VILNIUS UNIVERSITY

Evelina
ZAGORSKAITĖ

RECOGNITION OF MODIFIED CYTOSINE BY METHYL-DIRECTED RESTRICTION ENDONUCLEASES

DOCTORAL DISSERTATION

Physical sciences,
Biochemistry (04 P)

VILNIUS 2018

This dissertation was written between 2013 10 01 and 2017 10 01 at the Institute of Biotechnology, Vilnius University. The work was funded by the Research Council of Lithuania (MIP-027/2012, MIP-106/2015). Some travels were supported by iNEXT (proposal ID1812, funded by the Horizon 2020 programme of the European Union). Scholarship was received from the World Federation of Scientist, Switzerland.

The dissertation is defended on an external basis.

Scientific consultant:
dr. Giedrius Sasnauskas (Vilnius University, physical sciences, biochemistry – 04 P).

Dissertation Defence Panel:
Chairman – Prof. Dr. Saulius Serva (Vilnius University, physical sciences, biochemistry – 04P);
Members:
Dr. Edita Kriukienė (Vilnius University, physical sciences, biochemistry – 04P);
Dr. Zita Liutkevičiūtė (Catholic University of Leuven, Belgium, physical sciences, biochemistry – 04P);
Prof. Dr. Daumantas Matulis (Vilnius University, physical sciences, biochemistry – 04P).
Prof. Dr. Rolandas Meškys (Vilnius University, physical sciences, biochemistry – 04P).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 10:00 on 19th of December 2018 in room R402 of Vilnius University Life Science Center.

Address: Saulėtekio al. 7, Vilnius, Lithuania.

The text of this dissertation can be accessed at the libraries of Vilnius University as well as on the website of Vilnius University: www.vu.lt/lt/naujienos/ivykiu-kalendorius.

VILNIAUS UNIVERSITETAS

Evelina
ZAGORSKAITĖ

METILINTĄ DNR ATPAŽĮSTANČIŲ RESTRIKCIJOS ENDONUKLEAZIŲ SĄVEIKOS SU MODIFIKUOTU CITOZINU MECHANIZMAS

DAKTARO DISERTACIJA

Fiziniai mokslai,
Biochemija (04 P)

VILNIUS 2018

Disertacija rengta 2013 10 01 – 2017 10 01 Vilniaus universiteto Biotechnologijos institute. Mokslinius tyrimus rėmė Lietuvos mokslo taryba (MIP-027/2012, MIP-106/2015). Kai kurias keliones rėmė iNEXT programa. Stipendiją skyrė Pasaulio Mokslininkų Federacija, Šveicarija.

Disertacija ginama eksternu.

Mokslinis konsultantas:
dr. Giedrius Sasnauskas (Vilniaus universitetas, fiziniai mokslai, biochemija – 04 P)

Gynimo taryba:

Pirmininkas – **prof. dr. Saulius Serva** (Vilniaus universitetas, fiziniai mokslai, biochemija – 04P)
Nariai:
Dr. Edita Kriukienė (Vilniaus universitetas, fiziniai mosklai, biochemija – 04P);
Dr. Zita Liutkevičiūtė (Leveno katalikiškasis universitetas, Belgija, fiziniai mosklai, biochemija – 04P);
Prof. Dr. Daumantas Matulis (Vilniaus universitetas, fiziniai mosklai, biochemija – 04P);
Prof. Dr. Rolandas Meškys (Vilniaus universitetas, fiziniai mosklai, biochemija – 04P).

Disertacija bus ginama viešame Gynimo tarybos posėdyje 2018 m. gruodžio 19 d., 10:00 val., Gyvybės mokslų centro auditorijoje R402.

Adresas: Saulėtekio al. 7, Vilnius, Lietuva.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekose ir VU interneto svetainėje adresu: www.vu.lt/lt/naujienos/ivykiu-kalendorius.

CONTENTS

LIST OF ABBREVIATIONS

aa	amino acid
Ap	ampicillin
BER	base excision repair
bp	base pair(s)
BSA	bovine serum albumin
Cm	chloramphenicol
DNMT	DNA methyltransferase
ds	double stranded
DTT	1,4-dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
5mC	5-methylcytosine
4mC	N4-methylcytosine
5hmC	5-hydroxymethylcytosine
5fC	5-formylcytosine
5caC	5-carboxycytosine
5ghmC	5-glucosylhydroxymethylcytosine
6mA	N6-methyladenine
hmU	hydroxymethyluracil
J base	5-β-D-glucosylhydroxymethyluracil
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani bacterial culture medium
MBD	methyl-CpG binding domain
MES	2-(*N*-morpholino)ethanesulfonic acid
nt	nucleotide(s)
PAGE	polyacrylamide gel electrophoresis
REase	restriction endonuclease
RM	restriction-modification
SAXS	small-angle X-ray scattering
SDS	sodium dodecyl sulphate
SRA	SET and RING associated
ss	single stranded
vdW	van der Waals
TDG	thymine DNA glycosylase
TEMED	N,N,N’,N’- tetramethylethylenediamine
TET	Ten-Eleven-Translocation
TRD	transcriptional repression domain
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
wt/WT	wild type
wH	winged helix

INTRODUCTION

5-methylcytosine (5mC) and its oxidized forms 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) serve as epigenetic marks responsible for regulation of gene expression. The methylation levels are dynamic and vary in different types of eukaryotic cells during development, differentiation, aging, and disease (Irier and Jin, 2012; Smith and Meissner, 2013). In prokaryotes, cytosine methylation is primarily related to restriction-modification (RM) systems (Pingoud et al., 2014). In a typical scenario, restriction endonucleases (REases) recognize short specific DNA sequences and cleave foreign unmethylated DNA immediately upon its entry into the cell, while the host genomic DNA remains uncleaved due to methylation by the partner methyltransferase. Some bacteriophages bypass such RM systems by introducing their own DNA modification, including 5mC, 5hmC, or glucosylated 5hmC (5ghmC). As a response to phage DNA modifications, some bacteria have evolved methyl-directed restriction endonucleases, which recognize and cut modified phage DNA (Labrie et al., 2010).

The ability of proteins to discriminate between modified cytosine and unmodified is of key importance for both the epigenetic regulation and the activity of methyl-directed REases. Structural studies of eukaryotic 5mC and 5hmC binding domains revealed two different strategies for the modified base recognition. First, proteins that share the methyl-CpG binding domain (MBD1, MBD2, MBD4, and MeCP2), also a zinc-finger protein Kaiso, recognize modified cytosine in the context of a Watson-Crick base pair (Buck-Koehntop et al., 2012; Ho et al., 2008; Ohki et al., 2001; Otani et al., 2013; Scarsdale et al., 2011). Second, the SRA domains of UHRF1, UHRF2, KRYPTONITE and SUVH5 proteins flip out the modified base and place it in a protein pocket (Arita et al., 2008; Avvakumov et al., 2008; Du et al., 2014; Hashimoto et al., 2008; Rajakumara et al., 2011a; Zhou et al., 2014). Structurally related and likely also capable of base flipping are DNA binding domains found in REases of MspJI and PvuRts1I families that recognize modified cytosine in various sequence contexts (Cohen-Karni et al., 2011; Szwagierczak et al., 2011; Wang et al., 2011). The base flipping mechanism is also employed by structurally unrelated modification-dependent enzyme McrBC which recognizes 5mC, 5hmC and N4-methylcytosine (4mC) (Sutherland et al., 1992). Modified base in this case is recognized by the N-terminal domain of McrB subunit (McrB-N) (Sukackaite et al., 2012). However, the mechanism of modified cytosine recognition by many

uncharacterized eukaryotic and prokaryotic modified cytosine "readers" remains unknown (Iurlaro et al., 2013; Spruijt et al., 2013), as is the DNA recognition and reaction mechanism of the classical methyl-directed *E. coli* enzyme McrA (Revel, 1967), which has no *in vitro* DNA cleavage activity.

Given the importance of cytosine modifications and modified cytosine 'reader' proteins in diverse mammalian cellular processes, and the potential use of bacterial methyl-dependent endonucleases in epigenetic studies, we have dedicated this work to the studies of modified cytosine recognition mechanisms employed by various methyl-directed restriction endonucleases.

The specific aims of this study were:

1. To examine base flipping of modified cytosine in the solution by methyl-directed REases LpnPI, YkrI and BmeDI of MspJI and PvuRts1I families employing methods based on fluorescence and chemical modification, and assess the suitability of these methods for the analysis of the modified base extrusion.
2. To determine the mechanism of the modified cytosine and its sequence context recognition by LpnPI.
3. To investigate the relative affinity of WT McrB-N and its pocket mutants for DNA substrates with 5mC, 4mC, 5hmC, 5fC, and 5caC modifications.
4. To determine functions of N-terminal and C-terminal domains of McrA, and mechanism of modified cytosine recognition.

Scientific novelty

For the first time we demonstrate that MspJI- and PvuRt1I-like enzymes LpnPI, YkrI and BmeDI flip out modified base from dsDNA in solution. We also show that the suitability of various chemical and spectroscopic methods for extrahelical base detection depends on the properties of the protein pocket in question. For example, the positive signal for base flipping by the SRA domains of MspJI- and PvuRts1I-like enzymes was detected using $KMnO_4$ oxidation of extrahelical pyrimidines, while alternative methods, such as pyrrolocytosine fluorescence (base flipping confirmed for the McrBC DNA binding domain) and chloroacetaldehyde modification (base flipping confirmed for methyltransferases and Ecl18kI) gave no such signal. We also demonstrate for the first time that several LpnPI protein loops act as DNA binding / recognition modules, replacement of which results in altered sequence specificity. We present the first detailed study of McrBC specificity for

various cytosine modifications. Finally, this is the first study that demonstrates McrA nuclease activity *in vitro* and confirms functions of its N- and C-terminal domains.

Practical value

Since the mechanism of modified cytosine recognition by many potential eukaryotic and prokaryotic 5(h)mC "readers" is still unknown, simple fluorescence and chemical methods may prove useful in the mechanistic studies of structurally uncharacterized proteins implicated in cytosine status readout. The altered specificity of LpnPI for the 5mC sequence context paves the way for engineering of methyl-directed restriction enzymes which recognize modified cytosines in any sequence context, thereby expanding the toolbox of nucleases with potential applications in epigenome studies. Finally, fundamental characterization of McrA opens up opportunities for the use of this REase for practical purposes, e. g. engineering of artificial methyl-directed endonucleases or nickases.

The major findings presented for defense in this thesis:

1. Methyl-directed restriction endonucleases of MspJI and PvuRts1I families flip out modified cytosine from double stranded DNA.
2. MspJI family enzymes recognize 5-methylcytosine and its sequence context via several surface loops that act as separate DNA binding / recognition modules. Replacement of these modules generates enzyme variants with altered sequence specificity.
3. The affinity of McrBC DNA binding domain for different cytosine modifications can be altered by mutations in the flipped base binding pocket.
4. McrA *in vitro* has weak, Mn^{2+} ion-dependent endonuclease activity, which is partly dependent on DNA methylation.
5. McrA N-terminal domain is responsible for modified cytosine recognition, which occurs without base flipping.

1 LITERATURE OVERVIEW

1.1 DNA modifications

Every organism has a genome which is the storage of the genetic information and can be replicated and transmitted through generations. The genome is composed of long polynucleotide chains of purine and pyrimidine nucleotides: adenine, guanine, cytosine and thymine (A, G, C, T), which form G:C and A:T base pairs (Figure 1.1 A). Billions of these base pairs can be read differently in all types of cells because of epigenetic marking – a mechanism that allows living organisms to adapt to the changing environment and dynamically change genes expression in each cell without changing the primary DNA sequence (Kriukiene et al., 2012).

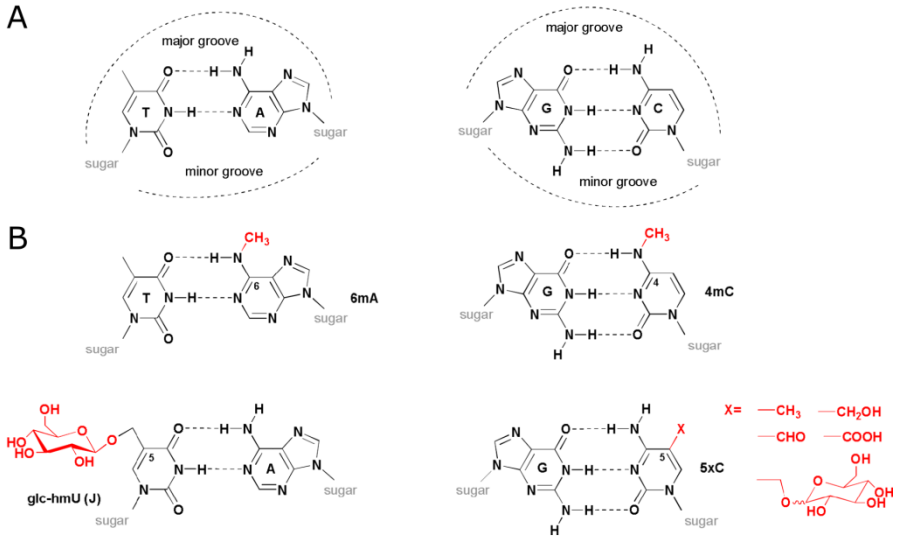


Figure 1.1 DNA base pairs and modifications of DNA bases. (A) A:T and G:C base pairs. (B) Base pairs with modified bases: N6-methyladenine, N4-methylcytosine, 5-β-D-glucosylhydroxymethyluracil and cytosine derivatives with substitution in 5th position: 5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxycytosine, and 5-glucosylhydroxymethylcytosine. Figure adapted from (Kriukiene et al., 2012)

DNA modifications are incorporated in genome after replication and include these modifications of DNA bases: 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), 5-carboxycytosine (5caC), N4-methylcytosine (4mC) and N6-methyladenine (6mA) (Figure 1.1 B).

1.1.1 5-methylcytosine

One of the best explored epigenetic mechanism is DNA methylation. The major function of DNA methylation is gene silencing. Epigenetic regulation by DNA methylation together with histone modifications plays an important role in development genes regulation, differentiation, tissue-specific genes expression regulation, silencing of transposable elements, genomic imprinting, X chromosome inactivation and other (Bird, 2002). Although epigenetic methylation does not change the original genetic content of the genome, the marks occur in the major groove of the DNA helix (Figure 1.1), where many DNA-binding proteins make contact with DNA. So there is an additional layer of information encoded in the genome.

Enzymes that methylate DNA are DNA methyltransferases, they transfer a methyl group from the cofactor S-adenosyl-L-methionine onto DNA (Figure 1.2) (Kriukiene et al., 2012).

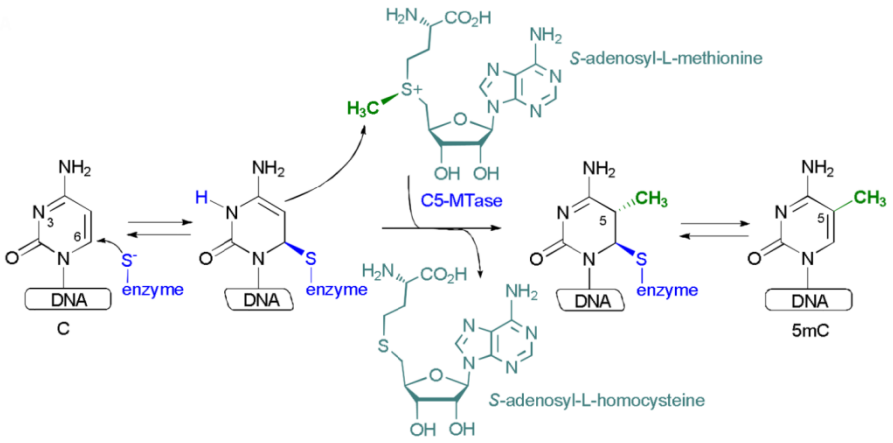


Figure 1.2 DNA cytosine-5 methyltransferase methylation mechanism. Reaction involves covalent addition of a nucleophile (a cysteine residue from an enzyme) at the C6 position of the cytosine ring. Enzyme catalyzes the transfer of methyl group from the cofactor S-adenosyl-L-methionine to the C5 position of the cytosine ring. Figure adapted from (Kriukiene et al., 2012).

In prokaryotes, DNA methyltransferases methylate nucleobases by replacing a hydrogen atom with a methyl group in the bases adenine N6- and cytosine N4- or C5- positions yielding, respectively, N6-methyladenine (6mA), N4-methylcytosine (4mC) or 5-methylcytosine (5mC) (Figure 1.1 B). Prokaryotic DNA methyltransferases are coupled with the accompanying restriction endonucleases with the same recognition sequence. These DNA modifications occur in specific sequences in a genome and the methylation

secure from self-cleavage by restriction endonucleases, which cleave foreign unmethylated DNA protecting bacteria from viruses and other foreign DNA (Pingoud et al., 2014).

In higher eukaryotes, the sole methylation product is 5mC, which accounts for ~1 % of total DNA bases in the human genome (Ehrlich et al., 1982). In mammalian genomes the majority of methylation occurs throughout all genome at CpG sites (70–80 %) with the exception of CpG islands (Bird et al., 1985). CpG islands are high density clusters which are not methylated at all stages of development and in all tissue types (Antequera and Bird, 1993). However, a small fraction of CpG islands is methylated during development and is associated with really stable silencing, e. g. genomic imprinting and X chromosome inactivation (Bird, 2002). DNA methylation is also found within sites other than CpG sequences, called non-CpG sites and includes methylation at CpA, CpT and CpC. Non-CpG methylation mostly found in human embryonic stem cells and brain tissue, and comprises 0.02 % of total 5mC in differentiated cells (Lister et al., 2009; Xie et al., 2012).

Naturally during formation of an embryo the methylation level decreases. The lowest level of DNA methylation is during early stages of development, it accounts ~30 % of the typical somatic level (Bird, 2002). Postreplicative *de novo* methylation restores normal levels of epigenetic patterns. There are three canonical DNA methyltransferases that methylate CpG sites in mammalian genome: DNMT1, DNMT3A and DNMT3B (Lyko, 2017) and recently the fourth DNA methyltransferase DNMT3C was found in mouse genome (Barau et al., 2016) (Figure 1.3).

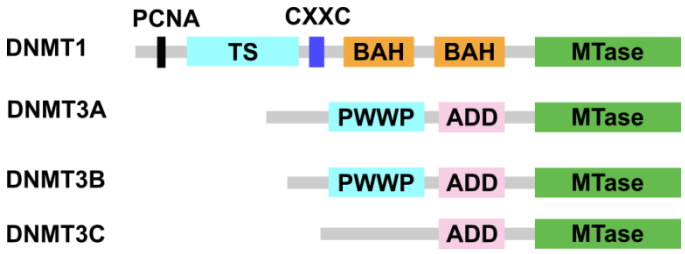


Figure 1.3 Schematic representation of mammalian DNA methyltransferases. DNMTs share a conserved methyltransferase domain, but differ in their N-terminal regulatory regions. DNMT1 contains a PCNA binding domain, a pericentric heterochromatin targeting sequence (TS), a CXXC domain, and two bromo adjacent homology domains (BAH). DNMT3A and 3B comprise of PWWP domain named after a conserved Pro-Trp-Trp-Pro motif and ATRX-DNMT3-DNMT3L (ADD) domain also known as the PHD (plant homeodomain) domain. DNMT3C similar to DNMT3A and DNMT3B, but absents PWWP domain. Figure adapted from (Barau et al., 2016; Kriukiene et al., 2012)

DNMT1 is called maintenance methyltransferase which methylates DNA during cell divisions maintaining the CpG methylation marks. *De novo* methylation occurs early in embryogenesis by DNA methyltransferases DNMT3A and DNMT3B. DNMT3C also is *de novo* DNA methyltransferase which methylates the promoters of evolutionarily young retrotransposons in the male germ line. Non-CpG methylation mechanism so far is poorly understood. However, there is evidence that DNMT3A and DNMT3B methyltransferases also methylate non-CpG sites (Laurent et al., 2010; Ramsahoye et al., 2000).

All four DNMTs are involved in the development of an embryo. After differentiation DNMTs expression level is much reduced. It may suggest that the DNA methylation pattern in differentiated cells is stable. However, in the mature mammalian brain tissue DNMTs expression level is sufficient suggesting that DNMTs and DNA methylation play an important role in the brain (Feng et al., 2005). DNA methylation patterns may undergo dynamic changes and can be reversible in a genome wide and locus specific manner. The identification of 5-hydroxymethylcytosine in postmitotic neurons suggests that 5hmC serves as an intermediate in the DNA demethylation pathway (Kriaucionis and Heintz, 2009).

1.1.2 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxycytosine

First 5-hydroxymethylcytosine was discovered in T-even bacteriophages. These phages incorporate 5hmC into their genome during DNA synthesis and then it is modified by phage α- and β-glucosyltransferases creating a highly glucosylated DNA containing 5-glucosylhydroxymethylcytosine (5ghmC) residues (Figure 1.4). These modifications help to infect bacteria protecting phage DNA from restriction (Warren, 1980). In eukaryotes the first hypermodified DNA base found in kinetoplastid flagellates and some unicellular flagellates is J base (5-β-D-glucosylhydroxymethyluracil, figure 1.1 B). It covers 1 % of all thymines and is found in repetitive sequences, mostly in telomeric repeats. In the first step of J base biosynthesis thymine is oxidized to hydroxymethyluracil (hmU) by the J-binding proteins JBP1 / JBP2 which are members of a large 2-oxoglutarate and Fe(II)-dependent oxygenase family (Borst and Sabatini, 2008). In the second step hmU is glucosylated. The functions of base J remain unknown but mostly it relates to gene silencing and repression of repetitive sequences.

In higher eukaryotes, 5hmC was first discovered several decades ago in the brain of mice, rats and frogs (Penn et al., 1972). Later it was thought that

5hmC is an *in vitro* artifact of chemical DNA oxidation and these results could not be reproduced. However, only in 2009 two independent groups confirmed that 5hmC exist in the mouse brain and embryonic stem cells (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009).

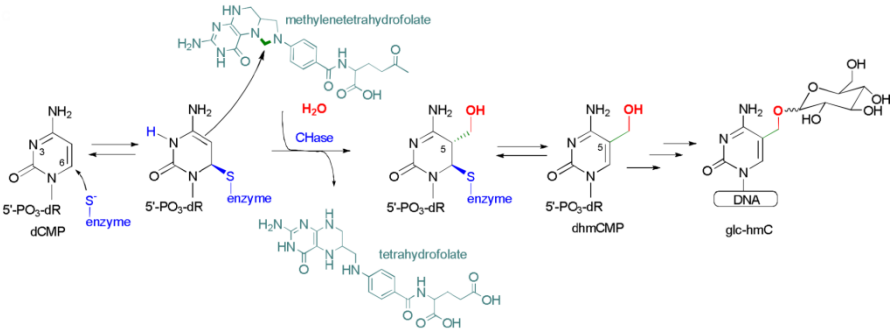


Figure 1.4 DNA hydroxymethylation and glucosylation by bacteriophage. dCMP is converted into 5-hydroxymethyl-dCMP (dhmCMP) by a bacteriophage deoxycytidylate hydroxymethylase, which transfers a methylene group from methylenetetrahydrofolate and adds the hydroxyl group to the transferred methylene group, producing dhmCMP and tetrahydrofolate. The hydroxymethylated nucleotide is incorporated into bacteriophage DNA and glucosylated by α- and β-glucosyltransferases yielding 5ghmC. Figure adapted from (Kriukiene et al., 2012)

Kriaucionis and Heinz using thin layer chromatography, high-pressure liquid chromatography, and mass spectrometry identified that 5hmC constitutes 0.6 % of total nucleotides in Purkinje cells and 0.2 % in granule cells. The other group identified JBP1 and JBP2 homologs in mammalian genome, the Ten-Eleven-Translocation (TET) proteins TET1, TET2 and TET3 (Figure 1.5 A) which convert 5mC to 5hmC (Tahiliani et al., 2009). TET proteins can further oxidize 5hmC to 5fC and 5caC which rapidly are processed by base excision repair machinery (Wu and Zhang, 2017).

5mC is a dynamic epigenetic mark which can be removed in two ways: through a passive demethylation or an active demethylation (Wu and Zhang, 2017). First, during DNA replication methylated cytosine marks can be diluted passively from the genome when there is lack of active maintenance methyltransferase. Second, active demethylation carried out by TET proteins which oxidize 5mC to 5hmC, 5fC and 5caC. Then these oxidized variants can be diluted during replication or excised by thymine DNA glycosylase (TDG) and repaired in the base excision repair (BER) pathway (Figure 1.5 B).

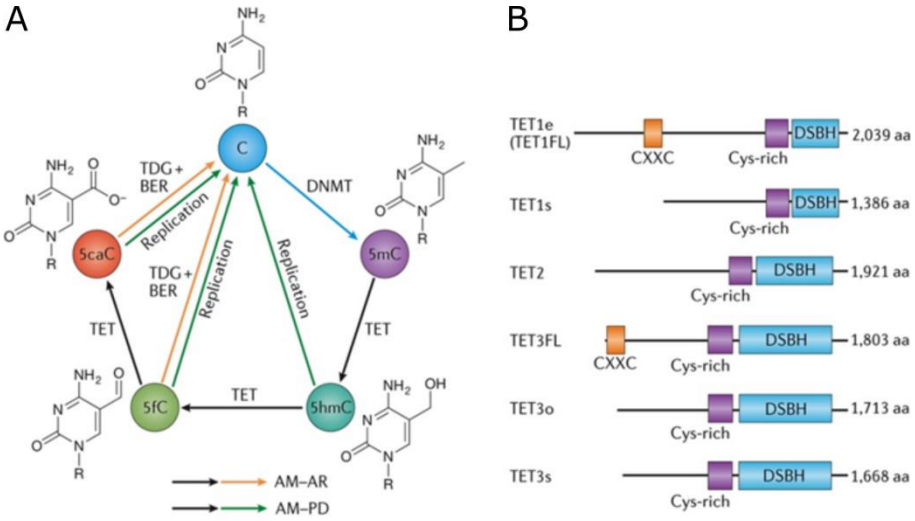


Figure 1.5 TET proteins and active DNA demethylation. (A) Domain structure of mouse TET proteins. Catalytic center is in the C-terminal domains: cysteine-rich and double-stranded β-helix (DSBH). Full-length TET1 (TET1FL) and TET3 have a CXXC domain at the N-terminus. TET1 and TET3 have multiple splicing isoforms. (B) Schematic representation of active DNA demethylation. DNMTs methylate cytosine yielding 5mC. TET proteins oxidize 5mC to 5hmC, 5fC and 5caC which can be excised by TDG and BER mechanism (the process of active modification–active removal (AM–AR)) or replication-dependent dilution of 5hmC, 5fC or 5caC (the process of active modification–passive dilution (AM–PD)). Figure adapted from (Wu and Zhang, 2017)

However 5hmC and its derivatives are not only intermediates in the active 5mC demethylation pathway. It was demonstrated that 5hmC and 5fC are a predominantly stable modifications and their levels differ from each other and 5mC (Figure 1.6), and are recognized by specific proteins (Bachman et al., 2014, 2015).

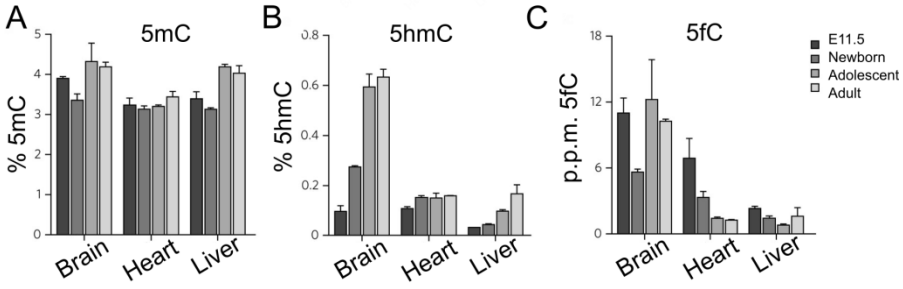


Figure 1.6 5mC, 5hmC and 5fC level variations in different tissues. Changes of global 5mC (A), 5hmC (B) and 5fC (C) levels during development in different mouse tissues. Figure adapted from (Bachman et al., 2015).

It was shown that 5hmC occurs in all mammalian tissues but distribution is tissue-specific (Figure 1.6 B) and global levels range between 0.005 % and 0.7 % of all cytosines (Bachman et al., 2014). 5fC is found in all tissues but also is specific for some of them (Figure 1.6 C) and relative levels differ from 0.02–0.002 % of total cytosines (Bachman et al., 2015). 5hmC play an important role in transcriptional regulation and maintenance of pluripotency of embryonic stem cells, neurogenesis, brain development and aging (Kriukiene et al., 2012). It was found that 5fC play important role in DNA replication (Ji et al., 2018).

1.1.3 N4-methylcytosine and N6-methyladenine

4mC and 6mA DNA modifications (Figure 1.1 B) are mostly found in prokaryotes (Ye et al., 2017). Their primarily function is to distinguish host DNA from foreign pathogenic DNA (Pingoud et al., 2014). Moreover, 6mA is established as epigenetic mark in bacteria and play important role in DNA repair, replication, cell cycle, repression of transposable elements, and transcription regulation (Fang et al., 2012; Wion and Casadesús, 2006). Recently, several studies reported that 6mA also was found in eukaryotic genomes, including *Chlamydomonas reinhardtii*, *Caenorhabditis elegans*, and *Drosophila melanogaster*, with epigenetic functions in gene regulation and development (Fu et al., 2015; Greer et al., 2015; Zhang et al., 2015).

1.2 Modified DNA recognition in eukaryotes

Covalent modifications of cytosine 5′ position change chemical groups used for protein-DNA interactions in the major groove (Figure 1.1). It is another layer of information that alters affinity and specificity of DNA binding by various proteins. Modified cytosine “readers” specifically recognize DNA modifications and initiates regulatory events which change gene expression and chromatin structure (Shimbo and Wade, 2016). These proteins recognize CpG sequences in DNA and act as transcription repressors and other effectors. Structural studies of eukaryotic methyl-binding domains revealed two different strategies for the modified base recognition. Proteins that share the methyl-CpG binding domain, including MeCP2, MBD1, MBD3, and MBD4, also a zinc-finger domain of protein Kaiso, recognize modified cytosine in the context of a Watson-Crick base pair (Buck-Koehntop et al., 2012; Ho et al., 2008; Ohki et al., 2001; Otani et al., 2013; Scarsdale et al., 2011). In contrast, the SRA domains of UHRF1, UHRF2, KRYPTONITE and SUVH5 proteins flip out the modified base and place it in a protein pocket (Arita et al., 2008; Avvakumov et al., 2008; Du et

al., 2014; Hashimoto et al., 2008; Rajakumara et al., 2011a; Zhou et al., 2014). The domain composition of each methyl-binding protein is unique (Figure 1.7) which provides opportunities to interact with different multiple targets resulting in epigenetic gene expression regulation.

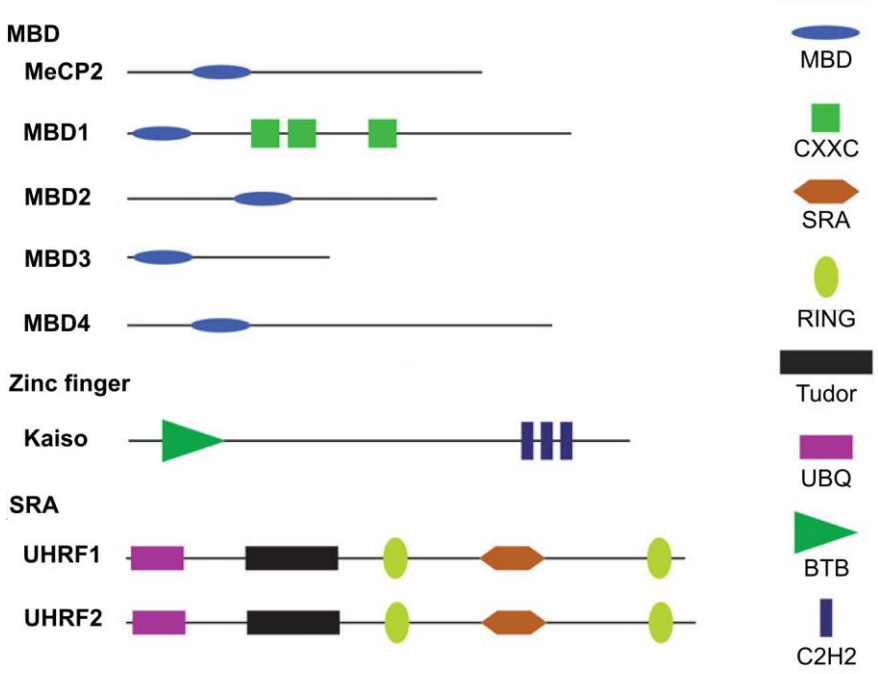


Figure 1.7 Methyl-binding proteins and their domains. MBD, Methyl-CpG-binding domain; CXXC, zinc finger domain; PWWP, "Pro-Trp-Trp-Pro" domain; SRA, SET- and RING-associated domain; RING, Really Interesting New Gene finger domain; Tudor domain; UBQ, ubiquitin-like domain; BTB BR-C, ttk, and bab domain; C2H2, zinc finger domain. Figure adapted from (Shimbo and Wade, 2016).

1.2.1 Methyl-CpG binding domain proteins

MBD (methyl-CpG binding domain) proteins recognize methylated CpG sites (5mCG/5mCG) through a conserved MBD domain and recruit chromatin-modifying enzymes, histone deacetylases, and transcriptional repressors to these methylated sites to silence genes (Jones et al., 1998; Nan et al., 1998). The amino acid sequence of each MBD protein is unique, except the MBD domain is conserved. Each protein target specific promoters and recruit specific regulatory complexes and has a critical role in various biological processes; mutations in MBD proteins are associated with various neurological and developmental disorders (Shimbo and Wade, 2016).

1.2.1.1 MeCP2 protein

The first characterized methyl-CpG binding protein was MeCP2 (Lewis et al., 1992). The protein consists of 492 amino acids, MBD domain located in the N-terminal region (Figure 1.7) and a transcriptional repression domain (TRD) is in the C-terminal region. MeCP2 recognizes single fully methylated CpG dinucleotide (Lewis et al., 1992; Nan et al., 1993). For high affinity DNA binding by MeCP2 the CpG dinucleotide should be flanked by four or more AT base pairs: 5mCG-$(A/T)_{\geq 4}$. Tandem Asx-ST motif contacts the phosphate backbone at the start of the A/T run (Klose et al., 2005).

The crystal structure of MeCP2 revealed that MeCP2 recognizes methyl-CpG using two arginine, aspartate and tyrosine residues: R111, R133, D121, and Y123 (Figure 1.8) (Ho et al., 2008).

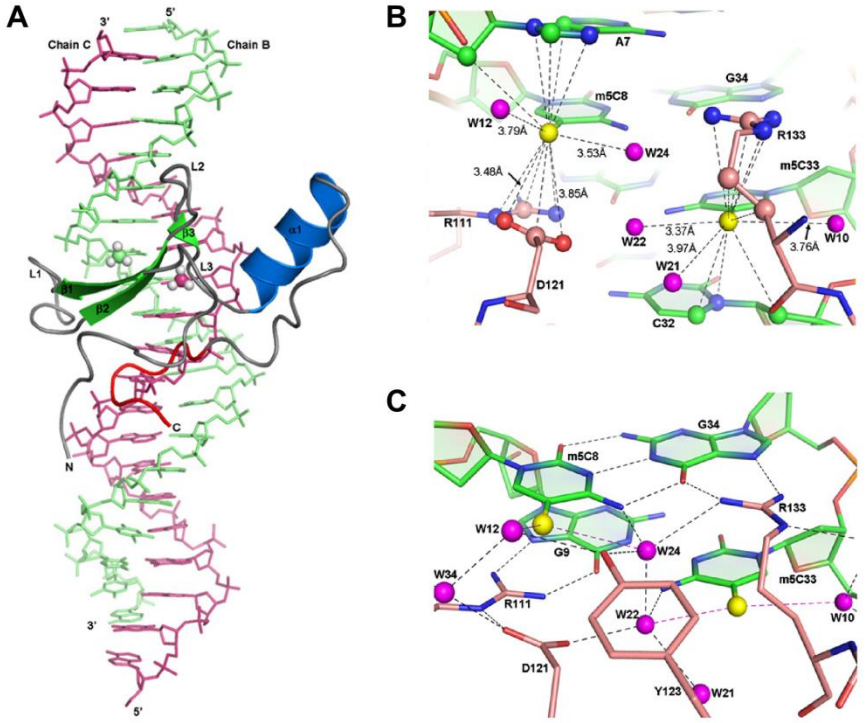


Figure 1.8 The crystal structure of MeCP2-MBD in complex with DNA. (A) The overall MeCP2-MBD-DNA structure. Methyl groups are shown as spheres. The β strands (green) and α helix (blue) are connected by the loops L1, L2 and L3. The tandem Asx-ST motif is in red. (B) Contacts with 5mC methyl groups. Black dashed lines show all nonbonded contacts to the 5mC methyl groups of less than 4 Å. The 5mC methyl groups are shown as yellow balls and water molecules W21, W22, and W24 as purple balls. (C) H-bonds (black dashed lines) with 5mCpG pair. Specific H-bonds formed between R111 and G9, also R133 and G34. Figure adapted from (Ho et al., 2008)

1.2.1.2 MBD1 protein

Other MBD proteins including MBD1 were found in homology searches using the first discovered MBD protein MeCP2 (Hendrich and Bird, 1998). MBD1 differs from other MBD proteins: the protein has not only MBD domain that interacts with DNA, but also has two or three CXXC zinc finger domains (Figure 1.7). MBD domain located in the N-terminal region and TRD domain is in the C-terminal region, CXXC zinc fingers domain is in the middle (Cross et al., 1997; Fujita et al., 1999). The zinc fingers can help MBD1 interact with unmethylated CpG dinucleotides (Jørgensen et al., 2004). However, the binding to unmethylated CpG sites was observed only if the MBD domain was deleted, suggesting that the recruitment of MBD1 to target genes firstly requires the MBD-methyl-CpG interaction (Baubec et al., 2013; Clouaire et al., 2010). MBD1 recognizes methyl-CpG within specific sequence context. MBD1 has a higher binding affinity for methyl-CpG sites with A at position +2 and T at position –1 / 2: T(G)5mCGCA (Clouaire et al., 2010). The methyl-CpG site context is important for target recognition by different MBD proteins (Shimbo and Wade, 2016).

The NMR structure analysis revealed that MBD1 interface is asymmetric and form a hydrophobic patch recognizing symmetrical methyl-CpG site (Figure 1.9) (Ohki et al., 2001).

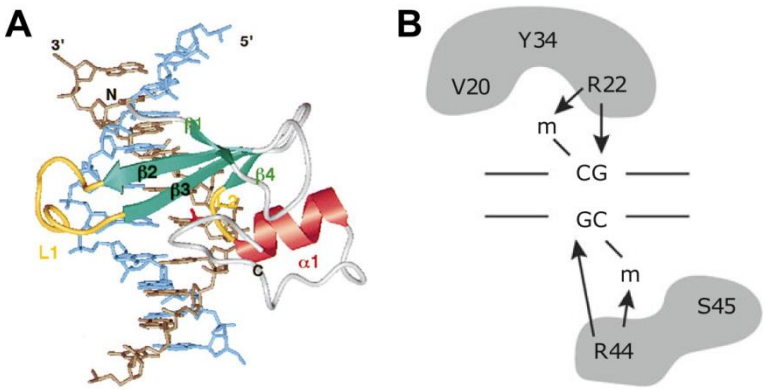


Figure 1.9 The methyl-CpG recognition by MBD1. (A) NMR structure of the MBD1-MBD-DNA complex. The β strands (green) and α helix (red) are connected by the loops L1and L2 (yellow). (B) Schematic representation of the methyl-CpG recognition by MBD1. The symmetrical DNA substrate – methyl-CpG pair – is recognized by an asymmetric protein recognition module. Two interaction surfaces make contacts with the methyl-CpG dinucleotide on each strand separately: V20, R22 and Y34 residues bind one strand, and R44 and S45 – the other strand. Figure adapted from (Ohki et al., 2001; Shimbo and Wade, 2016)

The two methyl groups are recognized by making contacts with MBD domain residues valine, two arginines, tyrosine and serine (V20, R22, R44, Y34, and S45) in the major groove (Figure 1.9 B). The methyl groups make hydrophobic interactions with MBD domain amino acid residues. The arginine residues form H-bonds with guanines, the tyrosine residue contacts one of the two cytosine methyl group (Ohki et al., 2001). The recognition of the methyl-CpG sites by MBD1 is very similar to the MeCP2. Interesting is that MBD1 binding methylated DNA causes an unstructured loop L1 to fold into a structured and make interactions with DNA backbone of one strand (Figure 1.9 A) (Ohki et al., 2001).

1.2.1.3 MBD2 and MBD3 proteins

MBD2 and MBD3 are components of the nucleosome remodeling and histone deacetylase complex NuRD/Mi-2 (Feng and Zhang, 2001). Also MBD2 was found in the MeCP1 complex which is large multisubunit complex and has similarities with NuRD complex (Feng and Zhang, 2001; Meehan et al., 1989). It was shown that NuRD is core complex which can be recruited either by MBD2 or other sequence-specific DNA binding proteins. In the MeCP1 complex MBD2 recruits NuRD/Mi-2 complex to methylated genes silencing (Feng and Zhang, 2001). Although MBD3 has a high sequence similarity to MBD2, their targets are different. MBD3 is the only member of MBD family proteins lacking specificity towards methyl-CpG due to mutations in the MBD domain (Saito and Ishikawa, 2002). It was suggested that MBD3 binds hydroxymethylated CpG dinucleotides (Yildirim et al., 2011). However binding experiments does not confirm affinity preference for hydroxymethylated CpG sites (Hashimoto et al., 2012a). MBD2 and MBD3 are localized to transcriptional start sites within CpG islands: MBD2 preferentially binds methylated CpG sites and is associated with these genes silencing; MBD3 localize to unmethylated CpG islands and is enriched at active promoters (Baubec et al., 2013). Also the biological roles of MBD2 and MBD3 are different. Knockout studies in mice show that knocking out MBD3 is lethal for embryos, meanwhile, MBD2 knockout mice stay alive having few defects (Hendrich et al., 2001).

The NMR structure of chicken MBD2 was determined with CpG target from the β-globin gene promoter (Figure 1.10) which is natural MBD2 target and this interaction mediates developmental transcriptional silencing in erythroid cells (Scarsdale et al., 2011).

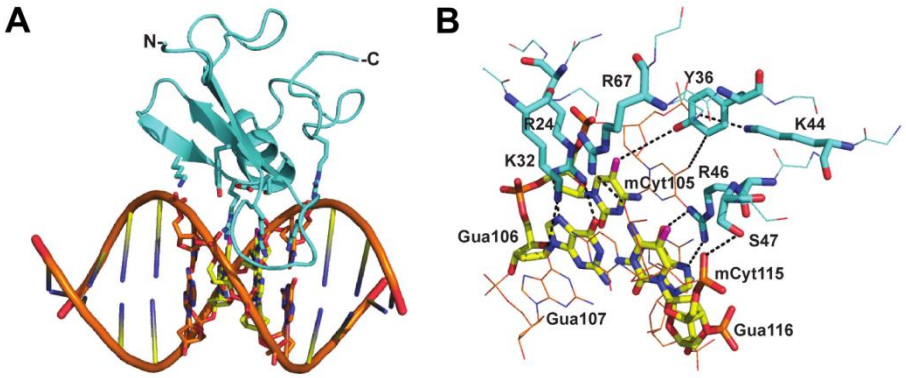


Figure 1.10 NMR structure of MBD2 methyl-binding domain bound to methylated DNA. (A) Overall structure of MBD2 (cyan) and DNA (blue / orange). (B) Contacts with 5mCpG. Base-specific contacts formed with R24, R46, Y36 and K32. Figure adapted from (Scarsdale et al., 2011)

The structural and dynamic studies showed that MBD2 recognizes methylated CpG target similarly to other MBD proteins. The methyl groups are recognized by a pair of arginine residues R24 and R46, and tyrosine residue Y36 (Figure 1.10 B). However, there are differences in the methylated DNA target recognition which leads to higher DNA affinity and selectivity for methylated versus unmethylated CpG dinucleotides. MBD2 binds the target sequence in a single orientation. It was observed that MBD2 prefer 5mCGG sequences because lysine residue (K32) makes direct contacts with +1 guanine base (Figure 1.10 B) (Scarsdale et al., 2011).

The MBD3 NMR structure analysis revealed that MBD domain is similar to other MBD proteins; however the loop involved in DNA binding is not structured as previously observed (Figure 1.11 A) (Cramer et al., 2014). In contrast to MBD2, which shows a strong preference for methylated CpG sites, NMR experiments with methylated, unmethylated and hydroxymethylated CpG dinucleotides showed that MBD3 preferentially localizes to methylated sites, less – to unmethylated sites, but the structure of MBD3 does not change while binding different targets. It was also shown that MBD3 does not distinguish between methylated and hydroxymethylated cytosine forms (Cramer et al., 2014).

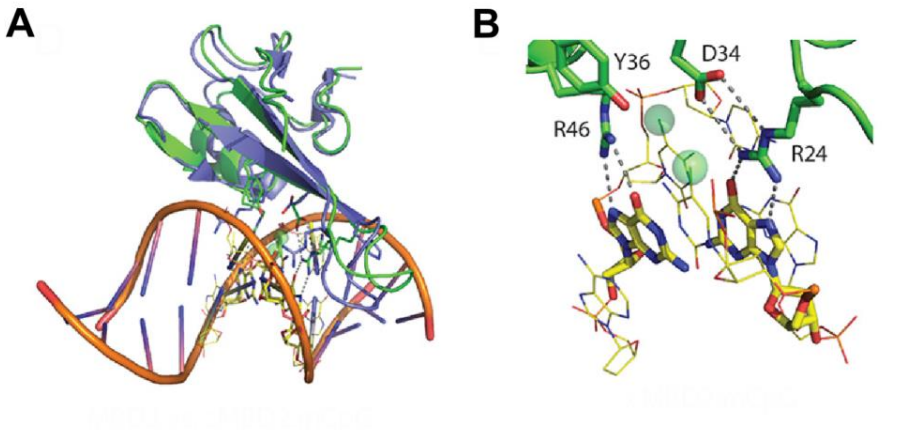


Figure 1.11 The NMR structures of MBD2 and MBD3 methyl-binding domains bound to methylated DNA (A) Overall structure MBD-DNA of MBD2 (green) vs MBD3 (blue). (B) Contacts with 5mCpG. Base-specific contacts formed with R24, R46, Y36 and D34. Green spheres indicate methyl groups. Figure adapted from (Cramer et al., 2014)

1.2.1.4 MBD4 protein

MBD4 contains two functional domains, C-terminal DNA glycosylase domain and N-terminal MBD domain. MBD4 is a unique member of MBD protein family, as it participates in DNA mismatch repair as T/G and U/G mismatch glycosylase, and in transcriptional repression through methyl-CpG binding (Hendrich et al., 1999; Kondo et al., 2005). MBD4 has high affinity for methylated CpG dinucleotides and also for T or U mismatches in CpG site, which are the products of the methylated or hydroxymethylated CpG deamination. Thus, MBD4 participates in DNA repair, and also in DNA demethylation (Figure 1.12) (Hashimoto et al., 2012b; Hendrich et al., 1999). The role of MBD4 in genomic integrity maintenance was demonstrated in MBD4-deficient mice (Millar et al., 2002). The deletion resulted in an increased frequency of C to T transitions at CpG sites. Moreover, mutations in human MBD4 protein was observed in various carcinomas with microsatellite instability (Riccio et al., 1999).

The crystal structure of the MBD of MBD4 was determined with symmetric methylated CpG substrate (5mCpG/5mCpG) and with asymmetric T/G mismatched substrate (5mCpG/TpG) (Figure 1.13 A) (Otani et al., 2013). The structures are similar to other MBD proteins with some minor alterations with the mismatched substrate. As in MeCP2, an important role in the target recognition is attributed to an extensive water network (Figure 1.13 B, C) (Ho et al., 2008; Otani et al., 2013). The methyl groups

are recognized by conserved arginine fingers (R84 and R106), aspartate residue (D94) and tyrosine residue (Y96) (Figure 1.13). Conserved tyrosine residue in MBD4 has a different orientation from other MBD proteins, as it is shifted from the DNA interface and makes water-mediated interactions and the hydroxyl group occupies a different position (Figure 1.13 B, C).

The crystal structure of another DNA recognizing domain of MBD4 – the glycosylase domain – revealed a different DNA recognition mechanism. It was shown that MBD4 glycosylase domain flips out the target nucleotide from dsDNA and arginine residue fills the left space (Hashimoto et al., 2012b). The solution structure analysis of full-length MBD4 protein revealed that MBD domain scans methylated CpG sites and facilitates the mismatch repair by the glycosylase domain (Walavalkar et al., 2014). The base flipping out of dsDNA mechanisms are discussed below.

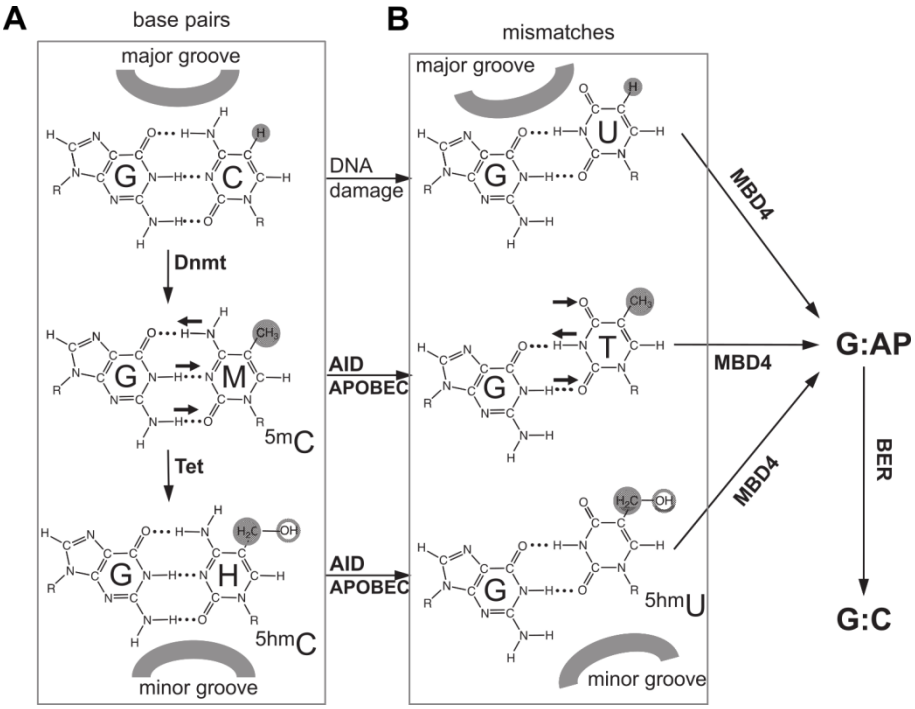


Figure 1.12 A putative pathway of DNA demethylation and mismatches repair by MBD4. (A) Cytosine methylation by DNMT and hydroxymethylation by Tet proteins. (B) DNA damage and deamination by AID (APOBEC superfamily proteins) generate mismatches which are removed by base excision by MBD4 linked to base excision repair (BER). Figure adapted from (Hashimoto et al., 2012a)

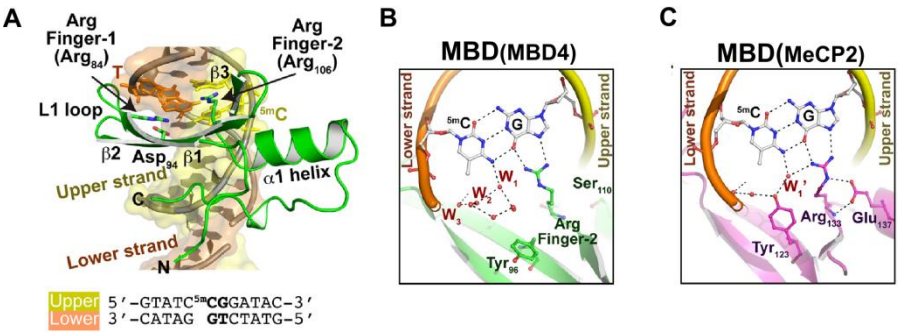


Figure 1.13 Structure of MBD4 methyl-binding domain with DNA and comparison with MBD of MeCP2. (A) Overall structure of MBD4-MBD with DNA containing 5mCG/TG. Two arginine fingers make base-specific contacts. The lower DNA strand containing a mismatched T is colored in orange, and the complementary upper strand is yellow. (B) MBD domain of MBD4 makes contacts through water molecules (W1-3), arginine fingers. Tyrosine orientation differs from MeCP2-MBD (C). Figure adapted from (Otani et al., 2013).

1.2.2 SET- and RING-associated domain proteins

SRA (SET and RING-associated) domain proteins recognize modified CpG sites in various sequence contexts through a conserved SRA domain and is responsible for the maintenance of global and local DNA methylation, recruiting DNMTs, histone modifying enzymes which lead to gene silencing (Bostick et al., 2007; Jackson et al., 2004; Sharif et al., 2007). SRA proteins consist of several domains with different activities, e. g. SUVH4, SUVH5, and SUVH6 contain histone methyltransferase domain and catalyze H3K9 dimethylation (H3K9me2) (Jackson et al., 2004). SRA domains use different mechanism for methyl-CpG recognition comparing with MBD proteins, SRA domains flip out modified base from DNA helix and accommodate in a protein pocket. The recognition occurs through several loops which reach DNA from both the major and the minor grooves.

1.2.2.1 UHRF1 protein

UHRF1, known as NP95 (nuclear protein 95) in mice and ICBP90 (inverted CCAATT box binding protein 90) in human, contains UBL (ubiquitin-like), PHD (plant homeo domain), RING (really interesting new gene), and SRA (SET- and RING-associated) domains (Figure 1.7) (Fujimori et al., 1998; Hopfner et al., 2000; Mousli et al., 2003). UHFR1 recognizes hemi-methylated CpG sites through the SRA domain and recruits *de novo* DNA methyltransferase (DNMT1) to the replication fork (Figure 1.14), where newly synthesized CpG site is methylated to maintain the

methylation status of the template strand (Bostick et al., 2007; Sharif et al., 2007). UHRF1 knockout in mice results in genome-wide DNA hypomethylation, which in turn causes embryonic lethality. Also it was shown that UHRF1 plays an important role in tumorigenesis by inducing genome-wide DNA hypomethylation (Shimbo and Wade, 2016).

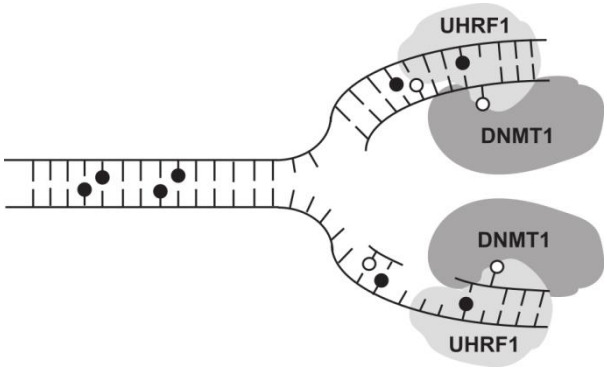


Figure 1.14 UHRF1 role in the maintenance of DNA methylation after replication. Symmetrically methylated CpG dinucleotides are shown as black circles on DNA strands. After DNA replication UHRF1 recognizes hemimethylated CpG sites (white circles), flips out the unmethylated base, and recruits DNMT1, which methylates the unmethylated cytosine. Figure adapted from (Defossez and Stancheva, 2011)

The crystal structures of SRA domain of UHRF1 (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008) revealed that hemi-methylated CpG site is recognized by the base flipping mechanism; 5mC is flipped out of dsDNA and accommodated in the protein pocket (Figure 1.15). The position of the extruded base is stabilized by stacking contacts by two tyrosine residues, hydrogen bonds and van der Waals (vdW) contacts (Figure 1.15 B). Binding of DNA by SRA domain induces formation of two different loops that are responsible for CpG recognition and base flipping. The two loops make contacts in both the major and the minor grooves: the NKR 'finger' (N, K and R are amino acid sequence in the loop) makes hydrogen bonds in the major groove with CpG sequence and the 'thumb' occupies the minor groove but makes no base-specific contacts with the DNA bases (Figure 1.15). The 'thumb' and the 'finger' interacts with each other through the free space left after 5mC extrusion out of the DNA helix (Avvakumov et al., 2008). Although the base is flipped out of the DNA helix, the DNA in the complex with SRA domain retains the regular B-form conformation (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). The UHRF1-SRA domain is the first known example of a protein,

which flips a base out of the DNA helix, but has no enzymatic activity (Arita et al., 2008), contrary to other base flipping enzymes, such as methyltransferases, REases and DNA repair enzymes (Hong and Cheng, 2016).

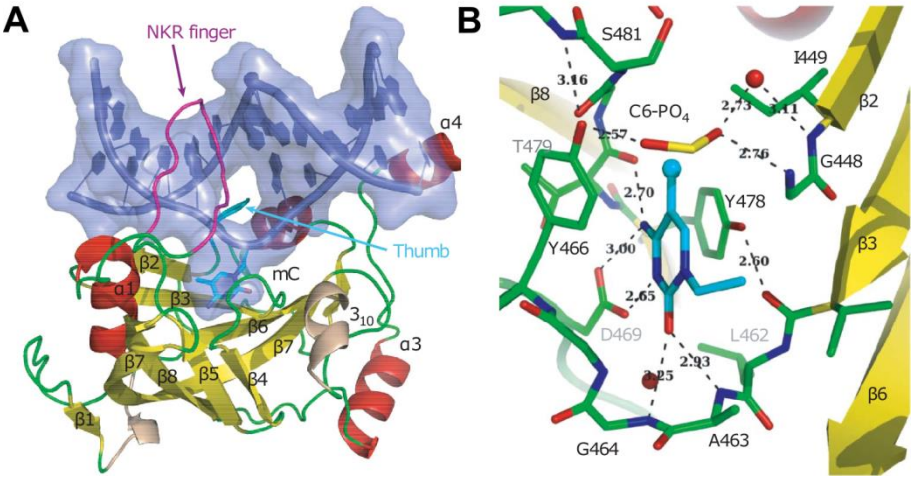


Figure 1.15 The crystal structure of UHRF1 SRA domain and DNA complex. (A) The SRA-DNA complex. The flexible loops, β-strands and α-helices are colored in green, yellow and red, respectively. NKR finger and thumb loop are colored in magenta and cyan. (B) Specific contacts in the protein pocket between 5mC and pocket amino acid residues. Water molecules are shown as red spheres. H-bonds are shown as dashed lines and labelled with distances (Å). Figure adapted from (Avvakumov et al., 2008)

In addition to DNMT1, UHRF1 is associated with other epigenetic regulators, such as H3K9 methyltransferase G9a, histone deacetylase and histone acetyltransferase Tip60 (Hu et al., 2011). UHRF1 also contains histone reader domains (Tudor and PHD domains), which recognize modified and unmodified histone modifications, such as H3K9me3, H3K4, and H3R2 (Nady et al., 2011; Rajakumara et al., 2011b). Thus, UHRF1 functions as a reader of multiple epigenetic modifications and connects DNA methylation with histone modifications.

1.2.2.2 UHRF2 protein

UHRF2 shares high sequence and structural similarity with UHRF1 (Zhou et al., 2014). Both proteins recognize hemi-methylated DNA and interact with DNMTs and histone methyltransferase G9a (Zhang et al., 2011). However, UHRF1 and UHRF2 are functionally different. First, UHRF2 specifically recognizes 5hmC, not 5mC. Next, UHRF2 prefers fully hydroxymethylated targets to hemi-hydroxymethylated, also opposite to

UHRF1 (Spruijt et al., 2013; Zhou et al., 2014). Finally, the proteins are expressed in different cells. UHRF1 is expressed in embryonic stem cells, while UHRF2 is expressed in differentiated cells, where the level of fully methylated DNA is high (Pichler et al., 2011; Zhang et al., 2011). It could explain the target preference for fully, not hemi-methylated CpG dinucleotides. Treatment of UHRF1-null embryonic stem cells with UHRF2 did not rescue the hypomethylation phenotype, which also confirms that these proteins have different functions (Pichler et al., 2011; Zhang et al., 2011).

The structure of UHRF2-SRA domain with 5hmC containing DNA (Zhou et al., 2014) revealed the same recognition mechanism for the modified base – 5hmC is extruded from the DNA helix and placed in the protein pocket (Figure 1.16).

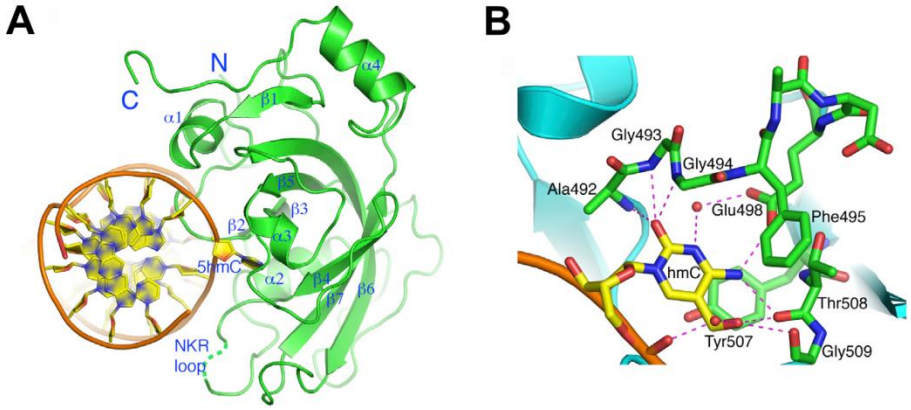


Figure 1.16 The crystal structure of UHRF2 SRA domain binding 5hmC containing DNA. (A) Overall UHRF2-SRA-DNA structure**.** The flipped 5hmC base is colored yellow. (B) Specific contacts in the protein pocket between 5hmC and pocket amino acid residues. Water molecules are shown as red spheres. H-bonds are shown as dashed lines. Figure adapted from (Zhou et al., 2014)

Despite the similarities, there are several unique features in the structure explaining functional differences between UHRF1 and UHRF2 SRA domains. The additional hydroxyl group of 5hmC makes a hydrogen bond with threonine (T508); and phenylalanine (F495) makes a larger binding pocket more suitable for 5hmC (Figure 1.16 B). Another difference is that UHRF2-SRA 'NKR finger' loop is disordered in the structure (Figure 1.16 A) (Zhou et al., 2014). The same feature was observed in the SUVH5 SRA structure, which can bind dual flipped-out DNA, and the loop flexibility is associated with fully modified DNA target binding (Rajakumara et al., 2011a; Zhou et al., 2014).

1.2.2.3 SUVH5 protein

SU(VAR)3-9 HOMOLOG 5 (SUVH5) is a member of SET proteins that share a conserved SRA domain (Baumbusch et al., 2001). In *Arabidopsis thaliana*, SUVH5 with SUVH4 and SUVH6 is required for methylation maintenance of the epigenetic marks in DNA (5mC) and histone tails (H3K9me2) (Ebbs and Bender, 2006; Jackson et al., 2004; Law and Jacobsen, 2010). In plants DNA methylation can occur in all sequence contexts: symmetric CG and CHG (where H=A, T, or C), and asymmetric CHH; meanwhile in mammals, DNA methylation occurs mostly in the symmetric CG sequences (Law and Jacobsen, 2010). SUVH5 SRA domain specifically recognizes 5mC and 5hmC in all sequence contexts, and prefer fully- over hemi-methylated sequences (Rajakumara et al., 2011a, 2016).

Crystal structures of the SUVH5 SRA domain was determined with DNA containing fully or hemi-methylated CG, and an asymmetric CHH contexts, also with fully-hydroxymethylated CG context (Rajakumara et al., 2011a, 2016). The structures revealed a dual flip-out mechanism: two bases from both strands (5mC / 5hmC or unmodified base) are simultaneously extruded from the DNA helix and are accommodated in the binding pockets of individual SRA domains (Figure 1.17) (Rajakumara et al., 2011a, 2016).

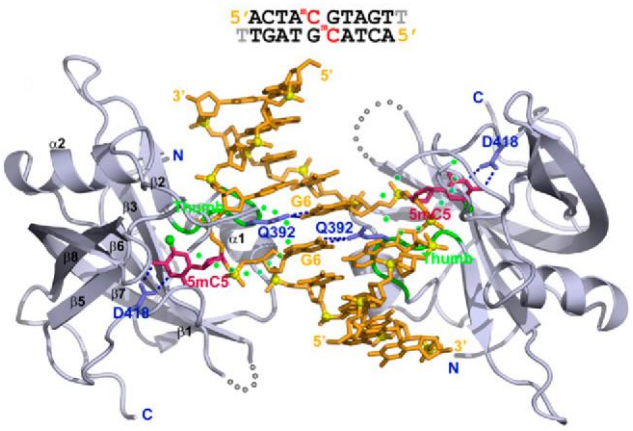


Figure 1.17 Crystal structure of the SUVH5 SRA domain bound to fully methylated CG DNA. Oligoduplex sequence with fully methylated CG dinucleotide in the center (colored red) is shown in the top. Dually flipped out 5mC are colored in purple. Methyl groups of 5mC are shown as green spheres. 'Thumb' loop and distorted 'NKR' loop segments are colored in green and dotted green, respectively. D418 makes H-bonds with 5mC in the binding pockets of individual SRA domains. Figure adapted from (Rajakumara et al., 2011a)

1.3 Modified DNA recognition in prokaryotes

1.3.1 Restriction endonucleases

Restriction endonucleases (REases) are enzymes that recognize short specific DNA sequences and cut DNA in or near the sequence. Restriction enzymes are found in bacteria, archaea and some viruses together with methyltransferases which recognize the same DNA sequence as the REase and methylate these sequences to ensure that genome of the bacterial cell remains intact. The main function of restriction-modification systems is to protect bacteria from viruses, plasmids and other foreign DNA (Pingoud et al., 2014). Also there is evidence that restriction enzymes are modulators of the frequency of genetic variation; and function as selfish mobile elements (Arber, 2000; Kobayashi, 2001). There are four main types (I, II, III and IV) of restriction endonucleases which differ in structure, DNA targets or cofactors (Roberts et al., 2003). The restriction enzymes database REBASE statistics show that there are 4466 characterized enzymes, Type II REases – 4310, Type I – 115, Type III – 22 and Type IV – 19 (Roberts et al., 2015).

1.3.2 Modification-dependent restriction endonucleases

In prokaryotes modification-dependent restriction endonucleases protect host bacteria from bacteriophages containing modified DNA. Phages under pressure from RM systems evolved mechanisms to incorporate modified DNA bases into their genomes and protect their own genome from restriction (Labrie et al., 2010). However, modification-dependent REases evolved as a response to modified bases in the bacteriophage genome in the co-evolutionary arm race between bacteria and bacteriophages (Figure 1.18). It is a diverse group of restriction enzymes, which includes Type IIM ("M" stands for Modified DNA) and Type IV REases, characterized by different subunit constitution, biochemical properties and restriction specificity. Type IIM enzymes cleave DNA at constant positions at or close to the recognition sequence; meanwhile, Type IV cleavage sites are not well defined (Roberts et al., 2003). Type IIM REases include DpnI, MspJI family, PvuRts1I family enzymes; Mcr systems McrBC and McrA are members of the Type IV REase family. Modification-dependent REases have many practical applications in gene engineering and in the analysis of eukaryotic epigenomes (Lippman et al., 2004, 2005; Szwagierczak et al., 2011; Wang et al., 2011; Zheng et al., 2010).

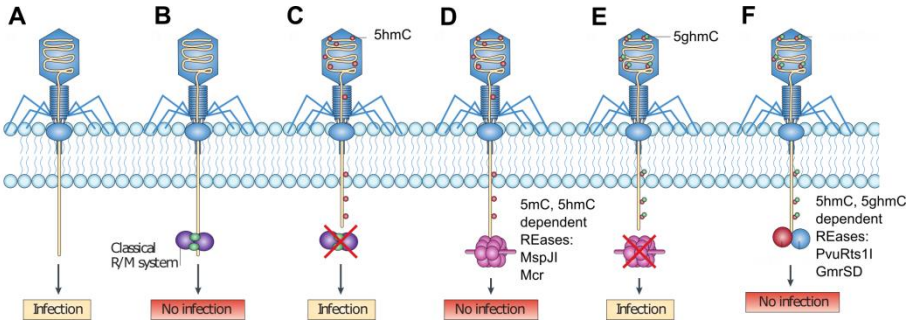


Figure 1.18 Co-evolutionary arms-race between bacteria and bacteriophages. (A) Phage infecting phage-sensitive bacteria. (B) Phage infecting phage-resistant bacteria with classical RM systems, which cleave invading phage DNA, thereby protecting host cell from infection. (C) Phage inserts 5hmC into its genome avoiding restriction. (D) Bacteria evolves methylation and hydroxymethylation dependent RM systems. (E) Phage acquires resistance to methylation and hydroxymethylation dependent RM systems through glucosylation of 5hmC. (F) Bacteria acquire restriction systems that cleave glucosylated DNA. Figure adapted from (Labrie et al., 2010)

1.3.2.1 Restriction endonuclease DpnI

One of the best characterized modification-dependent restriction enzyme is DpnI. It was isolated from *Streptococcus (Diplococcus) pneumoniae* (Lacks and Greenberg, 1975). In vivo DpnI protects bacteria from phages which are propagated in a Dam+ bacterial strain (de la Campa et al., 1988). Methyltransferase Dam methylates 6th position of adenine in a specific sequence GATC. Methylation of GATC sequence plays an important role in various cellular processes in bacteria, such as gene transcription, DNA mismatch repair, initiation of chromosome replication, nucleoid structure and virulence (Løbner-Olesen et al., 2005). DpnI recognizes G6mATC sequence and cleaves in the middle of it leaving blunt ends. The endonuclease prefers sequences methylated in both strands rather than hemimethylated (Lu et al., 2002). DpnI belongs to Type IIM restriction endonucleases (Roberts et al., 2003). Due to its unique properties DpnI are used in several molecular biology protocols, e. g. QuickChange site-directed mutagenesis (Zheng et al., 2004). DpnI hydrolyzes parental DNA molecules isolated from Dam+ hosts, which are without mutations, while newly synthesized PCR products with mutations are not methylated and remain uncleaved.

Crystal structure and biochemical analysis of DpnI revealed that it is a monomeric protein that consists of an N-terminal catalytic domain of the PD-(D/E)XK family and a C-terminal winged helix (wH) domain (Siwek et

al., 2012). The structures were solved with DNA bound to wH domain only, and with two duplexes bound to both domains (Figure 1.19 A) (Mierzejewska et al., 2014; Siwek et al., 2012). In the asymmetric complex composed of one enzyme molecule and two oligoduplexes both domains bind DNA, however only the catalytic domain can cleave DNA (Mierzejewska et al., 2014). The bases of the DNA duplexes in two different domains are paired canonically by Watson-Crick pairs including methylated bases. DNA binding makes the catalytic domain ordered and cover the DNA from both major and minor grooves (Figure 1.19 A). In the structure two methyl groups in the fully methylated substrate are very close to each other and they are bound together in one cleft of DpnI catalytic domain. The hydrophobic loop consisting of L129, R135 and W138 becomes organized when DNA is bound and Trp138 helps to recognize methylation making van der Waals contacts (Figure 1.19 B). Overall methylated target recognition by DpnI results from the combination of direct favorable interactions with the methyl groups, solvation / desolvation effects, indirect effects imposed by the proximity of the methyl groups on DNA conformation and crosstalk between the wH and catalytic domains (Mierzejewska et al., 2014).

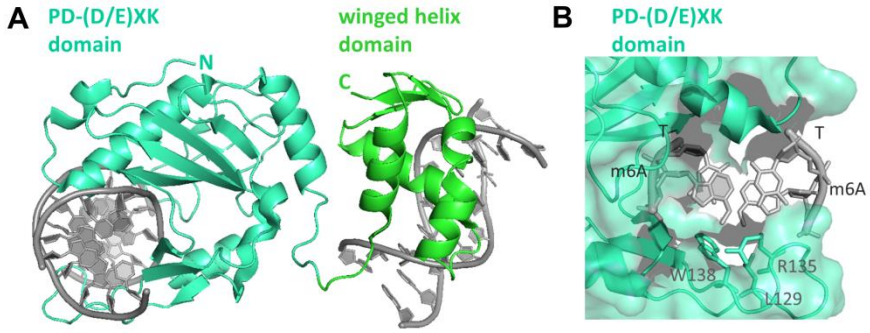


Figure 1.19 The crystal structure of DpnI. (A) The structure of DpnI with two DNA duplexes. PD-(D/E)XK and wH domains of the protein are shown in cyan and green, respectively, and the DNA is in gray; (B) Recognition of the 6mA by the catalytic domain of DpnI. The methyl groups are in contact with a long loop connecting adjacent antiparallel β-strands (PDB: 4KYW; (Mierzejewska et al., 2014)).

1.3.2.2 MspJI family restriction endonucleases

MspJI family restriction enzymes belong to Type IIM methyl-directed restriction endonucleases (Roberts et al., 2003). MspJI family include six characterized enzymes: MspJI, FspEI, LpnPI, AspBHI, RlaI and SgrTI (Cohen-Karni et al., 2011). These enzymes recognize 5mC and 5hmC (but

not 5ghmC) in various sequence contexts (Table 1.1) and cut both DNA strands 12/16 nt downstream of the modified cytosine (Zheng et al., 2010). MspJI enzymes are arranged as the N-terminal SRA-like domain and the C-terminal PD-(D/E)XK nuclease domain fusions (Cohen-Karni et al., 2011; Horton et al., 2012, 2014a). MspJI protects bacteria from phages which do not glucosylate DNA and after DNA synthesis, when DNA is not yet glucosylated.

Table 1.1 SRA domains and their recognition sequences.

Protein	Recognition sequence	Modification (mC)
UHRF1-SRA	5′-(mC)G-3′	5mC, 5hmC
UHRF2-SRA	5′-(mC)G-3′	5hmC > 5mC
SUVH5-SRA	5′-(mC)G-3′ / 5′-(mC)HH-3′	5mC
MspJI	5′-(mC)NNR-3′	5mC, 5hmC
AspBHI	5′-YS(mC)NS-3′	5mC, 5hmC
LpnPI	5′-C(mC)DG-3′	5mC, 5hmC
PvuRts1I	5′-(mC)-3′	5hmC, 5ghmC
AbaSI	5′-(mC)-3′	5hmC, 5ghmC

Table adapted from (Sasnauskas et al., 2015)

1.3.2.2.1 MspJI structure

Crystal structure and biochemical analysis of MspJI revealed that the protein forms an unusual homotetramer (Figure 1.20 A) (Horton et al., 2012, 2014b). Each monomer consist of two domains: N-terminal SRA-like domain linked with an ~10-residue connecting linker to the C-terminal PD-(D/E)XK endonuclease domain (Pingoud and Jeltsch, 2001). In the tetramer two dimers are in a different conformations: subunits A and B of the A:B dimer are located in ‘closed’ conformation where the binding and catalytic domains are close to each other; and subunits C and D in the C:D dimer are in ‘open’ conformation where the domains are farther apart with respect to each other (Figure 1.20 A). Two cleavage sites are formed from different subunits: A and D, also B and C. The cleavage site consists of two active sites facing each other for double-stranded DNA cleavage (Horton et al., 2014b).

The structures in the absence of DNA and in the presence of DNA are similar (Horton et al., 2012, 2014b). In the structure with DNA, each DNA molecule interacts with two adjacent tetramers and each tetramer interacts with two DNA molecules (Horton et al., 2014b). Only the ‘open’ conformation C-D dimer interacts with DNA (Figure 1.20 A). Authors propose this is due to crystal packing, rather than an intrinsic property of the protein. It was suggested that A-B dimer might act as allosteric activation

center (Horton et al., 2014b). However, specific interactions are formed only by the binding domain in the subunit C of one tetramer and non-specific interactions are formed by the subunit D of the symmetry-related tetramer. Both interactions, specific and non-specific, are involved in base flipping by SRA-like domain; in the specific interaction 5mC is flipped out of dsDNA (Figure 1.20 B) and in the non-specific interaction the G base is flipped. The base is flipped from the DNA helix and is accommodated in the SRA-like binding pocket and sandwiched between W101 and Y114 (Figure 1.20 B). In the pocket 5mC makes H-bonds with S90, D103, F115 and G116, and vdW contact with D117 (Figure 1.20 B). The residues F115 and G116 might be responsible for specificity for 5(h)mC but not C or T (Horton et al., 2014b). An unexpected feature is non-specific base flipping which might be a part of sequence-discrimination mechanism of MspJI like in some DNA repair enzymes which search for specific base by flipping (Parikh et al., 2000). MspJI lacks contacts in the major groove with 5mC therefore checking every base might be the case here (Horton et al., 2014b).

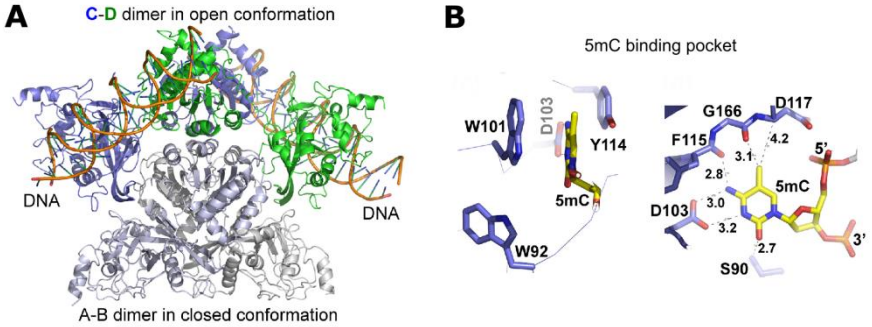


Figure 1.20 Crystal structure of MspJI-DNA complex. (A) MspJI co-crystal structure. One MspJI molecule interacts with two DNA molecules. Only C-D dimer makes contacts with DNA. A-B dimer is in a closed conformation, C-D dimer is in an open conformation. (B) The binding pocket of the SRA-like domain of MspJI contacts with flipped 5mC. Figure adapted from (Horton et al., 2012, 2014b).

Readout of the hemi-methylated 5mCNNR sequence by MspJI REase occurs through the minor and major groove contacts (Figure 1.21). Three loops are responsible for sequence recognition: loop-2B, loop-78 and loop-B3 (Figure 1.21). G65 from loop-B3 makes H-bond with the lonely guanine left after 5mC extrusion. Q33 from loop-2B interacts with base downstream from the flipped 5mC. Mutational and structural analysis showed that K173 from loop-78 is responsible for recognition of the third base downstream from the flipped 5mC which in the case of MspJI is a purine (R=A or G), and couples recognition and cleavage (Horton et al., 2014b).

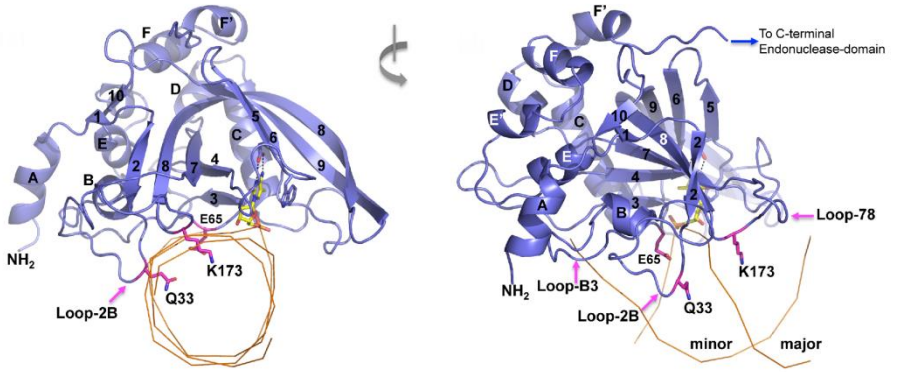


Figure 1.21 Sequence recognition by MspJI determinants. Loops-2B, -B3 and -78 are responsible for DNA recognition. Q33 and E65 are in the minor groove, K173 – in major groove. Flipped out 5mC is colored in yellow. Figure adapted from (Horton et al., 2014b).

1.3.2.2.2 DNA cleavage by MspJI

MspJI recognizes hemi-modified 5mCNNR or 5hmCNNR (R=purine, A or G) sequence and cuts 'top' strand 12 and 'bottom' strand 16 bases downstream from the modified cytosine (Figure 1.22 A) (Cohen-Karni et al., 2011; Zheng et al., 2010). Fully modified targets are recognized as two hemi-methylated sites each individually and cleavage occurs on both sides generating 4-nt 5′-overhangs (Figure 1.22 B) (Zheng et al., 2010). This unique feature enables application of MspJI in epigenomic studies, as it allows precise mapping of modifications in the genomic DNA. Cleavage of fully methylated CpG sites by MspJI generates ~32 bp fragments which can be sequenced to map the location of the modification on the DNA (Figure 1.22 B). MspJI family REases with different recognition specificities are promising tools for epigenomes studies.

Catalytic site consists of residues D334, Q355, D356 and K357 ($DX_{20}QAK$ motif – variant of the classical PD-(D/E)XK motif (Pingoud and Jeltsch, 2001)) and requires Mg^{2+} ions for activity (Horton et al., 2012; Zheng et al., 2010). In the co-crystal structure cleavage centers were not engaged because of one of two binding events with the same oligonucleotide was non-specific and instead of an interaction with catalytic center there was the interaction with recognition domain on the 'bottom' strand where the cleavage should be engaged (Figure 1.22 C). It might be due to crystal packing (Horton et al., 2014b).

