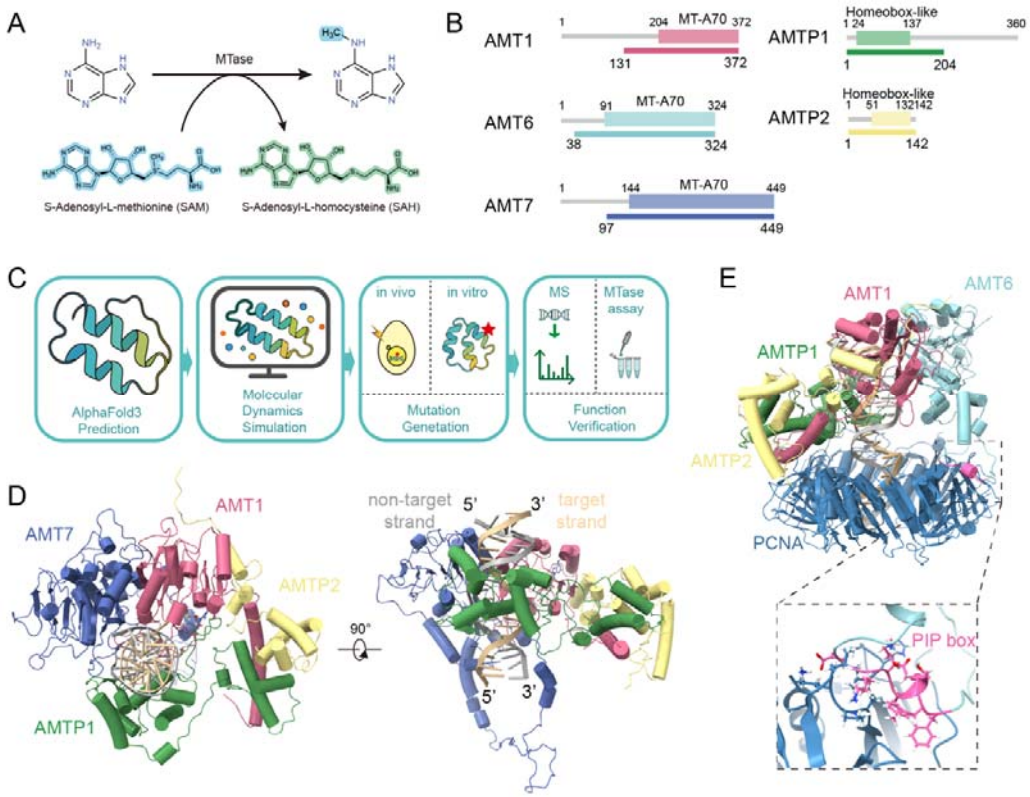

Figure 1. Overall structural organization of AMT1 holo-complex.

A. Methyl transfer from S-adenosyl-L-methionine (SAM) to adenine, generating N^6-methyladenine (6mA) and S-adenosyl-L-homocysteine (SAH).

B. Domain architecture of AMT1 complex subunits: AMT1, AMT6, AMT7, AMTP1, and AMTP2. The colored bars below represent sequences used in structural modeling.

C. Workflow for structural modeling and functional verification of AMT1 complex. After initial AF3 prediction, critical models were further optimized by MD simulations. Key residues underpinning predicted structural features were systematically mutated, and their functional relevance was verified by various in vitro and in vivo assays.

D. Front (left) and side (right) views of the AMT7 ternary complex (AMT7-full-SAH). Protein subunits AMT1, AMT7, AMTP1, AMTP2 are colored in pale violet red, pale denim, forest green and khaki, respectively. DNA target and non-target strands are colored in burly wood and silver, respectively.

E. Side view of the AMT6 ternary complex (AMT6-full-SAH) with tandemly loaded PCNA. AMT6 and PCNA are colored in light blue and steel blue, respectively. Expanded view: interactions between the PCNA-interaction protein (PIP) box (pink) of AMT6 and PCNA.

