

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Diverse Domains of (Cytosine-5)-DNA Methyltransferases: Structural and Functional Characterization

---

A. Yu. Ryazanova, L. A. Abrosimova,  
T. S. Oretskaya and E. A. Kubareva

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/52046>

---

## 1. Introduction

(Cytosine-5)-DNA methyltransferases (C5-DNA MTases) are enzymes which catalyze methyl group transfer from *S*-adenosyl-*L*-methionine (AdoMet) to C5 atom of cytosine residue in DNA. As a result, AdoMet is converted into *S*-adenosyl-*L*-homocysteine (AdoHcy). The recognition sites of C5-DNA MTases are usually short palindromic sequences (2–6 bp) in double-stranded DNA. One or both DNA strands can be methylated. The introduced methyl group is localized in the major groove of the DNA double helix and thus does not disrupt Watson–Crick interactions [1].

In prokaryotes, DNA methylation underlies several important processes, *e.g.* host and foreign DNA distinction as well as maternal and daughter strand discrimination that is vital for correction of replication errors in the newly synthesized DNA strand. DNA methylation is also responsible for DNA replication control and its interconnection with cell cycle [1]. The majority of the known DNA MTases are components of restriction–modification (R–M) systems which protect bacterial cells from bacteriophage infection. A typical R–M system consists of a MTase which modifies certain DNA sequences and a restriction endonuclease (RE) which hydrolyses DNA if these sequences remain unmethylated [2].

In eukaryotes, DNA methylation has diverse functions such as control of gene expression, regulation of genome imprinting, X-chromosome inactivation, genome defense from endog-

enous retroviruses and transposons, participation in development of immune system and in brain functioning. Anomalous methylation patterns in humans are associated with psychoses, immune system diseases and different cancers [3].

Different C5-DNA MTases share high similarity both in the primary and in the tertiary structure that enables their easy identification by bioinformatic tools. At the moment (July 2012), the Pfam database (<http://pfam.sanger.ac.uk/>) contains 5065 protein sequences that possess the characteristic domain of C5-DNA MTases (PF00145). Among this moiety, 3072 sequences (61%) contain the only domain PF00145 while the others have a duplication of this domain and/or other additional domains. The diversity of such domains fused in a single polypeptide with the C5-DNA MTase domain is rather wide (Table 1): there are MTases, RE, transcription factors, chromatin-associated domains *etc.*

To date, the structural characteristics of the catalytic domain from different MTases, the details of the catalytic mechanism and the biological functions of C5-DNA MTases from different organisms are summarized in a variety of reviews (for example, see [1, 4-8]). Therefore, these aspects are discussed here rather briefly. The present review is focused on the additional activities of C5-DNA MTases, on the structure and functions of the domains which are additional to the catalytic one. The data about C5-DNA MTases have not yet been summarized from this point of view.

## 2. The methyltransferase domain in prokaryotic and eukaryotic (cytosine-5)-DNA methyltransferases

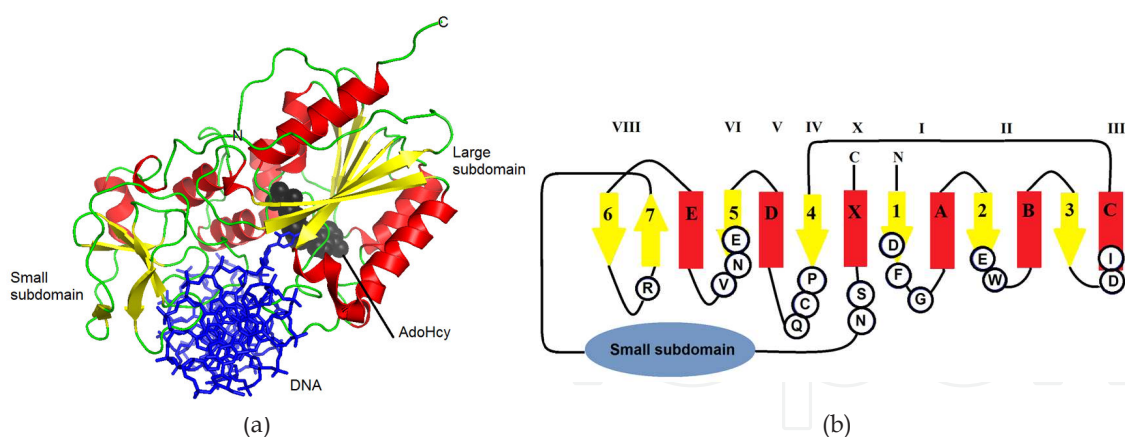
The most studied enzyme among the prokaryotic C5-DNA MTases is MTase HhaI (M.HhaI) from *Haemophilus haemolyticus*. It methylates the inner cytosine residue in the sequence 5'-GCGC-3'/3'-CGCG-5' (italicised). M.HhaI consists of only the MTase domain (Figure 1). The structural organization and catalytic mechanism of C5-DNA MTases were extensively studied using this enzyme as a model.

The catalytic domain of C5-DNA MTases consists of two subdomains, a large one and a small one, separated by a DNA-binding cleft. The tertiary structure of the large (catalytic) subdomain has a common structural core – a  $\beta$ -sheet that consists of 7  $\beta$ -strands and is flanked by 3  $\alpha$ -helices from each side. Six of seven  $\beta$ -strands have a parallel orientation, while the 7<sup>th</sup>  $\beta$ -strand is located between the 5<sup>th</sup> and the 6<sup>th</sup>  $\beta$ -strands in an antiparallel orientation (Figure 1, b). Thus, the large subdomain consists of 2 parts: the first one ( $\beta$ 1– $\beta$ 3) forms the AdoMet binding site while the second one ( $\beta$ 4– $\beta$ 7) forms the binding site for the target cytosine. The small subdomain contains a TRD region (target recognition domain) that has a unique sequence in each MTase and is responsible for the substrate specificity. The small subdomains of C5-DNA MTases vary substantially in size and spatial structure [1].

Domain name and number	Domain description
<b>Methyltransferase domains</b>	
<b>DNA_methylase (PF00145)</b>	C5-cytosine-specific DNA methyltransferase
Eco57I (PF07669)	Eco57I restriction–modification methyltransferase
Methyltransf_26 (PF13659)	Methyltransferase domain
MethyltransfD12 (PF02086)	D12 class N6-adenine-specific DNA methyltransferase
N6_Mtase (PF02384)	N6-DNA methyltransferase
N6_N4_Mtase (PF01555)	DNA methyltransferase
Cons_hypoth95 (PF03602)	Conserved hypothetical protein 95
EcoRI_methylase (PF13651)	Adenine-specific methyltransferase EcoRI
Dam (PF05869)	DNA N6-adenine-methyltransferase (Dam)
<b>Endonuclease domains</b>	
RE_Eco47II (PF09553)	Eco47II restriction endonuclease
RE_HaeII (PF09554)	HaeII restriction endonuclease
RE_HaeIII (PF09556)	HaeIII restriction endonuclease
RE_HpaII (PF09561)	HpaII restriction endonuclease
HNH_3(PF13392)	HNH endonuclease
Vsr (PF03852)	DNA mismatch endonuclease Vsr
DUF559 (PF04480)	Domain of unknown function
BsuBI_PstI_RE (PF06616)	BsuBI/PstI restriction endonuclease C-terminus
TaqI_C (PF12950)	TaqI-like C-terminal specificity domain
<b>Transcription regulators</b>	
HTH_3 (PF01381)	Helix–turn–helix (HTH)
HTH_17 (PF12728)	HTH
HTH_19 (PF12844)	HTH
HTH_23 (PF13384)	Homeodomain-like domain
HTH_26 (PF13443)	Cro/C1-type HTH DNA-binding domain
HTH_31 (PF13560)	HTH
MerR (PF00376)	MerR family regulatory protein
MerR_1 (PF13411)	MerR HTH family regulatory protein
DUF1870 (PF08965)	Domain of unknown function
PHD (PF00628)	PHD-finger (plant homeo domain)
<b>Domains of other DNA-operating enzymes</b>	
SNF2_N (PF00176)	SNF2 family N-terminal domain

Helicase_C (PF00271)	Helicase conserved C-terminal domain
Terminase_6 (PF03237)	Terminase-like family
RVT_1 (PF00078)	Reverse transcriptase (RNA-dependent DNA polymerase)
MutH (PF02976)	DNA mismatch repair enzyme MutH
DEDD_Tnp_IS110 (PF01548)	Transposase
DYW_deaminase (PF14432)	DYW family of nucleic acid deaminases
DNA_pol3_beta_2 (PF02767)	DNA polymerase III beta subunit, central domain
DNA_pol3_beta_3 (PF02768)	DNA polymerase III beta subunit, C-terminal domain
HhH-GPD (PF00730)	HhH-GPD superfamily base excision DNA repair protein
Transposase_20 (PF02371)	Transposase IS116/IS110/IS902 family
<b>Chromatin-associated domains</b>	
Chromo (PF00385)	CHRromatin Organisation MOdifier
PWWP (PF00855)	PWWP domain (conserved Pro-Trp-Trp-Pro motif)
BAH (PF01426)	BAH domain (bromo-adjacent homology)
DMAP_binding (PF06464)	DMAP1-binding domain
DNMT1-RFD (PF12047)	Cytosine specific DNA methyltransferase replication foci domain
zf-CXXC (PF02008)	CXXC zinc finger domain
RCC1_2 (PF13540)	Regulator of chromosome condensation (RCC1) repeat
<b>Others</b>	
Pkinase_Tyr (PF07714)	Protein tyrosine kinase
YTH (PF04146)	YT521-B-like domain
PPR (PF01535)	PPR repeat (pentatricopeptide repeat)
AOX (PF01786)	Alternative oxidase
Cullin (PF00888)	Cullin family
Cyt-b5 (PF00173)	Cytochrome b5-like heme/steroid binding domain
PALP (PF00291)	Pyridoxal-phosphate dependent enzyme
CH (PF00307)	Calponin homology (CH) domain
Dabb (PF07876)	Stress responsive A/B barrel domain
Hint_2 (PF13403)	Hint domain
DUF1152 (PF06626)	Domain of unknown function
DUF3444 (PF11926)	Domain of unknown function

**Table 1.** Domains existing in a single polypeptide chain with the C5-DNA MTase domain (PF00145) according to the Pfam database.

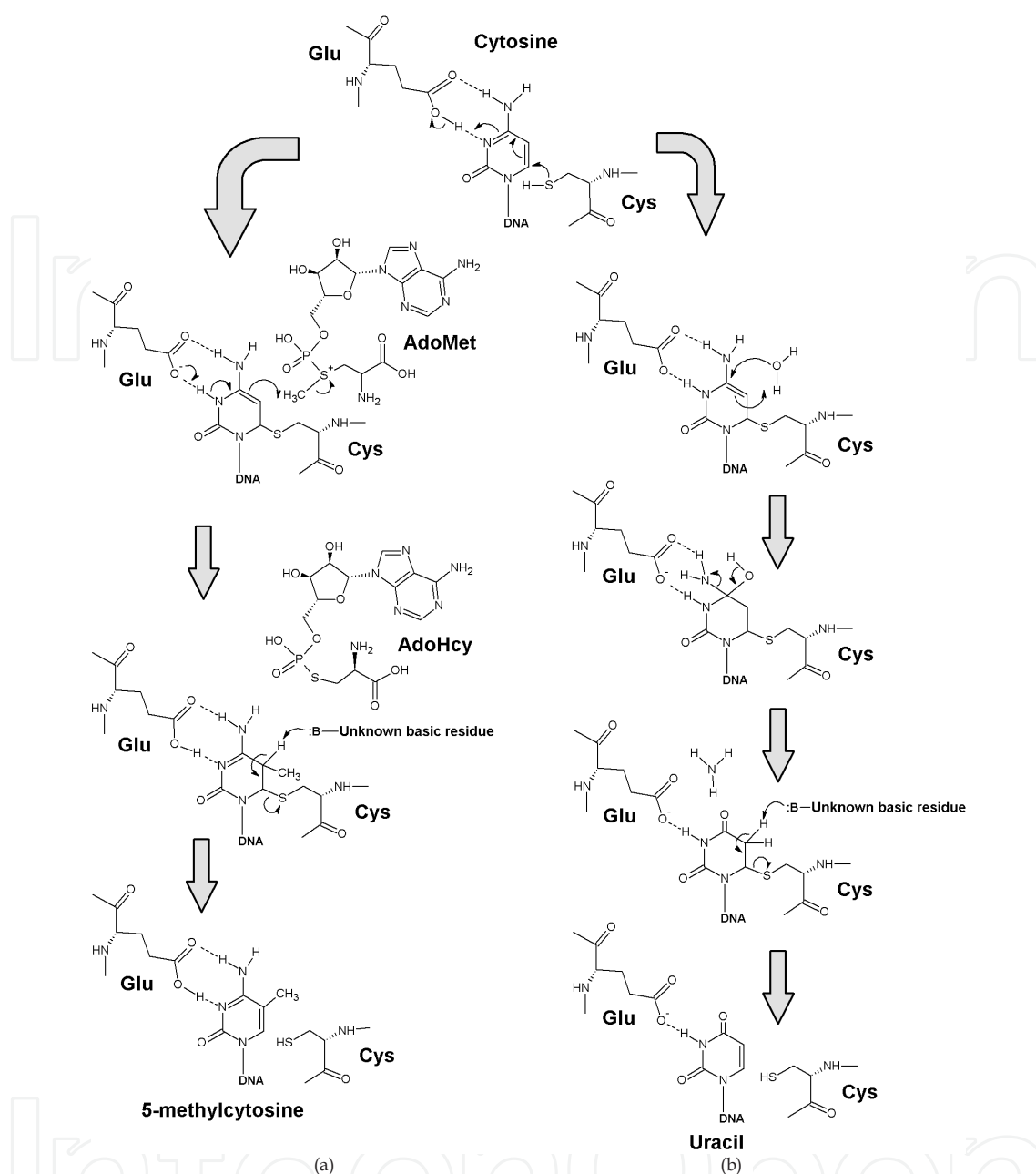


**Figure 1.** a) Secondary and tertiary structure typical for the C5-DNA MTase domain (M.HhaI in complex with DNA and AdoHcy, PDB code: 3mht).  $\alpha$ -Helices are depicted in pink,  $\beta$ -strands are in yellow, DNA is in blue. AdoHcy is in black, spacefill representation. (b) The structural core and the order of the conservative motifs in the large subdomain of C5-DNA MTases.  $\alpha$ -Helices are depicted as boxes,  $\beta$ -strands – as arrows. The conservative motifs are numbered using Roman numerals.

Mammalian DNA MTase Dnmt1 is responsible for maintenance methylation. Its main recognition site is monomethylated 5'-CG-3'/3'-GC-5' DNA fragment. The structures of M.HhaI and Dnmt1 in their complexes with DNA were compared [9]. Their catalytic subdomains are rather similar (the root mean square deviation of  $C_{\alpha}$  atoms is 2.0 Å over 218 aligned residues). However, the TRD primary and tertiary structures differ significantly between Dnmt1 and M.HhaI. The larger part of the Dnmt1 TRD is structurally isolated and stabilized by  $Zn^{2+}$  ion, the latter one being coordinated by three Cys and one His residues. A  $\beta$ -hairpin in the C-terminal part of the Dnmt1 TRD forms hydrophobic contacts with the catalytic subdomain and the BAH1 domain (see section 4.1.5). The side chains of a few residues (presumably arginine) in the catalytic subdomain make contacts with the phosphate groups that flank unmethylated 5'-CG-3'/3'-GC-5' sites. In the M.HhaI structure, the DNA is located in the cleft between the catalytic subdomain and the TRD, whereas the DNA in Dnmt1 complex is distant from the Dnmt1 catalytic center. This is likely to be connected with the fact that the activity of the Dnmt1 catalytic domain is regulated by the N-terminal part of the protein. An isolated Dnmt1 catalytic domain proved to be inactive [10-13].

### 3. The mechanism of DNA methylation

To catalyze the methylation reaction, a MTase binds DNA containing its recognition site and the cofactor AdoMet. Specific DNA-protein contacts are formed between the MTase and heterocyclic bases of the recognition site except the cytosine to be methylated. The target cytosine is methylated according to  $S_N2$  mechanism. The whole methylation process can be divided into 3 steps: a cytosine flipping out of the DNA double helix, a formation of a covalent enzyme-substrate intermediate and a methyl group transfer to the cytosine residue [14-23] (Figure 2).



**Figure 2.** a) Mechanism of cytosine C5 atom methylation by a C5-DNA MTase. (b) Mechanism of cytosine deamination as a possible side reaction of the methylation process.