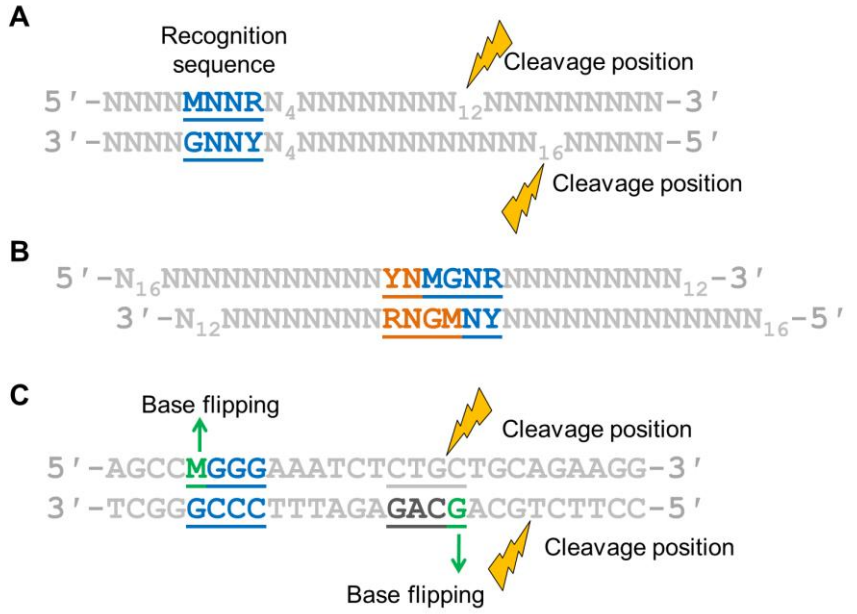


Figure 1.22 Sequence recognition and cleavage by MspJI. (A) MspJI cleaves DNA at N(12)/N(16) downstream from the modified cytosine; M=5mC, R=A or G, Y=T or C; (B) MspJI recognizes fully modified sequences as two hemi-methylated and cleaves fully modified CpG site generating 32 bp fragments with 4-nt 5′-overhangs; (C) In the co-crystal structure of MspJI with DNA oligonucleotide two bases are flipped out from specific (5mCGGG) and non-specific (GCAG) binding where the cleavage event should be engaged.

Despite the available co-crystal structure of MspJI, the DNA cleavage mechanism of MspJI is not yet fully understood. The proposed mechanism is based on modeling, biochemical and mutational analysis (Horton et al., 2012, 2014b). In the structure when two dimers A-B and C-D form tetramer, the catalytic sites from subunits A and D, also C and B get close forming two cleavage centers (each consists of two catalytic sites) for double-strand DNA cleavage. It was suggested that at least three different monomers of tetramer may be involved in recognition and cleavage: (i) DNA binding, (ii) cleavage in the same strand 12 nt downstream from modified cytosine, and (iii) the second strand cleavage 16 nt downstream (Horton et al., 2012, 2014b). In the co-crystal structure the cleavage center formed by A and D is positioned for double-strand DNA cleavage on ‘top’ strand by active center of D subunit and on ‘bottom’ strand by A (Figure 1.20 A). Modeling showed that DNA molecule bound specifically to recognition sequence of subunit D must curve for hydrolysis by cleavage center formed by C and B active sites (Horton et al., 2012, 2014b). The curved DNA molecule could be

accommodated in a channel on the surface of the protein between the binding sites and the catalytic centers.

For efficient DNA cleavage MspJI requires interaction with multiple recognition sites, either *in cis* or *in trans* (Zheng et al., 2010). Cleavage activity by MspJI reaches maximum when all four DNA recognition domains are specifically bound to recognition sequences (Horton et al., 2012, 2014b). In the co-crystal structure only 'open' conformation C-D dimer binding domains make contacts with DNA. It was suggested that 'closed' conformation A-B dimer might participate in the catalysis indirectly as allosteric activation center (Horton et al., 2014b).

Overall, MspJI tetramer has two DNA binding sites for recognition (C and D); two cleavage centers for double-strand DNA cleavage (active sites paired of A and D; C and B subunits); and two DNA binding sites for allosteric activation (A and B).

1.3.2.3 Mcr systems

Modification-dependent Type IV restriction enzymes McrA, Mrr and McrBC are found in strains of *E. coli* K12. These enzymes protect bacteria from T-even phages which has non-glucosylated 5hmC DNA and was named Rgl (for restricts glucose-less phages) (Revel, 1967). Later terminology has been changed to *mcr* (for modified cytosine restriction) and *rglA* and *rglB* were renamed *mcrA* and *mcrB*, respectively (Raleigh et al., 1989). The *mcrA* gene is located on the defective prophage-like element *e14* present in the *E. coli* K-12 genome and this element can be lost through exposure to ultraviolet light (Hiom and Sedgwick, 1991). *The mcrB*, *mcrC* and *mrr* (for methylated adenine recognition and restriction) genes are located in the same locus as type I REases *hsd* gene cluster. This region was named the Immigration Control Region because it protects bacteria from foreign nucleic acids "immigration" (Figure 1.23) (Bourniquel and Bickle, 2002).

Figure 1.23 The Immigration Control Region of *E. coli* K-12. In this region there are encoded three restriction enzymes systems: Type I RM system EcoKI, modification-dependent McrBC enzyme, and the Mrr enzyme.

Mcr and Mrr systems can be used to find plasmids coding DNA methyltransferases genes by inducing DNA-damage SOS response (Piekarowicz et al., 1991). DNA methylation studies (Burman et al., 1999;

Chotai and Payne, 1998; Lyko et al., 1999) were succeeded by microarray-based methods using McrBC endonuclease (Lippman et al., 2004, 2005).

1.3.2.3.1 Structure of restriction endonuclease McrBC

McrBC is a multi-subunit GTP-dependent complex which consists of two subunits: McrB which is responsible for modified DNA recognition, GTP binding and hydrolysis (Krüger et al., 1995; Pieper et al., 1997); and McrC which contains nuclease PD-(D/E)XK motif and is responsible for DNA cleavage (Figure 1.24) (Pieper and Pingoud, 2002).

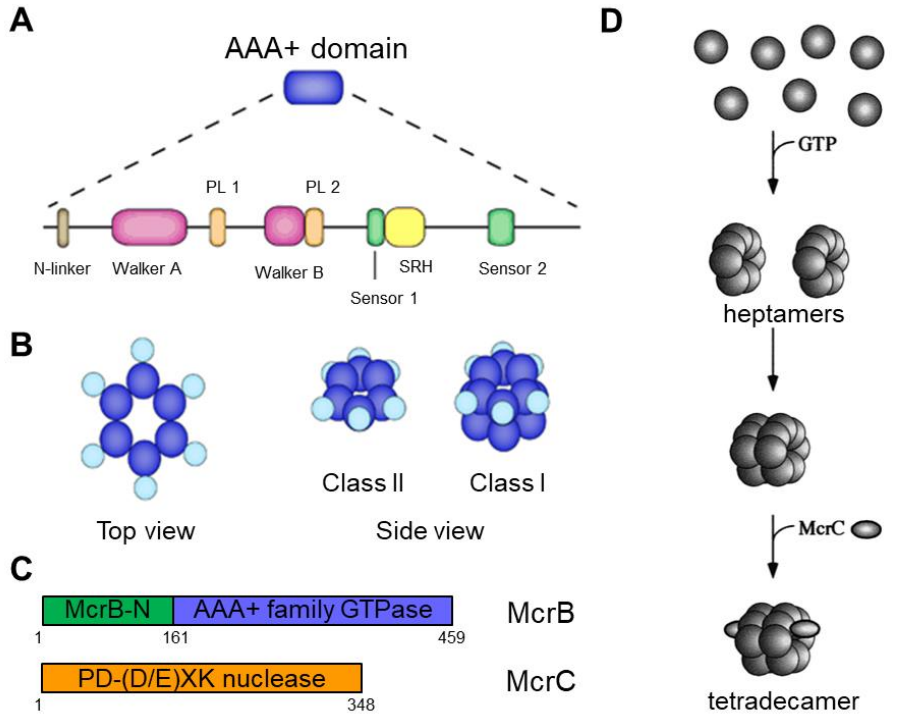


Figure 1.24 AAA+ domain containing proteins form oligomeric ring-shaped structures. (A) Conserved sequence motifs of the AAA+ domain which are responsible for ATP binding and hydrolysis (PL – pore loop, SRH – second region of homology). (B) Oligomerization. Class II AAA+ domains organize into hexamers and class I form double ring-shaped structures. (C) Domain composition of McrBC; (D) McrB oligomerizes into heptamers in the presence of GTP, and into tetradecamers in the presence of McrC. Figure adapted from (Panne et al., 2001; White and Lauring, 2007).

McrB harbors C-terminal AAA+ (ATPases Associated with various cellular Activities) domain which is found in motor proteins, chaperone-like ATPases (e. g. proteasome, dynein head, ClpX protease), proteins forming

complexes with DNA (e. g. prokaryotic DnaA, RuvB, NtrC, phage T4 clamp loader gp44/gp62) which are important in replication, recombination and transcription (Neuwald et al., 1999). AAA+ family proteins often are functional forming oligomeric ring-shaped structures (Figure 1.24 A) (White and Lauring, 2007). Also McrB form heptameric ring in the presence of Mg^{2+} and GTP, and in the presence of the McrC the two rings dimerize in the tetradecamer (Figure 1.24 B) (Panne et al., 2001).

N-terminal (1-161 aa) domain of McrB (Figure 1.24 C) recognizes modified DNA target (Gast et al., 1997). The co-crystal structure of the McrBC DNA binding domain McrB-N revealed that despite an unrelated tertiary structure, it follows the same mechanism for modified base recognition as the SRA domains: the base is flipped out into a protein pocket (Figure 1.25) (Sukackaite et al., 2012).The pocket accommodates 5mC, 4mC and 5hmC, but not 5ghmC, because the pocket is too small for glucosylated base (Sukackaite et al., 2012).

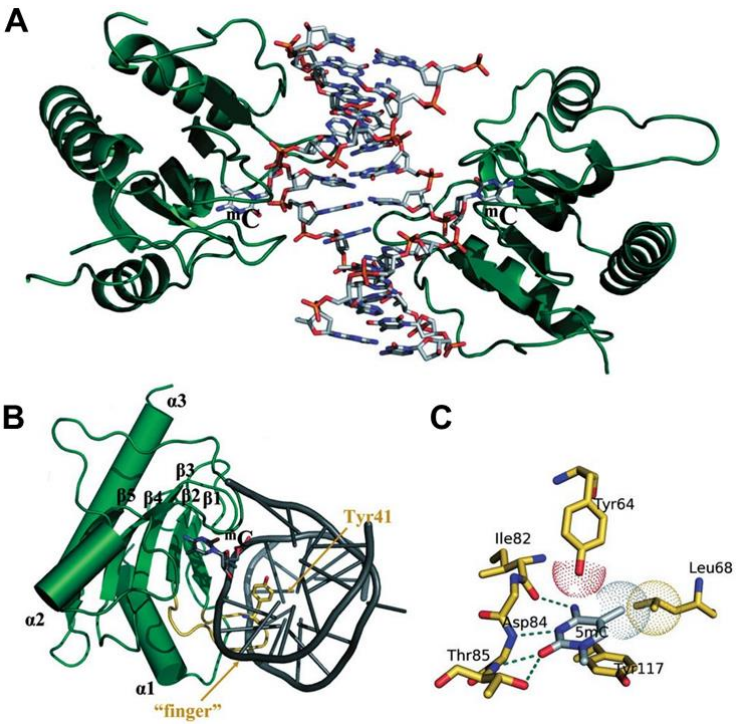


Figure 1.25. Crystal structure of McrB-N and DNA complex. (A) Overall structure of McrB-N bound methylated DNA. (B) The three loops participate in the sequence recognition from the minor groove; 'finger' loop and its Tyr41 colored in yellow. (C) 5mC binding pocket specific interactions. Figure adapted from (Sukackaite et al., 2012).

In the crystal two monomers are bound to A5mC sites on different strands (Figure 1.25 A). In the structure DNA is deformed, minor groove is wider and oligoduplex is bent by 29° towards the major groove. The three loops participate in the sequence recognition from the minor groove (Figure 1.25 B). The 'finger' loop sticks in the widened minor groove and Y41 intercalates into the DNA base stack flipping out 5mC (Figure 1.25 B). In the protein pocket Y117 stacks with the flipped out cytosine, the residues I82, D84 and T85 make H-bonds, Y64 and L68 make vdW interactions with extrahelical base (Figure 1.25 C). VdW interactions may be responsible for discrimination between the extrahelical 5mC and C bases and direct hydrogen bonding with 5mC explains incongruity of T, A and G (Sukackaite et al., 2012). Also the base flipping was confirmed in a solution with DNA oligoduplex containing fluorescent cytidine analogue, pyrrolo-dC (pyC) (Sukackaite et al., 2012).

1.3.2.3.2 DNA cleavage by McrBC

McrBC cleaves between two well-separated (30-3000 bp apart; however, optimal distance is 55-103 bp (Stewart and Raleigh, 1998)) recognition sites R5mC (where R = A or G; mC=5mC, 4mC or 5hmC) (Gast et al., 1997; Stewart and Raleigh, 1998; Sutherland et al., 1992). The cleavage occurs approximately at 10, 20, 30 (the main cut), 40 and 50 bp from one R5mC site. 10 bp periodicity represents helical repeat which shows that McrC attacks phosphodiester bonds presented on the surface of a large MrcBC complex (Pieper et al., 2002). DNA cleavage depends on McrB and McrC ratio which is optimal at 3-5:1 (McrB:McrC). In the cell this ratio is regulated by *mcrb* gene smaller product $McrB_S$ which is translating independently from full length product $McrB_L$ (L for large, 53 kDa; usually it is called McrB) (Krüger et al., 1992). $McrB_S$ (S for small, 34 kDa) is a smaller variant of McrB without N-terminal DNA recognition domain (1-161 amino acids) (Dila et al., 1990; Krüger et al., 1992; Ross et al., 1989). $McrB_S$ regulates McrB:McrC ratio binding excess McrC, thus forming protein complexes optimal for DNA cleavage (Figure 1.26) (Panne et al., 1998).

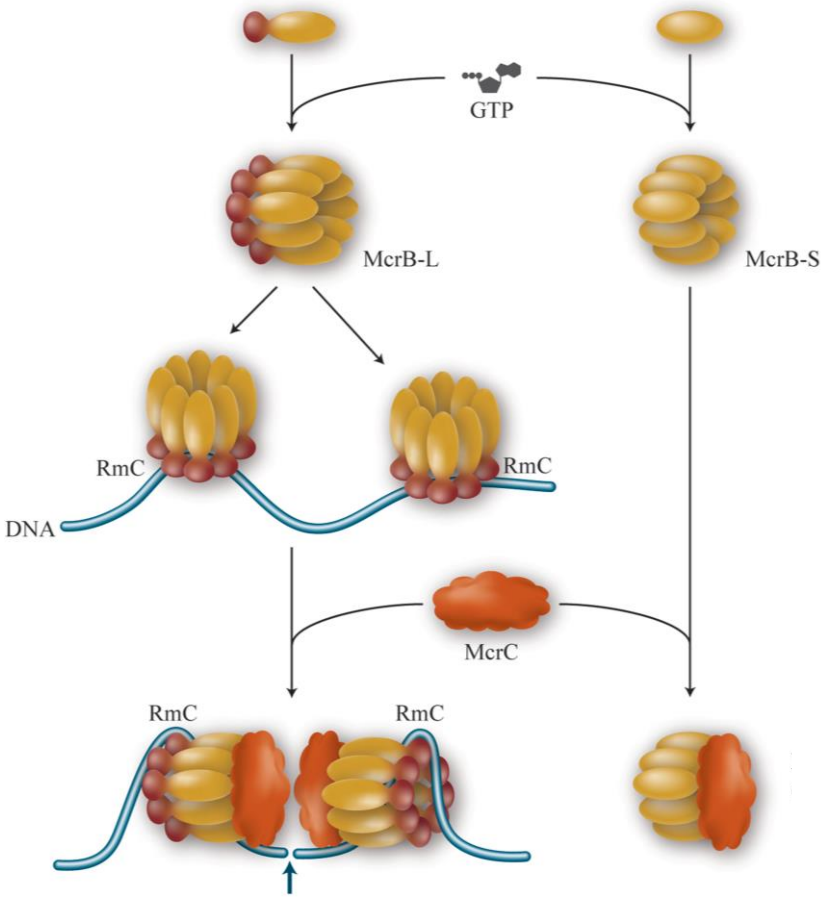


Figure 1.26 Model for the mechanism of action of McrBC. In the presence of GTP $McrB_L$ and $McrB_S$ form heptameric rings. $McrB_L$ heptameric rings bind specific R5mC sequences. In the presence of McrC two heptameric rings dimerize and form tetradecameric structure. DNA are translocating until two rings collide. Then two McrC dimerize and DNA is cleaved. $McrB_S$ regulates $McrB_L$:McrC ratio binding excess McrC. Figure adapted from (Loenen and Raleigh, 2014).

McrBC is the only one known nuclease which uses GTP but not ATP (Bourniquel and Bickle, 2002). Only GTP is required for DNA cleavage; neither GDP nor GMP, nor other NTP can be used instead of GTP; ATP inhibits reaction (Sutherland et al., 1992). GTP has several roles in McrBC: i) GTP and non-hydrolysable analogs (γ-S-GTP) increases McrB affinity for DNA (Krüger et al., 1995; Stewart et al., 2000); ii) GTP is required for McrB-DNA-McrC complex formation (Panne et al., 2001; Stewart et al., 2000); iii) McrB-McrC interaction increases GTPase activity 30 times (Pieper et al., 1997); iv) GTP hydrolysis is required for DNA translocation (Panne et al., 1999)

In the presence of GTP McrB form heptameric ring-shaped structure which binds DNA and then associates with McrC. It was suggested that DNA cleavage occurs when two DNA translocating McrBC complexes collide (Figure 1.26). GTP hydrolysis is required for DNA translocation and cleavage (Bourniquel and Bickle, 2002; Panne et al., 1999). However, some aspects of the mechanism of action of McrBC remain unclear such as translocation character or where is DNA – in or out of the ring.

1.3.2.3.3 Restriction endonuclease McrA

McrA (also known as EcoKMcrA) is a Type IV modification-dependent restriction system located on an easily excisable, defective prophage-like element *e14* present in the *E. coli* K-12 genome (Hiom and Sedgwick, 1991; Raleigh et al., 1989). Bioinformatics and mutagenesis analysis revealed that McrA might belong to the ßßα-Me finger superfamily of nucleases and consists of C-terminal HNH nuclease domain and unknown N-terminal a putative DNA recognition domain (Anton and Raleigh, 2004; Bujnicki et al., 2000). HNH motif is common in other restriction endonucleases, homing endonucleases and other nucleases involved in DNA recombination and repair (Cymerman et al., 2006; Gorbalenya, 1994; Stoddard, 2006).

Bioinformatics model of C-terminal domain showed that this region has a several metal coordinating regions: i) three histidine residues (H228, 252 and 256) might coordinate Mg^{2+} ion; ii) four cysteine residues (C207, 210, 248 and 251) form a putative zinc finger and might be involved in coordinating Zn^{2+} or some other divalent metal ion (Bujnicki et al., 2000).

In vivo McrA restricts 5mC and 5hmC containing DNA in appropriate sequence context (e. g. M.HpaII methylated C5mCGG context) (Fleischman et al., 1976; Raleigh and Wilson, 1986). EMSA experiments showed that McrA binds DNA substrates with symmetrically methylated and hemimethylated sequences Y5mCGR (where R=A or G; Y=C or T) (Mulligan et al., 2010). Unfortunately, McrA did not cleave the same substrates *in vitro* (Mulligan and Dunn, 2008). The mechanism of McrA restriction is still unsolved.

1.3.2.4 Glucosylation-dependent restriction endonucleases

Glucosyltransferases encoded by bacteriophages can modify 5-hydroxymethylated DNA producing 5-glucosylhydroxymethylcytosines. α-glucosyltransferase and β-glucosyltransferase incorporate hm-dCTP during replication using UDPG (uridine diphosphate glucose) produced by bacteria UDPG pyrophosphorylase. T-even phages in their genomes have different percentage combinations of 5-α-glucosylhydroxymethylcytosines (5αghmC)

and 5-β-glucosylhydroxymethylcytosines (5βghmC), e. g. T4 has 70 % of 5αghmC and 30 % of 5βghmC; other T-even phages have also βα-diglucosylated 5hmC (Kornberg et al., 1961). In bacteria evolved restriction endonucleases that recognize and cleave glucosylated DNA, such as PvuRts1I family enzymes and GmrSD complex (Figure 1.18 D) (Labrie et al., 2010). PvuRts1I family belongs to Type IIM REases (Roberts et al., 2003). PvuRts1I from plasmid Rts1 protects bacteria from T-even phages which have hydroxymethylated and glucosylated DNA, and discriminate against methylated DNA (Janosi et al., 1994). Another enzyme complex which restricts T-even phages glucosylated DNA is GmrSD (from glucose-modified restriction) composed of two components GmrS and GmrD proteins (Bair and Black, 2007). GmrSD cleaves only glucosylated DNA (5ghmC), but cannot cut methylated and hydroxymethylated DNA. It is Type IV restriction system and cut DNA in the presence of Ca^{2+} and UTP (Bair and Black, 2007). Some T4 type bacteriophages encode protein IPI which blocks GmrSD system and protects phage genome from restriction (Bair et al., 2007). Thus the arm race continues between bacteria and bacteriophages.

1.3.2.4.1 PvuRts1I family restriction endonucleases

PvuRts1I family restriction enzymes belong to Type IIM methyl-directed restriction endonucleases (Roberts et al., 2003). PvuRts1I family includes twenty characterized members, including PvuRts1I, AbaSI, BmeDI and others (Borgaro and Zhu, 2013). These enzymes recognize 5hmC as well as 5ghmC and for efficient cleavage require two cytosines on opposite strands separated by 21-22 nt and at least one of two must be modified (Janosi et al., 1994; Szwagierczak et al., 2011). PvuRts1I enzymes cleave double stranded DNA on the 3′ side of the modified cytosine on the top strand at $CN_{10–13}$ and on the bottom strand at $N_{9–11}G$ (Figure 1.27 A) (Szwagierczak et al., 2011).

PvuRts1I family enzymes have one more layer of specificity which includes not only modification of cytosine but also the glucose position. Some enzymes are specific for 5αghmC, and others for 5βghmC (Borgaro and Zhu, 2013). PvuRts1I family restriction enzymes specificity for 5hmC and 5ghmC, discrimination against 5mC and fixed cleavage position let provide tools for 5-hydroxymethylome profiles studies in mammalian genomes (Szwagierczak et al., 2011; Wang et al., 2011). For example, Aba-seq method has been used to map hydroxymethylome of mouse embryonic stem cells (Figure 1.27 B). This method has a lot of advantages such as high resolution, identification of very low 5hmC levels, a low quantity of the

sample DNA and no need of the chemical conversion as in the bisulphite method (Clark et al., 2006; Sun et al., 2013).

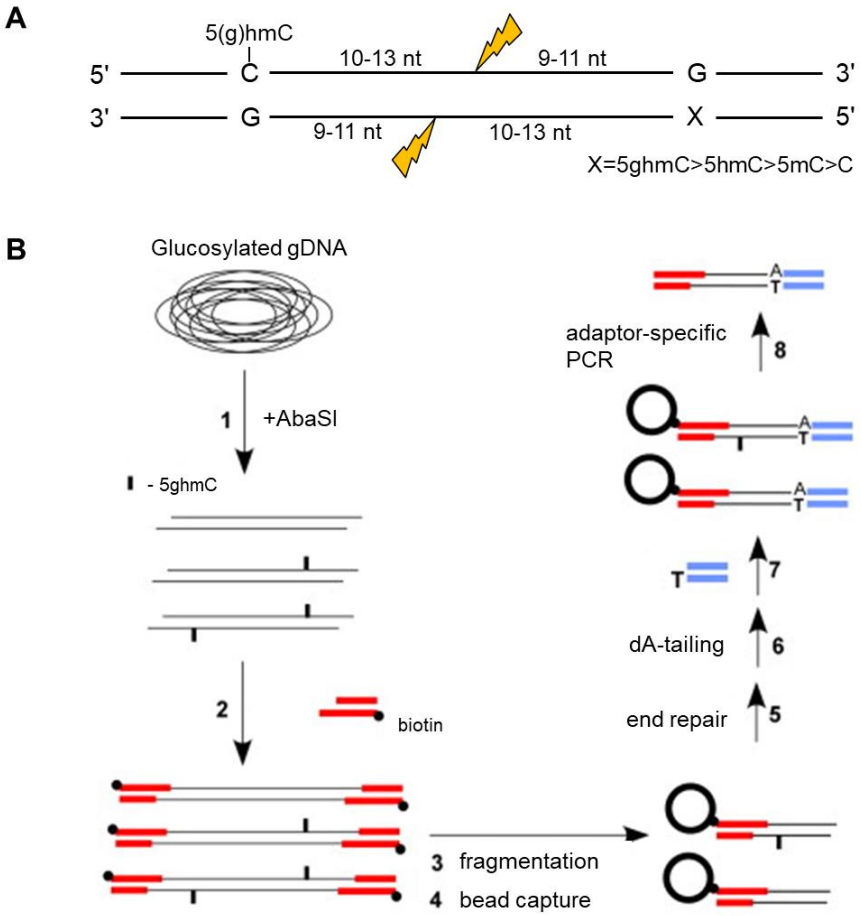


Figure 1.27 DNA cleavage by PvuRts1I family enzymes. (A) PvuRts1I enzymes cleave between two cytosines on the top strand at CN_{10-13} and on the bottom strand at $N_{9-11}G$. One of two cytosines must be 5hmC or 5ghmC. (B) Aba-seq method for hydroxymethylome mapping. Firstly, genomic DNA is glucosylated and then cleaved with AbaSI. Then fragmented DNA is captured, all needed adaptors ligated and after PCR the DNA can be sequenced. Figure adapted from (Sun et al., 2013)

1.3.2.4.2 PvuRts1I and AbaSI structures

It was shown that PvuRts1I family enzymes form dimers in solution (Borgaro and Zhu, 2013). However, PvuRts1I has been crystalized as monomer in the absence of DNA, meanwhile the crystal structure of AbaSI homodimer is present in the absence of DNA and homotetramer in the presence of DNA (Figure 1.28) (Horton et al., 2014c; Kazrani et al., 2014).

PvuRts1I monomeric structure has the same dimeric interface which is determined by the crystallographic symmetry. Bioinformatics analysis, modeling and structural data of PvuRts1I and AbaSI revealed that interaction interface between two molecules is formed by two α helices αA and αB and is mediated by hydrophobic interactions, hydrogen bonds and salt bridges (Figure 1.28 B) (Horton et al., 2014c; Kazrani et al., 2014; Shao et al., 2014).

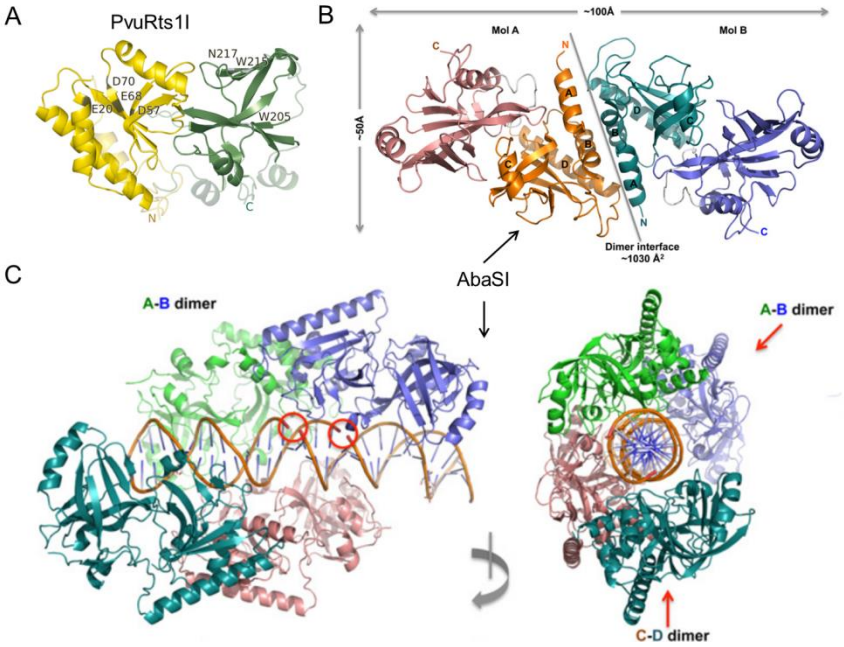


Figure 1.28 Crystal structures of PvuRts1I and AbaSI. (A) PvuRts1I monomer. Catalytic domain is in yellow and SRA domain is in green. Amino acids of active site and protein pocket are written. (B) AbaSI dimer. Two monomers of AbaSI A and B form dimer. Interaction interface between the two molecules is formed by two α helices αA and αB. Catalytic domains are in orange and cyan, SRA domains are in rose and blue. (C) AbaSI tetramer in the presence of DNA. Two dimers A-B and C-D form tetrameric structure and encircle DNA. Two views rotated ~90° along a vertical axis. Two red circles show missing phosphates which occurred in the DNA substrate preparation stage. Figure adapted from (Horton et al., 2014c; Kazrani et al., 2014; Shao et al., 2014).

The crystal structures of PvuRts1I family enzymes revealed that these enzymes are comprised of an N-terminal unusual PD-(D/E)XK endonuclease domain fused to a C-terminal SRA-like DNA-binding domain (Figure 1.29 A) (Bujnicki and Rychlewski, 2001; Horton et al., 2014c; Kazrani et al., 2014). In the PD-(D/E)XK motif of PvuRts1I aspartate take place instead of the lysine residue; in the AbaSI histidine substitute the lysine residue (Figure

1.29 B). However, another lysine residue found in proximity of the active site in both PvuRts1I (K17) and AbaSI (K23) (Figure 1.29 B) structures are important for activity, mutation of these lysine residues abolished activity (Horton et al., 2014c; Kazrani et al., 2014).

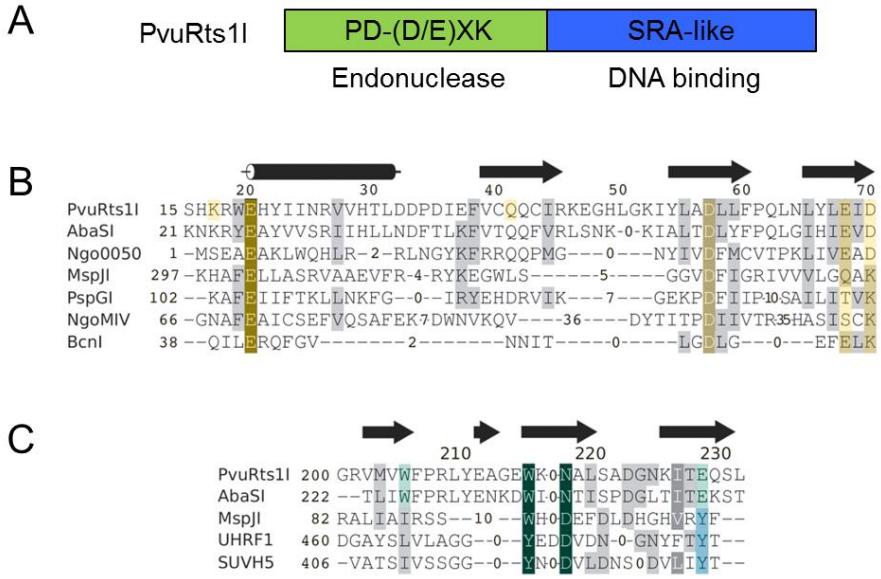


Figure 1.29 Domain composition and sequence alignment of PvuRts1I enzymes. (A) PvuRts1I family enzymes domain composition. (B) Sequence alignment of PD-(D/E)XK endonuclease domain of PvuRts1I and its homologs. (C) Sequence alignment of the SRA domain of PvuRts1I and its homologs. Figure adapted from (Kazrani et al., 2014).

Despite that AbaSI has been crystalized in the presence of DNA the modified base was not flipped out from DNA double helix (authors explain that it could be due to pre-flipped stage (Horton et al., 2014c)), and PvuRts1I has been crystalized as monomer, the C-terminal DNA binding domain of PvuRts1I family enzymes is homologous to eukaryotic UHRF1, SUVH5 and prokaryotic MspJI SRA domains which flip out modified base from double stranded DNA and accommodate it in the protein pocket (Figure 1.29 C) (Hashimoto et al., 2008; Horton et al., 2014b; Rajakumara et al., 2011a). The structures, structural models and mutational analysis of PvuRts1I family enzymes SRA-like domain suggest that these enzymes also flip out the modified cytosine. In the structures of the PvuRts1I enzymes there is a conserved pocket in which could fit 5hmC and 5ghmC (Horton et al., 2014c; Kazrani et al., 2014). However, the pocket of PvuRt1I enzymes differs from others in the SRA-like domains. In the PvuRts1I structure it is more rigid

because instead of the two glycine residues there are tyrosine and alanine residues. W205 and W215 might be responsible for flipped base stacking interactions. Y210, N217 and Q228 also play important role in the modified cytosine recognition in the pocket of the PvuRts1I (Kazrani et al., 2014; Shao et al., 2014). In the AbaSI structure there is also the conserved pocket named "4W pocket" because of the tryptophan residues which are in the pocket and might make some important contacts with the flipped out base. W234 and W224 might have role in the stacking interactions, while N236, Q247, R227 are candidates for the H-bonding (Figure 1.30 B) (Horton et al., 2014c). Another difference between PvuRts1I family SRA-like domains and eukaryotic SRA domains or MspJI family SRA-like domains is that the loops responsible for the sequence context recognition and base flipping ('NKR finger' and 'thumb' loop) are shorter and cannot reach some parts of the substrate as in the UHRF1 or SUVH5 (Figure 1.30 A) (Horton et al., 2014c; Kazrani et al., 2014; Shao et al., 2014).

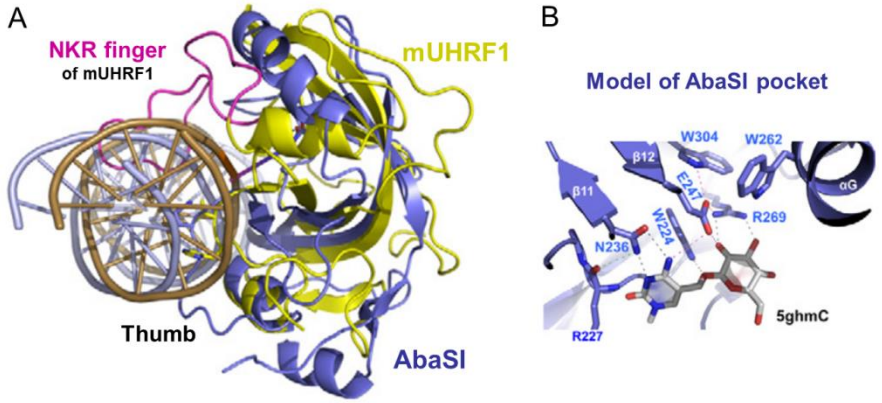


Figure 1.30 A model of AbaSI in complex with modified DNA. (A) Superimposition of SRA domains of AbaSI (in blue) and mUHRF1 (in yellow). NKR finger of mUHRF1 colored in magenta is in the major groove. DNA in the mUHRF1 colored in silver and in AbaSI in brown; (B) Model of AbaSI pocket with 5ghmC. Putative interactions are in dashed lines and amino acids are written. Figure adapted from (Horton et al., 2014c).

2 MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals

All chemicals used in this study were of the highest quality available.

2.1.2 Enzymes

T4 polynucleotide kinase; T4 DNA ligase; FastAP thermosensitive alkaline phosphatase; DNA polymerases; bovine serum albumin and restriction enzymes used in this study were purchased from Thermo Fisher Scientific. All these products were used according to recommendations of the manufacturer.

2.1.3 *E. coli* strains

ER2267: F′ proA+B+ lacIq Δ(lacZ)M15 zzf::mini-Tn10 (KanR)/ Δ(argF-lacZ)U169 glnV44 e14-(McrA-) rfbD1? recA1 relA1? endA1 spoT1? thi-1 Δ(mcrC-mrr)114::IS10

ER2566: F–λ–fhuA2[lon]ompT lacZ::T7 gene1 gal sulA11.(mcrC-mrr)114::IS10 R(mcr-73::miniTn10-TetS)2 R(zgb-210::Tn10)(TetS) endA1 [dcm].

BL21(DE3): F^- ompT gal dcm lon $hsdS_B(r_B^-m_B^-)$ λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) $[malB^+]_{K-12}(\lambda^S)$

DH10B: F^- endA1 $deoR^+$ recA1 galE15 galK16 nupG rpsL Δ(lac)X74 φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) Str^R λ^-

2.1.4 Buffers and solutions

Sonication buffer: 20 mM Tris-HCl (pH 7.5-8.0 at 25 °C), 500 mM NaCl, 7 mM 2-mercaptoethanol, 2 mM phenylmethylsulfonyl fluoride, 10-25 mM imidazole;

Column buffer: 10 mM Tris-acetate (pH 8.0 at 25 °C), 500 mM K-acetate;

Storage buffer I: 20 mM Tris-HCl (pH 7.5–8.5 at 25 °C), 200 mM KCl, 1 mM DTT, 50 % v/v glycerol;

Storage buffer II: 20 mM Tris-acetate (pH 7.5 at 25 °C), 250 mM K-acetate, 1 mM DTT and 50 % v/v glycerol;

Binding buffer I: 40 mM Tris-acetate (pH 8.3 at 25 °C), 0.1 mg/ml BSA, 10 % v/v glycerol;

Binding buffer II: 30 mM Mes-histidine (pH 6.3 at 25 °C), 0.1 mg/ml BSA, 10 % v/v glycerol;

Electrophoresis buffer I: 40 mM Tris-acetate (pH 8.3 at 25 °C);

Electrophoresis buffer II: 30 mM Mes-histidine (pH 6.3 at 25 °C);

Acrylamide solution: acrylamide/N,N'-methylenebisacrylamide (29:1 w/w);

Loading dye solution I: 95 % (v/v) formamide, 25 mM EDTA (pH 8.0 at 25 °C), 0.01 % bromphenol blue (w/v);

Loading dye solution II: 75 mM EDTA (pH 8.0 at 25 °C), 0.01 % bromphenol blue (w/v), 0.1 % SDS, 50 % (v/v) glycerol;

TBE buffer: 0.1 M Tris-H_3BO_3, 2 mM EDTA (pH 8.3 at 25 °C);

Reaction buffer I: 33 mM Tris-acetate (pH 8.0 at 25 °C), 66 mM K-acetate, 10 mM Mg-acetate, 0.1 mg/ml BSA;

Reaction buffer II: 50 mM Tris–HCl (pH 7.8 at 37 °C), 50 mM KCl, 0.1 mg/ml BSA;

NBE buffer: 100 mM H_3BO_3-NaOH, 15 mM CH_3COONa, 2 mM EDTA (pH 8.2 at 25 °C);

SDS loading solution: 100 mM Tris-HCl (pH 6.8 at 25 °C), 4 % SDS (w/v), 200 mM DTT, 20 % (v/v) glycerol;

4 % stacking gel: 125 mM Tris-HCl (pH 6.8 at 25 °C), 0.1 % SDS (w/v), 4 % acrylamide/N,N'-methylenebisacrylamide (37.5:1 w/w);

12 % separating gel: 375 mM Tris-HCl (pH 8.8 at 25 °C), 0.1 % SDS (w/v), 12 % acrylamide/N,N'-methylenebisacrylamide (37.5:1 (w/w);

SDS-PAGE buffer: 25 mM Tris, 190 mM glycine (pH 8.3 at 25 °C), 0.1 % SDS (w/v).

2.1.5 Plasmids and DNA

pTYB2 (Ap^R) (New England Biolabs) expression vector was used for cloning YkrI and BmeDI, pLATE11 (Ap^R) (Thermo Fisher Scientific) for LpnPI, LpnPI-N and mutants, pLATE31 (Ap^R) (ThermoFisher Scientific) for McrA, pBAD24 (Ap^R) for McrB-N.

pACYC184 (Cm^R) and pACYC184_M.HpaII (Cm^R) plasmids were used for plasmid restriction experiments with McrA.

Genomic DNA of *Legionella pneumophila* Philadelphia-1 (DSM No. 7513), *Yersinia kristensenii* (DSM No. 18543), and *Bacillus megaterium* (DSM No. 319) were purchased from Leibnitz Institute DSMZ (Germany).

2.1.6 DNA oligonucleotides

Oligoduplex substrates used in this study are listed in Table 2.1. Oligonucleotides with 5fC, 5caC and 4mC modifications were purchased

from IBA, all other oligonucleotides were from Metabion. Oligonucleotides were labeled with [γ-^{33}P]ATP or [γ-^{32}P]ATP (Hartmann Analytic, PerkinElmer) and T4 polynucleotide kinase (Thermo Fisher Scientific). Oligoduplexes were assembled by annealing the corresponding radiolabeled and unlabeled strands.

Table 2.1. Oligonucleotides used in this study.

Oligonucleotide	Sequence	Specification
16-M	5′-CGTAGC**M**TGGTCGATC-3′ 3′-GCATCGGACCAGCTAG-5′	A short cognate LpnPI oligoduplex. M designates 5mC. Modified base are shown in boldface, recognition sequence are underlined.
16-C	5′-CGTAGCCTGGTCGATC-3′ 3′-GCATCGGACCAGCTAG-5′	As 16-M but 5mC is replaced with an unmodified C.
16-P	5′-CGTAGC**P**TGGTCGATC-3′ 3′-GCATCGGACCAGCTAG-5′	As 16-M but 5mC is replaced with a pyC (P designates pyC).
16-T	5′-CGTAGC**T**TGGTCGATC-3′ 3′-GCATCG**G**ACCAGCTAG-5′	As 16-M but 5mC is replaced with a T. Cognate LpnPI oligoduplex with T-G mismatch (in boldface).
16-T-N	5′-CGTAGATCTTACGATC-3′ 3′-GCATCTGGAATGCTAG-5′	Noncognate LpnPI oligoduplex with a T-G mismatch (underlined).
30-M/ gC(mC)TG	5′-CCGTAGC**M**TGGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATCGGACCAGCTAGGATCGACCAGCGG-5′	An extended cognate LpnPI oligoduplex. M/mC designates 5mC.
30-C/ gCCTG	5′-CCGTAGCCTGGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATCGGACCAGCTAGGATCGACCAGCGG-5′	As 30-M but 5mC is replaced with an unmodified C.
30-P	5′-CCGTAGC**P**TGGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATCGGACCAGCTAGGATCGACCAGCGG-5′	As 30-M but 5mC is replaced with a pyC.
30-T	5′-CCGTAGC**T**TGGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATCG**G**ACCAGCTAGGATCGACCAGCGG-5′	As 30-M but 5mC is replaced with a T.
39-H/H	5′-CGACGATA**H**TTACCGGATAAGGCGCAATTAGATTACTTC-3′ 3′-GCTGCTATGAATGGCCTATTCCGCGTTAAT**H**TAATGAAG-5′	Optimal YkrI and BmeDI substrate with two 5hmC-G bp. H designates 5hmC.
39-M/H	5′-CGACGATA**M**TTACCGGATAAGGCGCAATTAGATTACTTC-3′ 3′-GCTGCTATGAATGGCCTATTCCGCGTTAAT**H**TAATGAAG-5′	As 39-H/H but one 5hmC is replaced with a 5mC
39-C/H	5′-CGACGATACTTACCGGATAAGGCGCAATTAGATTACTTC-3′ 3′-GCTGCTATGAATGGCCTATTCCGCGTTAAT**H**TAATGAAG-5′	As 39-H/H but one 5hmC is replaced with an unmodified C
39-P/H	5′-CGACGATA**P**TTACCGGATAAGGCGCAATTAGATTACTTC-3′ 3′-GCTGCTATGAATGGCCTATTCCGCGTTAAT**H**TAATGAAG-5′	As 39-H/H but one 5hmC is replaced with a pyC

See next page for Table 2.1 extension

Extension of the Table 2.1

39-T/H	5′-CGACGATA**T**TTACCGGATAAGGCGCAATTAGATTACTTC-3′ 3′-GCTGCTAT**G**AATGGCCTATTCCGCGTTAAT**H**TAATGAAG-5′	As 39-H/H but one 5hmC is replaced with a T generating mismatch.
39-H	5′-CGACGATATTTACCGGATAAGGCGCAATTAGATTACTTC-3′ 3′-GCTGCTATAAATGGCCTATTCCGCGTTAAT**H**TAATGAAG-5′	As 39-H/H but one 5hmC-G base pair is replaced with a T-A base pair
39-T	5′-CGACGATA**T**TTACCGGATAAGGCGCAATTAAATTACTTC-3′ 3′-GCTGCTAT**G**AATGGCCTATTCCGCGTTAATTTAATGAAG-5	As 39-T/H but the 5hmC-G base pair is replaced with a T-A base pair
31-C	5′-TGACCCACGCTCGCCCGGCGACACATTACGT-3′ 3′-ACTGGGTGCGAGCGGGCCGCTGTGTAATGCA-5′	Cognate Ecl18kI oligoduplex with a C-G central base pair
31-M	5′-TGACCCACGCTCGCC**M**GGCGACACATTACGT-3′ 3′-ACTGGGTGCGAGCGGGCCGCTGTGTAATGCA-5′	As 31-C but with a central 5mC-G base pair
tC(mC)TG	5′-CCGTA***T***C**M**TGGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCAT***A***GGACCAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the -2 bp (upstream of 5mC) is T:A. DNA base pairs that deviate from the reference substrate ‘gC(mC)TG’ are shown in ***bold italic*** typeface.
aC(mC)TG	5′-CCGTA***A***C**M**TGGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCAT***T***GGACCAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the -2 bp is A:T
cC(mC)TG	5′-CCGTA***C***C**M**TGGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCAT***G***GGACCAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the -2 bp is C:G
gG(mC)TG	5′-CCGTAG***G***M**M**TGGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATC***C***GACCAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the -1 bp is G:C
gT(mC)TG	5′-CCGTAG***T***M**M**TGGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATC***A***GACCAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the -1 bp is T:A
gA(mC)TG	5′-CCGTAG***A***M**M**TGGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATC***T***GACCAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the -1 bp is A:T
gC(mC)AG	5′-CCGTAGC**M*****A***GGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATCGG***T***CCAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the +1 bp (downstream of 5mC) is A:T
gC(mC)CG	5′-CCGTAGC**M*****C***GGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATCGG***G***CCAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the +1 bp is C:G
gC(mC)GG	5′-CCGTAGC**M*****G***GGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATCGG***C***CCAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the +1 bp is G:C
gC(mC)TC	5′-CCGTAGC**M**T***C***GTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATCGGA***G***CAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the +2 bp is C:G

See next page for Table 2.1 extension

Extension of the Table 2.1

gC(mC)TA	5′-CCGTAGC**M**T***A***GTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATCGGA***T***CAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the +2 bp is A:T
gC(mC)TT	5′-CCGTAGC**M**T***T***GTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATCGGA***A***CAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the +2 bp is T:A
gG(mC)TC	5′-CCGTAG***G***M**T*****C***GTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATC***C***GA***G***CAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the -2 bp is G:C and the +2 bp is C:G
gT(mC)TT	5′-CCGTAG***T***M**T*****T***GTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATC***A***GA***A***CAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the -2 bp is T:Aand the +2 bp is T:A
gA(mC)TA	5′-CCGTAG***A***M**T*****A***GTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATC***T***GA***T***CAGCTAGGATCGACCAGCGG-5′	As gC(mC)TG, but the -2 bp is A:T and the +2 bp is A:T
12_5mC	5′-AGCTA**X**CGGTCTC-3′ 3′-CGATGGCCAGAGT-5′	A 12 bp oligoduplex with 1 nt 5′-terminal extrusions and various cytosine variants at the position marked by ‘X’. Here X=5mC. McrB-N recognition sequence is underlined. Used in fluorescence anisotropy measurements (5′-terminus of the bottom strand contained a HEX modification).
12_C	5′-AGCTA**X**CGGTCTC-3′ 3′-CGATGGCCAGAGT-5′	X=C
12_5hmC	5′-AGCTA**X**CGGTCTC-3′ 3′-CGATGGCCAGAGT-5′	X=5hmC
12_5fC	5′-AGCTA**X**CGGTCTC-3′ 3′-CGATGGCCAGAGT-5′	X=5fC
12_5caC	5′-AGCTA**X**CGGTCTC-3′ 3′-CGATGGCCAGAGT-5′	X=5caC
12_4mC	5′-AGCTA**X**CGGTCTC-3′ 3′-CGATGGCCAGAGT-5′	X=4mC
13_C	5′-AGCTA**X**CGGTCTC-3′ 3′-TCGATGGCCAGAG-5′	Same as 12_X series, except that DNA strands form blunt-ended 13 bp duplexes. Oligoduplexes used in EMSA and EMSA competition assay. X=C
13_5mC	5′-AGCTA**X**CGGTCTC-3′ 3′-TCGATGGCCAGAG-5′	X=5mC
13_5hmC	5′-AGCTA**X**CGGTCTC-3′ 3′-TCGATGGCCAGAG-5′	X=5hmC

See next page for Table 2.1 extension

Extension of the Table 2.1

13_5fC	5′-AGCTA**X**CGGTCTC-3′ 3′-TCGATGGCCAGAG-5′	X=5fC
13_5caC	5′-AGCTA**X**CGGTCTC-3′ 3′-TCGATGGCCAGAG-5′	X=5caC
13_4mC	5′-AGCTA**X**CGGTCTC-3′ 3′-TCGATGGCCAGAG-5′	X=4mC
12_C/12_C	5′-ACCTC**X**GGTTCC-3′ 3′-TGGAGGCCAAGG-5′	12 bp oligoduplexes series for McrA (recognition sequence underlined). Used in EMSA and fluorescence measurements. X=C
12_5mC/12_C	5′-ACCTC**X**GGTTCC-3′ 3′-TGGAGGCCAAGG-5′	X=5mC
12_pC/12_C	5′-ACCTC**X**GGTTCC-3′ 3′-TGGAGGCCAAGG-5′	X=pyC
30_C/30_C	5′-AGACCCACGCTCAC**X**GGTTCCAGATTTATC-3′ 3′-TCTGGGTGCGAGTGG**X**CAAGGTCTAAATAG-5′	30 bp oligoduplexes series for McrA. Used for EMSA and DNA cleavage experiments X=C/C
30_5mC/30_C	5′-AGACCCACGCTCAC**X**GGTTCCAGATTTATC-3′ 3′-TCTGGGTGCGAGTGG**X**CAAGGTCTAAATAG-5′	X=5mC/C
30_5mC/30_5mC	5′-AGACCCACGCTCAC**X**GGTTCCAGATTTATC-3′ 3′-TCTGGGTGCGAGTGG**X**CAAGGTCTAAATAG-5′	X=5mC/5mC
T4, T5, T6, T7	5′-AGAC-3′ 5′-AGACC-3′ 5′-AGACCC-3′ 5′-AGACCCA-3′	Used as size markers to monitor top strand cleavage of the 30-mer oligoduplexes by McrA.
B22, B23, B24, B25, B26, B28	3′-TGGGTGCGAGTGGCCAAGGTCTAAATAG-5′ 3′-GGTGCGAGTGGCCAAGGTCTAAATAG-5′ 3′-GTGCGAGTGGCCAAGGTCTAAATAG-5′ 3′-TGCGAGTGGCCAAGGTCTAAATAG-5′ 3′-GCGAGTGGCCAAGGTCTAAATAG-5′ 3′-CGAGTGGCCAAGGTCTAAATAG-5′	Used as size markers to monitor bottom strand cleavage of the 30-mer oligoduplexes.
30′_C/30′_C	5′-GATTTATCAGACCCACGCTCAC**X**GGTTCCA-3′ 3′-CTAAATAGTCTGGGTGCGAGTGGCCAAGGT-5′	Alternative 30 bp oligoduplexes used for DNA cleavage experiments. X=C
30′_5mC/30′_C	5′-GATTTATCAGACCCACGCTCAC**X**GGTTCCA-3′ 3′-CTAAATAGTCTGGGTGCGAGTGGCCAAGGT-5′	X=5mC

See next page for Table 2.1 extension

Extension of the Table 2.1

T14′	5′-GATTTATCAGACCC-3′	Used as a size marker to monitor top strand cleavage of the 30′-mer oligoduplexes
B17′	3′-GTGCGAGTGGCCAAGGT-5′	Used as a size marker to monitor bottom strand cleavage of the 30′ -mer oligoduplexes
50_C/50_C	5′-CAGATTTATCAGACCCACGCTCAC**X**GGTTCCAGATTTATCGATGGTTAAC-3′ 3′-GTCTAAATAGTCTGGGTGCGAGTGGCCAAGGTCTAAATAGCTACCAATTG-5′	50 bp oligoduplex for McrA; Used in DNase I footprinting experiments. X=C
50_5mC/50_C	5′-CAGATTTATCAGACCCACGCTCAC**X**GGTTCCAGATTTATCGATGGTTAAC-3′ 3′-GTCTAAATAGTCTGGGTGCGAGTGGCCAAGGTCTAAATAGCTACCAATTG-5′	X=5mC

2.2 Methods

2.2.1 Electrophoresis

2.2.1.1 Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel electrophoresis (PAGE) was employed in gel-shift experiments (see 2.2.6.1). The gels consisted of 8 % Acrylamide solution in either Electrophoresis buffer I or II, polymerisation was initiated by adding TEMED and ammonium persulfate. The gel dimensions (length:width:thickness) were 22:15:0.1 cm. Prior to gel casting, one of the glass plates was processed with "bind silane" (3-methacryloxypropyltrimethoxysilane) and the other with "repeal silane" (5 % v/v dichlorodimethylsilane in $CHCl_3$). Samples of radiolabeled DNA and protein in either Binding buffer I or II were mixed in the wells of ELISA plate, left for 15 min at room temperature and then loaded on the gel. Electrophoresis was run at room temperature for 45-90 min at ~5 V/cm. Low power consumption during electrophoresis runs (~2 W or 110 V × ~18 mA per electrophoretic unit containing two gels and 1 l of electrophoretic buffer) ensured that the gels remained at room temperature (~22 °C), well below the melting temperature of the oligoduplexes used in the assay (>40 °C). After electrophoresis the glass treated with "repeal silane" was removed and the gel was dried on the glass treated with "bind silane" under a hot air flow. Radiolabeled DNA and protein-DNA complexes were detected in the dried gels using BAS-MS image plates (FujiFilm) and Cyclone phosphorimager (Perkin Elmer). The amounts of various DNA fragments were quantified with OptiQuant software (Perkin Elmer).

2.2.1.2 Denaturing polyacrylamide gel electrophoresis

Denaturing PAGE was employed in oligonucleotide DNA cleavage studies and DNase I footprinting experiments (see 2.2.7.1 and 2.2.8). Electrophoresis was performed in TBE buffer (see 2.1.4). The standard 22×15×0.1 cm gels consisted of 20 % Acrylamide solution (see 2.1.4) supplemented with 8.5 M urea in the electrophoretic buffer.

Samples of radiolabeled DNA were mixed with Loading dye solution I (see 2.1.4), placed for 4 min into a hot (95 °C) water bath and then placed on ice. Electrophoresis was run for 30 min at 30 V/cm without the samples and for another 1-2 hours with the DNA samples.

DNase I footprinting and some McrA cleavage experiments samples were analyzed on high resolution denaturing polyacrylamide gels (gel dimensions,

length:width:thickness, 50×20×0.02 cm). Gel composition was the same as described above. Electrophoresis was run for 20 min at 50 V/cm without the samples and for another 1-2 hours with the DNA samples. During electrophoresis, the gel was thermostated at 60 °C. Radiolabeled DNA was detected as described in 2.2.1.1.

2.2.1.3 Non-denaturing electrophoresis through agarose

Separation of different DNA fragments or different forms of plasmid DNA was performed in 0.8 % agarose gels in the NBE buffer (see 2.1.4) supplemented with 0.5 μg/ml ethidium bromide. DNA samples were mixed with 1/3 volume of Loading dye solution II (see 2.1.4) and electrophoresed at 3 V/cm until the bromphenol blue dye migrated for 3 cm. Digital images of the gels were taken with BioDocAnalyze gel documenting systems (Biometra).

2.2.1.4 Denaturing (SDS) polyacrylamide gel electrophoresis of proteins

Denaturing SDS-PAGE of proteins was employed to monitor protein purification steps and verify homogeneity of proteins. Protein samples were mixed at 1:1 (v/v) ratios with the SDS loading solution (see 2.1.4) and denatured at 95 °C for 5 min. Polyacrylamide gel comprised of 4 % stacking and 12 % separating gel (see 2.1.4) layers. Electrophoresis was run in SDS-PAGE buffer (see 2.1.4) at room temperature for 1-1.5 hours at ~30 V/cm. Gels were stained with Page Blue protein staining solution (Thermo Fisher Scientific).

2.2.2 Protein cloning and expression

The genes encoding LpnPI, YkrI, and BmeDI were amplified by PCR from the genomic DNA of *Legionella pneumophila* Philadelphia-1 (DSM No. 7513), *Yersinia kristensenii* (DSM No. 18543), and *Bacillus megaterium* (DSM No. 319), respectively. Genes of LpnPI and its N-terminal DNA binding domain (LpnPI-N, corresponds to 1-224 residues of the full length protein) were cloned into the pLATE11 expression vector. The first methionine in both proteins was replaced by a short hexahistidine tag (sequence MGHHHHHHG).

Genes encoding YkrI and BmeDI were cloned into the pTYB2 expression vector as C-terminal fusions with the self-cleavable chitin binding domain (CBD).

The expression vectors for McrA with C-terminal hexahistidine tag were made by ligating the *mcrA* gene into the pLATE31 vector. The pLATE31

plasmid was also used for the expression construct of the McrA N-terminal fragment (McrA-N, corresponds to 1-174 residues of the full length protein).

LpnPI and its mutants (generated by QuickChange Mutagenesis, see 2.2.3), YkrI, BmeDI and McrA were expressed in the *E. coli* strain ER2566. Cells were grown to OD_{600} 0.5–0.8 and induced with a final concentration of 0.2 mM IPTG at 16 °C overnight, harvested by centrifugation and stored at −20 °C.

The region coding for the 1–161 residues of McrB (McrB-N) was PCR-amplified from the pBBI/McrB template, introducing a hexahistidine tag at the C-terminus, and inserted into the pBAD24 vector (Sukackaite et al., 2012). WT McrB-N and mutant variants (generated by QuickChange Mutagenesis, see 2.2.3) were expressed in *E. coli* DH10B (ara-) strain. Cells were grown in LB medium at 37 °C to A_{600} 0.5–0.8, and protein expression was then induced by adding arabinose to a final 0.2 % concentration. After 4 h cells were collected by centrifugation and stored at -20 °C.

2.2.3 Mutagenesis

His-tagged full-length LpnPI and McrB-N mutant variants were generated by the QuickChange method (Zheng et al., 2004) using pLATE11_N-HisTag_LpnPI and pBAD_McrB-N_C-HisTag (Sukackaite et al., 2012) plasmids, respectively. The *E. coli* strain ER2566 (LpnPI) and ER2267 (McrB-N) was used as a transformation host. Competent cells were prepared and transformed with heterologous DNA by the $CaCl_2$ method (Sambrook et al., 1989). The mutations were confirmed by DNA sequencing of the entire gene.

2.2.4 Protein purification

For protein purification the frozen cells were sonicated in the Sonication buffer (for LpnPI: pH 7.5, +10 % v/v glycerol; McrB-N and McrA: pH 8.0, see 2.1.4). The cells expressing YkrI and BmeDI were sonicated in Column buffer (see 2.1.4) supplemented with 2 mM phenylmethylsulfonyl fluoride. Cleared lysates were collected after centrifugation at 40000 g for 1 h. The hexahistidine tag containing proteins (LpnPI, McrB-N and McrA) were purified by chromatography through HisTrap HP chelating column (GE Healthcare) using a linear gradient in imidazole and HiTrap Heparin HP column (GE Healthcare) using linear NaCl gradient. YkrI and BmeDI were purified using a chitin column (New England Biolabs, #S6651). For CBD cleavage, the column was flushed with the Column buffer containing 50 mM DTT and incubated at 4 °C overnight. After incubation with DTT fractions containing YkrI and BmeDI were eluted from the column using the Column

buffer and loaded on a HiTrap Heparin HP column and eluted using a buffer containing Tris-acetate (pH 7.6) and 100–1000 mM K-acetate. The fractions containing the purest proteins were pooled and dialyzed against Storage buffer I (LpnPI, LpnPI-N and mutants: pH 7.5-8.5; McrB-N: pH 8.0, +0.1 mM EDTA; McrA: pH 8.0, see 2.1.4) or Storage buffer II (YkrI and BmeDI, see 2.1.4) and stored at −20 °C. Wt Ecl18kI was purified as described by Tamulaitis et al. (Tamulaitis et al., 2002).

The purity of all purified proteins was higher than 95 % as judged by densitometric analysis of SDS-PAGE (see 2.2.1.4). Concentrations of the proteins were determined by absorbance at 280 nm using the theoretical extinction coefficients calculated with the ProtParam tool available at http://web.expasy.org/protparam/. All protein concentrations are expressed in terms of monomer if not stated otherwise.

2.2.5 Analytical gel filtration

The gel filtration was carried out at room temperature on an AKTA FPLC system using a Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated with 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.5 M KCl and 5 mM 2-mercaptoethanol. The samples (2 μM wt McrA dimer or 4 μM McrA-N) were prepared in 0.1 ml of the same buffer. Elution from the column (flow rate 0.5 ml/min) was monitored by measuring A220. The calibration line was generated by measuring the elution volumes of a series of standards of known molecular mass (Gel Filtration Calibration Kit from GE Healthcare). The molecular weights of proteins were calculated by interpolating its elution volumes onto the calibration line.

2.2.6 DNA binding studies

2.2.6.1 Electrophoretic mobility shift assay (EMSA)

DNA binding was analyzed by the electrophoretic mobility shift assay (EMSA) using ^{33}P-labeled oligoduplexes (Table 2.1). Radiolabeled DNA was incubated with different amounts of enzymes for 15 min in 20 μl of the either Binding buffer I or II (see 2.1.4) supplemented with 5 mM Ca-acetate (for LpnPI, YkrI and BmeDI binding) or 1 mM EDTA (LpnPI, YkrI, BmeDI and McrA) or 1 mM DTT (McrB-N). Free DNA and protein–DNA complexes were separated by non-denaturing polyacrylamide gel electrophoresis (see 2.2.1.1).

2.2.6.2 EMSA competition experiments

EMSA samples contained McrB-N protein (final concentration 0.5 μM), radiolabeled methylated 13_5mC/13 DNA (0.5 μM), and variable amounts (final concentrations 0.5-5.0 μM) of unlabeled competitor DNA (carrying unmodified C or various cytosine modifications, Table 2.1). Electrophoresis was run as described above (see 2.2.1.1). The amounts of the radiolabeled specific protein-DNA complexes (i. e., the complex with the 13_5mC/13 DNA) at different competitor concentrations were determined by densitometric analysis of phosphorimager images of EMSA gels using OptiQuant software. Competition data was analyzed as described in (Sasnauskas et al., 2018), providing dissociation constant for the protein – radiolabeled 13_5mC/13 DNA complex ($K_{D(5mC)}$, in μM), and dissociation constant for protein – unlabeled competitor DNA interaction (K_D, in μM). All competition experiments were performed in triplicate, and the results are reported as the average binding constants ($1/K_{D(5mC)}$ and $1/K_D$, in μM^{-1}) ± 1 SD.

2.2.6.3 Fluorescence anisotropy measurements

McrB-N anisotropy measurements were performed using oligoduplex substrates carrying a HEX (hexachloro-fluorescein, Table 2.1) label at the 5′-terminus of the unmodified DNA strand. The measurements were performed on a Fluoromax-3 spectrofluorimeter equipped with polarizers. Excitation wavelength was 530 nm (2 nm slit), emission wavelength was 570 nm (16 nm slit), measurements were performed at 25 °C in the measurement buffer (100 mM KCl, 20 mM Mes-KOH pH 6.0). Protein was titrated into the 160 μL cuvette containing the DNA and buffer (final protein concentrations were from 20 to 1600 nM). The measured anisotropy values were corrected for the background anisotropy signal measured with single-stranded HEX-labeled DNA and plotted against protein concentration. The resultant dependencies were described by the equation (1)

$$A([p]) = A_0 + A_1 \times [p]/(K_D + [p]) \quad (1),$$

where $A([p])$ is the observed anisotropy value at protein concentration [p], A_0 is anisotropy observed with free DNA, A_1 is the amplitude of anisotropy increase, and K_D is the dissociation constant of protein-DNA complex. Non-linear fitting was performed using KyPlot software (Version 2.0 beta 14) (Yoshioka, 2002) with A_0, A_1 and K_D as fit parameters. In cases where parameters A_1 and K_D were poorly defined by experimental data (weak protein interactions with certain DNA variants), amplitude A_1 was set to a fixed value determined with methylated (13_5mC/13) DNA. The results

are reported as the average binding constants ($1/K_D$, in μM^{-1}) determined from 3 separate titrations ± 1 SD. Different batches of McrB-N mutant V66N in this assay produced erroneous results (no detectable increase in the anisotropy signal).

2.2.7 DNA cleavage experiments

2.2.7.1 Reactions with oligonucleotide substrates

DNA hydrolysis reactions were performed by manually mixing radiolabeled oligoduplexes (1 nM for YkrI and BmeDI, 100-400 nM for LpnPI, 200 nM for McrA) with the enzyme (100 nM YkrI and BmeDI, 500 nM LpnPI, 500 nM McrA) in either Reaction buffer I (for YkrI, BmeDI and LpnPI, see 2.1.4) or Reaction buffer II (for McrA) supplemented with appropriate concentrations (0.01–10 mM) of $MgCl_2$, $MnCl_2$, $CoCl_2$, $CaCl_2$, NiCl2, CuSO4 or ZnCl2 (McrA) at either 25 °C (LpnPI), 15 °C (YkrI and BmeDI) or 37 °C (McrA). All samples were subjected to denaturing PAGE (see 2.2.1.2).

A single exponential was fitted to the substrate depletion data yielding the observed rate constant k_{obs}. The k_{obs} values are reported as an average value from two to four experiments ± 1 SD.

2.2.7.2 Plasmid and phage DNA cleavage assays

The mixtures contained 0.05 or 0.5 μM wt McrA or its H229A variant and 0.2 or 0.5 μg DNA (reaction volume 25 μl). Reactions were carried out using supercoiled pACYC184 plasmid; supercoiled pACYC184_M.HpaII plasmid (M.HpaII methylated DNA); single-stranded M13mp18 DNA (all isolated from either a dcm+ or dcm– *E. coli* strain); supercoiled plasmid pBR322; or phage lambda DNA (isolated from either Dam+/Dcm+ or Dam-/Dcm- *E. coli* strains). Cleavage was stopped by phenol extraction. Samples were analyzed by agarose electrophoresis (see 2.2.1.3).

2.2.8 DNase I footprinting

50 nM of ^{33}P-labeled unmethylated and hemimethylated 50 bp DNA (Table 2.1) was preincubated at 22°C with 0.25–0.5 μM of wt McrA homodimer or 0.5–1 μM McrA-N monomer in 10 μl of the Binding buffer I (see 2.1.4). After 15 min, 2.5 μl of DNase I (ThermoFisher Scientific, 0.004 U/μl) in the Binding buffer I supplemented with 5 mM $CaCl_2$ and 12.5 mM $MgCl_2$ were subsequently added. After 2 min at 22 °C the DNase I reactions were quenched and separated on a high-resolution denaturing polyacrylamide gel (see 2.2.1.2).

2.2.9 Plasmid restriction assay

BL21(DE3) (McrA$^-$) *E. coli* cells with no intrinsic antibiotic resistance were made competent. The cells were then transformed with pLATE31 plasmids (Ap^R) containing wt McrA, its H228A or H229A variants, its N-terminal fragment (residues 1–174) or REM14 (unrelated protein, UniProt ID: Q9FX77). The cells were made competent again using Ap selection (at all growing steps the media contained 1 % glucose). The cells containing each of the pLATE31 plasmids were transformed with pACYC184 plasmids (Cm^R) containing either no methyltransferase ('empty') or M.HpaII. The pACYC184 plasmids were amplified in ER2267 (McrA$^-$) *E. coli* strain and verified by sequencing. 30 ng (3 μl) was used for transformation of 0.1 ml competent cells. At the recovery step after heat-shock the cells were supplemented with 0.9 ml of LB medium containing 1 % glucose. After 1 h of shaking cells were gently sedimented and resuspended in 0.6 ml of glucose-free medium. 0.1 ml of transformants were spread on 3 types of LB-agar plates with (i) Ap, Cm, 1 % glucose; (ii) Ap, Cm; (iii) Ap, Cm and 0.2 mM IPTG.

2.2.10 Base flipping experiments

2.2.10.1 Pyrrolocytosine fluorescence measurements

Steady-state fluorescence measurements were acquired on a Fluoromax-3 (JY-Horiba Inc., New Jersey) spectrofluorimeter equipped with a Xe lamp. Sample temperatures were maintained at 25 °C. Emission spectra (440–460 nm) were recorded at an excitation wavelength of 350 nm with excitation and emission bandwidths of 5 nm. The samples contained 1.0–2.0 μM of protein and 0.2-0.5 μM of pyrrolocytosine-labeled DNA oligoduplex (Table 2.1) in either buffer containing 30 mM Mes and 30 mM histidine, pH 6.3 (LpnPI, YkrI, BmeDI) or 20 mM Tris-HCl, pH 8.0, and 100 mM KCl (McrA). Control spectra for the background correction were collected under identical conditions except that DNA lacking pyrrolocytosine modifications was used instead of the fluorescent DNA. The fluorescence emission value of the corrected spectrum was determined at 450 nm. The fluorescence intensity of the pyrrolocytosine-modified oligoduplex in the presence of the protein was divided by the fluorescence intensity in the absence of the protein. The resultant values (the fold change of pyrrolocytosine fluorescence upon addition of the protein) were plotted as an average value from several independent experiments ±1 SD.

2.2.10.2 Reactions with CAA

DNA modification with chloroacetaldehyde (CAA) was performed as described by Daujotyte et al. (Daujotyte et al., 2008). Briefly, 100 nM radiolabeled DNA was mixed with 2 µM Ecl18kI, BmeDI or YkrI in 20 µl of the Binding buffer I (see 2.1.4). Reactions were initiated by adding CAA to a final concentration of 0.5 M and were incubated for 1 h at 37 °C. Modified strand cleavage was performed by adding 100 µl of freshly diluted 1 M piperidine and heating at 90 °C for 30 min. DNA was precipitated with ethanol and DNA fragments were separated on high resolution denaturing polyacrylamide gels (see 2.2.1.2). The markers were generated using the standard A+G Maxam-Gilbert sequencing reactions.

2.2.10.3 Chemical display of flipped out thymine and 5-methylcytosine

Experiments with thymine-substituted substrates (Table 2.1) were performed as described by Serva et al. (Serva et al., 1998). Briefly, radiolabeled DNA (10 nM) and protein (100 nM) were mixed in the Binding buffer II (total volume 20 µl, see 2.1.4). The reactions were initiated by adding $KMnO_4$ to a final concentration of 2 mM, incubated for 5 min at 25 °C and stopped by adding 20 µl of the solution containing 1.5 M Na-acetate (pH 7.0) and 1 M 2-mercaptoethanol. DNA was then precipitated with ethanol, redissolved in 1 M piperidine, heated at 90°C for 30 min, precipitated with ethanol, and analyzed on a high-resolution denaturing polyacrylamide gel (see 2.2.1.2). 5-methylcytosine oxidation assay followed the same procedures, except that a 20 mM Na-acetate reaction buffer (pH 4.3 at 25°C, ref. (Fritzsche et al., 1987)) was used.

2.2.11 McrB-N crystallization and structure determination

Protein crystallization was performed by sitting drop vapor diffusion method at 293 K. The McrB-N protein in the storage buffer was transferred into a glycerol-free buffer (200 mM KCl, 10 mM Tris-HCl pH 7.5, 0.1 mM EDTA and 0.02 % NaN_3) using NAP5 columns (GE Healthcare). The protein was concentrated to 9 mg/ml (~400 µM) and mixed with appropriate DNA (the protein : DNA ratio, as in the previous study (Sukackaite et al., 2012), was kept from 2 : 1 to 2 : 1.1). Crystals with 5hmC/5hmC DNA were obtained by mixing 0.5 µl of the complex (200 µM protein and 100 µM DNA) with an equal volume of the crystallization buffer (10 mM Bis-Tris, pH 5.5, 0.2 M ammonium acetate and 14% PEG4000); crystals with 5fC/C and 4mC/C DNAs were obtained by mixing 1 µl of the appropriate complex (400 µM protein, 220 µM DNA) with an equal volume of the crystallization

buffer (0.2 M LiCl and 22 % PEG4000 for the 5fC/C DNA, 17 % PEG4000 for the 4mC/C DNA). In all cases, crystals appeared within 2 weeks and reached the maximum size within 2 months.

Diffraction data for the McrB-N complex with 5hmC/5hmC DNA were collected using an in-house X-ray source (Rigaku-MM-007HF) and a Raxis-IV++ detector. For cryoprotection, the crystals were transferred into a buffer containing 7 mM BisTris (pH 5.5 at 25 °C), 0.15 M ammonium acetate, 12.5 % PEG4000, and 27 % PEG400.

Diffraction data for the complex with 5fC/C DNA were collected on the I911-3 beamline (MAX-Lab, Lund), and data for the complex with 4mC/C DNA were collected on the P13 beamline (PETRAIII, DESY, Hamburg) without any additional cryoprotection.

Data were processed with XDS (Kabsch, 2010), SCALA, and TRUNCATE (Collaborative Computational Project, Number 4, 1994). The structure was solved by molecular replacement with MOLREP (Vagin and Teplyakov, 2010). The structure of McrB-N bound to unmodified DNA (PDB ID: 3ssc) was used as the search model. The DNA in each case was built anew using COOT (Emsley and Cowtan, 2004). In all cases, the DNA in the structure did not have a unique orientation, therefore, two alternative DNA orientations were built, each with an occupancy of 0.5. Structure refinement was performed with PHENIX (Afonine et al., 2012). Data collection and refinement statistics are summarized in (Zagorskaitė et al., 2018). Coordinates and structure factors of McrB-N structures with 5hmC/5hmC, 5fC/C, and 4mC/C DNAs were deposited into the RCSB protein data bank under PDB IDs 6GCD, 6GCE, and 6GCF, respectively.

3 RESULTS AND DISCUSSION

3.1 Base flipping by modification-dependent restriction endonucleases of MspJI and PvuRts1I families

SRA domains of UHRF1, UHRF2, KRYPTONITE, and SUVH5 proteins flip out the modified base and place it in a protein pocket (Figure 3.1) (Arita et al., 2008; Avvakumov et al., 2008; Du et al., 2014; Hashimoto et al., 2008; Rajakumara et al., 2011a; Zhou et al., 2014).

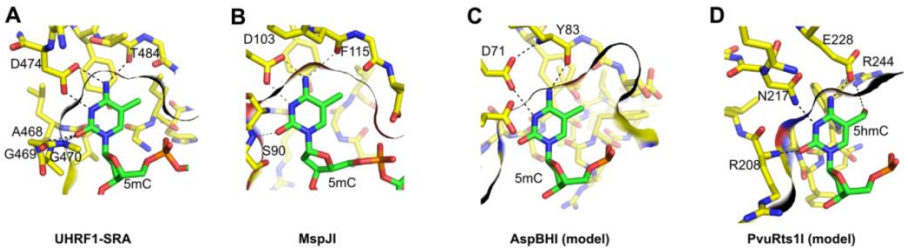


Figure 3.1. The modified cytosine binding pockets. (A-B) 5-methylcytosine recognition by the UHRF1 SRA domain (PDB ID 3fde) and the DNA recognition domain of MspJI endonuclease (PDB ID 4r28). The indicated protein pocket residues make base-specific contacts to the extrahelical base (Hashimoto et al., 2009; Horton et al., 2014b). (C-D) The models for the modified cytosine recognition by the DNA binding domains of AspBHI (PDB ID 4oc8) and PvuRts1I (PDB ID 4oq2). The indicated amino acid residues could form base-specific contacts to the extruded base. In each panel the dark line marks the boundaries of the protein pocket cut at the plane of the cytosine ring.

SRA-like domains were also identified in prokaryotes, where they serve as modules for the modified cytosine recognition by the modification-dependent restriction enzymes that protect host cells from infection by bacteriophages containing methylated, hydroxymethylated or glucosylated DNA. Despite the recently solved structures of MspJI, AspBHI (MspJI family), PvuRts1I and AbaSI (PvuRts1I family) enzymes (Horton et al., 2012, 2014b, 2014a, 2014c; Kazrani et al., 2014; Shao et al., 2014) only the co-crystal structure of the MspJI-DNA complex demonstrated that MspJI flips-out 5-methylcytosine into a protein pocket where a set of base-specific contacts are made (Figure 3.1 B) (Horton et al., 2014c). Mutational analysis and close structural resemblance of DNA binding domains of AspBHI, PvuRts1I and AbaSI to the SRA domains of MspJI and eukaryotic proteins (Figure 3.1 C-D) suggest that these enzymes also flip-out the modified cytosine (Horton et al., 2012, 2014a, 2014c; Kazrani et al., 2014; Shao et al.,

2014). However, for many eukaryotic and prokaryotic 5(h)mC-binding proteins the mechanism of modified cytosine recognition remains unknown. In the present study we tested several solution-based methods for fast and easy detection of extrahelical bases, including fluorescence measurements of the cytosine analog pyrrolocytosine and chemical modification of extrahelical pyrimidines with chloroacetaldehyde and $KMnO_4$.

3.1.1 DNA cleavage and binding by restriction enzymes of MspJI and PvuRts1I families

In this study we used three modification-dependent restriction endonucleases: LpnPI, YkrI and BmeDI. LpnPI belongs to the MspJI family and recognizes the DNA sequence 5′-CMDG-3′ (where M is 5mC or 5hmC, D – A, T or G) and cuts DNA 10/14 nt downstream from the recognition site (Zheng et al., 2010). It is closely related to the structurally characterized enzyme AspBHI (40 % identical and, 60 % similar amino acid residues (Zagorskaitė and Sasnauskas, 2014)), including nearly complete conservation of the presumed 5(h)mC binding pocket (Figure 3.1 C). LpnPI cleaved the cognate oligoduplex 30-M with a rate constant of 0.2 min^{-1}, but no cleavage was detected with an equivalent unmethylated oligoduplex 30-C (Figure 3.2 A). Discrimination between methylated and unmethylated DNA was also observed in EMSA: both LpnPI and the N-terminal LpnPI DNA binding domain (LpnPI-N) formed protein-DNA complexes with methylated DNA at much lower protein concentrations than with unmethylated DNA (Figure 3.2 B, C). Noteworthy, the discrimination of specific (methylated) vs non-specific DNA by LpnPI was stronger at pH 6.3 and less pronounced at pH 8.3 (Figure 3.2 D). This indirectly supports the model of extrahelical 5mC recognition in the binding pockets of MspJI-like enzymes, which requires protonation of the conserved aspartate (D103 in MspJI, D71 in both AspBHI and LpnPI, Figure 3.1 B-C) (Horton et al., 2014b, 2014a).

The YkrI and BmeDI display significant sequence similarities to the structurally characterized PvuRts1I-like family members PvuRts1I and AbaSI (Zagorskaitė and Sasnauskas, 2014). An optimal substrate for the PvuRts1I family enzymes consists of two 5hmC or 5ghmC nucleotides in the opposite DNA strands separated by a 20–22 bp DNA fragment (Szwagierczak et al., 2011; Wang et al., 2011).

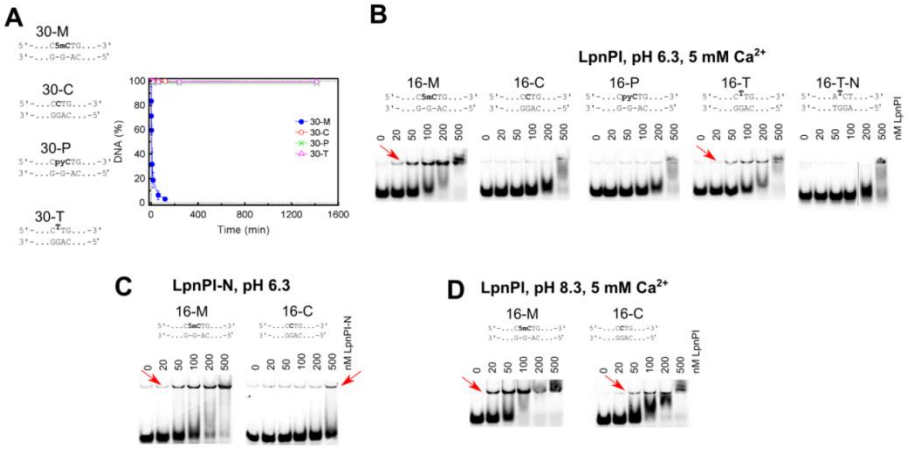


Figure 3.2 DNA cleavage and binding by LpnPI. The sequences in all panels depict recognition sites in the oligoduplex substrates. (A) DNA cleavage experiments. The reactions were performed with 500 nM enzyme (monomer) and 400 nM substrate at 25 °C. Time courses of the reactions are shown. The reaction rate constant for the 30-M substrate equals 0.20±0.05 min^{-1}. Reaction rate constants for other substrates were lower than 1×10^{-5} min^{-1}. (B) EMSA with LpnPI. DNA binding experiments were performed in a pH 6.3 buffer in the presence of 5 mM Ca^{2+} ions. The final substrate concentration was 10 nM, LpnPI concentrations (in terms of monomer) are indicated above the gel lanes. Red arrows mark the location of the protein-DNA complexes. (C) EMSA with LpnPI DNA binding domain (LpnPI-N). Experiments with the cognate (16-M) and noncognate (16-C) substrates were performed in a pH 6.3 buffer in the absence of Ca^{2+} ions. (D) EMSA with LpnPI in a pH 8.3 buffer in the presence of Ca^{2+} ions.