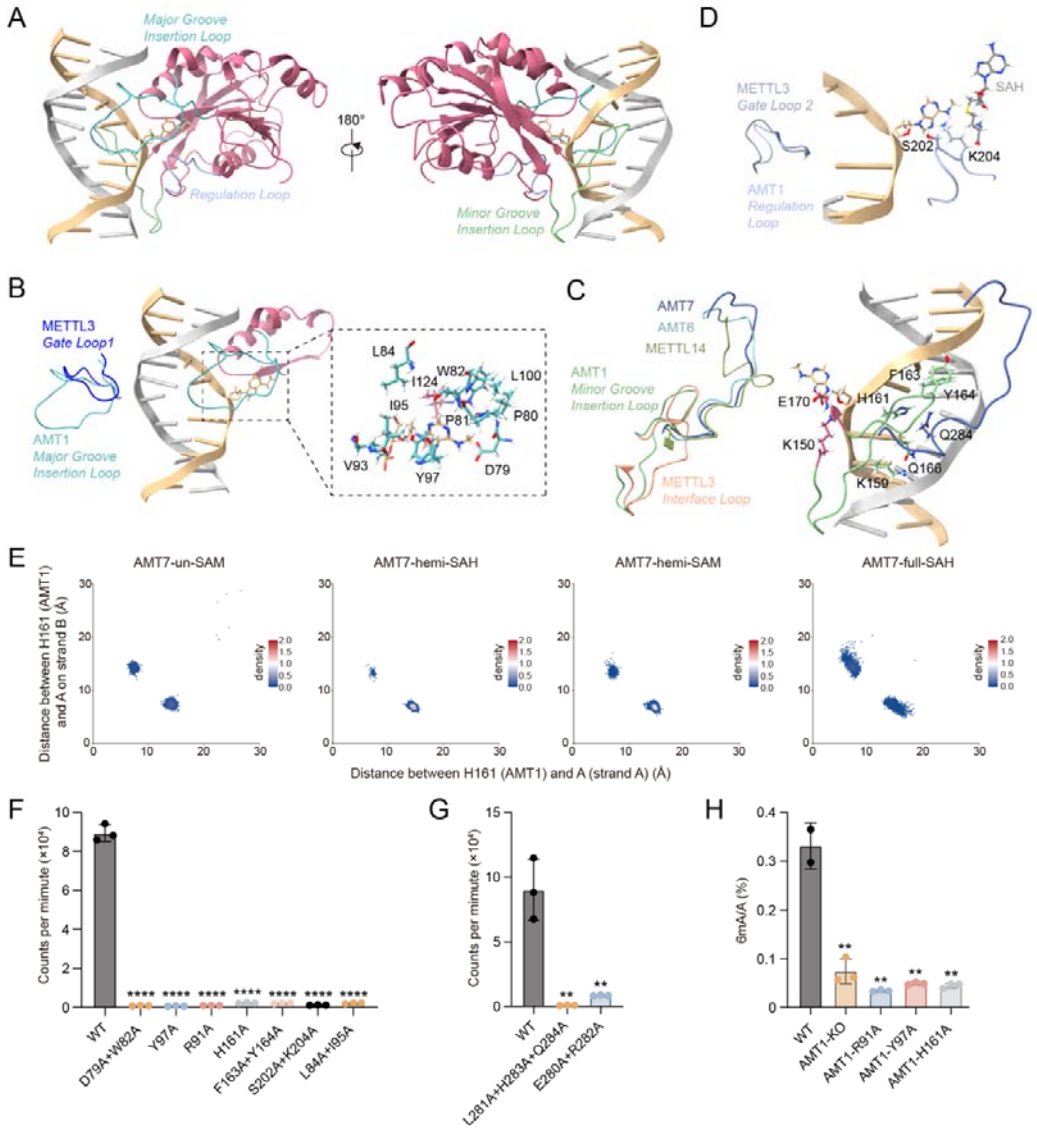

Figure 2. DNA binding by AMT1.

A. Opposite views of DNA binding by AMT1. AMT7 ternary complex (AMT7-full-SAH) in the base-flipping mode. Left/right: DNA major groove facing outward/inward. Three important AMT1 loops are highlighted: *Major Groove Insertion Loop* (saxe blue), *Minor Groove Insertion Loop* (dark sea green), and *Regulation Loop* (lavender). Extrahelical 6mA is shown in the stick model.

B. Close-up view of AMT1 *Major Groove Insertion Loop*. Left: structural

overlay of AMT1 *Major Groove Insertion Loop* and METTL3 *Gate Loop 1*. Right: DNA binding by AMT1 *Major Groove Insertion Loop*. Expanded view: residues critical for interacting with extrahelical 6mA.

C. Close-up view of AMT1 *Minor Groove Insertion Loop*. Left: structural overlay of AMT1 *Minor Groove Insertion Loop*, AMT6/AMT7 *Multiple Interaction Loop*, METTL3 *Interface Loop*, and METTL14 *Interface Loop*. Right: DNA binding by AMT1 *Minor Groove Insertion Loop* and AMT7 *Multiple Interaction Loop*.

D. Close-up view of AMT1 *Regulation Loop*. Left: structural overlay of AMT1 *Regulation Loop* and METTL3 *Gate Loop 2*. Right: AMT1 *Regulation Loop* interacting with substrate DNA (especially extrahelical 6mA) and cofactor SAH.

E. Conformational analysis tracking DNA binding by AMT1 *Minor Groove Insertion Loop*. AMT7 ternary complexes with different states of methylation (un: unmodified, hemi: hemi-methylation, full: full-methylation) and cofactor-binding (SAM/SAH). Scatterplot: distance between AMT1 H161 (C_α) and adenine (N^6) on DNA strand A/B (x/y-axis, Å); colored as per local density of AF3 models.