The target nucleotide flipping can occur spontaneously or with the help of an enzyme [20]. The catalytic loop (residues 81–100 in M.HhaI) shifts substantially towards the DNA almost simultaneously with the target cytosine flipping. As a result, the flipped out base occurs in close proximity with the cofactor molecule inside of a closed catalytic cavity [23–25].

A nucleophilic attack of M.HhaI Cys81 thiol group on the cytosine C6 atom results in a formation of M.HhaI–DNA conjugate. The Cys81 residue is a part of a conservative ProCys dipeptide of the protein (motif IV, Figure 2). The Glu119 residue from GluAsnVal tripeptide (motif



VI) protonates the cytosine N3 atom thus facilitating the nucleophilic attack by the thiol group. In the conjugate, the negative charge of the cytosine C5 atom causes its alkylation by the Ado-Met methyl group. Proton elimination from the C5 atom of the methylated cytosine and  $\beta$ -elimination of the Cys residue result in a breakdown of the covalent DNA–protein complex and restore aromaticity of the modified cytosine base (Figure 2, a). The rate constant for methyl group transfer catalyzed by M.HhaI is about  $0.14\text{--}0.26\text{ s}^{-1}$  [17, 18]. The following release of the reaction products is a rate-limiting step of the M.HhaI catalytic cycle [17, 18].

C5-DNA MTases HhaI and HpaII catalyze the methylation reaction in a distributive manner, *i.e.* dissociate from the DNA substrate after every catalytic act [1, 18]. Each one of these MTases is a component of a R–M system where the cognate RE searches for its recognition sites *via* scanning DNA by a linear diffusion mechanism. The MTase distributivity provides for the corresponding RE a possibility to bind an unmodified site before the MTase and to cleave it. On the other hand, C5-DNA MTase SssI [26] has no cognate RE and is able to methylate several recognition sites located in one DNA substrate in a processive manner, *i.e.* without dissociation from DNA after each catalytic act.

The affinity of prokaryotic C5-DNA MTases to their DNA ligands in the presence of non-reactive cofactor analogs increases in the following manner: dimethylated DNA  $\ll$  unmethylated DNA  $<$  monomethylated DNA [1, 2, 27, 28]. A similar correlation is observed for the eukaryotic enzyme Dnmt1. The Dnmt1 affinity to monomethylated 5′-CG-3′/3′-GC-5′ sites is 2–200 times higher than to unmethylated ones depending on the experimental conditions [13, 29–35]. Dnmt1 modifies monomethylated sites processively (more than 50 sites per one binding act on average). Unmethylated sites (mainly 5′-CCGG-3′) are methylated by Dnmt1 in a distributive manner [36].

Mammalian C5-DNA MTases Dnmt3a and Dnmt3b methylate presumably unmodified sites in DNA and are responsible for *de novo* DNA methylation during embryonic development. Dinucleotides 5′-CG-3′/3′-GC-5′ are the main recognition sites of Dnmt3a and Dnmt3b. These enzymes are also able to modify dinucleotides 5′-CA-3′ but 10–100 times less efficiently [37]. The efficiency of methylation by Dnmt3a and Dnmt3b is also dependent on the sequences flanking the recognition site: 5′-RCGY-3′ is the most frequently methylated site whereas 5′-YCGR-3′ is methylated with a lower efficiency (R and Y are purine and pyrimidine nucleosides respectively). The difference in the methylation rate of these sites can exceed 500 times [38]. The primary structures of the Dnmt3a and Dnmt3b catalytic domains share identity of 84%. In contrast to Dnmt1, isolated catalytic domains of Dnmt3a and Dnmt3b retain their activity [12, 39]. Dnmt3a methylates DNA in a distributive manner while Dnmt3b modifies DNA processively since the DNA binding center of Dnmt3b is more positively charged than DNA binding center of Dnmt3a [37, 39]. Interestingly, a Cys residue substitution in the catalytic ProCys motif of an isolated Dnmt3a catalytic domain does not totally abolish the enzyme activity but merely decreases it 2–6 times. This Cys residue is shown to take part in the DNA–Dnmt3a conjugate formation. However, the Dnmt3a catalytic domain loses its activity completely after a substitution of a Glu residue in the GluAsnVal motif. The active center of Dnmt3a seems to have an unusual conformation and the Cys residue perhaps does not have its optimal orientation. Therefore, the nucleophilic attack onto the cytosine C6 atom can be



performed by some other residue or hydroxyl ion. Post-translational modifications or interaction with other proteins are likely to be needed for the Dnmt3a activation [40].

Murine Dnmt3a is able to perform automethylation, a methyl group transfer to the Cys residue of its own catalytic center in the presence of AdoMet. This reaction is irreversible and rather slow but it can be activated by Dnmt3L. In the presence of a duplex containing 5'-CG-3'/3'-GC-5' sites Dnmt3a methylates the substrate DNA but not its own Cys residue. The automethylation seems to have a regulatory function that enables to inactivate excessive enzyme molecules in a cell. On the other hand, it can be just a side reaction which takes place in the absence of DNA [41].

A protein called Dnmt3L is also a member of Dnmt3 family. It is catalytically inactive but plays an important role as a stimulator of the Dnmt3a and Dnmt3b activity. A structure of a complex consisting of Dnmt3a and Dnmt3L C-terminal domain has been determined by X-ray crystallography (PDB-code 2qrv) [42]. The complex is a heterotetramer where the subunits are localized in the following order: Dnmt3L-Dnmt3a-Dnmt3a-Dnmt3L. Two active centers of Dnmt3a are localized nearby and probably can methylate two 5'-CG-3'/3'-GC-5' sites simultaneously. These recognition sites should be separated by 8–10 bp (about one turn of DNA double helix). Twelve murine genes which undergo maternal imprinting contain 5'-CG-3'/3'-GC-5' sites localized in such a manner. Moreover, highly methylated regions of human chromosome 21 possess 5'-CG-3'/3'-GC-5' sites separated by 9, 18 and 27 bp more frequently in comparison to the unmethylated regions [43]. The effective methylation of these regions could be determined by the proper distribution of 5'-CG-3'/3'-GC-5' sites. A substitution of Dnmt3a and Dnmt3L residues which do not take part in catalysis but are important for the interface formation (Dnmt3a-Dnmt3L or Dnmt3a-Dnmt3a interfaces) results in a suppression of Dnmt3a activity. This fact confirms the importance of the appropriate complex formation between Dnmt3a and Dnmt3L [42]. The orientation of the Dnmt3a-Dnmt3L complex relative to the substrate DNA is still unclear.

## 4. Additional functions of prokaryotic (cytosine-5)-DNA methyltransferases

### 4.1. DNA methyltransferases with multiple methyltransferase domains

Some C5-DNA MTases contain more than one MTase domain. The “additional” domain can belong to (cytosine-5)-DNA or (adenine-N6)-DNA MTases. Up to date, the Pfam database contains 5065 protein sequences that possess the characteristic domain of C5-DNA MTases (PF00145). Among them, 676 sequences (13%) contain two PF00145 domains and 42 sequences (1%) – even three PF00145 domains. Such structures might have arisen as fusions of two adjacent genes encoding different MTases. However, none of these proteins has been studied experimentally.

The ability of a single molecule to methylate both the cytosine C5 atom and the adenine N6 atom is typical for MTases that recognize asymmetric DNA sequences. This phenomenon

was demonstrated for M.Alw26I from *Acinetobacter lwoffii* RFL26 (recognition site 5'-GTCTC-3'/3'-CAGAG-5') and for M.Esp3I from *Hafnia alvei* RFL3 (recognition site 5'-CGTCTC-3'/3'-GCAGAG-5') [44]. Each of these proteins contains the N6-DNA MTase domain in its N-terminal part and the C5-DNA MTase domain – in its C-terminal part [45]. The ability of each domain to methylate its recognition site in the absence of the second domain has not been investigated.

A gene coding for another enzyme consisting of two MTase domains has been constructed from two genes of *Helicobacter pylori* 26695 [46]. These genes are located in tandem orientation and code for DNA MTases M.HpyAVIB and M.HpyAVIA. One nucleotide insertion before the stop codon results in a formation of a fused gene. A similar mutation has been occurred naturally in *H. pylori* D27 strain [47]. M.HpyAVIB methylates the cytosine C5 atom and M.HpyAVIA methylates the adenine N6 atom in the sequence 5'-CCTC-3'/3'-GGAG-5' (italicised). The obtained bifunctional protein contains the C5-DNA MTase domain and the N6-DNA MTase domain in its N-terminal and C-terminal parts, respectively. The both domains recognize in DNA the same sites as the initial proteins. The methylation kinetics and the properties of point mutants demonstrate that these domains function independently from each other. Each one of them contains its own catalytic and AdoMet binding motifs [46].

#### 4.2. Deamination of cytosine and 5-methylcytosine

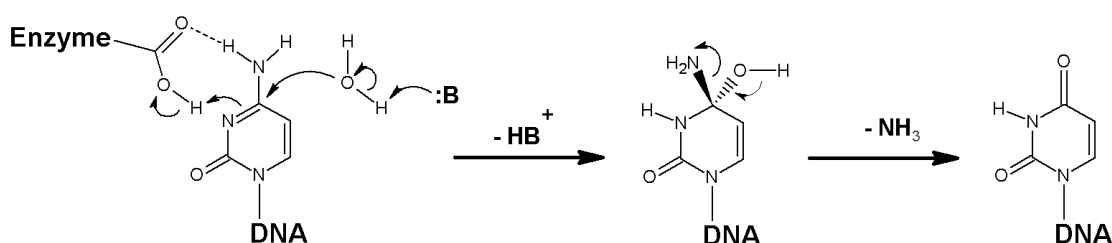
Prokaryotic C5-DNA MTases M.HpaII, M.HhaI, M.SssI, Dcm (from *E. coli*), M.EcoRII, and M.SsoII are shown to increase the rate of C → dU → T and m<sup>5</sup>C (5-methyl-2'-deoxycytidine) → T mutagenesis *in vitro* [48-55]. Some of them demonstrate the mutagenic activity (M.HpaII, M.EcoRII, and Dcm) also *in vivo* increasing the mutagenesis rate up to 50 times [50, 54, 56]. Interestingly, prokaryotic M.MspI does not stimulate cytosine deamination *in vitro* but its mutagenic effect is comparable with the effect of other C5-DNA MTases when M.MspI is expressed in *E. coli* cells [57]. Moreover, M.EcoRII is shown to catalyze 5-methylcytosine conversion into thymine [52].

The enzymatic catalysis of cytosine deamination is a side reaction of the methylation process. According to the standard mechanism (Figure 2), the cysteine thiol group performs a nucleophilic attack onto the cytosine C6 atom and at the same time the cytosine N3 atom gets protonated that altogether leads to the DNA-enzyme conjugate formation. As a result, the cytosine base aromaticity is disrupted and the C5 atom becomes negatively charged. However, the following step (the methyl group transfer to the C5 atom) becomes impossible in the absence of AdoMet. If water penetrates into the enzyme active center, hydroxylation of the cytosine C4 atom is likely to occur. These processes initiate the deamination reaction (Figure 2, b). Afterwards, the cytosine amino group is substituted with a carboxyl group and the base is converted into uracil [58, 59]. The presence of AdoMet or AdoHcy prevents water penetration into the active center and therefore inhibits deamination. A point mutant of M.HpaII incapable of AdoMet binding is a very effective catalyst of C → dU conversion [56].

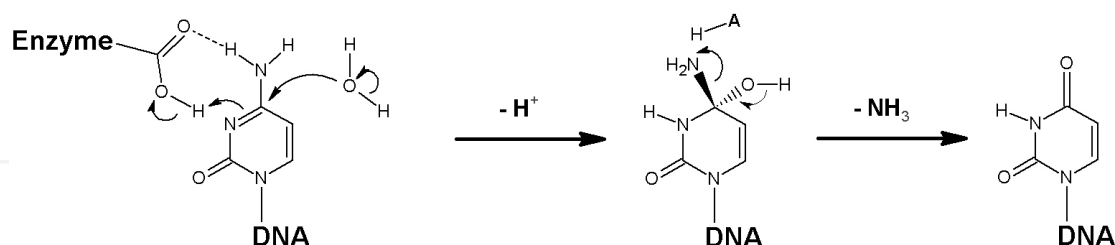
The AdoMet analogs such as sinefungin and 5'-amino-5'-deoxyadenosine can increase the rate of enzymatic deamination even in the presence of AdoMet and AdoHcy [53]. They seem

to trigger other reaction mechanisms which are not completely clarified yet. The supposed mechanisms include a water molecule direct attack onto the cytosine C4 atom. It becomes possible after the N3 atom protonation by a MTase which is a step of the methylation reaction. Two alternative mechanisms are suggested (Figure 3). According to the first one (mechanism 1), the hydroxyl group of 5'-amino-5'-deoxyadenosine activates a water molecule producing a hydroxyl ion which attacks the cytosine C4 atom (Figure 3). According to the mechanism 2, 5'-amino-5'-deoxyadenosine acts as an acid and protonates the cytosine N4 atom (Figure 3) thus facilitating the amino group elimination in the form of ammonium ion [59]. The mechanisms 1 and 2 are based on a water molecule direct attack onto the C4 atom and differ considerably from the others as they do not require a MTase interaction with the cytosine C6 atom. Therefore, a mutant form of a MTase which does not catalyze the methylation reaction should be able to catalyze these side reactions. Indeed, it has been shown experimentally for an M.EcoRII point mutant where catalytic Cys was substituted with Ala [59].

### Mechanism 1



### Mechanism 2



**Figure 3.** The supposed mechanisms of cytosine deamination catalyzed by an AdoMet analogue (sinefungin or 5'-amino-5'-deoxyadenosine). The AdoMet analogue might act as a base (mechanism 1) or as an acid (mechanism 2).

Both the methylation and the deamination reactions require flipping out of the cytosine residue. Thus, the longer cytosine base remains flipped out the faster these reactions go. The both reactions share a common intermediate – the enzyme–substrate conjugate (Figure 2). The usage of tritiated cytosine allows an estimation of tritium to hydrogen exchange rate at the cytosine C5 atom which can serve as a measure of the conjugate formation rate. In the absence of the cofactor, such an exchange catalyzed by murine Dnmt1 is slower than the one catalyzed by M.HhaI [60]. Thus, Dnmt1 forms the conjugate in the absence of AdoMet with a lower probability and therefore is less mutagenic than M.HhaI [60]. On the contrary, pro-

karyotic MTases catalyze effectively the deamination reaction in the absence of the cofactor [48, 53, 57].

The different rates of the covalent enzyme–substrate complex formation in the absence of the cofactor could reflect different physiological functions of prokaryotic and eukaryotic C5-DNA MTases. Limited nutrition decreases AdoMet amounts in prokaryotic cells. The mutations derived from the cytosine deamination might not be lethal for a bacterial cell and could help to prevent hydrolysis of cellular DNA by phage endonucleases. So, the ability of a C5-DNA MTase to catalyze deamination can turn out a physiological advantage for a bacterial cell [5]. On the contrary, mammalian cells are not likely to benefit from this kind of mutagenesis and therefore have developed mechanisms which provide low mutagenesis rate.

### 4.3. Topoisomerase activity

Two prokaryotic C5-DNA MTases are shown to have a topoisomerase activity, namely M.SssI from *Spiroplasma* MQI and M.MspI from *Moraxella* species. M.SssI is not a part of R–M system (there is no corresponding RE in *Spiroplasma* genome). As well as eukaryotic MTases, M.SssI modifies cytosine C5 atom in 5'-CG-3'/3'-GC-5' sequences [61]. In the presence of 10 mM Mg<sup>2+</sup>, M.SssI provides relaxation of negatively supercoiled plasmids which leads to accumulation of plasmids with different degree of supercoiling. The obtained set of plasmids is similar to the products of topoisomerase I from calf thymus. An ATP addition does not influence the topoisomerase activity of M.SssI. Since type II topoisomerases need ATP for the enzymatic activity, M.SssI can be regarded as a type I topoisomerase [62]. The MTase and the topoisomerase activities of M.SssI are functionally independent. The methylation process requires AdoMet, while the topoisomerase reaction demands Mg<sup>2+</sup> ions. The M.SssI conservative motifs IV and VIII share a certain similarity with topoisomerase sequences. In particular, the motif IV contains Tyr that is an important catalytic residue in topoisomerases. A more detailed analysis of M.SssI regions responsible for the topoisomerase activity has not been conducted.

