Mechanisms and biological roles of DNA methyltransferases and DNA methylation – from past achievements to future challenges

Renata Z. Jurkowska[1], & Albert Jeltsch[2]

1) School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff, CF10 3AX, UK

2) Institute of Biochemistry and Technical Biochemistry, Department of Biochemistry, Allmandring 31, University of Stuttgart, D-70569 Stuttgart, Germany

*Corresponding authors. E-mail: jurkowskar@cardiff.ac.uk, albert.jeltsch@ibtb.uni-stuttgart.de

List of abbreviations

DNA methyltransferase, MTase

Restriction/modification system, RM system

S-adenosyl-L-methionine, AdoMet

Ten-eleven translocation enzyme, TET enzyme

Single-molecule real-time sequencing, SMRT sequencing

Thin layer chromatography, TLC

High pressure liquid chromatography, HPLC

Abstract

DNA methylation and DNA methyltransferases (MTases) – the enzymes that introduce the methylation mark into the DNA– have been studied for almost 70 years. In this chapter, we review the key developments in the DNA methylation field that have led to our current understanding of the structures and mechanisms of DNA MTases. We discuss the essential biological roles of DNA methylation, including the discovery of DNA methylation, cloning and sequence analysis of the bacterial and eukaryotic MTases, and the elucidation of their structure, mechanism, regulation and molecular evolution. We describe genetics studies that contributed greatly to the evolving views on the role of DNA methylation in development and diseases, the invention of methods for the genome-wide analysis of DNA methylation, and the biochemical identification of DNA MTases and the TET enzyme family, which is involved in

DNA demethylation. We summarize the roles of MTases in bacterial epigenetics and the application of MTases in synthetic biology to generate artificial signaling systems. We finish by highlighting some critical questions for the next years of research in the field.

1 Discovery of DNA methylation

DNA from various sources contains the methylated bases C5-methylcytosine, N4-methylcytosine and N6-methyladenine in addition to the four standard nucleobases (Fig. 1A). Methylation of cytosine at the C5-position had been discovered in calf thymus DNA already in 1948 using paper chromatography experiments (Hotchkiss 1948), and 6-methyladenine was found in bacterial DNA in 1955 (Dunn, Smith 1955). N4-methylcytosine, the third and least common methylated base in bacterial DNA, was described for the first time in 1983 (Janulaitis et al. 1983). The methylation of nucleobases at all these positions places the methyl groups in the major groove of the double stranded B-DNA, where they do not interfere with the Watson/Crick base pairing, but can easily be detected by proteins interacting with the DNA (Fig. 1B). DNA methylation can, for example, directly prevent the readout of an AT base pair by glutamine residues in the major groove (Fig. 1C). By this and related processes, DNA methylation can control the binding of proteins to DNA and thereby regulate the expression of the genetic information. Hence, the methylation adds extra information to the DNA that is not encoded in the DNA sequence, and the methylated bases can be considered the 5th, 6th, and 7th letters of the genetic alphabet (Jeltsch 2002). The epigenetic toolbox has been further diversified with the discovery of the oxidized forms of 5-methylcytosine, viz. 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine, in the DNA of many species, including mammals; followed by the discovery of Ten-eleven translocation (TET) enzymes, responsible for generating them (Tahiliani et al. 2009; Kriaucionis, Heintz 2009; Munzel et al. 2010; Pfaffeneder et al. 2011; Ito et al. 2011; He et al. 2011). Despite the interesting properties of the identified methylated bases and their importance in living organisms, many years had passed after the initial discovery of DNA methylation until work with DNA methyltransferases, the enzymes that introduce this modification had been systematically started.

2 Discovery and early work on DNA MTases

DNA methyltransferases were initially discovered as parts of restriction/modification (RM) systems, which consist of a DNA methyltransferase and an associated restriction endonuclease (Arber, Dussoix 1962). S-adenosyl-L-methionine (AdoMet)-dependent DNA and RNA methylation activity was first described by Gold in 1963 (Gold et al. 1963) and a series of papers published by Gold in 1964 (Gold et al. 1964; Gold, Hurwitz 1964b; Hurwitz et al. 1964b, a; Gold, Hurwitz 1964a). The *E. coli* EcoDam (a solitary bacterial MTase that is not part of an RM system) was initially described in 1973 (Marinus, Morris 1973) and purified in 1982 (Herman, Modrich 1982). The first enzymatic studies with human and murine DNA MTases were reported in the late 70s and early 80s (Browne et al. 1977; Gruenbaum et al. 1982).

However, in the 70s and 80s, DNA MTases remained a kind of passengers in the ongoing molecular biology revolution, due to their functional and genetic association to restriction endonucleases, which were absolutely essential as analytical and cloning tools at this time (Arber, Linn 1969; Boyer 1971; Meselson et al. 1972). In addition, restriction endonucleases and DNA MTases constituted the first model systems to study the sequence specific DNA recognition, a process essential to the control of gene expression in all forms of life (Modrich 1982).

With the increasing commercial importance of restriction endonucleases, biotech companies were interested to shift the production procedures away from the purification of enzymes from the original bacterial strains towards recombinant expression of cloned enzymes. Therefore, cloning of restriction enzymes moved into the center of the scientific and economic interest. It was known that RM systems often reside on mobile genetic elements, with the genes coding for the methyltransferase and the endonuclease located next to each other. Hence, cloning of a DNA fragment containing the methyltransferase gene often led to the cloning of the restriction enzyme gene on the same DNA insert. In a procedure called "Hungarian trick", the group of Venetianer realized that the special properties of DNA methyltransferases could be exploited to selectively clone genes encoding these enzymes (Szomolanyi et al. 1980). This approach was based on the fact that after expression of a DNA MTase in cells, the enzyme modified its own encoding DNA. Hence, after shotgun cloning of bacterial genomes, the plasmids containing DNA inserts were isolated and cleaved with a restriction enzyme of interest. The protected DNA likely coded for a methyltransferase, which methylated DNA within the target region of the endonuclease and thereby prevented cleavage. After cloning of these protected inserts, it turned out that very often the gene for the restriction enzyme was found on the same piece of DNA next to the methyltransferase gene. Almost 20 years later, a similar coupling of genotype and phenotype after expression of DNA methyltransferases was applied by Tawfik and colleagues to develop a novel approach for protein engineering, which was based on the expression of libraries of MTase mutants in water/oil emulsions (Tawfik, Griffiths 1998).

3 DNA MTases contain conserved amino acids sequence motifs

The wide application of the above-described and related cloning procedures led to the cloning of hundreds of restriction enzymes together with their corresponding methyltransferases. Therefore, the group of bacterial DNA methyltransferases provided a rich source of enzymes known to recognize different DNA sequences for enzymatic, biochemical and evolutionary studies, which has led to many important insights and breakthrough discoveries (Wilson, Murray 1991; Pingoud, Jeltsch 1997; Pingoud et al. 2014). Comparisons of the amino acid sequences of DNA methyltransferases in the early days of multiple sequence alignments led to the discovery of up to ten amino acid motifs characteristic for cytosine-C5 methyltransferases (Posfai et al. 1989; Klimasauskas et al. 1989; Lauster et al. 1989) (Fig. 2).

In 1988, Bestor cloned the first mammalian DNA methyltransferase that shared extensive sequence similarity with the bacterial cytosine-C5 methyltransferases in its C-terminal catalytic part (Bestor et al. 1988). It was discovered that bacterial adenine-N6 methyltransferases contained conserved amino acid motifs as well (Fig. 2) (Lauster et al. 1987; Guschlbauer 1988) and some of the MTase motifs were shown to be part of general signature motifs of AdoMet dependent methyltransferases, including small molecule, protein and RNA methyltransferases (Kagan, Clarke 1994; Ingrosso et al. 1989). Although statistical methods were insufficient at that time, these studies led to the identification of the key catalytic regions both in adenine-N6 and cytosine-C5 methyltransferases. Many of the most conserved residues in both families of enzymes were shown to be directly involved in the catalytic process (Cheng 1995; Jeltsch 2002) and several amino acid motifs identified in the early alignment studies could later be connected to defined structural elements in the conserved methyltransferase fold (Malone et al. 1995) (Fig. 2).