F. In vitro MTase activity of AMT7 complex with indicated point mutations of AMT1. Wildtype (WT) AMT7 complex was included as positive control. Data are shown as mean±SD. Two-tailed Student's t-test: ****$p < 0.0001$.

G. In vitro MTase activity of AMT7 complex with indicated point mutations of AMT7. Data are shown as mean±SD. Two-tailed Student's t-test: **$p < 0.01$.

H. Mass spectrometry quantification of 6mA-to-A ratio (6mA/A, %) in WT (positive control), Δ*AMT1* (negative control), and mutants with indicated point mutations of AMT1 or AMT7. Data are shown as mean±SD. Two-tailed Student's t-test: **$p < 0.01$.

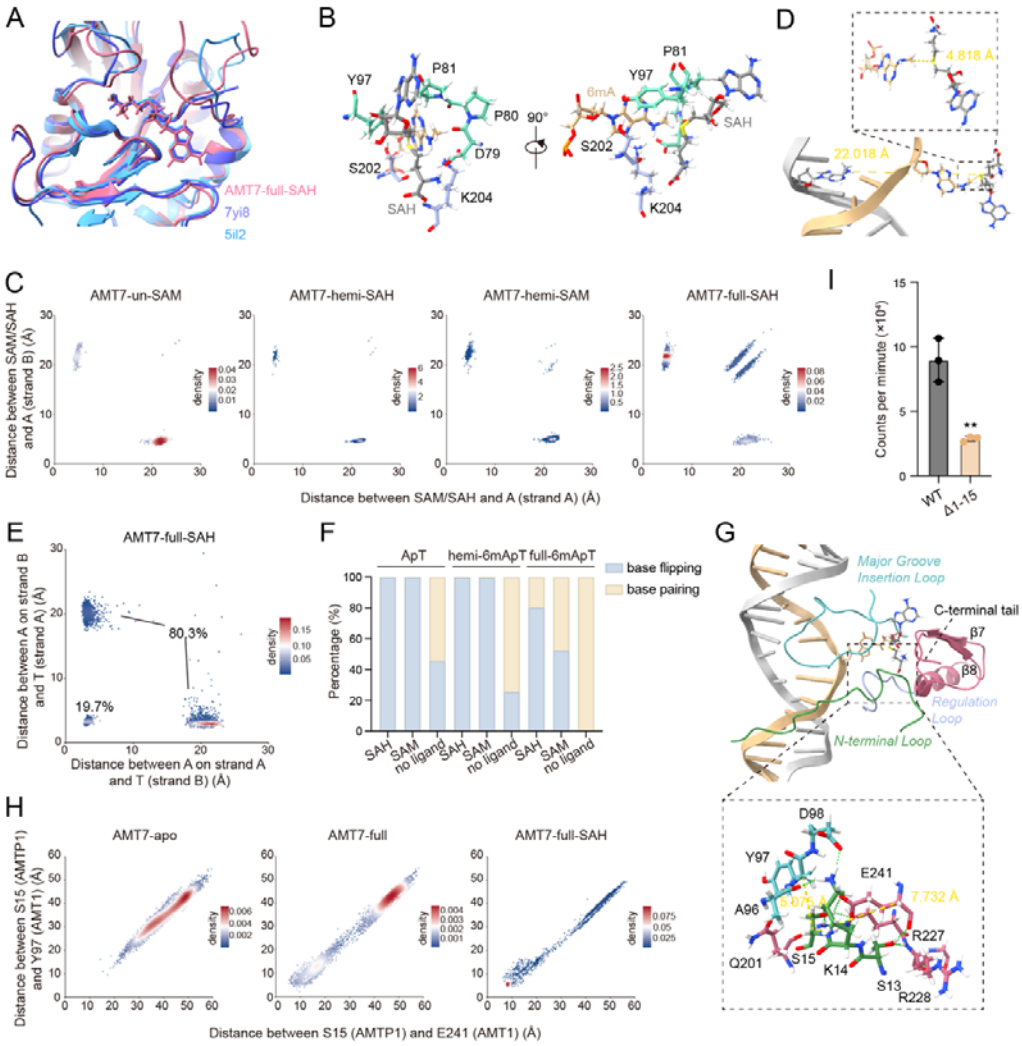

Figure 3. Structure-function relationship in AMT1 active site.

A. AMT1 cofactor-binding pocket loaded with SAH. Structural overlay of the AMT7 ternary holo-complex (AMT7-full-SAH, predicted) and AMT7 apo-complex (7yi8 and 5il2)[24,27].