Different speculations are proposed to explain why these two activities are combined in one M.SssI molecule. Firstly, the topoisomerase activity alters the supercoiling degree of plasmid DNA and thus could facilitate or complicate the cytosine flipping out of the DNA helix. Secondly, the methylation of 5'-CG-3'/3'-GC-5' sites by M.SssI could change the DNA structure. For example, a negatively supercoiled DNA region with large amount of methylated 5'-CG-3'/3'-GC-5' sites is likely to be converted into Z-form. For the B-form restoration, topoisomerase activity is necessary. Thirdly, the change in DNA topology perhaps influences the level of gene expression in *Spiroplasma*. Finally, the two different activities can be combined in one protein for the purpose of genome economy. *Spiroplasma* belongs to mycoplasmas – cellular organisms which have the smallest genome (from 600 to 1800 kbp). For a comparison, the T4 bacteriophage genome consists of 165 kbp and the *E. coli* genome – of 4600 kbp [62].

M.MspI is a part of MspI R–M system and methylates the first cytosine residue in the sequence 5'-CCGG-3'/3'-GGCC-5' [63]. The unique property of this MTase is its ability to bend DNA at 142 ± 4° upon its binding to the methylation site. This was demonstrated using 127



bp DNA duplex and has not been shown for any other MTase [64]. Unlike M.SssI, M.MspI has an N-terminal part responsible for the topoisomerase activity. This part consists of 107 residues and is located before the conservative motif I. There are two regions of M.MspI that share similarity with topoisomerase sequences: the residues 32–98 and the conservative motif VIII. In contrast to M.SssI, there is no similarity between the M.MspI conservative motif IV and the topoisomerase sequences. A mutant form of M.MspI lacking the first 34 residues retains the ability to methylate DNA but loses its topoisomerase activity. Mutant proteins M.MspI(W34A) and M.MspI(Y74A) also do not have topoisomerase activity but are still able to methylate the DNA substrate [65]. Additionally, the M.MspI C-terminal part contains a region (245–287 a.a.) that shares similarity with DNA ligase I active center. This is in accordance with the fact that the topoisomerase I activity includes ligation of the DNA strands [65].

#### 4.4. (Cytosine-5)-DNA methyltransferases as transcription factors

According to the Table 1, some C5-DNA MTases contain domains that can function as transcription factors. These domains are located in the N-terminal parts of the proteins and are followed by the MTase domains. The main structural element of these domains is a characteristic helix-turn-helix (HTH) motif that is also present in many transcription factors. To date, the Pfam database contains 68 sequences that include a domain with HTH motif followed by the C5-DNA MTase domain. Among them, 25 sequences belong to the HTH\_3 family (PF01381). The ability to downregulate expression of its own gene was shown experimentally only for 7 DNA MTases (M.MspI, M.EcoRII, M.ScrFIA, M1.LlaJI, M.Eco47II, M.SsoII, and M.Ecl18kI) [66–68]. Among them, M.SsoII and M.Ecl18kI are the most remarkable ones as they not only suppress the transcription of their own genes but also stimulate the transcription of their cognate RE genes.

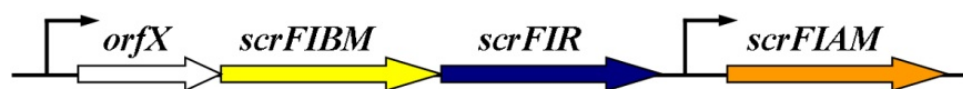
M.EcoRII from *E. coli* R245 strain methylates C5 atom of the second cytosine in the sequence 5'-CCWGG-3'/3'-GGWCC-5' (W = A or T). *In vitro* experiments demonstrate that M.EcoRII can bind both its methylation site and the promoter region of its own gene. The enzyme's binding site has been determined by footprinting with DNase I: M.EcoRII protects 47 nucleotides in the "top" strand and 49 nucleotides in the "bottom" strand from DNase I hydrolysis [69]. Thus, the binding site of M.EcoRII is located upstream of the MTase gene coding region and overlaps with its -10 and -35 promoter elements. This localization of M.EcoRII prevents RNA polymerase binding to the promoter and results in a suppression of the MTase gene transcription. The M.EcoRII binding site in the promoter region contains an imperfect inverted repeat (with 2 nucleotide substitutions). The repeat consists of two 11 bp sequences spaced by 12 bp. This kind of symmetry supposes the MTase to bind the promoter region as a dimer though this protein is a monomer in solution. Investigation of M.EcoRII catalytically inactive forms shows that the efficiency of the MTase binding to the promoter region does not depend on its ability to methylate substrate. These facts demonstrate that M.EcoRII consists of two domains: a catalytic domain and a domain responsible for the interaction with the promoter region [70].

R-M system Eco47II from *E. coli* RFL47 strain contains a MTase that is also able to downregulate the expression of its own gene. M.Eco47II methylates the cytosine C5 atom in the se-

quence 5'-GGNCC-3'/3'-CCNGG-5' (N = A, G, C, T). It remains unclear which one cytosine is modified. The M.Eco47II N-terminal part is predicted to contain an HTH motif and is demonstrated to be responsible for the transcription regulation but not for the MTase activity. Mutations introduced into the catalytic center of the enzyme result in suppression of the methylation activity but do not disrupt the regulatory function [68].

In the MspI R-M system, the *mspIM* and *mspIR* genes are transcribed divergently from the complementary DNA strands and their promoter regions (−35 elements) are separated by 6 bp. The regulatory site of M.MspI is located in the promoter region of the *mspIM* gene and contains a 12 bp inverted repeat. M.MspI protects from DNase I hydrolysis the region from −34 to +17 position in the “top” DNA strand and the region from −33 to +17 position in the “bottom” DNA strand (the numbers indicate the position relatively to the start point of the *mspIM* gene transcription). So, M.MspI interaction with the regulatory region prevents RNA polymerase binding to the promoter region and blocks transcription initiation from the *mspIM* gene. At the same time, M.MspI does not interact with the promoter region of the *mspIR* gene and does not interfere with the expression of RE MspI [71].

R-M system ScrFI from *Lactococcus lactis* subsp. *cremoris* also contains a C5-DNA MTase that regulates gene expression in its R-M system [66]. The RE gene (*scrFIR*) is flanked by two genes that code for MTases: *scrFIBM* and *scrFIAM*. The *scrFIAM* gene has its own promoter while the *scrFIBM* and *scrFIR* genes are transcribed together with a gene of unknown function – *orfX* (Figure 4). Both MTases from the ScrFI R-M system recognize and methylate the cytosine base in the sequence 5'-CCNGG-3'/3'-GGNCC-5' (it remains unknown which one cytosine is methylated). The biological sense of existing of two MTase genes is not clear. The N-terminal part of M.ScrFIA is predicted to contain an HTH motif [72]. M.ScrFIA is also shown to bind to the regulatory region – a 15 bp inverted repeat located before the transcription start point for the *scrFIAM* gene. M.ScrFIA binds to this region and inhibits the expression of its own gene [72].

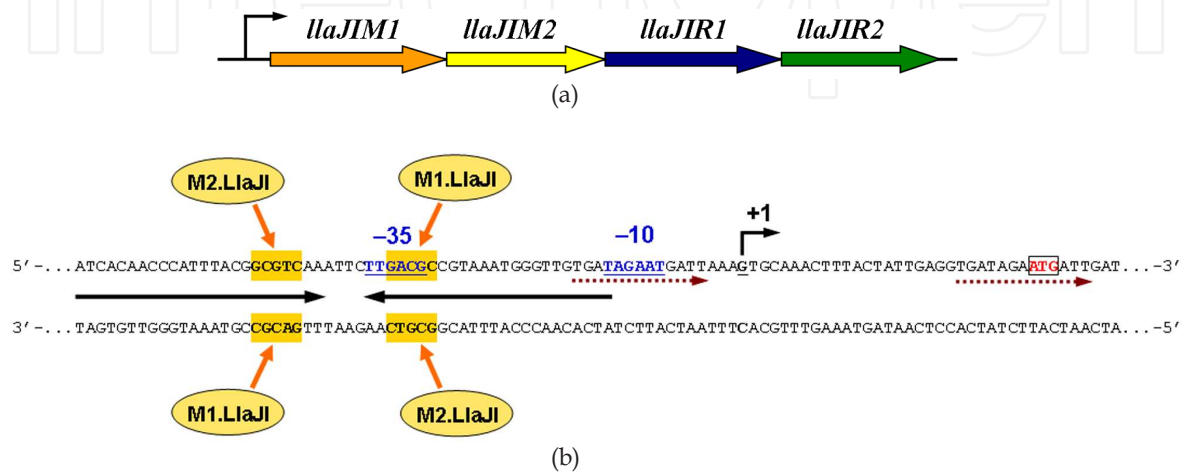


**Figure 4.** Genetic organization of the ScrFI R-M system. Thin arrows indicate the transcription initiation points.

R-M system LlaJI from *Lactococcus lactis* contains two C5-DNA MTases – M1.LlaJI and M2.LlaJI. These MTases have the same recognition site 5'-GACGC-3'/3'-CTGCG-5' but methylate different cytosine bases (italicised) in the “top” and the “bottom” DNA strands respectively [73, 74]. Genes coding for these MTases and two RE compose one operon and are transcribed from the same promoter (Figure 5, a). The promoter region contains a 24 bp inverted repeat that is a regulatory site for the MTases. This palindrome sequence contains two methylation sites of the LlaJI MTases one of which overlaps with −35 promoter element (Figure 5, b). At first, M2.LlaJI modifies both sites that enables binding of M1.LlaJI and methylation of cytosine bases in the opposite strand. Interaction of M1.LlaJI with the unmethylated substrate has not been demonstrated [73]. M2.LlaJI acts only as a modifying en-



zyme while M1.LlaJI has an additional capability of binding to the inverted repeat that contains -35 promoter element. The N-terminal part of M1.LlaJI is predicted to contain an HTH motif. The M1.LlaJI binding to the inverted repeat results in suppression of gene transcription for the whole LlaJI operon. This mechanism seems to be unique since methylation usually decreases the binding efficiency of a repressor with its operator [74]. The complicated regulatory mechanism is likely to enable fine tuning of transcriptional level in LlaJI R-M system [74].

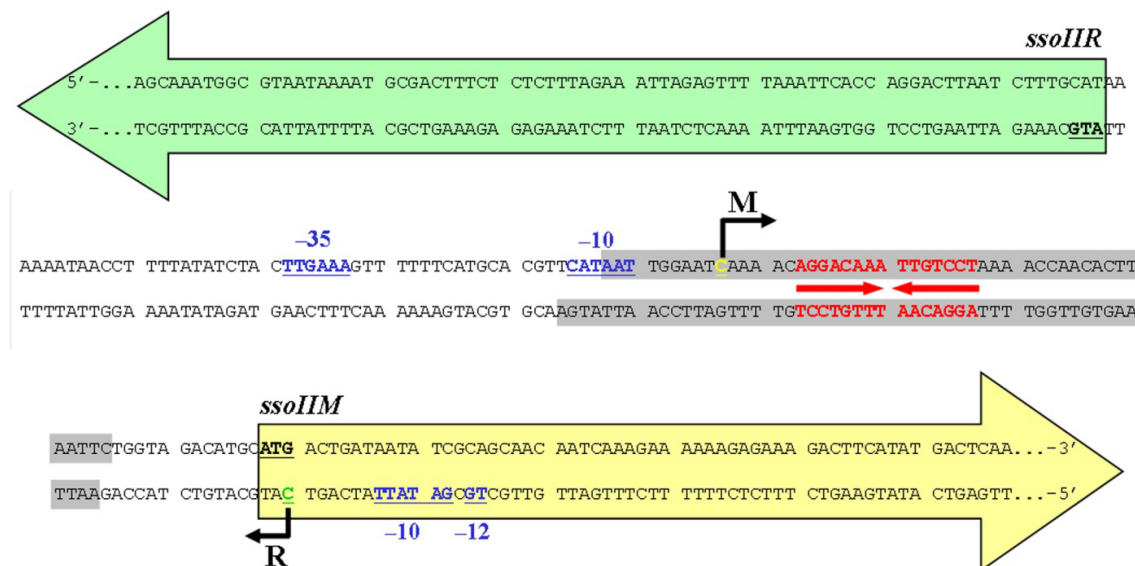


**Figure 5.** a) Genetic organization of the LlaJI R-M system. The thin arrow marks the transcription initiation point of the operon. (b) A fragment of the promoter region of the LlaJI R-M system. Two black arrows indicate the 24 bp inverted repeat. The methylation sites are highlighted with yellow. The transcription start point is marked by “+1”, the promoter -10 and -35 elements are in blue. The start codon is in red.

Transcription regulation has been studied most thoroughly in the R-M systems SsoII from *Shigella sonnei* 47 and Ecl18kI from *Enterobacter cloacae* 18k. The nucleotide sequences of these systems share 99% identity. The sequences of the intergenic regions are completely identical while the proteins differ in 1 amino acid residue. We will refer to such R-M systems as to SsoII-like ones. The genes of the SsoII R-M system are located divergently and spaced by an intergenic region of 109 bp (Figure 6). To investigate the regulation mechanism of the system, two plasmids have been constructed, with both possible combinations of the intergenic region and a *lacZ* gene which encodes  $\beta$ -galactosidase. Thus, the expression of the *lacZ* gene is under control of a promoter of the *ssoIIM* gene (pACYC-SsoIIM) or the *ssoIIR* gene (pACYC-SsoIIR). The  $\beta$ -galactosidase expression level is found to be 540 times higher in the cells containing the pACYC-SsoIIM plasmid in comparison to the other plasmid. Therefore, the expression from the *ssoIIM* promoter is much higher than from the *ssoIIR* promoter. When transformed with an additional plasmid where the *ssoIIM* gene is under its own promoter, the cells containing pACYC-SsoIIM demonstrate a 20-fold decrease of the  $\beta$ -galactosidase expression while the cells containing pACYC-SsoIIR display an 8-fold increase of it [75]. Thus, M.SsoII is shown to downregulate the expression of its own gene and to stimulate the expression of the cognate RE gene. The transformation of the cells with a plasmid encoding a mutant M.SsoII without its first 72 residues does not influence the  $\beta$ -galactosidase expres-

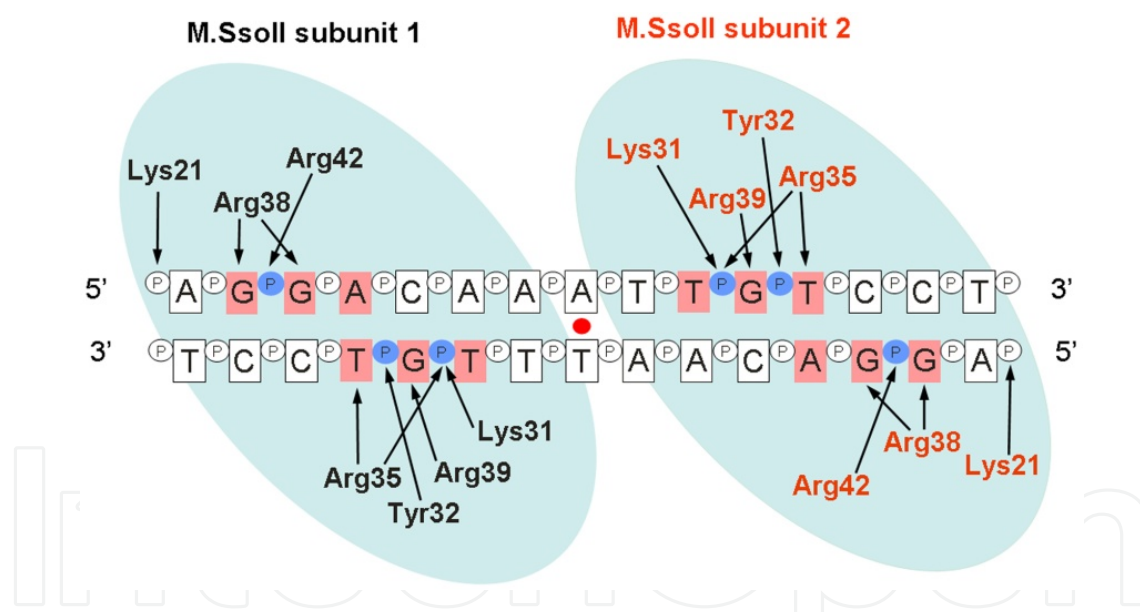
sion level. The same effect is observed after the transformation with a plasmid encoding M.NlaX – a protein homologous to the M.SsoII domain responsible for methylation. On the contrary, a plasmid encoding a fusion of the M.SsoII first 72 residues with the full-length M.NlaX gives the same effect as the plasmid encoding the full-length M.SsoII [75]. These experiments demonstrate that the M.SsoII ability to regulate transcription in the SsoII R–M system is determined by its N-terminal part (72 residues). This part is predicted to contain an HTH motif.