4 Structure and mechanism of DNA MTases

All DNA methyltransferases use AdoMet as a methyl group donor. Based on their mechanism, one can distinguish methyltransferases adding the methyl group to carbon or nitrogen atoms. The former group comprises cytosine-C5, the latter adenine-N6 and cytosine-N4 methyltransferases. All DNA MTases follow a ternary complex mechanism, where the catalytically competent complex consists of the enzyme, the DNA substrate and the AdoMet cofactor. In some enzymes, binding of the DNA substrate and the AdoMet is an ordered reaction, in other cases it is random. Wu and Santi studied the catalytic mechanism of cytosine-C5 methyltransferases and proposed in 1985 that it follows a Michael addition reaction, which is characterized by the formation of a covalent intermediate between the enzyme and the target base (Wu, Santi 1985, 1987) (Fig. 3A). Shortly afterwards, Santi and coworkers also showed that adenine methylation proceeds directly at the N6 position, despite the poor nucleophilicity of the N6 atom, and not by a transient transfer of the methyl group to the N1 followed by its shift to the N6 (Pogolotti et al. 1988). Seminal insights into the folding of the methyltransferases and the arrangement of their catalytic centers came with the first structure of a DNA methyltransferase (the bacterial M.HhaI enzyme) that was solved in 1993 (Cheng et al. 1993). In 1994, the publication of the first structure of a DNA methyltransferase (again M.HhaI) with its DNA substrate by Cheng and coworkers led to another conceptual breakthrough regarding the catalytic mechanism of DNA methyltransferases (Klimasauskas et al. 1994) (Fig. 4). It was observed that the target base for the methylation reaction was completely rotated out of the DNA helix and inserted into a catalytic pocket of the enzyme in a process called “base flipping”, which is necessary to allow for the close access of the catalytic residues described above to the substrate base. This unexpected and seminal discovery highlighted the flexibility of DNA and the dynamic processes that accompany enzymatic catalysis; research subjects that were intensively studied afterwards. Today, we appreciate base flipping as a universal process in DNA methylation, but also in other reactions occurring on DNA, including DNA repair (Roberts 1995; Roberts, Cheng 1998). Since then, numerous

crystal structures of DNA methyltransferases with bound substrate DNA provided additional confirmation for base flipping among bacterial and eucaryotic enzymes, reinforcing this pivotal discovery.

The first structure of an adenine-N6 MTase (M.TaqI) was published in 1994 as well (but without DNA) (Labahn et al. 1994), unexpectedly showing that both enzyme families contain a large catalytic domain with an identical fold, consisting of a similar six-stranded parallel ß-sheet with a seventh strand inserted in an antiparallel fashion between the fifth and sixth strands (Schluckebier et al. 1995) (Fig. 2). This fold is known today as the AdoMet dependent methyltransferase fold (Martin, McMillan 2002). The seven-stranded ß-sheet is flanked by α-helices creating two sub-domains with a Rossman fold architecture: one containing the binding site for the AdoMet and the other for the flipped base (Cheng 1995; Jeltsch 2002). In addition, all MTases contain a second less-conserved domain involved in DNA recognition. The first structure of an adenine-N6 MTase with DNA was solved in 2001 (Goedecke et al. 2001). It showed that the N6 of the flipped adenine is positioned in a tetrahedral environment of hydrogen bond donors provided by the conserved residues of the (DNS)PP(YFW) motif, suggesting that its nucleophilicity is increased by a change in hybridization from sp2 to sp3 (Fig. 3B). Cytosine-N4 MTases are believed to follow an analogous mechanism, based on the chemical similarity of the methyl-acceptor atom and the observation that the specificity of enzymes form these families overlap, i.e. that adenine-N6 MTases can also methylate cytosine at N4 and Cytosine-N4 MTases also methylate adenine (Jeltsch et al. 1999; Jeltsch 2001).

5 Molecular evolution of MTases

As described above, the conserved structure of the 7-ß strand MTases consists of two half domains with Rossman folds fused to each other. One of them mediates the AdoMet interaction, the second provides the binding sites for the methylation substrates, flipped nucleobases in the case of DNA MTases. The high structural similarity of all DNA MTases and the presence of the conserved motifs with similarities even between different groups of MTases, suggests that these enzymes are monophyletic. Presumably, the two subdomains originated from a duplication of a primordial AdoMet binding Rossman fold domain (Malone et al. 1995). Later, one subdomain continued to bind AdoMet, whereas the second diverged to generate the binding pockets for different methylation substrates including flipped cytosine and adenine bases, but also small molecules like catecholamine or amino acids like arginine, leading to the appearance of various groups of contemporary MTases specific for different methylation substrates. Moreover, the initial MTase ancestor has undergone several modifications during molecular evolution, including circular permutations (Jeltsch 1999; Bujnicki 2002), and in the case of DNA MTases, the insertion of diverse and unrelated DNA binding domain at different places in the consensus structure, leading to the creation of different classes of DNA MTases (Malone et al. 1995) (Fig. 2). In Eukaryotes, the catalytic domain of C5 MTases has taken part in diverse domain shuffling events and became fused with many other chromatin interacting domains found in animals and plants, including PWWP,

ADD, BAH, and Chromodomains. The acquisition of the various chromatin and protein/nuclei acid interacting modules by eukaryotic MTases allowed functional coupling and crosstalk between different layers of epigenetic information, including histone posttranslational modifications, chromatin remodeling and non-coding RNAs.

Unexpectedly, it turned out that the presence of 5-methylcytosine puts an evolutionary burden on organisms that affected the evolution of C5 MTases and shaped cellular DNA methylomes (Jeltsch 2002). One contributing factor is the accelerated deamination of 5-methylcytosine (2-4 times) compared to cytosine. The mutational threat of this reaction is further increased by the lower repair efficiency of T/G mismatches (arising from deamination of 5-methylcytosine), as compared to U/G mismatches (arising from deamination of unmethylated cytosine). Therefore, specific repair systems have been developed to handle TG mismatches, appearing in the sequence context of previously methylated cytosine bases. For example, in mammals, specific repair enzymes exist that act on T/G mismatches in a TpG context, including the 5-methylcytosine binding protein MBD4 (Hendrich et al. 1999). Similarly, in *E. coli* the VSR mismatch endonuclease triggers the repair at T/G mismatches originating from the deamination of methylated C<u>C</u>WGG sites, the target of the endogenous *E. coli* Dcm DNA-(cytosine C5)-MTase (Hennecke et al. 1991). Despite the existence of these and other specialised repair systems, methylation-mediated mutagenesis had a strong influence on the genomes evolution of vertebrates, where generally a strong depletion of CpG sites in the DNA sequences is observed (Cross et al. 1994). Recently, it has been shown that C5 MTases also generate 3-methylcytosine in a side reaction by methylating N3 atom of the cytosine base (Rosic et al. 2018). This modified base is toxic for the cell, because it interferes with RNA synthesis and DNA replication. It can be removed by members of the ALKB2 family of DNA alkylation repair enzymes, in an iron(II) and 2-oxoglutarate-dependent oxidation process. In fact, ALKB2 enzymes are apparently required in all species which express active C5-MTases, because an evolutionary analysis showed that these two enzyme families are very tightly connected, reflecting their essential functional link (Rosic et al. 2018). The disadvantages of the 5-methylcytosine methylation system may explain why C5-MTases were repeatedly lost and gained in the evolution of eukaryotes and why they are lacking in several species (Jeltsch 2010; de Mendoza et al. 2021). This effect may also explain why adenine methylation is far more abundant in the bacterial world than cytosine-C5 methylation.

6 Early views on the biological role of DNA methylation

Methylation of human and mammalian DNA at CpG sites was identified in the beginning of the 80s (Razin, Riggs 1980; Ehrlich, Wang 1981), in plants DNA methylation was found also in CNG sites (where N is any nucleotide) (Gruenbaum et al. 1981) and non-symmetric sites. However, the early 80s was a time when biology mainly focused on the detailed investigation of the so called “model organisms”. While this approach was extremely far-sighted and greatly contributed to the explosion in our understanding of the molecular basis of life, it did not come without risk, as illustrated by the general lack of appreciation for DNA methylation

around that time. Due to an unfortunate coincidence, many of the carefully selected model organisms like *S. cerevisiae*, *D. melanogaster* or *C. elegans* were lacking detectable DNA methylation. Consistently, there was a widespread belief that DNA methylation, although interesting, cannot be very important. On the other hand, it became clear that DNA methylation had an enormous influence on the human genome, when Bird discovered the existence of the CpG islands (Bird 1980; Bird et al. 1985), which are defined as regions of high density of CpG sites within the genome that was already known to be globally depleted of this dinucleotide (Swartz et al. 1962). It was realized that the depletion of the CpG sites from the bulk genome was indirectly due to the mutagenic effect of cytosine-C5 methylation (as described above), leading to the preservation of CpG sites only if they were unmethylated, as in CpG islands. Today, we know that DNA methylation systems are found in almost all organisms and the model organisms devoid of them listed above appear to be rather exceptions (Fig. 5).

7 Genetic studies on DNMTs in mammals

While models connecting DNA methylation with known epigenetic phenomena, gene expression and development (see for example (Riggs 1975; Holliday, Pugh 1975)) were developed, the general skepticism on the essential role of DNA methylation in human biology was only overcome with the discovery of the repressive function of DNA methylation on gene expression (Tazi, Bird 1990) and the finding that mice with a knock-out of DNMT1 (Li et al. 1992), the only mammalian DNMT known by that time, die during early embryonic development in the uterus. However, as often in science, this discovery led to the next question, because it turned out that DNMT1 knock-out cells were not completely devoid of DNA methylation (Lei et al. 1996), which opened a hunt for additional mammalian DNMTs. Researchers tried to purify additional DNA methyltransferases from human and mouse cells; in parallel, the rising flood of DNA sequences was searched for entries containing the characteristic DNA methyltransferase motifs described above. It was the bioinformatics approach that was successful at the end, leading to the discovery of the DNMT3A and DNMT3B enzymes in 1998 (Okano et al. 1998). Shortly afterwards, both MTases were also shown to be essential in mice (Okano et al. 1999). Soon after, genetic studies showed that DNMT3A together with DNMT3L (a catalytically inactive paralog of DNMT3A and DNMT3B) were needed to set impinging marks in the mouse germline (Bourc'his et al. 2001; Bourc'his, Bestor 2004; Hata et al. 2002; Kaneda et al. 2004). Recently, DNMT3C, a rodent specific paralog of DNMT3B essential for spermatogenesis and male fertility in mice , has been discovered, further diversifying the repertoire of mammalian MTases (Barau et al. 2016; Jain et al. 2017). It appears to have arisen from an ancient duplication of the DNMT3B gene that occurred ~46 million years ago in the in the Muroidea superfamily during rodent evolution and had been wrongly annotated as a nonfunctional pseudogene. (Barau et al. 2016; Jain et al. 2017).