B. Critical residues in the aperture separating cofactor SAH (near) and extrahelical 6mA (far).

C. Conformational analysis tracking base flipping into the active site. AMT7 ternary complexes with different states of methylation (un: unmodified, hemi: hemi-methylation, full: full-methylation) and cofactor-binding (SAM/SAH). Scatterplot: distance between cofactor (S_δ) and adenine

(N^6) on DNA strand A/B (x/y-axis, Å); colored as per local density of AF3 models.

D. Target strand 6mA flipping out of dsDNA and into the active site. Distance between target/non-target 6mA (N^6) and SAH (S_δ): 4.8/22.0 Å.

E. Conformational analysis tracking base flipping out of dsDNA in the AMT7 ternary complex (AMT7-full-SAH). Scatterplot: distance between adenine (N^6) on DNA strand A/B and thymine on DNA strand B/A (O^4) (x/y-axis, Å); colored as per local density of AF3 models.

F. The proportion of AF3 models in the base flipping or pairing mode for various AMT7 binary and ternary holo-complexes.

G. AMTP1 *N-terminal Loop* docking at the AMT1 active site. Expanded view: critical residues at the interface. Hydrogen bond: green dashed line. Distance between S15 (C_α) in AMTP1 and E241/Y97 (C_α) in AMT1: 7.7/5.1 Å.

H. Conformational analysis tracking AMTP1 *N-terminal Loop* docking at the AMT1 active site in the AMT7 apo-complex (AMT7), binary holo-complex (AMT7-full), and ternary holo-complex (AMT7-full-SAH). Scatterplot: distance between AMTP1 S15 (C_α) and AMT1 E241/Y97 (C_α) (x/y-axis, Å); colored as per local density of AF3 models.

I. In vitro MTase activity of AMT7 complex with AMTP1 N-terminal truncation (Δ1-15). Wildtype (WT) AMT7 complex was included as positive control. Data are shown as mean±SD. Two-tailed Student's t-test: $^{**}p < 0.01$.

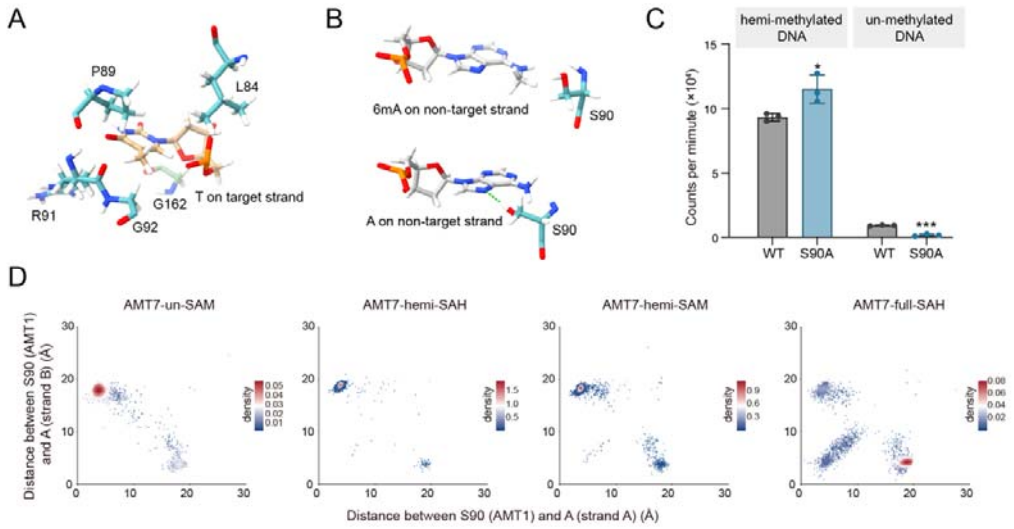

Figure 4. ApT duplex recognition by AMT1.

A. Target strand T recognition.

B. Non-target strand 6mA/A recognition. 6mA (top) and A (bottom) interact with alternative conformations of AMT1 S90. Hydrogen bond: green dashed line.

C. In vitro MTase activity of AMT7 complex with AMT1 S90A mutation. Wildtype (WT) AMT7 complex included as positive control. dsDNA substrates with hemi- or un-methylation were separately tested. Data are shown as mean±SD. Two-tailed Student's t-test: *$p<0.05$, ***$p<0.001$.

D. Conformational analysis tracking the distance between AMT1 S90 (C_α) and 6mA/A (N^6) on DNA strand A/B (x/y-axis, Å). AMT7 ternary complexes with different states of methylation (un: unmodified, hemi: hemi-methylation, full: full-methylation) and cofactor-binding (SAM/SAH). Colored as per local density of AF3 models.