M.SsoII is a typical C5-DNA MTase which modifies the second cytosine in the sequence 5'-CCN<sub>2</sub>GG-3'/3'-GGNCC-5'. Moreover, M.SsoII interacts with the intergenic region of the SsoII R–M system protecting from DNase I hydrolysis 48 nucleotides in the top strand and 52 nucleotides in the bottom strand [75]. The intergenic region contains a 15 bp inverted repeat (regulatory site) [76]. Seven guanine bases interacting with M.SsoII are identified using protection footprinting; six of them are located inside the regulatory site symmetrically relative to the central A•T pair (Figure 7) [76]. Interference footprinting experiments with formic acid, hydrazine, dimethyl sulfate, and *N*-ethyl-*N*-nitrosourea show 6 guanine, 2 adenine, and 4 thymine residues as well as 6 phosphate groups interacting with M.SsoII. These nucleotides form two symmetrically located clusters: 5'-GGA-3' and 5'-TGT-3' in each DNA strand of the regulatory site (Figure 7) [77]. Such a symmetrical interaction with the both halves of the palindromic site is typical for many regulatory proteins which bind to the operator sequence in a dimeric form and contain an  $\alpha$ -helix that interacts with the DNA major groove recognizing the specific sequence. M.SsoII as a transcription factor is supposed to have the same mechanism of functioning.



**Figure 6.** Genetic organization of the *SsoII* R–M system. The RE and MTase genes are depicted with green and yellow arrows respectively. The initiation codons are underlined. The region protected by M.SsoII from DNase I hydrolysis is highlighted with grey. The regulatory site is in red. The transcription initiation points of the RE and MTase genes are marked by black arrows, their promoter elements are in blue.

Arg35 or Arg38 substitution with Ala in M.Ecl18kI significantly impairs the protein binding to the regulatory site [79]. According to the computer simulation performed for M.SsoII, these residues belong to the second recognizing helix of the HTH motif [78]. Arg38 is supposed to form contacts with guanine bases of 5'-GGA-3' trinucleotide in one DNA strand while Arg35 can interact with thymine bases and DNA backbone of 5'-TGT-3' trinucleotide in the other strand (Figure 7). Amino acid substitutions in the M.Ecl18kI N-terminal part influence the ability of this protein to regulate transcription in the R-M system. However, there is no correlation between M.Ecl18kI affinity to the regulatory site and the amounts of the RNA transcripts [80]. Amino acid substitutions in the M.Ecl18kI N-terminal part increase the methylation activity of this enzyme in most of cases. There is also an inverse relationship: an M.SsoII point mutant which has Cys142 substituted with Ala is catalytically inactive but demonstrates an increased affinity to the regulatory site and effectively regulates transcription in the SsoII R-M system [79]. Thus, the interconnection between the two DNA binding sites is experimentally demonstrated for the SsoII-like DNA MTases. Moreover, it has been shown recently that M.SsoII binding to the regulatory site prevents its interaction with the methylation site. Thus, the two functions of the protein are mutually exclusive [67, 81].



**Figure 7.** Scheme of the contacts formed between the two N-terminal domains of M.SsoII and the regulatory DNA site. The heterocyclic bases and the phosphate groups interacting with M.SsoII identified by footprinting are marked by red and blue respectively.

The transcription start point of the RE gene is located in the beginning of the MTase gene in the SsoII and Ecl18kI R-M systems (Figure 6) [67]. The transcription start point of the MTase gene is located inside the region protected by M.SsoII from DNase I hydrolysis, 5 bp away from the regulatory site. The suppression of the MTase gene transcription is based on the competitive binding of the MTase and RNA polymerase with the intergenic region of the SsoII (or Ecl18kI) R-M system. The MTase interaction with the regulatory site does not inter-

ferre with RNA polymerase binding to the RE gene promoter. Thus, the RE gene is activated indirectly *via* averting of RNA polymerase binding with the MTase gene promoter [80].

## 5. Domains of eukaryotic DNA methyltransferases

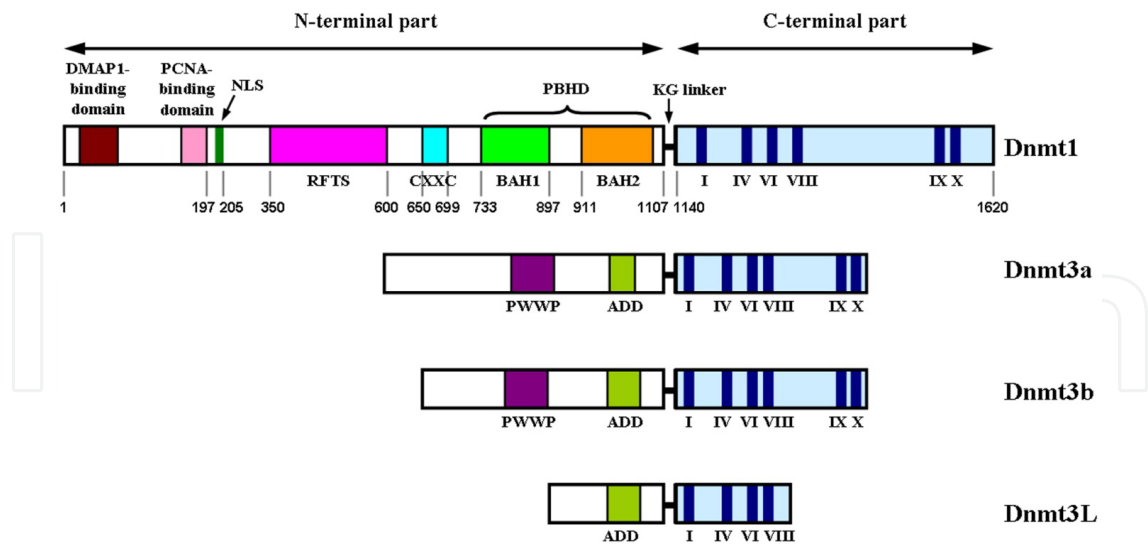
All known eukaryotic DNA MTases methylate C5 atom of cytosine. Eukaryotic MTases are usually multidomain proteins (Figure 8).

### 5.1. Functional domains of mammalian Dnmt1

Murine and human Dnmt1 consist of 1620 and 1616 amino acid residues respectively. The primary structures of these proteins share 85% identity. Dnmt1 molecule contains the following domains and functionally important regions (listed starting from the N-end, Figure 8):

1. charge-rich domain or DMAP1-binding domain (DMAP1 is DNA methyltransferase associated protein, a transcription repressor);
2. PCNA-binding domain, PBD;
3. at least three functionally independent nuclear localization signals, NLS [82];
4. RFTS domain (replication foci targeting sequence), also called TS (targeting sequence), RFD (replication foci domain) or TRF (targeting to replication foci) domain;
5. cysteine-rich Zn<sup>2+</sup>-binding domain, also called CXXC domain;
6. BAH1 and BAH2 (bromo-adjacent homology or bromo-associated homology) domains which are parts of the so-called PBHD domain (polybromo homology domain);
7. KG linker (consists of Lys and Gly residues) which connects N- and C-terminal parts of Dnmt1;
8. C-terminal catalytic domain.

The N-terminal part of Dnmt1 contains several domains that regulate the activity of the catalytic domain. Such a structural organization supposes that the Dnmt1 gene has arisen as fusion of a MTase gene with nonhomologous genes of other DNA-binding proteins [11, 83].

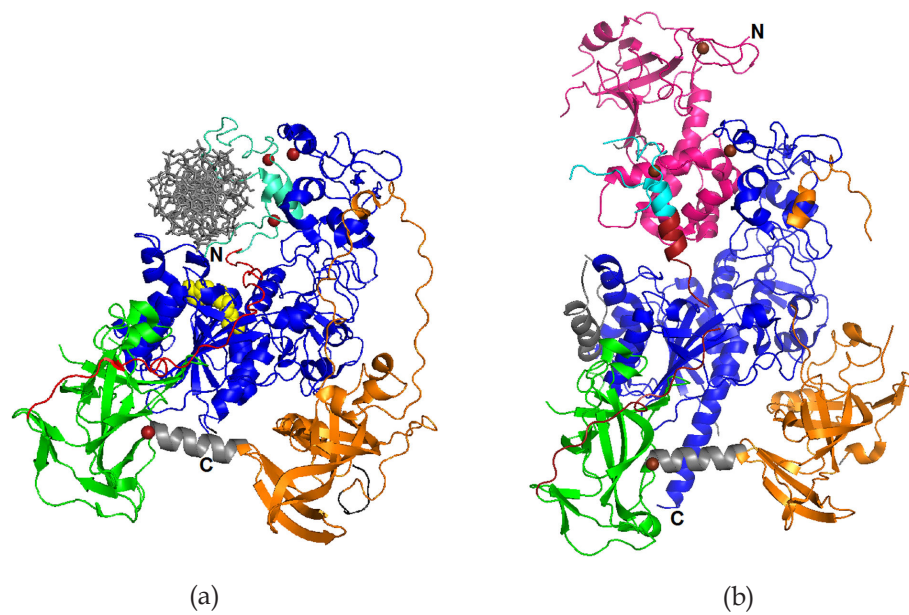


**Figure 8.** Domain architecture of mammalian C5-DNA MTases. The C-terminal domains are marked by a light blue filling. The other domains are signed. Roman numerals indicate the conservative motifs characteristic for C5-DNA MTases in the C-terminal domains. The borders of the Dnmt1 domains are shown according to [9].

Structures of two Dnmt1 complexes with DNA have been determined recently by X-ray crystallography: a fragment of murine Dnmt1 (residues 650–1602) with AdoHcy and a 19 bp DNA duplex containing two unmethylated 5′-CG-3′/3′-GC-5′ sites (PDB code: 3pt6) and a fragment of human Dnmt1 (residues 646–1600) with AdoHcy and the same duplex (PDB code: 3pta) [9]. The catalytic domain of Dnmt1 forms a core of the complex and makes contacts with the DNA on one side and with both BAH domains on the other side (Figure 9). AdoHcy molecule is located in the active center of the catalytic domain. The CXXC and BAH1 domains are located on different sides of the catalytic domain and are connected by a long CXXC–BAH1 linker. The BAH1 and BAH2 domains are located distantly from the bound DNA and are separated from each other by an  $\alpha$ -helical linker. The KG linker is disordered in the crystal. Different Dnmt1 domains are discussed further in this chapter.

It is worth to note that Dnmt1 targeting to replication foci is provided by three types of domains located in its N-terminal part: the PCNA binding domain, the RFTS domain, and the BAH domains [84–86]. Studying of Dnmt1 deletion mutants revealed 3 different DNA binding regions: the residues 1–343, the CXXC domain (residues 613–748), and the catalytic domain (residues 1124–1620) [13]. The catalytic domain binds preferentially monomethylated 5′-CG-3′/3′-GC-5′ sites while the fragment of residues 1–343 binds these sites independently of their methylation status [13]. The CXXC domain is shown to bind primarily dimethylated sites [13]. However, in the crystal structure the CXXC domain interacts with an unmethylated substrate [9].





**Figure 9.** Spatial structure of murine Dnmt1. (a) Dnmt1(650–1602) complexed with AdoHcy and a 19 bp DNA duplex containing two unmethylated 5'-CG-3'/3'-GC-5' sites (PDB code: 3pt6). (b) Dnmt1(291–1620) in apo-form (PDB code: 3av4). The RFTS, CXXC, BAH1, BAH2, and MTase domains are colored in magenta, cyan, green, orange, and blue, respectively.  $Zn^{2+}$  ions are in red and AdoHcy is in yellow, both in space-filling representation. DNA is in grey, in sticks representation. The autoinhibitory linker CXXC–BAH1 is colored in red. The borders of the domains are shown according to [9].

#### 5.1.1. DMAP1-binding domain

The first 120 residues of Dnmt1 can bind the transcription repressor DMAP1. This interaction provides DMAP1 presence in replication foci during the whole S phase of cell cycle. Moreover, Dnmt1 interacts directly with histone deacetylase 2 (HDAC2) during late S phase. Dnmt1, HDAC2, and DMAP1 form a complex *in vivo*. Since the direct interaction between HDAC2 and DMAP1 has not been demonstrated, Dnmt1 is likely to serve as a basis for this complex formation [88].

A special form of Dnmt1, Dnmt1o, is synthesized in oocytes. It lacks the first 118 residues that compose the DMAP1-binding domain. Dnmt1o is accumulated in oocyte cytoplasm and is transferred into nuclei of the 8-cell stage embryo where it is likely to be responsible for maintaining the methylation patterns of imprinted genes [82]. Dnmt1o is replaced by the regular Dnmt1 after implantation of the blastocyst. Homozygous mutant mice containing Dnmt1o in all somatic cells show a normal phenotype and have a normal level of genome methylation. This fact confirms the ability of Dnmt1o to perform all Dnmt1 functions. However, the Dnmt1o amounts and the corresponding enzymatic activity are much higher than the ones of Dnmt1. In heterozygous embryonic stem cells, the expression levels of Dnmt1o and Dnmt1 are the same. Though, in adult mice the Dnmt1o amount is 5 times higher than the Dnmt1 amount. So Dnmt1o seems to be more stable than Dnmt1 [89]. The DMAP1-bind-



ing domain is likely to decrease Dnmt1 stability *in vivo* and thus could be involved in Dnmt1 degradation.

#### 5.1.2. PCNA-binding domain

The Dnmt1 residues 163–174 are responsible for binding of PCNA (proliferating cell nuclear antigen, also known as processivity factor for DNA polymerase  $\delta$ ). Dnmt1 relocates to DNA replication foci when the cell enters S phase [84, 90]. Its binding with PCNA is observed in the regions of newly synthesized DNA in intact cells. This binding does not influence the MTase activity of Dnmt1 [85].

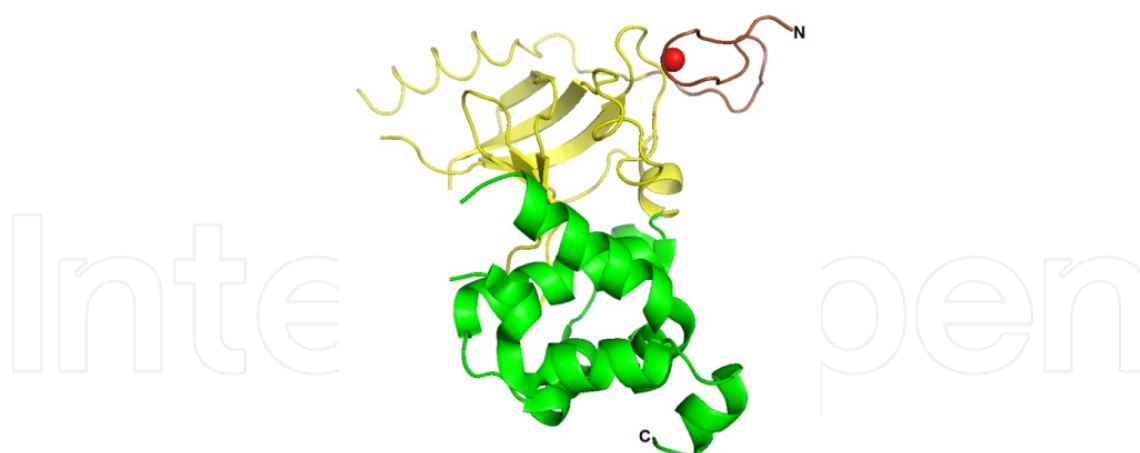
In mammalian cells, newly replicated DNA is rapidly packaged into nucleosomes to which histone H1 is added further [91, 92]. Histone H1 has a high affinity to methylated DNA regardless its nucleotide sequence [93] and can suppress the Dnmt1 enzymatic activity [94]. Therefore, the maintenance methylation should be performed before DNA is packaged in nucleosomes. Dnmt1 binding with PCNA probably underlies a special mechanism required for coordination of these processes in a cell [85].

#### 5.1.3. RFTS domain

Besides the presence in DNA replication foci during S phase (provided by the PCNA-binding domain) [84, 90], Dnmt1 is also associated with chromatin (mainly heterochromatin) from late S phase until early G1 phase [95]. This association is provided by the RFTS domain (replication foci targeting sequence) and does not depend on the methylation patterns specific for heterochromatin or on the histone binding proteins. Moreover, the association with chromatin does not depend on DNA replication, since it takes place in G2 phase and M phase *de novo* [95].

The RFTS domain inhibits the Dnmt1 binding with both free DNA and nucleosomal DNA. It functions as an intrinsic competitive inhibitor of Dnmt1 and can decrease its enzymatic activity up to 600 times [96]. The RFTS domain also inhibits the CXXC domain binding with nucleosomal DNA. The inhibition is observed for the isolated RFTS and CXXC domains as well as for the two domains in a single polypeptide chain. However, a deletion mutant containing both the CXXC domain and the catalytic domain is able to bind polynucleosomes in the presence of the isolated RFTS domain. Thus, the simultaneous presence of the two DNA binding domains seems to make the complex relatively resistant to exclusion of DNA by the RFTS domain [96].