Current biochemical and structural data indicate that the 5(g)hmC sites are recognized by the DNA binding domains, while the two nuclease domains form a dimer and perform DNA cleavage at the center of the connecting DNA fragment, i. e. 11 nt from each modified base (Horton et al., 2014c; Shao et al., 2014). Replacement of one 5hmC with a 5-methylcytosine, cytosine and a non-cytosine bases on a series of 39 bp substrates (oligoduplexes 39-H/H, 39-M/H, 39-C/H and 39-H respectively, Table 2.1) did not abolish their cleavage by YkrI and BmeDI, but decreased the reaction rate (Figure 3.3 A), suggesting that even a single DNA binding domain is enough to anchor the enzyme dimer to DNA via a 5hmC base. In this case the second YkrI / BmeDI DNA binding domain presumably makes contacts to the base located 20 bp downstream of the 5hmC nucleotide, and contributes to the enzyme-DNA complex stability depending on the structural similarity of the contacted base to 5(g)hmC. The cleavage data for YkrI is also complemented by EMSA experiments that show a gradual reduction in the amount of the specific enzyme-DNA complex as the second 5hmC in the optimal substrate is replaced with a 5-methylcytosine, cytosine

and a non-cytosine base (Figure 3.3 B). However, we were unable to demonstrate such differences in binding for BmeDI (Figure 3.3 C).

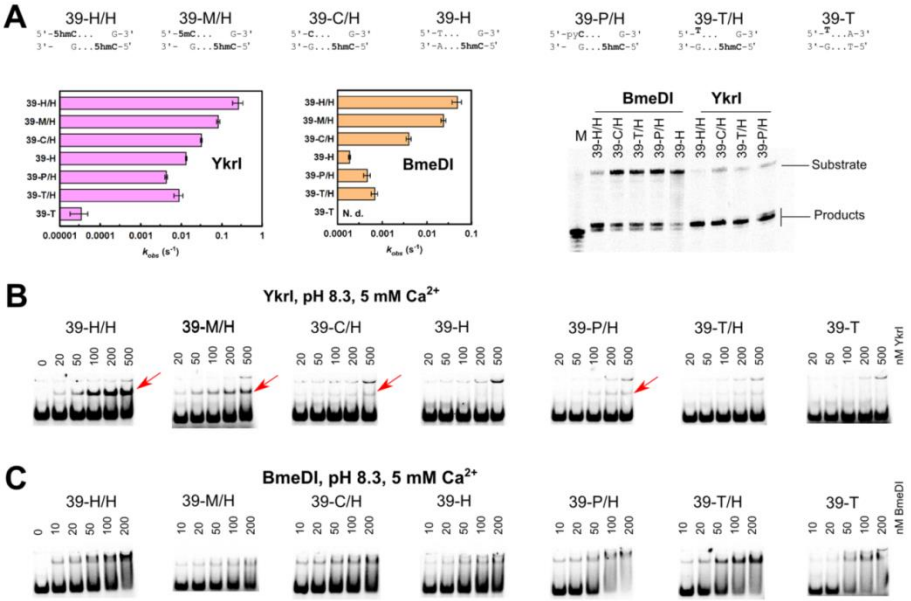


Figure 3.3. DNA cleavage and binding by YkrI and BmeDI. The sequences at the top of the figure schematically depict the 39-H/H (optimal substrate with two 5hmC bases), 39-M/H, 39-C/H, 39-H, 39-P/H, 39-T/H, and 39-T (one or both 5hmC-G base pairs replaced with a 5mC-G, C-G, T-A, pyrrolocytosine-G, and thymine-G base pairs, Table 2.1) oligoduplexes. (A) The observed first-order DNA cleavage rate constants. Cleavage reactions were performed with 1 nM substrate and 100 nM enzyme (monomer) at 15 °C. In our experimental setup, BmeDI cleavage of the 39-T oligoduplex was not detectable. Denaturing PAGE analysis of cleavage products formed with various DNA substrates is shown on the right-hand side. Gel lane 'M' contained a synthetic single-stranded oligonucleotide that corresponds to cleavage of the bottom strand 11 nt downstream of the 5hmC nucleotide. (B) EMSA with YkrI. DNA binding experiments were performed in a pH 8.3 buffer in the presence of 5 mM Ca^{2+} ions. The final DNA concentration was 1 nM, and YkrI concentrations are indicated above the gel lanes. Red arrows mark the location of the specific YkrI-DNA complexes. The upper band corresponds to the low mobility non-specific YkrI-DNA complex formed due to binding / aggregation of multiple protein molecules. (C) Electrophoretic mobility shift experiments with BmeDI. Reaction conditions were as in panel (B).

3.1.2 Pyrrolocytosine fluorescence measurements

Pyrrolocytosine (pyC) is a fluorescent cytosine analog that forms a stable base pair with guanine. The quantum yield of pyC fluorescence is sensitive to base unstacking (Berry et al., 2004), therefore pyC fluorescence measurements can be used to test the structural environment of a pyC base in

nucleic acids and protein-DNA complexes (Kuznetsov et al., 2012; Liu and Martin, 2001). Notably, pyC fluorescence was used to confirm base flipping in solution by the DNA binding domain of the McrBC enzyme (McrB-N) (Sukackaite et al., 2012). However, similar experiments with PvuRts1I were inconclusive, as no increase in pyC fluorescence was observed, despite the fact that PvuRts1I cleaved the pyC modified DNA (Kazrani et al., 2014).

An obvious prerequisite for binding and flipping of the pyC base is the ability of the enzyme to accommodate the flipped out base in the protein pocket. As this may be hindered due to the extra size of pyC in comparison to the unmodified cytosine or 5mC / 5hmC bases, we first tested the ability of LpnPI, YkrI and BmeDI to bind and cleave pyrrolocytosine-modified DNA. Upon replacement of 5mC with a pyC (substrate 16-P), the LpnPI and LpnPI-N binding to DNA became indistinguishable from the unmethylated DNA (Figure 3.2 B). The same was true for the pyC DNA cleavage (Figure 3.2 A). For YkrI and BmeDI we used a derivative of the optimal substrate with one 5hmC replaced with a pyC (substrate 39-P/H, Table 2.1). DNA binding and cleavage experiments performed with both enzymes indicated that the replacement of one 5hmC with a pyC compromises enzyme binding and activity to a similar extent as the replacement with an unmodified cytosine or a non-cytosine base (Figure 3.3). We therefore conclude that pyC is a poor 5mC / 5hmC substitute for LpnPI, YkrI and BmeDI. Indeed, modeling of the pyC base into the presumed binding pocket of AspBHI and PvuRts1I, preserving the H-bonding interactions with the conserved polar pocket residues, results in steric clashes (Figure 3.4 A). Not surprisingly, none of the proteins used in our study (LpnPI, YkrI and BmeDI) triggered fluorescence change of the pyC-containing DNAs (Figure 3.4 B). In contrast, almost no clashes are observed when pyC is modeled into the pocket of the DNA binding domain of McrBC (Figure 3.4 A), which readily binds pyC DNA and extrudes the modified base from the double helix (Sukackaite et al., 2012).

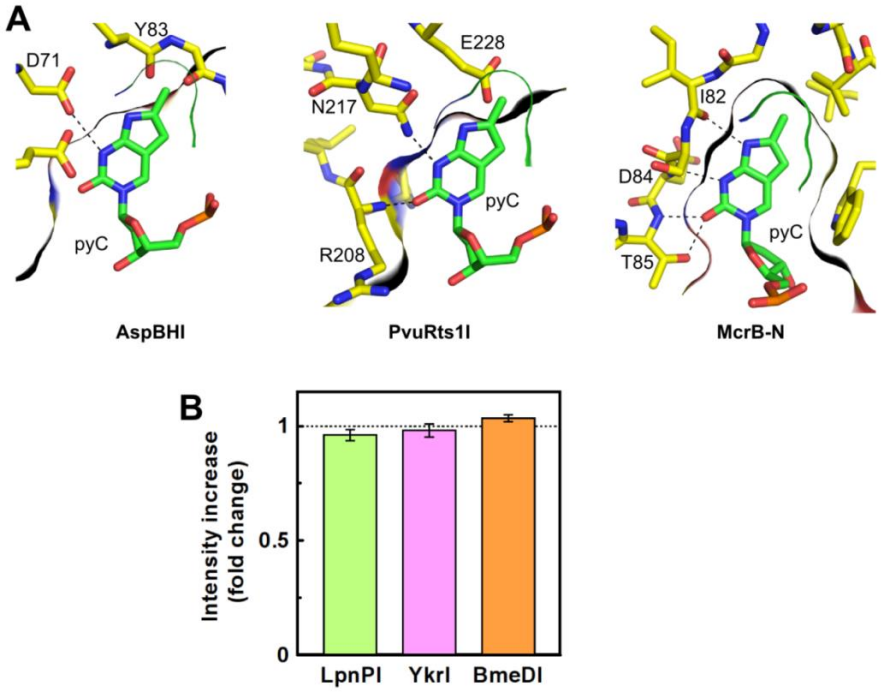


Figure 3.4 Experiments with pyrrolocytosine-substituted DNA. (A) The models of pyrrolocytosine base in the protein binding pockets of AspBHI, PvuRts1I, and the DNA binding domain of McrBC. The black and green lines mark the boundaries of the protein pockets (cut at the plane of the cytosine ring) and the pyrrolocytosine base, respectively. Only the McrBC domain accommodates the pyrrolocytosine base without steric clashes. (B) Pyrrolocytosine fluorescence measurements performed with LpnPI, YkrI and BmeDI. The graphs show the ratio of the cognate pyrrolocytosine-substituted oligoduplex (16-P for LpnPI and LpnPI-N, 39-P/H for YkrI and BmeDI) fluorescence intensity in the presence of the protein to the fluorescence intensity of the same oligoduplex in the absence of the protein.

3.1.3 Reactions with chloroacetaldehyde

Chloroacetaldehyde (CAA) is known to react with unpaired cytosine and adenine bases in DNA yielding 3, N4-ethenocytosine and 1, N6-ethenoadenine (Kuśmierek and Singer, 1982). Such modified residues can be detected by piperidine-induced strand cleavage. So far, the suitability of the CAA reaction was demonstrated for mapping of unmodified cytosine flipped out by several DNA cytosine-5 methyltransferases and restriction enzymes (Daujotyte et al., 2008). Since CAA also reacts with 5-methylcytosine (Oakeley et al., 1999), we asked if the same experimental setup could be used to detect extrahelical 5mC.

As a control we used endonuclease Ecl18kI. This base-flipping restriction enzyme recognizes the pseudosymmetric DNA site 5′-CCNGG-3′ and flips out the nucleotides of the central base pair (Bochtler et al., 2006) that become sensitive to CAA modification (Daujotyte et al., 2008). Ecl18kI binds DNA oligoduplexes with the central C-G and 5mC-G base pairs with comparable affinity both in the absence and in the presence of CAA (Figure 3.5 A). Nevertheless, enhanced DNA cleavage in the Ecl18kI complex after CAA / piperidine treatment was observed only for the unmodified central cytosine, but not for 5-methylcytosine (Figure 3.5 B). Thus, at least under standard reaction conditions used in our study, CAA cannot be used to detect extrahelical 5mC.

Therefore, we could only test if YkrI and BmeDI flip out the unmodified cytosine from the suboptimal substrate 39-C/H, which contains an unmodified cytosine base located ~20 bp away from the 5hmC base (Table 2.1). We rationalized that while one DNA binding domain of the dimeric enzyme is engaged in a high affinity interaction with the 5hmC site, the second DNA binding domain may interrogate the base ~20 bp downstream, in this case a cytosine, and this process may involve base flipping. This is supported by the observation that both YkrI and BmeDI cleave the 39-C/H substrate faster than the 39-H substrate, which lacks a cytosine base ~20 nt downstream of the 5hmC (Figure 3.3). However, neither YkrI nor BmeDI increased the reactivity of the target cytosine in the 39-C/H duplex with CAA (data not shown). Among other reasons for the lack of cytosine reactivity (no flipping, insufficient life-time of the flipped out base, and inactivation of the enzyme due to CAA treatment) is the mechanism for the extrahelical 5mC / 5hmC recognition by the SRA domains. In the solved co-crystal structures of the UHRF1 SRA domain and the MspJI REase, the Watson-Crick edge of the flipped out 5mC makes hydrogen bonds to the pocket residues (Figure 3.1 A-B) (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008; Horton et al., 2014b; Rajakumara et al., 2011a; Zhou et al., 2014). Conserved residues capable of hydrogen-bonding interactions with the Watson-Crick edge of cytosine derivatives are also present in both PvuRts1I (N217, E228, Figure 3.1 C) and AbaSI (N236, E247), and are conserved in YkrI / BmeDI (Zagorskaitė and Sasnauskas, 2014), suggesting that a cytosine base, had it been extruded from the DNA double helix, would be shielded from CAA due to the hydrogen-bonding interactions with the protein. In sharp contrast, Ecl18kI flips out both purine and pyrimidine bases, and binds them in a cavernous protein pocket without forming any base-specific contacts (Bochtler et al., 2006); this may explain

the efficiency of CAA modification of the extrahelical cytosine in the Ecl18kI-DNA complex (Daujotyte et al., 2008).

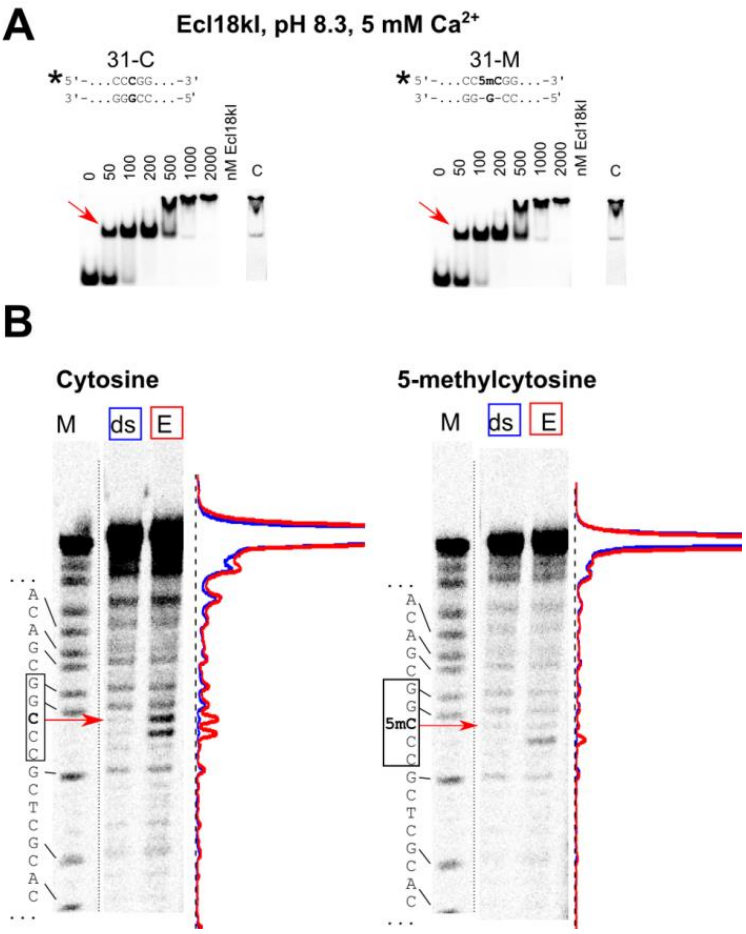


Figure 3.5 The chloroacetaldehyde modification assay with Ecl18kI restriction enzyme. Sequences at the top of the panel A schematically depict the 31-C (central base pair C-G) and 31-M (central base pair 5mC-G) oligoduplexes; the asterisk marks the radiolabel. (A) Ecl18kI binding to DNA oligoduplexes 31-C and 31-M in the pH 8.3 binding buffer in the presence of 5 mM Ca^{2+}. Final DNA concentration was 100 nM. Samples in gel lanes 'C' contained 1000 nM enzyme (dimer) and 500 mM CAA. Red arrows mark the position of the specific protein-DNA complexes. (B) DNA modification with CAA in the presence and in the absence of Ecl18kI. Red arrows mark the position of the central cytosine or 5-methylcytosine. Lanes 'M', the A+G markers of the radiolabeled strands; 'ds', 31-C (unmodified cytosine) and 31-M (5mC) oligoduplexes without the protein; 'E', 31-C and 31-M oligoduplexes + Ecl18kI. The normalized density profiles of individual lanes are shown at the bottom of the panel: 'ds' (blue), 'E' (red).

3.1.4 Permanganate oxidation of extrahelical pyrimidines

Under acidic conditions potassium permanganate oxidizes both thymine and 5-methylcytosine (Fritzsche et al., 1987). However, at physiologic pH this reaction is limited primarily to thymine. $KMnO_4$ treatment leads to conversion of pyrimidine bases to 5,6-dihydroxy-5,6-dihydropyrimidines (Hayatsu and Ukita, 1967); the oxidized bases undergo further degradation leading to cleavage of the phosphodiester backbone upon piperidine treatment. Since the oxidation reaction of the C5=C6 bond requires an access to the side of pyrimidine ring that is hidden in the double-stranded DNA, thymines and 5-methylcytosine in DNA helix are relatively resistant to permanganate oxidation compared to extrahelical pyrimidines. $KMnO_4$ was used to detect flipped-out thymines for cytosine and adenine DNA methyltransferases and a sequence-specific transposase (Bischerour and Chalmers, 2007; Reddy and Rao, 2000; Serva et al., 1998). We asked if the permanganate oxidation assay could help detect base-flipping by the modification-dependent restriction enzymes.

Since the $KMnO_4$ assay at near-neutral pH works only with the thymine bases, we made 5mC / 5hmC to thymine replacements in the standard LpnPI and PvuRts1I family substrates, thereby creating oligoduplexes with T-G mispairs (Table 2.1). EMSA experiments confirmed that LpnPI specifically binds the T-G mismatch substrate 16-T, albeit less tightly than the standard methylated duplex 16-M (Figure 3.2 B). However, we were unable to detect any T-substituted substrate cleavage by LpnPI, both in the standard reaction buffer (Figure 3.2 A) and under conditions mimicking the EMSA experiment (data not shown). On a control oligoduplex containing the T-G mismatch in a different sequence context (oligoduplex 16-T-N), we observed neither specific binding nor cleavage (Figure 3.2 B). Replacement of a single 5hmC base in the optimal YkrI / BmeDI oligoduplex 39-H/H with a thymine (substrate 39-T/H, Table 2.1) decreased the binding and cleavage of the substrate to a similar extent as the 5hmC-to-cytosine or the 5hmC-to-non-cytosine replacements (substrates 39-C/H and 39-H, Figure 3.3A-B), but did not change the cleavage position (BmeDI cleaves all substrates 11-12 nt, YkrI – 12 nt downstream from 5hmC, Figure 3.3 A); no YkrI binding was observed with the ‘non-cognate’ thymine-substituted oligoduplex 39-T (Figure 3.3 B). Taken together, the T-G mismatch is a poor substitute for a normal 5(h)mC-G base pair for all enzymes used in our study. A primary reason for this presumably is the direct read-out of the target base: the pockets for extrahelical base binding in the SRA domains, and their homologs in the MspJI / PvuRts1I-like enzymes, are optimized for the

specific hydrogen-bonding interactions with cytosine derivatives, but not thymine (Figure 3.1).

Surprisingly, incubation of the T-substituted substrates with LpnPI-N, LpnPI, YkrI, and BmeDI significantly increased the susceptibility of the mispaired thymine to $KMnO_4$ oxidation. A particularly strong enhancement in reactivity was observed with the 'cognate' mismatch substrate 16-T and LpnPI-N (Figure 3.6 A). Since the increase was not detectable with the 'non-cognate' T-G oligoduplex 16-T-N, we attribute the hyper-reactivity of the target thymine to the change in its environment induced by the specific binding of the modification-dependent enzyme. The increase in thymine reactivity of the 'cognate' T-G substrate 39-T/H upon its incubation with YkrI and BmeDI was less pronounced, but still clearly detectable (Figure 3.6 B). Almost no hyper-reactivity of the mispaired thymine was observed with the 'non-cognate' T-G oligoduplex 39-T (Figure 3.6 B), again implying that the change in the mispaired thymine environment observed with the 'cognate' thymine-substituted DNA was due to the specific enzyme interaction with the DNA.

The major drawback of the above assay is that it makes use of a mutated substrate. To strengthen the evidence for native modified cytosine flipping, we also performed $KMnO_4$ reactions on native 5mC-containing substrates. However, the $KMnO_4$ reacts with 5mC only at non-physiological pH (4.3), where the protonation state of both protein and DNA bases may interfere with protein function and the stability of the protein-DNA complex (nevertheless, examples of DNA-protein interaction studies performed at a very wide range of pH values are present in literature, e. g., ref. (Carey, 1988)). Though none of the enzymes used in our study showed any catalytic activity at pH 4.3, incubation of LpnPI-N with the cognate substrate 30-M, containing 5mC in the context of the LpnPI recognition site (Table 2.1), resulted in a significant enhancement of 5mC sensitivity to permanganate (Figure 3.6 C). This signal seems to be both LpnPI- and 5mC-specific, as control experiments performed on the 39-M/H oligoduplex containing 5mC in a different sequence context showed no enhancement in 5mC reactivity (though increase in reactivity was observed for some thymine residues, Figure 3.6 C).

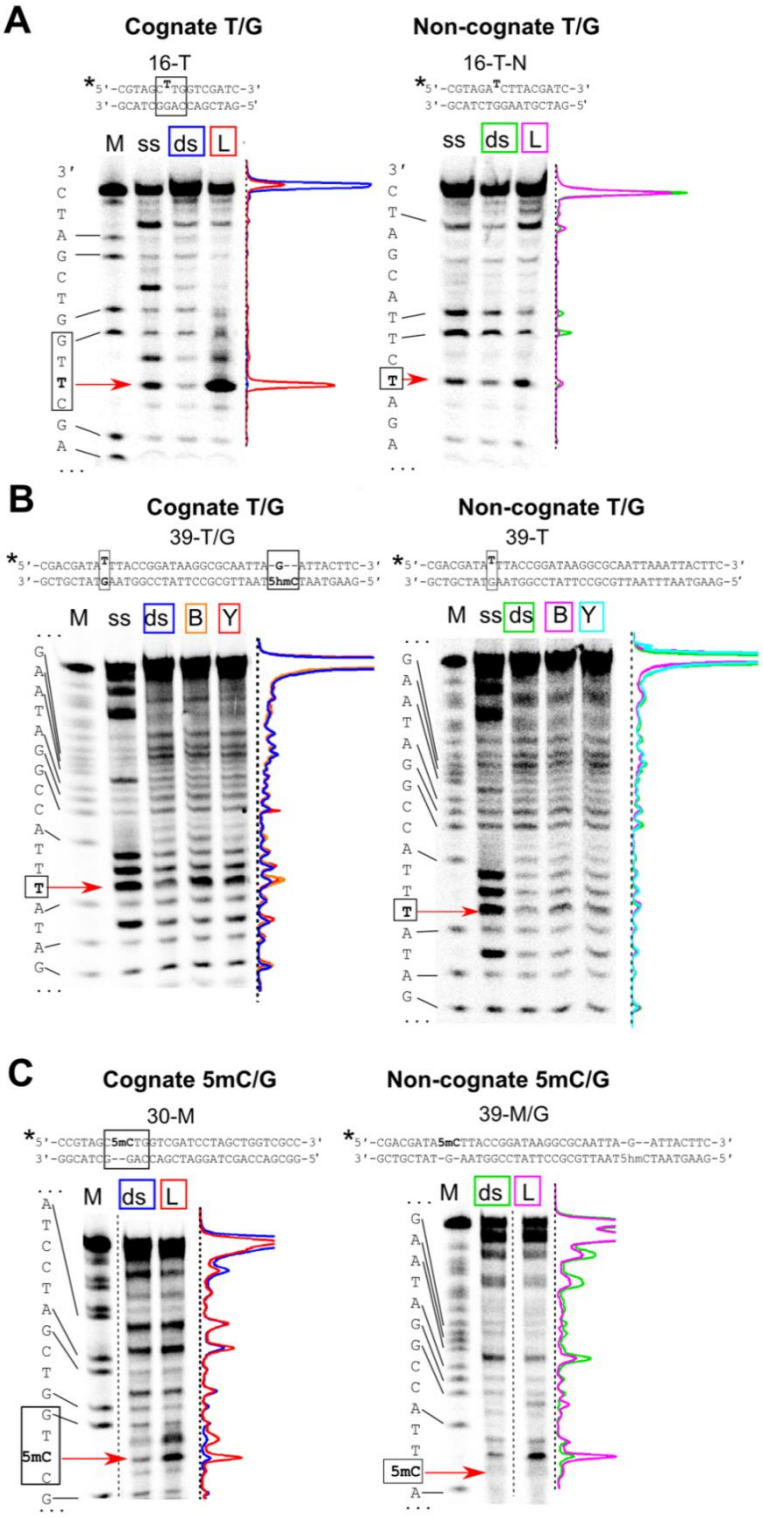

A
Cognate T/G
16-T
*5'-CGTAGCTGGTCGATC-3'
3'-GCATCGGACCAGCTAG-5'
M ss ds L
Non-cognate T/G
16-T-N
*5'-CGTAGATCTTACGATC-3'
3'-GCATCTGGAATGCTAG-5'
ss ds L
B
Cognate T/G
39-T/G
*5'-CGACGATATTACCGGATAAGGCGCAATTA-G--ATTACTTC-3'
3'-GCTGCTATGAATGGCCTATTCCGCGTTAAT5hmCTAATGAAG-5'
M ss ds B Y
Non-cognate T/G
39-T
*5'-CGACGATATTACCGGATAAGGCGCAATTAAATTACTTC-3'
3'-GCTGCTATGAATGGCCTATTCCGCGTTAATTTAATGAAG-5'
M ss ds B Y
C
Cognate 5mC/G
30-M
*5'-CCGTAGC5mCTGGTCGATCCTAGCTGGTCGCC-3'
3'-GGCATCG--GACCAGCTAGGATCGACCAGCGG-5'
M ds L
Non-cognate 5mC/G
39-M/G
*5'-CGACGATA5mCTTACCGGATAAGGCGCAATTA-G--ATTACTTC-3'
3'-GCTGCTAT-G-AATGGCCTATTCCGCGTTAAT5hmCTAATGAAG-5'
M ds L

Figure 3.6 Permanganate reactivity of pyrimidine bases in the protein-DNA complexes. Sequences at the top of the panels schematically depict the 'cognate' and the 'non-cognate' thymine-substituted substrates, the asterisk marks the ^{33}P radiolabel. Base pairs important for specific binding are in black boxes. Positions of the mispaired thymine and the 5-methylcytosine are marked with red arrows. (A) Thymine oxidation by $KMnO_4$ with or without LpnPI-N. 'M', the A+G markers of the radiolabeled 'cognate' and 'non-cognate' strands; 'ss', oxidation of the single-strand oligonucleotides; 'ds', double-stranded 16-T (cognate) and 16-T-N (noncognate) oligoduplexes without the protein; 'L', 16-T and 16-T-N oligoduplexes + LpnPI-N. Density profiles of individual lanes are shown: cognate DNA (blue), cognate DNA + LpnPI-N (red), non-cognate DNA (green), and non-cognate DNA + LpnPI-N (magenta). (B) Thymine oxidation by $KMnO_4$ in the presence of YkrI and BmeDI. 'M', the A+G markers of the radiolabeled 'cognate' and 'non-cognate' strands; 'ss', oxidation of the single-strand oligonucleotides; 'ds', 39-T/H (cognate) and 39-T (noncognate) oligoduplexes without the protein; 'Y', 39-T/H and 39-T oligoduplexes + YkrI; 'B', 39-T/H and 39-T oligoduplexes + BmeDI. Density profiles: cognate DNA (blue), cognate DNA + BmeDI and YkrI (orange and red, respectively), non-cognate DNA (green), non-cognate DNA + BmeDI and YkrI (magenta and cyan, respectively). (C) 5-methylcytosine oxidation by $KMnO_4$ at pH 4.3 with or without LpnPI-N. 'M', the A+G markers of the radiolabeled 'cognate' and 'non-cognate' strands; 'ds', double-stranded 30-M (cognate) and 39-M/H (noncognate) oligoduplexes without the protein; 'L', 30-M and 39-M/H oligoduplexes + LpnPI-N. Density profiles of individual lanes are colored as in panel (A).

When the 39-M/H oligoduplex (contains one 5mC and one 5hmC separated by 21 bp, Table 2.1), was incubated with YkrI and BmeDI, no changes in 5mC reactivity were observed (Zagorskaitė and Sasnauskas, 2014). Taken together, permanganate reactions are consistent with thymine (both LpnPI and YkrI / BmeDI) and 5-methylcytosine (LpnPI) flipping by the cytosine modification-dependent restriction enzymes, but the results are obtained either on a mutated substrate or at non-physiological pH.

3.1.5 Strengths and limitations of the solution-based methods for base flipping detection

In the present study we employed three established fluorescence and chemical-display based methods to test base flipping by the modification-dependent restriction endonucleases. Our results indicate that each assay has its strengths and limitations, and neither of them suits all 5mC / 5hmC-binding proteins. For example, the replacement of the target base with the fluorescent base pyrrolocytosine is a convenient method to detect base flipping by measuring the changes in pyrrolocytosine fluorescence intensity upon protein binding. In particular, it was successfully applied for the study of the modification-dependent enzyme McrBC (Sukackaite et al., 2012). However, for the method to work, the protein in question must have an

adequately sized protein pocket to accommodate the bulky pyrrolocytosine base. While this seems to be the case with the DNA binding domain of McrBC, modeling, DNA binding and cleavage studies indicate that the SRA-like domains of MspJI- and PvuRts1I-like enzymes used in our study do not tolerate such a replacement (Figures 3.2, 3.3 and 3.4). The second assay tested in our study, chloroacetaldehyde modification (Daujotyte et al., 2008), does not require non-natural base substitutions, but under conditions used in our study it worked only with unmodified cytosine (Figure 3.5).

The third base-flipping assay tested in our study makes use of the hyper-sensitivity of extrahelical pyrimidines to $KMnO_4$ oxidation. At near-neutral pH this method required replacement of the 5mC / 5hmC bases with thymines, thereby forming T-G mismatches. Due to the perturbed geometry of the mismatched base pairs, the unpaired thymines can themselves become hyper-reactive (Gogos et al., 1990). Fortunately, relatively low background signal observed in our experiments indicates that the accessibility of the mismatched thymines to the bulk solvent is limited, even though they are flanked from both sides with pyrimidine bases (oligoduplexes 16-T and 39-T/H, Table 2.1). Though similar in size to both 5mC and 5hmC, thymine was poorly tolerated by the modification-dependent enzymes used in our study, resulting in impaired binding and cleavage of the thymine-substituted substrates (Figures 3.2 and 3.3). Surprisingly, incubation of the thymine-substituted DNA with these proteins resulted in hyper-sensitivity of the mismatched thymine to $KMnO_4$ oxidation. A potential risk of using a mismatched oligoduplex is that the mismatch may induce additional conformational flexibility at or in the vicinity of the mispair that upon binding of a protein may result in stronger conformational changes as compared with those in a standard double-stranded DNA. However, the increase in reactivity in our experiments was localized to the target thymine, and was observed only with the ‘cognate’ thymine-substituted substrates (Figure 3.6), implying formation of native-like protein-DNA complexes with a flipped-out thymine.

Structural studies and modeling suggest that the flipped-out base in the protein pocket of the SRA-like domains is sandwiched between conserved polar, aromatic and hydrophobic residues (Figure 3.1). If the mispaired thymine occupies the same position as the 5mC / 5hmC residues, it should be shielded from $KMnO_4$ oxidation. Instead, the hyper-reactivity of the thymines suggests that the thymine base may be in a dynamic equilibrium between the intra-helical state and an ensemble of flipped-out states. Complete flipping and trapping of the thymine in the protein pocket, most

likely, is hindered by the failure of thymine to form hydrogen bonds with the pocket residues that are tailored for the direct read-out of the cytosine derivatives (Figure 3.1); on the other hand, opening of the T-G mismatch is much more easily achieved than the disruption of the native C-G base pair, thereby shifting the equilibrium towards the extrahelical states. Interestingly, the recently solved co-crystal structure of the MspJI-DNA complex (Horton et al., 2014b) revealed both specific base-flipping of a 5mC base by one DNA recognition domain, and non-specific flipping of a guanine residue by another DNA binding domain. In the latter case the extruded base occupies a slightly different position than 5mC, and the binding pocket remains in the more open conformation. Formation of such complex was interpreted as a possible target site search intermediate (Horton et al., 2014b). The extruded thymine detected in our study with LpnPI / YkrI / BmeDI could also occupy a similar non-native position that would permit the reaction with $KMnO_4$.

To strengthen the evidence for native modified cytosine flipping by the modification-dependent enzymes, we also probed the permanganate oxidation of 5mC bases at pH 4.3. Due to non-native conditions, this method was previously applied only for detection of 5mC in DNA (Fritzsche et al., 1987). Surprisingly, this assay revealed a significant sequence- and 5mC-specific 'positive' signal for LpnPI, which is consistent with 5mC flipping by this enzyme (whether the flipped-out base at such pH occupies the same position as in the native complex remains unknown). No signal was observed with YkrI and BmeDI, the most likely reason for the lack of the signal being the non-physiological pH, which may interfere with DNA binding by most proteins. Nevertheless, the $KMnO_4$ assay at low pH may work with some proteins, and therefore can prove useful in the studies of the 5mC recognition mechanism of other modified cytosine 'readers'.

Taken together, our study demonstrates the usefulness and limitations of several solution-based methods for the detection of flipped-out cytosine and its derivatives. Only one of the tested methods – permanganate oxidation of the extrahelical pyrimidines – provided evidence for base flipping by the modification-dependent restriction enzymes, implying that the lack of the 'positive signal' in one or even several assays does not exclude base flipping. Therefore, the DNA recognition mechanism of potential 5mC / 5hmC-binding proteins should be tested using a combination of all available methods. Nevertheless, the final proof or disproof for the base flipping mechanism would still require high resolution structures of protein-DNA complexes.

3.2 Structure-guided sequence specificity engineering of LpnPI

Available data show that the SRA-like fold is a versatile structural module that is used for the recognition of the modified cytosine in a different sequence context (Table 1.1). Subtle structural and size differences of the protein pocket may account for the discrimination of 5mC / 5hmC / 5ghmC bases by these domains (Horton et al., 2014c; Shao et al., 2014; Zhou et al., 2014). The recognition of the modified cytosine occurs only in a specific nucleotide context indicating tight coupling between the base flipping and recognition of the surrounding sequence. Structural and molecular mechanisms of sequence recognition by the SRA-like domains are still poorly understood. In the eukaryotic UHRF1-SRA domain specific for the 5mC residue in the CpG sequence context the 'base flip-promoting' ('thumb') and 'CpG recognition' ('NKR finger') loops penetrate into, respectively, major and minor DNA grooves, and make discriminating contacts to the CpG dinucleotide (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). In the target sites of the MspJI family enzymes the modified C is embedded into a variable, often degenerate nucleotide sequences that span up to 2 bp upstream and 3 bp downstream of the modified cytosine (Table 1.1). Readout of the 5mCNNR sequence by MspJI REase occurs through the minor groove contacts (Horton et al., 2014b), but the DNA recognition mechanism of other MspJI-like enzymes still has to be resolved. Therefore we concentrated on MspJI family enzyme LpnPI specificity studies trying understand and change structure determinants responsible for sequence recognition.

3.2.1 Crystal structure of LpnPI-N

LpnPI shares protein sequence similarities (Sasnauskas et al., 2015) with structurally characterized AspBHI (42 / 63 % identical / similar aa residues for the N-terminal domains) (Horton et al., 2014a) and MspJI (15 / 33 % identical / similar aa residues) enzymes of the MspJI family (Horton et al., 2012, 2014b).

The structure of LpnPI-N was solved at 2.1 Å resolution. The overall structure of LpnPI-N is very similar to that of the SRA-like DNA binding domain of AspBHI (Figure 3.7 A, B). The most interesting difference between AspBHI-N and LpnPI-N is the length and conformation of the Loop-2B (residues 21–31, correspond to AspBHI residues 22–33) involved in DNA binding. LpnPI-N is more compact than the corresponding domain of MspJI (224 versus 260 aa). The loops connecting β3–β4, β7–β8 strands and E–F helices are shorter in LpnPI-N by up to 5 aa; nevertheless LpnPI,

like AspBHI, contains an 8 aa insertion in the β8 strand that breaks it into two parts (Figure 3.7 A) (Horton et al., 2014a).

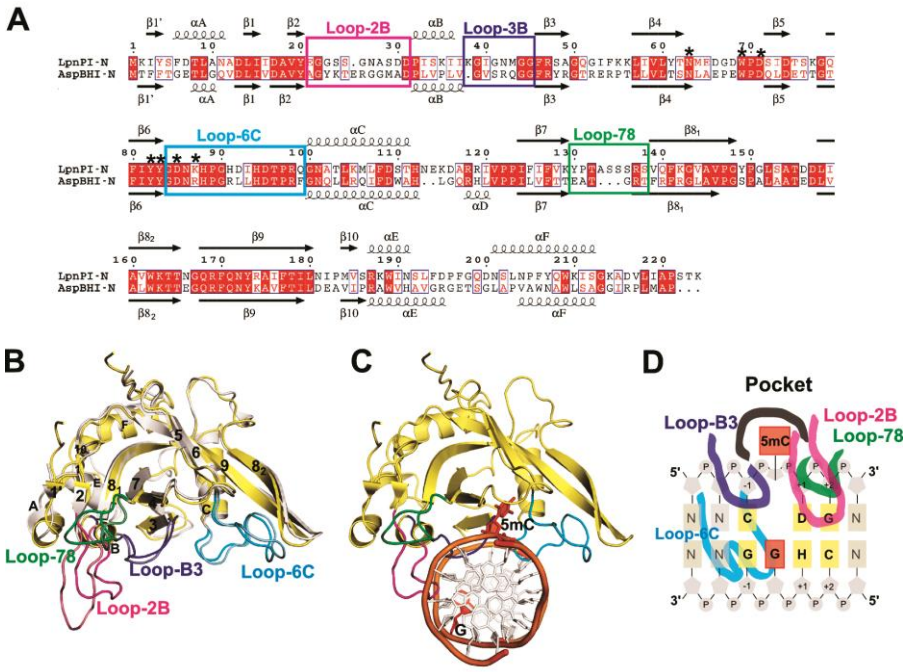


Figure 3.7 DNA recognition domain of restriction endonuclease LpnPI. (**A**) Sequence alignment of the N-terminal domains of LpnPI (LpnPI-N) and AspBHI (AspBHI-N). Secondary structure elements of LpnPI-N and AspBHI-N are numbered as in (Horton et al., 2014a). Residues that form the flipped-out base binding pocket are marked with asterisks. (**B**) Superimposition of LpnPI-N (in yellow) and AspBHI-N (in white; PDB 4OC8). The putative LpnPI / AspBHI DNA recognition loops are colored as follows: Loop-B3, blue / light blue; Loop-78, green / lime; Loop-2B, magenta / light magenta; Loop-6C, cyan / aquamarine. (**C**) The model of DNA-bound LpnPI-N, based on the crystal structure of DNA-bound UHRF1-SRA (PDB 3FDE). DNA recognition loops are colored as in (B), the flipped cytosine and the orphan intra-helical guanine are shown in red. (**D**) Schematic representation of LpnPI interactions with DNA. Protein loops and the 5mC:G base pair are colored as in panel (C); other bases comprising the LpnPI recognition site are shown in light orange and are numbered from '−1' (the bp upstream of 5mC) to '+2' (the 2nd bp downstream of 5mC).

3.2.2 DNA recognition determinants of LpnPI

The DNA-bound structures are available for several eukaryotic SRA domains (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008; Rajakumara et al., 2011a; Zhou et al., 2014) and the MspJI restriction endonuclease (Horton et al., 2014b). Since an overlay of LpnPI-N or

AspBHI-N with either the UHRF1-SRA or MspJI co-crystal structures places the DNA molecules and the flipped-out 5-methylcytosine bases in a similar position relative to LpnPI / AspBHI, we will further discuss the models of DNA-bound LpnPI-N and AspBHI-N based on the UHRF1-SRA-DNA structure (Figure 3.7 C, D). We will refer to the DNA base pairs 5′ (upstream) of the flipped cytosine as the '−1' and '−2' positions, and the base pairs 3′ (downstream) of the flipped base as the '+1', '+2' and '+3' positions.

The flipped-out cytosine binding pockets are similar in all SRA domains (Figure 3.8). The cytosine 5-methyl group in MspJI pocket is in van der Waals distance from D117, Y114 and W101 residues, and apparently makes a weak C–H…O hydrogen bond to the carbonyl oxygen of G116. These interactions may serve to distinguish modified cytosine from an unmodified base (Horton et al., 2014b). Equivalent positions in LpnPI and AspBHI are occupied by D85, Y82, W69, and G84 residues. The side walls of the MspJI pocket are formed by the residues W101, Y114 and D117, while D103, S90 and F115 make hydrogen bonds to the Watson-Crick edge of the modified cytosine. Equivalent residues in LpnPI are W69, Y82, D85, K87, D71, N63 and Y83 (W69, Y82, D85, R87, D71, N63, Y83 in AspBHI). Mutation of the AspBHI pocket residues D71, Y82 and D85 to alanine abolished the DNA cleavage activity (Horton et al., 2014a).

An alanine replacement of D71 in LpnPI had a similar effect: the reaction rate decreased more than 1000-fold (Table 3.1). A more conservative D71N replacement diminished the LpnPI DNA cleavage rate ~15-fold (Table 3.1). All DNA cleavage experiments were performed under the optimal reaction conditions (near equimolar enzyme and DNA concentrations; in agreement with the proposed mechanism for the MspJI reaction, which involves simultaneous interaction of the tetrameric enzyme with up to four cognate DNA molecules (Horton et al., 2012), the LpnPI reactions were much slower under enzyme excess conditions, (Sasnauskas et al., 2015)). We presume that the observed changes in the DNA cleavage rates under these reaction conditions are due to the altered DNA binding ability of LpnPI.

In the UHRF1–SRA–DNA complex structure, the vacant space left by the flipped-out base is filled in by the V451 residue from the 'base flipping-promotion' or 'thumb' loop (Figure 3.8) (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008); in MspJI–DNA complex, the E65 residue of the structurally equivalent Loop-B3 (loop between helix αB and strand β3) makes a hydrogen bond to the intra-helical orphaned guanine. The Loop-B3 in LpnPI contains residues N42 and M43 (Figure 3.8 B); M43 (Q43 in

AspBHI) is the likely candidate to fill the space left by the flipped-out cytosine, while the N42 (R42 in AspBHI) could make contacts to the orphan guanine or the −1 base pair from the minor groove side. In agreement with this model, the N42A mutation rendered the enzyme inactive, the M43Q mutation had little effect on the enzyme activity, and the M43A mutation decreased the DNA cleavage rate ~30-fold (Table 3.1). The glycine residue equivalent to the G41 in LpnPI is conserved in SgrTI and RlaI, but not in AspBHI, which has a serine at this position (Sasnauskas et al., 2015). Interestingly, the G41S replacement reduced LpnPI activity ~300-fold (Table 3.1). Presumably, the glycine residue contributes to Loop-B3 flexibility that is important for the LpnPI function.

In the LpnPI–DNA model, the DNA backbone on the 3′ side of the flipped cytosine is contacted by the Loop-78 (residues 130–137 between β7 and β8 strands, Figure 3.8 B). With a little change in a loop conformation, the side chains of S136 and R137 could make base-specific contacts in the major groove 3′ of the extrahelical cytosine; moreover, the K173 residue from an equivalent MspJI loop is the prime candidate for the purine base recognition at the +3 position in the MspJI target sequence (Horton et al., 2014b). DNA cleavage analysis of the alanine replacement mutants S136A and R137A showed that only R137 is critical for LpnPI function (Table 3.1).

The LpnPI Loop-2B (LpnPI residues 23–31 between β2 strand and αB helix) is positioned in the minor groove 3′ of the flipped cytosine. Structurally equivalent loops are present in both AspBHI and MspJI, but are much shorter in the eukaryotic SRA domains (Figure 3.8). To probe the role of Loop-2B residues in LpnPI function we made alanine replacements of the polar residues S25, N27 and D30. Only the N27A mutation abolished LpnPI activity, while the other two mutants displayed wt-like activity (Table 3.1). This is consistent with N27 playing a role in DNA binding / recognition.

The DNA on the 5′ side of the flipped cytosine is approached by the LpnPI Loop-6C (residues 84–99 in LpnPI / AspBHI, and 116–129 in MspJI). An equivalent ‘CpG recognition’ or ‘NKR finger’ loop in the eukaryotic SRA domains is significantly longer (e.g. 484–508 in UHRF1–SRA), adopts a different conformation, and makes base-specific contacts in the DNA major groove (Figure 3.8 D) (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). In our LpnPI–DNA model, only the conserved R98 residue of the Loop-6C is within an H-bonding distance to the −1 base. It may also contribute to proper positioning of the adjacent Loop-B3 residues. In agreement with this model, LpnPI mutant R98A was nearly inactive (Table 3.1).

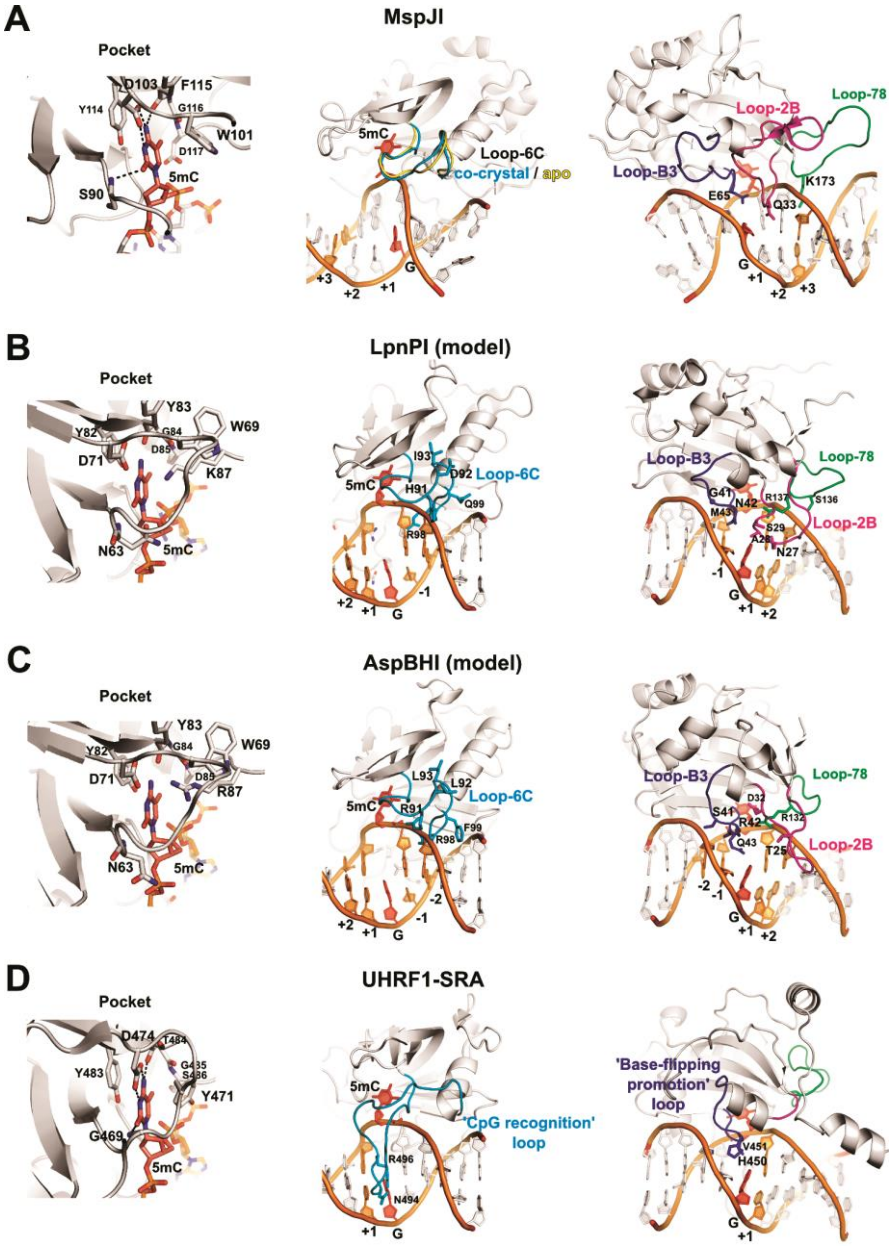


Figure 3.8 DNA recognition by SRA domains. The structures of the DNA-bound UHRF1-SRA and MspJI, and the apo-structures of MspJI, LpnPI-N and AspBHI-N (PDB ID: 3FDE, 4R28, 4F0Q, 4RZL, 4OC8) were superimposed with MultiProt (Shatsky et al., 2004). Equivalent DNA recognition elements in all panels are shown in identical orientation. Left: recognition of the flipped-out base in the protein pocket; center: Loop-6C or 'CpG recognition' / 'NKR finger' loop (cyan); right: Loop-B3 or 'base-flipping-promotion' loop (blue), Loop-B2 (magenta), and Loop-78 (green). In all panels the flipped-out base and the orphan intra-helical guanine are

colored red; other nucleotides comprising the specific recognition sequence of the corresponding protein are colored orange and are numbered from '−2' (the second bp upstream of 5mC) to '+3' (the third bp downstream of 5mC). (**A**) DNA recognition by MspJI. Loop-6C occupies a similar position both in the apo- and the DNA-bound structures and does not make base-specific contacts. Residues Q33, E65, and K173 from the '2B', 'B3', and '78' loops, respectively, are close to the DNA bases. (**B** and **C**) The models of DNA-bound LpnPI and AspBHI based on the co-crystal structure of UHRF1-SRA. Loop-C6 and Loop-B3 residues 41–43, 91–93, and 99 are different in LpnPI and AspBHI. LpnPI Loop-2B and Loop-78 residues 27–29 and 136–137 were mutated in the present study; AspBHI Loop-2B residues T25 and D32 are critical for the enzyme function (Horton et al., 2014a); AspBHI Loop-78 residue R132 overlaps with the critical LpnPI residue R137. (**D**) DNA recognition by the SRA domain of the eukaryotic UHRF1 protein. The loops equivalent to Loop-78 and Loop-2B in MspJI-like restriction endonucleases are colored green and magenta, respectively.

Table 3.1 Catalytic activity of LpnPI mutants

Mutation	k_{obs} (s^{-1})[a]	Activity (%)[b]
wt LpnPI	$(3.3 \pm 0.8) \times 10^{-3}$	100
5(h)mC binding pocket		
D71A	$(1.0 \pm 0.3) \times 10^{-6}$	0.03
D71N	$(2.0 \pm 0.4) \times 10^{-4}$	6
Loop-2B (contacts downstream of 5(h)mC)		
S25A	$(7.0 \pm 1.5) \times 10^{-3}$	200
N27A	No cleavage	<0.01
D30A	$(1.0 \pm 0.3) \times 10^{-3}$	30
Loop-B3 (adjacent to orphan guanine)		
G41S	$(1.0 \pm 0.1) \times 10^{-5}$	0.3
N42A	No cleavage	<0.01
M43A	$(1.0 \pm 0.2) \times 10^{-4}$	3
M43Q	$(1.6 \pm 0.1) \times 10^{-3}$	50
Loop-6C (contacts upstream of 5(h)mC)		
R98A	$(1.0 \pm 0.6) \times 10^{-6}$	0.03
Loop-78 (contacts downstream of 5(h)mC)		
R137A	$(0.7 \pm 0.3) \times 10^{-5}$	0.2
S136A	$(2.1 \pm 0.1) \times 10^{-3}$	60

[a]Oligoduplex DNA cleavage reactions were performed on the 'gC(mC)TG' substrate (Table 2.1) and the observed rate constants k_{obs} were determined by single-exponential fits. The lowest DNA cleavage rate measured in our assay is 3×10^{-7} s^{-1}. [b]The activity is expressed as the ratio k_{obs}(mutant)/k_{obs}(wt) × 100%.

3.2.3 The sequence specificity of LpnPI

The LpnPI recognition sequence provided in REBASE (Roberts et al., 2015) is 5′-C(mC)DG-3′ (where mC – modified C and D = A, G, or T), though 5′-S(mC)DS-3′ or 5′-(mC)DS-3′ (G>>C) sequence specificities have also been reported (Cohen-Karni et al., 2011). To analyze the sequence preference of LpnPI, we measured the LpnPI cleavage rates on a set of oligoduplex substrates (Table 2.1) that differ from the reference 'gC(mC)TG' substrate by 1 or 2 bp. Our results are consistent with LpnPI having a strong preference for the 5′-C(mC)DG-3′ recognition site (cleavage rate constant ~3×10^{-3} s^{-1}), albeit ~90-, ~500- and 1500-fold slower DNA cleavage was also observed with DNA sites 5′-G(mC)TG-3′, 5′-C(mC)TC-3′, and 5′-G(mC)TC-3′ (rate constants ~3×10^{-5}, ~6×10^{-6} and 2×10^{-6} s^{-1}, respectively, Figure 3.9 A, B). Therefore, the target site for the wt LpnPI can be defined as 5′-(C>>G)(mC)D(G>>C)-3′.

3.2.4 Sequence specificity engineering of LpnPI

SRA-like DNA binding domains of several MspJI family enzymes have closely related protein sequences (Sasnauskas et al., 2015) and recognize target sites partially overlapping with the 5′-C(mC)DG-3′ target of LpnPI, e.g. AspBHI (5′-YS(mC)NS-3′) and SgrTI (5′-C(mC)DS-3′) (Cohen-Karni et al., 2011). Since Loop-6C is the likely candidate for the recognition of the −1 or −1 / −2 nucleotides ('C' – LpnPI and SgrTI, 'YS' – AspBHI) (Figures 3.7 D and 3.8 B, C), we attempted to alter LpnPI sequence specificity by swapping the LpnPI Loop-6C (residues 91–99) with the equivalent loop of AspBHI (Figure 3.9 A). Our expectation was that the resultant LpnPI variant 'LpnPI-91RLL' would preferentially cleave DNA substrates with a 5′-YS-3′ dinucleotide in the −1 / −2 positions. Indeed, LpnPI-91RLL is more tolerant for a G in the −1 position than the wt enzyme (the ratio of cleavage rates between the 5′-C(mC)DG-3′ and 5′-G(mC)DG-3′ substrates dropped from 90- to 1.5-fold). However, LpnPI-91RLL, like the wt enzyme, has no preference for the −2 base pair (Figure 3.9 A). Loop-6C replacement also increased the tolerance for A and T substitutions in the −1 position: both the 5′-A(mC)DG-3′ and 5′-T(mC)DG-3′ substrates are refractory to the wt enzyme, but are slowly cleaved by LpnPI-91RLL (Figure 3.9 A). Surprisingly, the loop replacement also increased tolerance for substitutions in the downstream part of the recognition site, but to a lesser extent: the rate difference for the 5′-C(mC)DG-3′ and 5′-C(mC)DC-3′ substrates decreased from ~500- to ~20-fold (Figure 3.9 A). Thus, the relaxed specificity of LpnPI-91RLL for the −1 position at least in part stems from an overall

improvement of enzyme binding to methylated DNA. This may also account for a significant acceleration of the doubly substituted 5′-G(mC)DC-3′ substrate cleavage. Taken these data together, we define the preferred target site for LpnPI-91RLL cleavage as 5′-(S>>W)(mC)D(G>C)-3′.

The sequences of Loop-6C in LpnPI and AspBHI differ by four residues (positions 91–93 and 99, Figure 3.7 A). To identify the residue responsible for the altered LpnPI site preference, we have made LpnPI mutants H91R, D92L, I93L and Q99F, and tested their cleavage activity on the 5′-C(mC)DG-3′ and 5′-G(mC)DG-3′ substrates. We found that the site preference of single point mutants did not change (the rate difference was ~100-fold for H91R and Q99F mutants, and dropped only to ~50-fold for the D92L and I93L variants), suggesting that the change in LpnPI-91RLL specificity is due to simultaneous replacement of several residues. In the next step, we also replaced the LpnPI Loop-6C with a corresponding loop from MspJI (residues 123–130), which shows no specificity for the DNA sequence upstream of the modified base. However, the resultant LpnPI variant 'LpnPI-91VGL', despite the unperturbed secondary structure, was inactive (Sasnauskas et al., 2015).

Our LpnPI-DNA model suggests that Loop-2B contacts DNA to the 3′ side from the modified base (Figures 3.7 D and 3.8 B). To probe the role of Loop-2B residues in DNA recognition, we have made the following LpnPI variants:

- 'LpnPI-27HTG', in which three consecutive LpnPI residues (27-29) were replaced with equivalent SgrTI residues (34–36); our expectation was that this would relax the selectivity of LpnPI for the +2 base pair in the 5′-C(mC)DG-3′ sequence and enhance cleavage of the SgrTI recognition sequence 5′-C(mC)DS-3′;
- 'LpnPI-21AGY' – the 21–30 LpnPI residues were replaced with 21–31 AspBHI residues (Figure 3.7 A); our expectation was that this would enable cleavage at the AspBHI-like recognition sites with a 5′-NS-3′ dinucleotide at the +1 / +2 positions instead of the 5′-C(mC)DG-3′ sequence.

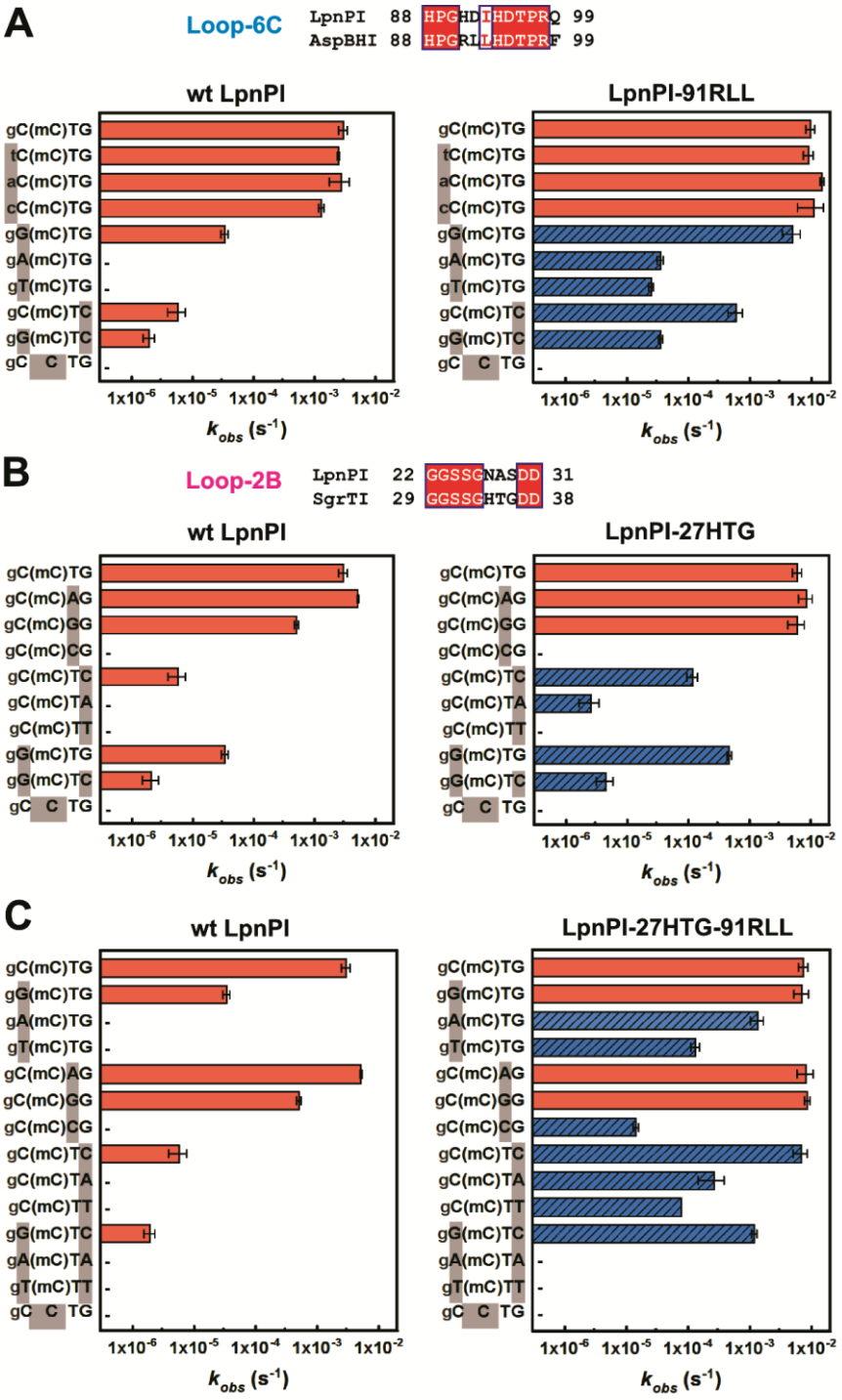


Figure 3.9 Recognition site preference of LpnPI. Oligoduplex DNA cleavage reactions were performed under standard reaction conditions and the observed rate constants k_{obs} were determined by single-exponential fits. The recognition sequences

of the DNA substrates are shown on the left-hand side of the graphs. '(mC)' stands for 5mC (the last substrate in each graph is the unmethylated control); sequence positions that differ from the reference 'gC(mC)TG' substrate are marked with grey boxes; full oligoduplex sequences are listed in Table 2.1. The reaction rates of LpnPI mutants that show increased cleavage due to loop replacement are marked by blue streaked bars; '−' marks undetectable cleavage (rate lower than 3×10^{-7} s^{-1}, the starting position of the x-axis). Alignments of the LpnP / AspBHI Loop-6C and the LpnPI / SgrTI Loop-2B that were replaced in the LpnPI-91RLL and LpnPI-27HTG are shown above panels A and B, respectively. (**A**) Wt enzyme and the LpnPI variant LpnPI-91RLL (Loop-6C replacement) on DNA substrates with variable sequence upstream and downstream of 5mC. (**B**) Wt enzyme and the LpnPI variant LpnPI-27HTG (Loop-2B replacement) on DNA substrates with variable sequence upstream and downstream of 5mC. (**C**) Wt enzyme and the 'double-swap' LpnPI variant LpnPI-27HTG-91RLL on DNA substrates with variable sequences upstream and downstream of 5mC.

As expected, the LpnPI-27HTG variant had an increased tolerance for a C in the +2 position (the rate difference for the 5′-C(mC)DG-3′ and 5′-C(mC)DC-3′ substrates dropped from ~500- to ~50-fold, Figure 3.9 B). The loop replacement also increased the cleavage rate of the 5′-C(mC)TA-3′ substrate, which was refractory to wt LpnPI. Unexpectedly, we have also observed improved cleavage of the 5′-G(mC)DC-3′ substrate with a substitution in the −1 position, as the rate difference for the 5′-C(mC)DG-3′ and 5′-G(mC)DC-3′ substrates decreased from ~90 to ~15-fold (Figure 3.9 B). Thus, relaxed recognition of the +2 base pair at least partially may be due to improved non-specific binding to methylated DNA. Taken these data together, we define the preferred target site for LpnPI-27HTG cleavage as 5′-(C>G)(mC)D(G>C>>A)-3′.

We also made point mutations at all three Loop-2B positions that differ between LpnPI and SgrTI (N27H, A28T and S29G). The ratio for the 5′-C(mC)DG-3′ and 5′-C(mC)DC-3′ cleavage rates for the S29G and A28T mutants was ~200 fold, while the N27H mutant was inactive. Intriguingly, the double mutant N27H+S29G regained full activity, suggesting that the N27H mutation needs extra space or flexibility provided by the S29G mutation. Moreover, the double N27H+S29G mutant displayed DNA cleavage properties akin to LpnPI-27HTG (the ratio for the 5′-C(mC)DG-3′ and 5′-C(mC)DC-3′ cleavage rates dropped to ~30-fold, (Sasnauskas et al., 2015)), indicating that N27 and S29 are the key Loop-2B residues involved in DNA binding.

Contrary to LpnPI-27HTG, the LpnPI-21AGY variant preserved wt-like specificity. Unlike AspBHI, it did not tolerate a cytosine at the +1 position, and had a strong preference for a G nucleotide at the +2 position (Sasnauskas et al., 2015). Involvement of Loop-2B in the recognition of the +1 base

therefore seems unlikely. Another plausible candidate for the +1 base pair recognition is Loop-78, which is 3 aa longer in LpnPI than in AspBHI (Figure 3.7 A). To test this hypothesis we have also constructed LpnPI variant 'LpnPI-133G', containing a shorter, AspBHI-like Loop-78 version (133–134 LpnPI residues replaced with a glycine, which is equivalent to AspBHI residue G131, Figure 3.7 A). Unfortunately, the resultant LpnPI variant LpnPI-133G was inactive on all substrates tested (data not shown).

Since Loop-6C and Loop-2B act as separate LpnPI DNA binding 'modules', we also produced a double-swap LpnPI variant 'LpnPI-27HTG-91RLL' containing the AspBHI Loop-6C (relaxes recognition of the -1, and to a lesser extent +2 positions) and the SgrTI Loop-2B (relaxes recognition of the +2, and to a lesser extent −1 positions). Our expectation was that the 'double-swap' LpnPI would readily cleave the 5′-SMDS-3′ site. DNA cleavage analysis confirmed this prediction: LpnPI-27HTG-91RLL cleaved 5′-G(mC)DG-3′ and 5′-C(mC)DC-3′ that differ by a single bp from the standard LpnPI recognition site, and the doubly-substituted 5′-G(mC)DC-3′ site, which is poorly tolerated by the wt LpnPI and the 'single-swap' variants (Figure 3.9 A-C). Simultaneous substitution of two loops apparently further relaxed LpnPI specificity for the −1 and +2 positions: the 'double-swap' LpnPI variant cleaved the 5′-A(mC)DG-3′ and 5′-T(mC)DG-3′ sites only ∼5- and ∼50-fold slower than the standard substrate (the rate differences are >10000- and ∼300-fold for the wt LpnPI and the '91RLL' variant, respectively, Figure 3.9 A); detectable cleavage (∼30–100-fold slower in comparison to the standard 5′-C(mC)DG-3′ substrate) was also observed for the 5′-C(mC)DA-3′ and 5′-C(mC)DT-3′ DNAs, which are both refractory or almost refractory to wt LpnPI and LpnPI-27HTG cleavage (Figure 3.9 B-C). However, no cleavage was detected for DNA substrates 5′-A(mC)DA-3′ and 5′-T(mC)DT-3′, indicating that the presence of two 'unfavorable' A:T base pairs in both the −1 and +2 positions is not tolerated (Figure 3.9 C). The recognition sequence of LpnPI-27HTG-91RLL can thus be defined as the combination of 5′-S(mC)(D>>C)-3′ and 5′-(mC)(D>>C)S-3′ recognition sites.

3.2.5 Overview of modified DNA sequence recognition by LpnPI

3.2.5.1 5(h)mC binding pocket

The modified cytosine binding pocket is conserved in SRA and SRA-like domains. Typically, the side walls of the pocket are built of aromatic side chains, which make stacking interactions with the extrahelical base, and polar residues, which make cytosine-specific H-bonds to the Watson-Crick

edge of the base (Figure 3.8). Pocket mutations of AspBHI, PvuRts1I and AbaSI proteins severely impaired DNA cleavage activity (Horton et al., 2014a, 2014c; Kazrani et al., 2014; Shao et al., 2014). LpnPI is no exception: the D71A mutation reduced the DNA cleavage rate ~1000-fold (Table 3.1). The co-crystal structure of MspJI and the models of DNA-bound AspBHI / LpnPI (Horton et al., 2014b, 2014a) predict that the pocket aspartate (D71 in LpnPI) must be protonated to form a H-bond with the N4 cytosine atom of the flipped-out base (Figure 3.8 B). Since a similar position in another cytosine modification-dependent enzyme PvuRts1I is occupied by an asparagine (N217), we also made the LpnPI mutant D71N. Surprisingly, even this conservative mutation reduced the LpnPI cleavage rate 10-fold (Table 3.1). Presumably, the structure of the pocket is highly optimized, therefore even slight perturbation of its geometry / H-bonding network has a detrimental effect on enzyme function.

3.2.5.2 DNA sequence recognition by Loop-B3

LpnPI Loop-B3 approaches DNA from the minor groove side. Structurally equivalent loops in SRA domains provide residues (e.g. V451 in UHRF1-SRA) that fill in the vacant space left by the flipped-out cytosine, and contribute to the recognition of the adjacent base pair (Figure 3.8 D) (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). In the MspJI–DNA co-crystal structure a similar position is occupied by E65, which contacts the orphan intra-helical guanine (Figure 3.8 A). Loop-B3 in AspBHI contains residues S41, R42 (both unique to AspBHI, (Sasnauskas et al., 2015)) and Q43; the same positions in LpnPI are occupied by G41, N42 and M43 (unique to LpnPI). Alanine replacement of the 42th residue in the 'B3' loops of both enzymes abolished their activity ((Horton et al., 2014a) and Table 3.1), suggesting direct involvement of R42 / N42 residues in orphan guanine or adjacent base pair recognition. Alanine mutations of residues Q43 / M43, which overlap with the UHRF1-SRA V451 residue, were less deleterious, while the LpnPI mutant M43Q displayed wt-like DNA cleavage activity (Table 3.1). Presumably, the main purpose of the bulky M43 / Q43 residues is to fill the space left by the flipped out 5mC rather than make base-specific contacts (Figure 3.8 B).

AspBHI has a preference for the −2 nucleotide to be a pyrimidine (C or T), while LpnPI and other related enzymes accept any nucleotide at this position (Cohen-Karni et al., 2011). In our current models of DNA-bound LpnPI and AspBHI (Figure 3.8 B, C) the closest residue to the −2 bp is G41 / S41. We speculate that lacking a side chain at the 41th position, LpnPI

accepts any nucleotide at the −2 position. In contrast, the same position in AspBHI is occupied by a serine, which could perform pyrimidine / purine discrimination, e.g. *via* a minor groove hydrogen bond to the N3 purine atom in the complementary strand (Figure 3.8 C). In agreement with this model, some AspBHI S41 mutants displayed altered site preference (Horton et al., 2014a). The G41S replacement in LpnPI decreased the cleavage activity ~300-fold, but did not change the base preference for the −2 position (data not shown). It cannot be excluded that other factors, including a subtle difference in the Loop-B3 conformation may contribute to the −2 base pair discrimination by AspBHI.

3.2.5.3 DNA sequence recognition by Loop-78

Loop-78 approaches DNA downstream of the modified base (Figure 3.8 C). In MspJI, the loop residue K173 is the primary candidate for the recognition of the +3 base pair, where MspJI has a strong preference for a purine (Horton et al., 2014b). Equivalent loops in LpnPI and AspBHI are shorter by five and eight residues, respectively (Sasnauskas et al., 2015). Nevertheless, our current model of DNA-bound LpnPI and mutational data (alanine replacement of Loop-78 residue R137 inactivates LpnPI, Table 4) both suggest that Loop-78 residues contact DNA. Whether these contacts are limited to the DNA backbone, or contribute to the specific recognition of DNA bases (e.g. discrimination of the +1 bp) currently remains unknown, as the replacement of LpnPI Loop-78 with an AspBHI-like shorter loop (LpnPI variant '133G'), despite the proper folding of the protein (Sasnauskas et al., 2015), rendered LpnPI inactive.

3.2.5.4 DNA sequence recognition by Loop-6C

In the models of DNA-bound LpnPI and AspBHI, Loop-6C approaches the 5′-part of the target sequence from the minor groove side (Figure 3.8 B, C). An equivalent 'CpG recognition' or 'NKR finger' loop in eukaryotic SRA domains is longer, and makes base-specific contacts in the major groove (Figure 3.8 D) (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). Here, we show that replacement of the LpnPI Loop-6C with an equivalent AspBHI loop (four amino acid mutations at positions 91–93 and 99) enables cleavage of the 5′-G(mC)DG-3′ site with a G base in the −1 position, accelerates cleavage of sites with A and T bases in the −1 position, and to a lesser extent improves cleavage of DNA with a C in the +2 position (Figure 3.9 A). This change in site preference could not be emulated by single Loop-6C mutants, indicating that several loop residues contribute to DNA recognition. Intriguingly, three out of four residues replaced in the

‘LpnPI-91RLL’ variant (positions 91–93) in our current model of DNA-bound LpnPI / AspBHI point away from the DNA (Figure 3.8 B, C). Direct contacts to DNA bases by these residues would require a change in Loop-6C conformation similar to that observed in UHRF1-SRA (Figure 3.8 D). However, in MspJI enzyme the Loop-6C occupies the same position both in the apo- and in the DNA-bound structures (Figure 3.8 A) (Horton et al., 2012, 2014b). The role of the 91–93 and 99 Loop-6C residues in the −1 bp recognition therefore remains undefined: some loop residues may contact DNA bases directly, but this would require an LpnPI / AspBHI-specific change in Loop-6C conformation upon DNA binding; alternatively, Loop-6C residues could contribute to the sequence recognition indirectly through interactions with other protein residues that make direct DNA contacts. An indirect role of Loop-6C residues in DNA recognition may also explain simultaneous relaxation of LpnPI interaction with substrates carrying substitutions both upstream (position −1) and downstream (position +2) of the methylated base. Another important fact is that LpnPI-91RLL, contrary to the donor enzyme AspBHI, had no preference for the −2 bp (Figure 3.9 A). Presumably, recognition of this base pair is performed by another AspBHI structural element, most likely the Loop-B3 (see 3.2.5.2).

3.2.5.5 DNA sequence recognition by Loop-2B

Loop-2B occupies the minor groove on the 3′ side of the modified base (Figure 3.8 A, C). Replacement of the LpnPI Loop-2B with an equivalent loop from SgrTI relaxed the specificity of the LpnPI-27HTG variant for the +2 position, thereby accelerating cleavage of 5′-C(mC)DC-3′ and 5′-C(mC)DA-3′ sites (Figure 3.9 B). Interestingly, cleavage of the 5′-G(mC)DG-3′ DNA, which carries a substitution in the −1 position, was also increased. This suggests that Loop-2B replacement may improve the overall affinity of the enzyme for the methylated DNA. Two out of three LpnPI residues replaced in the ‘27HTG’ variant, namely, N27 and S29, in the current apo-LpnPI / DNA model point away from the DNA and are located closer to the +1 rather to the +2 base pair (Figure 3.8 B). Presumably, upon DNA binding Loop-2B undergoes a conformational change that brings these residues closer to the +2 base pair. The AspBHI Loop-2B is longer by 1 aa and adopts a different conformation (Figures 3.7 B and 3.8 C); its importance for the enzyme function was also confirmed by mutagenesis (T25A and D32A mutations abolished AspBHI activity (Horton et al., 2014a)). Conversely, the glutamine Q33 from the MspJI Loop-2B that contacts the DNA bases 3′ to the flipped cytosine (Figure 3.8 D) is

dispensable for MspJI activity (Horton et al., 2014b). This is consistent with MspJI lacking any sequence preference for the +1 and +2 base pairs. An equivalent loop in eukaryotic SRA domains is much shorter and is not involved in base-specific DNA interactions (Figure 3.8 D).

In summary, we show here that LpnPI recognizes the context of the flipped cytosine via several surface loops that act as separate DNA binding / recognition modules. LpnPI is a promising model system for specificity engineering of modification-dependent restriction endonucleases, since it displays a significant plasticity of target site recognition, somewhat reminiscent of homing endonucleases (Chevalier and Stoddard, 2001). Indeed, though wt LpnPI is most active on the canonical site 5′-C(mC)DG-3′, it also cleaves at alternative 5′-G(mC)DG-3′ and 5′-C(mC)DC-3′, sites albeit at a reduced rate (Figure 3.9 A, B). The LpnPI loop engineering further shifted enzyme preference for alternative recognition sites. Most notably, the 'double-swap' LpnPI variant, which carries Loop-2B from AspBHI and Loop-6C from SgrTI, recognizes a shorter target sequence, which can be defined as either 5′-S(mC)(D>>C)-3′ or 5′-(mC)(D>>C)S-3′, and readily cuts the 5′-G(mC)DC-3′ site, which differs from the canonical recognition site 5′-C(mC)DG-3′ by two base pairs (Figure 3.9 C). The relaxed sequence specificity seems to be an intrinsic feature of MspJI family enzymes. From the practical point of view this means that results of a real-life DNA cleavage experiment (% DNA cleaved at a particular site) greatly depend on the enzyme / DNA concentrations and the reaction duration. For example, MspJI, AspBHI and LpnPI cleavage sites established under more favorable reaction conditions (with activator oligoduplex) were more 'relaxed' than recognition sequences determined under less favorable conditions (no activator duplex) (Cohen-Karni et al., 2011). Nevertheless, despite of promiscuous specificity for the target site surrounding the modified cytosine, the SRA-like domain has proved a surprisingly robust module for the modified cytosine DNA recognition: neither the wt LpnPI nor any 'swap' or mutant variants showed any activity on unmethylated DNA. The plasticity of the target site recognition intrinsic to the MspJI family enzymes and the stringent discrimination against unmethylated DNA provided by the SRA domain pave the way for engineering of an enzyme specific for the 5mC embedded in any sequence context. Such 5mC-specific enzyme would be a useful tool in genome methylation studies. Significant relaxation of LpnPI specificity presented here is a step towards this goal.

3.3 Recognition of modified cytosine variants by the DNA binding domain of McrBC

The McrBC complex recognizes 5mC, 5hmC or 4mC containing sequences, and cleaves DNA at a variable position from the recognition site. The crystal structure of the McrBC DNA binding domain (McrB-N) with methylated DNA revealed that despite an unrelated tertiary structure, it follows the same mechanism for modified base recognition as the SRA domains: the base is flipped out into a protein pocket. The base flipping also was confirmed by solution based method using fluorescent cytosine analog pyrrolocytosine (Sukackaite et al., 2012). Restriction enzyme McrBC is used as a sensitive molecular tool to detect methylated cytosines in DNA (Lippman et al., 2005), although its relative ability to recognize different cytosine modifications, including 5mC, 5hmC, 5fC, 5caC, and 4mC, has not been reported yet. We provided structural and biochemical study on the McrB-N preference for different cytosine modifications.

3.3.1 Novel structures of McrB-N

Crystal structures of McrB-N bound to DNA oligoduplexes containing 5hmC, 5fC and 4mC DNA modifications were solved at 1.55-1.8 Å resolution. It was shown that McrB-N flips-out 5hmC, 5fC and 4mC bases in a similar fashion to 5mC (Sukackaite et al., 2012) (Figure 3.10). In all cases the flipped out base occupies identical positions in the protein pocket, with the 02, N3 and N4 atoms forming H-bonds to backbone and side chain atoms of McrB-N residues T85, D84 and I82 (Figure 3.10).

The 5-methyl groups of 5hmC and 5fC make vdW contacts to the surrounding W49, Y64, L68 and Y117 residues. Due to a shifted location, the N4-methyl group of 4mC also makes vdW interactions to V66 and S120, but not to W49 (Figure 3.10). The hydroxyl group oxygen of 5hmC is located outside the plane of the aromatic ring, and makes vdW contacts to V66 and the hydroxyl oxygen of Y64. Formation of H-bond between 5hmC and Y64 hydroxyl groups is unlikely due to unfavorable geometry. In contrast, the formyl oxygen of 5fC is located in the plane of the aromatic ring, and in addition to the intramolecular H-bond to 5fC 4-aminogroup, forms vdW interactions with Y117, L68, Y64 and V66 residues (Figure 3.10).

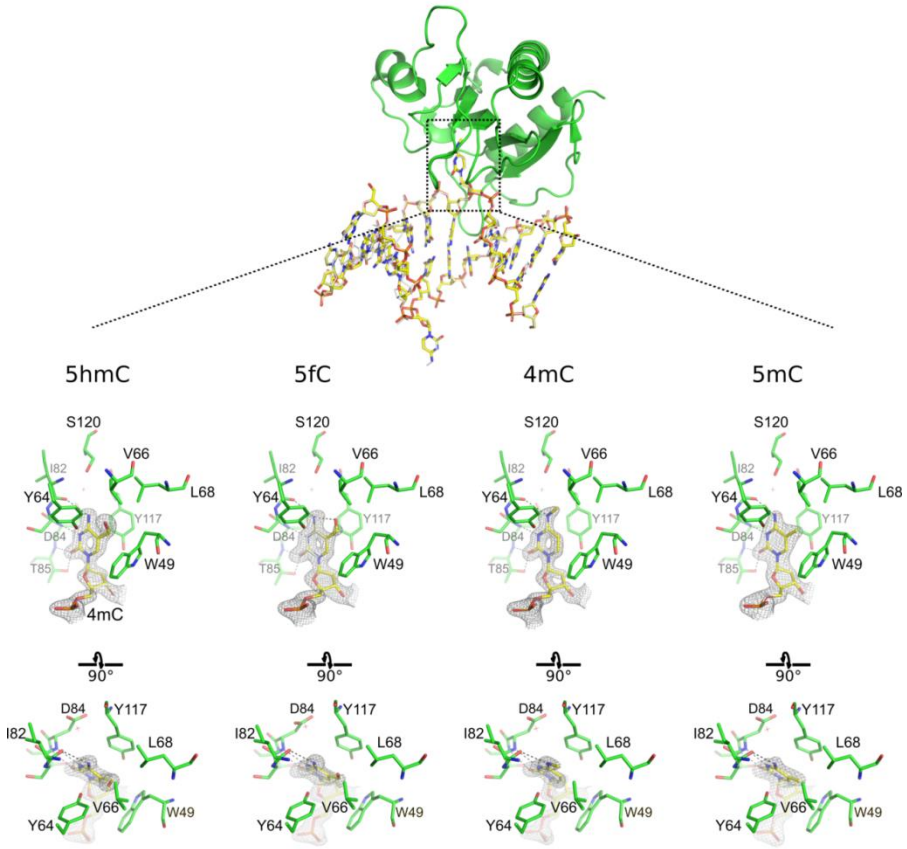


Figure 3.10 Structures of McrB-N with 5hmC, 5fC, 4mC and 5mC DNA. Top: the overall structure of McrB-N bound to 4mC DNA (PDB ID 6gcf). Only one protein subunit (chain A) is depicted; the DNA is shown in stick representation (yellow; the model of an alternative DNA orientation is shown as white sticks). Bottom: the flipped-out base binding pocket in the structures with 5hmC, 5fC and 4mC DNAs (PDB IDs 6gcd, 6gce and 6gcf, respectively); the structure with 5mC DNA (PDB ID 3ssc, ref. (Sukackaite et al., 2012)) is also shown. In all panels, the 2mFo-DFc electron density map contoured at a 2.0 σ (5hmC, 4mC) or 1.3 σ (5fC, 5mC) level is shown for the flipped-out nucleotides.