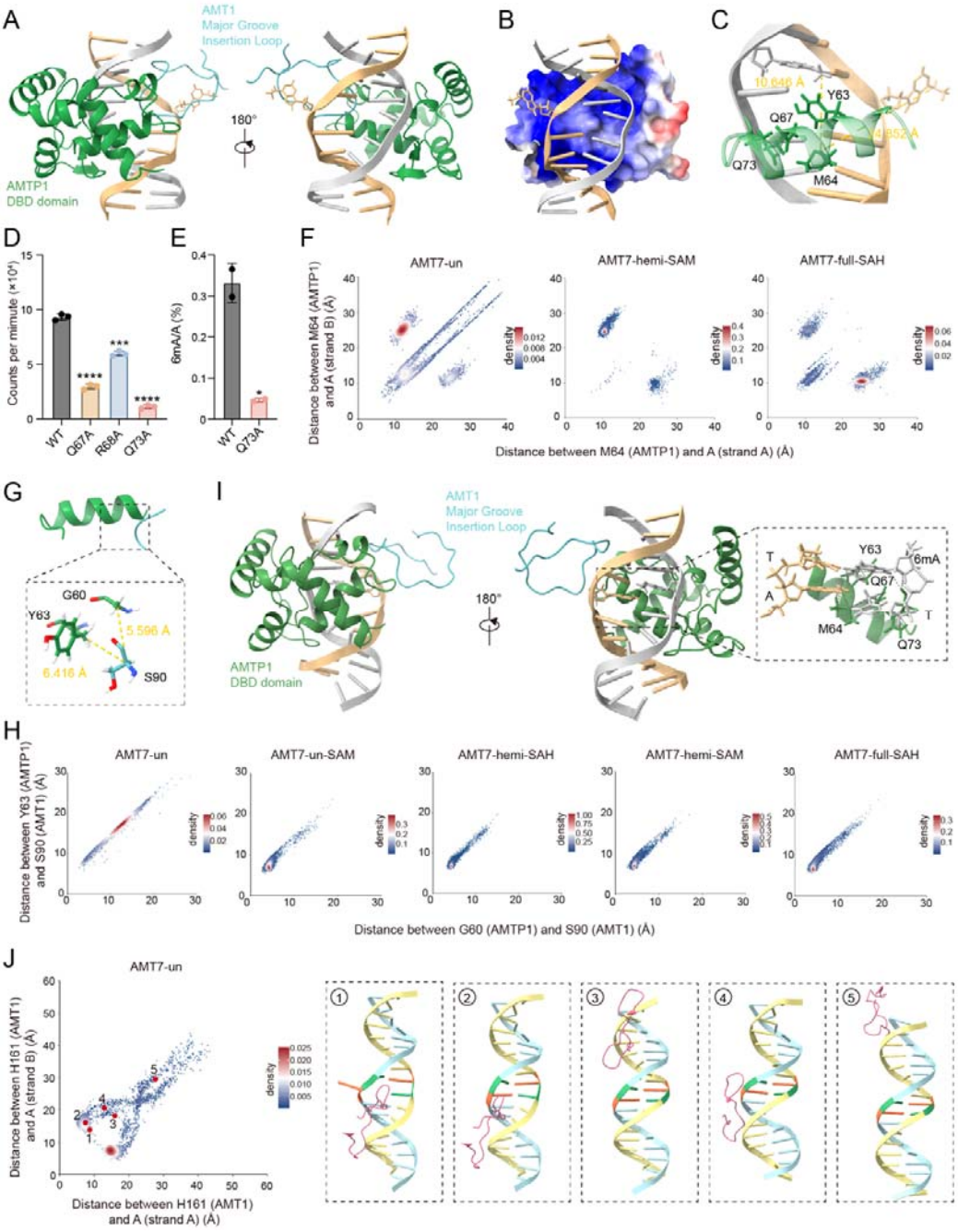

Figure 5. DNA binding by AMTP1.

A. Opposite views of cooperative binding by AMT1 *Major Groove Insertion Loop* and AMTP1 DNA-binding domain (DBD: 24-137). AMT7 ternary complex (AMT7-full-SAH) in the base-flipping mode. Left/right: DNA major groove facing outward/inward. Extrahelical 6mA is shown in the

stick model.

B. Electrostatic surface representation of AMTP1 DBD. Blue: positive. Red: negative. DNA major groove facing inward.

C. Ribbon representation of AMTP1 α3-helix, with stick model representation of residues critical for DNA binding. Distance between M64 (C_α) in AMTP1 and target/non-target 6mA (N^6): 4.9/10.6 Å.

D. In vitro MTase activity of AMT7 complex with indicated point mutations of AMTP1. Wildtype (WT) AMT7 complex was included as positive control. Data are shown as mean±SD. Two-tailed Student's t-test: ***$p < 0.001$, ****$p < 0.0001$.

E. Mass spectrometry quantification of 6mA-to-A ratio (6mA/A, %) in WT and AMTP1 Q73A mutant. Data are shown as mean±SD. Two-tailed Student's t-test: *$p < 0.05$.