DNA methylation provides organisms with an efficient epigenetic regulatory system, which is particularly important in multicellular organisms, because of their needs to develop stable

cellular differentiation states. It has been proposed that the development of powerful epigenetic systems, comprising DNA methyltransferases, demethylases and other enzyme systems introducing modifications on histones, had been a critical step in the evolution of multicellular life (Jeltsch 2013). Today, DNA methylation is recognized as an essential epigenetic mark that acts in concert with other chromatin modifications, like histone post-translational modifications, histone variants, or non-coding RNA. In mammals, DNA methylation is involved in the epigenetic processes, like imprinting and X-chromosome inactivation, but it also has global roles in the generation of heterochromatin, silencing of repeats and gene regulation during development and disease (Jurkowska et al. 2011).

8 Structure, function and regulation of mammalian DNA MTases

While time was progressing, structures of DNMT3A (Jia et al. 2007; Guo et al. 2015; Zhang et al. 2018), DNMT1 (Takeshita et al. 2011; Song et al. 2011; Song et al. 2012; Adam et al. 2020) and lately also DNMT3B (Gao et al. 2020; Lin et al. 2020) were published showing that complicated regulatory processes, including oligomerization, distinct chromatin contacts, conformational changes and auto-inhibition, all interplay to accurately control the activity of these enzymes. In 1997, targeting of DNMT1 to replication foci via its interaction with PCNA was discovered (Chuang et al. 1997), but later it became clear that the interaction of DNMT1 with UHRF1 is even more essential for the targeting and activity of DNMT1 (Bostick et al. 2007; Sharif et al. 2007). Furthermore, it was found that in addition to the indirect targeting by other complex partners, DNMTs directly interact with chromatin. DNMT3A, DNMT3B and DNMT3L use their ADD domains for binding to H3 tails unmethylated at K4 (Ooi et al. 2007; Zhang et al. 2010), DNMT3A and DNMT3B use their PWWP domains for binding to H3K36me2/3 (Dhayalan et al. 2010), and DNMT3A1 (an isoform of DNMT3A) binds with its UDR domain H2AK119ub1 (Weinberg et al. 2021). DNMT1 interacts with its replication foci targeting domain (RFTD) with ubiquitinated H3 tails (Nishiyama et al. 2013) and H3K9me3 (Ren et al. 2020) and via its BAH domain it binds H3K40me3 (Ren et al. 2021). Moreover, the principles of the regulation of the activity and stability of DNMTs via post-translational modifications begin to emerge (Esteve et al. 2011; Deplus et al. 2014), adding another fascinating topic to the study of these enzymes.

9 Discovery of TET enzymes

A similar changeful journey as in the field of DNA methylation was undertaken in the investigation of DNA demethylation, starting from the question of whether an active process of DNA demethylation might exist at all, leading to its discovery and the study of its mechanisms (Ooi, Bestor 2008). It was only in 2009, when the combination of powerful biochemical and bioinformatics approaches led to the discovery of the Ten-eleven Translocation (TET) enzymes (Tahiliani et al. 2009), which oxidize 5-methylcytosine to the hydroxymethyl, formyl or carboxyl state, and the discovery of these modified bases in human

DNA (Tahiliani et al. 2009; Kriaucionis, Heintz 2009; Munzel et al. 2010; Pfaffeneder et al. 2011; Ito et al. 2011; He et al. 2011). The exact role of these additional modified bases and the complete pathway of DNA demethylation is not yet fully understood (Wu, Zhang 2017), but it has been well-established that the TDG glycosylase can remove 5-formylcytosine and 5-carboxylcytosine, eventually leading to the generation of an unmethylated cytosine (He et al. 2011; Maiti, Drohat 2011). Moreover, it is becoming increasingly clear that the TET-derived oxidative cytosine bases constitute distinct epigenetic marks by themselves that can be read by specialized reader domains and confer biological functions (Song, Pfeifer 2016; Song et al. 2021). Currently, it is understood that the genome-wide and locus-specific DNA methylation level is determined by a steady-state reached through the combined action of MTases, demethylases and DNA replication (Jeltsch, Jurkowska 2014).

10 Methods for site-specific detection of DNA methylation

The detection of DNA methylation for a long time was based on the initial methods: TLC (followed initially by HPLC, and today by mass spectrometry), allowing for a quantitative overall genome methylation analysis, but without sequence resolution; and restriction digestion using enzymes sensitive to DNA methylation, allowing site-specific genome-wide analysis, but only at defined restriction sites. For cytosine-C5 methylation, this situation dramatically changed with the development of the bisulfite conversion method, which can be combined with a battery of downstream technologies to enable a genome-wide analysis of 5-methylcytosine at single nucleotide resolution (Frommer et al. 1992; Clark et al. 1994). This technology in concert with the breakthroughs in DNA sequencing technologies has enabled researchers starting in 2008 to provide first genome-wide DNA methylation maps of plant and mouse cells (Cokus et al. 2008; Lister et al. 2008; Meissner et al. 2008). Powerful variants of bisulfite sequencing, including oxidative bisulfite sequencing, have been developed, allowing detection of not only 5-methylcytosine, but also its oxidized forms at single base resolution (Booth et al. 2013). Nanopore sequencing is another emerging technique that has been applied for the direct detection of 5-methylcytosine in long DNA fragments (Laszlo et al. 2013; Lee et al. 2020; Sakamoto et al. 2020).

For N-methylation, it was only in 2010, almost 20 years after the discovery of the bisulfite technology, when the development of single-molecule real-time (SMRT) sequencing for the first time provided a method for the genome-wide analysis of 6-methyladenine at single nucleotide resolution (Flusberg et al. 2010). This invention was followed by a flurry of bacterial N6-adenine methylomes, including those of *E. coli* and *C. crescentus* (Sanchez-Romero et al. 2015), which provided novel insights into the role of DNA methylation in bacteria in defense mechanisms, cell division, gene expression, and DNA repair. Recently, nanopore sequencing has been applied for the direct detection of 6-methyladenine in bacterial DNA as well (McIntyre et al. 2019). Moreover, nitrite sequencing has been described for the base specific readout of 6-methyladenine by DNA sequencing (Mahdavi-Amiri et al. 2021). Conceptually similar to bisulfite sequencing used for detection of cytosine methylation, nitrite sequencing

is based on the selective conversion of adenine to hypoxanthine, which is blocked by adenine-N6 methylation.

11 DNA MTases and bacterial epigenetics

In bacteria, DNA methylation is involved in the control of DNA replication and repair, host defense by restriction/modification (RM) system and control of gene expression (Jeltsch 2002; Wion, Casadesus 2006; Casadesus, Low 2006; Sanchez-Romero, Casadesus 2020). The most known bacterial DNA MTases introduce N6-methyladenine and belong to the RM systems, which serve as defense systems to protect prokaryotes from bacteriophage infections (Matic et al. 1996). RM systems comprise a restriction endonuclease and a DNA MTase, with both enzymes recognizing the same target DNA sequence. In the most common type II RM systems, the DNA target sites are 4 to 8 base pair long palindromic sequences. The restriction endonuclease cleaves the DNA at the target sequence only in an unmethylated state, as found on an invading phage DNA during the early steps of infection. The bacterial host cell DNA is kept in a methylated state by the corresponding DNA MTase and thereby it is protected from cleavage (Pingoud, Jeltsch 2001). These systems are complemented by the adaptable CRISPR-Cas systems that have been discovered only recently as an additional defense system of bacteria and archaea against bacteriophages, which is independent of DNA methylation (Bhaya et al. 2011).

Aside from MTases being parts of RM systems, bacteria contain so called solitary or orphaned MTases that are not accompanied by a restriction enzyme. The *Escherichia coli* deoxyadenosine DNA methyltransferase (Dam) and the *Caulobacter crescentus* cell-cycle-regulated methyltransferase (CcrM) are two well-characterized examples of this type (Jeltsch 2002; Wion, Casadesus 2006; Casadesus, Low 2006; Sanchez-Romero, Casadesus 2020). *E. coli* Dam is involved in DNA mismatch repair, initiation of chromosome replication and regulation of gene expression, including the pap phase variation in uropathogenic *E. coli* (Marinus, Morris 1973; Lobner-Olesen et al. 2005; Low, Casadesus 2008; Marinus, Casadesus 2009; van der Woude 2011), indicating that DNA methylation patterns are involved in bacterial gene regulation. In all these processes, the main underlying principle is that DNA methylation alters protein-DNA interactions, because binding of repressor proteins to certain operator sites is impaired by DNA methylation. Although MTases are active in these bacteria, certain target sites can be protected against methylation by bound repressors, generating a methylation pattern. Under specific conditions, these patterns can be heritable and transmit epigenetic information used for example to regulate switching of cell states (Sanchez-Romero, Casadesus 2020). Of note, similar principles can also operate in eukaryotes, including mammals, as many transcriptional regulators show sensitivity to DNA methylation of their binding sites (Yin et al. 2017).