3.3.2 Pocket mutants

Distinct types of interactions made by alternative cytosine modifications prompted us to engineer pocket mutations that would alter the relative preference of McrB-N to different cytosine variants. We have made 3 groups of point mutations at conserved positions of McrB-N pocket:

#1 Y117L, Y64F – mutations that increase the size of the flipped base binding pocket;

#2 V66L, V66I – mutations that reduce the pocket size;

#3 V66D, V66N, V66Q, L68N, L68Q, S120K, S120E, S120Q – mutations that introduce additional H-bond donors and / or acceptors on various faces of the pocket surface.

Some of the #3 group mutants (V66D, L68Q, S120E,S120Q, S120K) could not be purified due to negligible yields; the yields of the #1 group mutant Y117L and #3 group mutants V66N, V66Q and L68N were also considerably reduced in comparison to WT McrB-N, but we were able to purify sufficient amounts of proteins for biochemical experiments.

The initial characterization of the mutant domains was performed using EMSA assay on DNA oligoduplexes containing 5mC, 5hmC, 5fC and 5caC modifications, or an unmodified C (Figure 3.11). We find that WT McrB-N readily interacts with 5mC, 5hmC and 4mC DNAs, but not with C, 5fC or 5caC DNA (Figure 3.11)). The group #1 mutant Y117L failed to form complexes with any of the tested DNAs, implying that large aromatic pocket residues that make stacking interactions with the flipped base are an absolute requirement for DNA binding by McrB-N. No binding was also observed with the V66Q mutant, suggesting that extra pocket space occupied by the long glutamine side chain may prevent base flipping with all cytosine variants. On the other hand, we find that #1 group mutant Y64F, #2 group mutants V66L and V66I, and #3 mutants V66N and L68N formed variable amounts of complexes with 5mC, 5hmC and 4mC DNAs, but none interacted with C, 5fC or 5caC DNAs (Figure 3.11).

3.3.3 Selectivity of WT McrB-N and pocket mutants to modified cytosine variants

More thorough examination of DNA binding affinities and relative preferences was performed with the WT domain and McrB-N mutants Y64F, V66I, V66L and V66N, employing either an EMSA-based competition assay (Figure 3.12 B), or fluorescence anisotropy measurements (Figure 3.12 C). Since the V66N mutant produced erroneous results in the fluorescence anisotropy assay (no increase in the anisotropy signal upon titration), we have also characterized the double mutant Y64F_V66N.

Competition experiments indicate that the preferred modification for WT McrB-N is N4-methylcytosine (dissociation constant $K_D \approx 0.1$ μM, $1/K_D \approx 10$ μM^{-1} Figure 3.12 B). Binding is 2.5-5-fold weaker to 5mC and 5hmC DNAs (K_Ds approx. 0.25 and 0.5 μM, respectively), and negligible to 5fC and 5caC ($K_D > 4$ μM, or $1/K_D < 0.25$ μM^{-1}). Qualitatively and quantitatively similar results, despite different buffer conditions, were also obtained using the anisotropy measurements (Figure 3.12 C).

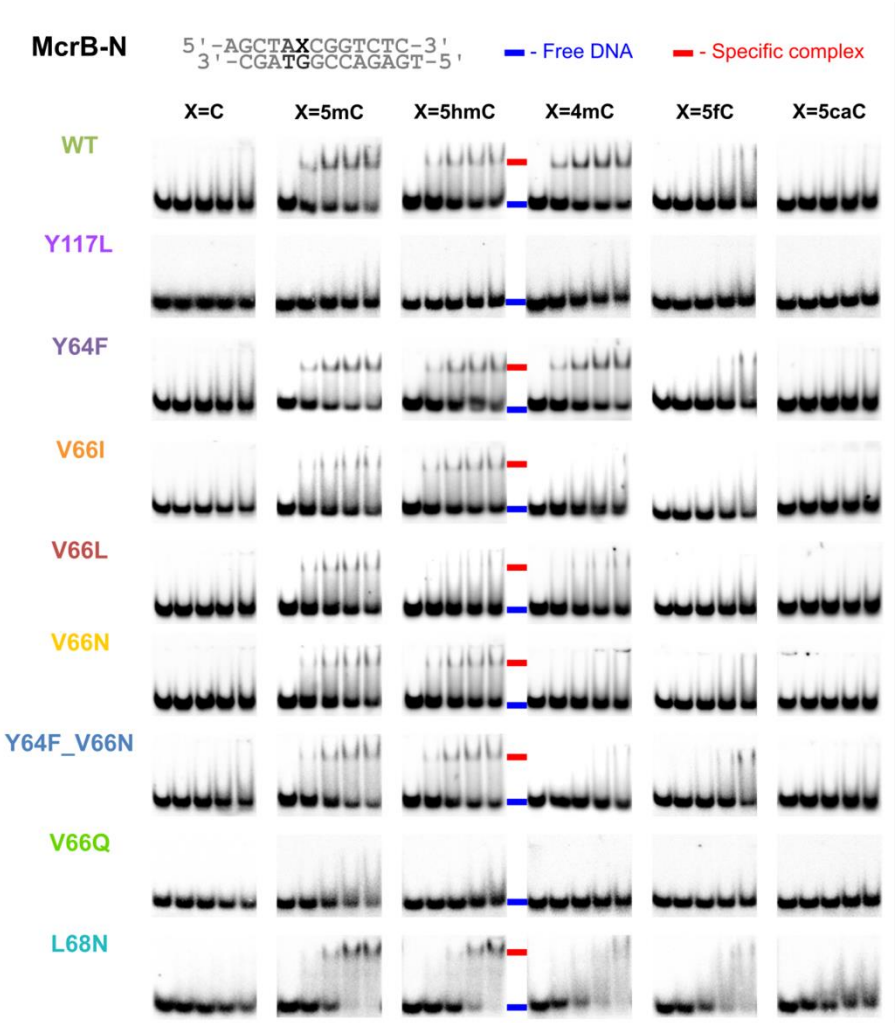


Figure 3.11 Characterization of pocket mutants by EMSA. Reactions contained 0, 0.5, 1, 2, 5 μM protein and 0.5 μM DNA carrying various cytosine variants. Experiments were performed as described in 2.2.6.1.

Mutation V66I, which reduces the flipped base binding pocket, decreased the affinity of McrB-N to 5mC, 5hmC and 4mC modifications to a similar extent (approx. 4-fold decrease in $1/K_D$ values, Figure 3.12 B, C), leaving the relative preferences for distinct cytosine variants unchanged. A similar decrease in affinity without well-defined changes in the relative preferences was also observed for the V66L mutant.

Mutation Y64F, which removes a single hydroxyl group from the pocket, had virtually no effect on the McrB-N interaction with 5mC and 5hmC

DNA, but 3-fold reduced the affinity to 4mC DNA (Figure 3.12 B, C). Even more pronounced discrimination against 4mC was observed for the V66N and Y64F_V66N mutants (binding to 4mC, $K_D \approx$ 2-4 μM, is comparable to 5fC and 5caC, Figure 3.12 B, C). The V66N mutation had a limited negative effect on 5mC DNA binding (the K_D increased approx. 2.5 fold to ≈0.6 μM, Figure 3.11 B), but had a positive effect on 5hmC binding (0.4-0.6 μM K_Ds, as determined in different assays, are either unchanged or up to 1.5-fold lower in comparison to WT McrB-N, Figure 3.12 B). This makes 5hmC the preferred DNA modification for the V66N and Y64F_V66N McrB-N mutants.

3.3.4 Relative affinity of McrB-N to different DNA modifications

Co-crystal structures of WT McrB-N were obtained with DNA oligoduplexes containing unmodified cytosine, 5mC (Sukackaite et al., 2012), 5hmC, 5fC and 4mC (this study). In all cases the (modified) cytosine base is placed into the protein pocket where it occupies an identical position, which allows formation of cytosine-specific hydrogen bonds between the Watson-Crick edge of the cytosine and the McrB-N pocket residues (Figure 3.10 A). Here we show that despite a similar DNA distortion, the affinity of McrB-N to DNA target sequences with various cytosine variants differs dramatically, with the values of equilibrium dissociation constants (K_Ds) ranging from 0.1 μM (4mC) to over 4 μM (C, 5fC, 5caC), Figure 3.12 B, C. We attribute these differences in affinity to 2 major factors: (i) different energetic cost for the disruption of the cytosine derivative:guanine base pair in duplex DNA; and (ii) different energetic gain for placing the flipped base in a hydrophobic protein pocket, defined by the energy of the newly formed van der Waals interactions and H-bonds, extent of steric clashes and the energetic cost for the flipped base and protein pocket dehydration.

The smallest interaction surface between the flipped base and the protein pocket, and, consequently, the lowest gain in vdW interaction energy upon base flipping, is observed for the unmodified cytosine C; this correlates with the lack of interactions with the C DNA (Figure 3.11) and lack of unmodified DNA cleavage by the McrBC complex (Sutherland et al., 1992). The specific-like binding and base flipping observed in the co-crystal structure with the C DNA (PDB ID: 3sse, (Sukackaite et al., 2012)) therefore must be a result of high protein / DNA concentrations used in the crystallization experiments and the effects of crystal packing forces.

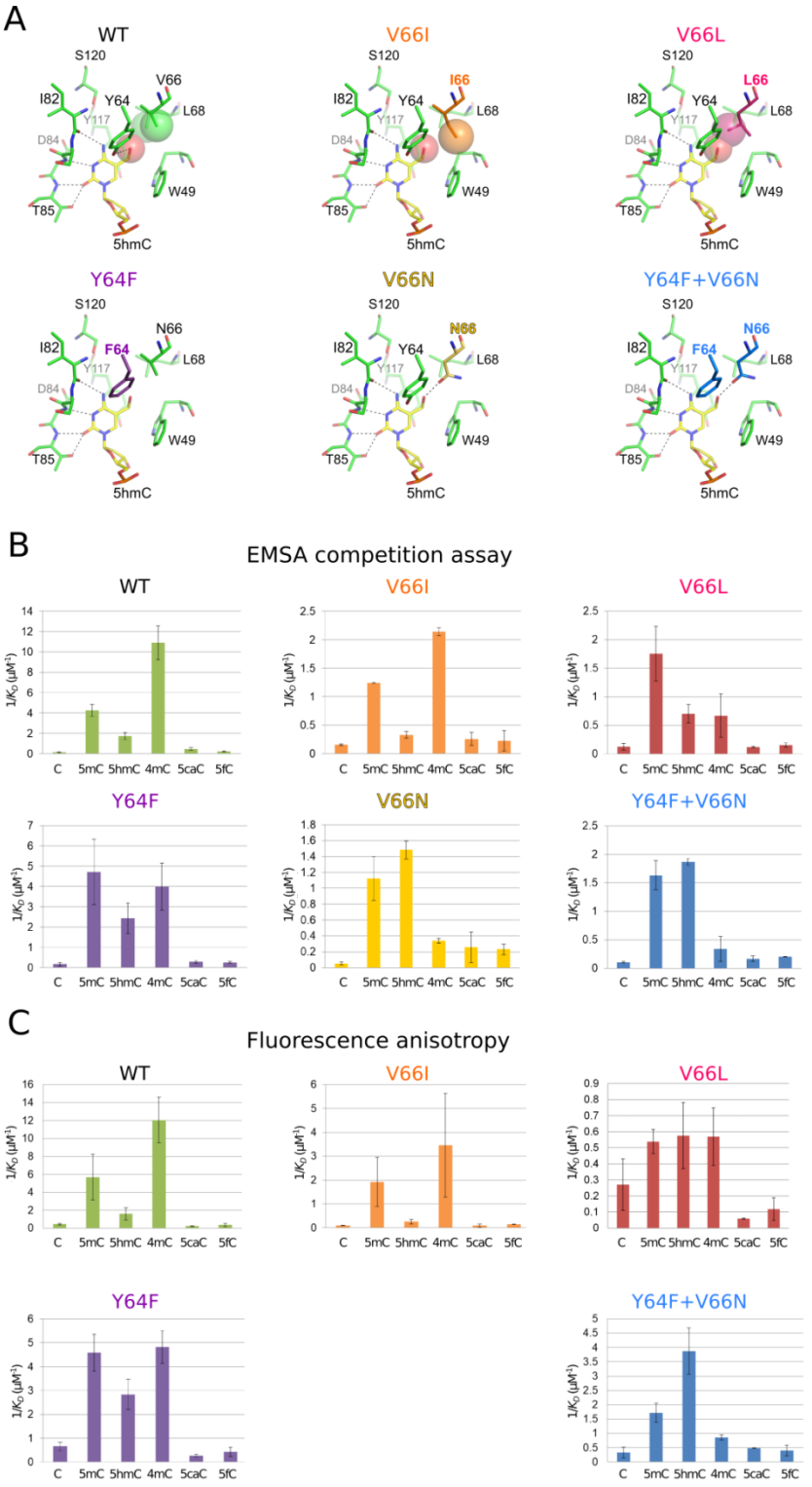

A
WT
V66I
V66L
Y64F
V66N
Y64F+V66N
S120
I82
Y64
Y117
D84
T85
W49
L68
V66
I66
L66
F64
N66
5hmC
B
EMSA competition assay
1/K_D (μM^-1)
C
5mC
5hmC
4mC
5caC
5fC
C
Fluorescence anisotropy

Figure 3.12 Selectivity of McrB-N pocket mutants. (A) Models of flipped-out 5hmC interactions with various McrB-N mutants. Experimentally determined DNA binding constants (in μM^{-1}) of McrB-N mutants with various DNA variants are shown in panels B (EMSA competition assay) and C (fluorescence anisotropy measurements). Error bars correspond to SD.

In contrast, cytosine variants 4mC, 5mC, 5hmC and 5fC have a larger contact surface with the pocket walls thanks to the extra methyl, hydroxymethyl, and formyl groups (Figure 3.10). In particular, the extra methyl groups of 4mC and 5mC make contacts to the pocket residues V66, L68 and S120 (4mC) or W49 (5mC). The approx. 2.5-fold difference in the K_Ds observed for 4mC and 5mC DNAs, which corresponds to a 2 kJ/mol difference in the binding energy ΔG, a value that does not exceed the energy of a single H-bond. Thus, the observed preference for 4mC must be due to small differences in the extent of vdW interactions, lower stability of the 4mC:G base pair as judged by lower melting temperature of 4mC-containingoligoduplexes (Butkus et al., 1987), and lower dehydration energy of 4mC compared to 5mC (due to the N4-methyl group 4mC is expected to form less H-bonds to water molecules that must be broken upon insertion into the protein pocket).

5hmC, 5fC and 5caC have larger modifications than 4mC and 5mC, but the extra size, and potentially larger interaction surface with the protein pocket do not translate into tighter binding (Figure 3.12). The affinity of McrB-N to 4mC DNA is 5-fold higher compared to 5hmC DNA (approx. 4 kJ/mol difference in ΔG), and more than 40-fold higher compared to 5fC and 5caC DNA (approx. 10 kJ/mol difference in ΔG, Figure 3.12 B, C). The solved structures with 5hmC and 5fC DNAs did not reveal any steric clashes between the flipped bases and the protein pocket; moreover, the extent of vdW interactions formed by the flipped 5hmC and 5fC bases is similar, neither base is capable of forming additional H-bonds, and none is reported to increase DNA duplex stability as judged by imino proton exchange rates (Szulik et al., 2015). This implies that slightly lower affinity of McrB-N to 5hmC DNA, and very low affinity to 5fC DNA must be due to other factors, primarily higher dehydration energy of polar hydroxymethyl and especially formyl groups, and possibly differences in the energies of vdW interactions due to distinct polarization properties of 5hmC and 5fC.

In contrast, the lack of interactions with 5caC DNA could simply be due to steric clashes (modeling, based on the McrB-N structure with 5fC DNA suggests that an additional 5caC carboxyl oxygen would be as close as 2.6 Å to the W49 residue), and increased stability of DNA duplex containing 5caC:guanine base pairs (Szulik et al., 2015).

3.3.5 Overview of McrB-N pocket mutations

We show that the relative preference of McrB-N to DNAs containing different cytosine modifications may be altered by replacement of residues forming the walls of the flipped base binding pocket. The most prominent effect is caused by the V66N mutation, which essentially abrogates McrB-N binding to 4mC DNA, and concomitantly has a positive effect on 5hmC DNA binding (Figure 3.12). The former effect is likely due to the fact that the N4-methyl group of 4mC is unable to form an energetically favorable interaction with the N66 asparagine; in contrast, 5hmC may form an additional H-bond (Figure 3.12 A).

3.3.6 Selectivity of the base-flipping 5(h)mC readers

The observed range of McrB-N affinities to cytosine derivatives, and the changes in the binding preferences due to mutations, are limited to approx. 40-fold (Figure 3.12). However, even such mediocre discriminating power is on par with, or exceeds these reported for eukaryotic SRA domains UHRF1 (a ~3.5-fold preference for 5mC compared to unmodified C (Zhou et al., 2014)) and UHRF2 (a *bona fide* 5hmC reader in eukaryotes (Spruijt et al., 2013), a ~1.5-fold preference for 5hmC compared to C (Zhou et al., 2014)). A much higher discriminating power, as judged from the activity measurements, is expected for the SRA-like domains found in the prokaryotic methyl-directed endonucleases of MspJI and PvuRts1I families (Wang et al., 2011; Zheng et al., 2010). For example, MspJI-like enzymes cut DNA containing 5mC and 5hmC modifications in a specific sequence context, and, like the McrBC complex, are inactive on unmodified DNA, while PvuRts1I-like enzymes have a 250-fold or higher preference for 5hmC or glucosylated 5hmC as compared to 5mC and unmodified C (Wang et al., 2011). Currently available structures of 5mC-bound MspJI enzyme (Horton et al., 2014b), and the apo-structures of PvuRts1I family (Horton et al., 2014c; Kazrani et al., 2014) enzymes, however, provide no obvious explanation for the higher selectivity of their SRA-like domains.

Despite identical positioning of the flipped base observed in the co-crystal structures of McrB-N with DNAs carrying various cytosine modifications, we find that McrB-N has relatively high affinity for 4mC, 5mC and 5hmC, but low affinity for 5fC. Strong discrimination against 5fC, which is almost isosteric to 5hmC, argues that modified cytosine recognition selectivity of McrB-N, and likely other base-flipping modified cytosine 'readers', in addition to intra-pocket contacts, is determined by other factors, such as the stability of the flipped base pair and the flipped base / protein

pocket dehydration energies. Low affinity to 5fC and 5caC *vs* high affinity to 5mC and 5hmC also implies that methylome profiling studies that utilized McrBC as a tool for DNA fragmentation (Burman et al., 1999; Chotai and Payne, 1998; Lippman et al., 2004, 2005; Lyko et al., 1999) reported the distribution of 5mC and 5hmC, but not 5fC and 5caC. It remains to be seen if McrB-N could be tailored for more selective recognition of individual cytosine variants.

3.4 Activity and structure of McrA

McrA from *E. coli* K strains restricts modified DNA only when the methylation is in an appropriate sequence context. The enzyme is very effective in restricting DNA that has been methylated by M.HpaII in C5mCGG (Card et al., 1990). Until now, it has been impossible to reproduce DNA cleavage by McrA *in vitro*, even using plasmids that are restricted in cells (e. g., plasmid with M.HpaII which methylates sequence C5mCGG) (Mulligan and Dunn, 2008). However, it has been shown that McrA exhibits methylation dependent DNA binding, which can be assayed biochemically. EMSAs suggest that the McrA specificity is Y5mCGR (Mulligan et al., 2010), i.e. slightly broader than required to cleave M.HpaII methylated DNA (Card et al., 1990). How McrA achieves its sequence and modification specificity is incompletely understood. Sequence analysis suggests the presence of an N-terminal domain that is not readily amenable to bioinformatics analysis, and a C-terminal domain that can be confidently classified as an HNH nuclease domain (Bujnicki et al., 2000). McrA is not only of interest in its own right, but it can also be regarded as the prototype of a larger family of related enzymes. BLASTP searches readily identify proteins from eu and archaebacteria as similar to McrA. However, in many cases, the similarity is clearly limited to the HNH domain. In the present study we characterized McrA biochemistry including the first demonstration of its *in vitro* activity and colleagues provided crystal structure of the enzyme in the absence of DNA (Czapinska et al., 2018).

3.4.1 McrA activity *in vitro*

We tested the catalytic activity of McrA using 30 nucleotide long oligoduplex substrates with unmodified, hemi- or fully methylated C5mCGG sites in two different positions; with either top or bottom strand radiolabeled (Table 2.1). The pilot cleavage reactions were performed with unmodified and hemimethylated substrates in a buffer supplemented with various amounts of divalent metal ions (Figure 3.13). Cleavage rates were highest

for the Mn^{2+} ions in the concentration range of 0.01 mM to 0.2 mM. In the presence of Mn^{2+}, it was possible to demonstrate the concerted cleavage of both DNA strands, most likely with a single nucleotide stagger (leading to 3′ overhangs).

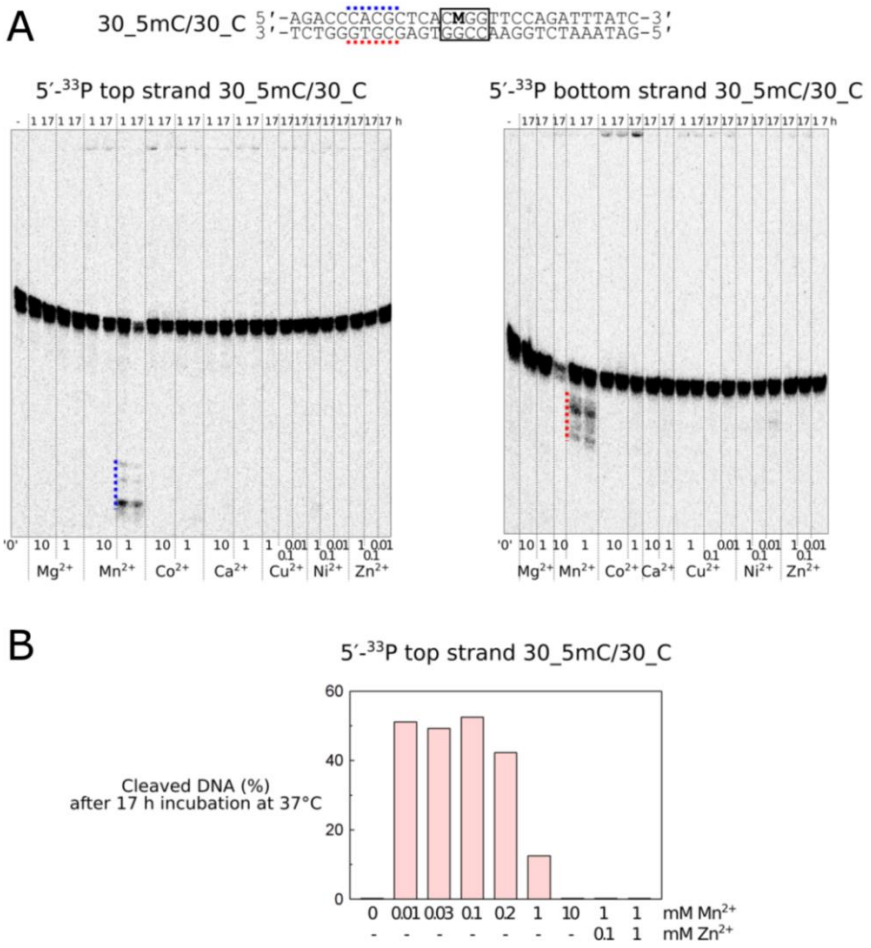


Figure 3.13 Catalytic activity of McrA in the presence of divalent metal ions. (A) Reactions of wt McrA (0.5 µM dimer) on 0.2 µM radiolabeled oligoduplex DNA 30_5mC/30_C (DNA sequence shown at the top, 'M' designates 5-methylcytosine). Incubation times in hours are indicated above the gel lanes, divalent metal ions and their concentrations in mM are shown below gel lanes. Sample '0' is untreated DNA. Approximate cleavage positions observed in the presence of Mn^{2+} are marked by blue (top strand) or red (bottom strand) dotted lines. (B) Quantification of 30_5mC/30_C DNA top strand cleavage by wt McrA at various Mn^{2+} and Zn^{2+} concentrations.

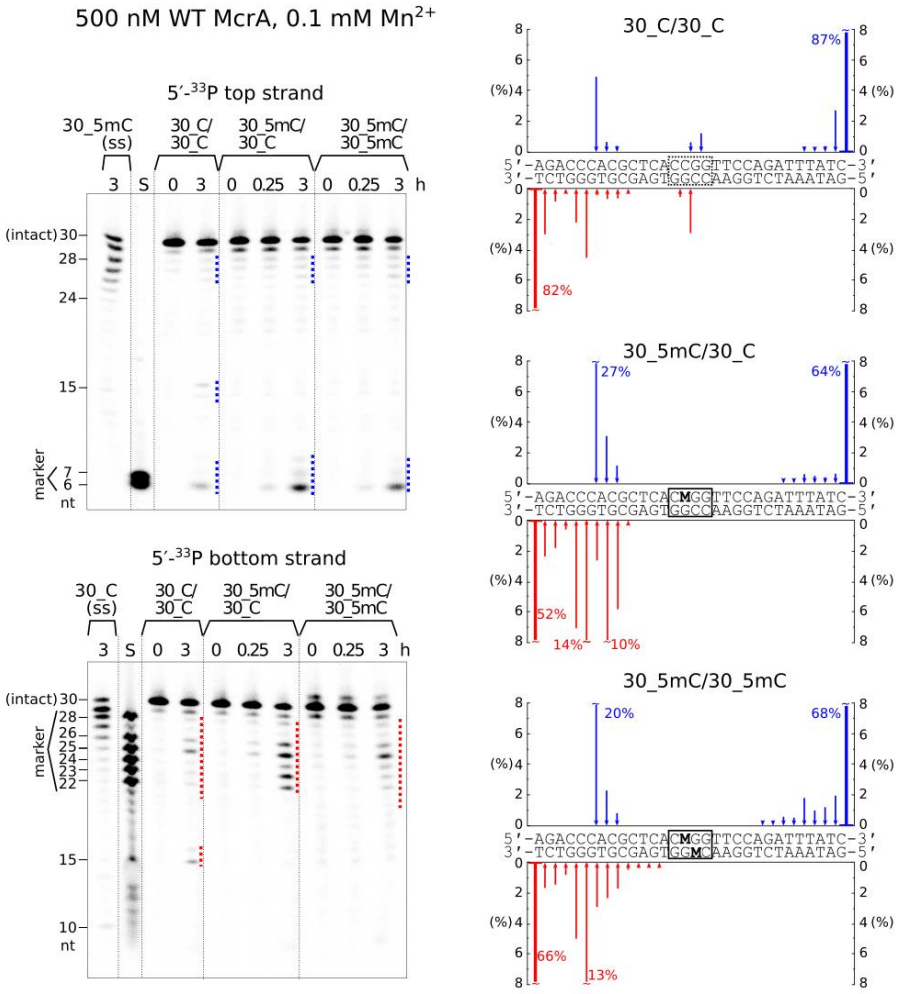


Figure 3.14 DNA cleavage by McrA. The reactions were performed with 0.5 µM wild type McrA (dimer) and 0.2 µM of unmodified, hemi- or fully methylated oligoduplex DNA or single stranded DNA, at 37 °C in a buffer supplemented with 0.1 mM Mn^{2+}. Top strand and bottom strand cleavage positions on the gels are marked by blue and red dotted lines, respectively. Gel lanes 'S' contained radiolabeled single-stranded oligonucleotides corresponding to the 5′-terminal fragments of the respective DNA strands (sizes in nucleotides are shown on the sides of the gels). The amounts of the bottom and top strand cleavage products after 3 hours of digestion are plotted as blue / red arrows along the oligoduplex sequences. The uncleaved oligonucleotides (the 0 hour lanes) contained a detectable fraction (approx. 10%) of shorter fragments. The reported cleavage data was corrected for the amount of these contaminants.

The extent of cleavage tended to be promoted by DNA methylation, but the effect was substrate dependent (Figure 3.14). The oligonucleotides were predominantly cleaved at multiple positions 5-15 nt upstream of the

C5mCGG sequence. The unmethylated DNA was cut at identical positions, with additional cleavage observed at the unmodified CCGG sites (Figures 3.14). Wild type McrA also displayed methylation independent exonuclease activity on single-stranded oligonucleotides (Figure 3.14).

Based on a prediction of the McrA active site (Bujnicki et al., 2000), the H228A, H229A, H252A and H256A variants of the enzyme were expected to be catalytically impaired or inactive. Indeed, none of these mutants generated any cleavage products (Figure 3.15), excluding a contaminating endonuclease as the source of the observed activity. Together, the data suggest that McrA is catalytically active *in vitro*, but the activity requires high enzyme concentrations, and in such conditions the enzyme exhibits only a moderate preference for methylated DNA.

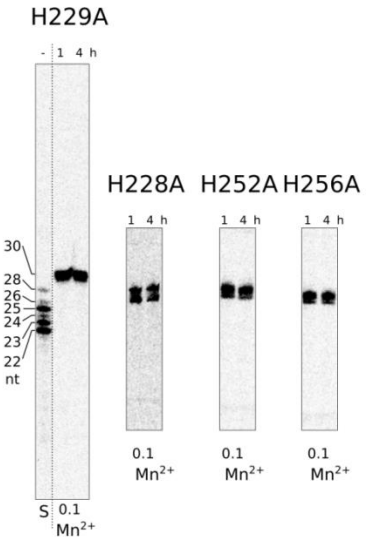


Figure 3.15 Inactivity of McrA mutants on oligoduplex DNA. The reactions contained 0.5 µM (dimer) McrA mutants and 0.2 µM 30_5mC/30_C DNA (radiolabel on the bottom strand) in a reaction buffer supplemented with 0.1 mM $MnCl_2$. Reactions were performed for up to 4 hours at 37 °C.

McrA was active not only on short linear DNA oligoduplexes, but also on supercoiled plasmids and phage λ DNA (Figure 3.16). Plasmid / phage DNA cleavage reactions required high (0.5 µM) enzyme concentrations, as the yield of cleavage products was much lower at 50 nM concentration. Under these conditions, there was surprisingly little effect of methylation in dcm (C5mCWGG) or HpaII (C5mCGG) context. Full digestion of the supercoiled unmethylated and methylated plasmids required 2 h incubation with 0.5 µM enzyme.

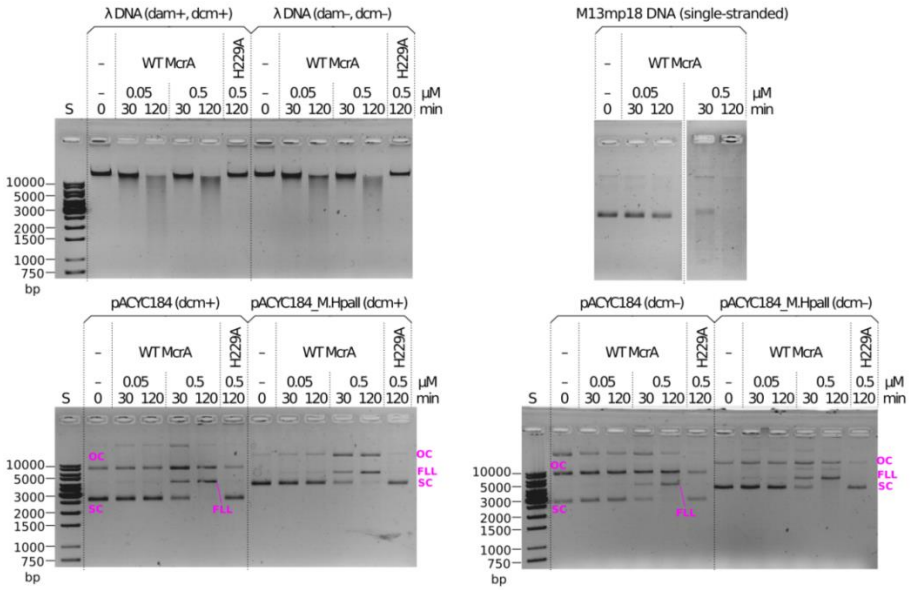


Figure 3.16 ***In vitro*** **phage and plasmid DNA cleavage by McrA.** The reactions were performed on phage λ DNA (isolated from either a (dam+, dcm+) or (dam–, dcm–) *E. coli* strain), single-stranded M13mp18 DNA, supercoiled plasmid pACYC184 and supercoiled plasmid pACYC184_M.HpaII, each plasmid isolated from either a dcm+ or dcm– *E. coli* strain. Gel lane 'S' contained DNA size marker. The positions of supercoiled, nicked (one or multiple nicks), and linear (a single double-strand break) DNA forms are marked as 'SC', 'OC' and 'FLL'. pACYC184 (dcm+ and dcm–) contained a detectable amount of heavier, presumably dimeric DNA forms. The reactions contained 0.05 or 0.5 μM wild type McrA, 0.5 μM H229A McrA variant and 0.5 μg / 25 μl DNA in a reaction buffer supplemented with 0.1 mM $MnCl_2$. The incubations were performed at 37 °C.

3.4.2 McrA nuclease activity for plasmid restriction

The weak *in vitro* activity of McrA made us question whether the catalytic activity of McrA was relevant for restriction in cells. In order to test this, we generated *E. coli* BL21(DE3) cells (McrA-) expressing wild type McrA, its H228A, H229A and N-terminal variants (all catalytically impaired), or a control protein (REM14), from a plasmid maintained in the cells using ampicillin selection. We also prepared pACYC184 plasmid, carrying a chloramphenicol resistance (Cm^R) gene, without insert ("empty"), or with an insert coding for M.HpaII. The CCGG sites in this vector were confirmed by HpaII digestion to be unmethylated and methylated, respectively. After transformation of the two plasmids into cells overexpressing McrA or its variants, cells were subjected to double antibiotic selection, and survivor colonies were counted. For low or moderate protein overexpression (glucose repression or no induction), full

restriction required both the M.HpaII methylation of the test plasmid, and the presence of an intact McrA active site. High protein overexpression did not affect the colony formation only when the plasmid bearing a gene for an unrelated control protein was used (Figure 3.17).

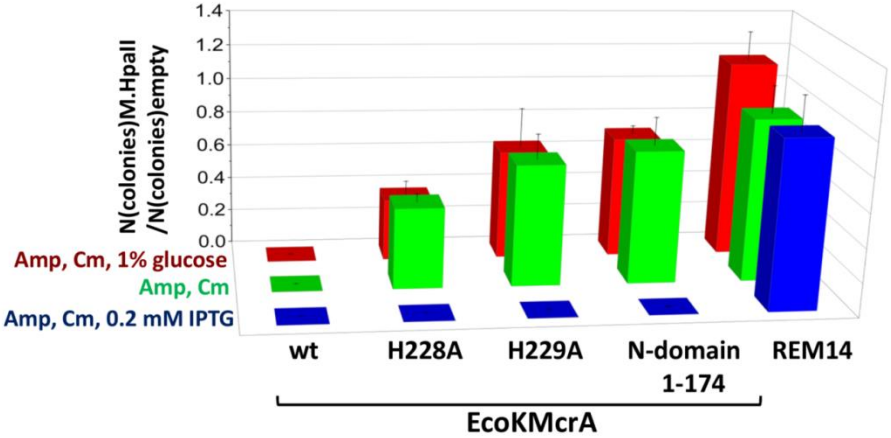


Figure 3.17 Plasmid restriction assay. Quantification of the restriction of empty and M.HpaII carrying plasmids by *E. coli* cells expressing either wild type McrA, two different active site variants of the enzyme, its N-terminal fragment lacking the catalytic domain or an unrelated protein (REM14). The pACYC184 plasmid carrying a chloramphenicol resistance gene was assayed in BL21(DE3) (McrA-). The plasmid was either empty and thus without cytosine modifications or carried a gene for M.HpaII methyltransferase and thus was 5mC modified in the C5mCGG sequence context assumed to be among the McrA targets.

3.4.3 McrA structure

Crystal structure of the McrA K196E variant was solved in the absence of DNA at 2.85 Å resolution (prof. M. Bochtler group, Warsaw). The McrA structure confirms the expected two-domain architecture of the enzyme, with a previously uncharacterized N-terminal domain, and a C-terminal HNH domain predicted by prior bioinformatics work (Figure 3.18 A). The nuclease domains clearly follow the two-fold symmetry; the N-terminal domains do not. This arrangement results in two distinct conformations of full length protomers, a compact and an extended one. The boundary was placed between the N- and C-terminal domains and the linker at the switch points between the structurally similar and dissimilar regions. This choice assigns residues 147-276 to the HNH domain and residues 1-143 to the N-terminal domain. The drastically different relative orientations of the two domains in the protomers are therefore due to different conformations of a short linker region spanning only a few residues (144-146) and folding into a short helix in the extended form of the protomer (Czapinska et al., 2018).

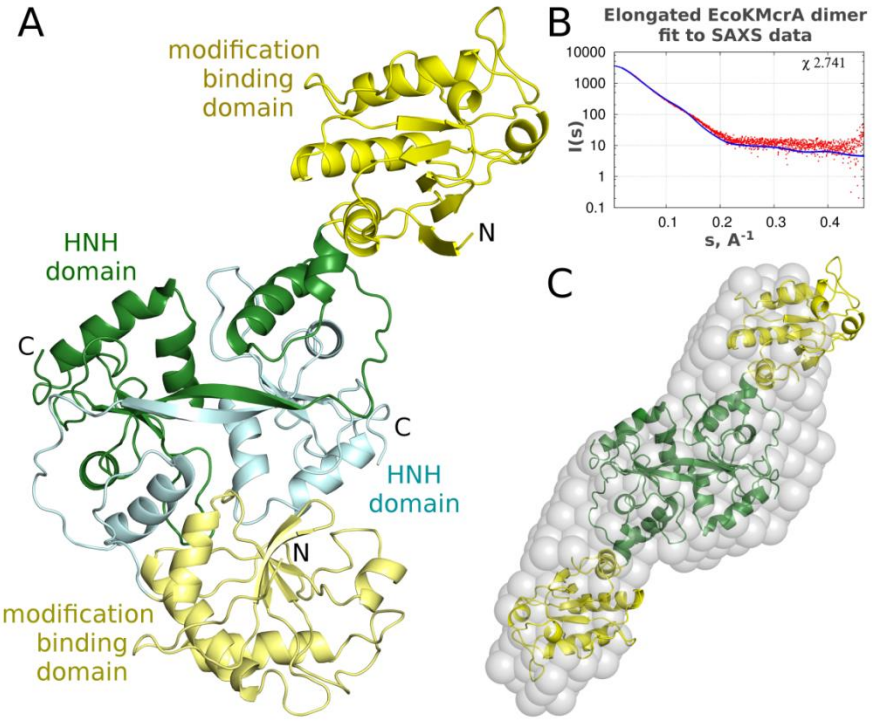


Figure 3.18 Structure of McrA. (A) Ribbon diagram of the McrA dimer in the asymmetric unit of the crystals, colored according to domains. The domain organization of each protomer of is shown below. (B) Comparison of a symmetrized model of the McrA dimer based on the more elongated protomer, with small-angle X-ray scattering (SAXS) data for the protein in the absence of DNA. (C) McrA model *ab initio* calculated from the SAXS data (grey spheres) overlaid with the symmetrized dimer that was used for the calculation of the predicted small angle X-ray scattering data presented in panel B.

The overall structure of McrA is in agreement with the previously reported dimeric form of the enzyme and suggests that the N-terminal domains on their own should be monomers. This was tested for the N-terminal McrA fragment (residues 1-174) by analytical sizing chromatography in high salt conditions, to prevent unspecific protein interactions with the column material. The fragment boundaries were selected prior to the structure solution and thus the N-terminal domain includes a domain swapped fragment of the HNH-domain. As a control, we confirmed that full-length McrA migrated as a dimer (apparent mass 54.4 kDa, compared to the expected mass for the dimer 64.5 kDa). In contrast, the N-terminal fragment eluted as a monomer (expected mass including His_6-tag 20.6 kDa, apparent mass 20.1 kDa). Experiments in the

presence of DNA could not be done meaningfully, because the required high ionic strength of the buffers (0.5 M) interfered with protein DNA binding, resulting in separate elution of the two components (Figure 3.19).

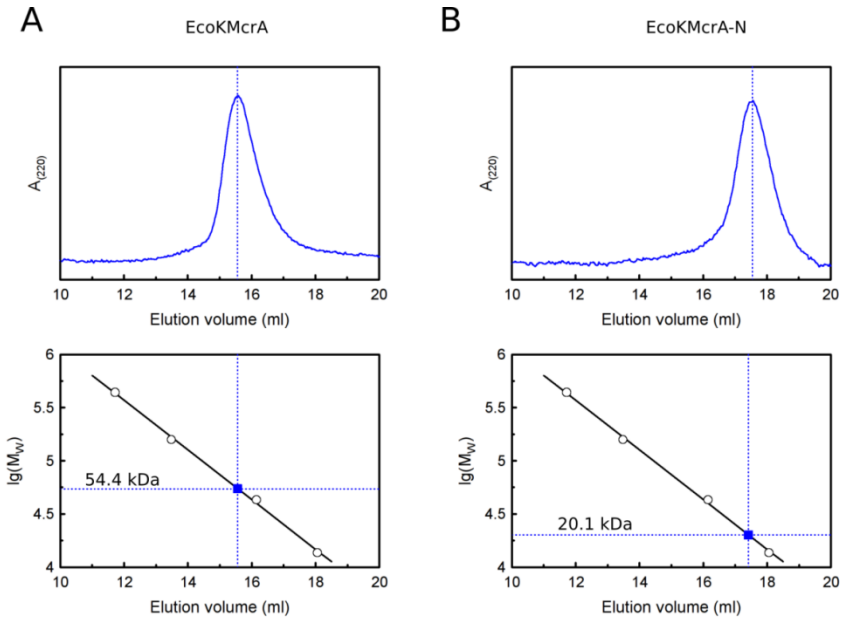


Figure 3.19 Gel filtration of wt McrA and its N-terminal fragment McrA-N. (A) The apparent MW of McrA determined by gel-filtration (15.55 ml elution volume, equivalent to 54.4 kDa, blue square) is close to the theoretical mass of McrA dimer (64.5 kDa). (B) The apparent MW of the N-terminal McrA fragment McrA-N (17.41 ml elution volume, equivalent to 20.1 kDa, blue square), is a close match to the theoretical mass of the McrA-N monomer (20.6 kDa). The proteins used as MW standards were ribonuclease A (13.7 kDa), ovalbumin (43 kDa), ferritin (440 kDa), aldolase (158 kDa).

In the crystal structure, the N-terminal domains pack very differently against the HNH-domain dimer. One domain interacts quite extensively, the other one makes contacts only to the crystallographic neighbors. The relative domain orientations of the two protomers differ by 150° (Czapinska et al., 2018). This suggested that the linkers are flexible and allow for multiple domain orientations. Alternatively, one of the observed conformations might result from the crystal packing effects.

To determine the solution conformations and to independently measure the molecular masses, we carried out small angle X-ray scattering (SAXS) experiments using full-length McrA, and its N-terminal fragment. There was calculated the *ab initio* shape of the enzyme in solution using SAXS data and the DAMMIF software (Franke and Svergun, 2009) with no symmetry

restraints. This shape was compared with the crystalized dimer, and the symmetrized dimers composed of the protomers in one conformation only. The symmetric, elongated dimer fitted best, judging from the real space fit and the χ parameters, which measure the agreement between calculated and observed scattering data in reciprocal space (Figs. 3.18 B, C). The SAXS masses for the full-length protein (between 58.7 kDa and 77.6 kDa, (Czapinska et al., 2018)) agreed with the expected mass of the dimer. The SAXS masses for McrA-N (between 16.8 kDa and 23.7 kDa, (Czapinska et al., 2018)) confirmed the monomeric state of the N-terminal fragment in isolation.

3.4.4 Modification specificity of the McrA N-terminal domain

The domain architecture of McrA is reminiscent of other methyl directed endonucleases, which are built as fusions of nuclease and modification binding domains. This makes the N-terminal domain of McrA the prime candidate for the recognition of methylated DNA. We carried out electrophoretic mobility shift experiments using the N-terminal fragment of McrA and the DNA that was previously used in the activity assay. The data indicated tighter binding of the methylated DNA compared to the unmodified DNA, confirming that McrA-N is the methylation specific domain (Figure 3.20 A). Methyl specific binding was also confirmed by the DNase I footprint of McrA-N. The protection from DNase I cleavage was observed only for the methylated DNA (Figure 3.20 B). The protected region around the 5mC was identical to the region covered by full-length McrA (Czapinska et al., 2018), but for the full-length enzyme, we additionally observed scattered footprint for unmethylated DNA. This suggests that McrA-N has higher overall selectivity for the methylated site than the full-length enzyme. The methyl-independent DNA binding observed with McrA is therefore likely due to the DNA interaction with the C-terminal HNH domains.

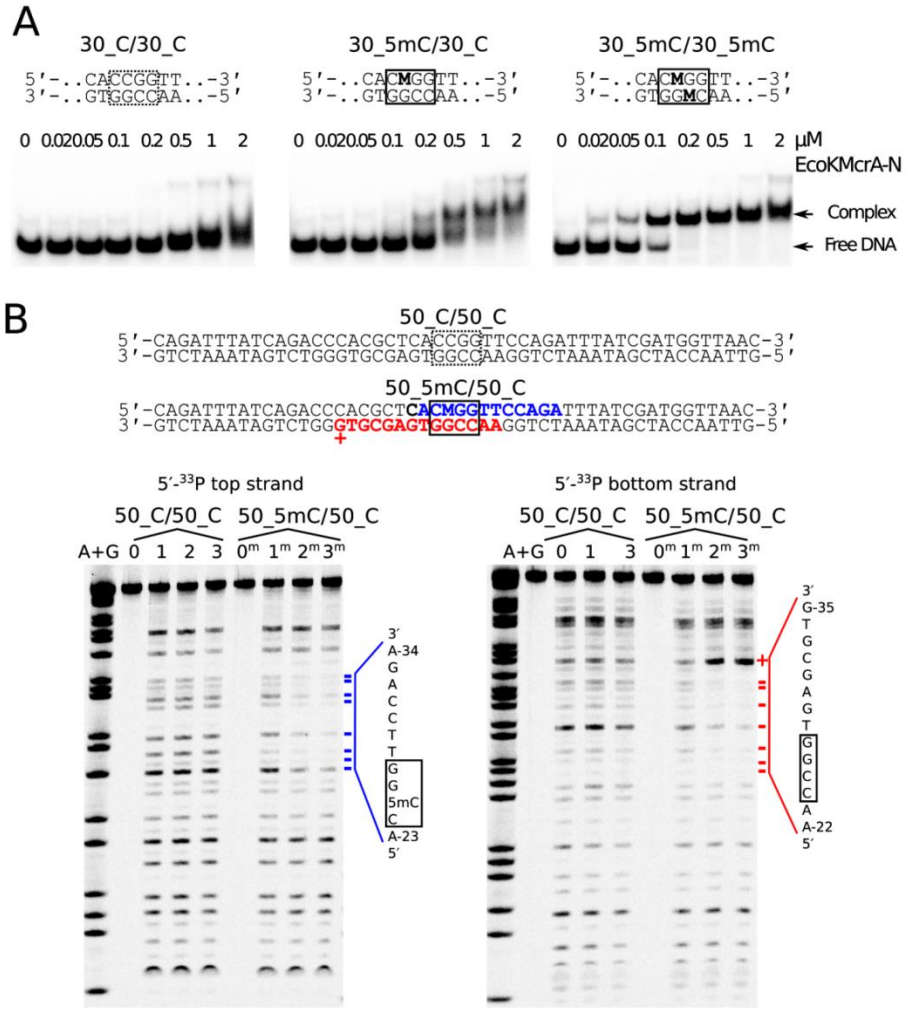


Figure 3.20 DNA binding by the N-terminal fragment of McrA (McrA-N). (A) EMSA. DNA was unmodified, hemi- or fully methylated (central sequences shown on top; 'M' denotes 5mC). DNA concentration was 50 nM, concentrations of McrA-N are indicated above gel lanes. Positions of free DNA and protein-DNA complexes are indicated. (B) DNase I footprint. Top: the region of the DNA protected from DNase I cleavage by McrA-N binding is shown in bold font. The bottom strand G-35 position that becomes more susceptible to DNase I treatment upon McrA-N binding is shown by '+'. Bottom: DNA protection by McrA-N. Gel lanes '0' contained untreated unmodified DNA, '1' – the DNA treated with DNase I in the absence of McrA-N, lanes '2' and '3' – the DNA treated with DNase I in the presence of 0.5 μM and 1 μM McrA-N, respectively. Lanes '0m', '1m', '2m' and '3m' contained analogous samples prepared with hemimethylated DNA. Size markers (lanes 'A+G') were generated using a standard Maxam-Gilbert sequencing reaction. Positions of the DNA protected from DNase I cleavage upon McrA-N binding are marked by blue (top strand) or red (bottom strand) dashes. The sequences of the protected regions are shown on the right-hand side of the gels.

3.4.5 No indication for modified base flipping by McrA in fluorescence assay using pyrrolocytosine

In order to check whether nucleotide flipping played a role in the interactions of the McrA N-terminal domain with DNA, we used a fluorescence based assay. Meaningful use of the pyC analogue requires that the enzyme accepts it instead of the physiologically relevant cytosine bases. In the case of McrA, a hemimodified 12-mer oligoduplex containing pyC (Table 2.1) bound more tightly than unmodified control, but less tightly than the duplex containing 5mC (Figure 3.21 A).

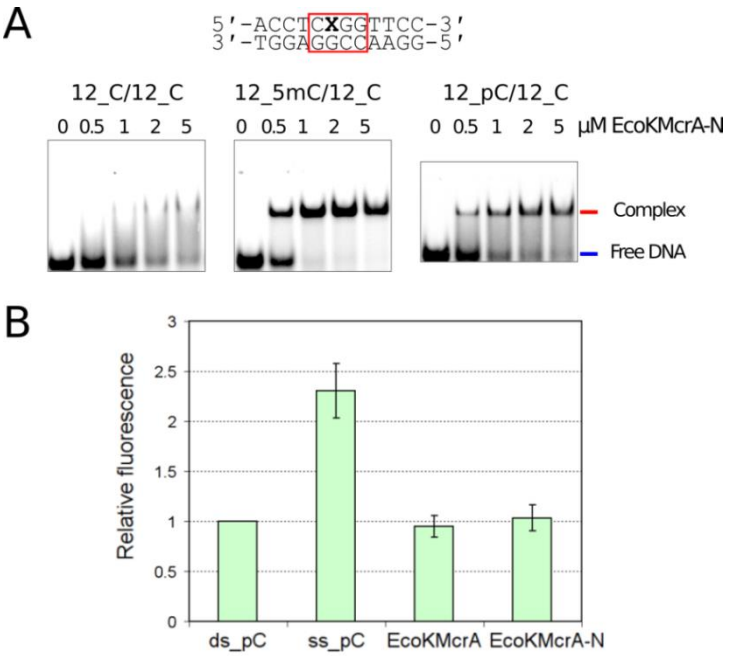


Figure 3.21 McrA-N interaction with pyC containing DNA. (A) EMSA performed with unmodified and hemimodified 12 bp DNA oligoduplexes. The DNA sequence is shown at the top of the panel, 'X' designates unmodified C (oligoduplex 12_C/12_C), 5mC (12_5mC/12_C), or pyrrolocytosine (12_pC/12_C). DNA concentration was 0.5 µM, protein concentrations are indicated above gel lanes. Experiments were performed in a 40 mM Tris-acetate pH 8.3 buffer. Positions of free DNA and the specific protein-DNA complex are marked. (B) pyC steady state fluorescence measurements in solution. The 'McrA-N' and 'McrA' samples contained the respective protein (1 µM McrA dimer, 2 µM McrA-N monomer) and 0.2 µM oligoduplex 12_pC/12_C. Control sample 'ds_pC' contained 0.2 µM of the double-stranded oligonucleotide 12_pC/12_C, the sample 'ss_pC' contained 0.2 µM of the modified (top) 12_pC strand. The emission intensities of the 'McrA-N', 'McrA' and 'ss_pC' samples were normalized against the 'ds_pC' sample and are presented as the average value of 3 independent measurements ±SD.

The dsDNA duplex containing pyC had very similar properties in the absence of the protein, and in the presence of either full length McrA or its N-terminal fragment. (Figure 3.21 B). This negative result speaks against nucleotide flipping, but could also be due to quenching of the flipped pyC fluorescence by aromatic amino acid residues of a binding pocket. According to the crystal structure, such a pocket is unlikely to be present in the N-terminal fragment of McrA. Therefore, the data suggest that the enzyme recognizes modified cytosine bases in the context of dsDNA without nucleotide flipping.

3.4.6 Two binding sites for the modified DNA in the McrA dimer

The structural data suggest that McrA should have one joint DNA binding site in the nuclease domain, and two separate modification dependent DNA binding domains. In order to verify this model, we first tested wild type McrA for preferential binding of methylated over unmethylated DNA in an EMSA assay, using either 30mer or 12mer duplexes. As observed for the N-terminal domain alone, binding was promoted by DNA methylation. However, in agreement with the DNase I footprinting results, for 30 bp long substrates the dependence on methylation was more pronounced for McrA-N than for the full length enzyme (Figure 3.22 A). We interpret these data as evidence that the nuclease domain contributes some unspecific affinity to DNA for the longer substrates, but not the short oligoduplexes that are largely covered by the N-terminal domain. EMSA experiments with full-length enzyme, but not McrA-N, showed the presence of two bands of retarded electrophoretic mobility (Figures 3.20 A and 3.22 A), as reported earlier, and attributed to a monomer dimer equilibrium (Mulligan et al., 2010). In the light of the biochemical and structural data, the two shifted bands more plausibly correspond to McrA with either one or two bound oligoduplexes. To test this model, we carried out EMSA experiments with modified oligoduplexes of two different lengths mixed in various ratios (Figure 3.22 B). We observed the presence of three methyl-dependent complexes of distinct mobilities, which is consistent with the two independent binding sites for methylated DNA in the McrA dimer.

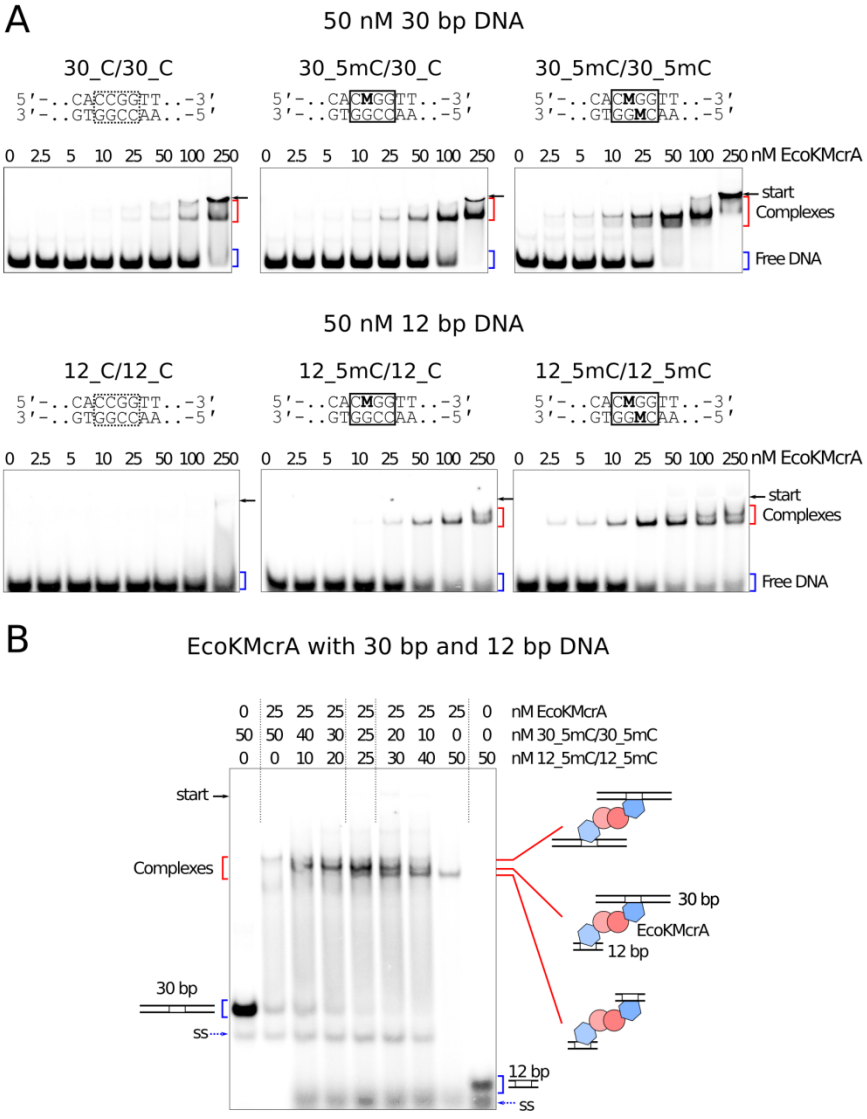


Figure 3.22 Electrophoretic mobility shift assay of DNA binding by McrA. (A) EMSA with unmodified, hemi- or fully methylated 30 or 12 bp DNA (central sequences shown on top; 'M' denotes 5mC), DNA concentration was 50 nM, concentrations of McrA dimer are indicated above gel lanes. Positions of free DNA and protein-DNA complexes are marked by blue and red brackets, respectively. The gels were run for 1 h at 5 V/cm. (B) EMSA with a mixture of radiolabeled 30 bp (30_5mC/30_5mC) and 12 bp (12_5mC/12_5mC) DNA oligoduplexes. Concentrations of McrA and DNAs are indicated above gel lanes. The gel was run for 3 h at 5 V/cm. Cartoons depict two types of unbound oligoduplexes (12 and 30 bp) and three types of protein-DNA complexes, containing two 30 bp, two 12 bp, or one of each DNAs. McrA N- and C-terminal domains are depicted as blue hexagons and red circles, respectively. Bands of the single-stranded forms of the respective oligonucleotides are marked by blue dotted lines.

3.4.7 Organization and activity of McrA

The N-terminal domain of McrA binds to DNA in a methylation dependent way. EMSA assays indicate that the affinity is highest for fully-methylated DNA, lower for hemi-methylated DNA, and lowest for unmethylated DNA. The structure of the N-terminal domain of McrA is surprisingly dissimilar from other modification specific DNA binding domains. In particular, the domain is not structurally similar to SRA domains and unrelated to the McrB-N domain responsible for the modification dependence of McrBC (Sukackaite et al., 2012). The DNA binding mode of the N-terminal fragment of McrA is still unknown. The pyrrolocytosine assay does not provide evidence for nucleotide flipping, although alone is not conclusive, however the structural data suggest that the enzyme recognizes modified cytosine bases in the context of dsDNA without nucleotide flipping.

The C-terminal domain of McrA is responsible for the nuclease activity. The domain has a typical HNH fold, with the expected active site residues, which are conserved in the McrA family. It is therefore surprising that demonstration of *in vitro* endonucleolytic activity has been so difficult. In our hands, McrA has weak, Mn^{2+} ion dependent endonuclease activity on double- and single-stranded DNA. The specific requirement for Mn^{2+} ions was unexpected, because many other HNH endonucleases have less stringent preference for the divalent metal cations in their active sites. Mn^{2+} ions may be available physiologically, thanks to mechanisms for active uptake, presumably from the host (Diaz-Ochoa et al., 2014). McrA generated nicks in both strands of the oligoduplex. The nicking of two DNA strands is consistent with concerted action resulting in double strand breaks (DSB). The most prominent nicks indicate 1 nucleotide 3′ cleavage stagger (Figure 3.13).

The overall organization of McrA with N-terminal modification specific, and C-terminal nuclease domain is reminiscent of the two-domain organization of the SRA-HNH (Han et al., 2015; Kisiala et al., 2018), PD-(D/E)XK-SRA (PvuRts1I) (Kazrani et al., 2014) and SRA-PD-(D/E)XK (MspJI) (Horton et al., 2012) modification specific restriction endonucleases. In all these enzymes, the largely unspecific nuclease domains mediate dimerization, and expose a single joint DNA binding site, whereas the modification specific domains expose separate binding sites for modified DNA. The crystal structure of McrA indicates that the nuclease domains of the enzyme indeed form a tight dimer, the conformation of which is compatible with concerted cleavage of two DNA strands. The results of the

gel filtration and SAXS experiments confirm that McrA is dimeric also in solution. Finally, the biochemical data indicate that each N-terminal domain binds double stranded DNA in modification dependent manner (Figure 3.22). However, the *in vitro* cleavage assay does not show a clear increase of the enzyme activity in the presence of two closely spaced modification sites (Czapinska et al., 2018).

Surprisingly, the McrA *in vitro* activity is weak, only partially methylation dependent. The activity of McrA on unmethylated DNA and the lack of dependence on two modification sites are probably a consequence of high protein concentrations required to achieve any DNA cleavage at all. Why such high protein concentrations and long incubation times are required remains puzzling, especially in the light of sequence and structural data that point to the conservation of the active site residues. The low *in vitro* activity of McrA also has to be reconciled with the efficacy of the enzyme as a restricting factor for plasmids with cytosine modifications. It is possible that an optimal substrate for *in vitro* experiments has not been identified. Such a substrate might only be created in cells, for example as a side product of DNA replication or transcription. Alternatively, McrA could require an activator or an effector nuclease, from the pool of host proteins. Our finding that the nuclease activity of McrA is required for efficient plasmid restriction makes it unlikely that the activity of the enzyme depends on the recruitment of another nuclease. The other possible explanations for the efficacy of McrA as a restricting factor remain to be explored.

CONCLUSIONS

1. A method based on $KMnO_4$ oxidation of extrahelical pyrimidines indicates that methyl-directed restriction endonucleases LpnPI, YkrI and BmeDI of MspJI and PvuRts1I families flip out modified cytosine from double stranded DNA.
2. MspJI family enzyme LpnPI recognizes 5-methylcytosine and its sequence context via several surface loops that act as separate DNA binding / recognition modules. Replacement of these loops by loops from homologous proteins AspBHI and SgrTI, which have different recognition sequences, generates LpnPI variants with altered sequence specificity.
3. McrBC DNA binding domain McrB-N recognizes cytosine variants with different modifications in the order of descending affinity: 4mC > 5mC > 5hmC >> 5fC. Preference for different cytosine modifications can be altered by mutations in the flipped base binding pocket.
4. McrA *in vitro* has weak, Mn^{2+} ion-dependent endonuclease activity, which is partly dependent on DNA methylation. DNA cleavage is performed by the C-terminal McrA HNH domain, which also is responsible for protein dimerization.
5. McrA N-terminal domain is responsible for modified cytosine recognition, which occurs without base flipping.

REFERENCES

1. Afonine, P. V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallogr. Sect. D Biol. Crystallogr. 68*, 352–367.
2. Antequera, F. and Bird, A. (1993). Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. U. S. A. 90*, 11995–11999.
3. Anton, B.P. and Raleigh, E.A. (2004). Transposon-Mediated Linker Insertion Scanning Mutagenesis of the Escherichia coli McrA Endonuclease. *J. Bacteriol. 186*, 5699–5707.
4. Arber, W. (2000). Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol. Rev. 24*, 1–7.
5. Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature 455*, 818–821.
6. Avvakumov, G. V, Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature 455*, 822–825.
7. Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., and Balasubramanian, S. (2014). 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem. 6*, 1049–1055.
8. Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H.E., Iurlaro, M., Reik, W., Murrell, A., and Balasubramanian, S. (2015). 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol. 11*, 555–557.
9. Bair, C.L. and Black, L.W. (2007). A Type IV Modification Dependent Restriction Nuclease that Targets Glucosylated Hydroxymethyl Cytosine Modified DNAs. *J. Mol. Biol. 366*, 768–778.
10. Bair, C.L., Rifat, D., and Black, L.W. (2007). Exclusion of Glucosyl-Hydroxymethylcytosine DNA Containing Bacteriophages Is Overcome by the Injected Protein Inhibitor IPI*. *J. Mol. Biol. 366*, 779–789.
11. Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science (80-.). 354*, 909–912.
12. Baubec, T., Ivánek, R., Lienert, F., and Schübeler, D. (2013). Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. *Cell 153*, 480–492.
13. Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I., Reuter, G., and Aalen, R.B. (2001). The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res. 29*, 4319–4333.
14. Berry, D.A., Jung, K.-Y., Wise, D.S., Sercel, A.D., Pearson, W.H., Mackie, H., Randolph, J.B., and Somers, R.L. (2004). Pyrrolo-dC and pyrrolo-C: fluorescent analogs of cytidine and 2′-deoxycytidine for the study of oligonucleotides. *Tetrahedron Lett. 45*, 2457–2461.

15. Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev. 16*, 6–21.
16. Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell 40*, 91–99.
17. Bischerour, J. and Chalmers, R. (2007). Base-flipping dynamics in a DNA hairpin processing reaction. *Nucleic Acids Res. 35*, 2584–2595.
18. Bochtler, M., Szczepanowski, R.H., Tamulaitis, G., Grazulis, S., Czapinska, H., Manakova, E., and Siksnys, V. (2006). Nucleotide flips determine the specificity of the Ecl18kI restriction endonuclease. *EMBO J. 25*, 2219–2229.
19. Borgaro, J.G. and Zhu, Z. (2013). Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases. *Nucleic Acids Res. 41*, 4198–4206.
20. Borst, P. and Sabatini, R. (2008). Base J: discovery, biosynthesis, and possible functions. *Annu. Rev. Microbiol. 62*, 235–251.
21. Bostick, M., Kim, J.K., Esteve, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science (80-.). 317*, 1760–1764.
22. Bourniquel, A.A. and Bickle, T.A. (2002). Complex restriction enzymes: NTP-driven molecular motors. *Biochimie 84*, 1047–1059.
23. Buck-Koehntop, B.A., Stanfield, R.L., Ekiert, D.C., Martinez-Yamout, M.A., Dyson, H.J., Wilson, I.A., and Wright, P.E. (2012). Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. *Proc. Natl. Acad. Sci. U. S. A. 109*, 15229–15234.
24. Bujnicki, J.M. and Rychlewski, L. (2001). Grouping together highly diverged PD-(D/E)XK nucleases and identification of novel superfamily members using structure-guided alignment of sequence profiles. *J. Mol. Microbiol. Biotechnol. 3*, 69–72.
25. Bujnicki, J.M., Radlinska, M., and Rychlewski, L. (2000). Atomic model of the 5-methylcytosine-specific restriction enzyme McrA reveals an atypical zinc finger and structural similarity to betabetaalphaMe endonucleases. *Mol. Microbiol. 37*, 1280–1281.
26. Burman, R.W., Yates, P.A., Green, L.D., Jacky, P.B., Turker, M.S., and Popovich, B.W. (1999). Hypomethylation of an expanded FMR1 allele is not associated with a global DNA methylation defect. *Am. J. Hum. Genet. 65*, 1375–1386.
27. Butkus, V., Klimasauskas, S., Petrauskiene, L., Maneliene, Z., Janulaitis, A., Minchenkova, L.E., and Schyolkina, A.K. (1987). Synthesis and physical characterization of DNA fragments containing N4-methylcytosine and 5-methylcytosine. *Nucleic Acids Res. 15*, 8467–8478.
28. Card, C.O., Wilson, G.G., Weule, K., Hasapes, J., Kiss, A., and Roberts, R.J. (1990). Cloning and characterization of the HpaII methylase gene. *Nucleic Acids Res. 18*, 1377–1383.
29. Carey, J. (1988). Gel retardation at low pH resolves trp repressor-DNA complexes for quantitative study. *Proc. Natl. Acad. Sci. U. S. A. 85*, 975–979.
30. Chevalier, B.S. and Stoddard, B.L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res. 29*, 3757–3774.
31. Chotai, K.A. and Payne, S.J. (1998). A rapid, PCR based test for differential molecular diagnosis of Prader-Willi and Angelman syndromes. *J. Med. Genet.*

35, 472–475.
32. Clark, S.J., Statham, A., Stirzaker, C., Molloy, P.L., and Frommer, M. (2006). DNA methylation: Bisulphite modification and analysis. *Nat. Protoc. 1*, 2353–2364.
33. Clouaire, T., de las Heras, J.I., Merusi, C., and Stancheva, I. (2010). Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic Acids Res. 38*, 4620–4634.
34. Cohen-Karni, D., Xu, D., Apone, L., Fomenkov, A., Sun, Z., Davis, P.J., Kinney, S.R.M., Yamada-Mabuchi, M., Xu, S., Davis, T., et al. (2011). The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proc. Natl. Acad. Sci. U. S. A. 108*, 11040–11045.
35. Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr. 50*, 760–763.
36. Cramer, J.M., Scarsdale, J.N., Walavalkar, N.M., Buchwald, W.A., Ginder, G.D., and Williams, D.C. (2014). Probing the Dynamic Distribution of Bound States for Methylcytosine-binding Domains on DNA. *J. Biol. Chem. 289*, 1294–1302.
37. Cross, S.H., Meehan, R.R., Nan, X., and Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat. Genet. 16*, 256–259.
38. Cymerman, I.A., Obarska, A., Skowronek, K.J., Lubys, A., and Bujnicki, J.M. (2006). Identification of a new subfamily of HNH nucleases and experimental characterization of a representative member, HphI restriction endonuclease. *Proteins 65*, 867–876.
39. Czapinska, H., Kowalska, M., Zagorskaitė, E., Manakova, E., Slyvka, A., Xu, S., Siksnys, V., Sasnauskas, G., and Bochtler, M. (2018). Activity and structure of EcoKMcrA. *Nucleic Acids Res.*
40. Daujotyte, D., Liutkeviciute, Z., Tamulaitis, G., and Klimasauskas, S. (2008). Chemical mapping of cytosines enzymatically flipped out of the DNA helix. *Nucleic Acids Res. 36*, e57–e57.
41. Defossez, P.-A. and Stancheva, I. (2011). Biological functions of methyl-CpG-binding proteins. *Prog. Mol. Biol. Transl. Sci. 101*, 377–398.
42. Diaz-Ochoa, V.E., Jellbauer, S., Klaus, S., and Raffatellu, M. (2014). Transition metal ions at the crossroads of mucosal immunity and microbial pathogenesis. *Front. Cell. Infect. Microbiol. 4*, 2.
43. Dila, D., Sutherland, E., Moran, L., Slatko, B., and Raleigh, E.A. (1990). Genetic and sequence organization of the mcrBC locus of Escherichia coli K-12. *J. Bacteriol. 172*, 4888–4900.
44. Du, J., Johnson, L.M., Groth, M., Feng, S., Hale, C.J., Li, S., Vashisht, A.A., Gallego-Bartolome, J., Wohlschlegel, J.A., Patel, D.J., et al. (2014). Mechanism of DNA Methylation-Directed Histone Methylation by KRYPTONITE. *Mol. Cell 55*, 495–504.
45. Ebbs, M.L. and Bender, J. (2006). Locus-Specific Control of DNA Methylation by the Arabidopsis SUVH5 Histone Methyltransferase. *PLANT CELL ONLINE 18*, 1166–1176.
46. Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res. 10*, 2709–2721.

47. Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr. 60*, 2126–2132.
48. Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M.C., Jabado, O.J., et al. (2012). Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. *Nat. Biotechnol. 30*, 1232–1239.
49. Feng, Q. and Zhang, Y. (2001). The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev. 15*, 827–832.
50. Feng, J., Chang, H., Li, E., and Fan, G. (2005). Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. *J. Neurosci. Res. 79*, 734–746.
51. Fleischman, R.A., Cambell, J.L., and Richardson, C.C. (1976). Modification and restriction of T-even bacteriophages. In vitro degradation of deoxyribonucleic acid containing 5-hydroxymethylctosine. *J. Biol. Chem. 251*, 1561–1570.
52. Franke, D. and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr. 42*, 342–346.
53. Fritzsche, E., Hayatsu, H., Igloi, G.L., Iida, S., and Kössel, H. (1987). The use of permanganate as a sequencing reagent for identification of 5-methylcytosine residues in DNA. *Nucleic Acids Res. 15*, 5517–5528.
54. Fu, Y., Luo, G.-Z., Chen, K., Deng, X., Yu, M., Han, D., Hao, Z., Liu, J., Lu, X., Doré, L.C., et al. (2015). N6-Methyldeoxyadenosine Marks Active Transcription Start Sites in Chlamydomonas. *Cell 161*, 879–892.
55. Fujimori, A., Matsuda, Y., Takemoto, Y., Hashimoto, Y., Kubo, E., Araki, R., Fukumura, R., Mita, K., Tatsumi, K., and Muto, M. (1998). Cloning and mapping of Np95 gene which encodes a novel nuclear protein associated with cell proliferation. *Mamm. Genome 9*, 1032–1035.
56. Fujita, N., Takebayashi, S., Okumura, K., Kudo, S., Chiba, T., Saya, H., and Nakao, M. (1999). Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. *Mol. Cell. Biol. 19*, 6415–6426.
57. Gast, F.U., Brinkmann, T., Pieper, U., Krüger, T., Noyer-Weidner, M., and Pingoud, A. (1997). The recognition of methylated DNA by the GTP-dependent restriction endonuclease McrBC resides in the N-terminal domain of McrB. *Biol. Chem. 378*, 975–982.
58. Gogos, J.A., Karayiorgou, M., Aburatani, H., and Kafatos, F.C. (1990). Detection of single base mismatches of thymine and cytosine residues by potassium permanganate and hydroxylamine in the presence of tetralkylammonium salts. *Nucleic Acids Res. 18*, 6807–6814.
59. Gorbalenya, A.E. (1994). Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. *Protein Sci. 3*, 1117–1120.
60. Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corrales, D., Hsu, C.-H., Aravind, L., He, C., and Shi, Y. (2015). DNA Methylation on N6-Adenine in C. elegans. *Cell 161*, 868–878.
61. Han, T., Yamada-Mabuchi, M., Zhao, G., Li, L., liu, G., Ou, H.-Y., Deng, Z., Zheng, Y., and He, X. (2015). Recognition and cleavage of 5-methylcytosine DNA by bacterial SRA-HNH proteins. *Nucleic Acids Res. 43*, 1147–1159.

62. Hashimoto, H., Horton, J.R., Zhang, X., Bostick, M., Jacobsen, S.E., and Cheng, X. (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature 455*, 826–829.
63. Hashimoto, H., Horton, J.R., Zhang, X., and Cheng, X. (2009). UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics 4*, 8–14.
64. Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012a). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res. 40*, 4841–4849.
65. Hashimoto, H., Zhang, X., and Cheng, X. (2012b). Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res. 40*, 8276–8284.
66. Hayatsu, H. and Ukita, T. (1967). The selective degradation of pyrimidines in nucleic acids by permanganate oxidation. *Biochem. Biophys. Res. Commun. 29*, 556–561.
67. Hendrich, B. and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol. 18*, 6538–6547.
68. Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature 401*, 301–304.
69. Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev. 15*, 710–723.
70. Hiom, K. and Sedgwick, S.G. (1991). Cloning and structural characterization of the mcrA locus of Escherichia coli. *J. Bacteriol. 173*, 7368–7373.
71. Ho, K.L., McNae, I.W., Schmiedeberg, L., Klose, R.J., Bird, A.P., and Walkinshaw, M.D. (2008). MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell 29*, 525–531.
72. Hong, S. and Cheng, X. (2016). DNA Base Flipping: A General Mechanism for Writing, Reading, and Erasing DNA Modifications. *Adv. Exp. Med. Biol. 945*, 321–341.
73. Hopfner, R., Mousli, M., Jeltsch, J.M., Voulgaris, A., Lutz, Y., Marin, C., Bellocq, J.P., Oudet, P., and Bronner, C. (2000). ICBP90, a novel human CCAAT binding protein, involved in the regulation of topoisomerase IIalpha expression. *Cancer Res. 60*, 121–128.
74. Horton, J.R., Mabuchi, M.Y., Cohen-Karni, D., Zhang, X., Griggs, R.M., Samaranayake, M., Roberts, R.J., Zheng, Y., and Cheng, X. (2012). Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease. *Nucleic Acids Res. 40*, 9763–9773.
75. Horton, J.R., Nugent, R.L., Li, A., Mabuchi, M.Y., Fomenkov, A., Cohen-Karni, D., Griggs, R.M., Zhang, X., Wilson, G.G., Zheng, Y., et al. (2014a). Structure and mutagenesis of the DNA modification-dependent restriction endonuclease AspBHI. *Sci. Rep. 4*, 4246.
76. Horton, J.R., Wang, H., Mabuchi, M.Y., Zhang, X., Roberts, R.J., Zheng, Y., Wilson, G.G., and Cheng, X. (2014b). Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. *Nucleic Acids Res. 42*, 12092–12101.
77. Horton, J.R., Borgaro, J.G., Griggs, R.M., Quimby, A., Guan, S., Zhang, X.,

Wilson, G.G., Zheng, Y., Zhu, Z., and Cheng, X. (2014c). Structure of 5-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA. *Nucleic Acids Res. 42*, 7947–7959.
78. Hu, L., Li, Z., Wang, P., Lin, Y., and Xu, Y. (2011). Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. *Cell Res. 21*, 1374–1378.
79. Irier, H.A. and Jin, P. (2012). Dynamics of DNA methylation in aging and Alzheimer's disease. *DNA Cell Biol. 31 Suppl 1*, S42-8.
80. Iurlaro, M., Ficz, G., Oxley, D., Raiber, E.-A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S., and Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol. 14*, R119.
81. Jackson, J.P., Johnson, L., Jasencakova, Z., Zhang, X., PerezBurgos, L., Singh, P.B., Cheng, X., Schubert, I., Jenuwein, T., and Jacobsen, S.E. (2004). Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. *Chromosoma 112*, 308–315.
82. Janosi, L., Yonemitsu, H., Hong, H., and Kaji, A. (1994). Molecular Cloning and Expression of a Novel Hydroxymethylcytosine-specific Restriction Enzyme (PvuRts1I) Modulated by Glucosylation of DNA. *J. Mol. Biol. 242*, 45–61.
83. Ji, S., Fu, I., Naldiga, S., Shao, H., Basu, A.K., Broyde, S., and Tretyakova, N.Y. (2018). 5-Formylcytosine mediated DNA–protein cross-links block DNA replication and induce mutations in human cells. *Nucleic Acids Res. 46*, 6455–6469.
84. Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet. 19*, 187–191.
85. Jørgensen, H.F., Ben-Porath, I., and Bird, A.P. (2004). Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Mol. Cell. Biol. 24*, 3387–3395.
86. Kabsch, W. (2010). XDS. *Acta Crystallogr. D. Biol. Crystallogr. 66*, 125–132.
87. Kazrani, A.A., Kowalska, M., Czapinska, H., and Bochtler, M. (2014). Crystal structure of the 5hmC specific endonuclease PvuRts1I. *Nucleic Acids Res. 42*, 5929–5936.
88. Kisiala, M., Copelas, A., Czapinska, H., Xu, S., and Bochtler, M. (2018). Crystal structure of the modification-dependent SRA-HNH endonuclease TagI. *Nucleic Acids Res.*
89. Klose, R.J., Sarraf, S.A., Schmiedeberg, L., McDermott, S.M., Stancheva, I., and Bird, A.P. (2005). DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell 19*, 667–678.
90. Kobayashi, I. (2001). Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res. 29*, 3742–3756.
91. Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. *Mol. Cell. Biol. 25*, 4388–4396.
92. Kornberg, S.R., Zimmerman, S.B., and Kornberg, A. (1961). Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected Escherichia coli. *J. Biol. Chem. 236*, 1487–1493.
93. Kriaucionis, S. and Heintz, N. (2009). The Nuclear DNA Base 5-

Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science (80-.). 324*, 929–930.
94. Kriukiene, E., Liutkeviciute, Z., and Klimasauskas, S. (2012). 5-Hydroxymethylcytosine - the elusive epigenetic mark in mammalian DNA. *Chem. Soc. Rev. 41*, 6916–6930.
95. Krüger, T., Grund, C., Wild, C., and Noyer-Weidner, M. (1992). Characterization of the mcrBC region of Escherichia coli K-12 wild-type and mutant strains. *Gene 114*, 1–12.
96. Krüger, T., Wild, C., and Noyer-Weidner, M. (1995). McrB: a prokaryotic protein specifically recognizing DNA containing modified cytosine residues. *EMBO J. 14*, 2661–2669.
97. Kuśmierek, J.T. and Singer, B. (1982). Chloroacetaldehyde-treated ribo- and deoxyribopolynucleotides. 1. Reaction products. *Biochemistry 21*, 5717–5722.
98. Kuznetsov, N.A., Vorobjev, Y.N., Krasnoperov, L.N., and Fedorova, O.S. (2012). Thermodynamics of the multi-stage DNA lesion recognition and repair by formamidopyrimidine-DNA glycosylase using pyrrolocytosine fluorescence--stopped-flow pre-steady-state kinetics. *Nucleic Acids Res. 40*, 7384–7392.
99. de la Campa, A.G., Springhorn, S.S., Kale, P., and Lacks, S.A. (1988). Proteins encoded by the DpnI restriction gene cassette. Hyperproduction and characterization of the DpnI endonuclease. *J. Biol. Chem. 263*, 14696–14702.
100. Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol. 8*, 317–327.
101. Lacks, S. and Greenberg, B. (1975). A deoxyribonuclease of Diplococcus pneumoniae specific for methylated DNA. *J. Biol. Chem. 250*, 4060–4066.
102. Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., et al. (2010). Dynamic changes in the human methylome during differentiation. *Genome Res. 20*, 320–331.
103. Law, J.A. and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet. 11*, 204–220.
104. Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell 69*, 905–914.
105. Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature 430*, 471–476.
106. Lippman, Z., Gendrel, A.-V., Colot, V., and Martienssen, R. (2005). Profiling DNA methylation patterns using genomic tiling microarrays. *Nat. Methods 2*, 219–224.
107. Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature 462*, 315–322.
108. Liu, C. and Martin, C.T. (2001). Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase. *J. Mol. Biol. 308*, 465–475.
109. Løbner-Olesen, A., Skovgaard, O., and Marinus, M.G. (2005). Dam methylation: coordinating cellular processes. *Curr. Opin. Microbiol. 8*, 154–160.
110. Loenen, W.A.M. and Raleigh, E.A. (2014). The other face of restriction:

modification-dependent enzymes. *Nucleic Acids Res. 42*, 56–69.
111. Lu, L., Patel, H., and Bissler, J.J. (2002). Optimizing DpnI digestion conditions to detect replicated DNA. *Biotechniques 33*, 316–318.
112. Lyko, F. (2017). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet. 19*, 81–92.
113. Lyko, F., Ramsahoye, B.H., Kashevsky, H., Tudor, M., Mastrangelo, M.A., Orr-Weaver, T.L., and Jaenisch, R. (1999). Mammalian (cytosine-5) methyltransferases cause genomic DNA methylation and lethality in Drosophila. *Nat. Genet. 23*, 363–366.
114. Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell 58*, 499–507.
115. Mierzejewska, K., Siwek, W., Czapinska, H., Kaus-Drobek, M., Radlinska, M., Skowronek, K., Bujnicki, J.M., Dadlez, M., and Bochtler, M. (2014). Structural basis of the methylation specificity of R.DpnI. *Nucleic Acids Res. 42*, 8745–8754.
116. Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R., and Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science 297*, 403–405.
117. Mousli, M., Hopfner, R., Abbady, A.-Q., Monté, D., Jeanblanc, M., Oudet, P., Louis, B., and Bronner, C. (2003). ICBP90 belongs to a new family of proteins with an expression that is deregulated in cancer cells. *Br. J. Cancer 89*, 120–127.
118. Mulligan, E.A. and Dunn, J.J. (2008). Cloning, purification and initial characterization of E. coli McrA, a putative 5-methylcytosine-specific nuclease. *Protein Expr. Purif. 62*, 98–103.
119. Mulligan, E.A., Hatchwell, E., McCorkle, S.R., and Dunn, J.J. (2010). Differential binding of Escherichia coli McrA protein to DNA sequences that contain the dinucleotide m5CpG. *Nucleic Acids Res. 38*, 1997–2005.
120. Nady, N., Lemak, A., Walker, J.R., Avvakumov, G. V, Kareta, M.S., Achour, M., Xue, S., Duan, S., Allali-Hassani, A., Zuo, X., et al. (2011). Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *J. Biol. Chem. 286*, 24300–24311.
121. Nan, X., Meehan, R.R., and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res. 21*, 4886–4892.
122. Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature 393*, 386–389.
123. Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E. V (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res. 9*, 27–43.
124. Oakeley, E.J., Schmitt, F., and Jost, J.P. (1999). Quantification of 5-methylcytosine in DNA by the chloroacetaldehyde reaction. *Biotechniques 27*, 744–746, 748–750, 752.
125. Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell 105*, 487–497.