F. Conformational analysis tracking DNA binding by AMTP1 DBD in AMT7 binary and ternary complexes (AMT7-un, AMT7-hemi, AMT7-hemi-SAH, AMT7-full-SAH). Scatterplot: distance between M64 (C_α) of AMTP1 and 6mA/A (N^6) on DNA strand A/B (x/y-axis, Å); colored as per local density of AF3 models.

G. Interactions between AMT1 *Major Groove Insertion Loop* and AMTP1 DBD α3-helix. In expanded view, distance between AMT1 S90 (C_α) and AMTP1 G60/Y63 (C_α): 5.6/6.4 Å.

H. Conformational analysis tracking interactions between AMT1 *Major Groove Insertion Loop* and AMTP1 DBD. AMT7 binary and ternary holo-complexes with different states of DNA methylation (un: unmodified, hemi: hemi-methylation, full: full-methylation) and cofactor-binding (SAM/SAH/none). Scatterplot: distance between AMT1 S90 (C_α) and AMTP1 G60/Y63 (C_α) (x/y-axis, Å); colored as per local density of AF3 models.

I. Opposite views of AMTP1-dominated binding of dsDNA. AMT7 binary complex (AMT7-hemi) in the base-pairing mode. Left/right: DNA major groove facing outward/inward. Note that AMT1 *Major Groove Insertion Loop* shifts away from the target site. Expanded view: AMTP1 DBD residues critical for recognizing the ApT duplex. Contacts are represented by grey dashed lines.

J. Conformational analysis tracking DNA binding by AMT1 *Minor Groove Insertion Loop* in an AMT7 binary complex (AMT7-un). Scatterplot: distance between AMT1 H161 (C_α) and 6mA/A (N^6) on DNA strand A/B (x/y-axis, Å); colored as per local density of AF3 models. Important conformations: 1) onsite binding in the base flipping mode, 2) onsite binding in the base pairing mode, 3) offsite binding in the major groove, 4) offsite binding in the minor groove, and 5) faraway offsite binding. Right panels: typical structures; only AMT1 *Minor Groove Insertion Loop* and dsRNA are shown. Strand A: blue. Strand B: yellow. Adenine: orange. Thymine: green. AMT1 *Minor Groove Insertion Loop*: pink. AMT1 H161: stick.