The RFTS domain contains a  $\text{Zn}^{2+}$ -binding motif followed by a  $\beta$ -barrel and an  $\alpha$ -helical bundle (Figure 10). Hydrophobic interactions of the RFTS domains provide Dnmt1 dimerization, although its functional importance is still unknown [97]. In murine Dnmt1(291–1620), the negatively charged RFTS domain penetrates deeply into the positively charged DNA-binding center of the catalytic domain and forms several hydrogen bonds inside it. Such a structural organization could explain the mechanism of DNA displacement from the catalytic center by means of competition with the RFTS domain [87].



**Figure 10.** Spatial structure of the RFTS domain from human Dnmt1 (PDB code: 3epz).  $\beta$ -Barrel is colored in yellow,  $\text{Zn}^{2+}$ -binding motif – in brown,  $\alpha$ -helical bundle – in green.  $\text{Zn}^{2+}$  ion is shown as a red sphere.

#### 5.1.4. CXXC domain

The cysteine-rich  $\text{Zn}^{2+}$ -binding domain contains several cysteine residues organized in CXXC motifs which provide a name to the domain. The CXXC domain of Dnmt1 is similar to cysteine-rich domains of other chromatin-associated proteins such as the MeCP2 protein, the CG binding protein (CGBP), the histone MTase MLL, and the histone demethylases JHDM1A and JHDM1B. This domain is shown to bind unmethylated 5'-CG-3'/3'-GC-5' sites *in vitro* in the case of MBD1, MLL, CGBP, JHDM1B, and Dnmt1 proteins [3]. The CXXC domain of Dnmt1 is crescent-shaped and contains 8 conservative catalytically important Cys residues [98]. These residues are clustered in two groups and bind two  $\text{Zn}^{2+}$  ions. In the structure of murine Dnmt1(650–1602) complexed with AdoHcy and 19 bp DNA (PDB code: 3pt6), all the specific contacts with DNA are formed by the CXXC domain which interacts with both major and minor grooves [9]. A loop region of the CXXC domain (Arg684-Ser685-Lys686-Gln687) penetrates into the major groove and forms contacts with heterocyclic bases and phosphate groups. The guanine bases of the 5'-CG-3' dinucleotide are recognized by the side chains of Lys686 and Gln687, whereas the cytosine bases are recognized by the backbone interactions of Ser685 and Lys686. Salt bridges are formed between the Arg side chains and the DNA backbone.

The CXXC domain is known to bind specifically unmethylated 5'-CG-3'/3'-GC-5' sites [98, 99]. The structural data confirm this type of specificity: a methyl group presence at the cytosine C5 atom would result in steric clashes between the DNA and the protein atoms [9]. The CXXC domain seems to bind newly synthesized unmethylated sites after DNA replication which would protect them from *de novo* methylation.

As mentioned above, Dnmt1 is a maintenance MTase that modifies mainly monomethylated sites. A deletion of the CXXC domain and the part of the CXXC–BAH1 linker results in a 7 times decrease of the Dnmt1 affinity to monomethylated DNA relatively to unmethylated

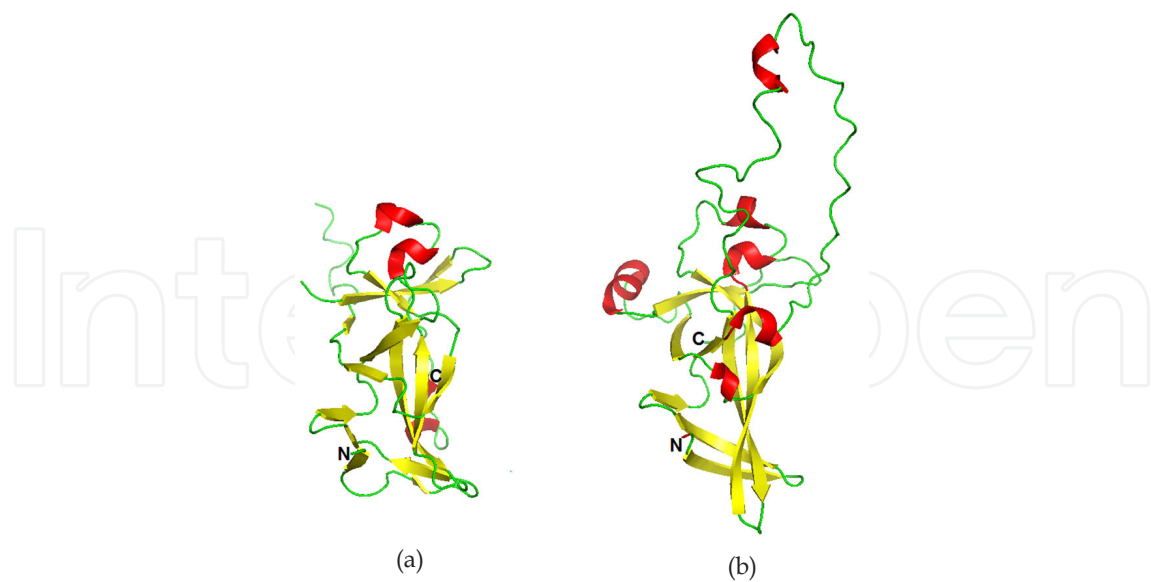
one. The same effect is observed after a substitution of two residues (K686A/Q687A) in the CXXC domain that form contacts with the guanine bases in the recognition site [9].

Addition of dimethylated DNA stimulates murine Dnmt1 to methylate unmodified sites. Such an allosteric activation of Dnmt1 results in lowering its specificity [13, 29, 34, 100]. This effect depends on the presence of  $\text{Zn}^{2+}$  ions and seems to be provided by the binding of the Dnmt1 residues 613–748 with dimethylated DNA [13]. This Dnmt1 region includes the CXXC domain. However, it remains unclear how the CXXC domain could bind dimethylated DNA.

#### 5.1.5. BAH domains

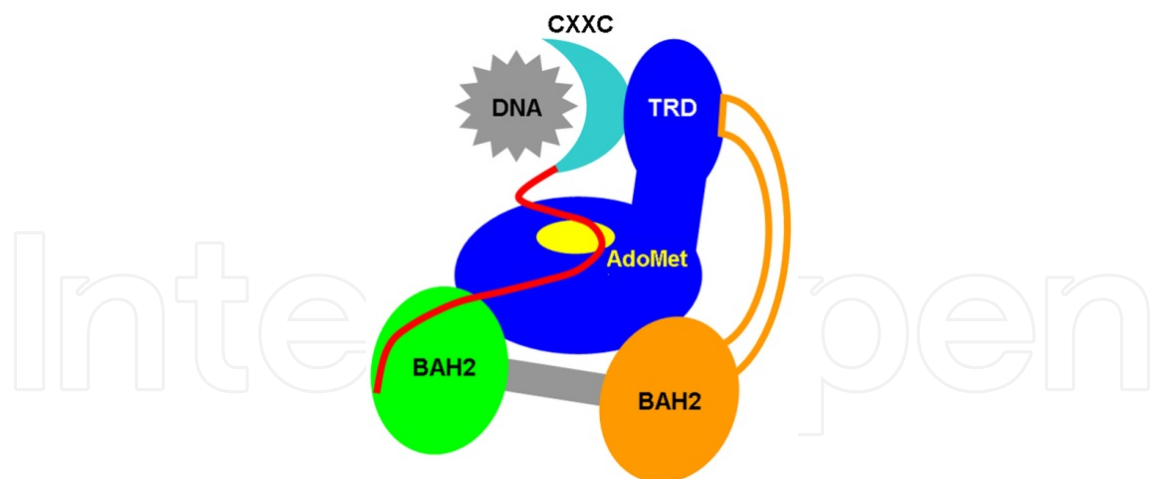
The BAH1 and BAH2 domains (bromo-adjacent homology) are the parts of the so-called PBHD domain (polybromo homology domain). This domain is present in some transcription regulators and is supposed to participate in protein–protein interactions that lead to gene repression. The BAH1 and BAH2 domains in Dnmt1 molecule are connected by an  $\alpha$ -helix and are arranged in a dumbbell shape (Figure 11). Three Cys and one His residues coordinate a  $\text{Zn}^{2+}$  ion that keeps the BAH1 domain near the linker  $\alpha$ -helix. Despite low similarity of the primary structures, the both BAH domains have the same fold (Figure 11): the N-terminal subdomain consists of three antiparallel  $\beta$ -strands, the following subdomain consists of five antiparallel  $\beta$ -strands. Some smaller  $\beta$ -strands and loops which are located further are not homologous in different BAH domains.

Both BAH domains are physically associated with the MTase domain. Seven  $\beta$ -strands of the catalytic subdomain and two  $\beta$ -strands of the BAH1 domain form a common  $\beta$ -sheet. The BAH2 domain has a long loop (BAH2–TRD loop) which interacts with the TRD region of the MTase domain (Figure 11). Perhaps, this interaction prevents the TRD binding to DNA in the complex of murine Dnmt1(650–1602) with AdoHcy and 19 bp DNA duplex (PDB code: 3pt6) (Figure 9). The BAH1 and BAH2 domains have large solvent-accessible surfaces and thus could serve as platforms for interaction with other proteins.



**Figure 11.** a) BAH1 and (b) BAH2 domains of murine Dnmt1 (PDB code: 3pt6) colored according to their secondary structure.  $\alpha$ -Helices are in red,  $\beta$ -strands are in yellow.

The structure of Dnmt1(650–1602) complex with 19 bp DNA duplex and AdoHcy (PDB-code 3pt6) suggests an autoinhibition mechanism of Dnmt1: the CXXC domain binding with DNA results in DNA removal from the active center. The negatively charged CXXC–BAH1 linker is located between the DNA and the active center and thus prevents DNA entrance into the catalytic pocket (Figure 12). In addition, the BAH2–TRD loop fixes the TRD apart from DNA preventing its interaction with the major groove [9].



**Figure 12.** A model of the Dnmt1 autoinhibition mechanism in maintenance DNA methylation. The domains are colored as in Figure 9. The CXXC domain and the autoinhibitory linker (in red) close the active center of the enzyme. Additionally, the BAH2–TRD loop (in orange) keeps the TRD away from the DNA.

#### 5.1.6. KG linker

The peptides containing the  $(\text{LysX})_n\text{Lys}$  sequences (where X = Gly, Ala or Lys) can effectively bind to Z-DNA and stabilize it even at low NaCl concentrations (10–150 mM) and physio-

logical pH. Such peptides can also induce B-DNA transition into Z-DNA. The efficacy of this process grows as the number of repeats ( $n$ ) increases [101, 102]. The sequences  $(\text{LysX})_n\text{Lys}$  are found in different proteins of plants, animals and unicellular eukaryotes. For example, they are present in the linkers which connect N- and C-terminal parts in eukaryotic MTases [102-104]. The number of repeats ( $n$ ) varies from 5 to 7 in this case. The sequences of the linker and the preceding protein region are highly conservative. Thus, the linker is supposed to have an important but still unknown function.

The most effective DNA transition into Z-form is observed for the sequences with alternating purine and pyrimidine nucleosides, including  $5'-(\text{CG})_n-3'$  and especially when the cytosine residues are methylated at the C5 atoms. Additionally, the transition into Z-form is stimulated by high ionic strength in solution and by DNA negative supercoiling [105]. The KG linker probably promotes Dnmt1 binding to  $5'-(\text{CG})_n-3'$  islands in Z-form. However, Dnmt1 is not able to methylate Z-DNA [106]. Perhaps, the KG linker participates in Dnmt1 targeting to the regions located behind a replication fork (due to their negative supercoiling when DNA polymerase has just passed) [3].

## 5.2. Functional domains of mammalian Dnmt3 family

Dnmt3 family includes C5-DNA MTases Dnmt3a and Dnmt3b, which are considered *de novo* MTases, and Dnmt3L – a catalytically inactive protein. There are also different isoforms of the proteins in the Dnmt3 family. Dnmt3a and Dnmt3b are expressed in embryonic cells and during gametogenesis [107]. Knockout of the corresponding genes suppresses *de novo* methylation. Mouse embryos where the Dnmt3b gene is knocked out die *in utero* while mouse embryos lacking Dnmt3a gene die soon after birth [108]. Dnmt3b methylates microsatellite repeats. In humans, Dnmt3b point mutations decreasing its enzymatic activity lead to a severe disease – ICF syndrome (immunodeficiency, centromere instability, facial abnormalities syndrome) [108-110].

Like Dnmt1, the Dnmt3 family enzymes consist of the N-terminal regulatory part and the C-terminal part containing the conservative motifs typical for C5-DNA MTases (Figure 8). The catalytic domains of Dnmt1 and Dnmt3 enzymes are homologous while the N-terminal parts share no similarity. Thus, these families seem to have been evolved from different prokaryotic predecessors [111, 112]. The intramolecular interactions between the N-terminal and the C-terminal parts are absent in Dnmt3a and Dnmt3b, in contrast to Dnmt1 [12].

The N-terminal parts of Dnmt3a and Dnmt3b contain two domains:

1. cysteine-rich domain ADD (ATRX–Dnmt3–Dnmt3L), also called PHD domain (plant homeodomain);
2. PWWP domain.

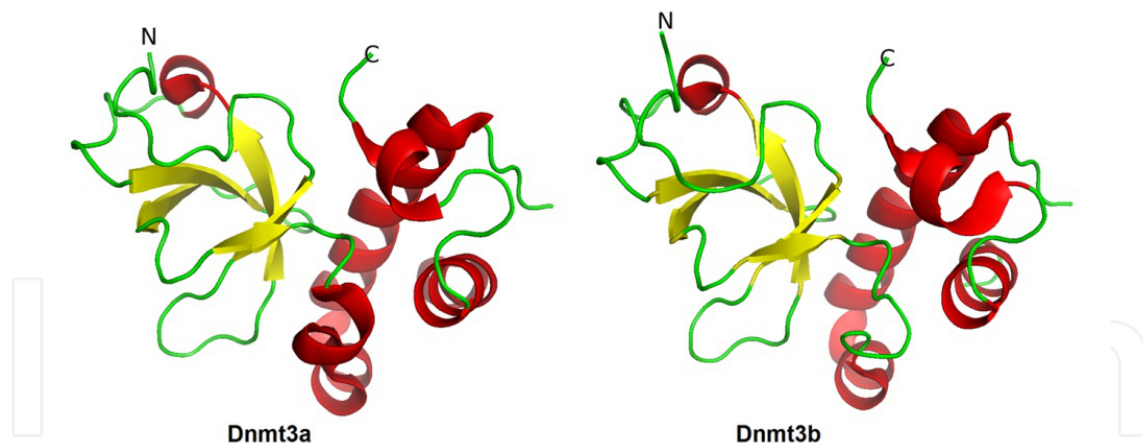
There is no PWWP domain in Dnmt3L. Moreover, Dnmt3L lacks some catalytic residues and the MTase motifs IX and X in the C-terminal domain. Dnmt3L functions as a stimulator of Dnmt3a and Dnmt3b enzymatic activity. Dnmt3L knockout mice are viable but the males are sterile while the females do not produce viable offspring [113-116].



### 5.2.1. PWWP domain

Dnmt3a and Dnmt3b contain a PWWP domain as well as some other chromatin-associated proteins. The PWWP domain is named after a conserved ProTrpTrpPro motif, though the first Pro is substituted with Ser in Dnmt3a and Dnmt3b. The second part of the motif is always the same, TrpPro [117]. PWWP domain along with chromo domain, Tudor and MBT domains belongs to the Tudor domain “Royal family” [118, 119]. The members of this family are shown to bind modified lysine residues of histones [119]. The PWWP domain consists of 100–130 amino acid residues. Its structure includes 3 motifs: a  $\beta$ -barrel core, an insertion between the second and the third  $\beta$ -strands, and a C-terminal  $\alpha$ -helical bundle (Figure 13). Three aromatic residues form a cleft in the center of the  $\beta$ -barrel that is a distinctive feature of the “Royal family”. The insertion motif varies in length and secondary structure among the different PWWP domains. The C-terminal  $\alpha$ -helical bundle can contain from 1 to 5  $\alpha$ -helices [120].

The PWWP domain of Dnmt3b has a positively charged surface with an approximate area of  $45 \times 32 \text{ \AA}^2$  and can bind DNA nonspecifically [117, 121]. The PWWP domain binds 30 bp duplexes with unmethylated, mono-, and dimethylated 5'-CG-3'/3'-GC-5' sites. However, it can also bind a nonspecific duplex of the same length with the same efficiency [117]. Deletion of the PWWP domain does not influence the Dnmt3b methylation efficiency *in vitro* [117]. On the contrary, the PWWP domain of Dnmt3a is almost unable to bind DNA [121].