126. Otani, J., Arita, K., Kato, T., Kinoshita, M., Kimura, H., Suetake, I., Tajima, S., Ariyoshi, M., and Shirakawa, M. (2013). Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. *J. Biol. Chem. 288*, 6351–6362.
127. Panne, D., Raleigh, E.A., and Bickle, T.A. (1998). McrBs, a modulator peptide for McrBC activity. *EMBO J. 17*, 5477–5483.
128. Panne, D., Raleigh, E.A., and Bickle, T.A. (1999). The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. *J. Mol. Biol. 290*, 49–60.
129. Panne, D., Müller, S.A., Wirtz, S., Engel, A., and Bickle, T.A. (2001). The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *EMBO J. 20*, 3210–3217.
130. Parikh, S.S., Putnam, C.D., and Tainer, J.A. (2000). Lessons learned from structural results on uracil-DNA glycosylase. *Mutat. Res. 460*, 183–199.
131. Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K., and Yura, R. (1972). The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J. 126*, 781–790.
132. Pichler, G., Wolf, P., Schmidt, C.S., Meilinger, D., Schneider, K., Frauer, C., Fellinger, K., Rottach, A., and Leonhardt, H. (2011). Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J. Cell. Biochem. 112*, 2585–2593.
133. Piekarowicz, A., Yuan, R., and Stein, D.C. (1991). A new method for the rapid identification of genes encoding restriction and modification enzymes. *Nucleic Acids Res. 19*, 1831–1835.
134. Pieper, U. and Pingoud, A. (2002). A Mutational Analysis of the PD...D/EXK Motif Suggests That McrC Harbors the Catalytic Center for DNA Cleavage by the GTP-Dependent Restriction Enzyme McrBC from *Escherichia coli* †. *Biochemistry 41*, 5236–5244.
135. Pieper, U., Brinkmann, T., Krüger, T., Noyer-Weidner, M., and Pingoud, A. (1997). Characterization of the interaction between the restriction endonuclease McrBC from E. coli and its cofactor GTP. *J. Mol. Biol. 272*, 190–199.
136. Pieper, U., Groll, D.H., Wünsch, S., Gast, F.-U., Speck, C., Mücke, N., and Pingoud, A. (2002). The GTP-Dependent Restriction Enzyme McrBC from *Escherichia coli* Forms High-Molecular Mass Complexes with DNA and Produces a Cleavage Pattern with a Characteristic 10-Base Pair Repeat †. *Biochemistry 41*, 5245–5254.
137. Pingoud, A. and Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Res. 29*, 3705–3727.
138. Pingoud, A., Wilson, G.G., and Wende, W. (2014). Type II restriction endonucleases--a historical perspective and more. *Nucleic Acids Res. 42*, 7489–7527.
139. Rajakumara, E., Law, J.A., Simanshu, D.K., Voigt, P., Johnson, L.M., Reinberg, D., Patel, D.J., and Jacobsen, S.E. (2011a). A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. *Genes Dev. 25*, 137–152.
140. Rajakumara, E., Wang, Z., Ma, H., Hu, L., Chen, H., Lin, Y., Guo, R., Wu, F., Li, H., Lan, F., et al. (2011b). PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Mol. Cell 43*, 275–284.

141. Rajakumara, E., Nakarakanti, N.K., Nivya, M.A., and Satish, M. (2016). Mechanistic insights into the recognition of 5-methylcytosine oxidation derivatives by the SUVH5 SRA domain. *Sci. Rep. 6*, 20161.
142. Raleigh, E.A. and Wilson, G. (1986). Escherichia coli K-12 restricts DNA containing 5-methylcytosine. *Proc. Natl. Acad. Sci. U. S. A. 83*, 9070–9074.
143. Raleigh, E.A., Trimarchi, R., and Revel, H. (1989). Genetic and physical mapping of the mcrA (rglA) and mcrB (rglB) loci of Escherichia coli K-12. *122*, 279–296.
144. Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U. S. A. 97*, 5237–5242.
145. Reddy, Y. V and Rao, D.N. (2000). Binding of EcoP15I DNA methyltransferase to DNA reveals a large structural distortion within the recognition sequence. *J. Mol. Biol. 298*, 597–610.
146. Revel, H.R. (1967). Restriction of nonglucosylated T-even bacteriophage: properties of permissive mutants of Escherichia coli B and K12. *Virology 31*, 688–701.
147. Riccio, A., Aaltonen, L.A., Godwin, A.K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Bassi, D.E., et al. (1999). The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. *Nat. Genet. 23*, 266–268.
148. Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.K., Dryden, D.T.F., Dybvig, K., et al. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res. 31*, 1805–1812.
149. Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2015). REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res. 43*, D298–D299.
150. Ross, T.K., Achberger, E.C., and Braymer, H.D. (1989). Nucleotide sequence of the McrB region of Escherichia coli K-12 and evidence for two independent translational initiation sites at the mcrB locus. *J. Bacteriol. 171*, 1974–1981.
151. Saito, M. and Ishikawa, F. (2002). The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J. Biol. Chem. 277*, 35434–35439.
152. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A laboratory manual (2nd ed.).
153. Sasnauskas, G., Zagorskaitė, E., Kauneckaitė, K., Tamulaitiene, G., and Siksnys, V. (2015). Structure-guided sequence specificity engineering of the modification-dependent restriction endonuclease LpnPI. *Nucleic Acids Res. 43*, 6144–6155.
154. Sasnauskas, G., Kauneckaite, K., and Siksnys, V. (2018). Structural basis of DNA target recognition by the B3 domain of Arabidopsis epigenome reader VAL1. *Nucleic Acids Res. 46*, 4316–4324.
155. Scarsdale, J.N., Webb, H.D., Ginder, G.D., and Williams, D.C. (2011). Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res. 39*, 6741–6752.
156. Serva, S., Weinhold, E., Roberts, R.J., and Klimasauskas, S. (1998). Chemical display of thymine residues flipped out by DNA methyltransferases. *Nucleic*

Acids Res. 26, 3473–3479.
157. Shao, C., Wang, C., and Zang, J. (2014). Structural basis for the substrate selectivity of PvuRts1I, a 5-hydroxymethylcytosine DNA restriction endonuclease. *Acta Crystallogr. D. Biol. Crystallogr. 70*, 2477–2486.
158. Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature 450*, 908–912.
159. Shatsky, M., Nussinov, R., and Wolfson, H.J. (2004). A method for simultaneous alignment of multiple protein structures. *Proteins 56*, 143–156.
160. Shimbo, T. and Wade, P.A. (2016). Proteins That Read DNA Methylation. In Advances in Experimental Medicine and Biology, pp. 303–320.
161. Siwek, W., Czapinska, H., Bochtler, M., Bujnicki, J.M., and Skowronek, K. (2012). Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI. *Nucleic Acids Res. 40*, 7563–7572.
162. Smith, Z.D. and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet. 14*, 204–220.
163. Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013). Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell 152*, 1146–1159.
164. Stewart, F.J. and Raleigh, E.A. (1998). Dependence of McrBC cleavage on distance between recognition elements. *Biol. Chem. 379*, 611–616.
165. Stewart, F.J., Panne, D., Bickle, T.A., and Raleigh, E.A. (2000). Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol. Biol. 298*, 611–622.
166. Stoddard, B.L. (2006). Homing endonuclease structure and function. *Q. Rev. Biophys. 38*, 49.
167. Sukackaite, R., Grazulis, S., Tamulaitis, G., and Siksnys, V. (2012). The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. *Nucleic Acids Res. 40*, 7552–7562.
168. Sun, Z., Terragni, J., Borgaro, J.G., Liu, Y., Yu, L., Guan, S., Wang, H., Sun, D., Cheng, X., Zhu, Z., et al. (2013). High-Resolution Enzymatic Mapping of Genomic 5-Hydroxymethylcytosine in Mouse Embryonic Stem Cells. *Cell Rep. 3*, 567–576.
169. Sutherland, E., Coe, L., and Raleigh, E.A. (1992). McrBC: a multisubunit GTP-dependent restriction endonuclease. *J. Mol. Biol. 225*, 327–348.
170. Szulik, M.W., Pallan, P.S., Nocek, B., Voehler, M., Banerjee, S., Brooks, S., Joachimiak, A., Egli, M., Eichman, B.F., and Stone, M.P. (2015). Differential Stabilities and Sequence-Dependent Base Pair Opening Dynamics of Watson–Crick Base Pairs with 5-Hydroxymethylcytosine, 5-Formylcytosine, or 5-Carboxylcytosine. *Biochemistry 54*, 1294–1305.
171. Szwagierczak, A., Brachmann, A., Schmidt, C.S., Bultmann, S., Leonhardt, H., and Spada, F. (2011). Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5–hydroxymethylcytosine. *Nucleic Acids Res. 39*, 5149–5156.
172. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL

partner TET1. *Science 324*, 930–935.
173. Tamulaitis, G., Solonin, A.S., and Siksnys, V. (2002). Alternative arrangements of catalytic residues at the active sites of restriction enzymes. *FEBS Lett. 518*, 17–22.
174. Vagin, A. and Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr. D. Biol. Crystallogr. 66*, 22–25.
175. Walavalkar, N.M., Cramer, J.M., Buchwald, W.A., Scarsdale, J.N., and Williams, D.C. (2014). Solution structure and intramolecular exchange of methyl-cytosine binding domain protein 4 (MBD4) on DNA suggests a mechanism to scan for mCpG/TpG mismatches. *Nucleic Acids Res. 42*, 11218–11232.
176. Wang, H., Guan, S., Quimby, A., Cohen-Karni, D., Pradhan, S., Wilson, G., Roberts, R.J., Zhu, Z., and Zheng, Y. (2011). Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. *Nucleic Acids Res. 39*, 9294–9305.
177. Warren, R.A. (1980). Modified bases in bacteriophage DNAs. *Annu. Rev. Microbiol. 34*, 137–158.
178. White, S.R. and Lauring, B. (2007). AAA+ ATPases: achieving diversity of function with conserved machinery. *Traffic 8*, 1657–1667.
179. Wion, D. and Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol. 4*, 183–192.
180. Wu, X. and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet. 18*, 517–534.
181. Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell 148*, 816–831.
182. Ye, P., Luan, Y., Chen, K., Liu, Y., Xiao, C., and Xie, Z. (2017). MethSMRT: an integrative database for DNA N6-methyladenine and N4-methylcytosine generated by single-molecular real-time sequencing. *Nucleic Acids Res. 45*, D85–D89.
183. Yildirim, O., Li, R., Hung, J.-H., Chen, P.B., Dong, X., Ee, L.-S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. *Cell 147*, 1498–1510.
184. Yoshioka, K. (2002). KyPlot as a Tool for Graphical Data Analysis. In Compstat, (Heidelberg: Physica-Verlag HD), pp. 37–46.
185. Zagorskaitė, E. and Sasnauskas, G. (2014). Chemical Display of Pyrimidine Bases Flipped Out by Modification-Dependent Restriction Endonucleases of MspJI and PvuRts1I Families. *PLoS One 9*, e114580.
186. Zagorskaitė, E., Manakova, E., and Sasnauskas, G. (2018). Recognition of modified cytosine variants by the DNA-binding domain of methyl-directed endonuclease McrBC. *FEBS Lett. 592*, 3335–3345.
187. Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., et al. (2015). N6-methyladenine DNA modification in Drosophila. *Cell 161*, 893–906.
188. Zhang, J., Gao, Q., Li, P., Liu, X., Jia, Y., Wu, W., Li, J., Dong, S., Koseki, H., and Wong, J. (2011). S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. *Cell Res. 21*, 1723–1739.
189. Zheng, L., Baumann, U., and Reymond, J.-L. (2004). An efficient one-step site-

directed and site-saturation mutagenesis protocol. *Nucleic Acids Res. 32*, e115.
190. Zheng, Y., Cohen-Karni, D., Xu, D., Chin, H.G., Wilson, G., Pradhan, S., and Roberts, R.J. (2010). A unique family of Mrr-like modification-dependent restriction endonucleases. *Nucleic Acids Res. 38*, 5527–5534.
191. Zhou, T., Xiong, J., Wang, M., Yang, N., Wong, J., Zhu, B., and Xu, R.-M. (2014). Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. *Mol. Cell 54*, 879–886.

VILNIAUS UNIVERSITETAS

Evelina
ZAGORSKAITĖ

METILINTĄ DNR ATPAŽĮSTANČIŲ RESTRIKCIJOS ENDONUKLEAZIŲ SĄVEIKOS SU MODIFIKUOTU CITOZINU MECHANIZMAS

DAKTARO DISERTACIJOS SANTRAUKA

Fiziniai mokslai,
Biochemija (04 P)

VILNIUS 2018

SUMMARY

TURINYS

SUTRUMPINIMŲ SĄRAŠAS

ar.	aminorūgštis
Ap	ampicilinas
bp	bazių pora
CAA	chloroacetaldehidas
Cm	chloramfenikolis
dg	dvigrandinė
DTT	1,4-ditiotreitolis
EJP	elektroforetinio judrumo poslinkis
H-ryšiai	vandeniliniai ryšiai
IPTG	izopropil-β-D-tiogalaktopiranozidas
JSA	jaučio serumo albuminas
5mC	5-metilcitozinas
4mC	N4-metilcitozinas
5hmC	5-hidroksimetilcitozinas
5fC	5-formilcitozinas
5caC	5-karboksicitozinas
5ghmC	5-glukozilhidroksimetilcitozinas
pyC	pirolocitozinas
M / (m)C	5mC arba 5hmC
MES	2-(*N*-morfolino)etanesulfoninė rūgštis
nt	nukleotidas
PAGE	poliakrilamidinio gelio elektroforezė
REazė	Restrikcijos endonukleazė
RM	Restrikcijos-modifikacijos (sistemos)
vg	viengrandinė
vdW	van der Waals (kontaktas)
wt/WT	laukinis tipas

ĮVADAS

5-metilcitozinas (5mC) ir jo oksiduotos formos 5-hidroksimetilcitozinas (5hmC), 5-formilcitozinas (5fC) ir 5-karboksicitozinas (5caC) veikia kaip epigenetiniai žymenys, kurie yra atsakingi už genų raiškos reguliaciją. Įvairių tipų ląstelėse citozino modifikacijų lygis kinta vystymosi, diferenciacijos, senėjimo ir ligų metu (Irier ir Jin, 2012; Smith ir Meissner, 2013). Prokariotuose citozino metilinimas daugiausiai yra susijęs su restrikcijos-modifikacijos (RM) sistemų veikla (Pingoud et al., 2014). Tipinės restrikcijos endonukleazės (REazės) atpažįsta trumpas specifines DNR sekas ir sukarpo į ląstelę patenkančią svetimą DNR, tuo tarpu metiltransferazės metilindamos tas pačias sekas nuo sukarpymo apsaugo šeimininko genominę DNR. Vienas iš mechanizmų, leidžiančių bakteriofagams apeiti RM sistemas, yra savos DNR modifikavimas, įvedant 5mC, 5hmC ar gliukozilintą 5hmC (5ghmC). Kaip atsakas į virusų DNR modifikacijas bakterijose išsivystė modifikuotą DNR karpančios REazės (Labrie et al., 2010)

Tiek eukariotų epigenetinei reguliacijai, tiek modifikuotą DNR karpančių REazių veikimui ypač svarbus baltymų gebėjimas tiksliai atskirti modifikuotą citoziną nuo nemodifikuoto. Remiantis struktūriniais eukariotinių 5mC ir 5hmC surišimo domenų tyrimais galima išskirti dvi modifikuotos bazės atpažinimo strategijas. Pirma, baltymai, kurie turi metil-CpG surišimo domeną (MBD1, MBD2, MBD4 ir MeCP2) bei cinko-pirštų baltymas Kaiso atpažįsta modifikuotą citoziną Watson-Crick bazių poros kontekste (Buck-Koehntop et al., 2012; Ho et al., 2008; Ohki et al., 2001; Otani et al., 2013; Scarsdale et al., 2011). Antra, UHRF1, UHRF2, KRYPTONITE ir SUVH5 baltymuose aptinkami SRA domenai išsuka modifikuotą bazę iš dgDNR ir patalpina ją baltymo kišenėje (Arita et al., 2008; Avvakumov et al., 2008; Du et al., 2014; Hashimoto et al., 2008; Rajakumara et al., 2011a; Zhou et al., 2014). Struktūriškai panašūs, ir, tikėtina, bazės išsukimo mechanizmą naudojantys DNR atpažinimo domenai taip pat randami MspJI ir PvuRts1I šeimų REazėse, atpažįstančiose modifikuotus citozino variantus įvairiuose sekos kontekstuose (Cohen-Karni et al., 2011; Szwagierczak et al., 2011; Wang et al., 2011). Bazės išsukimas taip pat yra būdingas struktūriškai negimininga modifikuotą DNR karpančiam fermentui McrBC, atpažįstančiam 5mC, 5hmC ir N4-metilcitoziną (4mC) turinčias sekas (Sutherland et al., 1992). Už bazės išsukimą šiuo atveju yra atsakingas McrB subvieneto N-gale esantis DNR atpažinimo domenas McrB-N (Sukackaite et al., 2012). Tačiau lieka daug

modifikuotus citozinus atpažįstančių baltymų, kurių veikimo mechanizmas yra nežinomas. Ypač daug necharakterizuotų modifikuotų citozino „skaitytojų“ yra eukariotuose (Iurlaro et al., 2013; Spruijt et al., 2013). Taip pat trūksta žinių ir apie kai kurių bakterinių modifikuotą DNR atpažįstančių nukleazių, pavyzdžiui, klasikinio *E. coli* fermento McrA (Revel, 1967) veikimo mechanizmą, nes šio baltymo aktyvumas *in vitro* nebuvo aptiktas.

Atsižvelgiant į citozinų metilinimo ir metilintus citozinus atpažįstančių baltymų svarbą daugelyje žinduolių ląstelėse vykstančių procesų, ir galimą bakterinių modifikuotus citozinus atpažįstančių endonukleazių panaudojimą epigenomo tyrimams, šį darbą skyrėme modifikuoto citozino atpažinimo mechanizmo tyrimams įvairiose metilintą DNR atpažįstančiose REazėse.

Pagrindiniai tikslai ir uždaviniai:

1. Ištirti MspJI ir PvuRts1I šeimoms priklausančių REazių LpnPI, YkrI ir BmeDI modifikuotos bazės išsukimą iš dgDNR tirpale panaudojant fluorescentinius ir cheminius tyrimo metodus bei įvertinti šių metodų tinkamumą modifikuotos bazės išsukimo tyrimams.
2. Ištirti LpnPI modifikuoto citozino ir jį supančio sekos konteksto atpažinimo mechanizmą.
3. Ištirti laukinio tipo ir mutantinių McrB-N variantų gebėjimą sąveikauti su 5mC, 4mC, 5hmC, 5fC ir 5caC modifikacijas turinčiais DNR taikiniais.
4. Ištirti McrA N-galinio ir C-galinio domenų funkcijas bei nustatyti modifikuoto citozino atpažinimo mechanizmą.

Mokslinis naujumas

Pirmą kartą parodyta, kad MspJI ir PvuRt1I šeimoms priklausančios REazės LpnPI, YkrI ir BmeDI išsuka modifikuotą bazę tirpale. Taip pat pademonstruota, kad įvairių cheminių ir spektroskopinių bazės išsukimo tyrimo metodų panaudojimo sėkmė priklauso nuo tiriamo baltymo kišenės savybių. Pavyzdžiui, bazės išsukimas SRA domeną turintinčiuose MspJI ir PvuRts1I baltymuose buvo nustatytas metodu, paremtu išsuktų pirimidinų oksidacija $KMnO_4$, tuo tarpu kiti metodai, kaip pirolocitozino fluorescencija (bazės išsukimas patvirtintas McrBC atveju) ir cheminė modifikacija choroacetaldehidu (bazės išsukimas patvirtintas metiltransferazėse ir REazėje Ecl18kI) nerodė teigiamo signalo su šiais baltymais. Pirmą kartą pademonstruota, kad LpnPI baltymo paviršinės kilpos veikia kaip atskiri DNR surišimo / atpažinimo moduliai, kuriuos keičiant galima keisti

fermento specifiškumą modifikuotą citoziną supančiai DNR sekai. Taip pat pirmą kartą ištirtas populiaraus epigenomo tyrimo įrankio McrBC gebėjimas atpažinti DNR su skirtingoms citozino modifikacijomis. Galiausiai, pirmą kartą nustatytas fermento McrA nukleazinis aktyvumas *in vitro* ir nustatytos jo domenų funkcijos.

Praktinė reikšmė

Daugelio eukariotinių ir prokariotinių baltymų 5(h)mC atpažinimo mechanizmas vis dar yra nežinomas, tad modifikuotos bazės išsukimo tirpale tyrimai fluorescentiniu ir cheminiais metodais gali būti naudingi tiriant nežinomos struktūros baltymų sąveikos su modifikuota DNR mechanizmą. LpnPI specifiškumo 5mC kontekstui keitimas atveria galimybes kurti baltymus, atpažįstančius modifikaciją kituose sekos kontekstuose ar netgi be konteksto, taip praplečiant šio ir gimininių fermentų taikymo epigenomo tyrimuose galimybes, o fundamentinis McrA charakterizavimas atveria galimybes panaudoti šią REazę praktiniams tikslams, pvz., konstruojant dirbtines endonukleazes ar nikazes.

Ginamieji teiginiai:

1. MspJI ir PvuRts1I šeimoms priklausančios REazės išsuka modifikuotą citoziną iš dvigrandinės DNR.
2. MspJI šeimos baltymai atpažįsta 5-metilcitoziną ir jį supantį sekos kontekstą per keletą paviršinių kilpų, veikiančių kaip atskiri DNR surišimo / atpažinimo moduliai. Keičiant šias kilpas gali būti keičiamas fermento specifiškumas.
3. McrBC DNR atpažinimo domeno McrB-N gimininingumas įvairioms citozino modifikacijomis gali būti keičiamas baltymo kišenės mutacijomis
4. McrA pasižymi silpnu *in vitro* aktyvumu, kuris priklauso nuo Mn^{2+} jonų ir dalinai nuo DNR metilinimo.
5. McrA N-galinis domenas atpažįsta modifikuotą citoziną turinčius DNR taikinius be bazės išsukimo.

1 MEDŽIAGOS IR TYRIMO METODAI

Reagentai, fermentai. Šiame tyrime visi naudoti reagentai buvo aukščiausios kokybės. T4 polinukleotido kinazė; T4 DNR ligazė; FastAP temperatūrai jautri šarminė fosfatazė; DNR polimerazės; JSA ir REazės buvo iš Thermo Fisher Scientific. Produktai naudoti pagal gamintojo rekomendacijas.

***E. coli* kamienai.** ER2267 kamienas naudotas klonavimo procedūroms ir plazmidžių padauginimui. ER2566, BL21(DE3), DH10B kamienai buvo naudoti baltymų raiškai.

Buferiai ir tirpalai. Ardymo buferis: 20 mM Tris-HCl (pH 7,5-8,0), 500 mM NaCl, 7 mM 2-merkaptoetanolis, 2 mM fenilmetilsulfonilfluoridas, 10-25 mM imidazolas; Kolonėlės buferis: 10 mM Tris-acetatas (pH 8,0), 500 mM K-acetatas; Saugojimo buferis I: 20 mM Tris-HCl (pH 7,5–8,5), 200 mM KCl, 1 mM DTT, 50 % v/v glicerolis; Saugojimo buferis II: 20 mM Tris-acetatas (pH 7,5), 250 mM K-acetatas, 1 mM DTT, 50 % v/v glicerolis; Susirišimo buferis I: 40 mM Tris-acetatas (pH 8,3), 0,1 mg/ml JSA, 10 % v/v glicerolio; Susirišimo buferis II: 30 mM Mes-histidinas (pH 6,3), 0,1 mg/ml JSA, 10 % v/v glicerolis; Elektroforezės buferis **I**: 40 mM Tris-acetatas (pH 8,3); Elektroforezės buferis II: 30 mM Mes-histidinas (pH 6,3); Akrilamido tirpalas: akrilamidas/N,N'-metilenbisakrilamidas (29:1 w/w); Mėginio dažas: 95 % (v/v) formamidas, 25 mM EDTA (pH 8,0), 0,01 % bromfenolio mėlynasis (w/v); TBE buferis: 0,1 M Tris-H_3BO_3, 2 mM EDTA (pH 8,3); Reakcijos buferis I: 33 mM Tris-acetatas (pH 8,0), 66 mM K-acetatas, 10 mM Mg-acetatas, 0,1 mg/ml JSA; Reakcijos buferis II: 50 mM Tris–HCl (pH 7,8), 50 mM KCl, 0,1 mg/ml JSA.

Plazmidės ir kita DNR. pTYB2 (Ap^R) (NEB), pLATE31, pLATE11 (Thermo Fisher Scientific) ir pBAD24 (Ap^R) raiškos vektoriai buvo naudojami baltymų klonavimui ir raiškai. pACYC184 (Cm^R) ir pACYC184_M.HpaII (Cm^R) plazmidės buvo naudojamos ribojimo eksperimentuose su McrA. Genominės DNR pirktos iš Leibnitz Instituto DSMZ.

DNR oligonukleotidai. Oligodupleksiniai substratai pateikti 1 lentelėje yra pirkti iš IBA ir Metabion. Oligonukleotidai žymėti su [γ-^{33}P]ATP ar [γ-^{32}P]ATP (Hartmann Analytic, PerkinElmer) ir T4 polinukleotido kinaze. Oligodupleksai gauti sulydant komplementarius žymėtą ir ne oligonukleotidus.

Nedenatūruojanti poliakrilamidinio gelio elektroforezė (PAGE). Nedenatūruojanti PAGE naudota EJP eksperimentuose. Geliai sudaryti iš 8 % Akrilamido tirpalo Elektroforezės buferyje I ar II, polimerizacija inicijuota pridedant TEMED ir amonio persulfato. Gelių dimensijos 22:15:0,1 cm. Vienas iš stiklų apdorojamas 3-metakriloksipropiltrimetoksisilanu, kitas – 5 % v/v dichlordimetilsilanu $CHCl_3$. Radioaktyvios DNR ir baltymų mėginiai sumaišyti Susirišimo buferyje I ar II ir palikti 15 min kambario temperatūroje bei užnešti ant gelio. Elektroforezės trukmė 45-90 min ~5 V/cm. Po elektroforezės stiklas, išteptas dichlordimetilsilanu, buvo nuimtas, gelis išdžiovintas ant likusio stiklo karšta oro srove. Radioaktyvūs DNR fragmentai aptikti naudojant fosforescentinius ekranus (FujiFilm) ir Cyclone fosfovaizdintuvą (Perkin Elmer). Vaizdai buvo apdorojami OptiQuant programa (Perkin Elmer).

Denatūruojanti PAGE. Denatūruojanti PAGE buvo naudojama DNR karpymo ir apsaugojimo nuo DNazės I eksperimentuose. Elektroforezė vykdyta TBE buferyje naudojant 22×15×0.1 cm arba 50×20×0,02 cm gelius. Geliai sudaryti iš 20 % Akrilamido tirpalo su 8 M karbamido elektroforezės buferyje. Radioaktyvios DNR mėginiai sumaišyti su Mėginio dažu, patalpinami į 95 °C 4 min ir atvėsinami ant ledo. Elektroforezė leidžiama 30 min 30 V/cm arba su didesniu geliu 50 V/cm be mėginių ir 1-2 val. su mėginiais. Aukštos raiškos gelis termostatuojamas 60 °C. Radioaktyvi DNR aptinkama kaip apibūdinta aukščiau.

Baltymų klonavimas, mutagenezė ir raiška. LpnPI, YkrI ir BmeDI genai buvo padauginti PGR būdu nuo *Legionella pneumophila*, *Yersinia kristensenii* ir *Bacillus megaterium* genominių DNR. LpnPI ir LpnPI-N (1-224 ar.) buvo klonuoti į pLATE11 raiškos vektorių. Pirmi metioninai buvo pakeisti trumpa šešių histidinų gryninimui skirtu inkaru. YkrI ir BmeDI genai klonuoti į pTYB2 raiškos vektorių su C-galiniu savaime nusikerpančiu chitino surišimo domenu (CBD). McrA ir McrA-N (1-174 ar.) buvo klonuotas į pLATE31 raiškos vektorių su C-galiniu šešių histidinų inkaru. LpnPI, McrB-N ir McrA mutantai gauti QuickChange metodu (Zheng et al., 2004) naudojant jų raiškos vektorius. Mutacijos patvirtintos sekoskaita. LpnPI, YkrI, BmeDI ir McrA baltymai, jų mutantai ir domenai buvo išreikšti *E. coli* kamiene ER2566. McrB-N baltymai išreikšti *E. coli* DH10B (ara-) kamiene. Ląstelės augintos 37 ºC iki A_{600} 0,5-0,8 ir indukuota 0,2 mM IPTG ar 0,2 % arabinoze bei auginta per naktį 16 ºC, o McrB-N 37 ºC 4 val. Biomasė surinkta centrifuguojant ir saugoma −20 ºC.

1 lentelė. Oligodupleksai naudoti šiame tyrime

Oligodupleksas	Seka	Apibūdinimas
16-M 16-C 16-P 16-T	5′-CGTAGC**X**TGGTCGATC-3′ 3′-GCATCGGACCAGCTAG-5′	16 bp LpnPI specifinis oligodupleksas. X=5mC/C/pyC/T/. M – 5mC, P – pyC. Modifikuota bazė paryškinta, atpažinimo seka pabraukta
16-T-N	5′-CGTAGATCTTACGATC-3′ 3′-GCATCTGGAATGCTAG-5′	Nespecifinis LpnPI oligodupleksas su T-G (pabraukta)
30-M 30-C 30-P 30-T	5′-CCGTAGC**X**TGGTCGATCCTAGCTGGTCGCC-3′ 3′-GGCATCGGACCAGCTAGGATCGACCAGCGG-5′	30 bp serija LpnPI, kaip ir 16 bp serija. M
39-H/H 39-M/H 39-C/H 39-P/H 39-T/H	5′-CGACGATA**X**TTACCGGATAAGGCGCAATTAGATTACTTC-3′ 3′-GCTGCTATGAATGGCCTATTCCGCGTTAAT**X**TAATGAAG-5′	39 bp YkrI ir BmeDI substratų serija: X/X=5hmC/5hmC; 5mC/5hmC; C/5hmC; pyC/5hmC; T/5hmC. H – 5hmC.
39-H	5′-CGACGATATTTACCGGATAAGGCGCAATTAGATTACTTC-3′ 3′-GCTGCTATAAATGGCCTATTCCGCGTTAAT**H**TAATGAAG-5′	Kaip 39-H/H, bet viena 5hmC-G bp pakeista į T-A
39-T	5′-CGACGATA**T**TTACCGGATAAGGCGCAATTAAATTACTTC-3′ 3′-GCTGCTAT**G**AATGGCCTATTCCGCGTTAATTTAATGAAG-5	Kaip 39-T/H, bet 5hmC-G bp pakeista T-A
31-C	5′-TGACCCACGCTCGCCCGGCGACACATTACGT-3′ 3′-ACTGGGTGCGAGCGGGCCGCTGTGTAATGCA-5′	Specifinis Ecl18kI oligodupleksas su C-G centrine bp
31-M	5′-TGACCCACGCTCGCC**M**GGCGACACATTACGT-3′ 3′-ACTGGGTGCGAGCGGGCCGCTGTGTAATGCA-5′	Kaip 31-C, bet centre 5mC-G

1 lentelės tęsinys kitame psl.

1 lentelės tęsinys

12_5mC 12_C 12_5hmC 12_5fC 12_5caC 12_4mC	5′-AGCTA**X**CGGTCTC-3′ 3′-CGATGGCCAGAGT-5′	12 bp oligodupleksai su 1 nt 5′-gale išsikišimu ir skirtingais citozino variantais: X=5mC/C/5hmC/5fC/5caC/4mC. McrB-N atpažinimo seka pabraukta. Naudoti anizotropijos matavimuose (apatinės grandinės 5′-gale yra HEX modifikacija)
13_5mC 13_C 13_5hmC 13_5fC 13_5caC 13_4mC	5′-AGCTA**X**CGGTCTC-3′ 3′-TCGATGGCCAGAG-5′	Taip pat kaip 12_X serija, bet dupleksai bukais galais ir 13 bp ilgio. Naudoti EJP eksperimentuose
12_C/12_C 12_5mC/12_C 12_pC/12_C	5′-ACCTC**X**GGTTCC-3′ 3′-TGGAGGCCAAGG-5′	12 bp oligodupleksų serija naudota EJP ir fluorescencijos su McrA tyrimuose (atpažinimo seka pabraukta)
30_C/30_C 30_5mC/30_C 30_5mC/30_5mC	5′-AGACCCACGCTCAC**X**GGTTCCAGATTTATC-3′ 3′-TCTGGGTGCGAGTGG**X**CAAGGTCTAAATAG-5′	McrA 30 bp oligodupleksų serija naudota EJP ir DNR karpymo eksperimentuose
50_C/50_C 50_5mC/50_C	5′-CAGATTTATCAGACCCACGCTCAC**X**GGTTCCAGATTTATCGATGGTTAAC-3′ 3′-GTCTAAATAGTCTGGGTGCGAGTGGCCAAGGTCTAAATAGCTACCAATTG-5′	50 bp oligodupleksai McrA apsaugojimo nuo DNazės I eksperimentuose

Baltymų gryninimas. Biomasė buvo suardyta Ardymo buferyje (LpnPI: pH 7,5, +10 % v/v glicerolio; McrB-N ir McrA: pH 8,0) naudojant ultragarsą. Ląstelės, kuriose buvo vykdoma YkrI ir BmeDI raiška, suardytos Kolonėlės buferyje su 2 mM fenilmetilsulfonilfluoridu. Lizatai surinkti centrifuguojant 40000 g 1 val. Šešių histidinų inkarą turintys baltymai (LpnPI, McrB-N ir McrA) buvo gryninami chromatografijos būdu per HisTrap HP kolonėlę (GE Healthcare), naudojant linijinį imidazolo gradientą, ir per HiTrap Heparin HP kolonėlę (GE Healthcare) naudojant linijinį NaCl gradientą. YkrI ir BmeDI buvo gryninti per chitininę kolonėlę (NEB). CBD nukirpimui kolonėlė buvo užpildyta Kolonėlės buferiu su 50 mM DTT ir inkubuota per naktį 4 ºC. Po inkubacijos frakcijos su YkrI ir BmeDI buvo eliuojamos su Kolonėlės buferiu ir užneštos ant HiTrap Heparin HP kolonėlės bei eliuojamos buferiu, turinčiu Tris-acetato (pH 7,6) ir 100-1000 mM K-acetato. Gryniausius baltymus turinčios frakcijos surinktos ir dializuotos prieš Saugojimo buferį I (LpnPI, LpnPI-N ir mutantai: pH 7,5-8,5; McrB-N: pH 8,0, +0,1 mM EDTA; McrA: pH 8,0) ar Saugojimo buferį II (YkrI ir BmeDI) bei saugotos −20 ºC.

Baltymų grynumas buvo didesnis nei 95 % sprendžiant iš SDS-PAGE. Koncentracijos suskaičiuotos pagal absorbcija ties 280 nm naudojant teorinius ekstincijos koeficientus, suskaičiuotus su ProtParam įrankiu http://web.expasy.org/protparam/. Visos koncentracijos pateiktos kaip monomero, jei neparašyta kitaip.

DNR surišimo tyrimai

EJP eksperimentai. DNR surišimas buvo tirtas EJP metodu naudojant ^{33}P-žymėtus oligodupleksus (1 lentelė). Radioaktyvi DNR buvo inkubuojama su skirtingomis baltymo koncentracijomis 15 min 20 µl Susirišimo buferyje I ar II papildytame 5 mM Ca-acetato (LpnPI, YkrI ir BmeDI), 1 mM EDTA (LpnPI, YkrI, BmeDI ir McrA) arba 1 mM DTT (McrB-N). Konkurenciniai EJP eksperimentai atlikti su McrB-N (0,5 µM), radioaktyvia 13_5mC/13 DNR (0,5 µM) bei skirtingais kiekiais nežymėtos konkuruojančios DNR su skirtingais citozino variantais (0,5-5,0 µM, 1 lentelė). Laisva DNR ir kompleksai su baltymais buvo atskirti nedenatūruojančia PAGE.

Konkurenciniai EJP geliai analizuoti kaip aprašyta (Sasnauskas, Kauneckaite ir Siksnys 2018), apskaičiuojant disociacijos konstantą žymėtos DNR-baltymo kompleksui ($K_{D(5mC)}$, µM) ir konkuruojančios nežymėtos DNR-baltymo kompleksui (K_D, µM). Visi eksperimentai pakartoti bent 3 kartus ir rezultatai pateikti kaip susirišimo konstantos vidurkis ($1/K_{D(5mC)}$ ir $1/K_D$, in μM^{-1}) ± 1 SD.

Fluorescencijos anizotropijos matavimai. McrB-N anizotropijos matavimai atlikti naudojant oligodupleksus, turinčius HEX žymę (heksachloro-fluoresceinas, 1 lentelė) nemodifikuotos grandinės 5′-gale. Matavimai atlikti su Fluoromax-3 spektrofluorimetru su poliarizatoriais. Sužadinimo bangos ilgis 530 nm (2 mm plyšys), emisijos – 570 nm (16 mm plyšys). Matavimai atlikti buferyje, turinčiame 100 mM KCl, 20 mM Mes-KOH (pH 6,0) 25 °C. Baltymas (20-1600 nM) titruotas į 160 µl kiuvetę su DNR. Gautos anizotropijos reikšmės buvo koreguotos pagal foninę anizotropiją, matuotą su vgDNR su HEX žyme. Rezultatai buvo apdoroti KyPlot programa (2.0 beta 14) (Yoshioka, 2002), panaudojant nelinijinę priklausomybę (1):

$$A([p]) = A_0 + A_1 \times [p]/(K_D + [p]) \quad (1),$$

kur $A([p])$ – stebėta anizotropijos reikšmė esant baltymo koncentracijai [p], A_0 – anizotropijos reikšmė su laisva DNR, A_1 – anizotropijos reikšmių padidėjimo amplitudė, K_D – disociacijos konstanta baltymo-DNR kompleksui. Esant silpnam susirišimui A_1 buvo parinkta pagal metilintos DNR surišimą (13_5mC/13). Rezultatai pateikti kaip susirišimo konstantos vidurkiai ($1/K_D$, μM^{-1}) bent iš 3 matavimų ± 1 SD.

Apsaugojimo nuo DNazės I eksperimentas. 50 nM ^{33}P-žymėta nemetilinta ar pusiau metilinta 50 bp DNR (1 lentelė) buvo preinkubuojama 22 °C su 0,25-0,5 µM McrA homodimeru ar 0,5-1 µM McrA-N monomeru Susirišimo buferyje I. Po 15min pridedama 2,5 µl DNazės I (Thermo Fisher Scientific, 0,004 U/µl) Susirišimo buferyje I su 5 mM $CaCl_2$ ir 12,5 mM $MgCl_2$. Po 2 min reakcija stabdoma ir mėginiai analizuojami aukštos raiškos denatūruojančiame PAGE.

DNR karpymo eksperimentai

Reakcijos su oligodupleksais. DNR hidrolizės reakcijos atliktos sumaišant radioaktyviai žymėtą oligodupleksą (1 nM DNR YkrI/BmeDI reakcijoms, 100-400 nM – LpnPI, 200 nM – McrA) su baltymu (100 nM YkrI, BmeDI; 500 nM LpnPI; 500 nM McrA) Reakcijos buferyje I (YkrI, BmeDI, LpnPI, žr. 1.1.4) arba II (McrA reakcijos papildytos 0,01-10 mM $MgCl_2$, $MnCl_2$, $CoCl_2$, $CaCl_2$, $NiCl_2$, $CuSO_4$ ar $ZnCl_2$) 25 °C (LpnPI), 15 °C (YkrI, BmeDI) ar 37 °C (McrA). Mėginiai analizuoti denatūruojančia PAGE.

Substratų nykimas buvo apibrėžtas eksponentine priklausomybe ir gautos greičio konstantos k_{obs}, suskaičiuotos iš kelių eksperimentų ± 1 SD. Aktyvumas buvo išreikštas santykiu k_{obs}(mutantas)/k_{obs}(wt) × 100 %.

Plazmidinės ir fagų DNR karpymas. McrA ar jo mutantas H229A (0,05 ar 0,5 µM) sumaišyta su 0,5 µg DNR (reakcijos tūris 25 µl). Reakcijos atliktos su superspiralizuotomis plazmidėmis pACYC184 ir pACYC184_M.HpaII, vgDNR M13mp18 (išskirti iš Dcm+ ar Dcm– *E. coli*

kamienų), plazmidine pBR322 ir fago λ DNR (iš Dam+/Dcm+ ar Dam–/Dcm– *E. coli* kamienų). Reakcijos stabdytos fenoliu. Mėginiai analizuoti agaroziniame gelyje.

Bazės išsukimo eksperimentai

Pirolocitozino fluorescencijos matavimai. Fluorescencijos matavimai atlikti Fluoromax-3 spektrofluorimetru su Xe lempa 25 °C. Emisijos spektras (440-460 nm) užrašytas sužadinant 350 nm bangos ilgiu su 5 nm plyšiais. Mėginiai turėjo 1,0-2,0 µM baltymo ir 0,2-0,5 µM pyC-žymėtos DNR (1 lentelė) buferyje su 30 mM Mes-histidino, pH 6,3 (LpnPI, YkrI, BmeDI) ar 20 mM Tris-HCl, pH 8,0, ir 100 mM KCl (McrA). Kontrolinis spektras fonui pakoreguoti buvo matuotas tomis pačiomis sąlygomis, tik be pyC modifikacijos. Fluorescencijos intensyvumo pokyčių reikšmės pridėjus baltymo buvo apskaičiuotos iš kelių eksperimentų ±1 SD.

Reakcijos su CAA. DNR modifikavimas chloroacetaldehidu (CAA) buvo atliktas kaip aprašyta (Daujotyte et al., 2008). 100 nM radioaktyvios DNR sumaišyta su 2 µM Ecl18kI, BmeDI ar YkrI 20 µl Susirišimo buferio I. Reakcijos pradėtos pridedant CAA iki 0,5 M ir inkubuojama 1 h 37 °C. Modifikuotos grandinės skaldymas atliktas pridedant 100 µl šviežiai pasigaminto 1 M piperidino ir kaitinama 90 °C 30 min. DNR buvo išskirta su etanoliu ir analizuojama aukštos raiškos denatūruojančia PAGE. Molekulinės masės standartai A+G gaminti standartine Maxam-Gilbert sekoskaitos reakcija.

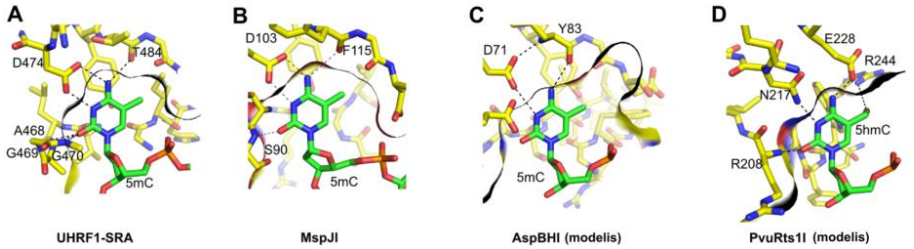
Reakcijos su $KMnO_4$. Eksperimentai su timinu-pakeistais substratais (1 lentelė) buvo atlikti kaip aprašyta (Serva et al., 1998). Radioaktyvi DNR (10 nM) ir baltymas (100 nM) sumaišyti Susirišimo buferyje II. Reakcijos inicijuotos pridedant $KMnO_4$ iki 2 mM bei inkubuota 5 min 25 °C, stabdyta 1,5 M Na-acetato (pH 7,0) ir 1 M 2-merkaptoetanolio tirpalu. Modifikuotos grandinės skaldymas atliktas kaip aprašyta aukščiau. 5mC oksidacija atlikta taip pat, tik naudotas buferis, turintis 20 mM Na-acetato (pH 4,3 (Fritzsche et al., 1987)).

2 REZULTATAI IR JŲ APTARIMAS

2.1 Bazės išsukimo tyrimai MspJI ir PvuRts1I šeimų restrikcijos endonukleazėse

SRA domenai UHRF1, UHRF2 ir SUVH5 baltymuose išsuka modifikuotą bazę iš dgDNR ir patalpina ją baltymo kišenėje (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008; Rajakumara et al., 2011a; Zhou et al., 2014) (2.1 A pav.). SRA tipo domenai prokariotuose

veikia kaip modifikuoto citozino atpažinimo moduliai nuo modifikacijų priklausomose REazėse, kurios apsaugo šeimininko ląsteles nuo bakteriofagų, turinčių 5mC, 5hmC ir 5ghmC savo DNR.



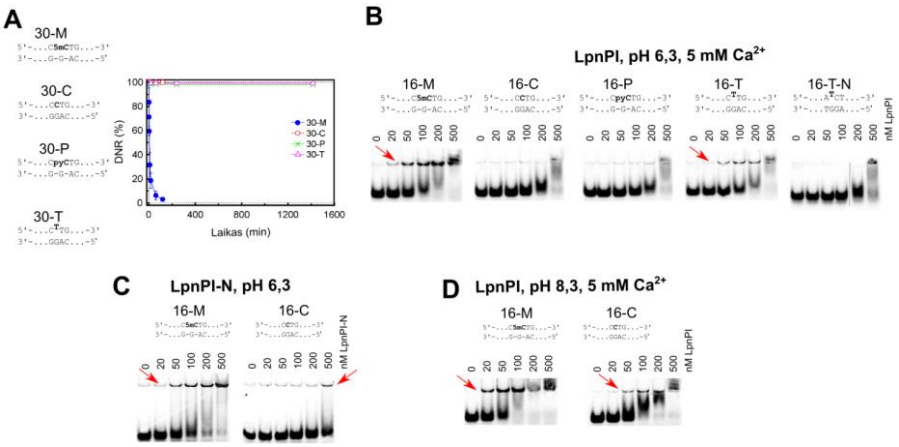
2.1 pav. Modifikuoto citozino surišimo kišenės. (A-B) 5mC atpažinimas UHRF1 SRA domene (PDB ID 3fde) ir MspJI DNR atpažinimo domene (PDB ID 4r28). Parodytos ar. sudaro specifinius kontaktus su išsukta baze (Hashimoto et al., 2009; Horton et al., 2014b). (C-D) Modifikuoto citozino atpažinimo modeliai su AspBHI (PDB ID 4oc8) ir PvuRts1I (PDB ID 4oq2) DNR atpažinimo domenais. Parodytos ar., kurios galėtų specifiškai sąveikauti su išsukta baze. Juoda linija žymi baltymo kišenės kraštus citozino žiedo plokštumoje.

Nepaisant to, kad yra nustatytos MspJI, AspBHI (MspJI šeima), PvuRts1I ir AbaSI (PvuRts1I šeima) struktūros (Horton et al., 2012, 2014b, 2014a, 2014c; Kazrani et al., 2014; Shao et al., 2014), tik MspJI atveju stebėtas bazės išsukimas (Horton et al., 2014c) (2.1 B pav.). Remiantis mutacine analize ir dideliu AspBHI, PvuRts1I bei AbaSI DNR atpažinimo domenų struktūriniu panašumu į MspJI bei eukariotinius SRA domenus, galima manyti, kad šie fermentai taip pat turėtų išsukti modifikuotą citoziną (Horton et al., 2012, 2014a, 2014c; Kazrani et al., 2014; Shao et al., 2014). Tačiau daugeliui eukariotinių ir prokariotinių baltymų 5(h)mC surišimo ir atpažinimo mechanizmas lieka nežinomas. Šiame tyrime mes išbandėme keletą tirpale veikiančių metodų, tokių kaip fluorescencijos matavimai su citozino analogu pirolocitozinu (pyC) bei cheminė išsuktų iš dgDNR pirimidinų modifikacija su chloroacetaldechidu (CAA) ir $KMnO_4$. Šie metodai padėtų greitai ir lengvai nustatyti išsuktas iš dgDNR bazes.

2.1.1 DNR karpymas ir sąveika su LpnPI, YkrI ir BmeDI

Šiame tyrime naudojome tris modifikuotą DNR karpančias REazes: LpnPI, YkrI ir BmeDI. LpnPI priklauso MspJI šeimai ir atpažįsta DNR seką 5′-CMDG-3′ (kur M yra 5mC arba 5hmC, D – A, T arba G) (Zheng et al., 2010). Ši REazė panaši į struktūriškai charakterizuotą baltymą AspBHI, įskaitant ir beveik pilnai konservatyvią numanomą 5(h)mC surišimo kišenę

(2.1 C pav.). LpnPI karpo 30-M oligodupleksą ~0,2 min^{-1} greičiu, tuo tarpu nemetilinto atitinkamo substrato 30-C – nekarpo (2.2 A pav.).

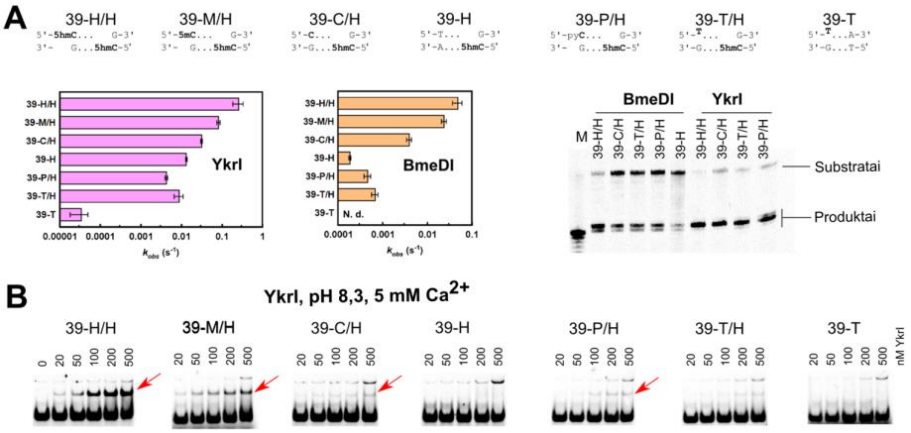


2.2 pav. DNR karpymas ir sąveika su LpnPI. Aprašytos atpažinimo sekos schematiškai vaizduoja oligodupleksus. (A) DNR karpymas. Reakcijos atliktos su 500 nM baltymo (monomero) ir 400 nM substrato 25 °C. Laikai parodyti grafike. Reakcijos greičio konstanta su 30-M substratu yra 0,20±0,05 min^{-1}. Reakcijos greičio konstantos su kitais substratais buvo žemesnės nei 1×10^{-5} min^{-1}. (B) EJP eksperimentai su LpnPI. DNR susirišimo eksperimentai atlikti pH 6,3 buferyje esant 5 mM Ca^{2+}. Galutinė substrato koncentracija buvo 10 nM, LpnPI koncentracija (monomero) pavaizduota virš gelių. Raudonos rodyklės parodo baltymo-DNR kompleksus. (C) EJP eksperimentai su LpnP-N. Eksperimentai su specifiniais (16-M) ir nespecifiniais (16-C) substratais atlikti pH 6,3 buferyje be Ca^{2+} jonų. (D) EJP eksperimentai su LpnPI pH 8,3 buferyje esant Ca^{2+}.

Tą patį diskriminavimą tarp metilinto ir nemetilinto substrato matome ir EJP eksperimentuose: ir LpnPI, ir N-galinis LpnPI DNR surišimo domenas (LpnPI-N) formavo blatymo-DNR kompleksus su metilinta DNR prie daug mažesnių koncentracijų nei su nemetilinta DNR (2.2 B, C pav.). Tačiau didesnis skirtumas tarp specifinės (metilintos) ir nespecifinės DNR surišimo buvo esant pH 6,3, esant pH 8,3 – skirtumas nebuvo ryškus (2.2 D pav.). Tai netiesiogiai patvirtina, kad MspJI šeimos baltymai atpažįsta išsuktą 5mC baltymo kišenėje, kurioje konservatyvūs aspartatai (D103 MspJI, D71 AspBHI ir LpnPI, 2.1 B, C pav.) turi būti protonizuoti (Horton et al., 2014b, 2014a).

YkrI ir BmeDI sekos yra panašios į struktūriškai charakterizuotų PvuRts1I šeimos atstovų PvuRts1I ir AbaSI sekas (Sasnauskas et al., 2015). PvuRts1I optimaliai kerpa DNR substratus, kuriuose yra du 5hmC arba 5ghmC nukleotidai priešingose grandinėse atskirti 20-22 bp (Szwagierczak et al., 2011; Wang et al., 2011). Iš biocheminių ir struktūrinių duomenų yra

žinoma, kad 5(g)hmC taikinius atpažįsta DNR surišimo domenai, o du nukleaziniai domenai formuoja dimerą ir kerpa DNR per vidurį, ~11 nt nuo taikinių (Horton et al., 2014c; Shao et al., 2014). Vieną iš 5hmC pakeitus 5mC, C ar kitu nukleotidu 39 bp substratų serijoje (oligodupleksai 39-H/H, 39-M/H, 39-C/H ir 39-H, 1 lentelė) YkrI ir BmeDI aktyvumas neprarandamas, tačiau reakcijos greitis mažėja (2.3 A pav.). Tai reiškia, kad užtenka ir vieno DNR surišimo domeno prijungiant baltymo dimerą prie DNR per 5hmC bazę. Tokiu atveju antrasis YkrI / BmeDI DNR surišimo domenas tikriausiai sudaro kontaktus su baze, kuri nutolusi ~20 bp po 5hmC nukleotido ir suteikia baltymo-DNR kompleksui papildomo stabilumo priklausomai nuo šios bazės struktūrinio panašumo į 5(g)hmC. YkrI DNR karpymas atitinka EJP eksperimentus, kuriuose matomas mažėjantis baltymo-DNR komplekso kiekis keičiant antrąją bazę 5mC, C ar kita baze (2.3 A pav.).



2.3 pav. YkrI ir BmeDI DNR karpymo ir DNR surišimo tyrimai. Sekos virš paveikslų schematiškai vaizduoja 39-H/H, 39-M/H, 39-C/H, 39-H, 39-P/H, 39-T/H ir 39-T (viena iš dviejų 5hmC-G bp pakeista 5mC-G, C-G, T-A, pyC-G ir T-G bp, 1 lentelė) oligodupleksus. (A) DNR karpymo greičio konstantos. Reakcijos atliktos su 1 nM substrato ir 100 nM baltymo (monomero) esant 15 °C. Mūsų naudotomis sąlygomis BmeDI nekarpė 39-T oligoduplekso. Denatūruojančios PAGE karpymo produktų, susidariusių su įvairiais DNR substratais, analizė pavaizduota dešinėje. Gelio paveiksle „M“ parodo sintetinį vg oligonukleotidą, kurio ilgis atitinka apatinės grandinės kirpimo produktą, 11 nt po 5hmC. (B) EJP eksperimentai su YkrI. DNR susirišimas buvo tirtas pH 8,3 buferyje esant 5 mM Ca^{2+}. DNR koncentracija 1nM, YkrI koncentracija pavaizduota virš gelių paveikslų. Raudonos rodyklės vaizduoja specifinį YkrI-DNR kompleksą. Aukštesnė juostelė parodo mažo judrumo nespecifinį YkrI-DNR kompleksą, susiformavusį dėl agregacijos.

2.1.2 Pirolocitozino fluorescencijos matavimai

Pirolocitozinas (pyC) yra fluorescentinis citozino analogas, kuris formuoja stabilias bazių poras su guaninu. Kvantinė pyC fluorescencijos išeiga yra jautri bazės išsukimui (Berry et al., 2004), todėl pyC fluorescencijos matavimai gali būti naudojami nustatant pyC struktūrinę aplinką nukleorūgštyse ir baltymų-DNR kompleksuose (Kuznetsov et al., 2012; Liu ir Martin, 2001). Šis metodas buvo sėkmingai pritaikytas tiriant modifikuotą DNR atpažįstantį baltymą McrBC (Sukackaite et al., 2012). Tačiau, kad šis metodas veiktų, tiriamas baltymas privalo turėti atitinkamo dydžio baltymo kišenę, kuri talpintų didesnę nei natūralūs citozino dariniai pyC bazę. Nors tai tiko McrBC DNR surišimo domeno atveju, tačiau modeliavimas (Zagorskaitė ir Sasnauskas, 2014), DNR susirišimo ir karpymo eksperimentai su LpnPI, BmeDI ir YkrI parodė, kad šių baltymų SRA domenai tokių pakeitimų netoleruoja (2.2, 2.3 pav.), tad ir fluorescencijos pokyčių nestebėjome su šių baltymų-DNR kompleksais (Zagorskaitė ir Sasnauskas, 2014).

2.1.3 Reakcijos su CAA

CAA reaguoja su nesuporuotomis citozino ir adenino bazėmis esančiomis DNR susidarant 3,N4-etenocitozinui ir 1,N6-etenoadeninui (Kuśmierek ir Singer, 1982). Tokie modifikuoti nukleotidai nustatomi piperidinu indukuojamu DNR skaldymu. CAA reakcija buvo panaudota nustatant nemodifikuoto citozino išsukimą su keletu metiltransferazių ir REazių (Daujotyte et al., 2008). Kadangi CAA reaguoja ir su 5mC (Oakeley et al., 1999), nusprendėme patikrinti, ar tokiu pačiu būdu galima nustatyti išsuktą 5mC. Deja mūsų naudotomis sąlygomis šis metodas veikė tik su nemodifikuotu citozinu, todėl neveikė nei su vienu mūsų tirtų baltymų (Zagorskaitė ir Sasnauskas, 2014).

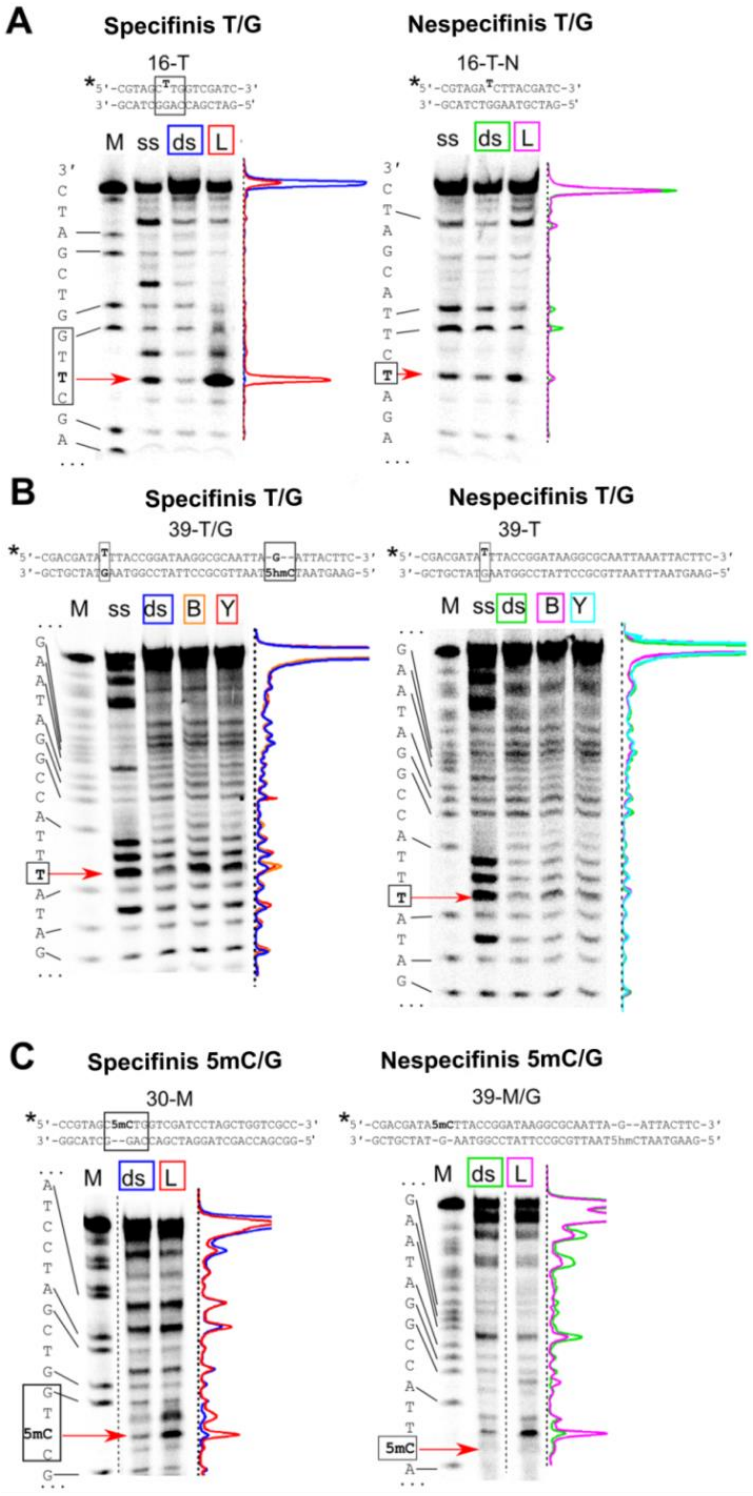
2.1.4 Iš dgDNR išsuktų pirimidinų oksidacija $KMnO_4$

Rūgštinėmis sąlygomis kalio permanganatas oksiduoja timiną ir 5mC (Fritzsche et al., 1987). Tačiau esant fiziologiniam pH ši reakcija vyksta tik su timinu. $KMnO_4$ oksiduoja pirimidino bazes iki 5,6-dihidroksi-5,6-dihidropirimidinų (Hayatsu ir Ukita, 1967); po to nukleorūgštį paveikus piperidinu oksiduotų bazių vietose yra skeliamas fosfodiesterinis karkasas. C5=C6 ryšio oksidacijai reikalingas priėjimas prie pirimidino žiedo krašto, kuris yra dgDNR, tad T ir 5mC DNR spiralėje yra atsparūs permanganato oksidacijai lyginant su išsuktais pirimidinais. $KMnO_4$ buvo panaudotas

nustatant timinų išsukimą iš dgDNR tiriant citozino ir adenino metiltransferazes ir sekai specifines transpozazes (Bischerour ir Chalmers, 2007; Reddy ir Rao, 2000; Serva et al. 1998). Nusprendėme patikrinti, ar oksidacija $KMnO_4$ galėtų būti pritaikoma nustatant bazės išsukimą tiriant modifikuotus taikinius atpažįstančias REazes.

Esant neutraliam pH šis metodas reikalauja 5mC / 5hmC pakeitimo timinais, kurie suformuoja T-G nesuporuotų nukleotidų porą. Dėl sutrikdytos geometrijos tokioje bazių poroje, nesuporuoti timinai gali savaime tapti reaktingesni (Gogos et al., 1990). Tačiau žemas foninis signalas mūsų eksperimentuose rodė, kad T-G nesuporuotų timinų prieinamumas tirpale yra mažas (oligodupleksai 16-T ir 39-T/H, 1 lentelė). Nors T yra panašaus dydžio į 5mC ir 5hmC, tačiau yra prastai toleruojami modifikuotą DNR atpažįstančių baltymų, naudotų mūsų tyrime – tiek susirišimas, tiek DNR kirpimas buvo susilpnėjęs su T-pakeistais substratais (2.2 ir 2.3 pav.). Nepaisant to, inkubuojant T-pakeistą DNR su LpnPI, YkrI ir BmeDI stebėjome nesuporuotų T padidintą jautrumą $KMnO_4$ oksidacijai. Naudojant tokius substratus iškyla rizika, kad nesuporuoti nt gali sukelti papildomą konformacinį lankstumą toje vietoje ar šalia, ir susirišant su baltymu gali įvykti didesni konformaciniai pokyčiai lyginant su standartine dgDNR. Tačiau mūsų eksperimentuose reaktingumo padidėjimas buvo stebėtas su taikinio timinu „specifiniame“ T-pakeistame substrate (2.4 A, B pav.), kas leidžia manyti, kad susiformavo natyvaus tipo baltymo-DNR kompleksas su išsuktu timinu.

Struktūriniai duomenys ir modeliavimas parodė, kad išsukta bazė patalpinama SRA tipo domenų kišenėje tarp konservatyvių polinių, aromatinių ir hidrofobinių ar. (2.1 pav.). Jeigu nesuporuotas T užima tą pačią poziciją kaip ir 5mC / 5hmC, jis turėtų būti apsaugotas nuo $KMnO_4$ oksidacijos. Tačiau padidintas T reaktingumas rodo, kad T bazė gali būti dinaminėje pusiausvyroje tarp neišsuktos būsenos ir įvairių dalinai išsuktų būsenų ansamblio. Pagrindinis trukdis, neleidžiantis pilnai išsukti ir uždaryti T baltymo kišenėje, tikriausiai yra tai, kad T nesudaro H-ryšių su kišenės ar., kurios skirtos tiesiogiai sąveikai su citozino dariniais (2.1 pav.); iš kitos pusės, T-G nesuporuotų nukleotidų porą atskirti yra daug lengviau nei natyvią C-G bp, tai galėtų paslinkti pusiausvyrą į išsuktą būseną. Įdomu tai, kad kristalinėje MspJI-DNR komplekso struktūroje (Horton et al., 2014b) matomas specifinis bazės išsukimas viename DNR susirišimo domene ir nespecifinis guanino išsukimas kitame DNR atpažinimo domene. Pastaruoju atveju išsukta bazė užima šiek tiek kitokią poziciją nei 5mC, o susirišimo kišenė išlieka atviresnės konformacijos.


A
Specifinis T/G
16-T
Nespecifinis T/G
16-T-N
M ss ds L
ss ds L
B
Specifinis T/G
39-T/G
Nespecifinis T/G
39-T
M ss ds B Y
C
Specifinis 5mC/G
30-M
Nespecifinis 5mC/G
39-M/G
M ds L
5mC

2.4 pav. Pirimidinų bazių, esančių baltymo-DNR komplekse, reakcijos su permanganatu. Sekos virš paveikslų schematiškai vaizduoja „specifinius“ ir „nespecifinius“ T-pakeistus substratus, žvaigždutė žymi radioaktyvų ^{33}P. Bp, svarbios specifiniam susirišimui, yra apibrauktos juodai. Nesuporuotų T ir 5mC pozicijos pažymėtos raudona rodykle. (A) T oksidacija $KMnO_4$ su ir be LpnPI-N. „M“ – A+G molekulinės masės žymuo, padarytas iš „specifinio“ ir „nespecifinio“ radioaktyviai žymėto substrato; „vg“, viengrandinio oligonukleotido oksidacija; „dg“, 16-T („specifinis“) ir 16-T-N („nespecifinis“) dg substratai be baltymo; „L“, 16-T ir 16-T-N oligodupleksai su LpnPI-N. Tankio profiliai individualiems takeliams parodyti specifinei DNR (mėlyna), specifinei DNR + LpnPI-N (raudona), nespecifinei DNR (žalia) ir nespecifinei + LpnPI-N (violetinė). (B) T oksidacija $KMnO_4$ esant YkrI ir BmeDI. „dg“, 39-T/H („specifinis“) ir 39-T („nespecifinis“) dg substratai be baltymo; „Y“, 39-T/H ir 39-T oligodupleksai + YkrI; „B“, 39-T/H ir 39-T oligodupleksai + BmeDI. Tankio profiliai: specifinė DNR (mėlyna), specifinė DNR + BmeDI ir YkrI (oranžinė ir raudona), nespecifinė DNR (žalia), nespecifinė DNR + BmeDI ir YkrI (violetinė ir žydra). (C) 5mC oksidacija $KMnO_4$ esant pH 4,3 su ir be LpnPI-N. 30-M – specifinis, 39-M/H – nespecifinis substratai. Kiti žymėjimai kaip (A).

Tokio komplekso susiformavimas buvo interpretuojamas kaip galima tarpinė taikinio paieškos būsena (Horton et al., 2014b). Išsuktas T stebimas mūsų eksperimentuose su LpnPI, YkrI, BmeDI galėtų užimti panašią ne natyvią poziciją, kuri yra prieinama $KMnO_4$.

Norėdami sustiprinti įrodymus, kad natyvus modifikuotas citozinas yra išsukamas modifikuotą DNR atpažįstančių baltymų, mes atlikome 5mC oksidaciją permanganatu esant pH 4,3. Dėl nefiziologinių sąlygų seniau šis metodas buvo pritaikytas 5mC nustatymui tik DNR (Fritzsche et al., 1987). Nepaisant to, stebėjome išaugusio reaktingumo signalą su specifiniu LpnPI substratu (2.4 C pav.), turinčiu 5mC atitinkamame sekos kontekste, tai rodo, kad LpnPI išsuka 5mC iš dgDNR (ar išsukta bazė užima tą pačią poziciją, kaip ir natyviame komplekse esant naudotam rūgštiniam pH, lieka neaišku). YkrI ir BmeDI atveju signalo nestebėjome, labiausiai tikėtina, dėl nefiziologinio pH, kuris sutrikdo daugumos baltymų susirišimą su DNR. Tačiau $KMnO_4$ metodas galėtų veikti tokiame žemame pH su tam tikrais baltymais ir galėtų būti naudingas tiriant kitų modifikuotą citoziną surišančių baltymų atpažinimo mechanizmą.

Apibendrinant, mūsų atliktas tyrimas atskleidė citozino ir jo darinių bazės išsukimui tirpale naudojamų metodų privalumus ir trūkumus. Tik vienas tyrime naudotas metodas – išorinių pirimidinų oksidacija permanganatu – parodė bazės išsukimą modifikuotą DNR atpažįstančiomis REazėmis. Tai reiškia, kad nesant teigiamo signalo naudojant vieną ar kelis metodus negalima teigti, kad bazės išsukimo nėra. Todėl DNR atpažinimo

mechanizmas potencialiems 5mC / 5hmC surišantiems baltymams turi būti tiriamas naudojant kombinaciją įvairių prieinamų metodų. Vis dėlto, galutinis įrodymas, kad bazė yra išsukama (arba neišsukama), yra eksperimentiškai nustatyta baltymo-DNR komplekso struktūra.

2.2 Išsukto 5-metilcitozino ir jo sekos konteksto atpažinimas LpnPI

SRA tipo domenai yra universalūs moduliai, kurie naudojami modifikuoto citozino atpažinime skirtinguose sekos kontekstuose (2.1 lentelė). Subtilūs baltymo kišenės struktūriniai ir dydžio skirtumai yra svarbūs atskiriant 5mC, 5hmC ir 5ghmC bazes šiuose domenuose (Horton et al., 2014c; Shao et al., 2014; Zhou et al., 2014). Daugeliu atvejų modifikuotas citozinas atpažįstamas tik specifiniame kontekste, kas parodo stiprų ryšį tarp bazės išsukimo ir aplinkinės sekos atpažinimo. Šie SRA domenų sekos atpažinimo mechanizmai yra mažai ištirti. Eukariotiniame UHRF1-SRA domene 5mC yra atpažįstamas CpG sekos kontekste: dvi kilpos (angl. *„thumb“* ir *„NKR finger“*) įeina į didįjį ir mažąjį griovius ir sudaro sekai specifinius kontaktus su CpG dinukleotidu (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). MspJI šeimos baltymų taikiniai aplink modifikuotą citoziną yra įvairaus dydžio, dažnai nevienareikšmės (angl. *degenerate*) sekos apimančios iki 2 bp prieš ir iki 3 bp už modifikuoto citozino (2.1 lentelė).

MspJI atpažįsta 5′-(mC)NNR-3′ seką, daugiausiai kontaktų sudaro iš mažojo griovio pusės (Horton et al., 2014b), tačiau molekulinis kitų MspJI šeimos baltymų DNR atpažinimo mechanizmas dar nėra nustatytas. Tad mes susikoncentravome į MspJI šeimos baltymo LpnPI specifiškumo tyrimus, norėdami suprasti ir keisti struktūrinius modulius, atsakingus už 5mC sekos konteksto atpažinimą.

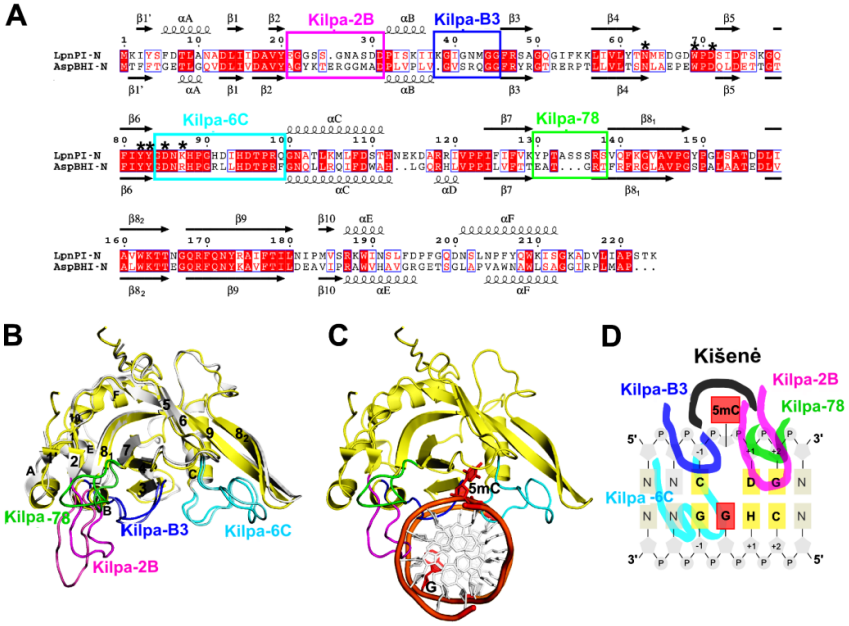
2.1 lentelė. SRA domenai ir jų atpažinimo sekos

Baltymas	Atpažinimo seka[a]	Modifikacija
UHRF1-SRA	5′-(mC)G-3′	5mC, 5hmC
UHRF2-SRA	5′-(mC)G-3′	5hmC > 5mC
SUVH5-SRA	5′-(mC)G-3′	5mC
	5′-(mC)HH-3′	
MspJI	5′-(mC)NNR-3′	5mC, 5hmC
AspBHI	5′-YS(mC)NS-3′	5mC, 5hmC
LpnPI	5′-C(mC)DG-3′	5mC, 5hmC
PvuRts1I	5′-(mC)-3′	5hmC, 5ghmC
AbaSI	5′-(mC)-3′	5hmC, 5ghmC

[a](mC) – modifikuotas citozinas; N –bet koks nt; D – A, T, G; Y – T, C; S – G, C; R – A, G. Lentelė pritaikyta iš (Sasnauskas et al., 2015).

2.2.1 LpnPI-N kristalinė struktūra

LpnPI baltymo seka panaši į struktūriškai charakterizuotų MspJI šeimos baltymų sekas (Sasnauskas et al., 2015): AspBHI (42 / 63 % vienodų / panašių ar. N-galiniame domene, (Horton et al., 2014a)) ir MspJI (~15 / 33 % vienodų / panašių ar. (Horton et al., 2012, 2014b)). Nustatyta LpnPI N-galinio SRA domeno (LpnPI-N) struktūra yra labai panaši į atitinkamo AspBHI domeno struktūrą (2.5 A pav.). Įdomiausias skirtumas tarp AspBHI-N ir LpnPI-N yra kilpos-2B, kuri dalyvauja susirišime su DNR, konformacija ir ilgis (21-31 ar., AspBHI atitinkamai 22-33 ar.).



2.5 pav. LpnPI DNR atpažinimo domenas. (A) LpnPI-N ir AspBHI-N sekų palyginys. Antrinės struktūros elementai sunumeruoti kaip (Horton et al., 2014a). Ar., kurios formuoja išsuktos bazės surišimo kišenę, pažymėtos žvaigždutėmis. (B) LpnPI-N (geltona) ir AspBHI-N (balta; PDB 4OC8) struktūrų perdengimas. Spėjamos LpnPI / AspBHI DNR atpažinimo kilpos nuspalvintos šiomis spalvomis: kilpa-B3, mėlyna; kilpa-78, žalia; kilpa-2B, violetinė; kilpa-6C, žydra. (C) LpnPI-N modelis su DNR, remiantis UHRF1-SRA-DNR struktūra (PDB 3FDE). DNR atpažinimo kilpos nuspalvintos kaip (B), išsuktas citozinas ir vidinis vienas likęs G nuspalvinti raudonai. (D) LpnPI schematinė sąveika su DNR. Baltymo kilpos ir 5mC:G bp nuspalvinta kaip (C); kitos bazės, esančios LpnPI atpažinimo sekoje nuspalvintos šviesiai oranžine ir sunumeruotos nuo „−1“ (bp prieš 5mC) iki „+2“ (antra bp už 5mC).

2.2.2 Išsukto 5mC atpažinimas surišimo kišenėje

Yra žinomos keleto eukariotinių SRA domenų (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008; Rajakumara et al., 2011a; Zhou et al., 2014) bei MspJI REazės struktūros su DNR (Horton et al., 2014b). Kadangi užklojant LpnPI-N arba AspBHI-N struktūras su UHRF1-SRA ar MspJI struktūromis DNR molekulė ir išsukta 5mC bazė atsiduria panašioje pozicijoje, aptarsime LpnPI-N bei AspBHI-N ir DNR modelius, kurie sudaryti remiantis UHRF1-SRA-DNR struktūra (2.5 C, D pav.). Bazes, kurios yra 5′ pusėje nuo išsukto citozino, vadinsime −1 ir −2 pozicijomis, 3′ pusėje – +1, +2 ir +3 pozicijomis.