12 Role of DNA methylation in cancer

In 1983, the first groups reported global hypomethylation of DNA in cancer cells (Feinberg, Vogelstein 1983; Gama-Sosa et al. 1983). Shortly afterwards, first examples of local hypermethylation at gene promoters of tumor suppressors, leading to their inactivation in cancer cells, were discovered (Baylin et al. 1986; Greger et al. 1989). It is now well-established that these two processes, global DNA hypomethylation and regional hypermethylation, occur in most tumor cells and are directly connected to the progression of the disease (Baylin 2012; Bergman, Cedar 2013). In 2010, it was discovered that somatic mutations in DNMT3A are prevalently observed in Acute myeloid leukemia (AML) patients, among them the R882H exchange was found with particularly high frequency (Yamashita et al. 2010). Later work has confirmed and extended this finding, and showed that mutations in DNMT3A are drivers of the disease process (Hamidi et al. 2015).

The frequent observation of the inactivation of tumor suppressor genes in cancers by hypermethylation has prompted the development of DNA methyltransferase inhibitors for clinical applications. This field was pioneered by Jones with the development of 5-azacytidine (Jones, Taylor 1980), which afterwards was confirmed to form an irreversible covalent complex with DNA methyltransferases (Santi et al. 1984). Later, cofactor analogs were also introduced to inhibit DNA methyltransferases (Reich, Mashhoon 1990). Today, many derivatives of these initial compounds have been developed and several are in clinical use for treatment of cancer and other diseases (Yang et al. 2010; Fahy et al. 2012). Further development led to the idea of combination treatments, where DNMT inhibitors could be combined with other antitumor drugs, leading to higher efficacy (Ren et al. 2021).

13 Application of MTases in artificial epigenetic systems

The field of synthetic biology emerged almost 20 years ago, based on several key developments in molecular biology techniques. Two reports in 2000 showing the construction and engineering of synthetic genetic circuits, a “toggle switch” and a “repressilator”, are now often considered hallmarks of the design of synthetic circuits and the field of synthetic biology in general (Gardner et al. 2000; Elowitz, Leibler 2000). Afterwards, engineered circuits became more diverse and versatile, with an increasing number of control elements integrated. In an artificial DNA methylation-based epigenetic memory system in *E.coli,* the CcrM DNA methyltransferase was combined with a designed zinc finger (ZnF), which only binds to unmethylated target sites (Maier et al. 2017). Thereby, it regulates an operon expressing a the CcrM MTase, which leads to positive feedback and stable switching of the system. With this system, an initial trigger, like the presence of arabinose, tetracyclin, or DNA damage is memorized in life *E. coli* cells (Maier et al. 2017; Ullrich et al. 2020). A switchable system based on *Salmonella enterica* and dam methyltransferase has been developed as well (Olivenza et al. 2019). Moreover, Park et al. described the development of a full artificial epigenetic system, based on adenine-N6 methylation in a human cells line (Park et al. 2019). It employs the dam MTase as a writer and the DpnI binding domain as reader of the GATC methylation. Notably,

it was shown to regulate gene expression in a DNA methylation-dependent manner and conferred the inheritance of the methylation states.
In 1997, Xu & Bestor developed the targeted methylation approach, a method in which a DNA MTase is fused to a DNA binding domain that targets the fusion protein to specific genomic loci and results in the introduction of DNA methylation at these sites (Xu, Bestor 1997). Later different programable DNA targeting domains, relying on C2H2 zinc fingers, the TAL effector arrays (TALE) or catalytically inactive CRISPR-Cas9 nuclease were fused to an epigenetic writer or reader domains have been employed to synthetically reprogram epigenetic and transcriptional changes in mammalian systems (Kungulovski, Jeltsch 2016; Jurkowska, Jurkowski 2019; Rots, Jeltsch 2018).

13 Conclusions and outlook

Although DNA MTases, the enzymes that introduce methylation into DNA have been intensively studied, the interest in these enzymes has remained high over many years (Fig. 6). This is due to an ever-growing importance of DNA methylation as an epigenetic modification in organismic development and human diseases. Despite decades of active research in the fields of DNA methylation and DNA methyltransferases, and progress in mapping the methylation landscapes at high resolution, many exciting questions still await answers and future challenges extend from our current level of knowledge. How is DNA methylation (and epigenetic information in general) deposited during organismic development and how is it maintained and altered if needed? How are DNA MTases regulated and targeted to achieve these goals? How does DNA methylation interact with other epigenetic systems in mammals, lower eukaryotes and even bacteria? Will it be soon possible to develop epigenetic antibacterial drugs, addressing processes like phase variation or drug resistance? How can we make use of epigenetic editing, including targeted DNA methylation, to repair aberrant, disease-causing epigenetic states and combat diseases like cancer? Which biological function(s) have the oxidized forms of 5-methylcytosine? Is there an active demethylation of N6-methyladenine in the cell? What is the role of this modification in other higher organisms? How can we use DNA methyltransferases in artificial epigenetic circuits more efficiently? We anticipate many more years of exciting research to come in the field of DNA methylation and the study of DNA MTases is an integral objective in this development.

References

Adam S, Anteneh H, Hornisch M, Wagner V, Lu J, Radde NE et al. DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. Nat Commun. 2020;11(1):3723. doi:10.1038/s41467-020-17531-8.

Arber W, Dussoix D. Host specificity of DNA produced by Escherichia coli. I. Host controlled modification of bacteriophage lambda. Journal of molecular biology. 1962;5:18-36.

Arber W, Linn S. DNA modification and restriction. Annual review of biochemistry. 1969;38:467-500. doi:10.1146/annurev.bi.38.070169.002343.

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y et al. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science. 2016;354(6314):909-12. doi:10.1126/science.aah5143.

Baylin SB. The cancer epigenome: its origins, contributions to tumorigenesis, and translational implications. Proceedings of the American Thoracic Society. 2012;9(2):64-5. doi:10.1513/pats.201201-001MS.

Baylin SB, Hoppener JW, de Bustros A, Steenbergh PH, Lips CJ, Nelkin BD. DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. Cancer research. 1986;46(6):2917-22.

Bergman Y, Cedar H. DNA methylation dynamics in health and disease. Nature structural & molecular biology. 2013;20(3):274-81. doi:10.1038/nsmb.2518.

Bestor T, Laudano A, Mattaliano R, Ingram V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. Journal of molecular biology. 1988;203(4):971-83.

Bhaya D, Davison M, Barrangou R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annual review of genetics. 2011;45:273-97. doi:10.1146/annurev-genet-110410-132430.

Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell. 1985;40(1):91-9.

Bird AP. DNA methylation and the frequency of CpG in animal DNA. Nucleic acids research. 1980;8(7):1499-504.

Booth MJ, Ost TW, Beraldi D, Bell NM, Branco MR, Reik W et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protoc. 2013;8(10):1841-51. doi:10.1038/nprot.2013.115.

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007;317(5845):1760-4. doi:10.1126/science.1147939.

Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004;431(7004):96-9. doi:10.1038/nature02886.

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. Science. 2001;294(5551):2536-9. doi:10.1126/science.1065848.

Boyer HW. DNA restriction and modification mechanisms in bacteria. Annual review of microbiology. 1971;25:153-76. doi:10.1146/annurev.mi.25.100171.001101.

Browne MJ, Turnbull JF, McKay EL, Adams RL, Burdon RH. The sequence specificity of a mammalian DNA methylase. Nucleic acids research. 1977;4(4):1039-45.

Bujnicki JM. Sequence permutations in the molecular evolution of DNA methyltransferases. BMC evolutionary biology. 2002;2:3.

Casadesus J, Low D. Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev. 2006;70(3):830-56. doi:10.1128/MMBR.00016-06.

Cheng X. Structure and function of DNA methyltransferases. Annual review of biophysics and biomolecular structure. 1995;24:293-318. doi:10.1146/annurev.bb.24.060195.001453.

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ. Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell. 1993;74(2):299-307.

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science. 1997;277(5334):1996-2000.

Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic acids research. 1994;22(15):2990-7.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008;452(7184):215-9. doi:10.1038/nature06745.

Cross SH, Charlton JA, Nan X, Bird AP. Purification of CpG islands using a methylated DNA binding column. Nature genetics. 1994;6(3):236-44. doi:10.1038/ng0394-236.

de Mendoza A, Poppe D, Buckberry S, Pflueger J, Albertin CB, Daish T et al. The emergence of the brain non-CpG methylation system in vertebrates. Nat Ecol Evol. 2021;5(3):369-78. doi:10.1038/s41559-020-01371-2.