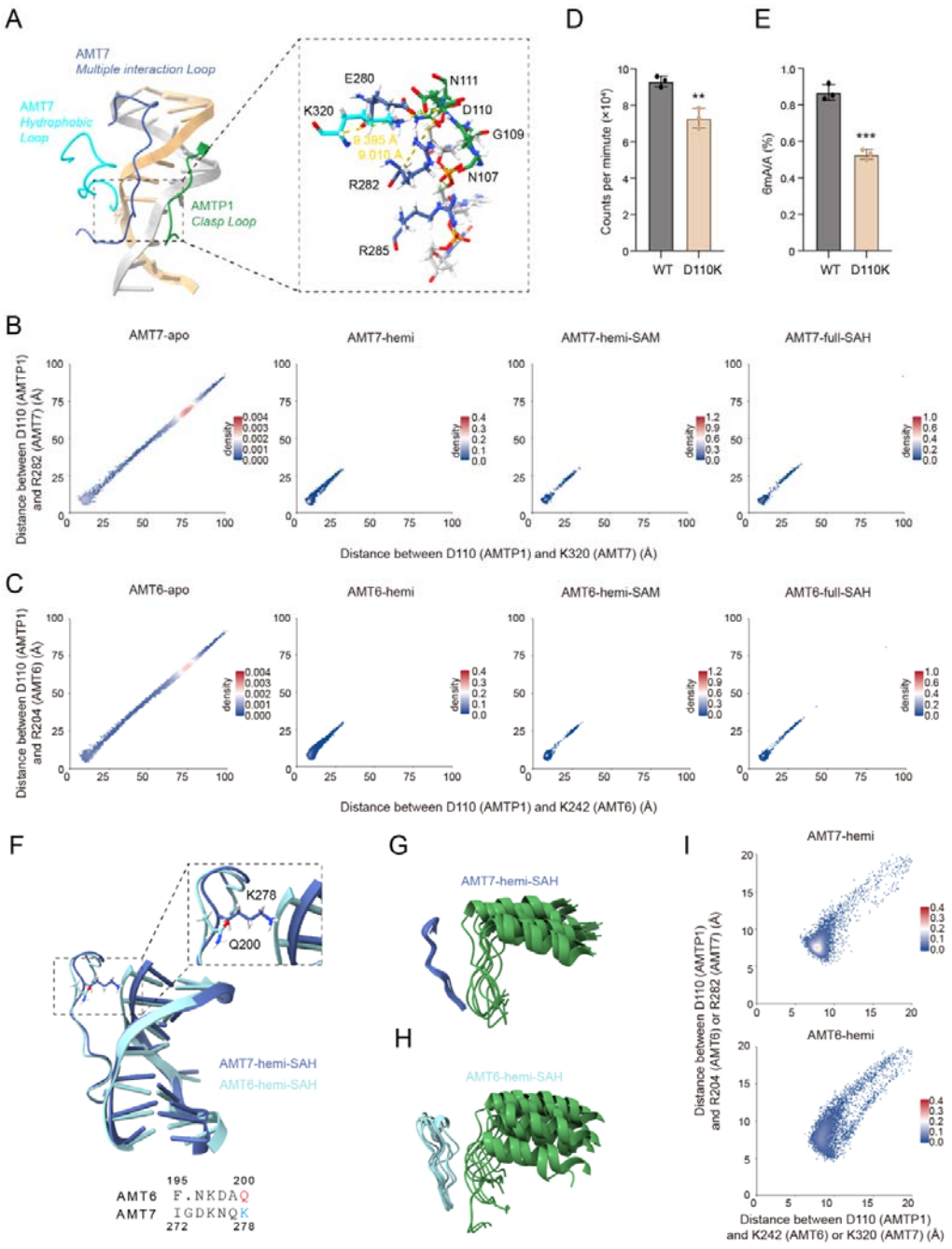

Figure 6. Interactions securing the closed form AMT1 complex.

A. The clasp securing the closed form AMT7 ternary complex (AMT7-full-SAH). Three important loops are highlighted: AMTP1 *Clasp Loop*, AMT7 *Multiple interaction Loop* and *Hydrophobic Loop*. Expanded view: residues critical for clasp formation are shown in the

stick model. Hydrogen bond: green dashed line. Distance between AMTP1 D110 (C_α) and AMT7 K320/R282 (C_α): 9.4/9.0 Å.

B. Conformational analysis tracking clasp formation in AMT7 apo- and holo-complexes (AMT7, AMT7-hemi, AMT7-hemi-SAM, AMT7-full-SAH). Scatterplot: distance between AMTP1 D110 (C_α) and AMT7 K320/R282 (C_α) (x/y-axis, Å); colored as per local density of AF3 models.

C. Conformational analysis tracking clasp formation in AMT6 apo- and holo-complexes (AMT6, AMT6-hemi, AMT6-hemi-SAM, AMT6-full-SAH). Scatterplot: distance between AMTP1 D110 (C_α) and AMT76 K242/R282 (C_α) (x/y-axis, Å); colored as per local density of AF3 models.

D. In vitro MTase activity of AMT7 complex with AMTP1 mutation. WT AMT7 complex was included as positive control. Data are shown as mean±SD. Two-tailed Student's t-test. **$p < 0.01$.

E. Mass spectrometry quantification of 6mA-to-A ratio (6mA/A, %) in WT and AMTP1 D110K mutant. Data are shown as mean±SD. Two-tailed Student's t-test. ***$p < 0.001$.

F. Structural overlay of *Multiple Interaction Loop* of AMT6 or AMT7. AMT6/AMT7 binary complexes (AMT6/AMT7-hemi) in the base pairing mode. Expanded view: AMT6/AMT7 variance region, with stick model representation for AMT7 K278 and AMT6 Q200 are represented with sticks. Bottom: AMT6/AMT7 local sequence alignment.

G. Structural overlay of AMT7 *Multiple Interaction Loop* (275-282) and AMTP1 DBD (107-126) of top 10 scored AF3 models for AMT7 binary complex (AMT7-hemi) in the base pairing mode.

H. Structural overlay of AMT6 *Multiple Interaction Loop* (197-204) and AMTP1 DBD (107-126) of top 10 scored AF3 models for AMT6 binary complex (AMT6-hemi) in the base pairing mode.

I. Expanded views of conformational analysis of clasp formation in AMT6/AMT7 binary complexes (AMT6/AMT7-hemi). Note preferential unraveling of the clasp in AMT6 complex.