Modifikuoto citozino surišimo kišenė yra konservatyvi SRA domenuose. Paprastai šoninės kišenės sienelės yra iš aromatinių šoninių grandinių, kurios sudaro stekingo sąveikas su išsukta baze, ir polinės ar., kurios sudaro citozinui specifinius H-ryšius su Watson-Crick bazės kraštu (2.6 pav.). Kišenės mutacijos AspBJI, PvuRts1I ir AbaSI baltymuose panaikina DNR kirpimo aktyvumą (Horton et al., 2014a, 2014c; Kazrani et al., 2014; Shao et al., 2014). LpnPI yra ne išimtis: D71A mutacija sumažina DNR kirpimo greitį ∼1000 kartų (2.2 lentelė). Iš MspJI-DNR komplekso struktūros ir AspBHI / LpnPI modelių su DNR spėjama, kad kišenės aspartatas (D71 LpnPI) turi būti protonizuotas, kad sudarytų H-ryšius su išsuktos bazės N4 citozino atomu (2.6 B pav.). Kadangi panašioje pozicijoje kitame modifikuotą citoziną surišančiame fermente PvuRts1I atsiduria asparaginas (N217), sukonstravome LpnPI mutantą D71N. Įdomu tai, kad netgi tokia konservatyvi mutacija sumažino LpnPI kirpimo greitį 10 kartų (2.2 lentelė). Tikriausiai kišenės struktūra yra labai optimizuota, tad netgi nedideli geometrijos, H-ryšių ar krūvių pasiskirstymo pokyčiai turi didelį efektą fermento funkcijai.

2.2.3 DNR sekos atpažinimas kilpa-B3

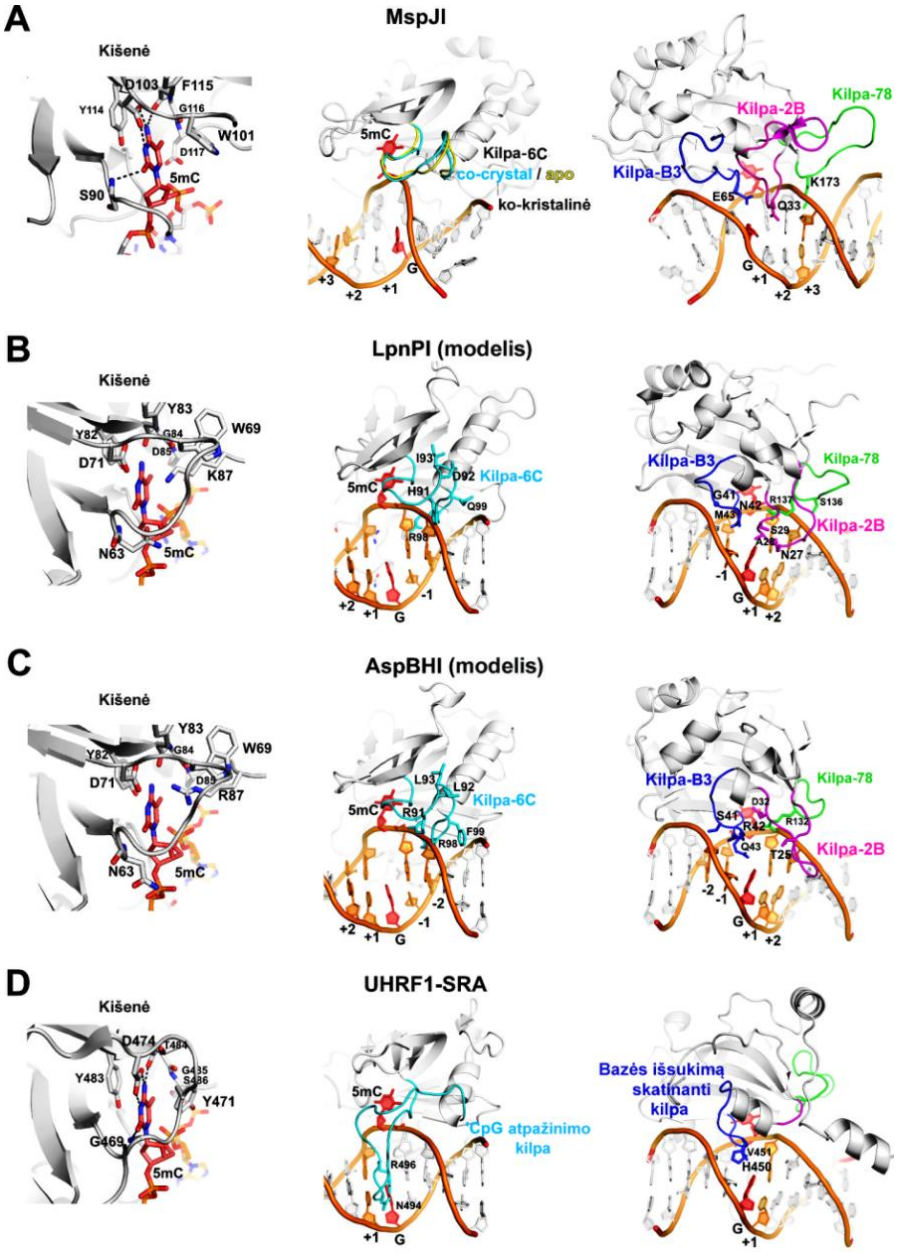
LpnPI kilpa-B3 DNR pasiekia iš mažojo griovio pusės. Struktūriškai lygiavertėse kitų SRA domenų kilpose esančios ar. (pvz., V451 UHRF1-SRA) užpildo laisvą vietą, likusią po citozino išsukimo, ir dalyvauja šalia esančių bp atpažinime (2.6 D pav.) (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). MspJI-DNR kristalinėje struktūroje panašią poziciją užima E65, kuri sąveikauja su likusiu nesuporuotu G (2.6 A pav.). AspBHI B3-kilpoje yra S41, R42 (abi ar. unikalios AspBHI) ir Q43; tose pačiose pozicijose LpnPI yra G41, N42 ir M43 (unikalios LpnPI). Pakeitus 42-ą poziciją abiejuose baltymuose B3-kilpoje šie praranda katalizinį aktyvumą ((Horton et al., 2014a), 2.2 lentelė), kas įrodo R42 / N42 tiesioginį

vaidmenį vieno likusio G ar šalia esančios bp atpažinime. Su UHRF1-SRA V451 persidengiančių Q43 / M43 ar. pakeitimas alaninu buvo mažiau žalingas, o LpnPI M43Q mutantas turėjo wt DNR kirpimo aktyvumą (2.2 lentelė). Todėl tikėtina, kad pagrindinis didelių M43 / Q43 ar. vaidmuo yra užpildyti erdvę, likusią po 5mC išsukimo, o ne sudaryti bazei specifinius kontaktus (2.6 B pav.).

AspBHI −2 pozicijoje teikia pirmenybę pirimidinui (C arba T), tuo tarpu LpnPI ir kiti susiję baltymai toleruoja bet kokį nt šioje pozicijoje (Cohen-Karni et al., 2011). Mūsų LpnPI bei AspBHI modeliuose su DNR (2.6 B, C pav.) artimiausios ar. −2 bp yra G41 / S41. Spėjame, kad neturint šoninės grandinės 41-oje pozicijoje, LpnPI toleruoja bet kokį nt −2 pozicijoje. Tuo tarpu AspBHI šią poziciją užima serinas, kuris galėtų dalyvauti pirimidino / purino atskyrime, pvz. per mažojo griovio H-ryšį su N3 purino atomu komplementarioje grandinėje (2.6 C pav.). Patvirtinant šį modelį kai kurie AspBHI S41 mutantai turėjo pakitusį specifiškumą (Horton et al., 2014a). G41S pakeitimas LpnPI sumažino kirpimo aktyvumą ∼300 kartų, bet nepakeitė specifiškumo −2 pozicijoje esančiai bazei (duomenys neparodyti). Todėl negalima atmesti kitų veiksnių, tame tarpe subtilių B3-kilpos konformacijos skirtumų, kurie gali lemti AspBHI ir LpnPI −2 bp atpažinimo skirtumus.

2.2 lentelė. LpnPI mutantų katalizinis aktyvumas

Mutacija	k_{obs} **(s^{-1})**	**Aktyvumas (%)**
wt LpnPI	$(3,3 \pm 0,8) \times 10^{-3}$	100
5(h)mC surišimo kišenė		
D71A	$(1,0 \pm 0,3) \times 10^{-6}$	0,03
D71N	$(2,0 \pm 0, 4) \times 10^{-4}$	6
Kilpa-2B (kontaktai po 5(h)mC)		
S25A	$(7,0 \pm 1,5) \times 10^{-3}$	200
N27A	Nekarpo	<0.01
D30A	$(1,0 \pm 0,3) \times 10^{-3}$	30
Kilpa-B3 (šalia vieno likusio G)		
G41S	$(1,0 \pm 0,1) \times 10^{-5}$	0,3
N42A	Nekarpo	<0,01
M43A	$(1,0 \pm 0,2) \times 10^{-4}$	3
M43Q	$(1,6 \pm 0,1) \times 10^{-3}$	50
Kilpa-6C (kontaktai prieš 5(h)mC)		
R98A	$(1,0 \pm 0,6) \times 10^{-6}$	0,03
Kilpa-78 (kontaktai po 5(h)mC)		
R137A	$(0,7 \pm 0,3) \times 10^{-5}$	0,2
S136A	$(2,1 \pm 0,1) \times 10^{-3}$	60



2.6 pav. DNR atpažinimas SRA domenuose. UHRFI-SRA ir MspJI struktūros su DNR ir MspJI, LpnPI-N bei AspBHI-N apo struktūros (PDB ID: 3FDE, 4R28, 4F0Q, 4RZL, 4OC8). Ekvivalentiški DNR atpažinimo elementai visur pavaizduoti vienoda orientacija. Kairėje –išsuktos bazės atpažinimas kišenėje; viduryje – kilpos-6C ar „CpG atpažinimo“/„NKR piršto“ kilpos (žydra); dešinėje – kilpos-B3 ar bazės išsukimą skatinančios kilpos (mėlyna), kilpos-B2 (violetinė) ir kilpos-78 (žalia). Išsukta bazė ir likęs vienas guaninas nuspalvinti raudonai; kiti nt, įeinantys į atpažinimo seką, nuspalvinti oranžine spalva ir sunumeruoti nuo −2 (antra bp prieš

5mC) iki +3 (trečia bp už 5mC). (A) DNR atpažinimas MspJI. Kilpa-6C užima panašią poziciją tiek apo, tiek struktūroje su DNR ir nesudaro specifinių kontaktų. Q33, E65 ir K173 iš „2B“, „B3“ ir „78“ kilpų yra arti DNR bazių. (B ir C) LpnPI ir AspBHI modeliai su DNR, remiantis UHRFI-SRA-DNR struktūra. Kilpos-C6 ir kilpos-B3 ar. 41-43, 91-93 ir 99 skiriasi LpnPI ir AspBHI. LpnPI kilpos-2B ir kilpos-78 27–29 ir 136–137 ar. buvo išmutuotos šiame tyrime; AspBHI kilpos-2B T25 ir D32 ar. yra kritinės baltymo funkcijai (Horton et al., 2014a); AspBHI kilpos-78 R132 persidengia su kritine LpnPI ar. R137. (D) DNR atpažinimas eukariotiniame UHRFI SRA domene. Kilpos ekvivalentiškos kilpai-78 ir kilpai-2B MspJI tipo REazėse nuspalvintos žalia ir violetine spalvomis.

2.2.4 DNR sekos atpažinimas kilpa-78

Kilpa-78 priartėja prie DNR 3′ pusėje nuo modifikuotos bazės (2.6 C pav.). MspJI kilpos K173 yra labiausiai tikėtinas kandidatas +3 bp atpažinimui, kur MspJI teikia pirmenybę purinui (Horton et al., 2014b). Lygiavertės kilpos LpnPI ir AspBHI yra trumpesnės atitinkamai 5 ir 8 ar. (Sasnauskas et al., 2015). Nepaisant to, mūsų dabartinis LpnPI-DNR komplekso modelis ir mutantų analizė (R137 pakeitus alaninu kilpoje-78 LpnPI tampa neaktyviu, 2.2 lentelė) rodo, kad ši kilpa kontaktuoja su DNR. Tačiau ar šie kontaktai yra tik su DNR karkasu, ar sąveika vyksta ir su bazėmis (pvz., +1 bp atskyrimas) lieka neaišku, nes LpnPI kilpos-78 pakeitimas AspBHI trumpesne kilpa (LpnPI 133G variantas) baltymą išaktyvuoja (Sasnauskas et al., 2015).

2.2.5 LpnPI sekos specifiškumas ir jo keitimas

Norėdami išanalizuoti LpnPI atpažįstamus taikinius, nustatėme LpnPI kirpimo greičius su įvairiais oligodupleksais, kurie skiriasi 1 arba 2 bazių poromis nuo REBASE pateikiamo taikinio gC(mC)TG (30-M, 1 lentelė). Mūsų rezultatai patvirtina, kad LpnPI optimali seka yra 5′-C(mC)DG-3′ (kirpimo greičio konstanta ~3×10^{-3} s^{-1}), tačiau LpnPI ~90, ~500 ir 1500 kartų lėčiau kirpo ir substratus su sekomis 5′-G(mC)TG-3′, 5′-C(mC)TC-3′ ir 5′-G(mC)TC-3′ (greičio konstantos ~3×10^{-5}, ~6×10^{-6} ir 2×10^{-6} s^{-1}, 2.7 A, B pav.). Taigi LpnPI patikslinta taikinio seka yra 5′-(C>>G)(mC)D(G>>C)-3′.

Kelių MspJI šeimos REazių SRA domenai turi labai panašias baltymo sekas (Sasnauskas et al., 2015) ir atpažįsta persidengiančius taikinius su LpnPI. Pavyzdžiui, LpnPI 5′-C(mC)DG-3′ taikinys persidengia su AspBHI (5′-YS(mC)NS-3′) bei SgrTI (5′-C(mC)DS-3′) ′) (Cohen-Karni et al., 2011). Norėdami ištirti LpnPI kilpos-6C ir kilpos-2B vaidmenis atpažįstant DNR

taikinius pabandėme pakeisti LpnPI kilpas atitinkamomis AspBHI ar SgrTI kilpomis ir sukonstravome tokius LpnPI variantus:

- LpnPI-91RLL, kuriame pakeitėme LpnPI kilpą-6C (91-93 ir 99 ar.) atitinkama kilpa AspBHI (2.7 A pav.); buvo tikėtasi, kad toks LpnPI variantas galėtų kirpti DNR substratus su 5′-YS-3′ dinukleotidu −1, −2 pozicijose
- LpnPI-27HTG, kuriame pakeitėme LpnPI kilpos-2B 27-29 ar. atitinkamomis SgrTI kilpos 34-36 ar. (2.7 B pav.); buvo tikėtasi, kad tai sumažins LpnPI selektyvumą +2 bazių porai sekoje 5′-C(mC)DG-3′ ir fermentas kirps SgrTI atpažinimo seką 5′-C(mC)DS-3′;
- LpnPI-27HTG-91RLL, kuris turėjo AspBHI kilpą-6C bei SgrTI kilpą-2B (2.7 C pav.); tikėtasi, kad toks dvigubai pakeistas variantas kirps 5′-SMDS-3′ seką.

2.2.6 DNR sekos atpažinimas kilpa-6C

LpnPI ir AspBHI modeliuose su DNR kilpa-6C pasiekia 5′ pusę nuo taikinio sekos iš mažojo griovio pusės (2.6 B, C pav.). Lygiavertė „CpG atpažinimo“ ir „NKR piršto“ kilpa eukariotiniuose SRA domenuose yra ilgesnė ir sudaro bazei specifinius kontaktus iš didžiojo griovio pusės (2.6 D pav.) (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). LpnPI-DNR modelyje tik konservatyvus R98 iš kilpos-6C yra H-ryšio atstumu nuo –1 bazės. Tai galėtų padėti tolimesnei kilpai-B3 užimti tinkamą poziciją. Šiam modeliui neprieštarauja tai, kad LpnPI mutantas R98A buvo beveik neaktyvus (2.2 lentelė).

Taip pat parodėme, kad pakeitus LpnPI kilpą-6C atitinkama AspBHI kilpa (keturių ar. mutacijos 91-93 ir 99 pozicijose), kerpamas 5′-G(mC)DG-3′ substratas su G baze −1 pozicijoje ir pagerėja taikinių su A ir T bazėmis −1 pozicijoje kirpimas, taip pat pastebimai pagerėja DNR kirpimas su C +2 pozicijoje (2.7 A pav.). Toks specifiškumo pasikeitimas neatsiranda pakeitus pavienes kilpos-6C ar., tai reiškia, kad už DNR atpažinimą yra atsakingos kelios ar.

Įdomu tai, kad trys iš keturių ar., pakeistų LpnPI-91RLL variante (91-93), LpnPI / AspBHI modelyje su DNR yra nusisukusios nuo DNR (2.6 B, C pav.). Kad susidarytų tiesioginiai kontaktai su DNR bazėmis kilpa-6C turėtų keisti konformaciją, panašiai kaip UHRF1-SRA (2.6 D pav.). Tačiau MspJI baltyme kilpa-6C atsiduria toje pačioje pozicijoje tiek apo, tiek struktūroje su DNR (2.6 A pav.) (Horton et al., 2012, 2014b).

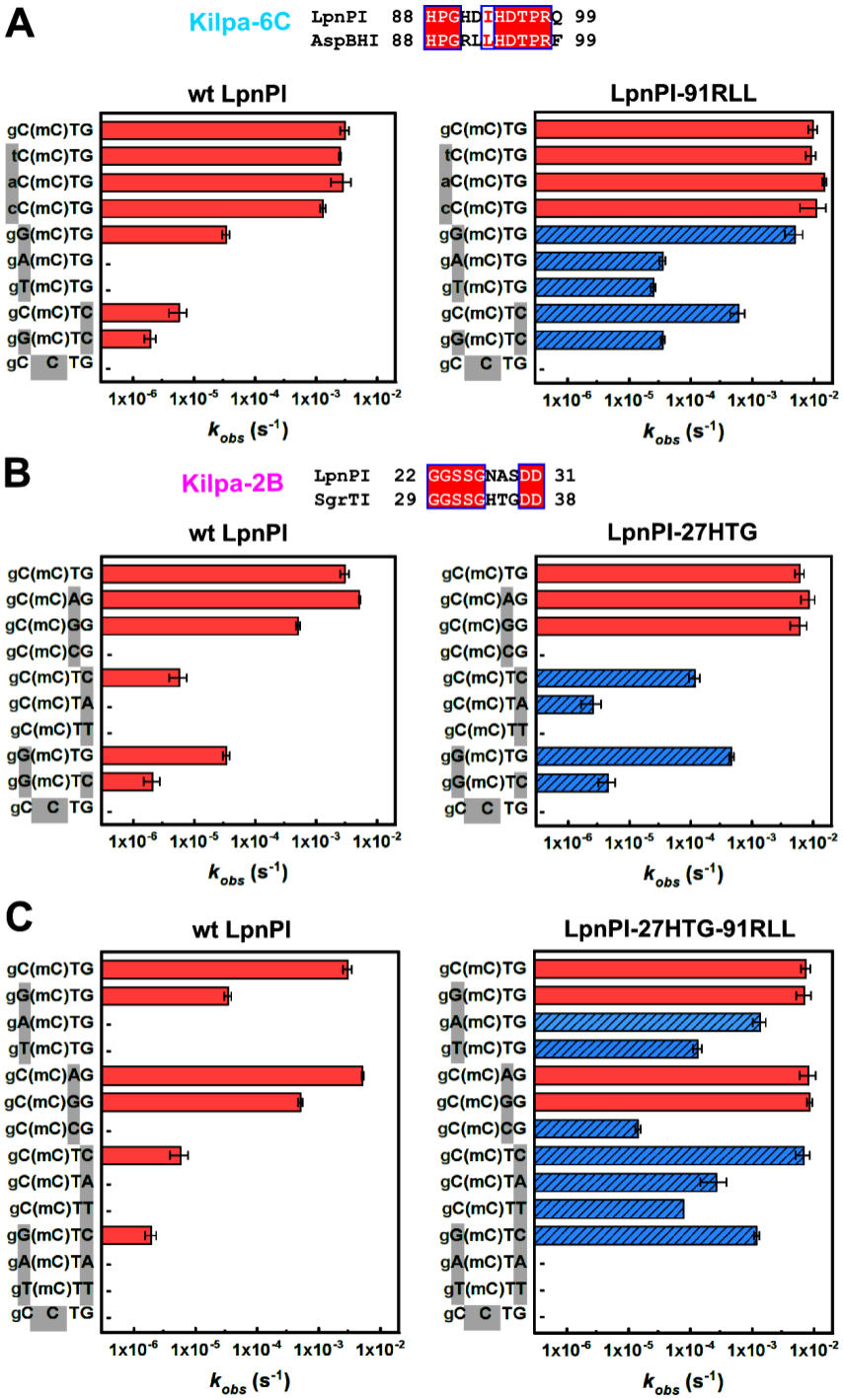
Kilpos-6C ar. 91-93 ir 99 vaidmuo −1 bp atpažinime lieka neaiškus: kai kurios kilpos ar. galėtų sąveikauti su DNR tiesiogiai, bet būtų reikalingi

LpnPI / AspBHI specifiniai konformaciniai pokyčiai; arba kilpa-6C galėtų netiesiogiai dalyvauti sekos atpažinime per sąveiką su kitomis ar., kurios sudaro tiesioginius kontaktus su DNR. Netiesioginis kilpos-6C vaidmuo paaiškintų relaksuotą substratų su pakeistomis −1 ir +2 pozicijomis atpažinimą. Kitas svarbus faktas yra tai, kad LpnPI-91RLL, priešingai nei donorinis baltymas AspBHI, neturėjo −2 bp specifiškumo (2.7 A pav.). Tikriausiai šios bp atpažinime AspBHI atsakingas kitas struktūrinis elementas, labiausiai tikėtina, kad kilpa-B3 (žiūrėti aukščiau).

2.2.7 DNR sekos atpažinimas kilpa-2B

Kilpa-2B atsiduria mažajame griovyje 3′ pusėje nuo modifikuotos bazės (2.6 A, C pav.). Norėdami nustatyti šios kilpos vaidmenį LpnPI funkcijai alaninu pakeitėme polines ar. S25, N27 ir D30. Tik N27A mutantas prarado aktyvumą, tuo tarpu kitų dviejų mutantų aktyvumas buvo toks pat kaip ir wt LpnPI (2.2 lentelė). Tai parodo, kad N27 vaidina svarbų vaidmenį DNR surišime / atpažinime.

Šios kilpos pakeitimas LpnPI į atitinkamą SgrTI kilpą (LpnPI-27HTG) relaksuoja +2 pozicijos specifiškumą, pagreitėja 5′-C(mC)DC-3′ ir 5′-C(mC)DA-3′ taikinių kirpimas (2.7 B pav.). Įdomu tai, kad padidėjo ir 5′-G(mC)DG-3′ DNR, kurioje pakeista −1 pozicija, kirpimo greitis. Tai rodo, kad kilpos-2B pakeitimas bendrai pagerina baltymo gimiringumą metilintai DNR. Dvi iš trijų ar., kurios LpnPI pakeistos į 27HTG variantą, N27 ir S29, apo-LpnPI ir LpnPI-DNR modelyje yra nusisukę nuo DNR ir atsiduria arčiau +1 nei +2 bp (2.6 B pav.). Tikriausiai jungiantis su DNR kilpa-2B patiria konformacinius pokyčius, kurie priartina šias ar. arčiau +2 bp. AspBHI kilpa-2B yra ilgesnė 1 ar. ir kitokios konformacijos (2.5 B, 2.6 C pav.); jos svarba baltymo funkcijai buvo patvirtinta mutageneze (R25A ir D32A mutacijos panaikina AspBHI aktyvumą (Horton et al., 2014a)). Tuo tarpu glutaminas Q33 iš MspJI kilpos-2B sąveikauja su DNR bazėmis 3′ pusėje nuo išsukto citozino (2.6 D pav.) ir yra nebūtinas MspJI aktyvumui (Horton et al., 2014b). Tai neprieštarauja, kad MspJI neturi specifiškumo +1 ir +2 bp. Lygiavertė kilpa eukariotiniuose SRA domenuose yra daug trumpesnė ir nedalyvauja specifiniuose kontaktuose su DNR (2.6 D pav.).



2.7 pav. LpnPI wt ir kilpų mutantų atpažinimo sekos ir jų atpažinimo pokyčiai. DNR kirpimo reakcijos atliktos standartinėmis sąlygomis, nustatytos greičio konstantos k_{obs} gautos iš eksponentinės priklausomybės kreivių. DNR atpažinimo

sekos nurodytos kairėje grafikų pusėje. Paskutinis substratas – nemetilinta kontrolė; sekų pozicijos, kurios skiriasi nuo gC(mC)TG substrato, yra nuspalvintos pilkai. LpnPI mutantų reakcijos greičiai, kurie padidėjo dėl kilpos pakeitimo, pažymėti dryžuotais stulpeliais; „–“ žymi nenustatomą greitį (mažesnį nei 3×10^{-7} s^{-1}, pradinė x ašies pozicija). Virš A ir B dalių parodyti LpnPI, AspBHI kilpų-6C ir LpnPI, SgrTI kilpų-2B, kurios buvo pakeistos LpnPI-91RLL ir LpnPI-27HTG, palyginiai. (A) Wt LpnPI ir LpnPI-91RLL variantas (kilpos-6C pakeitimas) su DNR substratais, turinčiais skirtingas sekas prieš ir už 5mC. (B) Wt LpnPI ir LpnPI-27HTG (kilpos-2B pakeitimas) su įvairiais DNR substratais. (C) Wt LpnPI ir LpnPI-27HTG-91RLL (dvigubas kilpų pakeitimas) su įvairiais substratais.

2.2.8 LpnPI modulinė sekos atpažinimo sistema

Apibendrinant, mes parodėme, kad LpnPI atpažįsta sekos kontekstą aplink išsuktą citoziną per keletą paviršinių kilpų, kurios veikia kaip atskiri DNR surišimo / atpažinimo moduliai. LpnPI yra perspektyvi modulinė sistema modifikuotą DNR atpažįstančių REazių specifiškumo keitimui, nes gali būti lanksčiai keičiamas atpažinimo taikinys, panašiai kaip introninėse (angl. *homing*) endonukleazėse (Chevalier ir Stoddard, 2001). Nors wt LpnPI didžiausias aktyvumas yra su kanoniniu taikiniu 5′-C(mC)DG-3′, tačiau baltymas taip pat kerpa alternatyvias sekas 5′-G(mC)DG-3′ ir 5′-C(mC)DC-3′, nors ir lėčiau (2.7 A, B pav.). Ryškiausiai alternatyvių atpažinimo taikinių karpymo pagerėjimas matomas dvigubai pakeistame LpnPI variante, kuris turėjo kilpą-2B iš AspBHI ir kilpą-6C iš SgrTI: šio LpnPI varianto atveju atpažinimo seka sutrumpėjo iki 5′-S(mC)(D>>C)-3′ ar 5′-(mC)(D>>C)S-3′ kombinacijos, taip pat buvo kerpamas 5′-G(mC)DC-3′ taikinys, kuris skiriasi nuo kanoninio atpažinimo taikinio 5′-C(mC)DG-3′ dvejomis bp (2.7 C pav.).

Relaksuotas sekos atpažinimas yra būdinga MspJI šeimos baltymų savybė. Vis dėlto, nepaisant įvairaus specifiškumo taikiniui aplink modifikuotą citoziną, SRA domenai yra stebėtinai geri moduliai modifikuoto citozino DNR atpažinimui: nei wt LpnPI, nei pakeistomis kilpomis ar mutantiniai variantai nerodė jokio aktyvumo su nemetilinta DNR. MspJI šeimos REazių SRA domenams būdingas 5mC konteksto atpažinimo lankstumas ir gebėjimas griežtai atskirti metilintą nuo nemetilintos DNR atveria kelią fermentų, atpažįstančių 5mC bet kokiame sekos kontekste, inžinerijai. Toks 5mC specifinis fermentas būtų naudingas įrankis genomų metilinimo tyrimuose. Reikšmingas LpnPI specifiškumo atpalaidavimas yra žingsnis link šio tikslo.

2.3 Įvairių citozino variantų atpažinimas McrBC DNR surišimo domene

McrBC kompleksas atpažįsta 5mC, 5hmC ir 4mC turinčias sekas ir kerpa DNR įvairiose vietose nuo atpažinimo sekos. Kristalinė McrBC DNR atpažinimo domeno McrB-N struktūra su metilinta DNR (Sukackaite et al., 2012) parodė, kad nepaisant to, kad jo tretinė struktūra nėra panaši į SRA domenų, McrB-N taip pat išsuka modifikuotą bazę bei patalpina ją į baltymo kišenę. Bazės išsukimas taip pat buvo patvirtintas tirpale panaudojant pyC (Sukackaite et al., 2012). REazė McrBC yra naudojama kaip jautrus įrankis nustatyti metilintus citozinus DNR (Lippman et al., 2004, 2005). Tačiau kaip McrB-N atpažįsta kitas modifikacijas, tokias kaip 5hmC ir 4mC, bet ir 5fC bei 5caC, iki šiol nebuvo parodyta. Taigi, mes atlikome struktūrinius ir biocheminius McrB-N tyrimus, norėdami nustatyti bei keisti gimiringumą įvairioms citozino modifikacijoms.

2.3.1 Naujos McrB-N struktūros

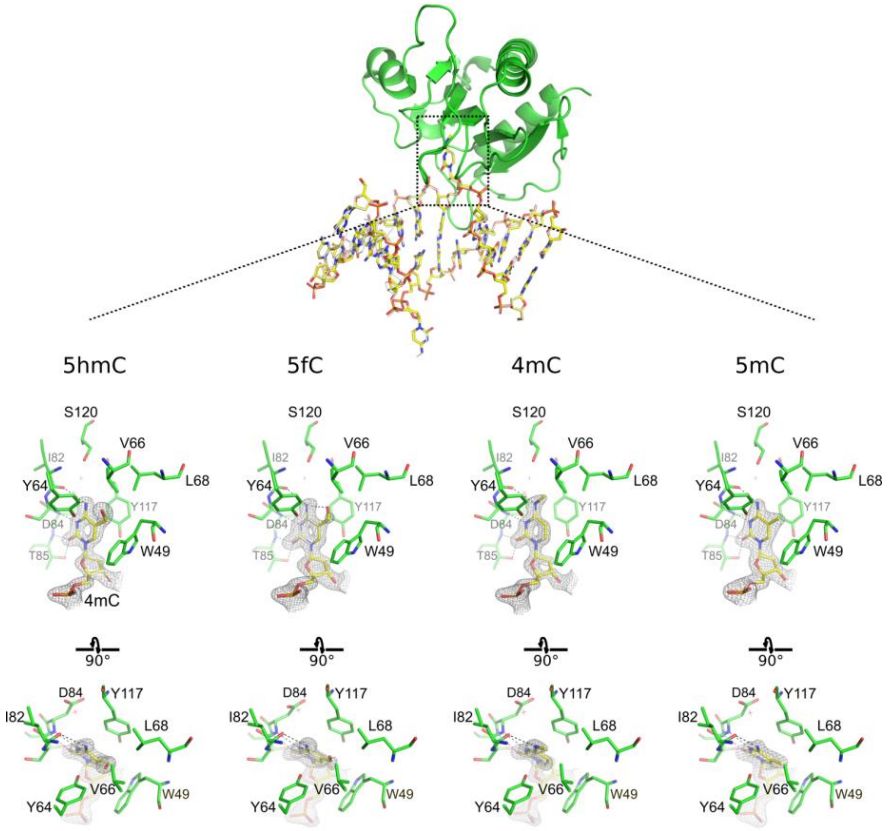
McrB-N struktūros su DNR oligodupleksais, turinčiais 5hmC, 5fC ir 4mC modifikacijas, buvo išspręstos 1,55–1,8 Å skiriamąja geba. Parodyta, kad McrB-N išsuka 5hmC, 5fC ir 4mC bazes iš dgDNR panašiai kaip ir 5mC (Sukackaite et al., 2012) (2.8 pav.). Visose struktūrose išsukta bazė atsiduria toje pačioje vietoje baltymo kišenėje, formuoja H-ryšius tarp bazės O2, N3 ir N4 atomų ir baltymo pagrindinės grandinės ir T85, D84 ir I82 ar. šoninių grandinių atomų (2.8 pav.).

5hmC ir 5fC 5-padėtyje esančios grupės sudaro vdW kontaktus su W49, Y64, L68 ir Y117. Tuo tarpu 4mC N4-metilo grupė taip pat sudaro vdW kontaktus su V66 ir S120, bet ne W49 (2.8 pav.). Hidroksilo grupės deguonis iš 5hmC nėra aromatinio žiedo plokštumoje ir sudaro vdW kontaktus su V66 ir Y64 hidroksilo grupe (H-ryšio su Y64 negalėtų sudaryti dėl nepalankios geometrijos). Tuo tarpu formilo grupės deguonis iš 5fC atsiduria aromatinio žiedo plokštumoje ir sudaro vidinį H-ryšį su 5fC 4-amino grupe bei formuoja vdW kontaktus su Y117, L68, Y64 ir V66 (2.8 pav.).

2.3.2 WT McrB-N ir kišenės mutantų selektyvumas modifikuotiems citozino variantams

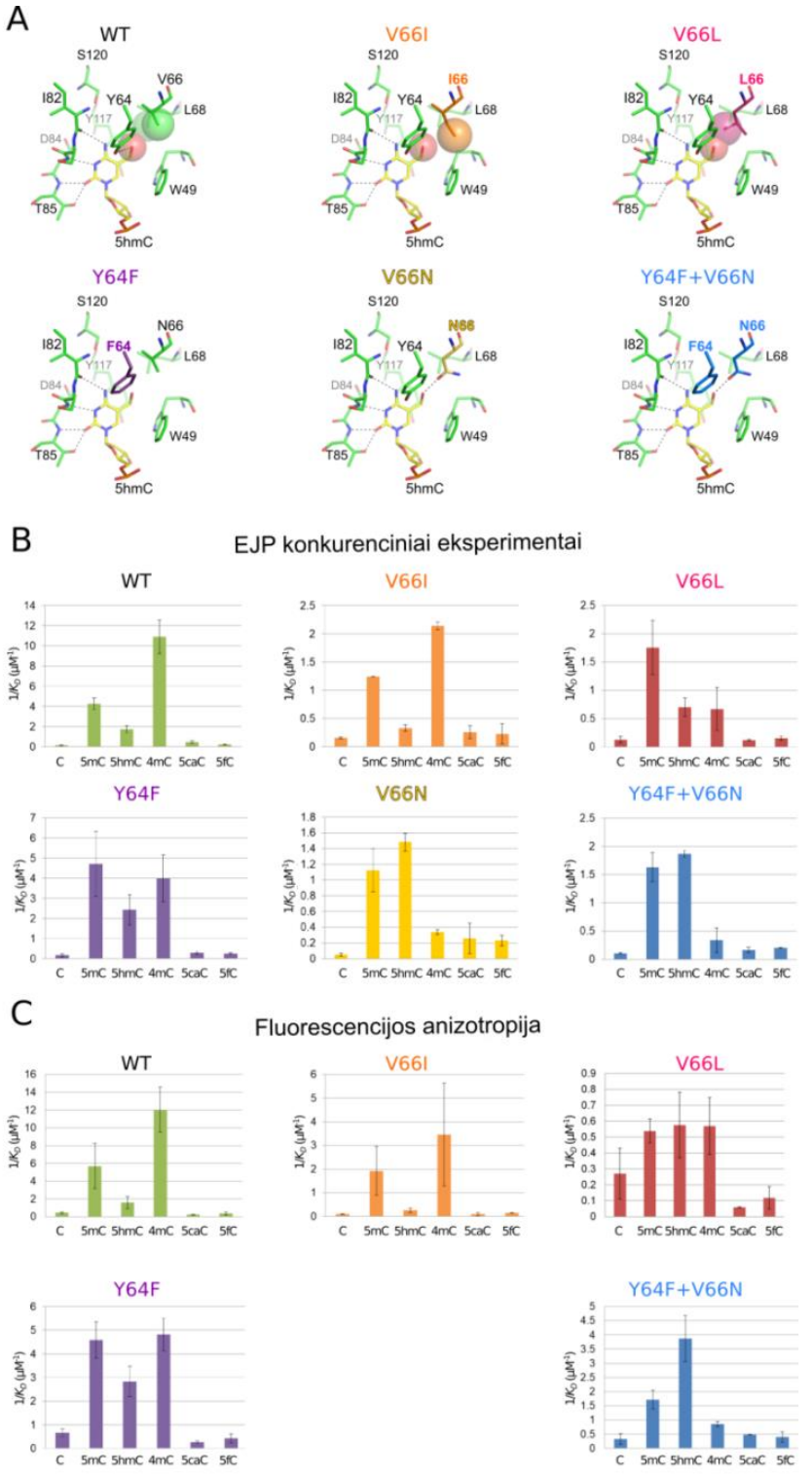
Nuodugniai buvo ištirtas WT McrB-N ir mutantų Y64F, V66I, V66L ir V66N DNR surišimo gimiringumas ir santykinis specifiškumas įvairioms citozino modifikacijoms naudojant EJP eksperimentais paremtus įvairias

citozino modifikacijas turinčių substratų konkurencinius tyrimus (2.9 B pav.) ir fluorescencijos anizotropijos matavimus (2.9 C pav.). Kadangi V66N mutanto fluorescencijos anizotropijos matavimuose signalas nebuvo stebėtas, dar charakterizavome ir dvigubą mutantą Y64F_V66N.



2.8 pav. McrB-N struktūros su DNR, turinčia 5hmC, 5fC, 4mC ir 5mC. Viršuje: McrB-N struktūra su 4mC DNR (PDB ID 6gcf). Parodytas tik vienas baltymo subvienetas (grandinė A); DNR pavaizduota pagaliukais (geltona). Apačia: išsuktos bazės surišimo kišenės struktūros su 5hmC, 5fC ir 4mC DNR (PDB ID 6gcd, 6gce ir 6gcf); taip pat parodyta struktūra su 5mC DNR (PDB ID 3ssc, (Sukackaite et al., 2012)). Išsuktiem nt parodytas elektronų tankio žemėlapis 2.0 σ (5hmC, 4mC) ir 1.3 σ (5fC, 5mC).

EJP eksperimentai parodė, kad WT McrB-N teikia pirmenybę 4mC (disociacijos konstanta $K_D \approx$ 0,1 μM, $1/K_D \approx 10$ μM^{-1}, 2.9 B pav.). Susirišimas su 5mC ar 5hmC yra 2,5-5 karto silpnesnis (K_Ds apytiksliai 0,25 ir 0,5 μM, atitinkamai), o su 5fC ir 5caC - nereikšmingas ($K_D >$ 4 μM, $1/K_D <$ 0,25 μM^{-1}). Kokybiškai ir kiekybiškai panašūs rezultatai buvo gauti ir anizotropijos matavimuose (2.9 C pav.).


A
WT
V66I
V66L
Y64F
V66N
Y64F+V66N
S120
I82
Y64
Y117
D84
T85
W49
L68
5hmC
V66
I66
L66
F64
N66
B
EJP konkurenciniai eksperimentai
C
Fluorescencijos anizotropija
C
5mC
5hmC
4mC
5caC
5fC

2.9 pav. McrB-N kišenės mutantų selektyvumas. (A) McrB-N mutantų modeliai su išsuktu 5hmC. Eksperimentiškai nustatytos McrB-N mutantų DNR su įvairiomis citozino modifikacijomis surišimo konstantos (μM^{-1}) (B) nustatytos EJP eksperimentais bei (C) fluorescencijos anizotropijos matavimais.

Mutacija V66I, kuri sumažina išsukamos bazės susirišimo kišenę, sumažino McrB-N gimiringumą 5mC, 5hmC ir 4mC modifikacijoms panašiu dydžiu (apytiksliai 4 kartus sumažėjo $1/K_D$ vertės, 2.9 B, C pav.), palikdama santykinį specifiškumą skirtingiems citozino variantams nepakitusį. Panašus gimiringumo sumažėjimas be santykinio specifiškumo pokyčio buvo stebėtas ir su V66L mutantu. Mutacija Y64F, kuri pašalina vieną hidroksilo grupę iš kišenės, neturėjo poveikio McrB-N sąveikai su 5mC ir 5hmC DNR, tačiau 3 kartus sumažino gimiringumą 4mC DNR (2.9 B, C pav.). Dar ryškesnis gimiringumo sumažėjimas 4mC buvo stebėtas su V66N ir Y64F_V66N mutantais (susirišimas su 4mC, $K_D \approx$ 2-4 μM, yra palyginamas su 5fC bei 5caC, 2.9 B, C pav.). V66N mutacija turėjo nedidelį neigiamą poveikį 5mC DNR surišimui (K_D padidėjo maždaug 2,5 karto iki ≈0,6 μM, 2.9 B pav.), tačiau turėjo teigiamą poveikį 5hmC surišimui (0,4-0,6 μM K_Ds, kaip matome abiejuose eksperimentuose, yra arba nepakitusi, arba 1,5 karto mažesnė lyginant su WT McrB-N, 2.9 B pav.). Tad V66N ir Y64F_V66N McrB-N mutantai teikia pirmenybę 5hmC turinčiai DNR.

Apibendrinant, parodėme, kad McrB-N santykinis specifiškumas DNR, turinčiai skirtingas citozino modifikacijas, gali būti keičiamas keičiant kišenės sienelių ar. Didžiausią efektą sukėlė V66N mutacija, kuri panaikino McrB-N susirišimą su 4mC DNR ir turėjo teigiamą efektą 5hmC DNR surišimui (2.9 pav.). Šį efektą tikriausiai sukėlė tai, kad 4mC N4-metilo grupė negali formuoti energetiškai palankių sąveikų su N66 asparaginu; tuo tarpu 5hmC galėtų formuoti papildomą H-ryšį (2.9 A pav.).

2.3.3 5(h)mC bazę išsukančių baltymų selektyvumas

Stebėti McrB-N K_D verčių DNR su įvairiais citozino dariniais santykiai ir jų pokyčiai dėl mutacijų neviršija 40 kartų (2.9 pav.). Tačiau net toks palyginti nedidelis skirtumas yra didesnis nei žinomų eukariotinių SRA domenų atveju. Pavyzdžiui, UHRF1-SRA 5mC turinčią DNR riša tik ~3,5 karto stipriau nei nemodifikuotą DNR (Zhou et al., 2014), o UHRF2 (5hmC atpažįstantis eukariotinis baltymas (Spruijt et al., 2013)), DNR su 5hmC suriša tik ~1,5 karto geriau nei DNR su C (Zhou et al., 2014)). Geresniu gebėjimu skirti nemodifikuotą ir modifikuotą citozinus, remiantis aktyvumo matavimais, turėtų pasižymėti SRA tipo domenus turinčios prokariotinės metilintą DNR atpažįstančios MspJI ir PvuRts1I šeimų REazės (Wang et al.,

2011; Zheng et al., 2010). Pvz., MspJI tipo baltymai, tame tarpe aukščiau mūsų darbe tirtas LpnPI, kerpa DNR, turinčią 5mC ir 5hmC modifikacijas specifiniame sekos kontekste, ir, kaip ir McrBC kompleksas, yra neaktyvūs su nemodifikuota DNR, tuo tarpu PvuRts1I tipo baltymai 250 ir daugiau kartų gimininges ni 5hmC ir 5ghmC lyginant su 5mC ir nemodifikuotu C (Wang et al., 2011). Šiuo metu žinoma MspJI struktūra su 5mC turinčia DNR (Horton et al., 2014b) ir kelios PvuRts1I šeimos baltymų apo struktūros (Horton et al., 2014c; Kazrani et al., 2014) negali paaiškinti didesnio jų SRA domenų selektyvumo lyginant su eukariotiniais SRA domenais.

Nepaisant identiškų vietų, kuriose atsiduria išsukta bazė McrB-N struktūrose su DNR, turinčia įvairias citozino modifikacijas, McrB-N turi didelį gimininguma 4mC, 5mC ir 5hmC, bet mažą – 5fC. Silpna sąveika su 5fC, kuris yra beveik izosteriškas 5hmC, rodo, kad McrB-N, o tuo pačiu ir kitų citoziną išsukančių baltymų, selektyvumas įvairiems citozino variantams yra lemiamas ne tik vidinių kišenės kontaktų, bet ir kitų veiksnių, tokių kaip išsuktos bazių poros stabilumo bei išsuktos bazės / baltymo kišenės dehidratacijos energijų. Žemas McrB-N gimininguma s 5fC ir 5caC, bei aukštas gimininguma s 5mC ir 5hmC taip pat leidžia teigti, kad metilomos profilių tyrimuose, kuriuose kaip įrankis DNR fragmentavimui buvo naudotas McrBC (Burman et al., 1999; Chotai ir Payne, 1998; Lippman et al., 2004, 2005; Lyko et al., 1999), buvo nustatytas 5mC ir 5hmC pasiskirstymas, o 5fC ar 5caC registruoti nebuvo. Lieka atviras klausimas, ar McrB-N galėtų būti pritaikytas selektyvesniam individualių citozino variantų atpažinimui.

2.4 McrA aktyvumas ir struktūra

E. coli K kamienų fermentas McrA (taip pat žinomas kaip EcoKMcrA) riboja modifikuotos DNR patekimą į ląsteles tik tuomet, kai metilinimas yra atitinkamame sekos kontekste. Fermentas efektyviai riboja DNR, kuri buvo metilinta M.HpaII sekoje C5mCGG (Card et al., 1990). Iki šiol nebuvo nustatytas McrA aktyvumas *in vitro,* netgi naudojant plazmides, kurios yra ribojamos ląstelėse (pvz., plazmidę su M.HpaII) (Mulligan ir Dunn, 2008). Tačiau buvo parodyta, kad McrA rišasi specifiškai su metilinta DNR. EJP eksperimentais nustatytas McrA specifiškumas yra Y5mCGR (Mulligan et al., 2010). Kaip McrA atpažįsta seką ir modifikaciją, dar nėra žinoma. Sekos analizė parodė, kad McrA sudaryta iš N-galinio domeno, kurio seka neturi panašumų su žinomais baltymais, bei C-galinio HNH nukleazinio domeno (Bujnicki et al., 2000). Tad McrA gali būti naujos didesnės fermentų šeimos

prototipas. Bioinformatinė analizė randa baltymų iš eubakterijų ir archėjų panašių į McrA. Tačiau dauguma atvejų panašumas apsiriboja tik HNH domenu. Bendradarbiaudami su M. Bochtler grupe Varšuvoje, atlikome struktūrinius (Czapinska et al., 2018) ir biocheminius McrA tyrimus, tame tarpe pirmą kartą pademonstravome šio fermento *in vitro* aktyvumą.

2.4.1 McrA aktyvumas *in vitro*

Mes ištyrėme katalizinį McrA aktyvumą naudodami 30 bp ilgio oligodupleksus su nemodifikuota, pusiau ir pilnai modifikuota C5mCGG seka (1 lentelė). Pirmiausia DNR kirpimo reakcijos buvo atliktos su nemodifikuotu ir pusiau modifikuotu substratu buferyje, papildytame įvairiais kiekiais dvivalenčių metalo jonų. DNR karpymas stebėtas tik su Mn^{2+}, didžiausi greičiai buvo su 0.01-0.2 mM Mn^{2+} (Czapinska et al., 2018).

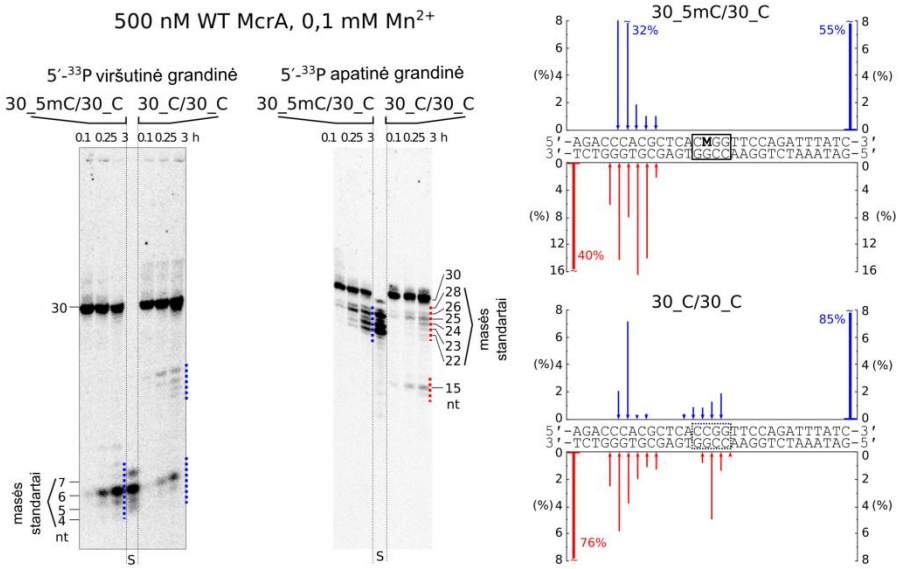
Stipresnis karpymas stebėtas su metilintais substratais, tačiau McrA lėtai karpė ir nemetilintą DNR (2.10 pav.). Visi oligodupleksai buvo kerpami keletoje pozicijų 5-15 nt nuo C5mCGG sekos. Nemetilinta DNR buvo kerpama tose pačiose vietose su papildoma kirpimo vieta nemodifikuotame CCGG taikinyje (2.10 pav.). McrA taip pat veikė kaip nuo metilinimo nepriklausoma 3′-5′ egzonukleazė ant vg oligonukleotido (2.10 pav.).

McrA aktyvumas stebėtas ne tik naudojant trumpus DNR oligodupleksus, bet ir superspiralizuotas plazmides bei fago λ DNR (Czapinska et al., 2018). Superspiralizuotos nemetilintos ir metilintos plazmidės buvo pilnai sukarpytos po 2 h esant 0,5 µM baltymo.

2.4.2 McrA struktūra

McrA K196E varianto kristalinė struktūra buvo išspręsta be DNR 2,85 Å skiriamąja geba (prof. M. Bochtler grupė, Varšuva). Struktūra patvirtino spėjamą dviejų domenų baltymo architektūrą su prieš tai necharakterizuotu N-galiniu domenu ir HNH C-galiniu domenu (Bujnicki et al., 2000) (2.12 A pav.). C-galinis HNH domenas (147-276 ar.) su N-galiniu domenu (1-143 ar.) jungiasi per lankstų jungtuką (144-146 ar.) (2.11 pav.).

McrA sudaro homodimerinę struktūrą. Kadangi dimerizacija vyksta tik per C-galinius domenus, galima spėti, kad atskiras N-galinis domenas turėtų būti monomeras. Tai buvo patvirtinta gelfiltracijos metodu su McrA N-galinio domeno fragmentu (1-174 ar.). Kaip kontrolė naudotas pilno ilgio McrA elgėsi kaip dimeras (gauta masė 54,4 kDa, teorinė dimero masė 64,5 kDa), o N-galinis fragmento eliucijos tūris atitiko monomerą (teorinė masė su His_6-inkaru 20,6 kDa, nustatyta masė – 20,1 kDa) (Czapinska et al., 2018).



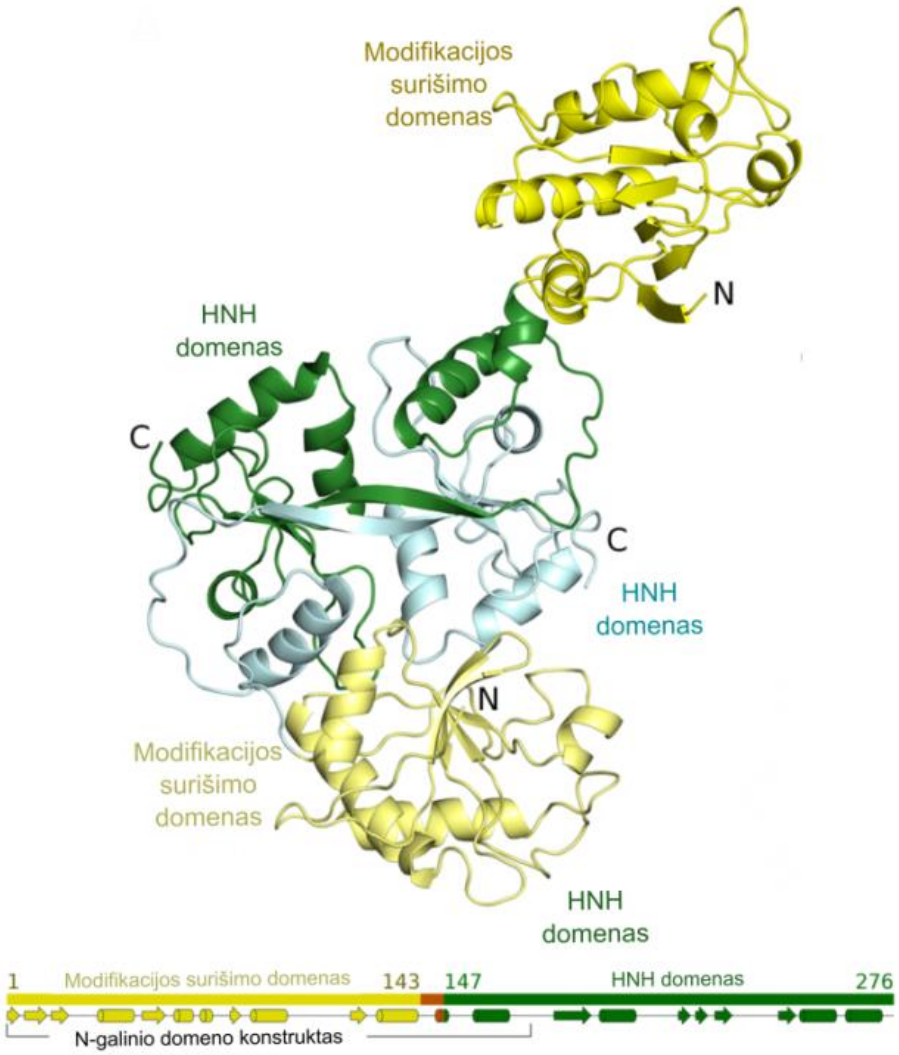
2.10 pav. DNR karpymas McrA. Reakcijos atliktos su 0,5 µM wt McrA (dimeras) ir 0,2 µM nemodifikuotu, pusiau ir pilnai metilintu oligodupleksu ir vgDNR 37 °C buferyje su 0,1 mM Mn^{2+}. Viršutinės ir apatinės grandinės kirpimo pozicijos geliuose pažymėtos mėlyna ir raudona punktyrinėmis linijomis. Takelyje „S“ yra radioaktyviai žymėti vg oligonukleotidai, kurių dydis atitinka produktų dydį (nt dydžiai parašyti šalia gelių nuotraukų). Viršutinės ir apatinės grandinių kirpimo produktų kiekiai po 3 h karpymo reakcijos yra parodyti mėlyna / raudona rodyklėmis išilgai oligodupleksų sekos dešinėje pusėje.

2.4.3 McrA N-galinio domeno specifiškumas modifikuotą citoziną turinčiai DNR

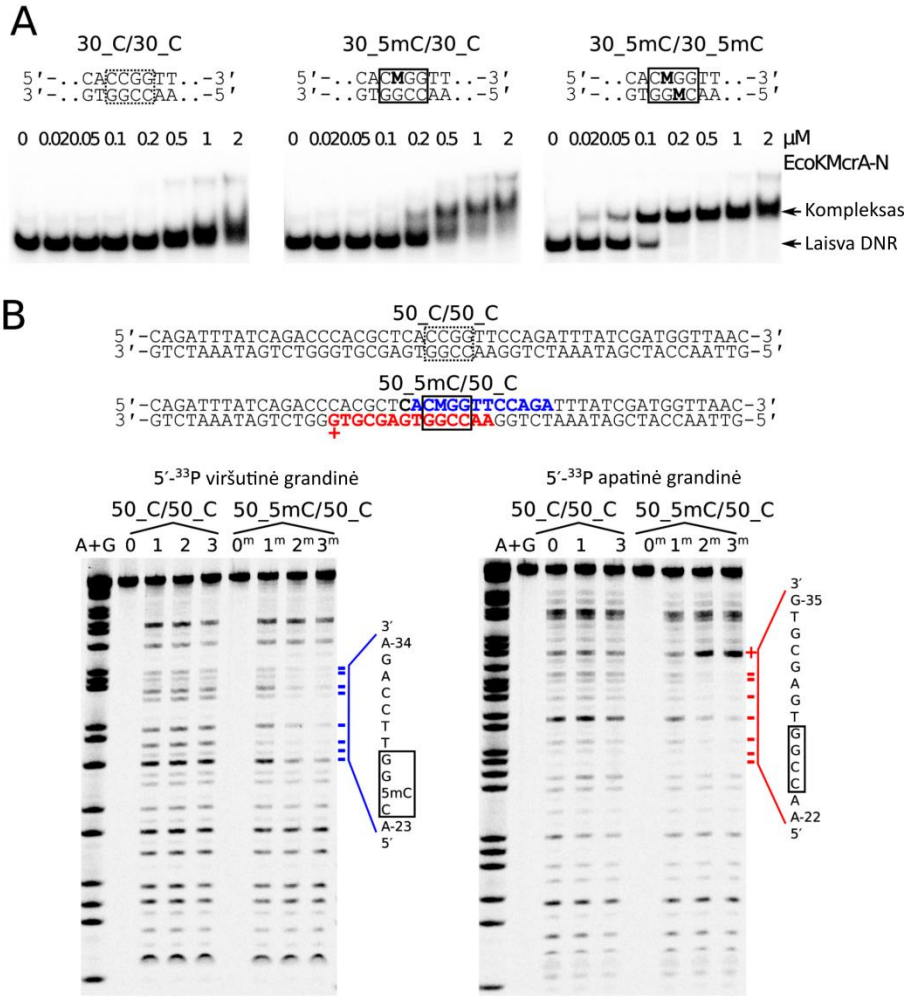
McrA domeninė architektūra primena kitas metilintą DNR atpažįstančias REazes, kurios yra sudarytos iš nukleazinio ir modifikaciją atpažįstančio domenų. Tad McrA N-galinis domenas yra pagrindinis kandidatas metilintos DNR atpažinimui. Atlikome EJP eksperimentus panaudodami N-galinį fragmentą. Duomenys rodo, kad susirišimas su metilinta DNR yra stipresnis lyginant su nemodifikuota DNR. Tai patvirtina, kad McrA-N yra metilinimui specifiškas domenas (2.12 A pav.).

Metilinimui specifinis susirišimas taip pat buvo patvirtintas tiriant DNR apsaugojimą nuo DNazės I (angl. „*DNAse I footprinting*“). McrA-N apsaugo nuo DNazės I karpymo tik metilintą DNR (2.12 B pav.). Apsaugotas regionas aplink 5mC buvo identiškas regionui, kurį apsaugojo pilno ilgio McrA, tačiau pilno ilgio baltymas papildomai apsaugojo ir kai kurias nemetilintos DNR pozicijas (duomenys neparodyti). Tai rodo, kad McrA-N

turi didesnį bendrą selektyvumą metilintiems taikiniams nei pilno ilgio baltymas. Nuo metilo nepriklausomas DNR susirišimas stebimas su McrA tikriausiai yra dėl C-galinio HNH domeno nespecifinės sąveikos su DNR.



2.11 pav. McrA struktūra. (A) McrA dimeras asimetriniame kristalo vienete (viršuje) ir domeninė sandara (apačioje), skirtingomis spalvomis nuspalvinti domenai ir linkeris.



2.12 pav. McrA-N sąveika su DNR. (A) EJP eksperimentai. DNR (50 nM) nemodifikuota, pusiau ir pilnai metilinta (centrinė seka parodyta viršuje, „M“ – 5mC). McrA-N koncentracijos nurodytos virš gelių takelių. Rodyklėmis parodytos laisvos DNR ir baltymo-DNR kompleksų pozicijos. (B) Apsaugojimo nuo DNazės I eksperimentas. Viršuje: DNR substrato seka. Paryškintas DNR regionas, kurį McrA-N apsaugo nuo DNazės I karpymo. „+“ pažymėta apatinės grandinės G-35 pozicija, kuri tampa prieinamesnė DNazei I esant McrA-N. Apačioje: DNR apsaugojimas McrA-N. Takelyje „0“ buvo nemodifikuota DNR nepaveikta DNaze I, „1“ – DNR paveikta DNaze I, „2“ ir „3“ – DNR paveikta DNaze I esant 0,5 μM ir 1 μM McrA-N. Takeliai „0m“, „1m“, „2m“ ir „3m“ analogiški prieš tai aprašytiems, tik eksperimentai buvo atlikti su pusiau metilinta DNR. Dydžio standartas (A+G) buvo gautas standartine Maxam-Gilbert sekoskaitos reakcija. Pozicijos, apsaugotos nuo DNazės I karpymo esant McrA-N, pažymėtos mėlynai (viršutinė grandinė) ir raudonai (apatinė). Apsaugotos sritys parodytos dešinėje.

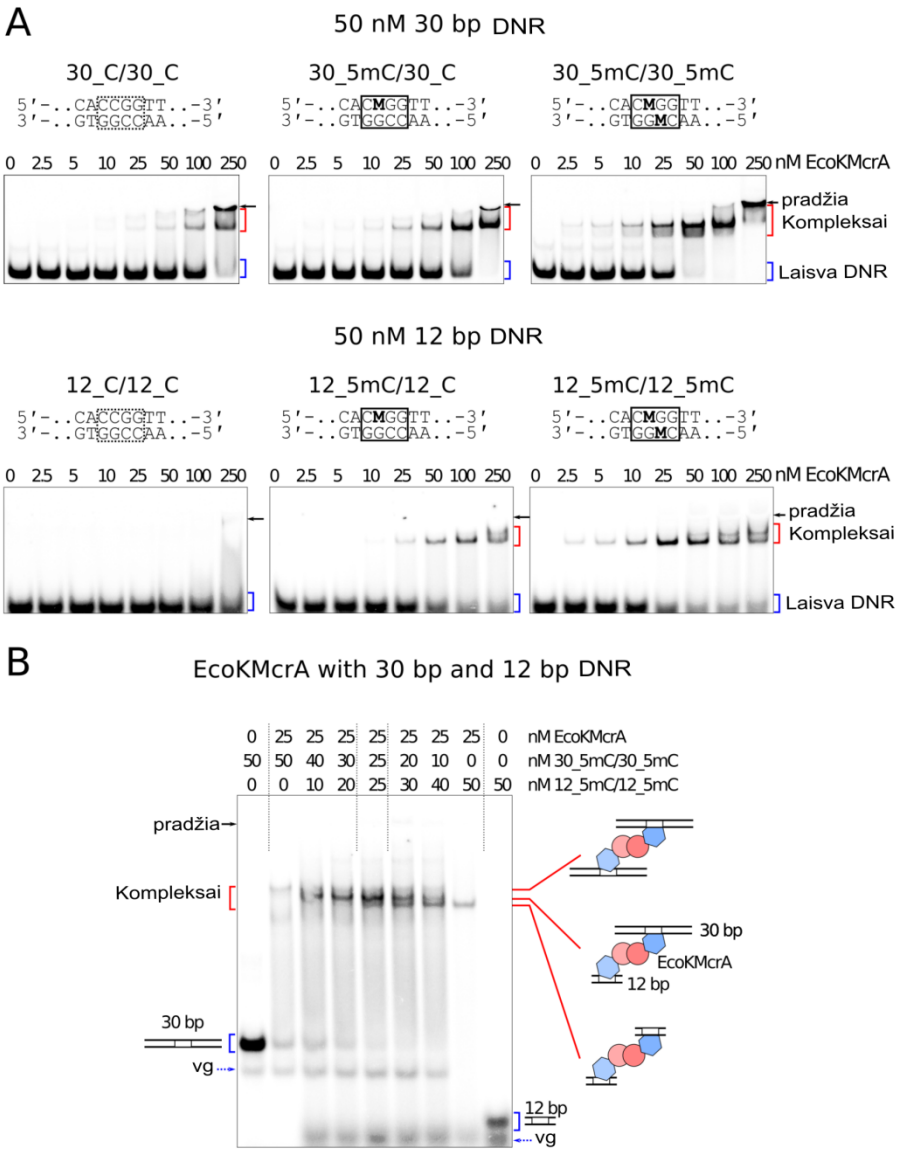
2.4.4 McrA bazės išsukimo tyrimas naudojant fluorescentinį metodą su pyC

Norėdami patikrinti, ar McrA išsuka bazę, panaudojome fluorescentinį metodą su pyC. McrA rišasi su pusiau modifikuotu 12 bp oligodupleksu, turinčiu pyC, stipriau nei su nemodifikuota kontrole, tačiau silpniau nei su dupleksu, turinčiu 5mC (Czapinska et al., 2018). Tačiau pyC turintis substratas turėjo panašias fluorescentines savybes tiek nesant baltymo, tiek su pilno ilgio McrA ir N-galiniu fragmentu (Czapinska et al., 2018). Neigiamas rezultatas galėtų reikšti, kad bazė nėra išsukama, tačiau gali reikšti ir tai, kad išsukto pyC fluorescencija yra gesinama aromatinių kišenės ar. Kadangi kristalinėje struktūroje tokia kišenė nėra matoma, galima teigti, kad McrA atpažįsta modifikuotą citozino bazę dgDNR kontekste be bazės išsukimo.

2.4.5 McrA modifikuotos DNR surišimo centrai

Struktūriniai duomenys rodo, kad McrA turėtų turėti vieną jungtinį DNR surišimo centrą nukleaziniame domene ir du atskirus modifikuotos DNR surišimo domenus (Czapinska et al., 2018). Norėdami patvirtinti šį modelį, pirmiausiai patikrinome WT McrA susirišimą su metilintais ir nemetilintais 30 ir 12 bp dupleksais EJP eksperimentuose. N-galinio domeno susirišimą skatino metilinimas. 30 bp ilgio substrato susirišimo priklausomybė nuo metilinimo buvo labiau išreikšta su McrA-N nei su pilno ilgio baltymu (2.13 A pav.), kas neprieštarauja DNazės I pėdsakų rezultatams. Šie duomenys rodo, kad nuklеazė pasižymi silpna nespecifine sąveika su ilgesniais (pvz., 30 bp) DNR substratais, tačiau ne su trumpais (pvz., 12 bp) oligodupleksais, kurių visą paviršių gali uždengti N-galinis domenas.

Pilno ilgio baltymas, bet ne McrA-N, EJP eksperimentuose sudarė du skirtingus kompleksus, nes buvo matomos dvi skirtingų judrumų juostelės (2.13 A pav.). Šis reiškinys jau buvo aprašytas anksčiau, ir buvo aiškinamas monomero-dimero pusiausvyra (Mulligan et al., 2010). Remiantis mūsų biocheminiais ir struktūriniais duomenimis, du kompleksai galėtų atitikti McrA homodimerus su vienu ar dviem oligodupleksais. Norėdami patikrinti šį modelį, atlikome EJP eksperimentus su dviejų skirtingų ilgių modifikuotais oligodupleksais, sumaišytais įvairiais santykiais (2.13 B pav.). Stebėjome tris skirtingo judrumo kompleksus, kas patvirtina, kad McrA dimeras vienu metu gali surišti du metilintus DNR taikinius.



2.13 pav. McrA sąveika su DNR. (A) EJP eksperimentai su nemodifikuota, pusiau ir pilnai metilinta 30 ir 12 bp DNR (atpažinimo sekos parodytos, M žymi 5mC), DNR koncentracija 50 nM, McrA dimero koncentracija pažymėta virš gelių takelių. Laisva DNR pažymėta mėlynai, baltymo-DNR kompleksai – raudonai. (B) EJP eksperimentai su žymėta 30 bp (30_5mC/30_5mC) ir 12 bp (12_5mC/12_5mC) DNR. Koncentracijos pažymėtos virš gelių takelių. Pieštame paveiksliuke parodyta dviejų tipų nesurišta DNR (12 ir 30 bp) ir trijų tipų baltymo-DNR kompleksai, turintys dvi 30 bp, dvi 12 bp ir dvi skirtingas DNR. McrA N- ir C-galiniai domenai pavaizduoti mėlynais šešiakampiais ir raudonais apskritimais. Vg oligonukleotidas parodytas mėlyna punktyrine rodykle.

2.4.6 McrA aktyvumas ir sandara

McrA N-galinis domenas rišasi specifiškai su metilinta DNR. EJP eksperimentai parodė, kad didžiausias gimiringumas yra pilnai metilintai DNR, mažesnis – pusiau metilintai DNR, o mažiausias – nemetilintai. McrA N-galinio domeno struktūra skiriasi nuo kitų modifikuotą DNR atpažįstančių domenų. Šis domenas nėra struktūriškai panašus į SRA domenus ar McrB-N domeną, atsakingą už modifikacijos atpažinimą McrBC (Sukackaite et al., 2012). McrA N-galinio fragmento DNR surišimo ir atpažinimo mechanizmas vis dar lieka nežinomas.

C-galinis domenas yra atsakingas už nukleazinį McrA aktyvumą. Domenas turi tipinę HNH domenų struktūrą su atitinkamomis ar., kurios yra konservatyvios McrA šeimoje. Tad yra keista, kodėl *in vitro* aktyvumas iki šiol nebuvo stebėtas. Mums pavyko pademonstruoti silpną nuo Mn^{2+} jonų priklausomą McrA endonukleazinį aktyvumą su dg ir vg DNR. Netikėta tai, kad baltymo aktyvumui reikia būtent Mn^{2+} jonų, nes dauguma HNH endonukleazių turi mažiau griežtą selektyvumą metalo katijonams aktyviajame centre.

Bendra McrA sandara su N-galiniu modifikacijai specifiniu ir C-galiniu nukleaziniu domenu primena įvairių kitų modifikuotą DNR atpažįstančių dvidomeninių REazių sandarą, pvz. SRA-HNH (Han et al., 2015; Kisiala et al., 2018), PD-(D/E)XK-SRA (PvuRts1I) (Kazrani et al., 2014) ir SRA-PD-(D/E)XK (MspJI) (Horton et al., 2012). Visų šių fermentų atveju nespecifinis nukleazinis domenas yra atsakingas už dimerizaciją susidarant vienam bendram DNR surišimo centrui, tuo tarpu modifikacijai specifiniai domenai turi atskirus modifikuotos DNR surišimo centrus.

Deja, McrA *in vitro* aktyvumas yra silpnas, tik dalinai priklausomas nuo metilinimo, ir stebimas tik esant aukštoms baltymo koncentracijoms. Kodėl reikalinga tokia didelė baltymo koncentracija ir ilgi inkubacijos laikai lieka neaišku, ypač turint omeny nepažeistą ir tipinį HNH aktyvųjį centrą (Czapinska et al., 2018). Mažas McrA *in vitro* aktyvumas nesiderina su baltymo efektyvumu apsaugant *E. coli* nuo citozino modifikacijas turinčių plazmidžių *in vivo*. Todėl gali būti, kad optimalus McrA substratas *in vitro* eksperimentuose nebuvo atrastas. Galbūt toks substratas susidaro tik ląstelėse DNR replikacijos ar transkripcijos metu. McrA taip pat galėtų reikėti aktyvatoriaus ar efektorinės nukleazės iš ląstelės baltymų. Tačiau mūsų *in vivo* tyrimai (Czapinska et al., 2018) rodo, kad McrA *in vivo* veikimui būtinas nepažeistas jo katalizės centras, tad ląstelės šeimininko nukleazės panaudojimas yra mažai tikėtinas. Molekulinis McrA veikimo mechanizmas tokiu būdu lieka iki galo neaiškus.

IŠVADOS

1. Cheminiu metodu, kuris remiasi išsuktų pirimidinų oksidacija $KMnO_4$, buvo patvirtinta, kad MspJI ir PvuRts1I šeimoms priklausančios REazės LpnPI, YkrI ir BmeDI išsuka modifikuotą citoziną iš dvigrandinės DNR.
2. MspJI šeimos baltymas LpnPI atpažįsta 5mC ir jį supantį sekos kontekstą per keletą paviršinių kilpų, veikiančių kaip atskiri DNR surišimo / atpažinimo moduliai. Keičiant šias kilpas homologiškų baltymų AspBHI ir SgrTI, kurie atpažįsta skirtingas sekas, kilpomis gali būti keičiamas fermento specifiškumas modifikuotą citoziną supančiai sekai.
3. McrBC DNR atpažinimo domenas McrB-N atpažįsta DNR substratus su įvairiomis citozino modifikacijomis, o pagal sąveikos stiprumą citozino variantai išsidėsto tvarka 4mC > 5mC > 5hmC >> 5fC. Šią sąveiką galima keisti baltymo kišenės mutacijomis.
4. McrA *in vitro* pasižymi silpnu, nuo Mn^{2+} jonų priklausomu endonukleaziniu aktyvumu, kuris tik dalinai priklauso nuo DNR metilinimo. DNR kirpimą atlieka McrA C-galinis HNH domenas, kuris taip pat yra atsakingas už baltymo dimerizaciją.
5. McrA N-galinis domenas yra atsakingas už modifikuotą citoziną turinčių taikinių atpažinimą, kuris vyksta be bazės išsukimo.

LITERATŪROS SĄRAŠAS

1. Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455, 818–821.
2. Avvakumov, G. V, Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455, 822–825.
3. Berry, D.A., Jung, K.-Y., Wise, D.S., Sercel, A.D., Pearson, W.H., Mackie, H., Randolph, J.B., and Somers, R.L. (2004). Pyrrolo-dC and pyrrolo-C: fluorescent analogs of cytidine and 2′-deoxycytidine for the study of oligonucleotides. Tetrahedron Lett. 45, 2457–2461.
4. Bischerour, J. and Chalmers, R. (2007). Base-flipping dynamics in a DNA hairpin processing reaction. Nucleic Acids Res. 35, 2584–2595.
5. Buck-Koehntop, B.A., Stanfield, R.L., Ekiert, D.C., Martinez-Yamout, M.A., Dyson, H.J., Wilson, I.A., and Wright, P.E. (2012). Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc. Natl. Acad. Sci. U. S. A. 109, 15229–15234.
6. Bujnicki, J.M., Radlinska, M., and Rychlewski, L. (2000). Atomic model of the 5-methylcytosine-specific restriction enzyme McrA reveals an atypical zinc finger and structural similarity to betabetaalphaMe endonucleases. Mol. Microbiol. 37, 1280–1281.
7. Burman, R.W., Yates, P.A., Green, L.D., Jacky, P.B., Turker, M.S., and Popovich, B.W. (1999). Hypomethylation of an expanded FMR1 allele is not associated with a global DNA methylation defect. Am. J. Hum. Genet. 65, 1375–1386.
8. Card, C.O., Wilson, G.G., Weule, K., Hasapes, J., Kiss, A., and Roberts, R.J. (1990). Cloning and characterization of the HpaII methylase gene. Nucleic Acids Res. 18, 1377–1383.
9. Chevalier, B.S. and Stoddard, B.L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res. 29, 3757–3774.
10. Chotai, K.A. and Payne, S.J. (1998). A rapid, PCR based test for differential molecular diagnosis of Prader-Willi and Angelman syndromes. J. Med. Genet. 35, 472–475.
11. Cohen-Karni, D., Xu, D., Apone, L., Fomenkov, A., Sun, Z., Davis, P.J., Kinney, S.R.M., Yamada-Mabuchi, M., Xu, S., Davis, T., et al. (2011). The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. Proc. Natl. Acad. Sci. U. S. A. 108, 11040–11045.
12. Czapinska, H., Kowalska, M., Zagorskaitė, E., Manakova, E., Slyvka, A., Xu, S., Siksnys, V., Sasnauskas, G., and Bochtler, M. (2018). Activity and structure of EcoKMcrA. Nucleic Acids Res.
13. Daujotyte, D., Liutkeviciute, Z., Tamulaitis, G., and Klimasauskas, S. (2008). Chemical mapping of cytosines enzymatically flipped out of the DNA helix. Nucleic Acids Res. 36, e57–e57.
14. Du, J., Johnson, L.M., Groth, M., Feng, S., Hale, C.J., Li, S., Vashisht, A.A., Gallego-Bartolome, J., Wohlschlegel, J.A., Patel, D.J., et al. (2014). Mechanism of DNA Methylation-Directed Histone Methylation by KRYPTONITE. Mol. Cell 55, 495–504.
15. Fritzsche, E., Hayatsu, H., Igloi, G.L., Iida, S., and Kössel, H. (1987). The use of permanganate as a sequencing reagent for identification of 5-methylcytosine

residues in DNA. Nucleic Acids Res. 15, 5517–5528.
16. Gogos, J.A., Karayiorgou, M., Aburatani, H., and Kafatos, F.C. (1990). Detection of single base mismatches of thymine and cytosine residues by potassium permanganate and hydroxylamine in the presence of tetralkylammonium salts. Nucleic Acids Res. 18, 6807–6814.
17. Han, T., Yamada-Mabuchi, M., Zhao, G., Li, L., liu, G., Ou, H.-Y., Deng, Z., Zheng, Y., and He, X. (2015). Recognition and cleavage of 5-methylcytosine DNA by bacterial SRA-HNH proteins. Nucleic Acids Res. 43, 1147–1159.
18. Hashimoto, H., Horton, J.R., Zhang, X., Bostick, M., Jacobsen, S.E., and Cheng, X. (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455, 826–829.
19. Hashimoto, H., Horton, J.R., Zhang, X., and Cheng, X. (2009). UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. Epigenetics 4, 8–14.
20. Hayatsu, H. and Ukita, T. (1967). The selective degradation of pyrimidines in nucleic acids by permanganate oxidation. Biochem. Biophys. Res. Commun. 29, 556–561.
21. Ho, K.L., McNae, I.W., Schmiedeberg, L., Klose, R.J., Bird, A.P., and Walkinshaw, M.D. (2008). MeCP2 binding to DNA depends upon hydration at methyl-CpG. Mol. Cell 29, 525–531.
22. Horton, J.R., Mabuchi, M.Y., Cohen-Karni, D., Zhang, X., Griggs, R.M., Samaranayake, M., Roberts, R.J., Zheng, Y., and Cheng, X. (2012). Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease. Nucleic Acids Res. 40, 9763–9773.
23. Horton, J.R., Wang, H., Mabuchi, M.Y., Zhang, X., Roberts, R.J., Zheng, Y., Wilson, G.G., and Cheng, X. (2014b). Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. Nucleic Acids Res. 42, 12092–12101.
24. Horton, J.R., Nugent, R.L., Li, A., Mabuchi, M.Y., Fomenkov, A., Cohen-Karni, D., Griggs, R.M., Zhang, X., Wilson, G.G., Zheng, Y., et al. (2014a). Structure and mutagenesis of the DNA modification-dependent restriction endonuclease AspBHI. Sci. Rep. 4, 4246.
25. Horton, J.R., Borgaro, J.G., Griggs, R.M., Quimby, A., Guan, S., Zhang, X., Wilson, G.G., Zheng, Y., Zhu, Z., and Cheng, X. (2014c). Structure of 5-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA. Nucleic Acids Res. 42, 7947–7959.
26. Irier, H.A. and Jin, P. (2012). Dynamics of DNA methylation in aging and Alzheimer's disease. DNA Cell Biol. 31 Suppl 1, S42-8.
27. Iurlaro, M., Ficz, G., Oxley, D., Raiber, E.-A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S., and Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. Genome Biol. 14, R119.
28. Kazrani, A.A., Kowalska, M., Czapinska, H., and Bochtler, M. (2014). Crystal structure of the 5hmC specific endonuclease PvuRts1I. Nucleic Acids Res. 42, 5929–5936.
29. Kisiala, M., Copelas, A., Czapinska, H., Xu, S., and Bochtler, M. (2018). Crystal structure of the modification-dependent SRA-HNH endonuclease TagI. Nucleic Acids Res.
30. Kuśmierek, J.T. and Singer, B. (1982). Chloroacetaldehyde-treated ribo- and deoxyribopolynucleotides. 1. Reaction products. Biochemistry 21, 5717–5722.
31. Kuznetsov, N.A., Vorobjev, Y.N., Krasnoperov, L.N., and Fedorova, O.S. (2012). Thermodynamics of the multi-stage DNA lesion recognition and repair by

formamidopyrimidine-DNA glycosylase using pyrrolocytosine fluorescence--stopped-flow pre-steady-state kinetics. Nucleic Acids Res. 40, 7384–7392.
32. Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8, 317–327.
33. Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature 430, 471–476.
34. Lippman, Z., Gendrel, A.-V., Colot, V., and Martienssen, R. (2005). Profiling DNA methylation patterns using genomic tiling microarrays. Nat. Methods 2, 219–224.
35. Liu, C. and Martin, C.T. (2001). Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase. J. Mol. Biol. 308, 465–475.
36. Lyko, F., Ramsahoye, B.H., Kashevsky, H., Tudor, M., Mastrangelo, M.A., Orr-Weaver, T.L., and Jaenisch, R. (1999). Mammalian (cytosine-5) methyltransferases cause genomic DNA methylation and lethality in Drosophila. Nat. Genet. 23, 363–366.
37. Mulligan, E.A. and Dunn, J.J. (2008). Cloning, purification and initial characterization of E. coli McrA, a putative 5-methylcytosine-specific nuclease. Protein Expr. Purif. 62, 98–103.
38. Mulligan, E.A., Hatchwell, E., McCorkle, S.R., and Dunn, J.J. (2010). Differential binding of Escherichia coli McrA protein to DNA sequences that contain the dinucleotide m5CpG. Nucleic Acids Res. 38, 1997–2005.
39. Oakeley, E.J., Schmitt, F., and Jost, J.P. (1999). Quantification of 5-methylcytosine in DNA by the chloroacetaldehyde reaction. Biotechniques 27, 744–746, 748–750, 752.
40. Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell 105, 487–497.
41. Otani, J., Arita, K., Kato, T., Kinoshita, M., Kimura, H., Suetake, I., Tajima, S., Ariyoshi, M., and Shirakawa, M. (2013). Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. J. Biol. Chem. 288, 6351–6362.
42. Pingoud, A., Wilson, G.G., and Wende, W. (2014). Type II restriction endonucleases--a historical perspective and more. Nucleic Acids Res. 42, 7489–7527.
43. Rajakumara, E., Law, J.A., Simanshu, D.K., Voigt, P., Johnson, L.M., Reinberg, D., Patel, D.J., and Jacobsen, S.E. (2011). A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes Dev. 25, 137–152.
44. Reddy, Y. V and Rao, D.N. (2000). Binding of EcoP15I DNA methyltransferase to DNA reveals a large structural distortion within the recognition sequence. J. Mol. Biol. 298, 597–610.
45. Revel, H.R. (1967). Restriction of nonglucosylated T-even bacteriophage: properties of permissive mutants of Escherichia coli B and K12. Virology 31, 688–701.
46. Sasnauskas, G., Zagorskaitė, E., Kauneckaitė, K., Tamulaitiene, G., and Siksnys, V. (2015). Structure-guided sequence specificity engineering of the modification-dependent restriction endonuclease LpnPI. Nucleic Acids Res. 43, 6144–6155.
47. Sasnauskas, G., Kauneckaite, K., and Siksnys, V. (2018). Structural basis of DNA target recognition by the B3 domain of Arabidopsis epigenome reader VAL1. Nucleic Acids Res. 46, 4316–4324.