Deplus R, Blanchon L, Rajavelu A, Boukaba A, Defrance M, Luciani J et al. Regulation of DNA methylation patterns by CK2-mediated phosphorylation of Dnmt3a. Cell reports. 2014;8(3):743-53. doi:10.1016/j.celrep.2014.06.048.

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S et al. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. The Journal of biological chemistry. 2010;285(34):26114-20. doi:10.1074/jbc.M109.089433.

Dunn DB, Smith JD. Occurrence of a new base in the deoxyribonucleic acid of a strain of Bacterium coli. Nature. 1955;175(4451):336-7.

Ehrlich M, Wang RY. 5-Methylcytosine in eukaryotic DNA. Science. 1981;212(4501):1350-7.

Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature. 2000;403(6767):335–8. doi:10.1038/35002125.

Esteve PO, Chang Y, Samaranayake M, Upadhyay AK, Horton JR, Feehery GR et al. A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. Nature structural & molecular biology. 2011;18(1):42-8. doi:10.1038/nsmb.1939.

Fahy J, Jeltsch A, Arimondo PB. DNA methyltransferase inhibitors in cancer: a chemical and therapeutic patent overview and selected clinical studies. Expert opinion on therapeutic patents. 2012;22(12):1427-42. doi:10.1517/13543776.2012.729579.

Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301(5895):89-92.

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nature methods. 2010;7(6):461-5. doi:10.1038/nmeth.1459.

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(5):1827-31.

Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW et al. The 5-methylcytosine content of DNA from human tumors. Nucleic acids research. 1983;11(19):6883-94.

Gao L, Emperle M, Guo Y, Grimm SA, Ren W, Adam S et al. Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. Nat Commun. 2020;11(1):3355. doi:10.1038/s41467-020-17109-4.

Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in Escherichia coli. Nature. 2000;403(6767):339-42. doi:10.1038/35002131.

Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E. Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. Nature structural biology. 2001;8(2):121-5. doi:10.1038/84104.

Gold M, Hausmann R, Maitra U, Hurwitz J. The Enzymatic Methylation of Rna and DNA. 8. Effects of Bacteriophage Infection on the Activity of the Methylating Enzymes. Proceedings of the National Academy of Sciences of the United States of America. 1964;52:292-7.

Gold M, Hurwitz J. The Enzymatic Methylation of Ribonucleic Acid and Deoxyribonucleic Acid. V. Purification and Properties of the Deoxyribonucleic Acid-Methylating Activity of Escherichia Coli. The Journal of biological chemistry. 1964a;239:3858-65.

Gold M, Hurwitz J. The Enzymatic Methylation of Ribonucleic Acid and Deoxyribonucleic Acid. Vi. Further Studies on the Properties of the Deoxyribonucleic Acid Methylation Reaction. The Journal of biological chemistry. 1964b;239:3866-74.

Gold M, Hurwitz J, Anders M. The Enzymatic Methylation of Rna and DNA, Ii. On the Species Specificity of the Methylation Enzymes. Proceedings of the National Academy of Sciences of the United States of America. 1963;50(1):164-9.

Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Human genetics. 1989;83(2):155-8.

Gruenbaum Y, Cedar H, Razin A. Substrate and sequence specificity of a eukaryotic DNA methylase. Nature. 1982;295(5850):620-2.

Gruenbaum Y, Naveh-Many T, Cedar H, Razin A. Sequence specificity of methylation in higher plant DNA. Nature. 1981;292(5826):860-2.

Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X et al. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature. 2015;517(7536):640-4. doi:10.1038/nature13899.

Guschlbauer W. The DNA and S-adenosylmethionine-binding regions of EcoDam and related methyltransferases. Gene. 1988;74(1):211-4.

Hamidi T, Singh AK, Chen T. Genetic alterations of DNA methylation machinery in human diseases. Epigenomics. 2015;7(2):247-65. doi:10.2217/epi.14.80.

Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development. 2002;129(8):1983-93.

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333(6047):1303-7. doi:10.1126/science.1210944.

Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature. 1999;401(6750):301-4. doi:10.1038/45843.

Hennecke F, Kolmar H, Brundl K, Fritz HJ. The vsr gene product of E. coli K-12 is a strand- and sequence-specific DNA mismatch endonuclease. Nature. 1991;353(6346):776-8. doi:10.1038/353776a0.

Herman GE, Modrich P. Escherichia coli dam methylase. Physical and catalytic properties of the homogeneous enzyme. The Journal of biological chemistry. 1982;257(5):2605-12.

Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975;187(4173):226-32.

Horton JR, Liebert K, Bekes M, Jeltsch A, Cheng X. Structure and substrate recognition of the Escherichia coli DNA adenine methyltransferase. Journal of molecular biology. 2006;358(2):559-70. doi:10.1016/j.jmb.2006.02.028.

Hotchkiss RD. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. The Journal of biological chemistry. 1948;175(1):315-32.

Hurwitz J, Gold M, Anders M. The Enzymatic Methylation of Ribonucleic Acid and Deoxyribonucleic Acid. 3. Purification of Soluble Ribonucleic Acid-Methylating Enzymes. The Journal of biological chemistry. 1964a;239:3462-73.

Hurwitz J, Gold M, Anders M. The Enzymatic Methylation of Ribonucleic Acid and Deoxyribonucleic Acid. Iv. The Properties of the Soluble Ribonucleic Acid-Methylating Enzymes. The Journal of biological chemistry. 1964b;239:3474-82.

Ingrosso D, Fowler AV, Bleibaum J, Clarke S. Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes. Common sequence motifs for protein, DNA, RNA, and small molecule S-adenosylmethionine-dependent methyltransferases. The Journal of biological chemistry. 1989;264(33):20131-9.

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300-3. doi:10.1126/science.1210597.

Jain D, Meydan C, Lange J, Claeys Bouuaert C, Lailler N, Mason CE et al. rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. PLoS genetics. 2017;13(8):e1006964. doi:10.1371/journal.pgen.1006964.

Janulaitis A, Klimasauskas S, Petrusyte M, Butkus V. Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. FEBS letters. 1983;161(1):131-4.

Jeltsch A. Circular permutations in the molecular evolution of DNA methyltransferases. Journal of molecular evolution. 1999;49(1):161-4.

Jeltsch A. The cytosine N4-methyltransferase M.PvuII also modifies adenine residues. Biological chemistry. 2001;382(4):707-10. doi:10.1515/BC.2001.084.

Jeltsch A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem : a European journal of chemical biology. 2002;3(4):274-93.

Jeltsch A. Molecular biology. Phylogeny of methylomes. Science. 2010;328(5980):837-8. doi:10.1126/science.1190738.

Jeltsch A. Oxygen, epigenetic signaling, and the evolution of early life. Trends in biochemical sciences. 2013;38(4):172-6. doi:10.1016/j.tibs.2013.02.001.

Jeltsch A, Christ F, Fatemi M, Roth M. On the substrate specificity of DNA methyltransferases. adenine-N6 DNA methyltransferases also modify cytosine residues at position N4. The Journal of biological chemistry. 1999;274(28):19538-44.

Jeltsch A, Jurkowska RZ. New concepts in DNA methylation. Trends in biochemical sciences. 2014;39(7):310-8. doi:10.1016/j.tibs.2014.05.002.

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature. 2007;449(7159):248-51. doi:10.1038/nature06146.

Jones PA, Taylor SM. Cellular differentiation, cytidine analogs and DNA methylation. Cell. 1980;20(1):85-93.

Jurkowska RZ, Jurkowski TP. Establishment, Erasure and Synthetic Reprogramming of DNA Methylation in Mammalian Cells. In: Jurga S, Barciszewski J, editors. The DNA, RNA, and Histone Methylomes. Cham: Springer International Publishing; 2019. p. 1-26.

Jurkowska RZ, Jurkowski TP, Jeltsch A. Structure and function of mammalian DNA methyltransferases. Chembiochem : a European journal of chemical biology. 2011;12(2):206-22. doi:10.1002/cbic.201000195.

Kagan RM, Clarke S. Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. Archives of biochemistry and biophysics. 1994;310(2):417-27. doi:10.1006/abbi.1994.1187.

Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E et al. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature. 2004;429(6994):900-3. doi:10.1038/nature02633.

Klimasauskas S, Kumar S, Roberts RJ, Cheng X. HhaI methyltransferase flips its target base out of the DNA helix. Cell. 1994;76(2):357-69.

Klimasauskas S, Timinskas A, Menkevicius S, Butkiene D, Butkus V, Janulaitis A. Sequence motifs characteristic of DNA[cytosine-N4]methyltransferases: similarity to adenine and cytosine-C5 DNA-methylases. Nucleic acids research. 1989;17(23):9823-32.

Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929-30. doi:10.1126/science.1169786.

Kungulovski G, Jeltsch A. Epigenome Editing: State of the Art, Concepts, and Perspectives. Trends in genetics : TIG. 2016;32(2):101-13. doi:10.1016/j.tig.2015.12.001.

Labahn J, Granzin J, Schluckebier G, Robinson DP, Jack WE, Schildkraut I et al. Three-dimensional structure of the adenine-specific DNA methyltransferase M.Taq I in complex with the cofactor S-adenosylmethionine. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(23):10957-61.