**Figure 13.** Spatial structures of the PWWP domains from human Dnmt3a (PDB code: 3llr) and Dnmt3b colored according to their secondary structure (PDB code: 3qkj).  $\alpha$ -Helices are in red,  $\beta$ -strands are in yellow.

The PWWP domains of Dnmt3a and Dnmt3b are necessary for targeting these MTases to pericentromeric heterochromatin [121-123]. Deletions in the PWWP domain change the protein distribution in a nucleus and therefore result in its disability to methylate satellite repeats. However, such mutants are catalytically active since they are able to methylate DNA in other regions [121]. In humans, S282P point mutation in the PWWP domain of Dnmt3b

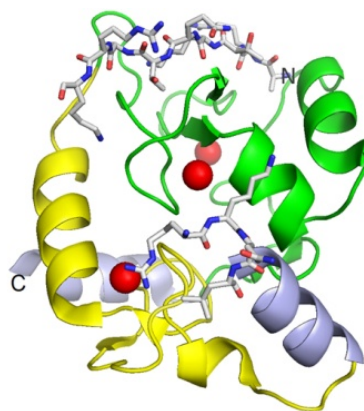


causes the ICF syndrome [124]. In this case, the disease is likely to be caused by the enzyme improper distribution in nuclei rather than by its insufficient catalytic activity.

The PWWP domain of Dnmt3a specifically binds trimethylated Lys36 of H3 histone (H3K36m3) that increases the Dnmt3a ability to methylate nucleosomal DNA [125]. Distribution of methylated sequences in a genome correlates with the presence of H3K36m3 [126, 127]. DNA methylation and H3K36m3 serve as marks for histone deacetylation and the following gene suppression [128]. There is no crystal structure of Dnmt3a or Dnmt3b PWWP domain in complex with a histone protein or a peptide. The resolved spatial structures have a cleft that is supposed to bind methylated lysine residues. This cleft in the crystal contains a molecule of Bis-tris buffer (bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane) that is situated similar to the lysine in complexes of other PWWP domains with histone peptides containing di- or trimethylated lysine residues [120].

### 5.2.2. ADD domain

ADD domain is found only in the following proteins: ATRX (alpha thalassemia/mental retardation syndrome X-linked), Dnmt3a, Dnmt3b, and Dnmt3L. Thus, the domain is called ADD (ATRX–Dnmt3–Dnmt3L) [129]. To date, the crystal structures of ADD domain from Dnmt3a, Dnmt3L, and ATRX proteins are resolved. The structural organization of the domain remains the same in all the cases: it contains two C4-type zinc fingers [129–131]. One of them is similar to GATA binding protein 1 (GATA1) and the other one – to plant homeodomain (PHD).



**Figure 14.** Structure of a fusion protein containing the ADD domain and a fragment of histone H3 (residues 1–9, shown in sticks representation), PDB-code: 3a1b. GATA1-like finger is in yellow, PHD finger is in green.  $\text{Zn}^{2+}$  ions are shown as red spheres. The disordered linker peptide that connects the C-terminus of the histone H3 fragment and the N-terminus of ADD is not shown.

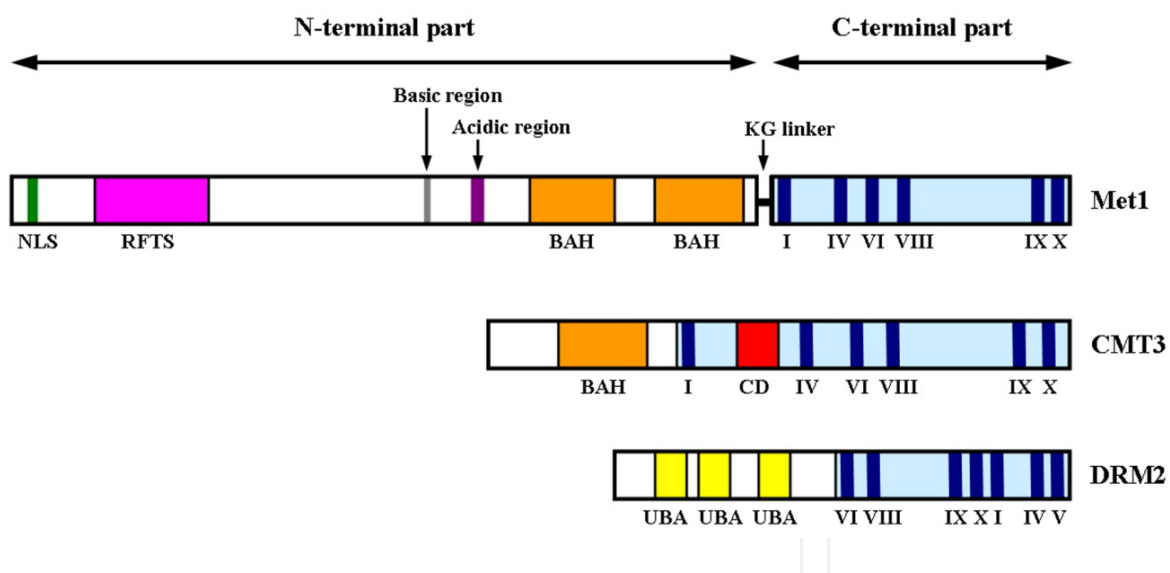
The ADD domain contains many cysteine residues which bind  $\text{Zn}^{2+}$  providing thus the Dnmt3a interaction with many other proteins such as transcription factors PU.1, Myc, RP58, histone deacetylase HDAC1, heterochromatin protein HP1, histone MTases SUV39H1, SETDB1 and EZH2, methyl-CG-binding protein Mbd3, and chromatin remodeling factor Brg1 [3]. The functions of most of these interactions remain unclear. Additionally, the ADD

domains of Dnmt3a, Dnmt3b, and Dnmt3L are shown to interact specifically with the N-terminal part of H3 histone when its Lys4 is not modified. This interaction stimulates Dnmt3a to methylate DNA [130-132]. Thus, the ADD domain of C5-DNA MTases can induce DNA methylation in response to specific histone modifications.

### 5.3. Domains of plant DNA methyltransferases

Plant DNA MTases are very diverse but are studied much less than the mammalian DNA MTases up to now [133]. In particular, none of the plant MTases has been crystallized.

DNA MTase Met1 from *Arabidopsis thaliana* is quite similar to mammalian Dnmt1 (Figure 15). The C-terminal domains of these enzymes which are responsible for methylation share 50% identity. Both proteins possess an extended N-terminal part that is connected with the C-terminal domain *via* KG linker. The N-terminal parts of Met1 and Dnmt1 share 24% identity [134]. There are four similar genes encoding Met1 in *A. thaliana* whereas only one gene encodes the mammalian Dnmt1. Genomes of *Daucus carota* and *Zea mays* contain two Met1 homologs. The N-terminal part of Met1 has two BAH domains. These domains seem to serve as a platform for protein–protein interactions which result in inhibition of gene expression thus providing an interconnection of DNA methylation, replication, and transcription regulation [135].



**Figure 15.** Domain architecture of different C5-DNA MTases from *A. thaliana*. Met1 – MTase similar to mammalian Dnmt1, CMT3 – chromomethyltransferase 3, DRM2 – MTase with the circular permutation of conservative motifs. The catalytic domains are marked by a light blue filling. The other domains are signed in the bottom. Roman numerals indicate the conservative motifs of C5-DNA MTases.

Dnmt3 family homologs are found in plants also. They form a DRM family (domains rearranged methyltransferase). The peculiarity of these enzymes is a circular permutation of their conservative motifs: the motifs VI–X are followed by the motifs I–V. The circular permutation of the conservative motifs is also found in some prokaryotic enzymes such as C5-DNA MTase BssHII [136]. Another characteristic of the DRM proteins is their ability to

methyrate 5'-CHG-3' and 5'-CHH-3' sites (H = A, C or T) *de novo* in an RNA-dependent manner. Such DNA methylation is supposed to take place in the presence of short RNAs that guide methylation of homologous DNA [4]. The DRM1 MTase from *Nicotiana tabacum* is shown to avoid cytosine methylation in 5'-CG-3'/3'-GC-5' sites rather than specifically recognize any sequence. A structural basis of such an unusual functioning is not yet clarified [137].

Proteins of the DRM family are found only in flowering plants [4]. The C-terminal domains of DRM proteins share 28% identity with Dnmt3a or Dnmt3b and contain the same catalytic conservative motifs. Cysteine-rich regions of mammalian MTases (the RFTS, the CXXC, and the ADD domains) are found neither in proteins of DRM family nor in other plant homologs of C5-DNA MTases [112]. The N-terminal part of the DRM MTases contains several UBA domains (ubiquitin-associated). The presence of UBA domains seems to provide a link between DNA methylation and ubiquitin-mediated protein degradation. These domains could promote degradation of DRM molecules at specific points of the cell cycle [112]. The UBA domains are not found in any other families of DNA MTases. The UBA domains of DRM2 MTase in *A. thaliana* were experimentally shown to be required for normal RNA-directed DNA methylation. Perhaps, these domains are essential for proper localization of MTases in the cell [138].

C5-DNA MTases of chromomethylase (or chromomethyltransferase, CMT) family modify 5'-CNG-3'/3'-GNC-5' sites (N = A, C, G or T) in plant genomes. These MTases contain chromo domains that were identified as conserved sequences between the II and the IV MTase motifs (Figure 15). The chromo domain (chromatin organization modifier) consists of 50 amino acid residues and contains three  $\beta$ -strands and a perpendicularly packed  $\alpha$ -helix. This folding type belongs to OB class (oligonucleotide/oligosaccharide binding fold) and is considered to be evolutionary very old. The chromo domain is responsible for DNA binding [139]. This domain is originally found in polycomb-group proteins where it is important for the protein association with heterochromatin [140]. Therefore, CMT are thought to modify heterochromatin. The chromo domain is not found in other C5-DNA MTases. The *A. thaliana* genome contains 3 members of the CMT family: CMT1, CMT2, and CMT3. The members of this family are also found in genomes of *Oryza sativa* and *Brassica oleracea*. Function of CMT1 and CMT2 are still unknown while CMT3 seems to methylate 5'-CNG-3'/3'-GNC-5' sites [4]. CMT3 deficiency in *A. thaliana* results in loss of DNA methylation in centromeric regions and also leads to retrotransposon activation [141, 142].

The enzymes from CMT family contain one BAH domain in its N-terminal region in contrast to the MTases from Met1 family that possess two BAH domains.

## 6. Conclusion

Precise time and space coordination of different molecular events underlies development of all living organisms, unicellular as well as multicellular. Synchronization of molecular processes can take place at a transcriptional level (when the same DNA-binding protein regulates expression of several genes) or at a post-translational level (when a multifunctional protein participates in different processes). Multifunctionality of a protein can be based on the presence of several domains in a single polypeptide chain. For example, mammalian Dnmt1, besides the catalytic domain which provides DNA methylation, contains several other domains responsible for Dnmt1 cellular localization, its interaction with other proteins, and regulation of its catalytic domain activity. Among prokaryotes, multidomain proteins are less common. Nevertheless, some bacterial DNA methyltransferases contain additional domains which are responsible for transcription regulation, topoisomerase activity *etc.*

As shown in this review, the structural and functional features of the additional domains in C5-DNA MTases are studied yet insufficiently. On the basis of the existing data, it is impossible to draw a decisive conclusion on the effect of the additional domains onto the methylating activity of C5-DNA MTases. Evidently, a complex research of multifunctional DNA MTases with multidomain organization would be most promising.

## Acknowledgements

The authors would like to thank Mrs. Anna Nazarenko for her technical assistance. The work was supported by the Russian Foundation for Basic Research (grants no. 10-04-01578 and 12-04-32103).

## Author details

A. Yu. Ryazanova<sup>1</sup>, L. A. Abrosimova<sup>2</sup>, T. S. Oretskaya<sup>1,3</sup> and E. A. Kubareva<sup>3\*</sup>

\*Address all correspondence to: [kubareva@belozersky.msu.ru](mailto:kubareva@belozersky.msu.ru)

1 Chemistry Department, Lomonosov Moscow State University, Moscow, Russia

2 Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia

3 Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia

## References

- [1] Jeltsch, A. (2002). Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases. *A European Journal of Chemical Biology Chem Bio Chem*, 3(4), 274-293.
- [2] Gromova, E. S., & Khoroshaev, A. V. (2003). Prokaryotic DNA Methyltransferases: the Structure and the Mechanism of Interaction with DNA. *Molekulyarnaya Biologiya*, 37(2), 300-314.
- [3] Jurkowska, R. Z., Jurkowski, T. P., & Jeltsch, A. (2011). Structure and Function of Mammalian DNA Methyltransferases. *A European Journal of Chemical Biology Chem-BioChem*, 12(2), 206-222.
- [4] Goll, M. G., & Bestor, T. H. (2005). Eukaryotic Cytosine Methyltransferases. *Annual Review of Biochemistry*, 74, 481-514.
- [5] Svedruzic, Z. M. (2008). Mammalian Cytosine DNA Methyltransferase Dnmt1: Enzymatic Mechanism, Novel Mechanism-Based Inhibitors, and RNA-Directed DNA Methylation. *Current Medicinal Chemistry*, 15(1), 92-106.
- [6] Jeltsch, A. (2006). Molecular Enzymology of Mammalian DNA Methyltransferases. *Current Topics in Microbiology and Immunology*, 301, 203-225.
- [7] Cheng, X., & Roberts, R. J. (2001). AdoMet-Dependent Methylation, DNA Methyltransferases and Base Flipping. *Nucleic Acids Research*, 29(18), 3784-3795.
- [8] Cheng, X., & Blumenthal, R. M. (2008). Mammalian DNA Methyltransferases: a Structural Perspective. *Structure*, 16(3), 341-350.
- [9] Song, J., Rechko, O., Bestor, T. H., & Patel, D. J. (2011). Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation. *Science*, 331(6020), 1036-1040.
- [10] Zimmermann, C., Guhl, E., & Graessmann, A. (1997). Mouse DNA Methyltransferase (MTase) Deletion Mutants that Retain the Catalytic Domain Display Neither de Novo nor Maintenance Methylation Activity *in vivo*. *The Journal of Biological Chemistry*, 272(5), 393-405.
- [11] Margot, J. B., Aguirre-Arteta, A. M., Di Giacco, B. V., Pradhan, S., Roberts, R. J., Cardoso, M. C., & Leonhardt, H. (2000). Structure and Function of the Mouse DNA Methyltransferase Gene: Dnmt1 Shows a Tripartite Structure. *Journal of Molecular Biology*, 297(2), 293-300.
- [12] Margot, J. B., Ehrenhofer-Murray, A. E., & Leonhardt, H. (2003). Interactions within the Mammalian DNA Methyltransferase Family. *BMC Molecular Biology*, 4, 7.
- [13] Fatemi, M., Hermann, A., Pradhan, S., & Jeltsch, A. (2001). The Activity of the Murine DNA Methyltransferase Dnmt1 Is Controlled by Interaction of the Catalytic Domain with the N-Terminal Part of the Enzyme Leading to an Allosteric Activation of the Enzyme after Binding to Methylated DNA. *Journal of Molecular Biology*, 309(5), 1189-1199.