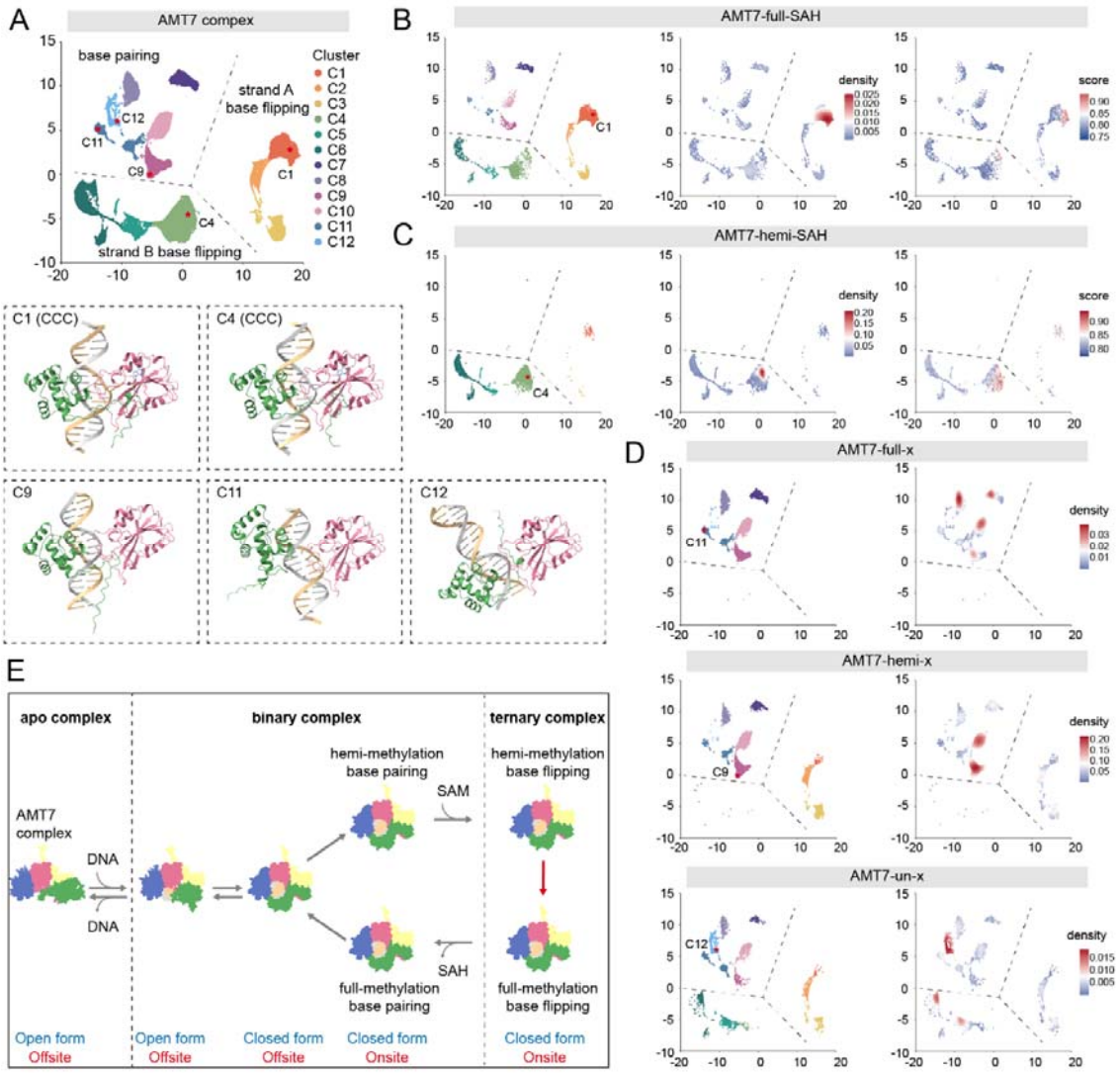

Figure 7. Reaction pathway for DNA adenine methylation catalyzed by AMT1 complex.

A. UMAP of all AMT7 holo-complexes. 12 clusters, featuring distinct structural characteristics, are demarcated and differentially colored. Demarcations separating the base pairing mode and the base flipping mode are shown; the latter is further divided by selection of the target strand for base flipping (A/B). Below: representative structures in cluster C1 (CCC), C4 (CCC), C9 (AMTP1-dominated binding), C11 (AMTP1 offsite binding), and C12 (AMT1 offsite binding); only AMTP1, AMT1 MT-A70 domain, and dsDNA are shown.

B. UMAP of AMT7 ternary complex (AMT7-full-SAH). Colored by cluster (left), density (middle), and AF3 score (right). Representative structure for C1.
C. UMAP of AMT7 ternary complex (AMT7-hemi-SAH). Colored by cluster (left), density (middle), and AF3 score (right). Representative structure for C4.
D. UMAP of all AMT7 binary complexes. AMT7-full (top), AMT7-hemi (middle), and AMT7-un (bottom). Colored by cluster (left) and density (right). Representative structures for C9, C11, and C12.
E. A simplified Reaction pathway for processive DNA adenine methylation catalyzed by AMT7 complex. See text for details.

Supplemental tables

Table S1. AF3 models with their ipTM scores.

Table S2. Percentage of base flipping and base pairing in various AF3 input settings.

Table S3. Key inter-atomic distances used to sample the conformational space of AMT1 complexes.

Supplemental files

File S1. AMT1 sequence alignment.

File S2. AMT6/7 sequence alignment.

File S3. AMTP1 sequence alignment.