Laszlo AH, Derrington IM, Brinkerhoff H, Langford KW, Nova IC, Samson JM et al. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(47):18904-9. doi:10.1073/pnas.1310240110.

Lauster R, Kriebardis A, Guschlbauer W. The GATATC-modification enzyme EcoRV is closely related to the GATC-recognizing methyltransferases DpnII and dam from E. coli and phage T4. FEBS letters. 1987;220(1):167-76.

Lauster R, Trautner TA, Noyer-Weidner M. Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. Journal of molecular biology. 1989;206(2):305-12.

Lee I, Razaghi R, Gilpatrick T, Molnar M, Gershman A, Sadowski N et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. Nature methods. 2020;17(12):1191-9. doi:10.1038/s41592-020-01000-7.

Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R et al. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development. 1996;122(10):3195-205.

Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell. 1992;69(6):915-26.

Lin CC, Chen YP, Yang WZ, Shen JCK, Yuan HS. Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. Nucleic acids research. 2020;48(7):3949-61. doi:10.1093/nar/gkaa111.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell. 2008;133(3):523-36. doi:10.1016/j.cell.2008.03.029.

Lobner-Olesen A, Skovgaard O, Marinus MG. Dam methylation: coordinating cellular processes. Current opinion in microbiology. 2005;8(2):154-60. doi:10.1016/j.mib.2005.02.009.

Low DA, Casadesus J. Clocks and switches: bacterial gene regulation by DNA adenine methylation. Current opinion in microbiology. 2008;11(2):106-12. doi:10.1016/j.mib.2008.02.012.

Mahdavi-Amiri Y, Chung KCK, Hili R. Single-nucleotide resolution of N-6-adenine methylation sites in DNA and RNA by nitrite sequencing. Chem Sci. 2021;12(2):606-12. doi:10.1039/d0sc03509b.

Maier JAH, Möhrle R, Jeltsch A. Design of synthetic epigenetic circuits featuring memory effects and reversible switching based on DNA methylation. Nature communications. 2017;8:15336. doi:10.1038/ncomms15336.

Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. The Journal of biological chemistry. 2011;286(41):35334-8. doi:10.1074/jbc.C111.284620.

Malone T, Blumenthal RM, Cheng X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. Journal of molecular biology. 1995;253(4):618-32. doi:10.1006/jmbi.1995.0577.

Marinus MG, Casadesus J. Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. FEMS microbiology reviews. 2009;33(3):488-503. doi:10.1111/j.1574-6976.2008.00159.x.

Marinus MG, Morris NR. Isolation of deoxyribonucleic acid methylase mutants of Escherichia coli K-12. Journal of bacteriology. 1973;114(3):1143-50.

Martin JL, McMillan FM. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Current opinion in structural biology. 2002;12(6):783-93.

Matic I, Taddei F, Radman M. Genetic barriers among bacteria. Trends Microbiol. 1996;4(2):69-72. doi:10.1016/0966-842X(96)81514-9.

McIntyre ABR, Alexander N, Grigorev K, Bezdan D, Sichtig H, Chiu CY et al. Single-molecule sequencing detection of N6-methyladenine in microbial reference materials. Nat Commun. 2019;10(1):579. doi:10.1038/s41467-019-08289-9.

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008;454(7205):766-70. doi:10.1038/nature07107.

Meselson M, Yuan R, Heywood J. Restriction and modification of DNA. Annual review of biochemistry. 1972;41:447-66. doi:10.1146/annurev.bi.41.070172.002311.

Modrich P. Studies on sequence recognition by type II restriction and modification enzymes. CRC critical reviews in biochemistry. 1982;13(3):287-323.

Munzel M, Globisch D, Bruckl T, Wagner M, Welzmiller V, Michalakis S et al. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew Chem Int Ed Engl. 2010;49(31):5375-7. doi:10.1002/anie.201002033.

Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature. 2013;502(7470):249-53. doi:10.1038/nature12488.

Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99(3):247-57.

Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nature genetics. 1998;19(3):219-20. doi:10.1038/890.

Olivenza DR, Nicoloff H, Sanchez-Romero MA, Cota I, Andersson DI, Casadesus J. A portable epigenetic switch for bistable gene expression in bacteria. Scientific reports. 2019;9(1):11261. doi:10.1038/s41598-019-47650-2.

Ooi SK, Bestor TH. The colorful history of active DNA demethylation. Cell. 2008;133(7):1145-8. doi:10.1016/j.cell.2008.06.009.

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448(7154):714-7. doi:10.1038/nature05987.

Park M, Patel N, Keung AJ, Khalil AS. Engineering Epigenetic Regulation Using Synthetic Read-Write Modules. Cell. 2019;176(1-2):227-38 e20. doi:10.1016/j.cell.2018.11.002.

Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew Chem Int Ed Engl. 2011;50(31):7008-12. doi:10.1002/anie.201103899.

Pingoud A, Jeltsch A. Recognition and cleavage of DNA by type-II restriction endonucleases. European journal of biochemistry / FEBS. 1997;246(1):1-22.

Pingoud A, Jeltsch A. Structure and function of type II restriction endonucleases. Nucleic acids research. 2001;29(18):3705-27. doi:10.1093/nar/29.18.3705.

Pingoud A, Wilson GG, Wende W. Type II restriction endonucleases--a historical perspective and more. Nucleic acids research. 2014;42(12):7489-527. doi:10.1093/nar/gku447.

Pogolotti AL, Jr., Ono A, Subramaniam R, Santi DV. On the mechanism of DNA-adenine methylase. The Journal of biological chemistry. 1988;263(16):7461-4.

Posfai J, Bhagwat AS, Posfai G, Roberts RJ. Predictive motifs derived from cytosine methyltransferases. Nucleic acids research. 1989;17(7):2421-35.

Razin A, Riggs AD. DNA methylation and gene function. Science. 1980;210(4470):604-10.

Reich NO, Mashhoon N. Inhibition of EcoRI DNA methylase with cofactor analogs. The Journal of biological chemistry. 1990;265(15):8966-70.

Ren W, Fan H, Grimm SA, Guo Y, Kim JJ, Yin J et al. Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. Proceedings of the National Academy of Sciences of the United States of America. 2020;117(31):18439-47. doi:10.1073/pnas.2009316117.

Ren W, Fan H, Grimm SA, Kim JJ, Li L, Guo Y et al. DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation. Nat Commun. 2021;12(1):2490. doi:10.1038/s41467-021-22665-4.

Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenetics and cell genetics. 1975;14(1):9-25.

Roberts RJ. On base flipping. Cell. 1995;82(1):9-12.

Roberts RJ, Cheng X. Base flipping. Annual review of biochemistry. 1998;67:181-98. doi:10.1146/annurev.biochem.67.1.181.

Rosic S, Amouroux R, Requena CE, Gomes A, Emperle M, Beltran T et al. Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. Nature genetics. 2018;50(3):452-9. doi:10.1038/s41588-018-0061-8.

Rots MG, Jeltsch A. Editing the Epigenome: Overview, Open Questions, and Directions of Future Development. Methods Mol Biol. 2018;1767:3-18. doi:10.1007/978-1-4939-7774-1_1.

Sakamoto Y, Sereewattanawoot S, Suzuki A. A new era of long-read sequencing for cancer genomics. J Hum Genet. 2020;65(1):3-10. doi:10.1038/s10038-019-0658-5.

Sanchez-Romero MA, Casadesus J. The bacterial epigenome. Nat Rev Microbiol. 2020;18(1):7-20. doi:10.1038/s41579-019-0286-2.

Sanchez-Romero MA, Cota I, Casadesus J. DNA methylation in bacteria: from the methyl group to the methylome. Current opinion in microbiology. 2015;25:9-16. doi:10.1016/j.mib.2015.03.004.

Santi DV, Norment A, Garrett CE. Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. Proceedings of the National Academy of Sciences of the United States of America. 1984;81(22):6993-7.

Schluckebier G, O'Gara M, Saenger W, Cheng X. Universal catalytic domain structure of AdoMet-dependent methyltransferases. Journal of molecular biology. 1995;247(1):16-20. doi:10.1006/jmbi.1994.0117.

Seeman NC, Rosenberg JM, Rich A. Sequence-specific recognition of double helical nucleic acids by proteins. Proceedings of the National Academy of Sciences of the United States of America. 1976;73(3):804-8.

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007;450(7171):908-12. doi:10.1038/nature06397.

Song G, Wang G, Luo X, Cheng Y, Song Q, Wan J et al. An all-to-all approach to the identification of sequence-specific readers for epigenetic DNA modifications on cytosine. Nat Commun. 2021;12(1):795. doi:10.1038/s41467-021-20950-w.

Song J, Pfeifer GP. Are there specific readers of oxidized 5-methylcytosine bases? Bioessays. 2016;38(10):1038-47. doi:10.1002/bies.201600126.

Song J, Rechkoblit O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science. 2011;331(6020):1036-40. doi:10.1126/science.1195380.

Song J, Teplova M, Ishibe-Murakami S, Patel DJ. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science. 2012;335(6069):709-12. doi:10.1126/science.1214453.

Swartz MN, Trautner TA, Kornberg A. Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. The Journal of biological chemistry. 1962;237:1961-7.