- [14] Wu, J. C., & Santi, D. V. (1987). Kinetic and Catalytic Mechanism of HhaI Methyltransferase. *Journal of Biological Chemistry*, 262(10), 4778-4786.
- [15] Klimasauskas, S., Szyperski, T., Serva, S., & Wüthrich, K. (1998). Dynamic Modes of the Flipped-Out Cytosine During HhaI Methyltransferase-DNA Interactions in Solution. *The EMBO Journal*, 17(1), 317-324.
- [16] Wang, P., Brank, A. S., Banavali, N. K., Nicklaus, M. C., Marquez, V. E., Christman, J. K., & MacKerell, A. D. (2000). Use of Oligodeoxyribonucleotides with Conformationally Constrained Abasic Sugar Targets to Probe the Mechanism of Base Flipping by HhaI DNA (Cytosine C5)-Methyltransferase. *Journal of the American Chemical Society*, 122(50), 12422-12434.
- [17] Lindstrom, W. M. Jr., Flynn, J., & Reich, N. O. (2000). Reconciling Structure and Function in HhaI DNA Cytosine-C-5 Methyltransferase. *Journal of Biological Chemistry*, 275(7), 4912-4919.
- [18] Vilkaitis, G., Merkiene, E., Serva, S., Weinhold, E., & Klimašauskas, S. (2001). The Mechanism of DNA Cytosine-5 Methylation. Kinetic and Mutational Dissection of HhaI Methyltransferase. *Journal of Biological Chemistry*, 276(24), 20924-20934.
- [19] Svedruzic, Z. M., & Reich, N. O. (2004). The Mechanism of Target Base Attack in DNA Cytosine Carbon 5 Methylation. *Biochemistry*, 43(36), 11460-11473.
- [20] Daujotyte, D., Serva, S., Vilkaitis, G., Merkiene, E., Venclovas, C., & Klimasauskas, S. (2004). HhaI DNA Methyltransferase Uses the Protruding Gln237 for Active Flipping of Its Target Cytosine. *Structure*, 12(6), 1047-1055.
- [21] Horton, J. R., Ratner, G., Banavali, N. K., Huang, N., Choi, Y., Maier, M. A., Marquez, V. E., MacKerell, A. D., & Cheng, X. (2004). Caught in the Act: Visualization of an Intermediate in the DNA Base-Flipping Pathway Induced by HhaI Methyltransferase. *Nucleic Acids Research*, 32(13), 3877-3886.
- [22] Merkiene, E., & Klimašauskas, S. (2005). Probing a Rate-Limiting Step by Mutational Perturbation of AdoMet Binding in the HhaI Methyltransferase. *Nucleic Acids Research*, 33(1), 307-315.
- [23] Gerasimaitė, R., Merkienė, E., & Klimašauskas, S. (2011). Direct Observation of Cytosine Flipping and Covalent Catalysis in a DNA Methyltransferase. *Nucleic Acids Research*, 39(9), 3771-3780.
- [24] Estabrook, R. A., Nguyen, T. T., Fera, N., & Reich, N. O. (2009). Coupling Sequence-Specific Recognition to DNA Modification. *Journal of Biological Chemistry*, 284(34), 22690-22696.
- [25] Matje, D. M., Coughlin, D. F., Connolly, B. A., Dahlquist, F. W., & Reich, N. O. (2011). Determinants of Precatalytic Conformational Transitions in the DNA Cytosine Methyltransferase M.HhaI. *Biochemistry*, 50(9), 1465-1473.



- [26] Renbaum, P., & Razin, A. (1992). Mode of Action of the Spiroplasma CpG Methylase M.SssI. *FEBS Letters*, 313(3), 243-247.
- [27] Dryden, D. T. F. (1999). Bacterial DNA Methyltransferases. S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions. In: Cheng X., Blumenthal R.M. (ed.) Singapore: World Scientific Publishing, 283-340.
- [28] O'Gara, M., Roberts, R. J., & Cheng, X. (1996). A Structural Basis for the Preferential Binding of Hemimethylated DNA by HhaI DNA Methyltransferase. *Journal of Molecular Biology*, 263(4), 597-606.
- [29] Christman, J. K., Sheikhnejad, G., Marasco, C. J., & Sufrin, J. R. (1995). Methyl-2'-deoxycytidine in Single-stranded DNA Can Act in Cis to Signal de Novo DNA Methylation. *Proceedings of the National Academy of Science*, 92(16), 7347-7351.
- [30] Tollefsbol, T. O., & Hutchison, C. A. 3rd. (1995). Mammalian DNA (Cytosine-5)-Methyltransferase Expressed in *Escherichia coli*, Purified and Characterized. *Journal of Biological Chemistry*, 270(31), 18543-18550.
- [31] Tollefsbol, T. O., & Hutchison, C. A. 3rd. (1997). Control of Methylation Spreading in Synthetic DNA Sequences by the Murine DNA Methyltransferase. *Journal of Molecular Biology*, 269(4), 494-504.
- [32] Flynn, J., Glickman, J. F., & Reich, N. O. (1996). Murine DNA Cytosine-C5 Methyltransferase: Pre-steady- and Steady-State Kinetic Analysis with Regulatory DNA Sequences. *Biochemistry*, 35(23), 7308-7315.
- [33] Pradhan, S., Bacolla, A., Wells, R. D., & Roberts, R. J. (1999). Recombinant Human DNA (Cytosine-5) Methyltransferase. I. Expression, Purification, and Comparison of de Novo and Maintenance Methylation. *Journal of Biological Chemistry*, 274(46), 33002-33010.
- [34] Fatemi, M., Hermann, A., Gowher, H., & Jeltsch, A. (2002). Dnmt3a and Dnmt1 Functionally Cooperate During de Novo Methylation of DNA. *European Journal of Biochemistry*, 269(20), 4981-4984.
- [35] Hermann, A., Goyal, R., & Jeltsch, A. (2004). The Dnmt1 DNA-(Cytosine-C5)-Methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites. *Journal of Biological Chemistry*, 279(46), 48350-48359.
- [36] Goyal, R., Reinhardt, R., & Jeltsch, A. (2006). Accuracy of DNA Methylation Pattern Preservation by the Dnmt1 Methyltransferase. *Nucleic Acids Research*, 34(4), 1182-1188.
- [37] Gowher, H., & Jeltsch, A. (2001). Enzymatic Properties of Recombinant Dnmt3a DNA Methyltransferase from Mouse: the Enzyme Modifies DNA in a Non-Processive Manner and Also Methylates non-CpG Sites. *Journal of Molecular Biology*, 309(5), 1201-1208.

- [38] Handa, V., & Jeltsch, A. (2005). Profound Flanking Sequence Preference of Dnmt3a and Dnmt3b Mammalian DNA Methyltransferases Shape the Human Epigenome. *Journal of Molecular Biology*, 348(5), 1103-1112.
- [39] Gowher, H., & Jeltsch, A. (2002). Molecular Enzymology of the Catalytic Domains of the Dnmt3a and Dnmt3b DNA Methyltransferases. *Journal of Biological Chemistry*, 277(23), 20409-20414.
- [40] Reither, S., Li, F., Gowher, H., & Jeltsch, A. (2003). Catalytic Mechanism of DNA-(Cytosine-C5)-Methyltransferases Revisited: Covalent Intermediate Formation Is not Essential for Methyl Group Transfer by the Murine Dnmt3a Enzyme. *Journal of Molecular Biology*, 329(4), 675-684.
- [41] Siddique, A. N., Jurkowska, R. Z., Jurkowski, T. P., & Jeltsch, A. (2011). Auto-Methylation of the Mouse DNA-(Cytosine C5)-Methyltransferase Dnmt3a at Its Active Site Cysteine Residue. *The FEBS Journal*, 278(12), 2055-2063.
- [42] Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., & Cheng, X. (2007). Structure of Dnmt3a Bound to Dnmt3L Suggests a Model for de Novo DNA Methylation. *Nature*, 449(7159), 248-251.
- [43] Zhang, Y., Rohde, C., Tierling, S., Jurkowski, T. P., Bock, C., Santacruz, D., Ragozin, S., Reinhardt, R., Groth, M., Walter, J., & Jeltsch, A. (2009). DNA Methylation Analysis of Chromosome 21 Gene Promoters at Single Base Pair and Single Allele Resolution. *PLoS Genetics*, 5(3), e1000438.
- [44] Bitinaite, J., Maneliene, Z., Menkevicius, S., Klimasauskas, S., Butkus, V., & Janulaitis, A. (1992). Alw26I, Eco31I and Esp3I- Type IIS Methyltransferases Modifying Cytosine and Adenine in Complementary Strands of the Target DNA. *Nucleic Acids Research*, 20(19), 4981-4985.
- [45] Bitinaite, J., Mitkaite, G., Dauksaite, V., Jakubauskas, A., Timinskas, A., Vaisvila, R., Lubys, A., & Janulaitis, A. (2002). Evolutionary Relationship of Alw26I, Eco31I and Esp3I, Restriction Endonucleases that Recognise Overlapping Sequences. *Molecular Genetics and Genomics*, 267(5), 664-672.
- [46] Kumar, R., & Rao, D. N. (2011). A Nucleotide Insertion Between Two Adjacent Methyltransferases in *Helicobacter pylori* Results in a Bifunctional DNA Methyltransferase. *Biochemical Journal*, 433(3), 487-495.
- [47] Chan, S. H., Opitz, L., Higgins, L., O'Loane, D., & Xu, S. Y. (2010). Cofactor Requirement of HpyAV Restriction Endonuclease. *PLoS One*, 5(2), e9071.
- [48] Shen, J. C., Rideout, W. M., & Jones, P. A. (1992). High Frequency Mutagenesis by a DNA Methyltransferase. *Cell*, 71(7), 1073-1080.
- [49] Wyszynski, M., Gabbara, S., & Bhagwat, A. S. (1994). Cytosine Deaminations Catalysed by DNA Cytosine Methyltransferases Are Unlikely to Be the Major Cause of Mutational Hot-Spots at Sites of Cytosine Methylation in *E. coli*. *Proceedings of the National Academy of Science*, 91(4), 1574-1578.

- [50] Bandaru, B., Wyszynski, M., & Bhagwat, A. S. (1995). HpaII Methyltransferase Is Mutagenic in *Escherichia coli*. *Journal of Bacteriology*, 177(10), 2950-2952.
- [51] Yang, A. S., Shen, J. C., Zingg, J. M., Mi, S., & Jones, P. A. (1995). HhaI and HpaII DNA Methyltransferases Bind DNA Mismatches, Methylate Uracil and Block DNA Repair. *Nucleic Acids Research*, 23(8), 1380-1387.
- [52] Yebra, M. J., & Bhagwat, A. S. (1995). A Cytosine Methyltransferase Converts 5-Methylcytosine in DNA to Thymine. *Biochemistry*, 34(45), 14752-14757.
- [53] Zingg, J. M., Shen, J. C., Yang, A. S., Rapoport, H., & Jones, P. A. (1996). Methylation Inhibitors Can Increase the Rate of Cytosine Deamination by (Cytosine-5)-DNA Methyltransferase. *Nucleic Acids Research*, 24(16), 3267-3275.
- [54] Bandaru, B., Gopal, J., & Bhagwat, A. S. (1996). Overproduction of DNA Cytosine Methyltransferase Causes Methylation and C-> T Mutations at non-Canonical Sites. *The Journal of Biological Chemistry*, 271(13), 7851-7859.
- [55] Vorob'eva, O. V. (2004). (Cytosine-5)-DNA Methyltransferase SsoII as a Bifunctional Protein: Study of Its Interaction with the Methylation Site and with the Promoter Region of the Genes in the SsoII Restriction-Modification System. PhD thesis. Lomonosov Moscow State University.
- [56] Shen, J. C., Zingg, J. M., Yang, A. S., Schmutte, C., & Jones, P. A. (1995). A Mutant HpaII Methyltransferase Functions as a Mutator Enzyme. *Nucleic Acids Research*, 23(21), 4275-4282.
- [57] Zingg, J. M., Shen, J. C., & Jones, P. A. (1998). Enzyme-Mediated Cytosine Deamination by the Bacterial Methyltransferase M.MspI. *Biochemical Journal*, 332(1), 223-30.
- [58] Selker, E. U. (1990). Premeiotic Instability of Repeated Sequences in *Neurospora crassa*. *Annual Review of Genetics*, 24, 579-613.
- [59] Sharath, A. N., Weinhold, E., & Bhagwat, A. S. (2000). Reviving a Dead Enzyme: Cytosine Deaminations Promoted by an Inactive DNA Methyltransferase and an S-Adenosylmethionine Analogue. *Biochemistry*, 39(47), 14611-14616.
- [60] Svedruzic, Z. M., & Reich, N. O. (2005). DNA Cytosine C5 Methyltransferase Dnmt1: Catalysis-Dependent Release of Allosteric Inhibition. *Biochemistry*, 44(27), 9472-9485.
- [61] Renbaum, P., Abrahamove, D., Fainsod, A., Wilson, G. G., Rottem, S., & Razin, A. (1990). Cloning, Characterization, and Expression in *Escherichia coli* of the Gene Coding for the CpG DNA Methylase from *Spiroplasma* sp. Strain MQ1 (M.SssI). *Nucleic Acids Research*, 18(5), 1145-1152.
- [62] Matsuo, K., Silke, J., Gramatikoff, K., & Schaffner, W. (1994). The CpG-Specific Methylase SssI Has Topoisomerase Activity in the Presence of  $Mg^{2+}$ . *Nucleic Acids Research*, 22(24), 5354-5359.
- [63] Walder, R. Y., Langtimm, C. J., Catterjee, R., & Walder, J. A. (1983). Cloning of the MspI Modification Enzyme. *The Journal of Biological Chemistry*, 258(2), 1235-1241.

- [64] Dubey, A. K., & Bhattacharya, S. K. (1997). Angle and Locus of the Bend Induced by the MspI DNA Methyltransferase in a Sequence-Specific Complex with DNA. *Nucleic Acids Research*, 25(10), 2025-2029.
- [65] Bhattacharya, S. K., & Dubey, A. K. (2002). The N-Terminus of m5C-DNA Methyltransferase MspI Is Involved in Its Topoisomerase Activity. *European Journal of Biochemistry*, 269(10), 2491-2497.
- [66] Nagornykh, M. O., Bogdanova, E. S., Protsenko, A. S., Zakharova, M. V., & Severinov, K. V. (2008). Regulation of Gene Expression in Type II Restriction-Modification Systems. *Genetika*, 44, 1-10.
- [67] Protsenko, A., Zakharova, M., Nagornykh, M., Solonin, A., & Severinov, K. (2009). Transcription Regulation of Restriction-Modification System Ecl18kI. *Nucleic Acids Research*, 37(16), 5322-5330.
- [68] Stankevicius, K., & Timinskas, A. (1996). Expression Autoregulation of the Eco47II Methyltransferase Gene. *Biologija*, 0, 54-56.
- [69] Som, S., & Friedman, S. (1994). Regulation of EcoRII Methyltransferase: Effect of Mutations on Gene Expression and in Vitro Binding to the Promoter Region. *Nucleic Acids Research*, 22(24), 5347-5353.
- [70] Som, S., & Friedman, S. (1993). Autogenous Regulation of the EcoRII Methylase Gene at the Transcriptional Level: Effect of 5-Azacytidine. *The EMBO Journal*, 12(11), 4297-4303.
- [71] Som, S., & Friedman, S. (1997). Characterization of the Intergenic Region which Regulates the MspI Restriction-Modification System. *Journal of Bacteriology*, 179(3), 964-967.
- [72] Butler, D., & Fitzgerald, G. F. (2001). Transcriptional Analysis and Regulation of Expression of the ScrFI Restriction-Modification System of *Lactococcus lactis* subsp. *cremoris* UC503. *Journal of Bacteriology*, 183(15), 4668-4673.
- [73] O'Driscoll, J., Glynn, F., Cahalane, O., O'Connell-Motherway, M., Fitzgerald, G. F., & Van Sinderen, D. (2004). Lactococcal Plasmid pNP40 Encodes a Novel, Temperature-Sensitive Restriction-Modification System. *Applied and Environmental Microbiology*, 70(9), 5546-5556.
- [74] O'Driscoll, J., Fitzgerald, G. F., & van Sinderen, D. (2005). A Dichotomous Epigenetic Mechanism Governs Expression of the LlaJI Restriction/Modification System. *Molecular Microbiology*, 57(6), 1532-1544.
- [75] Karyagina, A., Shilov, I., Tashlitskii, V., Khodoun, M., Vasil'ev, S., Lau, P. C. K., & Nikolskaya, I. (1997). Specific Binding of SsoII DNA Methyltransferase to Its Promoter Region Provides the Regulation of SsoII Restriction-Modification Gene Expression. *Nucleic Acids Research*, 25(11), 2114-2120.