48. Scarsdale, J.N., Webb, H.D., Ginder, G.D., and Williams, D.C. (2011). Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. Nucleic Acids Res. 39, 6741–6752.
49. Serva, S., Weinhold, E., Roberts, R.J., and Klimasauskas, S. (1998). Chemical display of thymine residues flipped out by DNA methyltransferases. Nucleic Acids Res. 26, 3473–3479.
50. Shao, C., Wang, C., and Zang, J. (2014). Structural basis for the substrate selectivity of PvuRts1I, a 5-hydroxymethylcytosine DNA restriction endonuclease. Acta Crystallogr. D. Biol. Crystallogr. 70, 2477–2486.
51. Smith, Z.D. and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. 14, 204–220.
52. Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013). Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. Cell 152, 1146–1159.
53. Sukackaite, R., Grazulis, S., Tamulaitis, G., and Siksnys, V. (2012). The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. Nucleic Acids Res. 40, 7552–7562.
54. Sutherland, E., Coe, L., and Raleigh, E.A. (1992). McrBC: a multisubunit GTP-dependent restriction endonuclease. J. Mol. Biol. 225, 327–348.
55. Szwagierczak, A., Brachmann, A., Schmidt, C.S., Bultmann, S., Leonhardt, H., and Spada, F. (2011). Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5–hydroxymethylcytosine. Nucleic Acids Res. 39, 5149–5156.
56. Wang, H., Guan, S., Quimby, A., Cohen-Karni, D., Pradhan, S., Wilson, G., Roberts, R.J., Zhu, Z., and Zheng, Y. (2011). Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. Nucleic Acids Res. 39, 9294–9305.
57. Yoshioka, K. (2002). KyPlot as a Tool for Graphical Data Analysis. In Compstat, (Heidelberg: Physica-Verlag HD), pp. 37–46.
58. Zagorskaitė, E. and Sasnauskas, G. (2014). Chemical Display of Pyrimidine Bases Flipped Out by Modification-Dependent Restriction Endonucleases of MspJI and PvuRts1I Families. PLoS One 9, e114580.
59. Zheng, L., Baumann, U., and Reymond, J.-L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res. 32, e115.
60. Zheng, Y., Cohen-Karni, D., Xu, D., Chin, H.G., Wilson, G., Pradhan, S., and Roberts, R.J. (2010). A unique family of Mrr-like modification-dependent restriction endonucleases. Nucleic Acids Res. 38, 5527–5534.
61. Zhou, T., Xiong, J., Wang, M., Yang, N., Wong, J., Zhu, B., and Xu, R.-M. (2014). Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol. Cell 54, 879–886.

ACKNOWLEDGEMENTS

I am very grateful to my supervisor dr. Giedrius Sasnauskas for educational discussions, motivation, critical remarks and help with the preparation of publications and this dissertation.

I am very appreciative to prof. Virginijus Šikšnys for opportunity to work in the Department of Protein-DNA Interactions.

I would like to thank to the people who contributed in this work: dr. Giedrius Sasnauskas, dr. Giedrė Tamulaitienė and dr. Elena Manakova for the crystal structures and models of DNA binding domains of LpnPI and McrB, also dr. Elena Manakova for SAXS experiments with McrA, Kotryna Kauneckaitė for part of mutagenesis and cleavage experiments with LpnPI and LpnPI-N.

I am very thankful for Honorata Czapinska and prof. Matthias Bochtler's group (International Institute of Molecular and Cell Biology, Warsaw, Poland) for McrA crystal structure and the cooperation writing publication.

I wish to thank to all the colleagues in the Department of Protein-DNA Interactions for their support, advices, warm atmosphere and help with experiments.

I am very appreciative to my family and friends for support and motivation.

APPENDIX: LIST OF PUBLICATIONS (1, 2, 3, 4)

The results of this thesis were presented in following publications and conferences.

Publications:

1. **Zagorskaitė E** and Sasnauskas G (2014) Chemical display of pyrimidine bases flipped out by modification-dependent restriction endonucleases of MspJI and PvuRts1I families. PLoS One, 9, e114580.
2. Sasnauskas G, **Zagorskaitė E**, Kauneckaitė K, Tamulaitiene G and Siksnys V (2015) Structure-guided sequence specificity engineering of the modification-dependent restriction endonuclease LpnPI. Nucleic Acids Res., 43, 6144-55.
3. Czapinska H, Kowalska M, **Zagorskaite E**, Manakova E, Slyvka A, Xu SY, Siksnys V, Sasnauskas G and Bochtler M (2018) Activity and structure of EcoKMcrA. Nucleic Acids Res., 46, 9829–9841.
4. **Zagorskaitė E**, Manakova E and Sasnauskas G (2018) Recognition of modified cytosine variants by the DNA-binding domain of methyl-directed endonuclease McrBC. FEBS Lett., 592, 3335-3345.

Conferences:

1. **E. Zagorskaitė,** and G. Sasnauskas. Base flipping by methyl-directed restriction endonucleases. 13th international conference of Lithuanian Biochemical Society, Birštonas, Lithuania, 2014.06.17-20.
2. **E. Zagorskaitė,** G. Tamulaitienė, K. Kauneckaitė, G. Sasnauskas, and V. Siksnys. DNA recognition mechanism of the cytosine modification-dependent restriction endonuclease LpnPI. The Seventh NEB Meeting on DNA Restriction and Modification, Gdansk, Poland, 2015.08.24-29.
3. **E. Zagorskaitė,** G. Tamulaitienė, K. Kauneckaitė, G. Sasnauskas, and V. Siksnys. DNA recognition mechanism of the cytosine modification-dependent restriction endonuclease LpnPI. 14th international conference of Lithuanian Biochemical Society, Druskininkai, Lithuania, 2016.06.27-30.

APPENDIX 1

Chemical display of pyrimidine bases flipped out by modification-dependent restriction endonucleases of MspJI and PvuRts1I families

Zagorskaitė E and Sasnauskas G

PLoS One, 2014, volume 9, e114580
https://doi.org/10.1371/journal.pone.0114580

PLOS ONE

RESEARCH ARTICLE

Chemical Display of Pyrimidine Bases Flipped Out by Modification-Dependent Restriction Endonucleases of MspJI and PvuRts1I Families

Evelina Zagorskaitė, Giedrius Sasnauskas*

Department of Protein–DNA Interactions, Institute of Biotechnology, Vilnius University, Vilnius, Lithuania

*gsasnaus@ibt.lt

OPEN ACCESS

Citation: Zagorskaitė E, Sasnauskas G (2014) Chemical Display of Pyrimidine Bases Flipped Out by Modification-Dependent Restriction Endonucleases of MspJI and PvuRts1I Families. PLoS ONE 9(12): e114580. doi:10.1371/journal.pone.0114580

Editor: Shuang-yong Xu, New England Biolabs, Inc., United States of America

Received: October 8, 2014

Accepted: November 11, 2014

Published: December 8, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Research Council of Lithuania (Grant MIP-027/2012 to GS). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

The epigenetic DNA modifications 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) in eukaryotes are recognized either in the context of double-stranded DNA (e.g., by the methyl-CpG binding domain of MeCP2), or in the flipped-out state (e.g., by the SRA domain of UHRF1). The SRA-like domains and the base-flipping mechanism for 5(h)mC recognition are also shared by the recently discovered prokaryotic modification-dependent endonucleases of the MspJI and PvuRts1I families. Since the mechanism of modified cytosine recognition by many potential eukaryotic and prokaryotic 5(h)mC "readers" is still unknown, a fast solution based method for the detection of extrahelical 5(h)mC would be very useful. In the present study we tested base-flipping by MspJI- and PvuRts1I-like restriction enzymes using several solution-based methods, including fluorescence measurements of the cytosine analog pyrrolocytosine and chemical modification of extrahelical pyrimidines with chloroacetaldehyde and $KMnO_4$. We find that only $KMnO_4$ proved an efficient probe for the positive display of flipped out pyrimidines, albeit the method required either non-physiological pH (4.3) or a substitution of the target cytosine with thymine. Our results imply that DNA recognition mechanism of 5(h)mC binding proteins should be tested using a combination of all available methods, as the lack of a positive signal in some assays does not exclude the base flipping mechanism.

Introduction

5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) are important epigenetic modifications of mammalian and plant DNA. The methylation and hydroxymethylation levels are dynamic and vary in different types of cells during development, differentiation, aging, and disease [1, 2]. Structural studies of eukaryotic 5mC/5hmC binding domains revealed two different strategies for the modified base recognition. Proteins that share the methyl-CpG binding domain, including MBD1, MBD2, MBD4, and MeCP2, also a zinc-finger protein Kaiso, recognize modified cytosine in the context of a Watson-Crick base pair [3–7]. In contrast, the SRA (SET and RING-associated) domains of UHRF1, UHRF2, and SUVH5 proteins flip out the modified base and place it in a protein pocket [8–12] (Fig. 1A).

In prokaryotes modification-dependent restriction endonucleases protect host bacteria from bacteriophages containing modified DNA. The McrBC complex recognizes 5mC, 5hmC or 4-methylcytosine-containing sequences, and cleaves DNA at a variable position from the recognition site. The crystal structure of the McrBC DNA binding domain (McrB-N) revealed that despite an unrelated tertiary structure, it follows the same mechanism for modified base recognition as the SRA domains: the base is flipped out into a protein pocket [13]. MspJI-like enzymes recognize 5mC and 5hmC modifications in various sequence contexts (for example, 59-5mCNNR-39for MspJI) and cleave top and bottom DNA strands 12/16 nucleotides downstream of the modified base. PvuRts1I family enzymes recognize DNA substrates with 5hmC or glucosylated 5hmC (5ghmC) modifications and, unlike other modification-dependent enzymes, discriminate against substrates with 5mC [14, 15]. Due to strict specificity for unmodified DNA and fixed cleavage positions, both MspJI- and PvuRts1I-like enzymes are used as molecular tools for single-base resolution mapping of 5mC and 5hmC modifications in eukaryotic genomes [16, 17]. The recently solved structures of MspJI, AspBHI (MspJI family), PvuRts1I and AbaSI (PvuRts1I family) enzymes [18–23] revealed that these enzymes are comprised of a PD-(D/E)XK nuclease domain fused to a SRA-like DNA binding domain (DBD). The co-crystal structure of the MspJI-DNA complex demonstrated that MspJI flips-out 5-methylcytosine into a protein pocket where a set of base-specific contacts are made (Fig. 1B) [21]. Mutational analysis and close structural resemblance of DNA binding domains of AspBHI, PvuRts1I and AbaSI to the SRA domains of MspJI and eukaryotic proteins (Fig. 1C-D) suggest that these enzymes also flip-out the modified cytosine [18, 20–23].

However, for many eukaryotic 5(h)mC-binding proteins identified in the recent pull-down and mass-spectrometry studies [24, 25], the mechanism of modified cytosine recognition remains unknown. Mechanistic studies of these proteins would benefit from a fast solution-based method for the detection of extrahelical 5(h)mC. Here we used modification-dependent restriction endonucleases of MspJI and PvuRts1I families as a test system to assess the performance of several solution-based methods for extrahelical pyrimidine detection, including

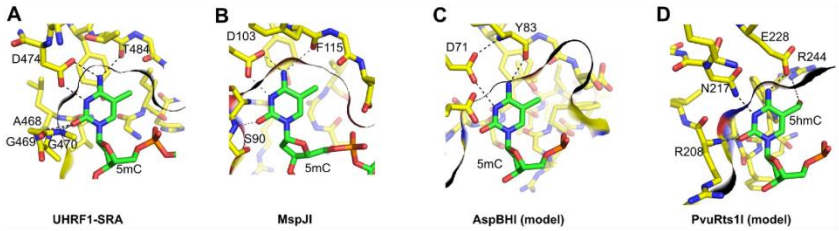


Figure 1. The modified cytosine binding pockets. (A-B) 5-methylcytosine recognition by the UHRF1 SRA domain (PDB ID 3fde) and the DNA recognition domain of MspJI endonuclease (PDB ID 4r28). The indicated protein pocket residues make base-specific contacts to the extrahelical base [19,45]. (C-D) The models for the modified cytosine recognition by the DNA binding domains of AspBHI (PDB ID 4oc8) and PvuRts1I (PDB ID 4oq2, see Materials and Methods for details). The indicated amino acid residues could form base-specific contacts to the extruded base. In the case of AspBHI, this would require protonation of the D71 residue. In each panel the dark line marks the boundaries of the protein pocket cut at the plane of the cytosine ring.

doi:10.1371/journal.pone.0114580.g001

fluorescence measurements of the cytosine analog pyrrolocytosine and chemical display of extrahelical bases [26].

Materials and Methods

DNA oligonucleotides

Oligoduplex substrates used in this study are listed in Table 1. Oligonucleotides with 5-hydroxymethylcytosine modifications were purchased from IBA, all other oligonucleotides were from Metabion. Oligonucleotides were 59-labeled with [c-^{33}P]ATP or [c-^{32}P]ATP (Hartmann Analytic) and T4 polynucleotide kinase (Thermo Fisher Scientific). Oligoduplexes were assembled by annealing the corresponding radiolabeled and unlabeled strands.

Proteins

The genes encoding LpnPI, YkrI, and BmeDI were amplified by PCR from the genomic DNA of Legionella pneumophila Philadelphia-1 (DSM No. 7513), Yersinia kristensenii (DSM No. 18543), and Bacillus megaterium (DSM No. 319), respectively. Genomic DNA was purchased from Leibnitz Institute DSMZ (Germany). Genes of LpnPI and its N-terminal DNA binding domain (LpnPI-N, corresponds to 1-224 residues of the full length protein) were cloned into the pLATE11 expression vector (Thermo Fisher Scientific). The first methionine in both proteins was replaced by a short hexahistidine tag (sequence MGHHHHHHG). Genes encoding YkrI and BmeDI were cloned into the pTYB2 expression vector as C-terminal fusions with the self-cleavable chitin binding domain. All proteins were expressed in the E. coli strain ER2566 (New England Biolabs). Cells were grown to OD_{600} 0.5–0.8 and induced with a final

Table 1. Oligonucleotide substrates.

Duplex	Sequence [a]	Specification
16-M	59-CGTAGCMTGGTCGATC-39	
	39-GCATCGGACCAGCTAG-59	A short cognate LpnPI oligoduplex
16-C	59-CGTAGCCTGGTCGATC-39	
	39-GCATCGGACCAGCTAG-59	As 16-M but 5mC is replaced with an unmodified cytosine
16-P	59-CGTAGCPTGGTCGATC-39	
	39-GCATCGGACCAGCTAG-59	As 16-M but 5mC is replaced with a pyrrolocytosine
16-T	59-CGTAGCTTGGTCGATC-39	
	39-GCATCGGACCAGCTAG-59	As 16-M but 5mC is replaced with a thymine
16-T-N	59-CGTAGATCTTACGATC-39	
	39-GCATCTGGAATGCTAG-59	Noncognate LpnPI oligoduplex with a T-G mismatch.
30-M	59-CCGTAGCMTGGTCGATCCTAGCTGGTCGCC-39	
	39-GGCATCGGACCAGCTAGGATCGACCAGCGG-59	An extended cognate LpnPI oligoduplex
30-C	59-CCGTAGCCTGGTCGATCCTAGCTGGTCGCC-39	
	39-GGCATCGGACCAGCTAGGATCGACCAGCGG-59	As 30-M but 5mC is replaced with an unmodified cytosine
30-P	59-CCGTAGCPTGGTCGATCCTAGCTGGTCGCC-39	
	39-GGCATCGGACCAGCTAGGATCGACCAGCGG-59	As 30-M but 5mC is replaced with a pyrrolocytosine
30-T	59-CCGTAGCTTGGTCGATCCTAGCTGGTCGCC-39	
	39-GGCATCGGACCAGCTAGGATCGACCAGCGG-59	As 30-M but 5mC is replaced with a thymine
39-H/H	59-CGACGATAHTTACCGGATAAGGCGCAATTAGATTACTTC-39	
	39-GCTGCTATGAATGGCCTATTCCGCGTTAATHTAATGAAG-59	Optimal YkrI and BmeDI substrate with two 5hmC-G base pairs
39-M/H	59-CGACGATAMTTACCGGATAAGGCGCAATTAGATTACTTC-39	
	39-GCTGCTATGAATGGCCTATTCCGCGTTAATHTAATGAAG-59	As 39-H/H but one 5hmC is replaced with a 5-methylcytosine
39-C/H	59-CGACGATACTTACCGGATAAGGCGCAATTAGATTACTTC-39	
	39-GCTGCTATGAATGGCCTATTCCGCGTTAATHTAATGAAG-59	As 39-H/H but one 5hmC is replaced with an unmodified cytosine
39-P/H	59-CGACGATAPTTACCGGATAAGGCGCAATTAGATTACTTC-39	
	39-GCTGCTATGAATGGCCTATTCCGCGTTAATHTAATGAAG-59	As 39-H/H but one 5hmC is replaced with a pyrrolocytosine
39-T/H	59-CGACGATATTTACCGGATAAGGCGCAATTAGATTACTTC-39	
	39-GCTGCTATGAATGGCCTATTCCGCGTTAATHTAATGAAG-59	As 39-H/H but one 5hmC is replaced with a thymine
39-H	59-CGACGATATTTACCGGATAAGGCGCAATTAGATTACTTC-39	
	39-GCTGCTATAAATGGCCTATTCCGCGTTAATHTAATGAAG-59	As 39-H/H but one 5hmC-G base pair is replaced with a T-A base pair
39-T	59-CGACGATATTTACCGGATAAGGCGCAATTAAATTACTTC-39	
	39-GCTGCTATGAATGGCCTATTCCGCGTTAATTTAATGAAG-59	As 39-T/H but the 5hmC-G base pair is replaced with a T-A base pair
31-C	59-TGACCCACGCTCGCCCGGCGACACATTACGT-39	
	39-ACTGGGTGCGAGCGGGCCGCTGTGTAATGCA-59	Cognate Ecl18kI oligoduplex with a C-G central base pair

Table 1. Cont.

Duplex	Sequence [a]	Specification
31-M	59-TGACCCACGCTCGCCMGGCGACACATTACGT-39	
	39-ACTGGGTGCGAGCGGGCCGCTGTGTAATGCA-59	As 31-C but with a central 5mC-G base pair

[a]M designates 5-methylcytosine (5mC), H designates 5-hydroxymethylcytosine (5hmC), P designates pyrrolocytosine. Modified bases and mismatched base pairs are shown in boldface, recognition sequences are underlined.

doi:10.1371/journal.pone.0114580.t001

concentration of 0.2 mM IPTG at 16°C overnight, harvested by centrifugation and stored at 2 20°C.

The cells expressing LpnPI and LpnPI-N were sonicated in a buffer containing 20 mM Tris-HCl, pH 7.5, 450 mM NaCl, 10% v/v glycerol, and 7 mM 2-mercaptoethanol. Cleared lysates were collected after centrifugation at 40000 g for 1 h. Both proteins were purified by chromatography through HisTrap HP chelating and HiTrap Heparin HP columns and a gel-filtration column HiLoad 16/600 Superdex 200 pg (GE Healthcare).

YkrI and BmeDI were purified using a chitin column (New England Biolabs) as described by Wang et al. [15], and subsequently were loaded on a HiTrap Heparin HP column and eluted using a buffer containing Tris-acetate (pH 7.6) and 100–1000 mM potassium acetate.

Purified LpnPI and LpnPI-N were stored at 2 20°C in a buffer containing 20 mM Tris-HCl (pH 7.5 for LpnPI-N, pH 8.5 for LpnPI), 200 mM KCl, 1 mM DTT, and 50% v/v glycerol. The YkrI and BmeDI storage buffer contained 20 mM Tris-acetate (pH 7.5), 250 mM potassium-acetate, 1 mM DTT and 50% v/v glycerol.

Wt Ecl18kI was purified as described previously [27]. The purity of all proteins was higher than 95% as judged by SDS-PAGE. Protein concentrations were determined from A_{280} measurements using the theoretical extinction coefficients calculated with the ProtParam tool available at http://web.expasy.org/protparam/. All protein concentrations are expressed in terms of monomer if not stated otherwise.

Electrophoretic mobility shift assay

DNA binding was analysed by the electrophoretic mobility shift assay (EMSA) using ^{33}P-labeled oligoduplexes. DNA (final concentration 1, 10 or 100 nM) was incubated with the protein (final concentrations varied from 5 to 1000 nM) for 15 min in 20 ml of the binding buffer containing either 40 mM Tris-acetate (pH 8.3 at 25°C) or 30 mM Mes-histidine (pH 6.3 at 25°C), 5 mM calcium-acetate, 0.1 mg/ml BSA and 10% v/v glycerol. Free DNA and protein–DNA complexes were separated by electrophoresis through 8% acrylamide gels (29:1 acrylamide/bisacrylamide) in either 40 mM Tris–acetate, pH 8.3, or 30 mM Mes-histidine, pH 6.3, all with 5 mM calcium-acetate for 60-90 min at 5 V/cm. In some cases the binding and the electrophoresis buffers were devoid of calcium-acetate and

instead were supplemented with 1 mM EDTA. Radiolabeled DNA and protein-DNA complexes were detected and quantified using the Cyclone phosphorimager and the OptiQuant software (Packard Instrument) [28].

DNA cleavage experiments

DNA hydrolysis reactions were performed by manually mixing radiolabeled oligoduplexes (1 nM for YkrI and BmeDI, 400 nM for LpnPI) with the enzyme (100 nM YkrI and BmeDI, 500 nM LpnPI) in the Reaction Buffer (33 mM Tris-acetate, pH 8.0, 66 mM K-acetate, 10 mM Mg-acetate, 0.1 mg/ml BSA) at either 25°C (LpnPI) or 15°C (YkrI and BmeDI). Samples (8 ml) were collected at timed intervals and quenched by mixing with 12 ml of the loading dye solution (25 mM EDTA, pH 8.0, 95% v/v formamide, 0.01% bromphenol blue). Enzyme activity measurements at low pH were performed in a buffer containing 10 mM Mg-acetate, 50 mM Na-acetate, pH 4.3, and 0.1 mg/ml BSA. Reaction products were separated by denaturing polyacrylamide gel electrophoresis. The gels contained 20% 29:1 acrylamide/bis-acrylamide with 8 M urea in standard tris-borate-edta (TBE) buffer, electrophoresis was performed for 1–2 h at 30 V/cm. Radiolabeled DNA was detected and quantified as described above. A single exponential was fitted to the substrate depletion data yielding the observed rate constant k_{obs}. The k_{obs} values were plotted as an average value from 2–4 experiments i 1 standard error.

Pyrrolocytosine fluorescence measurements

Steady-state fluorescence measurements were acquired on a Fluoromax-3 (Jobin Yvon) spectrofluorimeter equipped with a Xe lamp. Sample temperatures were maintained at 25°C. Emission spectra (440–460 nm) were recorded at an excitation wavelength of 350 nm with excitation and emission bandwidths of 5 nm. The samples contained 1.0–2.0 mM of protein and 0.5 mM of pyrrolocytosine-labeled DNA oligoduplex (Table 1) in a buffer containing 30 mM Mes and 30 mM histidine (pH 6.3). Control spectra for the background correction were collected under identical conditions except that DNA lacking pyrrolocytosine modifications was used instead of the fluorescent DNA. The fluorescence emission value of the corrected spectrum was determined at 450 nm. The fluorescence intensity of the pyrrolocytosine-modified oligoduplex in the presence of the protein was divided by the fluorescence intensity in the absence of the protein. The resultant values (the fold change of pyrrolocytosine fluorescence upon addition of the protein) were plotted as an average value from several independent experiments i 1 standard error.

Models of 5mC, 5hmC and pyC binding

The structures of the N-terminal domain of AspBHI (PDB ID 4oc8, chain A, residues 2-216), C-terminal domain of PvuRts1I (PDB ID 4oq2, chain A, residues 145–290) and the protein-DNA complex of UHRF1 SRA domain (PDB ID 3fde,

chains ADE) where overlayed using Multiprot [29]. This procedure placed the 5mC base of the UHRF1-SRA DNA into the putative binding pockets of AspBHI and PvuRts1I (an overlay based on the MspJI-DNA structure, PDB ID 4r28, placed the 5mC base in a similar position). To remove minor steric clashes, the 5mC nucleotide in the AspBHI pocket was manually moved by 0.5 Å away from the R87 residue, by 1.0 Å from the Y83 residue and by 1.0 Å towards the D71 residue; in the PvuRts1I structure 5mC was moved by 1.1 Å away from the N217 residue and by 1.0 Å towards the W215 residue. In the resultant structures and in the structure of the McrBC DNA binding domain with methylated DNA (PDB ID 3ssc, chains ACD) the 5mC base was converted into a pyrrolocytosine residue using the 'builder' function of PyMOL (The PyMOL Molecular Graphics System, Version 1.4.1 Schrödinger, LLC), using 1.4 Å bond lengths for the C-C and C-N bonds in the 5-atom aromatic ring and a 1.5 Å C-C bond length for the extra-ring methyl group. A similar procedure yielded 5hmC base in the binding pocket of PvuRts1I.

Reactions with CAA

DNA modification with chloroacetaldehyde (CAA) was performed as described by Daujotyte et al. [30]. Briefly, 100 nM radiolabeled DNA was mixed with 2 mM Ecl18kI, BmeDI or YkrI in 20 ml of the binding buffer (40 mM Tris-acetate, pH 8.3 at 25 °C, 5% glycerol, 0.1 mg/ml BSA). Reactions were initiated by adding CAA to a final concentration of 0.5 M and were incubated for 1 h at 37 °C. Modified strand cleavage was performed by adding 100 ml of freshly diluted 1 M piperidine and heating at 90 °C for 30 min. DNA was precipitated with ethanol and resuspended in 8 ml of the loading dye solution (see above). DNA fragments were separated on high resolution denaturing polyacrylamide gels. The markers were generated using the standard A+G (formic acid) Maxam-Gilbert sequencing reactions.

Chemical display of flipped out thymine and 5-methylcytosine

Experiments with thymine-substituted substrates were performed as described by Serva et al. [31]. Briefly, radiolabeled DNA (10 nM) and protein (100 nM) were mixed in the binding buffer (30 mM Mes-histidine, pH 6.3 at 25 °C, 5% glycerol, 0.1 mg/ml BSA, total volume 20 ml). The reactions were initiated by adding $KMnO_4$ to a final concentration of 2 mM, incubated for 5 min at 25 °C and stopped by adding 20 ml of the solution containing 1.5 M Na-acetate (pH 7.0) and 1 M 2-mercaptoethanol. DNA was then precipitated with ethanol, redissolved in 1 M piperidine, heated at 90 °C for 30 min, precipitated with ethanol, dissolved in 8 ml of the loading dye solution, and analyzed on a high-resolution denaturing polyacrylamide gel. 5-methylcytosine oxidation assay followed the same procedures, except that a 20 mM sodium-acetate reaction buffer (pH 4.3 at 25 °C, ref. [32]) was used.

Results

Modification-dependent enzymes

In this study we used three modification-dependent restriction endonucleases: LpnPI, YkrI and BmeDI. LpnPI belongs to the MspJI family and recognizes the DNA sequence 59-CMDG-39(where M is 5mC or 5hmC, D – A, T or G) [33]. It is closely related to the structurally characterized enzyme AspBHI (. 40% identical and , 60% similar amino acid residues, S1A Figure), including nearly complete conservation of the presumed 5(h)mC binding pocket, (Fig. 1C). LpnPI cleaved the cognate oligoduplex 30-M with a rate constant of , 0.2 $min^{2\,1}$, but no cleavage was detected with an equivalent unmethylated oligoduplex 30-C (Fig. 2A). Discrimination between methylated and unmethylated DNA was also observed in electrophoretic mobility shift experiments (EMSA): both LpnPI and the N-terminal LpnPI DNA binding domain (LpnPI-N) formed protein-DNA complexes with methylated DNA at much lower protein concentrations than with unmethylated DNA (Fig. 2B,C). Noteworthy, the discrimination of specific (methylated) vs non-specific DNA by LpnPI was stronger at pH 6.3 and less pronounced at pH 8.3 (Fig. 2D). This indirectly supports the model of extrahelical 5mC recognition in the binding pockets of MspJI-like enzymes, which requires protonation of the conserved aspartate (D103 in MspJI, D71 in both AspBHI and LpnPI, Fig. 1B-C) [19,22].

The YkrI and BmeDI display significant sequence similarities to the structurally characterized PvuRts1I-like family members PvuRts1I and AbaSI (S1B Figure). An optimal substrate for the PvuRts1I family enzymes consists of two 5hmC or 5ghmC nucleotides in the opposite DNA strands separated by a 20–22 bp DNA fragment [14,15]. Current biochemical and structural data indicate that the 5(g)hmC sites are recognized by the DNA binding domains, while the two nuclease domains form a dimer and perform DNA cleavage at the center of the connecting DNA fragment, i. e. , 11 nt from each modified base [21,23]. Replacement of one 5hmC with a 5-methylcytosine, cytosine and a non-cytosine bases on a series of 39 bp substrates (oligoduplexes 39-H/H, 39-M/H, 39-C/H and 39-H respectively, Table 1) did not abolish their cleavage by YkrI and BmeDI, but decreased the reaction rate (Fig. 3A), suggesting that even a single DNA binding domain is enough to anchor the enzyme dimer to DNA via a 5hmC base. In this case the second YkrI/BmeDI DNA binding domain presumably makes contacts to the base located , 20 bp downstream of the 5hmC nucleotide, and contributes to the enzyme-DNA complex stability depending on the structural similarity of the contacted base to 5(g)hmC. The cleavage data for YkrI is also complemented by EMSA experiments that show a gradual reduction in the amount of the specific enzyme-DNA complex as the second 5hmC in the optimal substrate is replaced with a 5-methylcytosine, cytosine and a non-cytosine base (Fig. 3B). However, we were unable to demonstrate such differences in binding for BmeDI (Fig. 3C).

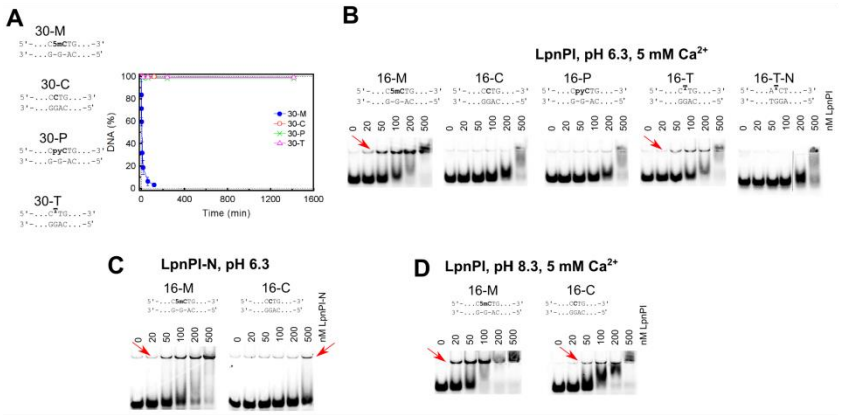


Figure 2. DNA cleavage and binding by LpnPI. The sequences in all panels depict recognition sites in the oligoduplex substrates. (A) DNA cleavage experiments. The reactions were performed with 500 nM enzyme (monomer) and 400 nM substrate at 25°C. Time courses of the reactions are shown. The reaction rate constant for the 30-M substrate equals 0.20 ± 0.05 min^{-1}. Reaction rate constants for other substrates were lower than 16 10^{-5} min^{-1}. (B) Electrophoretic mobility shift assay with LpnPI. DNA binding experiments were performed in a pH 6.3 buffer in the presence of 5 mM Ca^{2+} ions. The final substrate concentration was 10 nM, LpnPI concentrations (in terms of monomer) are indicated above the gel lanes. Red arrows mark the location of the protein-DNA complexes. (C) Electrophoretic mobility shift assay with LpnPI DNA binding domain (LpnPI-N). Experiments with the cognate (16-M) and non-cognate (16-C) substrates were performed in a pH 6.3 buffer in the absence of Ca^{2+} ions. (D) Electrophoretic mobility shift assay with LpnPI in a pH 8.3 buffer in the presence of Ca^{2+} ions.

doi:10.1371/journal.pone.0114580.g002

Pyrrolocytosine fluorescence measurements

Pyrrolocytosine (pyC) is a fluorescent cytosine analog that forms a stable base pair with guanine. The quantum yield of pyC fluorescence is sensitive to base unstacking [34], therefore pyC fluorescence measurements can be used to test the structural environment of a pyC base in nucleic acids and protein-DNA complexes [35, 36]. Notably, pyC fluorescence was used to confirm base flipping in solution by the DNA binding domain of the McrBC enzyme (McrB-N) [13]. However, similar experiments with PvuRts1I were inconclusive, as no increase in pyC fluorescence was observed, despite the fact that PvuRts1I cleaved the pyC-modified DNA [20].

An obvious prerequisite for binding and flipping of the pyC base is the ability of the enzyme to accommodate the flipped out base in the protein pocket. As this may be hindered due to the extra size of pyC in comparison to the unmodified cytosine or 5mC/5hmC bases, we first tested the ability of LpnPI, YkrI and BmeDI to bind and cleave pyrrolocytosine-modified DNA.

Upon replacement of 5mC with a pyC (substrate 16-P), the LpnPI and LpnPI-N binding to DNA became indistinguishable from the unmethylated DNA (Fig. 2B). The same was true for the pyC DNA cleavage (Fig. 2A). For YkrI and

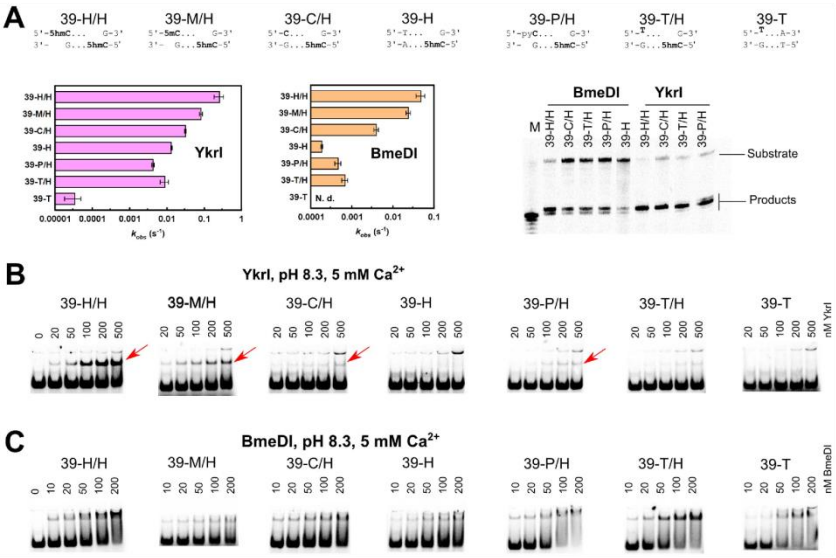


Figure 3. DNA cleavage and binding by YkrI and BmeDI. The sequences at the top of the figure schematically depict the 39-H/H (optimal substrate with two 5hmC bases), 39-M/H, 39-C/H, 39-H, 39-P/H, 39-T/H, and 39-T (one or both 5hmC-G base pairs replaced with a 5mC-G, C-G, T-A, pyrrolocytosine-G, and thymine-G base pairs, Table 1) oligoduplexes. (A) The observed first-order DNA cleavage rate constants. Cleavage reactions were performed with 1 nM substrate and 100 nM enzyme (monomer) at 15°C. In our experimental setup, BmeDI cleavage of the 39-T oligoduplex was not detectable. Denaturing PAGE analysis of cleavage products formed with various DNA substrates is shown on the right-hand side. Gel lane 'M' contained a synthetic single-stranded oligonucleotide that corresponds to cleavage of the bottom strand 11 nt downstream of the 5hmC nucleotide. (B) Electrophoretic mobility shift assay with YkrI. DNA binding experiments were performed in a pH 8.3 buffer in the presence of 5 mM Ca^{2+} ions. The final DNA concentration was 1 nM, and YkrI concentrations are indicated above the gel lanes. Red arrows mark the location of the specific YkrI-DNA complexes. The upper band corresponds to the low-mobility non-specific YkrI-DNA complex formed due to binding/aggregation of multiple protein molecules. (C) Electrophoretic mobility shift experiments with BmeDI. Reaction conditions were as in panel (B).

doi:10.1371/journal.pone.0114580.g003

BmeDI we used a derivative of the optimal substrate with one 5hmC replaced with a pyC (substrate 39-P/H, Table 1). DNA binding and cleavage experiments performed with both enzymes indicated that the replacement of one 5hmC with a pyC compromises enzyme binding and activity to a similar extent as the replacement with an unmodified cytosine or a non-cytosine base (Fig. 3). We therefore conclude that pyC is a poor 5mC/5hmC substitute for LpnPI, YkrI and BmeDI. Indeed, modeling of the pyC base into the presumed binding pocket of AspBHI and PvuRts1I, preserving the H-bonding interactions with the conserved polar pocket residues, results in steric clashes (S2A Figure). Not surprisingly, none of the proteins used in our study (LpnPI, YkrI and BmeDI) triggered fluorescence change of the pyC-containing DNAs (S2B Figure). In contrast, almost no clashes

are observed when pyC is modeled into the pocket of the DNA binding domain of McrBC (S2A Figure), which readily binds pyC DNA and extrudes the modified base from the double helix [13].

Reactions with chloroacetaldehyde

Chloroacetaldehyde (CAA) is known to react with unpaired cytosine and adenine bases in DNA yielding 3, N4-ethenocytosine and 1, N6-ethenoadenine [37]. Such modified residues can be detected by piperidine-induced strand cleavage. So far, the suitability of the CAA reaction was demonstrated for mapping of unmodified cytosine flipped out by several DNA cytosine-5 methyltransferases and restriction enzymes [30]. Since CAA also reacts with 5-methylcytosine [38], we asked if the same experimental setup could be used to detect extrahelical 5mC. As a control we used endonuclease Ecl18kI. This base-flipping restriction enzyme recognizes the pseudosymmetric DNA site 59-CCNGG-39 and flips out the nucleotides of the central base pair [39] that become sensitive to CAA modification [30]. Ecl18kI binds DNA oligoduplexes with the central C-G and 5mC-G base pairs with comparable affinity both in the absence and in the presence of CAA (S3A Figure). Nevertheless, enhanced DNA cleavage in the Ecl18kI complex after CAA/piperidine treatment was observed only for the unmodified central cytosine, but not for 5-methylcytosine (S3B Figure). Thus, at least under standard reaction conditions used in our study, CAA can not be used to detect extrahelical 5mC. Therefore, we could only test if YkrI and BmeDI flip out the unmodified cytosine from the suboptimal substrate 39-C/H, which contains an unmodified cytosine base located , 20 bp away from the 5hmC base (Table 1). We rationalized that while one DNA binding domain of the dimeric enzyme is engaged in a high affinity interaction with the 5hmC site, the second DNA binding domain may interrogate the base , 20 bp downstream, in this case a cytosine, and this process may involve base flipping. This is supported by the observation that both YkrI and BmeDI cleave the 39-C/H substrate faster than the 39-H substrate, which lacks a cytosine base , 20 nt downstream of the 5hmC (Fig. 3). However, neither YkrI nor BmeDI increased the reactivity of the target cytosine in the 39-C/H duplex with CAA (data not shown). Among other reasons for the lack of cytosine reactivity (no flipping, insufficient life-time of the flipped out base, inactivation of the enzyme due to CAA treatment) is the mechanism for the extrahelical 5mC/5hmC recognition by the SRA domains. In the solved co-crystal structures of the UHRF1 SRA domain and the MspJI REase, the Watson-Crick edge of the flipped out 5mC makes hydrogen bonds to the pocket residues (Fig. 1A-B) [8–12, 19]. Conserved residues capable of hydrogen-bonding interactions with the Watson-Crick edge of cytosine derivatives are also present in both PvuRts1I (N217, E228, Fig. 1C) and AbaSI (N236, E247), and are conserved in YkrI/BmeDI (S1B Figure), suggesting that a cytosine base, had it been extruded from the DNA double helix, would be shielded from CAA due to the hydrogen-bonding interactions with the protein. In sharp contrast, Ecl18kI flips out both purine and pyrimidine bases, and binds them in a cavernous protein pocket without forming any base-specific

contacts [39]; this may explain the efficiency of CAA modification of the extrahelical cytosine in the Ecl18kI-DNA complex [30].

Permanganate oxidation of extrahelical pyrimidines

Under acidic conditions potassium permanganate oxidizes both thymine and 5-methylcytosine [32]. However, at physiologic pH this reaction is limited primarily to thymine. $KMnO_4$ treatment leads to conversion of pyrimidine bases to 5,6-dihydroxy-5,6-dihydropyrimidines [40]; the oxidized bases undergo further degradation leading to cleavage of the phosphodiester backbone upon piperidine treatment. Since the oxidation reaction of the C55 C6 bond requires an access to the side of pyrimidine ring that is hidden in the double-stranded DNA, thymines and 5-methylcytosine in DNA helix are relatively resistant to permanganate oxidation compared to extrahelical pyrimidines. $KMnO_4$ was used to detect flipped-out thymines for cytosine and adenine DNA methyltransferases and a sequence-specific transposase [31, 41, 42]. We asked if the permanganate oxidation assay could help detect base-flipping by the modification-dependent restriction enzymes.

Since the $KMnO_4$ assay at near-neutral pH works only with the thymine bases, we made 5mC/5hmC to thymine replacements in the standard LpnPI and PvuRts1I family substrates, thereby creating oligoduplexes with T-G mispairs (Table 1). EMSA experiments confirmed that LpnPI specifically binds the T-G mismatch substrate 16-T, albeit less tightly than the standard methylated duplex 16-M (Fig. 2B). However, we were unable to detect any T-substituted substrate cleavage by LpnPI, both in the standard reaction buffer (Fig. 2A) and under conditions mimicking the EMSA experiment (data not shown). On a control oligoduplex containing the T-G mismatch in a different sequence context (oligoduplex 16-T-N), we observed neither specific binding nor cleavage (Fig. 2B). Replacement of a single 5hmC base in the optimal YkrI/BmeDI oligoduplex 39-H/H with a thymine (substrate 39-T/H, Table 1) decreased the binding and cleavage of the substrate to a similar extent as the 5hmC-to-cytosine or the 5hmC-to-non-cytosine replacements (substrates 39-C/H and 39-H, Fig. 3A-B), but did not change the cleavage position (BmeDI cleaves all substrates 11-12 nt, YkrI – 12 nt downstream from 5hmC, Fig. 3A); no YkrI binding was observed with the 'non-cognate' thymine-substituted oligoduplex 39-T (Fig. 3B). Taken together, the T-G mismatch is a poor substitute for a normal 5(h)mC-G base pair for all enzymes used in our study. A primary reason for this presumably is the direct read-out of the target base: the pockets for extrahelical base binding in the SRA domains, and their homologs in the MspJI/PvuRts1I-like enzymes, are optimized for the specific hydrogen-bonding interactions with cytosine derivatives, but not thymine (Fig. 1).

Surprisingly, incubation of the T-substituted substrates with LpnPI-N, LpnPI, YkrI, and BmeDI significantly increased the susceptibility of the mispaired thymine to $KMnO_4$ oxidation. A particularly strong enhancement in reactivity was observed with the 'cognate' mismatch substrate 16-T and LpnPI-N (Fig. 4A).

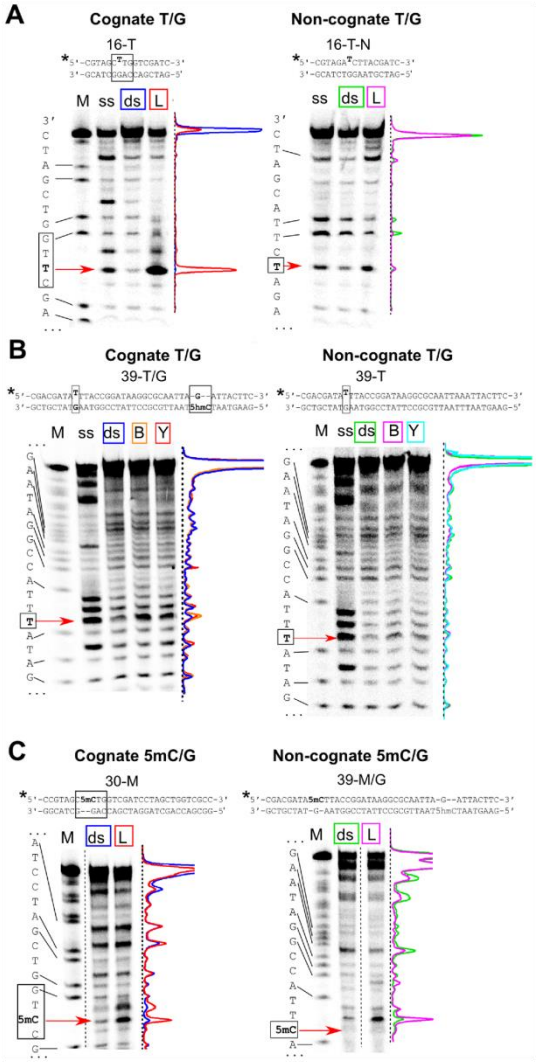

A
Cognate T/G
16-T
Non-cognate T/G
16-T-N
M ss ds L
ss ds L
B
Cognate T/G
39-T/G
Non-cognate T/G
39-T
M ss ds B Y
C
Cognate 5mC/G
30-M
Non-cognate 5mC/G
39-M/G
M ds L
5mC

Figure 4. Permanganate reactivity of pyrimidine bases in the protein-DNA complexes. Sequences at the top of the panels schematically depict the 'cognate' and the 'non-cognate' thymine-substituted substrates, the asterisk marks the ^{33}P radiolabel. Base pairs important for specific binding are in black boxes. Positions of the mispaired thymine and the 5-methylcytosine are marked with red arrows. (A) Thymine oxidation by $KMnO_4$ with or without LpnPI-N. 'M', the A+G markers of the radiolabeled 'cognate' and 'non-cognate' strands; 'ss', oxidation of the single-strand oligonucleotides; 'ds', double-stranded 16-T (cognate) and 16-T-N (noncognate) oligoduplexes without the protein; 'L', 16-T and 16-T-N oligoduplexes + LpnPI-N. Density profiles of individual lanes are shown: cognate DNA (blue), cognate DNA + LpnPI-N (red), non-cognate DNA (green), and non-cognate DNA + LpnPI-N (magenta). (B) Thymine oxidation by $KMnO_4$ in the presence of YkrI and BmeDI. 'M', the A+G markers of the radiolabeled 'cognate' and 'non-cognate' strands; 'ss', oxidation of the single-strand oligonucleotides; 'ds', 39-T/H (cognate) and 39-T (noncognate) oligoduplexes without the protein; 'Y', 39-T/H and 39-T oligoduplexes + YkrI; 'B', 39-T/H and 39-T oligoduplexes + BmeDI. Density profiles: cognate DNA (blue), cognate DNA + BmeDI and YkrI (orange and red, respectively), non-cognate DNA (green), non-cognate DNA + BmeDI and YkrI (magenta and cyan, respectively). (C) 5-methylcytosine oxidation by $KMnO_4$ at pH 4.3 with or without LpnPI-N. 'M', the A+G markers of the radiolabeled 'cognate' and 'non-cognate' strands; 'ds', double-stranded 30-M (cognate) and 39-M/H (noncognate) oligoduplexes without the protein; 'L', 30-M and 39-M/H oligoduplexes + LpnPI-N. Density profiles of individual lanes are colored as in panel (A).

doi:10.1371/journal.pone.0114580.g004

Since the increase was not detectable with the 'non-cognate' T-G oligoduplex 16-T-N, we attribute the hyper-reactivity of the target thymine to the change in its environment induced by the specific binding of the modification-dependent enzyme. The increase in thymine reactivity of the 'cognate' T-G substrate 39-T/H upon its incubation with YkrI and BmeDI was less pronounced, but still clearly detectable (Fig. 4B). Almost no hyper-reactivity of the mispaired thymine was observed with the 'non-cognate' T-G oligoduplex 39-T (Fig. 4B), again implying that the change in the mispaired thymine environment observed with the 'cognate' thymine-substituted DNA was due to the specific enzyme interaction with the DNA.

The major drawback of the above assay is that it makes use of a mutated substrate. To strengthen the evidence for native modified cytosine flipping, we also performed $KMnO_4$ reactions on native 5mC-containing substrates. However, the $KMnO_4$ reacts with 5mC only at non-physiological pH (4.3), where the protonation state of both protein and DNA bases may interfere with protein function and the stability of the protein-DNA complex (nevertheless, examples of DNA-protein interaction studies performed at a very wide range of pH values are present in literature, e. g., ref. [43]). Though none of the enzymes used in our study showed any catalytic activity at pH 4.3, incubation of LpnPI-N with the cognate substrate 30-M, containing 5mC in the context of the LpnPI recognition site (Table 1), resulted in a significant enhancement of 5mC sensitivity to permanganate (Fig. 4C). This signal seems to be both LpnPI- and 5mC-specific, as control experiments performed on the 39-M/H oligoduplex containing 5mC in a different sequence context showed no enhancement in 5mC reactivity (though increase in reactivity was observed for some thymine residues, Fig. 4C). When the 39-M/H oligoduplex (contains one 5mC and one 5hmC separated by 21 bp, Table 1), was incubated with YkrI and BmeDI, no changes in 5mC reactivity were observed (S4 Figure). Taken together, permanganate reactions are consistent with thymine (both LpnPI and YkrI/BmeDI) and 5-methylcytosine (LpnPI) flipping by

the cytosine modification-dependent restriction enzymes, but the results are obtained either on a mutated substrate or at non-physiological pH.

Discussion

In the present study we employed three established fluorescence and chemical-display based methods to test base flipping by the modification-dependent restriction endonucleases. Our results indicate that each assay has its strengths and limitations, and neither of them suits all 5mC/5hmC-binding proteins. For example, the replacement of the target base with the fluorescent base pyrrolocytosine is a convenient method to detect base flipping by measuring the changes in pyrrolocytosine fluorescence intensity upon protein binding. In particular, it was successfully applied for the study of the modification-dependent enzyme McrBC [13]. However, for the method to work, the protein in question must have an adequately sized protein pocket to accommodate the bulky pyrrolocytosine base. While this seems to be the case with the DNA binding domain of McrBC, modeling, DNA binding and cleavage studies indicate that the SRA-like domains of MspJI- and PvuRts1I-like enzymes used in our study do not tolerate such a replacement (Figs. 2, 3 and S2 Figure). The second assay tested in our study, chloroacetaldehyde modification [30], does not require non-natural base substitutions, but under conditions used in our study it worked only with unmodified cytosine (S3 Figure).

The third base-flipping assay tested in our study makes use of the hyper-sensitivity of extrahelical pyrimidines to $KMnO_4$ oxidation. At near-neutral pH this method required replacement of the 5mC/5hmC bases with thymines, thereby forming T-G mismatches. Due to the perturbed geometry of the mismatched base pairs, the unpaired thymines can themselves become hyper-reactive [44]. Fortunately, relatively low background signal observed in our experiments indicates that the accessibility of the mismatched thymines to the bulk solvent is limited, even though they are flanked from both sides with pyrimidine bases (oligoduplexes 16-T and 39-T/H, Table 1). Though similar in size to both 5mC and 5hmC, thymine was poorly tolerated by the modification-dependent enzymes used in our study, resulting in impaired binding and cleavage of the thymine-substituted substrates (Figs. 2 and 3). Surprisingly, incubation of the thymine-substituted DNA with these proteins resulted in hyper-sensitivity of the mismatched thymine to $KMnO_4$ oxidation. A potential risk of using a mismatched oligoduplex is that the mismatch may induce additional conformational flexibility at or in the vicinity of the mispair that upon binding of a protein may result in stronger conformational changes as compared with those in a standard double-stranded DNA. However, the increase in reactivity in our experiments was localized to the target thymine, and was observed only with the 'cognate' thymine-substituted substrates (Fig. 4), implying formation of native-like protein-DNA complexes with a flipped-out thymine.

Structural studies and modeling suggest that the flipped-out base in the protein pocket of the SRA-like domains is sandwiched between conserved polar, aromatic and hydrophobic residues (Fig. 1). If the mispaired thymine occupies the same position as the 5mC/5hmC residues, it should be shielded from $KMnO_4$ oxidation. Instead, the hyper-reactivity of the thymines suggests that the thymine base may be in a dynamic equilibrium between the intra-helical state and an ensemble of flipped-out states. Complete flipping and trapping of the thymine in the protein pocket, most likely, is hindered by the failure of thymine to form hydrogen bonds with the pocket residues that are tailored for the direct read-out of the cytosine derivatives (Fig. 1); on the other hand, opening of the T-G mismatch is much more easily achieved than the disruption of the native C-G base pair, thereby shifting the equilibrium towards the extrahelical states. Interestingly, the recently solved co-crystal structure of the MspJI-DNA complex [19] revealed both specific base-flipping of a 5mC base by one DNA recognition domain, and non-specific flipping of a guanine residue by another DNA binding domain. In the latter case the extruded base occupies a slightly different position than 5mC, and the binding pocket remains in the more open conformation. Formation of such complex was interpreted as a possible target site search intermediate [19]. The extruded thymine detected in our study with LpnPI/YkrI/BmeDI could also occupy a similar non-native position that would permit the reaction with $KMnO_4$.

To strengthen the evidence for native modified cytosine flipping by the modification-dependent enzymes, we also probed the permanganate oxidation of 5mC bases at pH 4.3. Due to non-native conditions, this method was previously applied only for detection of 5mC in DNA [32]. Surprisingly, this assay revealed a significant sequence- and 5mC-specific 'positive' signal for LpnPI, which is consistent with 5mC flipping by this enzyme (whether the flipped-out base at such pH occupies the same position as in the native complex remains unknown). No signal was observed with YkrI and BmeDI, the most likely reason for the lack of the signal being the non-physiological pH, which may interfere with DNA binding by most proteins. Nevertheless, the $KMnO_4$ assay at low pH may work with some proteins, and therefore can prove useful in the studies of the 5mC recognition mechanism of other modified cytosine 'readers'.

Taken together, our study demonstrates the usefulness and limitations of several solution-based methods for the detection of flipped-out cytosine and its derivatives. Only one of the tested methods – permanganate oxidation of the extrahelical pyrimidines – provided evidence for base flipping by the modification-dependent restriction enzymes, implying that the lack of the 'positive signal' in one or even several assays does not exclude base flipping. Therefore, the DNA recognition mechanism of potential 5mC/5hmC-binding proteins should be tested using a combination of all available methods. Nevertheless, the final proof or disproof for the base flipping mechanism would still require high resolution structures of protein-DNA complexes.

Supporting Information

S1 Figure. Modification-dependent endonucleases used in the study. (A) Alignment of the MspJI family member LpnPI with the structurally characterized enzyme AspBHI. Numbering of AspBHI secondary structure elements is taken from [22]. (B) Alignment of the PvuRts1I family members YkrI and BmeDI with the structurally characterized enzymes AbaSI and PvuRts1I. Numbering of AbaSI secondary structure elements is taken from [21]. In both panels green squares mark residues forming the walls of the putative flipped-out base binding pocket; black triangles mark pocket residues that are predicted to contact the Watson-Crick edge of the flipped-out base; stars mark the catalytic centers. The figure was generated with ESPript [46].
doi:10.1371/journal.pone.0114580.s001 (TIF)

S2 Figure. Experiments with pyrrolocytosine-substituted DNA. (A) The models of pyrrolocytosine base in the protein binding pockets of AspBHI, PvuRts1I, and the DNA binding domain of McrBC (see Materials and Methods for details). The black and green lines mark the boundaries of the protein pockets (cut at the plane of the cytosine ring) and the pyrrolocytosine base, respectively. Only the McrBC domain accommodates the pyrrolocytosine base without steric clashes. (B) Pyrrolocytosine fluorescence measurements performed with LpnPI, YkrI and BmeDI. The graphs show the ratio of the cognate pyrrolocytosine-substituted oligoduplex (16-P for LpnPI and LpnPI-N, 39-P/H for YkrI and BmeDI) fluorescence intensity in the presence of the protein to the fluorescence intensity of the same oligoduplex in the absence of the protein.
doi:10.1371/journal.pone.0114580.s002 (TIF)

S3 Figure. The chloroacetaldehyde modification assay with Ecl18kI restriction enzyme. Sequences at the top of the panel A schematically depict the 31-C (central base pair C-G) and 31-M (central base pair 5mC-G) oligoduplexes; the asterisk marks the radiolabel. (A) Ecl18kI binding to DNA oligoduplexes 31-C and 31-M in the pH 8.3 binding buffer in the presence of 5 mM Ca^{2+}. Final DNA concentration was 100 nM. Samples in gel lanes 'C' contained 1000 nM enzyme (dimer) and 500 mM CAA. Red arrows mark the position of the specific protein-DNA complexes. (B) DNA modification with CAA in the presence and in the absence of Ecl18kI. Red arrows mark the position of the central cytosine or 5-methylcytosine. Lanes 'M', the A+G markers of the radiolabeled strands; 'ds', 31-C (unmodified cytosine) and 31-M (5mC) oligoduplexes without the protein; 'E', 31-C and 31-M oligoduplexes + Ecl18kI. The normalized density profiles of individual lanes are shown at the bottom of the panel: 'ds' (blue), 'E' (red).
doi:10.1371/journal.pone.0114580.s003 (TIF)

S4 Figure. 5-methylcytosine oxidation by $KMnO_4$ at pH 4.3 with or without YkrI and BmeDI. Sequence at the top of the image schematically depict the substrate, the asterisk marks the ^{33}P radiolabel. Base pairs 5mC-G and 5hmC-G important for specific binding are in black boxes. Position of the 5-methylcytosine in the autoradiograph is marked with a red arrow. 'M', the A+G marker of the

radiolabeled substrate strand; 'ds', double-stranded 39-M/H oligoduplex without the protein; 'Y', 39-M/H oligoduplex + YkrI; 'B', 39-M/H oligoduplex + BmeDI. The normalized density profiles of individual lanes are shown: 39-M/H DNA (blue), 39-M/H DNA + BmeDI and YkrI (orange and red, respectively).
doi:10.1371/journal.pone.0114580.s004 (TIF)

Acknowledgments

The authors are grateful to prof. Saulius Klimašauskas for discussions and suggestions, Kotryna Kauneckaitė for help with DNA cleavage experiments and dr. Gintautas Tamulaitis for the Ecl18kI sample.

Author Contributions

Conceived and designed the experiments: EZ GS. Performed the experiments: EZ GS. Analyzed the data: EZ GS. Contributed reagents/materials/analysis tools: EZ GS. Wrote the paper: GS.

References

1. Irier HA, Jin P (2012) Dynamics of DNA methylation in aging and Alzheimer's disease. DNA Cell Biol 31 Suppl 1: S42–8.
2. Hahn MA, Szabó PE, Pfeifer GP (2014) 5-Hydroxymethylcytosine: A stable or transient DNA modification? Genomics doi: 10.1016/j.ygeno.2014.08.015.
3. Ohki I, Shimotake N, Fujita N, Jee J, Ikegami T, et al. (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell 105: 487–497.
4. Scarsdale JN, Webb HD, Ginder GD, Williams DC (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. Nucleic Acids Res 39: 6741–6752.
5. Otani J, Arita K, Kato T, Kinoshita M, Kimura H, et al. (2013) Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. J Biol Chem 288: 6351–6362.
6. Ho KL, McNae IW, Schmiedeberg L, Klose RJ, Bird AP, et al. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. Mol Cell 29: 525–531.
7. Buck-Koehntop BA, Stanfield RL, Ekiert DC, Martinez-Yamout MA, Dyson HJ, et al. (2012) Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc Natl Acad Sci U S A 109: 15229–15234.
8. Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455: 818–821.
9. Avvakumov G V, Walker JR, Xue S, Li Y, Duan S, et al. (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455: 822–825.
10. Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, et al. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455: 826–829.
11. Zhou T, Xiong J, Wang M, Yang N, Wong J, et al. (2014) Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol Cell 54: 879–886.
12. Rajakumara E, Law JA, Simanshu DK, Voigt P, Johnson LM, et al. (2011) A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes Dev 25: 137–152.

13. Sukackaite R, Grazulis S, Tamulaitis G, Siksnys V (2012) The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. Nucleic Acids Res 40: 7552–7562.

14. Szwagierczak A, Brachmann A, Schmidt CS, Bultmann S, Leonhardt H, et al. (2011) Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine. Nucleic Acids Res 39: 5149–5156.

15. Wang H, Guan S, Quimby A, Cohen-Karni D, Pradhan S, et al. (2011) Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. Nucleic Acids Res 39: 9294–9305.

16. Huang X, Lu H, Wang J-W, Xu L, Liu S, et al. (2013) High-throughput sequencing of methylated cytosine enriched by modification-dependent restriction endonuclease MspJI. BMC Genet 14: 56.

17. Sun Z, Terragni J, Borgaro JG, Liu Y, Yu L, et al. (2013) High-Resolution Enzymatic Mapping of Genomic 5-Hydroxymethylcytosine in Mouse Embryonic Stem Cells. Cell Rep 3: 567–576.

18. Horton JR, Mabuchi MY, Cohen-Karni D, Zhang X, Griggs RM, et al. (2012) Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease. Nucleic Acids Res 40: 9763–9773.

19. Horton JR, Wang H, Mabuchi MY, Zhang X, Roberts RJ, et al. (2014) Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. Nucleic Acids Res: doi: 10.1093/nar/gku871.

20. Kazrani AA, Kowalska M, Czapinska H, Bochtler M (2014) Crystal structure of the 5hmC specific endonuclease PvuRts1I. Nucleic Acids Res 42: 5929–5936.

21. Horton JR, Borgaro JG, Griggs RM, Quimby A, Guan S, et al. (2014) Structure of 5-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA. Nucleic Acids Res 42: 7947–7959.

22. Horton JR, Nugent RL, Li A, Mabuchi MY, Fomenkov A, et al. (2014) Structure and mutagenesis of the DNA modification-dependent restriction endonuclease AspBHI. Sci Rep 4: 4246.

23. Shao C, Wang C, Zang J (2014) Structural basis for the substrate selectivity of PvuRts1I, a 5-hydroxymethylcytosine DNA restriction endonuclease. Acta Crystallogr D Biol Crystallogr 70: 2477–2486.

24. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, et al. (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell 152: 1146–1159.

25. Iurlaro M, Ficz G, Oxley D, Raiber E-A, Bachman M, et al. (2013) A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. Genome Biol 14: R119.

26. Klimašauskas S, Liutkevičiūtė Z (2009) Experimental Approaches to Study DNA Base Flipping. In: Grosjean H, editor. DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution. Landes Bioscience. pp. 37–50.

27. Tamulaitis G, Solonin AS, Siksnys V (2002) Alternative arrangements of catalytic residues at the active sites of restriction enzymes. FEBS Lett 518: 17–22.

28. Sasnauskas G, Halford SE, Siksnys V (2003) How the BfiI restriction enzyme uses one active site to cut two DNA strands. Proc Natl Acad Sci U S A 100: 6410–6415.

29. Shatsky M, Nussinov R, Wolfson HJ (2004) A method for simultaneous alignment of multiple protein structures. Proteins 56: 143–156.

30. Daujotyte D, Liutkevičiute Z, Tamulaitis G, Klimašauskas S (2008) Chemical mapping of cytosines enzymatically flipped out of the DNA helix. Nucleic Acids Res 36.

31. Serva S, Weinhold E, Roberts RJ, Klimasauskas S (1998) Chemical display of thymine residues flipped out by DNA methyltransferases. Nucleic Acids Res 26: 3473–3479.

32. Fritzsche E, Hayatsu H, Igloi GL, Iida S, Kössel H (1987) The use of permanganate as a sequencing reagent for identification of 5-methylcytosine residues in DNA. Nucleic Acids Res 15: 5517–5528.

33. Zheng Y, Cohen-Karni D, Xu D, Chin HG, Wilson G, et al. (2010) A unique family of Mrr-like modification-dependent restriction endonucleases. Nucleic Acids Res 38: 5527–5534.

34. Berry DA, Jung K-Y, Wise DS, Sercel AD, Pearson WH, et al. (2004) Pyrrolo-dC and pyrrolo-C: fluorescent analogs of cytidine and 29deoxycytidine for the study of oligonucleotides. Tetrahedron Lett 45: 2457–2461.

35. Liu C, Martin CT (2001) Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase. J Mol Biol 308: 465–475.

36. Kuznetsov NA, Vorobjev YN, Krasnoperov LN, Fedorova OS (2012) Thermodynamics of the multi-stage DNA lesion recognition and repair by formamidopyrimidine-DNA glycosylase using pyrrolocytosine fluorescence–stopped-flow pre-steady-state kinetics. Nucleic Acids Res 40: 7384–7392.

37. Kuśmierek JT, Singer B (1982) Chloroacetaldehyde-treated ribo- and deoxyribopolynucleotides. 1. Reaction products. Biochemistry 21: 5717–5722.

38. Oakeley EJ, Schmitt F, Jost JP (1999) Quantification of 5-methylcytosine in DNA by the chloroacetaldehyde reaction. Biotechniques 27: 744–746, 748–750, 752.

39. Bochtler M, Szczepanowski RH, Tamulaitis G, Grazulis S, Czapinska H, et al. (2006) Nucleotide flips determine the specificity of the Ecl18kI restriction endonuclease. EMBO J 25: 2219–2229.

40. Hayatsu H, Ukita T (1967) The selective degradation of pyrimidines in nucleic acids by permanganate oxidation. Biochem Biophys Res Commun 29: 556–561.

41. Bischerour J, Chalmers R (2007) Base-flipping dynamics in a DNA hairpin processing reaction. Nucleic Acids Res 35: 2584–2595.

42. Reddy Y V, Rao DN (2000) Binding of EcoP 15I DNA methyltransferase to DNA reveals a large structural distortion within the recognition sequence. J Mol Biol 298: 597–610.

43. Carey J (1988) Gel retardation at low pH resolves trp repressor-DNA complexes for quantitative study. Proc Natl Acad Sci U S A 85: 975–979.

44. Gogos JA, Karayiorgou M, Aburatani H, Kafatos FC (1990) Detection of single base mismatches of thymine and cytosine residues by potassium permanganate and hydroxylamine in the presence of tetralkylammonium salts. Nucleic Acids Res 18: 6807–6814.

45. Hashimoto H, Horton JR, Zhang X, Cheng X (2009) UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. Epigenetics 4: 8–14.

46. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42: W320–4.

APPENDIX 2

Structure-guided sequence specificity engineering of the modification-dependent restriction endonuclease LpnPI

Sasnauskas G, Zagorskaitė E, Kauneckaitė K, Tamulaitiene G and Siksnys V

Nucleic Acids Research, 2015, volume 43, pp. 6144-55
https://doi.org/10.1093/nar/gkv548

6144–6155 *Nucleic Acids Research, 2015, Vol. 43, No. 12* *Published online 22 May 2015*
doi: 10.1093/nar/gkv548

Structure-guided sequence specif city engineering of the modif cation-dependent restriction endonuclease LpnPI

Giedrius Sasnauskas*, Evelina Zagorskaitė, Kotryna Kauneckaitė, Giedre Tamulaitiene and Virginijus Siksnys*

Department of Protein–DNA Interactions, Institute of Biotechnology, Vilnius University, Graiciuno 8, LT-02241 Vilnius, Lithuania

Received February 23, 2015; Revised May 12, 2015; Accepted May 13, 2015

ABSTRACT

The eukaryotic Set and Ring Associated (SRA) domains and structurally similar DNA recognition domains of prokaryotic cytosine modif cation-dependent restriction endonucleases recognize methylated, hydroxymethylated or glucosylated cytosine in various sequence contexts. Here, we report the apo-structure of the N-terminal SRA-like domain of the cytosine modif cation-dependent restriction enzyme LpnPI that recognizes modif ed cytosine in the 5′-C(mC)DG-3′ target sequence (where mC is 5-methylcytosine or 5-hydroxymethylcytosine and D = A/T/G). Structure-guided mutational analysis revealed LpnPI residues involved in base-specif c interactions and demonstrated binding site plasticity that allowed limited target sequence degeneracy. Furthermore, modular exchange of the LpnPI specif city loops by structural equivalents of related enzymes AspBHI and SgrTI altered sequence specif city of LpnPI. Taken together, our results pave the way for specif city engineering of the cytosine modif cation-dependent restriction enzymes.

INTRODUCTION

5-Methylcytosine (5mC) and its oxidized derivatives, primarily 5-hydroxymethylcytosine (5hmC), are epigenetic marks in eukaryotic cells. The readout of these epigenetic marks is mediated by proteins that specif cally recognize modif ed cytosine variants and strictly discriminate against unmodif ed cytosine. Structural studies of eukaryotic 5mC/5hmC binding proteins revealed two different strategies for the modif ed cytosine recognition. Proteins that share the methyl-CpG binding domain as exemplif ed by MBD (methyl binding domain) proteins and a zinc-f nger protein Kaiso, recognize modif ed C in the context of a Watson–Crick base pair (1–4). The SRA (SET and RING-associated) proteins, exemplif ed by UHRF1, UHRF2 and SUVH5 methyl-binding domains, extrude the modif ed base from DNA helix and place it in a protein pocket for discrimination (5–10).

SRA-like domains were recently identif ed in prokaryotes, where they serve as modules for the modif ed cytosine recognition by the modif cation-dependent restriction enzymes that protect host cells from infection by bacteriophages containing methylated, hydroxymethylated or glucosylated DNA. MspJI family enzymes recognize 5mC and 5hmC in various sequence contexts and cut both DNA strands 12/16 nt downstream of the modif ed cytosine. They are arranged as the N-terminal SRA-like domain and the C-terminal PD-(D/E)XK nuclease domain fusions (11–13). PvuRts1I family enzymes recognize DNA containing 5hmC or glucosylated 5-hydroxymethylcytosine (5ghmC). The optimal substrate for PvuRts1I is a DNA duplex with two 5(g)hmC bases in the opposite strands separated by a ∼20 bp fragment of unmodif ed DNA. The cleavage occurs at the center of this fragment, e.g. 11/9 nt away from the modif ed cytosines (14,15). The PvuRts1I and MspJI-family enzymes share PD-(D/E)XK nuclease and SRA-like DNA recognition domains but the domain order is being permuted.

Available data show that the SRA-like fold is a versatile structural module that is used for the recognition of the modif ed cytosine in a different sequence context (Table 1). The modif ed cytosine base is extruded outside the DNA helix and accommodated in a pocket of SRA and SRA-like proteins. Subtle structural and size differences of the protein pocket may account for the discrimination of 5mC/5hmC/5ghmC bases by these domains (8,16–17). The recognition of the modif ed cytosine occurs only in a specif c nucleotide context indicating tight coupling between the base f ipping and recognition of the surrounding sequence. Structural and molecular mechanisms of sequence

*To whom correspondence should be addressed. Tel: +370 5 2602111; Fax: +370 5 2602116; Email: gsasnaus@ibt.lt
Correspondence may also be addressed to Virginijus Siksnys. Tel: +370 5 2602108; Fax: +370 5 2602116; Email: siksnys@ibt.lt

Table 1. Structurally characterized SRA domains and their recognition sequences

Protein	Recognition site[a]	Base modification	PDB ID	References
UHRF1-SRA	5′-(mC)G-3′	5mC, 5hmC	2ZO0, 2ZO1, 3CLZ, 2ZKD, 2ZKE, 2ZKD	(5–7,10)
UHRF2-SRA	5′-(mC)G-3′	5hmC > 5mC	4PW5, 4PW6, 4PW7	(8,10)
SUVH5-SRA	5′-(mC)G-3′ 5′-(mC)HH-3′	5mC	3Q0B, 3Q0C, 3Q0D	(9)
MspJI	5′-(mC)NNR-3′	5mC, 5hmC	4R28, 4F0Q, 4F0P	(11–12,18)
AspBHI	5′-YS(mC)NS-3′	5mC, 5hmC	4OC8	(11,13)
LpnPI	5′-C(mC)DG-3′	5mC, 5hmC	4RZL	this work, (11)
PvuRts1I	5′-(mC)-3′	5hmC, 5ghmC	4OQ2, 4OKY	(14–15,17,34)
AbaSI	5′-(mC)-3′	5hmC, 5ghmC	4PAR, 4PBA, 4PBB	(15,16)

[a](mC) – modified cytosine; N – any nucleotide; D – A, T, or G; Y – T or C; S – G or C; R – A or G.

recognition by the SRA-like domains are still poorly understood. In the eukaryotic UHRF1-SRA domain specific for the 5mC residue in the CpG sequence context the 'base flip-promoting' ('thumb') and 'CpG recognition' ('NKR' finger) loops penetrate into, respectively, major and minor DNA grooves, and make discriminating contacts to the CpG dinucleotide (5–7). In the target sites of the MspJI family enzymes the modified C is embedded into a variable, often degenerate nucleotide sequences that span up to 2 bp upstream and 3 bp downstream of the modified cytosine (Table 1). Readout of the 5′-(mC)NNR-3′ sequence by MspJI REase occurs through the minor groove contacts (18), but the DNA recognition mechanism of other MspJI-like enzymes still has to be resolved. In the present study we report the crystal structure of the N-terminal DNA binding domain of the LpnPI restriction endonuclease (LpnPI-N), which recognizes the modified cytosine in the sequence context 5′-C(mC)DG-3′ (where D – A/T/G, Table 1), and provide mutational/loop-swapping analysis that supports the structural model for the sequence recognition. Our findings pave the way for specificity engineering of the modification-dependent restriction endonucleases.

MATERIALS AND METHODS

Protein expression, purification and crystallization

The N-terminal LpnPI DNA binding domain LpnPI-N (residues 2–224 of the full-length protein), wt LpnPI and its mutants were expressed and purified as described in (19). In all constructs the first methionine was replaced with a $(His)_6$-tag (MGHHHHHHG). According to the mass-spec analysis, the purified LpnPI-N protein did not preserve the N-terminal methionine. The yield of mutant LpnPI variants was comparable to that of the wt enzyme (varied within a factor of two). The folding of all proteins at the secondary structure level was very similar, as demonstrated by the far-UV CD spectra of wt LpnPI and mutants (Supplementary Figure S1).

Protein concentrations were estimated spectrophotometrically using extinction coefficients of 35410/M/cm (LpnPI-N), 49850/M/cm (LpnPI and most mutants/loop-swapping variants), and 51340/M/cm (the loop-swapping variant '21AGY'), and are expressed in terms of monomer if not stated otherwise. All extinction coefficients were calculated using the ProtParam tool (http://web.expasy.org/protparam/).

Protein crystallization was performed by sitting drop vapor diffusion method at 291 K. The LpnPI-N protein in 10 mM Tris-HCl (pH 7.5 at 25°C), 200 mM KCl, 0.1 mM EDTA and 0.02% NaN_3 was concentrated to 5.5 mg/ml (~220 M) and mixed with 0.4 volume of the crystallization solution (160 mM $(NH_4)_2SO_4$, 80 mM HEPES pH 7.5, and 20% w/v PEG 3350). Crystals appeared after 5 days and reached the maximum size in 1 month.

Data collection and structure determination

Crystal diffraction data for the LpnPI-N protein were collected at 100 K (no extra cryo-protection used) at the MAX II synchrotron I911-3 beamline on a CCD detector. Data were processed with XDS (20), SCALA and TRUNCATE (21). The structure was solved using the molecular replacement protocol of Auto-Rickshaw (22) and the structure of the AspBHI DNA binding domain (PDB ID 4OC8, chain A, residues 6–22, 31–85, 97–208) as the starting model. Molecular replacement procedure during the Auto-Rickshaw run was performed with MOLREP (23), rigid-body refinement was conducted using CNS (24), density modification was performed using PIRATE (21), model was built with ARP/wARP (25), and structure refinement was performed with REFMAC (26) and PHENIX (27). The obtained model was manually rebuilt using COOT (28) and refined using PHENIX (phenix.refine 1.9_1692) (27). During refinement NCS restraints between the two protein subunits present in the asymmetric unit were used. Data collection and refinement statistics are shown in Table 2.

Structure analysis

The structures of the N-terminal domain of AspBHI (PDB ID: 4OC8, chain A, residues 2–216), C-terminal domain of PvuRts1I (PDB ID: 4OQ2, chain A, residues 145–290), LpnPI-N (PDB ID: 4RZL, chain A), the protein-DNA complex of UHRF1-SRA (PDB ID: 3FDE, chains ADE), the protein-DNA complex of MspJI (PDB ID: 4R28, chain C, residues 8–263), and apo-MspJI (PDB ID: 4F0Q, chain A, residues 8–263) where overlaid using Multiprot (29). The LpnPI-N interfaces in the crystal were analyzed using the PDBePISA web-server (30), alignments were generated with ESPript (31).

Table 2. Data collection and refinement statistics

Data collection statistics	
Space group	$P6_1$
A (Å)	82.037
B (Å)	82.037
C (Å)	152.829
Wavelength	1.0012
X-ray source	MAX II I911-3 beamline
Total reflections	438 186
Unique reflections	33 984
Resolution range (Å)	41.4-2.1
Completeness (%) (last shell)	100 (100)
Multiplicity (last shell)	12.9 (12.8)
I/σ (last shell)	21.5 (5.1)
R(merge) (%) (last shell)	10.1 (52.4)
B(iso) from Wilson (Å^2)	21.68
Refinement statistics	
Resolution range (Å)	41.019–2.10
Reflections work/test	60 566/6585
Protein atoms	3490
Solvent molecules	422
R-factor (%)	16.6
R-free (%)	19.9
R.M.S.D. bond lengths (Å)	0.010
R.M.S.D. angles (°)	1.092
Ramachandran core region (%)	97.56
Ramachandran allowed region (%)	2.44
Ramachandran disallowed region (%)	0

DNA oligonucleotides

All oligonucleotides were purchased from Metabion. Oligoduplex substrates used in this study are listed in Table 3. Oligonucleotides were 5′-labeled with [γ-^{33}P]ATP (Hartmann Analytic) and T4 polynucleotide kinase (Thermo Fisher Scientific). Oligoduplexes were assembled by annealing the corresponding radiolabeled and unlabeled strands.

Mutagenesis

His-tagged full-length LpnPI mutant variants were generated by the QuickChange method (32). The *Escherichia coli* strain ER2566 was used as a transformation host. The mutations were confirmed by DNA sequencing of the entire gene.

DNA cleavage experiments

DNA hydrolysis reactions were performed by manually mixing radiolabeled oligoduplexes (100–400 nM) with the enzyme (500 nM) in the Reaction Buffer (33 mM Tris-acetate, pH 8.0, 66 mM K-acetate, 10 mM Mg-acetate, 0.1 mg/ml BSA) at 25°C. Samples were collected at timed intervals and quenched by mixing with the loading dye solution (25 mM EDTA, pH 8.0, 95% v/v formamide, 0.01% bromphenol blue). Reaction products were separated by denaturing polyacrylamide gel electrophoresis. The gels contained 20% 29:1 acrylamide/bis-acrylamide with 8 M urea in a standard Tris-borate–EDTA (TBE) buffer, electrophoresis was performed for 1–2 h at 30 V/cm. Radiolabeled DNA was detected and quantified using Cyclone phosphorimager and OptiQuant software. A single exponential was fitted to the substrate depletion data yielding the observed rate constant k_{obs}. The k_{obs} values are reported as an average value from two to four experiments ± 1 standard error of the mean (SEM). The minimal cleavage rate detectable in our experimental setup was 3×10^{-7} s^{-1} (corresponds to 2% substrate cleaved during ∼20 h incubation). Wt enzyme cleaved the top (methylated) and the bottom strands of the reference 'gC(mC)TG' substrate (Table 3) with comparable rates (data not shown). Cleavage of most other DNA substrates and LpnPI variants was monitored using DNA substrates with the radiolabeled top strand.

RESULTS

Overall structure of LpnPI-N

LpnPI recognizes 5mC or 5hmC in the sequence context 5′-C-5(h)mC-(A/T/G)-G-3′, and cuts DNA 10/14 nt downstream from the recognition site (11). It shares protein sequence similarities (Supplementary Figure S2) with structurally characterized AspBHI (42/63% identical/similar aa residues for the N-terminal domains) (13) and MspJI (∼15/33% identical/similar aa residues) enzymes of the MspJI family (12,18).

The structure of LpnPI-N was solved at 2.1 Å resolution (Table 2). The asymmetric unit contains two almost identical protein molecules (RMSD < 0.5 Å over 209 C_α atoms), with slightly larger deviations observed only at the N-termini and two flexible loops (residues 24–26 and 51–54). Both protein subunits make similar contacts in the crystal. The largest contact surface (∼750 Å^2) is formed between the N- and C-termini of both A and B subunits that encircle the 160–170 αE-hairpins of the symmetry related A/B subunits (Supplementary Figure S3). This 'pinching' interaction, however, is not functionally important, since LpnPI-N is a monomer in solution (Supplementary Figure S4).