Szomolanyi E, Kiss A, Venetianer P. Cloning the modification methylase gene of Bacillus sphaericus R in Escherichia coli. Gene. 1980;10(3):219-25.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930-5. doi:10.1126/science.1170116.

Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A et al. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proceedings of the National Academy of Sciences of the United States of America. 2011;108(22):9055-9. doi:10.1073/pnas.1019629108.

Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. Nature biotechnology. 1998;16(7):652-6. doi:10.1038/nbt0798-652.

Tazi J, Bird A. Alternative chromatin structure at CpG islands. Cell. 1990;60(6):909-20.

Ullrich T, Weirich S, Jeltsch A. Development of an epigenetic tetracycline sensor system based on DNA methylation. PLoS One. 2020;15(5):e0232701. doi:10.1371/journal.pone.0232701.

van der Woude MW. Phase variation: how to create and coordinate population diversity. Current opinion in microbiology. 2011;14(2):205-11. doi:10.1016/j.mib.2011.01.002.

Weinberg DN, Rosenbaum P, Chen X, Barrows D, Horth C, Marunde MR et al. Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands. Nature genetics. 2021;53(6):794-800. doi:10.1038/s41588-021-00856-5.

Wilson GG, Murray NE. Restriction and modification systems. Annual review of genetics. 1991;25:585-627. doi:10.1146/annurev.ge.25.120191.003101.

Wion D, Casadesus J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat Rev Microbiol. 2006;4(3):183-92. doi:10.1038/nrmicro1350.

Wu JC, Santi DV. On the mechanism and inhibition of DNA cytosine methyltransferases. Progress in clinical and biological research. 1985;198:119-29.

Wu JC, Santi DV. Kinetic and catalytic mechanism of HhaI methyltransferase. The Journal of biological chemistry. 1987;262(10):4778-86.

Wu X, Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet. 2017;18(9):517-34. doi:10.1038/nrg.2017.33.

Xu GL, Bestor TH. Cytosine methylation targetted to pre-determined sequences. Nature genetics. 1997;17(4):376-8. doi:10.1038/ng1297-376.

Yamashita Y, Yuan J, Suetake I, Suzuki H, Ishikawa Y, Choi YL et al. Array-based genomic resequencing of human leukemia. Oncogene. 2010;29(25):3723-31. doi:10.1038/onc.2010.117.

Yang X, Lay F, Han H, Jones PA. Targeting DNA methylation for epigenetic therapy. Trends in pharmacological sciences. 2010;31(11):536-46. doi:10.1016/j.tips.2010.08.001.

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 2017;356(6337). doi:10.1126/science.aaj2239.

Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I et al. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic acids research. 2010;38(13):4246-53. doi:10.1093/nar/gkq147.

Zhang ZM, Lu R, Wang P, Yu Y, Chen D, Gao L et al. Structural basis for DNMT3A-mediated de novo DNA methylation. Nature. 2018;554(7692):387-91. doi:10.1038/nature25477.

Figures and figure legends legends

Figure 1: Molecules related to DNA methylation. A) Structures of the methylated bases that occur in DNA and of the AdoMet cofactor, the universal donor for all DNA methylation reactions. B) Space fill model of the structure of B-DNA with a methylated CpG site. The methyl groups are shown in green in the major groove of the DNA. C) Example of the major groove readout of an AT base pair by Gln as proposed by Seeman and colleagues in 1976 (Seeman et al. 1976). This contact is disrupted by methylation of the A at the N6-position.

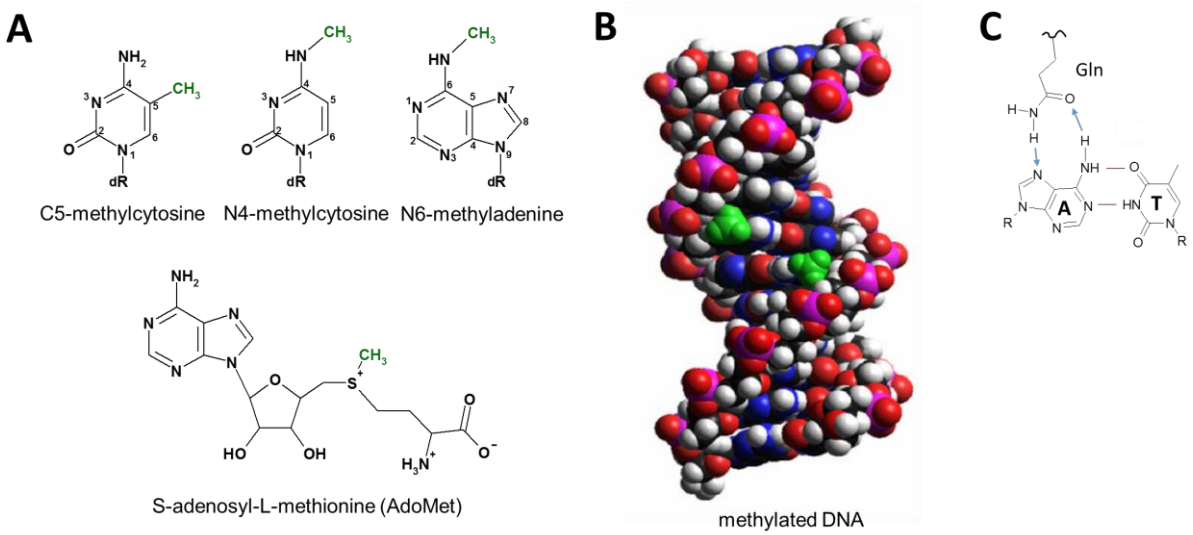

Figure 2: Topological scheme of the universal AdoMet dependent DNA MTase fold. It consists of two Rossman fold half-domains with several conserved amino acid sequences, one subdomain forming the binding site for the AdoMet and the second for the flipped base. DNA recognition is mediated by a DNA binding domain, which is variable in sequence and structure. The linear arrangements of the functional elements varies between different DNA MTases by circular permutation, generating several characteristic subgroups of DNA MTases. A) Schematic representation of the general structure of the DNA MTase fold. B) General structure of Cytosine-C5 MTases. C-D) General structure of three subgroups of Adenine-N6 and Cytosine-N4 MTases.

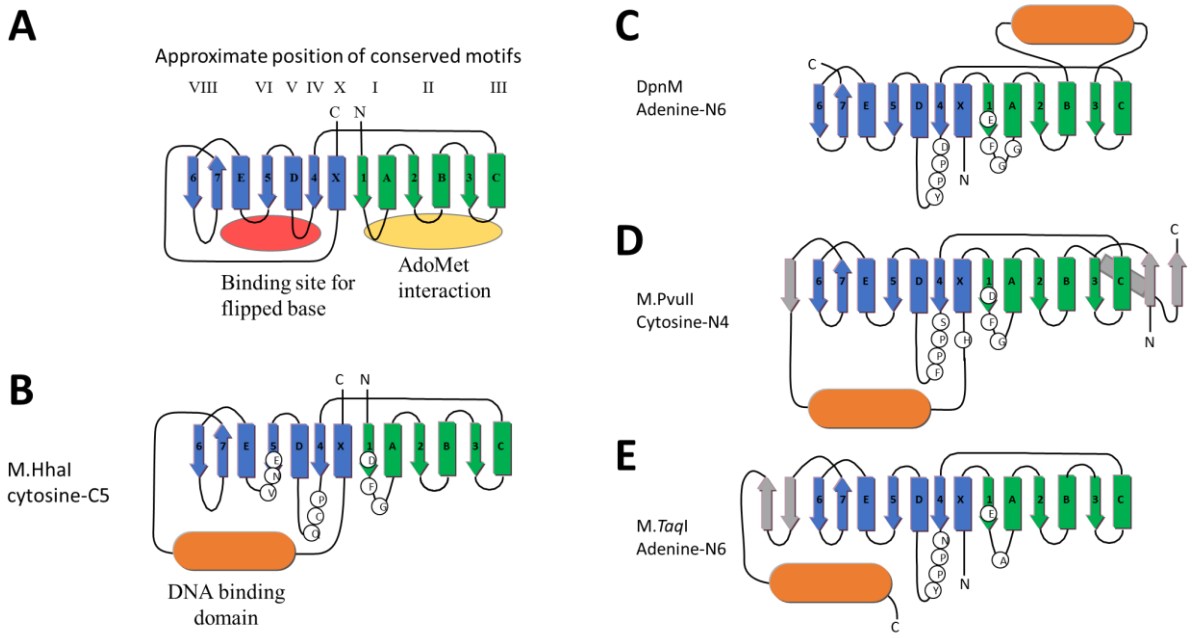

Figure 3: Catalytic mechanism of DNA MTases. A) Mechanism of DNA-(cytosine C5)-MTases. B) Mechanism of DNA-(adenine N6)-MTases.