- [76] Shilov, I., Tashlitsky, V., Khodoun, M., Vasil'ev, S., Alekseev, Y., Kuzubov, A., Kubareva, E., & Karyagina, A. (1998). DNA-methyltransferase SsoII Interaction with Own Promoter Region Binding Site. *Nucleic Acids Research*, 26(11), 2659-2664.
- [77] Vorob'eva, O. V., Vasil'ev, S. A., Kariagina, A. S., Oretskaia, T. S., & Kubareva, E. A. (2000). Analysis of Contacts between DNA and Protein in a Complex of SsoII Methyltransferase-Promoter Region of the Gene for the SsoII Restriction-Modification System. *Molekulyarnaya Biologiya*, 34(6), 1074-1080.
- [78] Karyagina, A. S., Alexeevski, A. V., Golovin, A. V., Spirin, S. A., Vorob'eva, O. V., & Kubareva, E. A. (2003). Computer Modeling of Complexes of (C5-Cytosine)-DNA Methyltransferase SsoII with Target and Regulatory DNAs. *Biophysics*, 48(1), S45-S55.
- [79] Fedotova, E. A., Protsenko, A. S., Zakharova, M. V., Lavrova, N. V., Alekseevsky, A. V., Oretskaya, T. S., Karyagina, A. S., Solonin, A. S., & Kubareva, E. A. (2009). SsoII-like DNA-methyltransferase Ecl18kI: Interaction between Regulatory and Methylating Functions. *Biochemistry*, 74(1), 85-91.
- [80] Fedotova, E. A. (2006). Peculiarities of Gene Expression Regulation in the SsoII Restriction-Modification System. PhD thesis. Lomonosov Moscow State University.
- [81] Ryazanova, A. Yu. (2012). Structural and Functional Characterization of (Cytosine-5)-DNA Methyltransferase SsoII and Its Complexes with DNA Ligands. PhD thesis. Lomonosov Moscow State University.
- [82] Cardoso, M. C., & Leonhardt, H. (1999). DNA Methyltransferase Is Actively Retained in the Cytoplasm During Early Development. *The Journal of Cell Biology*, 147(1), 25-32.
- [83] Bestor, T. H. (1992). Activation of Mammalian DNA Methyltransferase by Cleavage of a Zn Binding Regulatory Domain. *The EMBO Journal*, 11(7), 2611-2617.
- [84] Leonhardt, H., Page, A. W., Weier, H. U., & Bestor, T. H. (1992). A Targeting Sequence Directs DNA Methyltransferase to Sites of DNA Replication in Mammalian Nuclei. *Cell*, 71(5), 865-873.
- [85] Chuang, L. S., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G., & Li, B. F. (1997). Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1. *Science*, 277(5334), 1996-2000.
- [86] Liu, Y., Oakeley, E. J., Sun, L., & Jost, J. P. (1998). Multiple Domains Are Involved in the Targeting of the Mouse DNA Methyltransferase to the DNA Replication Foci. *Nucleic Acids Research*, 26(4), 1038-1045.
- [87] Takeshita, K., Suetake, I., Yamashita, E., Suga, M., Narita, H., Nakagawa, A., & Tajima, S. (2011). Structural Insight into Maintenance Methylation by Mouse DNA Methyltransferase 1 (Dnmt1). *Proceedings of the National Academy of Science*, 108(22), 9055-9059.
- [88] Rountree, M. R., Bachman, K. E., & Baylin, S. B. (2000). Dnmt1 binds HDAC2 and a New Co-Repressor, DMAP1, to Form a Complex at Replication Foci. *Nature Genetics*, 25(3), 269-277.



- [89] Ding, F., & Chaillet, J. R. (2002). In Vivo Stabilization of the Dnmt1 (Cytosine-5)-Methyltransferase Protein. *Proceedings of the National Academy of Science*, 99(23), 14861-14866.
- [90] Margot, J. B., Cardoso, M. C., & Leonhardt, H. (2001). Mammalian DNA Methyltransferases Show Different Subnuclear Distributions. *Journal of Cellular Biochemistry*, 83(3), 373-379.
- [91] Krude, T. (1995). Chromatin. Nucleosome Assembly During DNA Replication. *Current Biology*, 5(11), 1232-1234.
- [92] Gasser, R., Koller, T., & Sogo, J. M. (1996). The Stability of Nucleosomes at the Replication Fork. *Journal of Molecular Biology*, 258(2), 224-239.
- [93] McArthur, M., & Thomas, J. O. (1996). A Preference of Histone H1 for Methylated DNA. *The EMBO Journal*, 15(7), 1705-1714.
- [94] Carotti, D., Funiciello, S., Lavia, P., Caiafa, P., & Strom, R. (1996). Different Effects of Histone H1 on de Novo DNA Methylation in Vitro Depend on Both the DNA Base Composition and the DNA Methyltransferase. *Biochemistry*, 35(36), 11660-11667.
- [95] Easwaran, H. P., Schermelleh, L., Leonhardt, H., & Cardoso, M. C. (2004). Replication-Independent Chromatin Loading of Dnmt1 During G2 and M Phases. *EMBO Reports*, 5(12), 1181-1186.
- [96] Syeda, F., Fagan, R. L., Wean, M., Avvakumov, G. V., Walker, J. R., Xue, S., Dhe-Paganon, S., & Brenner, C. (2011). The Replication Focus Targeting Sequence (RFTS) Domain Is a DNA-competitive Inhibitor of Dnmt1. *Journal of Biological Chemistry*, 286(17), 15344-15351.
- [97] Fellingner, K., Rothbauer, U., Felle, M., Längst, G., & Leonhardt, H. (2009). Dimerization of DNA Methyltransferase 1 Is Mediated by Its Regulatory Domain. *Journal of Cellular Biochemistry*, 106(4), 521-528.
- [98] Pradhan, M., Estève, P. O., Chin, H. G., Samaranayake, M., Kim, G. D., & Pradhan, S. (2008). CXXC Domain of Human Dnmt1 Is Essential for Enzymatic Activity. *Biochemistry*, 47(38), 10000-10009.
- [99] Lee, J. H., Voo, K. S., & Skalnik, D. G. (2001). Identification and Characterization of the DNA Binding Domain of CpG-Binding Protein. *Journal of Biological Chemistry*, 276(48), 44669-44676.
- [100] Bacolla, A., Pradhan, S., Roberts, R. J., & Wells, R. D. (1999). Recombinant Human DNA (Cytosine-5) Methyltransferase. II. Steady-State Kinetics Reveal Allosteric Activation by Methylated DNA. *Journal of Biological Chemistry*, 274(46), 33011-33019.
- [101] Takeuchi, H., Hanamura, N., Hayasaka, H., & Harada, I. (1991). B-Z Transition of Poly(dG-m<sup>5</sup>dC) Induced by Binding of Lys-Containing Peptides. *FEBS Letters*, 279(2), 253-255.

- [102] Takeuchi, H., Hanamura, N., & Harada, I. (1994). Structural Specificity of Peptides in Z-DNA Formation and Energetics of the Peptide-Induced B-Z Transition of Poly(dG-m<sup>5</sup>dC). *Journal of Molecular Biology*, 236(2), 610-617.
- [103] Krzyzaniak, A., Siatecka, M., Szyk, A., Mucha, P., Rekowski, P., Kupryszewski, G., & Barciszewski, J. (2000). Specific Induction of Z-DNA Conformation by a Nuclear Localization Signal Peptide of Lupin Glutaminyl tRNA Synthetase. *Molecular Biology Reports*, 27(1), 51-54.
- [104] Kim, Y. G., Park, H. J., Kim, K. K., Lowenhaupt, K., & Rich, A. (2006). A Peptide with Alternating Lysines Can Act as a Highly Specific Z-DNA Binding Domain. *Nucleic Acids Research*, 34(17), 4937-4942.
- [105] Rich, A., Nordheim, A., & Wang, A. H. (1984). The Chemistry and Biology of Left-Handed Z-DNA. *Annual Review of Biochemistry*, 53-791.
- [106] Bestor, T. (1987). Supercoiling-Dependent Sequence Specificity of Mammalian DNA Methyltransferase. *Nucleic Acids Research*, 15(9), 3835-3843.
- [107] Okano, M., Xie, S., & Li, E. (1998). Cloning and Characterization of a Family of Novel Mammalian DNA (Cytosine-5) Methyltransferases. *Nature Genetics*, 19(3), 219-220.
- [108] Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for de Novo Methylation and Mammalian Development. *Cell*, 99(3), 247-257.
- [109] Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M., & Gartler, S. M. (1999). The DNMT3B DNA Methyltransferase Gene Is Mutated in the ICF Immunodeficiency Syndrome. *Proceedings of the National Academy of Science*, 96(25), 14412-14417.
- [110] Xu, G. L., Bestor, T. H., Bourc'his, D., Hsieh, C. L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J. J., & Viegas-Péquignot, E. (1999). Chromosome Instability and Immunodeficiency Syndrome Caused by Mutations in a DNA Methyltransferase Gene. *Nature*, 402(6758), 187-191.
- [111] Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W. W., Okumura, K., & Li, E. (1999). Cloning, Expression and Chromosome Locations of the Human DNMT3 Gene Family. *Gene*, 236(1), 87-95.
- [112] Cao, X., Springer, N. M., Muszynski, M. G., Phillips, R. L., Kaeppler, S., & Jacobsen, S. E. (2000). Conserved Plant Genes with Similarity to Mammalian de Novo DNA Methyltransferases. *Proceedings of the National Academy of Science*, 97(9), 4979-4984.
- [113] Hata, K., Okano, M., Lei, H., & Li, E. (2002). Dnmt3L Cooperates with the Dnmt3 Family of de Novo DNA Methyltransferases to Establish Maternal Imprints in Mice. *Development*, 129(8), 1983-1993.
- [114] Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the Establishment of Maternal Genomic Imprints. *Science*, 294(5551), 2536-2539.

- [115] Bourc'his, D., & Bestor, T. H. (2004). Meiotic Catastrophe and Retrotransposon Reactivation in Male Germ Cells Lacking Dnmt3L. *Nature*, 431(7004), 96-99.
- [116] Webster, K. E., O'Bryan, M. K., Fletcher, S., Crewther, P. E., Aapola, U., Craig, J., Harrison, D. K., Aung, H., Phutikanit, N., Lyle, R., Meachem, S. J., Antonarakis, S. E., de Kretser, D. M., Hedger, M. P., Peterson, P., Carroll, B. J., & Scott, H. S. (2005). Meiotic and Epigenetic Defects in Dnmt3L-Knockout Mouse Spermatogenesis. *Proceedings of the National Academy of Science*, 102(11), 4068-4073.
- [117] Qiu, C., Sawada, K., Zhang, X., & Cheng, X. (2002). The PWWP Domain of Mammalian DNA Methyltransferase Dnmt3b Defines a New Family of DNA-Binding Folds. *Nature Structural & Molecular Biology*, 9(3), 217-224.
- [118] Maurer-Stroh, S., Dickens, N. J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., & Ponting, C. P. (2003). The Tudor Domain 'Royal Family': Tudor, Plant Agenet, Chromo, PWWP and MBT Domains. *Trends in Biochemical Sciences*, 28(2), 69-74.
- [119] Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., & Patel, D. J. (2007). How Chromatin-Binding Modules Interpret Histone Modifications: Lessons from Professional Pocket Pickers. *Nature Structural & Molecular Biology*, 14(11), 1025-1040.
- [120] Wu, H., Zeng, H., Lam, R., Tempel, W., Amaya, M. F., Xu, C., Dombrowski, L., Qiu, W., Wang, Y., & Min, J. (2011). Structural and Histone Binding Ability Characterizations of Human PWWP Domains. *PLoS One*, 6(6), e18919.
- [121] Chen, T., Tsujimoto, N., & Li, E. (2004). The PWWP Domain of Dnmt3a and Dnmt3b Is Required for Directing DNA Methylation to the Major Satellite Repeats at Pericentric Heterochromatin. *Molecular and Cellular Biology*, 24(20), 9048-9058.
- [122] Bachman, K. E., Rountree, M. R., & Baylin, S. B. (2001). Dnmt3a and Dnmt3b Are Transcriptional Repressors that Exhibit Unique Localization Properties to Heterochromatin. *The Journal of Biological Chemistry*, 276(34), 32282-32287.
- [123] Ge, Y. Z., Pu, M. T., Gowher, H., Wu, H. P., Ding, J. P., Jeltsch, A., & Xu, G. L. (2004). Chromatin Targeting of de Novo DNA Methyltransferases by the PWWP Domain. *The Journal of Biological Chemistry*, 279(24), 25447-25454.
- [124] Shirohzu, H., Kubota, T., Kumazawa, A., Sado, T., Chijiwa, T., Inagaki, K., Suetake, I., Tajima, S., Wakui, K., Miki, Y., Hayashi, M., Fukushima, Y., & Sasaki, H. (2002). Three Novel Dnmt3B Mutations in Japanese Patients with ICF Syndrome. *American Journal of Medical Genetics*, 112(1), 31-37.
- [125] Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., & Jeltsch, A. (2010). The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. *The Journal of Biological Chemistry*, 285(34), 26114-26120.
- [126] Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-Scale DNA Methylation Maps of Pluripotent and Differentiated Cells. *Nature*, 454(7205), 766-770.

- [127] Hodges, E., Smith, A. D., Kendall, J., Xuan, Z., Ravi, K., Rooks, M., Zhang, M. Q., Ye, K., Bhattacharjee, A., Brizuela, L., McCombie, W. R., Wigler, M., Hannon, G. J., & Hicks, J. B. (2009). High Definition Profiling of Mammalian DNA Methylation by Array Capture and Single Molecule Bisulfite Sequencing. *Genome Research*, 19(9), 1593-1605.
- [128] Lee, J. S., & Shilatifard, A. (2007). A Site to Remember: H3K36 Methylation a Mark for Histone Deacetylation. *Mutation Research*, 618(1-2), 130-134.
- [129] Argentaro, A., Yang, J. C., Chapman, L., Kowalczyk, M. S., Gibbons, R. J., Higgs, D. R., Neuhaus, D., & Rhodes, D. (2007). Structural Consequences of Disease-Causing Mutations in the ATRX-DNMT3-DNMT3L (ADD) Domain of the Chromatin-Associated Protein ATRX. *Proceedings of the National Academy of Science*, 104(29), 11939-11944.
- [130] Ooi, S. K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X., & Bestor, T. H. (2007). Dnmt3L Connects Unmethylated Lysine 4 of Histone H3 to de Novo Methylation of DNA. *Nature*, 448(7154), 714-717.
- [131] Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., & Shirakawa, M. (2009). Structural Basis for Recognition of H3K4 Methylation Status by the DNA Methyltransferase 3A ATRX-Dnmt3-Dnmt3L Domain. *EMBO Reports*, 10(11), 1235-1241.
- [132] Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., & Jeltsch, A. (2010). Chromatin Methylation Activity of Dnmt3a and Dnmt3a/3L Is Guided by Interaction of the ADD Domain with the Histone H3 Tail. *Nucleic Acids Research*, 38(13), 4246-4253.
- [133] Pavlopoulou, A., & Kossida, S. (2007). Plant Cytosine-5 DNA Methyltransferases: Structure, Function, and Molecular Evolution. *Genomics*, 90(4), 530-541.
- [134] Finnegan, E. J., & Dennis, E. S. (1993). Isolation and Identification by Sequence Homology of a Putative Cytosine Methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Research*, 21(10), 2383-2388.
- [135] Callebaut, I., Courvalin, J. C., & Mornon, J. P. (1999). The BAH (Bromo-Adjacent Homology) Domain: a Link Between DNA Methylation, Replication and Transcriptional Regulation. *FEBS Letters*, 446(1), 189-193.
- [136] Xu, S., Xiao, J., Posfai, J., Maunus, R., & Benner, J. 2nd. (1997). Cloning of the BssHII Restriction-Modification System in *Escherichia coli*: BssHII Methyltransferase Contains Circularly Permuted Cytosine-5 Methyltransferase Motifs. *Nucleic Acids Research*, 25(20), 3991-3994.
- [137] Wada, Y., Ohya, H., Yamaguchi, Y., Koizumi, N., & Sano, H. (2003). Preferential de Novo Methylation of Cytosine Residues in non-CpG Sequences by a Domains Rearranged DNA Methyltransferase from Tobacco Plants. *The Journal of Biological Chemistry*, 278(43), 42386-42393.

- [138] Henderson, I. R., Deleris, A., Wong, W., Zhong, X., Chin, H. G., Horwitz, G. A., Kelly, K. A., Pradhan, S., & Jacobsen, S. E. (2010). The de Novo Cytosine Methyltransferase DRM2 Requires Intact UBA Domains and a Catalytically Mutated Paralog DRM3 During RNA-Directed DNA Methylation in *Arabidopsis thaliana*. *PLoS Genetics*, 6(10), e1001182.
- [139] Eissenberg, J. C. (2001). Molecular Biology of the Chromo Domain: an Ancient Chromatin Module Comes of Age. *Gene*, 275(1), 19-29.
- [140] Henikoff, S., & Comai, L. (1998). A DNA Methyltransferase Homolog with a Chromodomain Exists in Multiple Polymorphic Forms in *Arabidopsis*. *Genetics*, 149(1), 307-318.
- [141] Lindroth, A. M., Cao, X., Jackson, J. P., Zilberman, D., McCallum, C. M., Henikoff, S., & Jacobsen, S. E. (2001). Requirement of CHROMOMETHYLASE3 for Maintenance of CpXpG Methylation. *Science*, 292(5524), 2077-2080.
- [142] Bartee, L., Malagnac, F., & Bender, J. (2001). Arabidopsis CMT3 Chromomethylase Mutations Block non-CG Methylation and Silencing of an Endogenous Gene. *Genes and Development*, 15(14), 1753-1758.