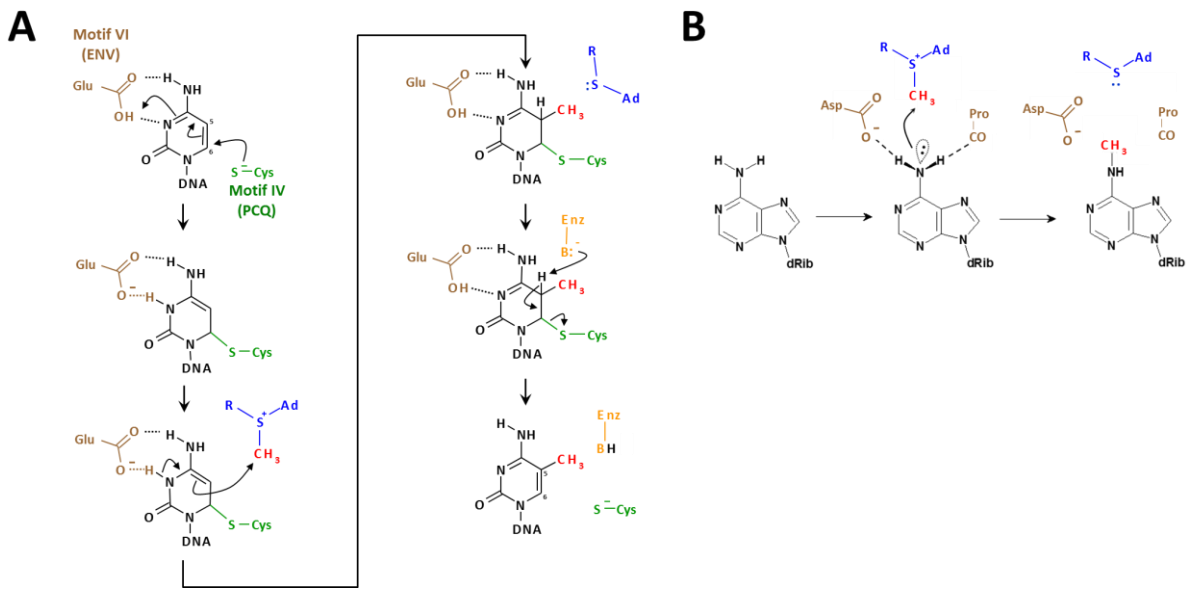

Figure 4: Ribbon model of the structure of the EcoDam DNA MTase (Horton et al. 2006). The AdoMet and the base binding subdomains are shown in green and blue, respectively. AdoMet is displayed in space filled form in yellow. The DNA binding domain is colored in orange. A) Structure of the EcoDam-AdoMet complex. B) Structure of the complex of EcoDam with bound substrate DNA (red, the flipped adenine base is shown in black).

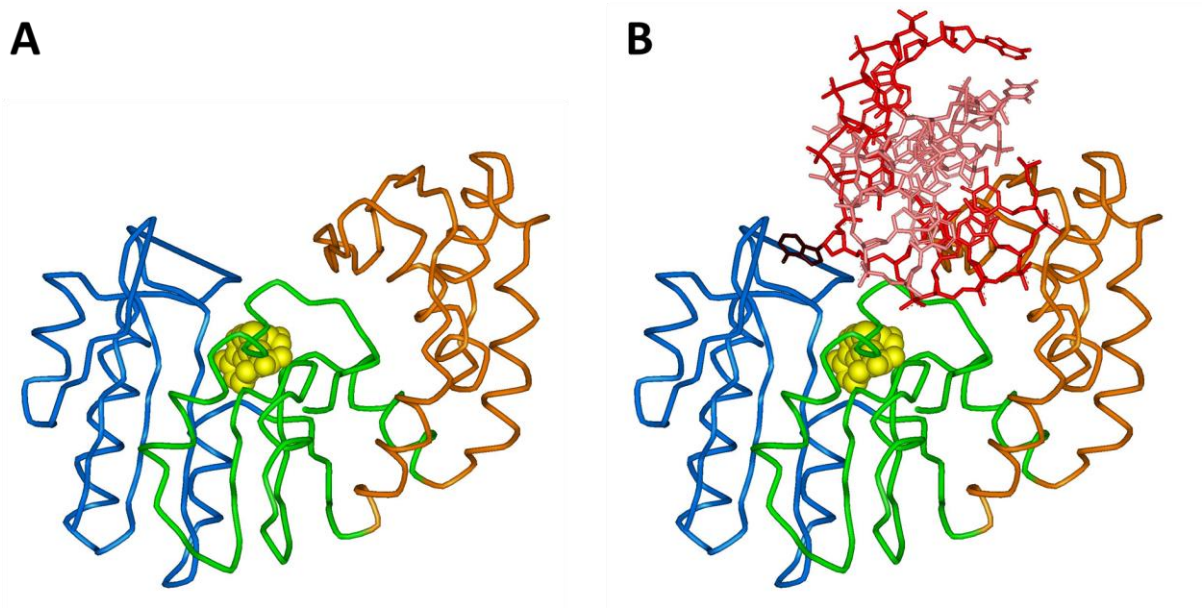

Figure 5: Phylogenic distribution of DNA methylation systems and DNA MTases. The distribution of MTases of the DNMT1 (red), DNMT3 (blue), and chromomethylase families (green) are shown in several characteristic species. Red circles denote plant Met1 homologs, diamonds enzymes of the fungal Dim-2 families, and squares DNMT1 homologs. Blue circles denote plant DRM homologs, and squares DNMT3 enzymes. DNA methylation data were averaged as described in (Jeltsch 2010) and shown for CpG (red), CHH (blue) and CNG (green). CHG methylation is shown only for plants. The phylogenetic tree was generated with National Center for Biotechnology Information taxonomy and the Interactive Tree of Life. Reproduced from (Jeltsch 2010) with modifications. Reprinted with permission from AAAS.

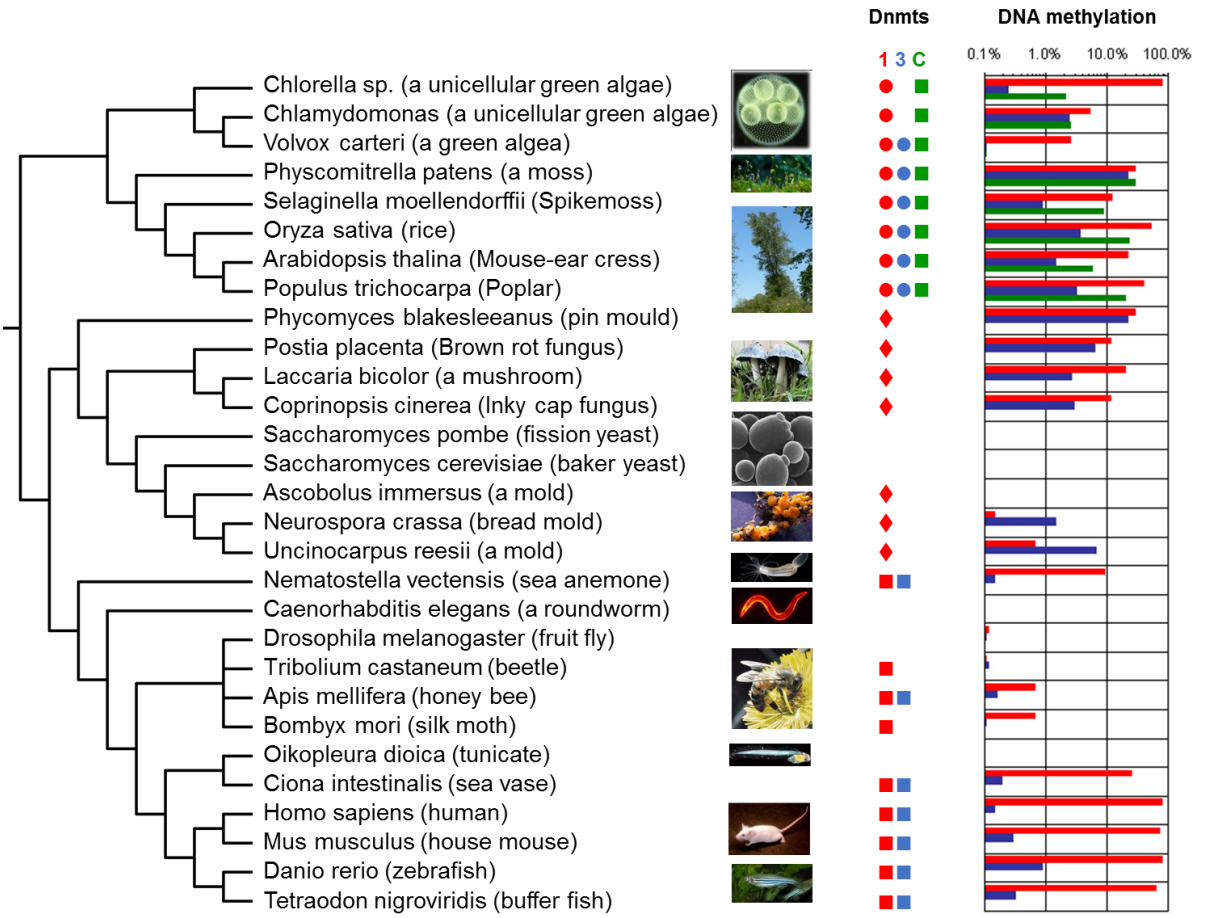

Figure 6: Number of PubMed entries with the term “DNA” and “Methyltransferase” in title or abstract (as of Dec, 2021).