The overall structure of LpnPI-N is very similar to that of the SRA-like DNA binding domain of AspBHI (RMSD 1 Å over 170 CA atoms, Figure 1A and B). This allowed us to solve the LpnPI-N structure by molecular replacement using the AspBHI-N structure as an initial model (see Materials and Methods for details). The most interesting difference between AspBHI-N and LpnPI-N is the length and conformation of the Loop-2B (residues 21–31, correspond to AspBHI residues 22–33) involved in DNA binding (see below). LpnPI-N is more compact than the corresponding domain of MspJI (224 versus 260 aa). The loops connecting αC–αD, αG–αH strands and E–F helices are shorter in LpnPI-N by up to 5 aa; nevertheless LpnPI, like AspBHI, contains an 8 aa insertion in the αH strand that breaks it into two parts (13) (Figure 1A).

DNA recognition determinants of LpnPI

The DNA-bound structures are available for several eukaryotic SRA domains (5–9) and the MspJI restriction endonuclease (18). Since an overlay of LpnPI-N or AspBHI-N with either the UHRF1-SRA or MspJI co-crystal structures places the DNA molecules and the flipped-out 5-methylcytosine bases in a similar position relative to LpnPI/AspBHI, we will further discuss the models of DNA-bound LpnPI-N and AspBHI-N based on the UHRF1-SRA-DNA structure (Figure 1C and D). We will

Table 3. Oligoduplex substrates

Oligoduplex	Sequence[a]	Specification[b]
gC(mC)TG	5 -CCGTAGC**5**TGGTCGATCCTAGCTGGTCGCC-3 3 -GGCATCGGACCAGCTAGGATCGACCAGCGG-5	Oligoduplex with a standard LpnPI recognition site; the reference substrate in DNA cleavage studies
tC(mC)TG	5 -CCGTA***T***C**5**TGGTCGATCCTAGCTGGTCGCC-3 3 -GGCAT***A***GGACCAGCTAGGATCGACCAGCGG-5	As gC(mC)TG, but the -2 bp is T:A
aC(mC)TG	5 -CCGTA***A***C**5**TGGTCGATCCTAGCTGGTCGCC-3 3 -GGCAT***T***GGACCAGCTAGGATCGACCAGCGG-5	The -2 bp is A:T
cC(mC)TG	5 -CCGTA***C***C**5**TGGTCGATCCTAGCTGGTCGCC-3 3 -GGCAT***G***GGACCAGCTAGGATCGACCAGCGG-5	The -2 bp is C:G
gG(mC)TG	5 -CCGTAG***G***5TGGTCGATCCTAGCTGGTCGCC-3 3 -GGCATC***C***GACCAGCTAGGATCGACCAGCGG-5	The -1 bp is G:C
gT(mC)TG	5 -CCGTAG***T***5TGGTCGATCCTAGCTGGTCGCC-3 3 -GGCATC***A***GACCAGCTAGGATCGACCAGCGG-5	The -1 bp is T:A
gA(mC)TG	5 -CCGTAG***A***5TGGTCGATCCTAGCTGGTCGCC-3 3 -GGCATC***T***GACCAGCTAGGATCGACCAGCGG-5	The -1 bp is A:T
gC(mC)AG	5 -CCGTAGC**5*A***GGTCGATCCTAGCTGGTCGCC-3 3 -GGCATCGG***T***CCAGCTAGGATCGACCAGCGG-5	The +1 bp is A:T
gC(mC)CG	5 -CCGTAGC**5*C***GGTCGATCCTAGCTGGTCGCC-3 3 -GGCATCGG***G***CCAGCTAGGATCGACCAGCGG-5	The +1 bp is C:G
gC(mC)GG	5 -CCGTAGC**5*G***GGTCGATCCTAGCTGGTCGCC-3 3 -GGCATCGG***C***CCAGCTAGGATCGACCAGCGG-5	The +1 bp is G:C
gC(mC)TC	5 -CCGTAGC**5*T******C***GTCGATCCTAGCTGGTCGCC-3 3 -GGCATCGGA***G***CAGCTAGGATCGACCAGCGG-5	The +2 bp is C:G
gC(mC)TA	5 -CCGTAGC**5*T******A***GTCGATCCTAGCTGGTCGCC-3 3 -GGCATCGGA***T***CAGCTAGGATCGACCAGCGG-5	The +2 bp is A:T
gC(mC)TT	5 -CCGTAGC**5*T******T***GTCGATCCTAGCTGGTCGCC-3 3 -GGCATCGGA***A***CAGCTAGGATCGACCAGCGG-5	The +2 bp is T:A
gG(mC)TC	5 -CCGTAG***G***5T***C***GTCGATCCTAGCTGGTCGCC-3 3 -GGCATC***C***GA***G***CAGCTAGGATCGACCAGCGG-5	The -2 bp is G:C and the +2 bp is C:G
gT(mC)TT	5 -CCGTAG***T***5T***T***GTCGATCCTAGCTGGTCGCC-3 3 -GGCATC***A***GA***A***CAGCTAGGATCGACCAGCGG-5	The -2 bp is T:A and the +2 bp is T:A
gA(mC)TA	5 -CCGTAG***A***5T***A***GTCGATCCTAGCTGGTCGCC-3 3 -GGCATC***T***GA***T***CAGCTAGGATCGACCAGCGG-5	The -2 bp is A:T and the +2 bp is A:T
gCCTG	5 -CCGTAGC***C***TGGTCGATCCTAGCTGGTCGCC-3 3 -GGCATCGGACCAGCTAGGATCGACCAGCGG-5	As gC(mC)TG, but 5mC is replaced with an unmodified cytosine

[a] '**5**' designates 5-methylcytosine; the DNA regions recognized by LpnPI are underlined; DNA base pairs that deviate from the reference substrate 'gC(mC)TG' are shown in ***bold italic*** typeface.
[b] Base pairs upstream of 5mC are numbered −1 and −2; base pairs downstream of 5mC are numbered +1 and +2.

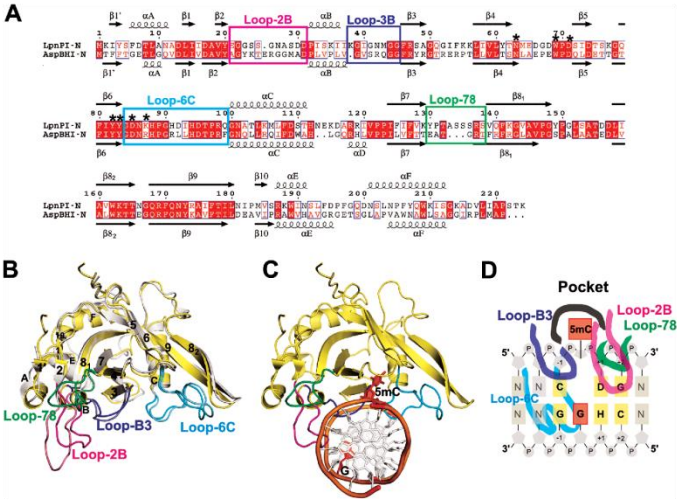


Figure 1. DNA recognition domain of restriction endonuclease LpnPI. (**A**) Sequence alignment of the N-terminal domains of LpnPI (LpnPI-N) and AspBHI (AspBHI-N). Secondary structure elements of LpnPI-N and AspBHI-N are numbered as in (13). Residues that form the flipped-out base binding pocket are marked with asterisks. Alignment was generated with ESPript (31). (**B**) Superimposition of LpnPI-N (in yellow) and AspBHI-N (in white; PDB 4OC8). The putative LpnPI/AspBHI DNA recognition loops are colored as follows: Loop-B3, blue/light blue; Loop-78, green/lime; Loop-2B, magenta/light magenta; Loop-6C, cyan/aquamarine. (**C**) The model of DNA-bound LpnPI-N, based on the crystal structure of DNA-bound UHRF1-SRA (PDB 3FDE). DNA recognition loops are colored as in (B), the flipped cytosine and the orphan intra-helical guanine are shown in red. (**D**) Schematic representation of LpnPI interactions with DNA. Protein loops and the 5mC:G base pair are colored as in panel (C); other bases comprising the LpnPI recognition site are shown in light orange and are numbered from '−1' (the bp upstream of 5mC) to '+2' (the 2nd bp downstream of 5mC).

refer to the DNA base pairs 5′ (upstream) of the flipped cytosine as the '−1' and '−2' positions, and the base pairs 3′ (downstream) of the flipped base as the '+1', '+2' and '+3' positions.

The flipped-out cytosine binding pockets are similar in all SRA domains (Figure 2). The cytosine 5-methyl group in MspJI pocket is in van der Waals distance from D117, Y114 and W101 residues, and apparently makes a weak C–H…O hydrogen bond to the carbonyl oxygen of G116. These interactions may serve to distinguish modified cytosine from an unmodified base (18). Equivalent positions in LpnPI and AspBHI are occupied by D85, Y82, W69, and G84 residues. The side walls of the MspJI pocket are formed by the residues W101, Y114 and D117, while D103, S90 and F115 make hydrogen bonds to the Watson-Crick edge of the modified cytosine. Equivalent residues in LpnPI are W69, Y82, D85, K87, D71, N63 and Y83 (W69, Y82, D85, R87, D71, N63, Y83 in AspBHI). Mutation of the AspBHI pocket residues D71, Y82 and D85 to alanine abolished the DNA cleavage activity (13).

An alanine replacement of D71 in LpnPI had a similar effect: the reaction rate decreased more than 1000-fold (Table 4). A more conservative D71N replacement diminished the LpnPI DNA cleavage rate ∼15-fold (Table 4). All DNA cleavage experiments were performed under the optimal reaction conditions (near equimolar enzyme and DNA concentrations; in agreement with the proposed mechanism for the MspJI reaction, which involves simultaneous interaction of the tetrameric enzyme with up to four cognate DNA molecules (12), the LpnPI reactions were much slower under enzyme excess conditions, Supplementary Figure S5). We presume that the observed changes in the DNA cleavage rates under these reaction conditions are due to the altered DNA binding ability of LpnPI.

In the UHRF1–SRA–DNA complex structure, the vacant space left by the flipped-out base is filled in by the V451 residue from the 'base flipping-promotion' or 'thumb' loop (5–7) (Figure 2); in MspJI–DNA complex, the E65 residue of the structurally equivalent Loop-B3 (loop between helix αB and strand β3) makes a hydrogen bond to the intra-helical orphaned guanine. The Loop-B3 in LpnPI contains residues N42 and M43 (Figure 2B); M43 (Q43 in AspBHI) is the likely candidate to fill the space left by the flipped-out cytosine, while the N42 (R42 in AspBHI) could make contacts to the orphan guanine or the −1 base pair from the minor groove side. In agreement with this model, the N42A mutation rendered the enzyme inactive, the M43Q mutation had little effect on the enzyme activity, and the M43A mu-

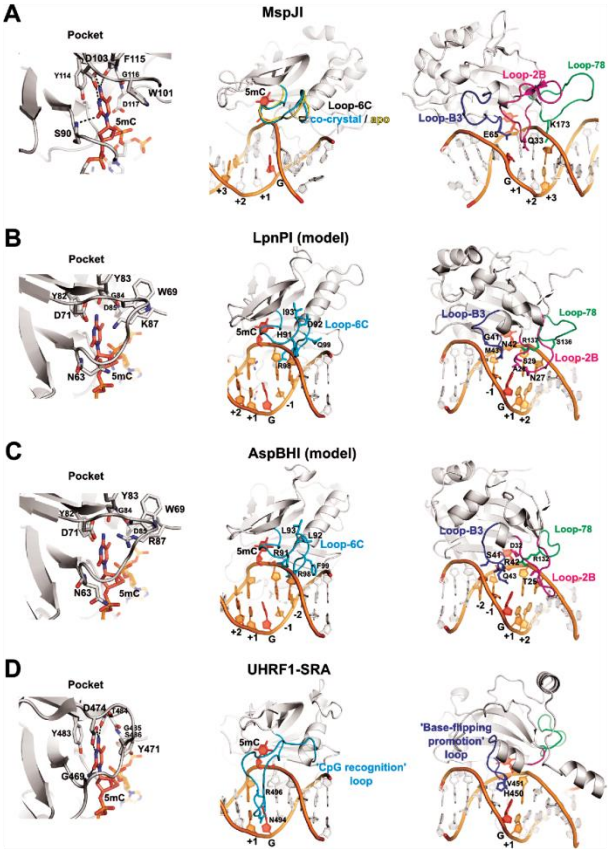


Figure 2. DNA recognition by SRA domains. The structures of the DNA-bound UHRF1-SRA and MspJI, and the apo-structures of MspJI, LpnPI-N and AspBHI-N (PDB ID: 3FDE, 4R28, 4F0Q, 4RZL, 4OC8) were superimposed with MultiProt (29). Equivalent DNA recognition elements in all panels are shown in identical orientation. Left: recognition of the flipped-out base in the protein pocket; center: Loop-6C or 'CpG recognition'/'NKR finger' loop (cyan); right: Loop-B3 or 'base-flipping-promotion' loop (blue), Loop-B2 (magenta), and Loop-78 (green). In all panels the flipped-out base and the orphan intra-helical guanine are colored red; other nucleotides comprising the specific recognition sequence of the corresponding protein are colored orange and are numbered from '−2' (the second bp upstream of 5mC) to '+3' (the third bp downstream of 5mC). (**A**) DNA recognition by MspJI. Loop-6C occupies a similar position both in the apo- and the DNA-bound structures and does not make base-specific contacts. Residues Q33, E65, and K173 from the '2B', 'B3', and '78' loops, respectively, are close to the DNA bases. (**B** and **C**) The models of DNA-bound LpnPI and AspBHI based on the co-crystal structure of UHRF1-SRA. Loop-C6 and Loop-B3 residues 41–43, 91–93, and 99 are different in LpnPI and AspBHI. LpnPI Loop-2B and Loop-78 residues 27–29 and 136–137 were mutated in the present study; AspBHI Loop-2B residues T25 and D32 are critical for the enzyme function (13); AspBHI Loop-78 residue R132 overlaps with the critical LpnPI residue R137. (**D**) DNA recognition by the SRA domain of the eukaryotic UHRF1 protein. The loops equivalent to Loop-78 and Loop-2B in MspJI-like restriction endonucleases are colored green and magenta, respectively.

Table 4. Catalytic activity of LpnPI mutants

Mutation	k_{obs} (s^{-1})[a]	Activity (%)[b]
wt LpnPI	(3.3 ± 0.8) × 10^{-3}	100
5(h)mC binding pocket		
D71A	(1.0 ± 0.3) × 10^{-6}	0.03
D71N	(2.0 ± 0. 4) × 10^{-4}	6
Loop-2B (contacts downstream of 5(h)mC)		
S25A	(7.0 ± 1.5) × 10^{-3}	200
N27A	No cleavage	<0.01
D30A	(1.0 ± 0.3) × 10^{-3}	30
Loop-B3 (adjacent to orphan guanine)		
G41S	(1.0 ± 0.1) × 10^{-5}	0.3
N42A	No cleavage	<0.01
M43A	(1.0 ± 0.2) × 10^{-4}	3
M43Q	(1.6 ± 0.1) × 10^{-3}	50
Loop-6C (contacts upstream of 5(h)mC)		
R98A	(1.0 ± 0.6) × 10^{-6}	0.03
Loop-78 (contacts downstream of 5(h)mC)		
R137A	(0.7 ± 0.3) × 10^{-5}	0.2
S136A	(2.1 ± 0.1) × 10^{-3}	60

[a]Oligoduplex DNA cleavage reactions were performed on the 'gC(mC)TG' substrate (Table 3) and the observed rate constants k_{obs} were determined by single-exponential fits (see Materials and Methods for details). The lowest DNA cleavage rate measured in our assay is 3 × 10^{-7} s^{-1}.
[b]The activity is expressed as the ratio k_{obs}(mutant)/k_{obs}(wt) × 100%.

tation decreased the DNA cleavage rate ∼30-fold (Table 4). The glycine residue equivalent to the G41 in LpnPI is conserved in SgrTI and RlaI, but not in AspBHI, which has a serine at this position (Supplementary Figure S2). Interestingly, the G41S replacement reduced LpnPI activity ∼300-fold (Table 4). Presumably, the glycine residue contributes to Loop-B3 flexibility that is important for the LpnPI function.

In the LpnPI–DNA model, the DNA backbone on the 3′ side of the flipped cytosine is contacted by the Loop-78 (residues 130–137 between αU7 and αN8 strands, Figure 2B). With a little change in a loop conformation, the side chains of S136 and R137 could make base-specific contacts in the major groove 3′ of the extrahelical cytosine; moreover, the K173 residue from an equivalent MspJI loop is the prime candidate for the purine base recognition at the +3 position in the MspJI target sequence (18). DNA cleavage analysis of the alanine replacement mutants S136A and R137A showed that only R137 is critical for LpnPI function (Table 4).

The LpnPI Loop-2B (LpnPI residues 23–31 between αN2 strand and αB helix) is positioned in the minor groove 3′ of the flipped cytosine. Structurally equivalent loops are present in both AspBHI and MspJI, but are much shorter in the eukaryotic SRA domains (Figure 2). To probe the role of Loop-2B residues in LpnPI function we made alanine replacements of the polar residues S25, N27 and D30. Only the N27A mutation abolished LpnPI activity, while the other two mutants displayed wt-like activity (Table 4). This is consistent with N27 playing a role in DNA binding/recognition.

The DNA on the 5′ side of the flipped cytosine is approached by the LpnPI Loop-6C (residues 84–99 in LpnPI/AspBHI, and 116–129 in MspJI). An equivalent 'CpG recognition' or 'NKR finger' loop in the eukaryotic SRA domains is significantly longer (e.g. 484–508 in UHRF1–SRA), adopts a different conformation, and makes base-specific contacts in the DNA major groove (5–7) (Figure 2D). In our LpnPI–DNA model, only the conserved R98 residue of the Loop-6C is within an H-bonding distance to the −1 base. It may also contribute to proper positioning of the adjacent Loop-B3 residues. In agreement with this model, LpnPI mutant R98A was nearly inactive (Table 4).

The sequence specificity of LpnPI

The LpnPI recognition sequence provided in REBASE (33) is 5′-C(mC)DG-3′ (where mC – modified C and D = A/G/T), though 5′-S(mC)DS-3′ or 5′-(mC)DS-3′ (G>>C) sequence specificities have also been reported (11). To analyze the sequence preference of LpnPI, we measured the LpnPI cleavage rates on a set of oligoduplex substrates (Table 3) that differ from the reference 'gC(mC)TG' substrate by 1 or 2 bp. Our results are consistent with LpnPI having a strong preference for the 5′-C(mC)DG-3′ recognition site (cleavage rate constant ∼3 × 10^{-3} s^{-1}), albeit ∼90-, ∼500- and 1500-fold slower DNA cleavage was also observed with DNA sites 5′-G(mC)TG-3′, 5′-C(mC)TC-3′, and 5′-G(mC)TC-3′ (rate constants ∼3 × 10^{-5}, ∼6 × 10^{-6} and 2 × 10^{-6} s^{-1}, respectively, Figure 3A and B). Therefore, the target site for the wt LpnPI can be defined as 5′-(C>>G)(mC)D(G>>C)-3′.

Sequence specificity engineering of LpnPI

SRA-like DNA binding domains of several MspJI family enzymes have closely related protein sequences (Supplementary Figure S2) and recognize target sites partially overlapping with the 5′-C(mC)DG-3′ target of LpnPI, e.g. AspBHI (5′-YS(mC)NS-3′) and SgrTI (5′-C(mC)DS-3′) (11). Since Loop-6C is the likely candidate for the recognition of the −1 or −1/−2 nucleotides ('C' – LpnPI and SgrTI, 'YS' – AspBHI) (Figures 1D and 2B, C), we attempted to alter LpnPI sequence specificity by swapping the LpnPI Loop-6C (residues 91–99) with the equivalent loop of AspBHI (Figure 3A). Our expectation was that the resultant LpnPI variant 'LpnPI-91RLL' would preferentially cleave DNA substrates with a 5′-YS-3′ dinucleotide in the −1/−2 positions. Indeed, LpnPI-91RLL is more tolerant for a G in the −1 position than the wt enzyme (the ratio of cleavage rates between the 5′-C(mC)DG-3′ and 5′-G(mC)DG-3′ substrates dropped from 90- to 1.5-fold). However, LpnPI-91RLL, like the wt enzyme, has no preference for the −2 base pair (Figure 3A). Loop-6C replacement also increased the tolerance for A and T substitutions in the −1 position: both the 5′-A(mC)DG-3′ and 5′-T(mC)DG-3′ substrates are refractory to the wt enzyme, but are slowly cleaved by LpnPI-91RLL (Figure 3A). Surprisingly, the loop replacement also increased tolerance for substitutions in the downstream part of the recognition site, but to a lesser extent: the rate difference for the 5′-C(mC)DG-3′ and 5′-C(mC)DC-3′ substrates decreased from ∼500- to ∼20-fold (Figure 3A). Thus, the relaxed specificity of LpnPI-91RLL for the −1 position at least in part stems from an overall improvement of enzyme binding

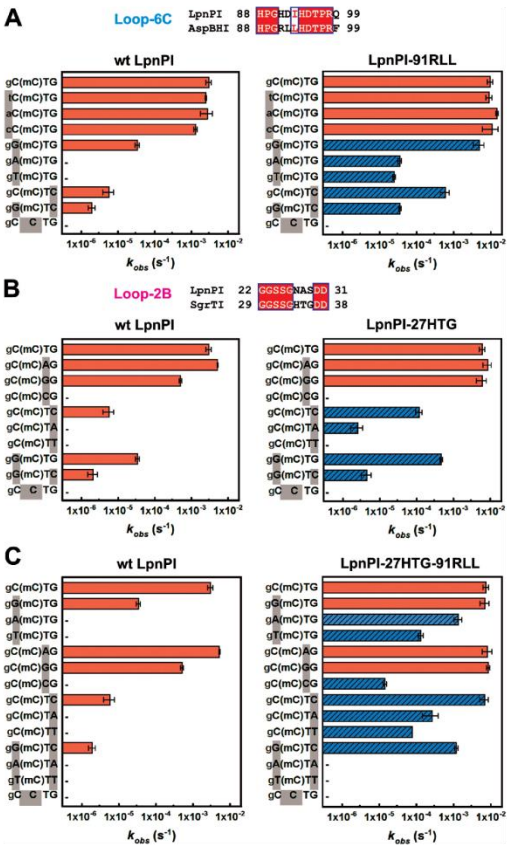


Figure 3. Recognition site preference of LpnPI. Oligoduplex DNA cleavage reactions were performed under standard reaction conditions and the observed rate constants k_{obs} were determined by single-exponential f ts (see Materials and Methods for details). The recognition sequences of the DNA substrates are shown on the left-hand side of the graphs. '(mC)' stands for 5mC (the last substrate in each graph is the unmethylated control); sequence positions that differ from the reference 'gC(mC)TG' substrate are marked with grey boxes; full oligoduplex sequences are listed in Table 3. The reaction rates of LpnPI mutants that show increased cleavage due to loop replacement are marked by blue streaked bars; '−' marks undetectable cleavage (rate lower than 3×10^{-7} s^{-1}, the starting position of the x-axis). Alignments of the LpnPI/AspBHI Loop-6C and the LpnPI/SgrTI Loop-2B that were replaced in the LpnPI-91RLL and LpnPI-27HTG are shown above panels A and B, respectively. (**A**) Wt enzyme and the LpnPI variant LpnPI-91RLL (Loop-6C replacement) on DNA substrates with variable sequence upstream and downstream of 5mC. (**B**) Wt enzyme and the LpnPI variant LpnPI-27HTG (Loop-2B replacement) on DNA substrates with variable sequence upstream and downstream of 5mC. (**C**) Wt enzyme and the 'double-swap' LpnPI variant LpnPI-27HTG-91RLL on DNA substrates with variable sequences upstream and downstream of 5mC.

to methylated DNA. This may also account for a significant acceleration of the doubly substituted 5 -G(mC)DC-3 substrate cleavage. Taken these data together, we define the preferred target site for LpnPI-91RLL cleavage as 5 -(S>>W)(mC)D(G>C)-3 .

The sequences of Loop-6C in LpnPI and AspBHI differ by four residues (positions 91–93 and 99, Figure 1A). To identify the residue responsible for the altered LpnPI site preference, we have made LpnPI mutants H91R, D92L, I93L and Q99F, and tested their cleavage activity on the 5 -C(mC)DG-3 and 5 -G(mC)DG-3 substrates. We found that the site preference of single point mutants did not change (the rate difference was ∼100-fold for H91R and Q99F mutants, and dropped only to ∼50-fold for the D92L and I93L variants), suggesting that the change in LpnPI-91RLL specificity is due to simultaneous replacement of several residues. In the next step, we also replaced the LpnPI Loop-6C with a corresponding loop from MspJI (residues 123–130), which shows no specificity for the DNA sequence upstream of the modified base. However, the resultant LpnPI variant 'LpnPI-91VGL', despite the unperturbed secondary structure, was inactive (Supplementary Figures S1D and S6A).

Our LpnPI-DNA model suggests that Loop-2B contacts DNA to the 3 side from the modified base (Figures 1D and 2B). To probe the role of Loop-2B residues in DNA recognition, we have made the following LpnPI variants:

(i) 'LpnPI-27HTG', in which three consecutive LpnPI residues (27–29) were replaced with equivalent SgrTI residues (34–36) (Supplementary Figure S2); our expectation was that this would relax the selectivity of LpnPI for the +2 base pair in the 5 -C(mC)DG-3 sequence and enhance cleavage of the SgrTI recognition sequence 5 -C(mC)DS-3 ;
(ii) 'LpnPI-21AGY' – the 21–30 LpnPI residues were replaced with 21–31 AspBHI residues (Figure 1A); our expectation was that this would enable cleavage at the AspBHI-like recognition sites with a 5 -NS-3 dinucleotide at the +1/+2 positions instead of the 5 -C(mC)DG-3 sequence.

As expected, the LpnPI-27HTG variant had an increased tolerance for a C in the +2 position (the rate difference for the 5 -C(mC)DG-3 and 5 -C(mC)DC-3 substrates dropped from ∼500- to ∼50-fold, Figure 3B). The loop replacement also increased the cleavage rate of the 5 -C(mC)TA-3 substrate, which was refractory to wt LpnPI. Unexpectedly, we have also observed improved cleavage of the 5 -G(mC)DC-3 substrate with a substitution in the −1 position, as the rate difference for the 5 -C(mC)DG-3 and 5 -G(mC)DC-3 substrates decreased from ∼90 to ∼15-fold (Figure 3B). Thus, relaxed recognition of the +2 base pair at least partially may be due to improved non-specific binding to methylated DNA. Taken these data together, we define the preferred target site for LpnPI-27HTG cleavage as 5 -(C>G)(mC)D(G>C>>A)-3 .

We also made point mutations at all three Loop-2B positions that differ between LpnPI and SgrTI (N27H, A28T and S29G). The ratio for the 5 -C(mC)DG-3 and 5 -C(mC)DC-3 cleavage rates for the S29G and A28T mutants was ∼200 fold, while the N27H mutant was inactive. Intriguingly, the double mutant N27H+S29G regained full activity, suggesting that the N27H mutation needs extra space or flexibility provided by the S29G mutation. Moreover, the double N27H+S29G mutant displayed DNA cleavage properties akin to LpnPI-27HTG (the ratio for the 5 -C(mC)DG-3 and 5 -C(mC)DC-3 cleavage rates dropped to ∼30-fold, data not shown), indicating that N27 and S29 are the key Loop-2B residues involved in DNA binding.

Contrary to LpnPI-27HTG, the LpnPI-21AGY variant preserved wt-like specificity. Unlike AspBHI, it did not tolerate a cytosine at the +1 position, and had a strong preference for a G nucleotide at the +2 position (Supplementary Figure S6B). Involvement of Loop-2B in the recognition of the +1 base therefore seems unlikely. Another plausible candidate for the +1 base pair recognition is Loop-78, which is 3 aa longer in LpnPI than in AspBHI (Figure 1A). To test this hypothesis we have also constructed LpnPI variant 'LpnPI-133G', containing a shorter, AspBHI-like Loop-78 version (133–134 LpnPI residues replaced with a glycine, which is equivalent to AspBHI residue G131, Figure 1A). Unfortunately, the resultant LpnPI variant LpnPI-133G was inactive on all substrates tested (data not shown).

Since Loop-6C and Loop-2B act as separate LpnPI DNA binding 'modules', we also produced a double-swap LpnPI variant 'LpnPI-27HTG-91RLL' containing the AspBHI Loop-6C (relaxes recognition of the -1, and to a lesser extent +2 positions) and the SgrTI Loop-2B (relaxes recognition of the +2, and to a lesser extent −1 positions). Our expectation was that the 'double-swap' LpnPI would readily cleave the 5 -SMDS-3 site. DNA cleavage analysis confirmed this prediction: LpnPI-27HTG-91RLL cleaved 5 -G(mC)DG-3 and 5 -C(mC)DC-3 that differ by a single bp from the standard LpnPI recognition site, and the doubly-substituted 5 -G(mC)DC-3 site, which is poorly tolerated by the wt LpnPI and the 'single-swap' variants (Figure 3A-C). Simultaneous substitution of two loops apparently further relaxed LpnPI specificity for the −1 and +2 positions: the 'double-swap' LpnPI variant cleaved the 5 -A(mC)DG-3 and 5 -T(mC)DG-3 sites only ∼5- and ∼50-fold slower than the standard substrate (the rate differences are >10000- and ∼300-fold for the wt LpnPI and the '91RLL' variant, respectively, Figure 3A); detectable cleavage (∼30–100-fold slower in comparison to the standard 5 -C(mC)DG-3 substrate) was also observed for the 5 -C(mC)DA-3 and 5 -C(mC)DT-3 DNAs, which are both refractory or almost refractory to wt LpnPI and LpnPI-27HTG cleavage (Figure 3B-C). However, no cleavage was detected for DNA substrates 5 -A(mC)DA-3 and 5 -T(mC)DT-3 , indicating that the presence of two 'unfavorable' A:T base pairs in both the -1 and +2 positions is not tolerated (Figure 3C). The recognition sequence of LpnPI-27HTG-91RLL can thus be defined as the combination of 5 -S(mC)(D>>C)-3 and 5 -(mC)(D>>C)S-3 recognition sites.

DISCUSSION

Here, we report the apo-structure of the DNA recognition domain of the cytosine modification-dependent restriction endonuclease LpnPI (LpnPI-N). The overall structure of LpnPI-N is very similar to the DNA binding domain of

AspBHI (Figure 1B) (13). Despite the structural similarity, the recognition sequences of LpnPI [5 -C(mC)DG-3] and AspBHI [5 -YS(mC)NS-3] differ. This raises a question how different sequence specificity is achieved in the conserved structural scaffold of the MspJI enzyme family. To this end we built a model of DNA-bound LpnPI (Figures 1C and 2B) and performed mutational analysis of LpnPI structural elements involved in DNA contacts, including the modified cytosine binding pocket and four surface loops.

5(h)mC binding pocket

The modified cytosine binding pocket is conserved in SRA and SRA-like domains. Typically, the side walls of the pocket are built of aromatic side chains, which make stacking interactions with the extrahelical base, and polar residues, which make cytosine-specific H-bonds to the Watson-Crick edge of the base (Figure 2). Pocket mutations of AspBHI, PvuRts1I and AbaSI proteins severely impaired DNA cleavage activity (13,16–17,34). LpnPI is no exception: the D71A mutation reduced the DNA cleavage rate ∼1000-fold (Table 4). The co-crystal structure of MspJI and the models of DNA-bound AspBHI/LpnPI (13,18) predict that the pocket aspartate (D71 in LpnPI) must be protonated to form a H-bond with the N4 cytosine atom of the flipped-out base (Figure 2B). Since a similar position in another cytosine modification-dependent enzyme PvuRts1I is occupied by an asparagine (N217), we also made the LpnPI mutant D71N. Surprisingly, even this conservative mutation reduced the LpnPI cleavage rate 10-fold (Table 4). Presumably, the structure of the pocket is highly optimized, therefore even slight perturbation of its geometry/H-bonding network has a detrimental effect on enzyme function.

Loop-B3

LpnPI Loop-B3 approaches DNA from the minor groove side. Structurally equivalent loops in SRA domains provide residues (e.g. V451 in UHRF1-SRA) that fill in the vacant space left by the flipped-out cytosine, and contribute to the recognition of the adjacent base pair (5–7) (Figure 2D). In the MspJI–DNA co-crystal structure a similar position is occupied by E65, which contacts the orphan intra-helical guanine (Figure 2A). Loop-B3 in AspBHI contains residues S41, R42 (both unique to AspBHI, Supplementary Figure S2) and Q43; the same positions in LpnPI are occupied by G41, N42 and M43 (unique to LpnPI). Alanine replacement of the 42th residue in the 'B3' loops of both enzymes abolished their activity ((13) and Table 4), suggesting direct involvement of R42/N42 residues in orphan guanine or adjacent base pair recognition. Alanine mutations of residues Q43/M43, which overlap with the UHRF1-SRA V451 residue, were less deleterious, while the LpnPI mutant M43Q displayed wt-like DNA cleavage activity (Table 4). Presumably, the main purpose of the bulky M43/Q43 residues is to fill the space left by the flipped out 5mC rather than make base-specific contacts (Figure 2B).

AspBHI has a preference for the -2 nucleotide to be a pyrimidine (C or T), while LpnPI and other related enzymes accept any nucleotide at this position (11). In our current models of DNA-bound LpnPI and AspBHI (Figure 2B and C) the closest residue to the −2 bp is G41/S41. We speculate that lacking a side chain at the 41th position, LpnPI accepts any nucleotide at the −2 position. In contrast, the same position in AspBHI is occupied by a serine, which could perform pyrimidine/purine discrimination, e.g. *via* a minor groove hydrogen bond to the N3 purine atom in the complementary strand (Figure 2C). In agreement with this model, some AspBHI S41 mutants displayed altered site preference (13). The G41S replacement in LpnPI decreased the cleavage activity ∼300-fold, but did not change the base preference for the −2 position (data not shown). It can not be excluded that other factors, including a subtle difference in the Loop-B3 conformation may contribute to the −2 base pair discrimination by AspBHI.

Loop-78

Loop-78 approaches DNA downstream of the modified base (Figure 2C). In MspJI, the loop residue K173 is the primary candidate for the recognition of the +3 base pair, where MspJI has a strong preference for a purine (18). Equivalent loops in LpnPI and AspBHI are shorter by five and eight residues, respectively (Supplementary Figure S2). Nevertheless, our current model of DNA-bound LpnPI and mutational data (alanine replacement of Loop-78 residue R137 inactivates LpnPI, Table 4) both suggest that Loop-78 residues contact DNA. Whether these contacts are limited to the DNA backbone, or contribute to the specific recognition of DNA bases (e.g. discrimination of the +1 bp) currently remains unknown, as the replacement of LpnPI Loop-78 with an AspBHI-like shorter loop (LpnPI variant '133G'), despite the proper folding of the protein (Supplementary Figure S1C), rendered LpnPI inactive.

Loop-6C

In the models of DNA-bound LpnPI and AspBHI, Loop-6C approaches the 5 -part of the target sequence from the minor groove side (Figure 2B and C). An equivalent 'CpG recognition' or 'NKR finger' loop in eukaryotic SRA domains is longer, and makes base-specific contacts in the major groove (Figure 2D) (5–7). Here, we show that replacement of the LpnPI Loop-6C with an equivalent AspBHI loop (four amino acid mutations at positions 91–93 and 99) enables cleavage of the 5 -G(mC)DG-3 site with a G base in the −1 position, accelerates cleavage of sites with A and T bases in the −1 position, and to a lesser extent improves cleavage of DNA with a C in the +2 position (Figure 3A). This change in site preference could not be emulated by single Loop-6C mutants, indicating that several loop residues contribute to DNA recognition. Intriguingly, three out of four residues replaced in the 'LpnPI-91RLL' variant (positions 91–93) in our current model of DNA-bound LpnPI/AspBHI point away from the DNA (Figure 2B and C). Direct contacts to DNA bases by these residues would require a change in Loop-6C conformation similar to that observed in UHRF1-SRA (Figure 2D). However, in MspJI enzyme the Loop-6C occupies the same position both in the apo- and in the DNA-bound structures (12,18) (Figure 2A). The role of the 91–93 and 99 Loop-6C residues in

the −1 bp recognition therefore remains undefined: some loop residues may contact DNA bases directly, but this would require an LpnPI/AspBHI-specific change in Loop-6C conformation upon DNA binding; alternatively, Loop-6C residues could contribute to the sequence recognition indirectly through interactions with other protein residues that make direct DNA contacts. An indirect role of Loop-6C residues in DNA recognition may also explain simultaneous relaxation of LpnPI interaction with substrates carrying substitutions both upstream (position -1) and downstream (position +2) of the methylated base. Another important fact is that LpnPI-91RLL, contrary to the donor enzyme AspBHI, had no preference for the −2 bp (Figure 3A). Presumably, recognition of this base pair is performed by another AspBHI structural element, most likely the Loop-B3 (see above).

Loop-2B

Loop-2B occupies the minor groove on the 3′ side of the modified base (Figure 2A and C). Replacement of the LpnPI Loop-2B with an equivalent loop from SgrTI relaxed the specificity of the LpnPI-27HTG variant for the +2 position, thereby accelerating cleavage of 5′-C(mC)DC-3′ and 5′-C(mC)DA-3′ sites (Figure 3B). Interestingly, cleavage of the 5′-G(mC)DG-3′ DNA, which carries a substitution in the −1 position, was also increased. This suggests that Loop-2B replacement may improve the overall affinity of the enzyme for the methylated DNA. Two out of three LpnPI residues replaced in the '27HTG' variant, namely, N27 and S29, in the current apo-LpnPI/DNA model point away from the DNA and are located closer to the +1 rather to the +2 base pair (Figure 2B). Presumably, upon DNA binding Loop-2B undergoes a conformational change that brings these residues closer to the +2 base pair. The AspBHI Loop-2B is longer by 1 aa and adopts a different conformation (Figures 1B and 2C); its importance for the enzyme function was also confirmed by mutagenesis (T25A and D32A mutations abolished AspBHI activity (13)). Conversely, the glutamine Q33 from the MspJI Loop-2B that contacts the DNA bases 3′ to the flipped cytosine (Figure 2D) is dispensable for MspJI activity (18). This is consistent with MspJI lacking any sequence preference for the +1 and +2 base pairs. An equivalent loop in eukaryotic SRA domains is much shorter and is not involved in base-specific DNA interactions (Figure 2D).

In summary, we show here that LpnPI recognizes the context of the flipped cytosine via several surface loops that act as separate DNA binding/recognition modules. LpnPI is a promising model system for specificity engineering of modification-dependent restriction endonucleases, since it displays a significant plasticity of target site recognition, somewhat reminiscent of homing endonucleases (35). Indeed, though wt LpnPI is most active on the canonical site 5′-C(mC)DG-3′, it also cleaves at alternative 5′-G(mC)DG-3′ and 5′-C(mC)DC-3′, sites albeit at a reduced rate (Figure 3A and B). The LpnPI loop engineering further shifted enzyme preference for alternative recognition sites. Most notably, the 'double-swap' LpnPI variant, which carries Loop-2B from AspBHI and Loop-6C from SgrTI, recognizes a shorter target sequence, which can be defined as either 5′-S(mC)(D>>C)-3′ or 5′-(mC)(D>>C)S-3′, and readily cuts the 5′-G(mC)DC-3′ site, which differs from the canonical recognition site 5′-C(mC)DG-3′ by two base pairs (Figure 3C). The relaxed sequence specificity seems to be an intrinsic feature of MspJI family enzymes. From the practical point of view this means that results of a real-life DNA cleavage experiment (% DNA cleaved at particular site) greatly depend on the enzyme/DNA concentrations and the reaction duration. For example, MspJI, AspBHI and LpnPI cleavage sites established under more favorable reaction conditions (with activator oligoduplex) were more 'relaxed' than recognition sequences determined under less favorable conditions (no activator duplex) (11). Nevertheless, despite of promiscuous specificity for the target site surrounding the modified cytosine, the SRA-like domain has proved a surprisingly robust module for the modified cytosine DNA recognition: neither the wt LpnPI nor any 'swap' or mutant variants showed any activity on unmethylated DNA. The plasticity of the target site recognition intrinsic to the MspJI family enzymes and the stringent discrimination against unmethylated DNA provided by the SRA domain pave the way for engineering of an enzyme specific for the 5mC embedded in any sequence context. Such 5mC-specific enzyme would be a useful tool in genome methylation studies. Significant relaxation of LpnPI specificity presented here is a step towards this goal.

ACCESSION NUMBER

Coordinates and structure factors of LpnPI-N are deposited under PDB ID 4RZL.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Authors acknowledge MAX-lab staff for the help with beamline operation, and Dr Saulius Gražulis for the help with data collection. G.S. and G.T. also acknowledge the travel support provided by the 'Baltic Science Link' project coordinated by the Swedish Research Council. K.K. acknowledges the support from the project 'Promotion of Student Scientific Activities' (VP1-3.1-ŠMM-01-V-02-003) from the Research Council of Lithuania.

FUNDING

Research Council of Lithuania [MIP-027/2012 to G.S.]. Funding for open access charge: Research Council of Lithuania.
Conflict of interest statement. None declared.

REFERENCES

1. Ohki,I., Shimotake,N., Fujita,N., Jee,J., Ikegami,T., Nakao,M. and Shirakawa,M. (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*, **105**, 487–497.

2. Scarsdale,J.N., Webb,H.D., Ginder,G.D. and Williams,D.C. (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res.*, **39**, 6741–6752.
3. Otani,J., Arita,K., Kato,T., Kinoshita,M., Kimura,H., Suetake,I., Tajima,S., Ariyoshi,M. and Shirakawa,M. (2013) Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. *J. Biol. Chem.*, **288**, 6351–6362.
4. Buck-Koehntop,B.A., Stanfield,R.L., Ekiert,D.C., Martinez-Yamout,M.A., Dyson,H.J., Wilson,I.A. and Wright,P.E. (2012) Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 15229–15234.
5. Hashimoto,H., Horton,J.R., Zhang,X., Bostick,M., Jacobsen,S.E. and Cheng,X. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*, **455**, 826–829.
6. Arita,K., Ariyoshi,M., Tochio,H., Nakamura,Y. and Shirakawa,M. (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature*, **455**, 818–821.
7. Avvakumov,G.V., Walker,J.R., Xue,S., Li,Y., Duan,S., Bronner,C., Arrowsmith,C.H. and Dhe-Paganon,S. (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, **455**, 822–825.
8. Zhou,T., Xiong,J., Wang,M., Yang,N., Wong,J., Zhu,B. and Xu,R-M (2014) Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. *Mol. Cell*, **54**, 879–886.
9. Rajakumara,E., Law,J.A., Simanshu,D.K., Voigt,P., Johnson,L.M., Reinberg,D., Patel,D.J. and Jacobsen,S.E. (2011) A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. *Genes Dev.*, **25**, 137–152.
10. Spruijt,C.G., Gnerlich,F., Smits,A.H., Pfaffeneder,T., Jansen,P.W.T.C., Bauer,C., Münzel,M., Wagner,M., Müller,M., Khan,F. *et al.* (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, **152**, 1146–1159.
11. Cohen-Karni,D., Xu,D., Apone,L., Fomenkov,A., Sun,Z., Davis,P.J., Kinney,S.R.M., Yamada-Mabuchi,M., Xu,S., Davis,T. *et al.* (2011) The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11040–11045.
12. Horton,J.R., Mabuchi,M.Y., Cohen-Karni,D., Zhang,X., Griggs,R.M., Samaranayake,M., Roberts,R.J., Zheng,Y. and Cheng,X. (2012) Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease. *Nucleic Acids Res.*, **40**, 9763–9773.
13. Horton,J.R., Nugent,R.L., Li,A., Mabuchi,M.Y., Fomenkov,A., Cohen-Karni,D., Griggs,R.M., Zhang,X., Wilson,G.G., Zheng,Y. *et al.* (2014) Structure and mutagenesis of the DNA modification-dependent restriction endonuclease AspBHI. *Sci. Rep.*, **4**, 4246.
14. Szwagierczak,A., Brachmann,A., Schmidt,C.S., Bultmann,S., Leonhardt,H. and Spada,F. (2011) Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **39**, 5149–5156.
15. Wang,H., Guan,S., Quimby,A., Cohen-Karni,D., Pradhan,S., Wilson,G., Roberts,R.J., Zhu,Z. and Zheng,Y. (2011) Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **39**, 9294–9305.
16. Horton,J.R., Borgaro,J.G., Griggs,R.M., Quimby,A., Guan,S., Zhang,X., Wilson,G.G., Zheng,Y., Zhu,Z. and Cheng,X. (2014) Structure of 5-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA. *Nucleic Acids Res.*, **42**, 7947–7959.
17. Shao,C., Wang,C. and Zang,J. (2014) Structural basis for the substrate selectivity of PvuRts1I, a 5-hydroxymethylcytosine DNA restriction endonuclease. *Acta Crystallogr. D Biol. Crystallogr.*, **70**, 2477–2486.
18. Horton,J.R., Wang,H., Mabuchi,M.Y., Zhang,X., Roberts,R.J., Zheng,Y., Wilson,G.G. and Cheng,X. (2014) Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. *Nucleic Acids Res.*, **42**, 12092–12101.
19. Zagorskaitė,E. and Sasnauskas,G. (2014) Chemical display of pyrimidine bases flipped out by modification-dependent restriction endonucleases of MspJI and PvuRts1I families. *PLoS One*, **9**, e114580.
20. Kabsch,W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
21. CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
22. Panjikar,S., Parthasarathy,V., Lamzin,V.S., Weiss,M.S. and Tucker,P.A. (2005) Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 449–457.
23. Vagin,A. and Teplyakov,A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 22–25.
24. Brünger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
25. Langer,G., Cohen,S.X., Lamzin,V.S. and Perrakis,A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.*, **3**, 1171–1179.
26. Murshudov,G.N., Skubák,P., Lebedev,A.A., Pannu,N.S., Steiner,R.A., Nicholls,R.A., Winn,M.D., Long,F. and Vagin,A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 355–367.
27. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.
28. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
29. Shatsky,M., Nussinov,R. and Wolfson,H.J. (2004) A method for simultaneous alignment of multiple protein structures. *Proteins*, **56**, 143–156.
30. Xu,Q., Canutescu,A.A., Wang,G., Shapovalov,M., Obradovic,Z. and Dunbrack,R.L. (2008) Statistical analysis of interface similarity in crystals of homologous proteins. *J. Mol. Biol.*, **381**, 487–507.
31. Robert,X. and Gouet,P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, **42**, W320–W324.
32. Zheng,L., Baumann,U. and Reymond,J.-L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.*, **32**, e115.
33. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2010) REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234–D236.
34. Kazrani,A.A., Kowalska,M., Czapinska,H. and Bochtler,M. (2014) Crystal structure of the 5hmC specific endonuclease PvuRts1I. *Nucleic Acids Res.*, **42**, 5929–5936.
35. Chevalier,B.S. and Stoddard,B.L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.*, **29**, 3757–3774.

APPENDIX 3

Activity and structure of EcoKMcrA

Czapinska H, Kowalska M, Zagorskaite E, Manakova E, Slyvka A, Xu SY, Siksnys V, Sasnauskas G and Bochtler M

Nucleic Acids Research, 2018, volume 46, pp. 9829–9841
https://doi.org/10.1093/nar/gky731

Nucleic Acids Research, 2018 **1**
doi: 10.1093/nar/gky731

Activity and structure of EcoKMcrA

Honorata Czapinska[1], Monika Kowalska[1], Evelina Zagorskaitė[2], Elena Manakova[2], Anton Slyvka[1], Shuang-yong Xu[3], Virginijus Siksnys[2], Giedrius Sasnauskas[2,*] and Matthias Bochtler[1,4,*]

[1]International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland, [2]Institute of Biotechnology, Vilnius University, Saulėtekio av. 7, 10257 Vilnius, Lithuania, [3]New England Biolabs, Inc. 240 County Road, Ipswich, MA 01938, USA and [4]Institute of Biochemistry and Biophysics PAS, Pawinskiego 5a, 02-106 Warsaw, Poland

Received May 18, 2018; Revised July 27, 2018; Editorial Decision July 30, 2018; Accepted August 07, 2018

ABSTRACT

***Escherichia coli* McrA (EcoKMcrA) acts as a methylcytosine and hydroxymethylcytosine dependent restriction endonuclease. We present a biochemical characterization of EcoKMcrA that includes the first demonstration of its endonuclease activity, small angle X-ray scattering (SAXS) data, and a crystal structure of the enzyme in the absence of DNA. Our data indicate that EcoKMcrA dimerizes via the anticipated C-terminal HNH domains, which together form a single DNA binding site. The N-terminal domains are not homologous to SRA domains, do not interact with each other, and have separate DNA binding sites. Electrophoretic mobility shift assay (EMSA) and footprinting experiments suggest that the N-terminal domains can sense the presence and sequence context of modified cytosines. Pyrrolocytosine fluorescence data indicate no base flipping. *In vitro*, EcoKMcrA DNA endonuclease activity requires Mn^{2+} ions, is not strictly methyl dependent, and is not observed when active site variants of the enzyme are used. In cells, EcoKMcrA specifically restricts DNA that is modified in the correct sequence context. This activity is impaired by mutations of the nuclease active site, unless the enzyme is highly overexpressed.**

INTRODUCTION

Restriction enzymes (REases) provide a genetic barrier for incoming DNA. Typically, such systems rely on DNA methylation patterns to distinguish self from non-self. Most REases identify invading DNA by a lack of modifications, while leaving genomic DNA untouched. However, there is also a growing group of REases which do the exact opposite and recognize invading DNA by its modifications.

The *Escherichia coli* immigration control systems for modified DNA were initially identified genetically based on the susceptibility of bacterial strains to phages containing modified DNA, or to plasmids that were methylated in certain sequence contexts. Genetic studies identified *modified cytosine restriction* A and B (*mcrA* and *mcrB*) as barriers for DNA modified in CCGG (M.HpaII) and RGCGCY (M.HaeII) sequence contexts, respectively (1). The same genes had previously been named *rglA* and *rglB*, where *rgl* stands for 'restricts glucose less phage'. The nomenclature was based on the observation that ablation of these genes affected propagation of glucosylation deficient phages with 5-hydroxymethylcytosine (5hmC) residues incorporated in the DNA, but not glucosylation proficient phages with DNA containing glucosyl-5hmC (2). Later, mutants in *mrr*, already known to affect restriction of methylated DNA, were also found to impair restriction of cytosine modified DNA in the *mcrA-mcrB* double negative background (3).

Since their genetics driven discovery, much information has accumulated on the biochemical properties of the modification dependent *E. coli* restriction endonucleases. Mrr is a remote homologue of the MspJI family of REases (4), which use an SRA domain for interaction with methylated DNA and a PD-(D/E)XK domain (the exact motif of this family is PD-QXK) for DNA cleavage (5–7). McrB consists of a AAA+ domain (8), and a structurally characterized specificity determining domain (McrB-N) (9) and assembles into a protein complex with the PD-(D/E)XK endonuclease McrC (10). The mechanistic basis for McrA modification dependence remains entirely unclear, making the enzyme an attractive target for structural studies.

McrA from *E. coli* K strains (EcoKMcrA) restricts modified DNA only when the methylation is in an appropriate sequence context. For example, it does not restrict DNA that has been methylated by Dcm in the C5mCWGG context, and therefore can co-exist with Dcm in *E. coli*. In contrast, the enzyme is very effective in restricting DNA that has been methylated by M.HpaII in the C5mCGG context. Until now, it has been impossible to reproduce DNA cleavage by EcoKMcrA in the test tube, even using plasmids that are restricted in cells (11). However, it has been shown that

*To whom correspondence should be addressed. Tel: +48 22 5970732; Fax: +48 22 5970715; Email: mbochtler@iimcb.gov.pl
Correspondence may also be addressed to Giedrius Sasnauskas. Tel: +370 5 2602111; Fax: +370 5 2602116; Email: gsasnaus@ibt.lt

the enzyme exhibits methylation dependent DNA binding, which can be assayed biochemically. Electrophoretic mobility shift assays (EMSAs) suggest that the EcoKMcrA specificity is Y5mCGR (12), i.e. slightly broader than required to cleave M.HpaII methylated DNA (13). How the enzyme achieves its sequence and modification specificity is incompletely understood. EcoKMcrA has the highest affinity to fully (symmetrically) methylated DNA, but the enzyme also binds hemimethylated DNA more than 10-fold tighter than unmodified DNA. In 'natural' DNA substrates, the base opposite the 5mC is always a guanine. In synthetic substrates, replacement of the G base by A, C, T or U inhibits binding, and only inosine is tolerated (12).

EcoKMcrA, like many other REases, is a dimer according to size exclusion chromatography (11). Sequence analysis suggests the presence of an N-terminal domain that is not readily amenable to bioinformatics analysis and was implicated in the modification readout (14), and a C-terminal domain that can be confidently classified as an HNH nuclease (15). HNH domains are present in many restriction and homing endonucleases, as well as Cas9 proteins. They are built around a structural Zn^{2+} ion, and a second divalent catalytic metal cation. In EcoKMcrA, the Zn^{2+} ligands, the core NLNL -Me motif with active site metal ligands and the nucleophile activating histidine are all predicted to be present. Therefore, the difficulties to demonstrate its activity *in vitro* have been surprising.

EcoKMcrA is not only of interest in its own right, but it can also be regarded as the prototype of a larger family of related enzymes. BLASTP searches readily identify proteins from eubacteria and archaea as similar to EcoKMcrA. However, in many cases, the similarity is clearly limited to the HNH domain. When only proteins with similarity to EcoKMcrA spanning the entire protein region are considered, the range of host organisms narrows down to the -proteobacteria, more specifically *Enterobacteriaceae* and a few additional species such as *Aeromonas*, *Kushneria* and *Vibrio*. Interestingly, the bacteria harboring clear EcoKMcrA orthologues are vertebrate commensals or pathogens. This finding supports the idea, originally based on the sequence specificity of the enzyme, that EcoKMcrA may also act as a barrier against the influx of CpG methylated DNA from the host (12).

Outside the -proteobacteria, ScoMcrA from the actinobacterial *Streptomyces coelicolor* has been described in some detail (16). The classification of ScoMcrA as an McrA protein is based on clear sequence similarity of the C-terminal nuclease domains. The N-terminal domains are rather different judging from amino acid sequence. ScoMcrA was originally discovered for its restricting activity on phosphorothioated DNA (16). The enzyme has type IV REase activity in the test tube, and has been shown to cleave 12–16 nucleotides away from a 5mC site and 16–28 nucleotides away from a site of phosphorothioation (16). It is still unknown whether the activity of ScoMcrA against phosphorothioated DNA is shared by the enzyme from *E. coli*.

Here, we present a biochemical characterization of EcoKMcrA that includes the first demonstration of its *in vitro* activity and the crystal structure of the enzyme in the absence of DNA together with small-angle X-ray scattering (SAXS) data obtained in solution. The data show that EcoKMcrA dimerizes via the HNH domains and has three separate DNA binding sites, one in the central channel formed by the nuclease domains, and two in the N-terminal domains. EMSA and DNase I footprinting data identify the N-terminal domains as the modification specific regions of EcoKMcrA.

MATERIALS AND METHODS

Cloning

The expression vectors for EcoKMcrA with N- and C-terminal His_6-tags (MGHHHHHHEF-EcoKMcrA$_{1-277}$ and EcoKMcrA$_{1-277}$-GHHHHHHG) were made by ligating the *ecokmcrA* gene into the pET15bm and pLATE31 plasmids, respectively (ThermoFisher Scientific). The pLATE31 plasmid was also used for the expression construct of the EcoKMcrA N-terminal fragment (EcoKMcrA$_{1-174}$-GHHHHHHG, henceforth dubbed 'EcoKMcrA-N'). Mutations were introduced using the 'quick change' protocol (17). Sequencing confirmed the intended sequences for all EcoKMcrA variants except for the N-terminally histidine tagged protein, where the presence of a K196E mutation was revealed.

Overproduction

Cells were grown at 37°C to OD_{600} of 0.7, induced by the addition of IPTG (0.2 mM for EcoKMcrA-His_6 and EcoKMcrA-N-His_6, and 0.5 mM for His_6-EcoKMcrA). Protein expression was carried out either overnight at 16°C (EcoKMcrA-His_6 and EcoKMcrA-N-His_6) or for 4 h at 22°C (His_6-EcoKMcrA).

Purification

EcoKMcrA-His_6 and EcoKMcrA-N-His_6 variants were purified by chromatography through HisTrap HP chelating and HiTrap Heparin HP columns (GE Healthcare). For His_6-EcoKMcrA purification, a gel filtration in 50 mM Tris–HCl pH 7.6, 200 mM NaCl, 1 mM EDTA, 1 mM DTT replaced the Heparin step. Proteins were concentrated by ultrafiltration, and stored in 20 mM Tris–HCl, 200 mM KCl, 1 mM DTT and 50% (v/v) glycerol, pH 8.0. Protein concentrations were determined from A_{280} measurements using the theoretical extinction coefficients calculated with the ProtParam tool available at http://web.expasy.org/protparam.

Oligonucleotides

Oligoduplex substrates used in this study are listed in Supplementary Table S1. Radioactive labeling was performed with [-^{33}P]ATP (PerkinElmer) or [-^{32}P]ATP (Hartmann Analytics) and T4 polynucleotide kinase (ThermoFisher Scientific). Oligoduplexes were assembled by annealing of the corresponding radiolabeled and unlabeled strands.

Electrophoretic mobility shift analysis

DNA binding of EcoKMcrA and EcoKMcrA-N was analyzed using ^{33}P-labeled 12 or 30 bp oligoduplexes. DNA

in the final concentration of 0.05–0.5 M was obtained by mixing 2–4 nM of radiolabeled DNA with an appropriate amount of unlabeled oligoduplex. It was incubated with the protein (final concentrations varied from 0.02 to 2 M) for 15 min in 20 l of the binding buffer (40 mM Tris-acetate, pH 8.3 at 25°C, 0.1 mg/ml BSA) supplemented with 10% (v/v) glycerol and 1 mM EDTA. Free DNA and protein-DNA complexes were separated by electrophoresis through 8% w/v polyacrylamide gels (29:1 acrylamide/bisacrylamide, gel length:width:thickness 22:15:0.1 cm) in 40 mM Tris-acetate (pH 8.3) for 45 min at 5 V/cm. Low power consumption during electrophoresis runs (~2 W or 110 V × ~18 mA per electrophoretic unit containing two gels and 1 l of electrophoretic buffer) ensured that the gels remained at room temperature (~22°C), well below the melting temperature of the oligoduplexes used in the assay (>40°C). Radiolabeled DNA and protein–DNA complexes were detected using the Cyclone phosphorimager and the OptiQuant software (Packard Instrument).

Oligonucleotide cleavage experiments

DNA cleavage experiments were performed at 37°C in the reaction buffer (50 mM KCl, 50 mM Tris–HCl, pH 7.8 at 37°C and 0.1 mg/ml BSA) supplemented with appropriate concentrations (0.01–10 mM) of $MgCl_2$, $MnCl_2$, $CoCl_2$, $CaCl_2$, $NiCl_2$, $CuSO_4$ or $ZnCl_2$. In the oligoduplex cleavage assay, the reaction mixtures contained 0.5 M EcoKMcrA (dimer) and 0.2 M DNA. At timed intervals the reactions were stopped by addition of the denaturing loading dye solution (95% v/v formamide, 25 mM EDTA) and analyzed by denaturing PAGE.

Plasmid restriction assay

BL21(DE3) ($McrA^-$) *E. coli* cells with no intrinsic antibiotic resistance were made competent. The cells were then transformed with pLATE31 plasmids (Amp^R) containing wt EcoKMcrA, its H228A or H229A variants, its N-terminal fragment (residues 1–174) or REM14 (unrelated protein, UniProt ID: Q9FX77). The cells were made competent again using Amp selection (at all growing steps the media contained 1% glucose). The cells containing each of the pLATE31 plasmids were transformed with pACYC184 plasmids (Cm^R) containing either no methyltransferase ('empty') or M.HpaII. The pACYC184 plasmids were amplified in ER2267 ($McrA^-$) *E. coli* strain and verified by sequencing. 30 ng (3 l) was used for transformation of 0.1 ml competent cells. At the recovery step after heat-shock the cells were supplemented with 0.9 ml of LB medium containing 1% glucose. After 1 h of shaking cells were gently sedimented and resuspended in 0.6 ml of glucose-free medium. 0.1 ml of transformants were spread on 3 types of LB-agar plates with (i) Amp, Cm, 1% glucose; (ii) Amp, Cm; (iii) Amp, Cm and 0.2 mM IPTG.

DNase I footprinting

50 nM of ^{33}P-labeled unmethylated and hemimethylated 50 bp DNA was preincubated at 22°C with 0.25–0.5 M of wt EcoKMcrA homodimer or 0.5–1 M EcoKMcrA-N monomer in 10 l of the binding buffer. After 15 min, 2.5 l of DNase I (ThermoFisher Scientific, 0.004 U/ l) in the binding buffer supplemented with 5 mM $CaCl_2$ and 12.5 mM $MgCl_2$ were subsequently added. After 2 min at 22°C the DNase I reactions were quenched by the addition of the denaturing loading dye solution (95% v/v formamide, 75 mM EDTA and bromophenol blue). Samples were separated on a high-resolution denaturing polyacrylamide gel (gel concentration 20% w/v, acrylamide:bis-acrylamide ratio 29:1, urea concentration 8 M, gel temperature 60°C, standard Tris–borate–EDTA buffer).

Small angle X-ray scattering (SAXS) experiments

The small angle X-ray scattering data for EcoKMcrA and EcoKMcrA-N were collected at the P12 EMBL beamline of the PETRA III ring of the DESY synchrotron in Hamburg, Germany (18), equipped with Pilatus 2M detector (Dectris). Detector to sample distance was 3.1 m. X-ray wavelength was 0.124 nm. The capillary temperature was set to 20°C, sample changer temperature was 10°C. 20 frames with exposure time 0.05 s were collected for each sample. Data were collected in the range from 0.076 to 4.665 nm^{-1}. Key SAXS data parameters are summarized in Supplementary Table S2 and SAXS-based estimates of molecular mass are listed in Supplementary Table S3. All samples were transferred into the buffer containing 20 mM Tris–HCl pH 7.5, 200 mM KCl, 0.1 mM EDTA, 0.01% (w/v) sodium azide and 1 mM DTT using desalting columns NAP5 (GE Healthcare). Samples of 30 l of EcoKMcrA-N as well as full length protein were mixed equimolarly with hemimethylated 12 bp DNA oligoduplex. EcoKMcrA and EcoKMcrA-N were step-wise concentrated by ultrafiltration and centrifuged 10–15 min before data collection. The ATSAS 2.7.1 (r6669) software suite (19) was used for data analysis. Indirect Fourier transform and parametrization of the data was performed using GNOM (Version 5.0) (20). DAMMIN (21) and DAMMIF (22) programs were used for *ab initio* shape determination. Ten runs of each program were averaged by DAMAVER package (23) and the most probable shape was determined. Models were overlaid with SUBCOMB (24). The comparison of the SAXS data with the crystal structure was performed with CRYSOL (25). The molecular masses were estimated using SAXS data measured in a relative scale by SAXSMoW (v2.1) implemented as server (http://saxs.ifsc.usp.br) (26). MW was also evaluated from Porod volumes calculated by DATPOROD and dummy atom modeling programs (DAMMIN and DAMMIF) and using the method of Rambo and Tainer (27) implemented in DATVC.

Crystallization and data collection

The EcoKMcrA K196E variant was crystallized in the absence of DNA by the vapor diffusion technique in sitting drops from 1:1 mixtures of the protein solution in 50 mM Tris–HCl, pH 7.6, 200 mM NaCl, 1 mM EDTA, 1 mM DTT with the reservoir buffer containing 0.1 M HEPES–CsOH, pH 7.3, 30% MPD. The diffraction data were collected at 1.22021 Å wavelength at the P14 beamline

of the DESY synchrotron in Hamburg and at the 1.2833 and 1.2777 Å wavelengths at the MX 14.2 beamline of the BESSY synchrotron in Berlin. All data were processed with XDS (28). The data collected at the PETRA ring of DESY reached the highest 2.85 Å resolution.

Structure determination and refinement

The molecular replacement attempts proved unsuccessful and the structure was solved with the help of experimental phasing. The data collected at DESY and BESSY were not isomorphous enough to allow for a successful MAD approach. The SAD phasing based on the highest resolution data collected at 1.22021 Å and performed with the help of autoSHARP (29) which internally uses the SHELXC/D/E pipeline (30) proved most efficient. The procedure originally found two strong zinc peaks which were later complemented with additional two weaker metal peaks. The distances between the four metal sites pointed to the presence of a two-fold symmetry axis. Two-fold symmetry was expected based on the dimeric character of the EcoKMcrA enzyme but did not show up as a peak in the 180° self-rotation map. We determined the symmetry operator and averaged the preliminary density maps with the help of DM program (31). The resulting maps were input to the LJS routine of autoSHARP that alternatingly uses BUCCANEER (32) and PARROT (33) programs for model building and density modification. The resulting model indicated that only the C-terminal domain that anchors the metal ions followed the two fold symmetry. The model was manually improved with the N- and C-terminal domains of the better resolved protomer. The resulting structure was refined with BUSTER (Global Phasing Ltd, Cambridge, United Kingdom) and rebuilt with BUCCANEER (32). The alternating cycles of manual model improvement and further automatic rebuilding led to the assembly of the complete two protomers. The structure was refined with REFMAC (34) and COOT (35) programs. The final refinement statistics are presented in Supplementary Table S4. The atomic coordinates and the corresponding structure factors were deposited at PDB under the 6GHC accession code.

RESULTS

EcoKMcrA activity *in vitro*

EcoKMcrA restricts M.HpaII methylated plasmids in cells, but so far all attempts to detect this activity *in vitro* have failed (11,12). We tested the catalytic activity of EcoKMcrA using 30 nucleotide long deoxyoligoduplex substrates with unmodified, hemi- or fully methylated C5mCGG sites in two different positions, with either top or bottom strand radiolabeled (Supplementary Table S1). The pilot cleavage reactions were performed with hemimethylated substrates in a buffer supplemented with various amounts of divalent metal ions (Supplementary Figure S1). Cleavage rates were highest for the Mn^{2+} ions in the concentration range of 0.01–0.2 mM. In the presence of Mn^{2+}, it was possible to demonstrate the concerted cleavage of both DNA strands, most likely with a single nucleotide stagger (leading to 3 overhangs). The extent of cleavage tended to be promoted by DNA methylation, but the effect was substrate dependent (Figure 1 and Supplementary Figure S2). The oligonucleotides were predominantly cleaved at multiple positions 5–15 nt upstream of the C5mCGG sequence. The unmethylated DNA was cut at identical positions, with additional cleavage observed at the unmodified CCGG sites (Figure 1 and Supplementary Figure S2). Wild type EcoKMcrA also displayed methylation independent exonuclease activity on single-stranded oligonucleotides (Figure 1).

Based on a prediction of the EcoKMcrA active site (15), the H228A, H229A, H252A and H256A variants of the enzyme were expected to be catalytically impaired or inactive. Indeed, none of these mutants generated any cleavage products (Supplementary Figure S3), excluding a contaminating endonuclease as the source of the observed activity. Together, the data suggest that EcoKMcrA is catalytically active *in vitro*, but the activity requires high enzyme concentrations, and in such conditions the enzyme exhibits only a moderate preference for methylated DNA.

EcoKMcrA was active not only on short linear DNA oligoduplexes, but also on supercoiled plasmids and phage DNA (Supplementary Figure S4). Plasmid/phage DNA cleavage reactions required high (0.5 M) enzyme concentrations, as the yield of cleavage products was much lower at 50 nM concentration. Under these conditions, there was surprisingly little effect of methylation in dcm (C5mCWGG) or HpaII (C5mCGG) context. Full digestion of the supercoiled unmethylated and methylated plasmids required 2 h incubation with 0.5 M enzyme.

The low activity of EcoKMcrA and non-stringent preference for methylated DNA prompted us to test the *in vitro* activity of the enzyme using further substrates. The enzyme was active on PCR products created with a modified dNTP mix containing 5-hydroxymethylcytidine triphosphate, but it also cleaved the DNA of the Lambda and T4GT7 phages (the latter is a hydroxymethylation- and glucosylation-deficient variant of T4 phage) (Supplementary Figure S5A). However, control digestions using MspJI (5mCNNRN9/), HpaII (C/CGG) and MluCI (/AATT) endonucleases suggested that T4GT7 and Lambda contained, in addition to the predominant unmodified cytosine, also 5mC, resulting from the passage through the Dcm^+ cells (Supplementary Figure S5B). DNA from T4 phage was resistant to digestion, demonstrating that EcoKMcrA does not cleave DNA with glucosylated 5hmC bases. The control digestions with H228A variant of the enzyme indicated only residual activity.

Next, we compared the cleavage of modified and unmodified substrates in isolation and in a competition assay. PCR DNAs containing either cytosine or 5hmC were tested separately and in a mixture additionally including 5mC modified XP12 phage DNA. In competition experiments as well as for isolated substrates, EcoKMcrA digested predominantly the modified DNA, whereas the unmodified DNA was barely affected. In this assay, the residual activity of the H228A variant was not observed. Controls with HpaII and MspJI exhibited the expected cleavage preferences (Supplementary Figure S6).

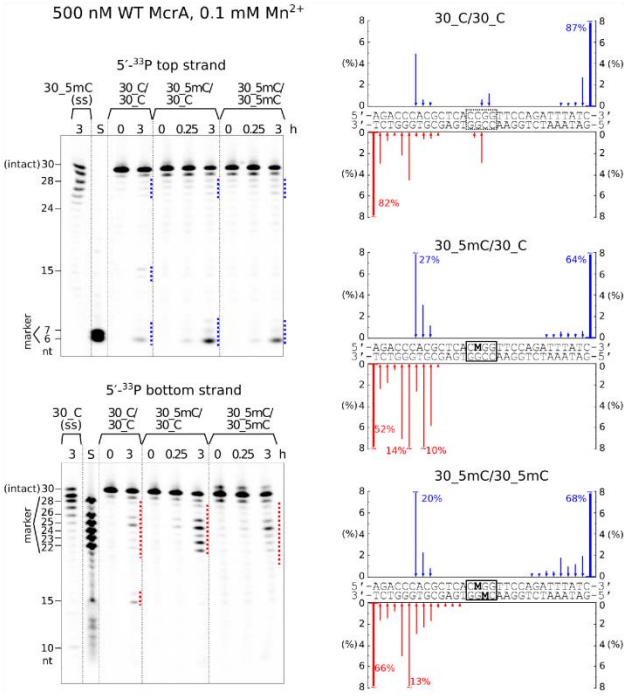


Figure 1. DNA cleavage by EcoKMcrA. The reactions were performed with 0.5 M wild type EcoKMcrA (dimer) and 0.2 M of unmodified, hemi- or fully methylated oligoduplex DNA or single stranded DNA, at 37°C in a buffer supplemented with 0.1 mM Mn^{2+}. Top strand and bottom strand cleavage positions on the gels are marked by blue and red dotted lines, respectively. Gel lanes 'S' contained radiolabeled single-stranded oligonucleotides corresponding to the 5′-terminal fragments of the respective DNA strands (sizes in nucleotides are shown on the sides of the gels). The amounts of the bottom and top strand cleavage products after 3 hours of digestion are plotted as blue/red arrows along the oligoduplex sequences. The uncleaved oligonucleotides (the 0 hour lanes) contained a detectable fraction (~10%) of shorter fragments. The reported cleavage data was corrected for the amount of these contaminants.

EcoKMcrA nuclease activity is required for plasmid and phage restriction

The weak *in vitro* activity of EcoKMcrA made us question whether the catalytic activity of EcoKMcrA was relevant for restriction in cells. In order to test this, we generated *E. coli* BL21(DE3) cells (McrA$^-$) expressing wild type EcoKMcrA, its H228A, H229A and N-terminal variants (all catalytically impaired), or a control protein (REM14), from a plasmid maintained in the cells using ampicillin selection. We also prepared pACYC184 plasmid, carrying a chloramphenicol resistance (Cm^R) gene, without insert ('empty'), or with an insert coding for M.HpaII. The CCGG sites in this vector were confirmed by HpaII digestion to be unmethylated and methylated, respectively. After transformation of the two plasmids into cells overexpressing EcoKMcrA or its variants, cells were subjected to double antibiotic selection, and survivor colonies were counted. For low or moderate protein overexpression (glucose repression or no induction), full restriction required both the M.HpaII methylation of the test plasmid, and the presence of an intact EcoKMcrA active site. High protein overexpression did not affect the colony formation only when the plasmid bearing a gene for an unrelated control protein was used (Figure 2 and Supplementary Figure S7).

The dependence of restriction on the catalytic activity of EcoKMcrA was further confirmed using a phage restriction assay with Lambda (partial dcm modification), T4gt (5hmC containing) and T4 (g5hmC containing) phages. T4 and Lambda phages were not significantly restricted for any

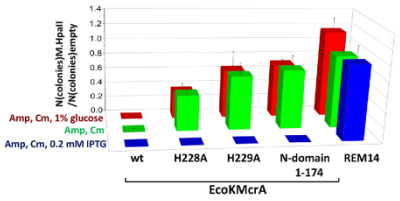


Figure. 2. Plasmid restriction assay. Quantification of the restriction of empty and M.HpaII carrying plasmids by *E. coli* cells expressing either wild type EcoKMcrA, two different active site variants of the enzyme, its N-terminal fragment lacking the catalytic domain or an unrelated protein (REM14). The pACYC184 plasmid carrying a chloramphenicol resistance gene was assayed in BL21(DE3) (McrA$^-$). The plasmid was either empty and thus without cytosine modifications or carried a gene for M.HpaII methyltransferase and thus was 5mC modified in the C5mCGG sequence context assumed to be among the EcoKMcrA targets.

tested phage titer. In contrast, T4gt phage was efficiently restricted for all tested titers by C2566 *E. coli* cells expressing wt EcoKMcrA. In contrast, the catalytically impaired variants (H228A and H229A), and the N-terminal fragment of the enzyme, only prevented plaque formation for low phage titer (Supplementary Figure S8).

EcoKMcrA structure

Crystallization of EcoKMcrA in the presence of dsDNA was not successful. However, we could grow crystals of the K196E variant in the absence of DNA. Crystals diffracted to 2.85 Å resolution at synchrotron beamlines (Supplementary Table S4). The anomalous signal resulting from the presence of Zn^{2+} ions was used for structure solution by experimental phasing at high energy remote wavelength (SAD approach).

The EcoKMcrA structure confirms the expected two-domain architecture of the enzyme, with a previously uncharacterized N-terminal domain, and a C-terminal HNH domain predicted by prior bioinformatic work (Figure 3A). The nuclease domains clearly follow the two-fold symmetry, the N-terminal domains do not. This arrangement results in two distinct conformations of full length protomers, a compact and an extended one. We place the boundary between the N- and C-terminal domains and the linker at the switch points between the structurally similar and dissimilar regions. This choice assigns residues 147–276 to the HNH domain and residues 1–143 to the N-terminal domain. The drastically different relative orientations of the two domains in the protomers are therefore due to different conformations of a short linker region spanning only a few residues (144–146) and folding into a short helix in the extended form of the protomer (Supplementary Figure S9).

The overall structure of EcoKMcrA is in agreement with the previously reported dimeric form of the enzyme and suggests that the N-terminal domains on their own should be monomers. This was tested for the N-terminal EcoKMcrA fragment (residues 1–174) by analytical sizing chromatography in high salt conditions, to prevent unspecific protein interactions with the column material. The fragment boundaries were selected prior to the structure solution and thus the N-terminal domain includes a domain swapped fragment of the HNH-domain. As a control, we confirmed that full-length EcoKMcrA migrated as a dimer (apparent mass 54.4 kDa, compared to the expected mass for the dimer 64.5 kDa). In contrast, the N-terminal fragment eluted as a monomer (expected mass including His_6-tag 20.6 kDa, apparent mass 20.1 kDa). Experiments in the presence of DNA could not be done meaningfully, because the required high ionic strength of the buffers (0.5 M) interfered with protein DNA binding, resulting in separate elution of the two components (Supplementary Figure S10).

In the crystal structure, the N-terminal domains pack very differently against the HNH-domain dimer. One domain interacts quite extensively, the other one makes contacts only to the crystallographic neighbors. According to DynDom (36), the relative domain orientations of the two protomers differ by 150° (Supplementary Figure S9). This suggested that the linkers are flexible and allow for multiple domain orientations. Alternatively, one of the observed conformations might result from the crystal packing effects.

To determine the solution conformations and to independently measure the molecular masses, we carried out small angle X-ray scattering (SAXS) experiments using full-length EcoKMcrA, and its N-terminal fragment. Data were collected both in the absence of DNA and in the presence of a 12-mer oligoduplex containing a single hemimethylated CCGG site (two oligoduplexes per EcoKMcrA dimer) (Supplementary Figure S11, Supplementary Table S2). We further calculated the *ab initio* shape of the enzyme in solution using SAXS data and the DAMMIF software (22) with no symmetry restraints. This shape was compared with the crystallized dimer, and the symmetrized dimers composed of the protomers in one conformation only. The symmetric, elongated dimer fitted best, judging from the real space fit and the parameters, which measure the agreement between calculated and observed scattering data in reciprocal space (Figure 3BC and Supplementary Figure S12). The SAXS masses for the full-length protein (between 58.7 and 77.6 kDa) agreed with the expected mass of the dimer even better than the gel filtration data. The SAXS masses for EcoKMcrA-N (between 16.8 and 23.7 kDa) confirmed the monomeric state of the N-terminal fragment in isolation (Supplementary Table S3).

EcoKMcrA N-terminal domain structure

The N-terminal domain of EcoKMcrA is organized around the six-stranded antiparallel β-sheet. The edge strand β_6 is much shorter than the strands in the center of the sheet, as expected, perhaps to avoid aggregation. The presence and length of strand β_4 might result from the N-terminal His_6 tag. Strands $\beta_4-\beta_2$, $\beta_2-\beta_3$, $\beta_4-\beta_5$ are connected by hairpins. Among these, the hairpin between strands $\beta_4-\beta_5$ is notable because it contains two short α-helices (α_2 and α_3), which pack approximately against the sheet. The α-helices of the $\beta_3-\alpha_1-\beta_4$ and $\beta_5-\alpha_4-\beta_6$ motifs, and another helix at the C-terminus of the domain (α_5) also pack against the sheet, but on the other side (Figure 4AB).

Overall, the N-terminal domain of EcoKMcrA is al-

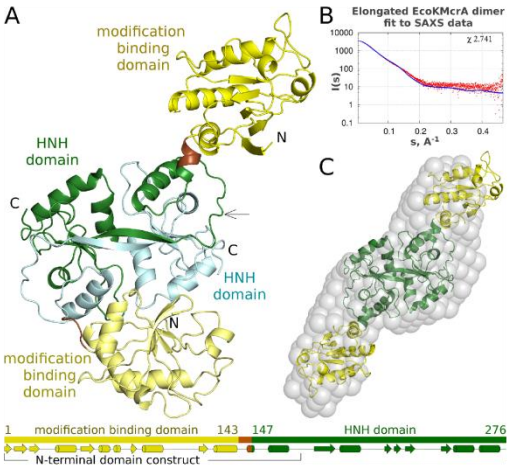


Figure 3. Structure of EcoKMcrA. (**A**) Ribbon diagram of the EcoKMcrA dimer in the asymmetric unit of the crystals, colored according to domains. The domain organization of each protomer of is shown below. (**B**) Comparison of a symmetrized model of the EcoKMcrA dimer based on the more elongated protomer, with small-angle X-ray scattering (SAXS) data for the protein in the absence of DNA. (**C**) EcoKMcrA model *ab initio* calculated from the SAXS data (grey spheres) overlaid with the symmetrized dimer that was used for the calculation of the predicted small angle X-ray scattering data presented in panel B.

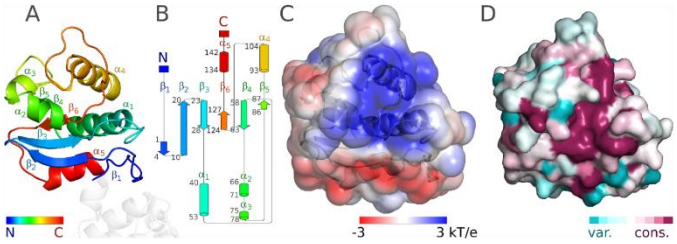


Figure 4. Structure of the N-terminal domain of EcoKMcrA. (**A**) Ribbon diagram in rainbow coloring from N- to C-terminus (blue to red). (**B**) Topology diagram generated with Pro-origami (50). (**C**) Electrostatic potential at the solvent excluded surface, blue represents positively charged regions, red negatively charged regions. (**D**) ConSurf (37) JTT conservation scores were mapped to the solvent accessible surface. Coloring is from cyan (poorly conserved regions) to magenta (for highly conserved regions). Panels A, C and D show the domain in the same orientation.

most charge neutral (18 arginines and lysines, versus 18 glutamates and aspartates, sequence based isoelectric point pI 6.9). However, when the charges or the electrostatic potential are mapped to the protein surface, a clear pattern emerges. The side of the protein distal to the linker is generally positively charged, whereas the side more proximal to the linker is negatively charged. The main regions of positive charge are the α_5–β_4–α_6 motif (94–120) and the α_4–α_5 hairpin (70–81). Altogether, this patch can be described as a groove between helices β_1 on one side and β_2 and β_4 on the other (wrapping from one face of the α-sheet to the other). We strongly suspect that this groove of EcoKMcrA is a DNA binding region (Figure 4C).

Analysis of the sequence conservation of full length EcoKMcrA homologues shows that only about half of the positively charged residues are conserved. We carried out an unbiased analysis of sequence conservation in the N-terminal domain with the help of the ConSurf server (37). We used default settings, i.e. we did not require the presence of an HNH domain accompanying the N-domain in the orthologues. Among the surface regions, the positively

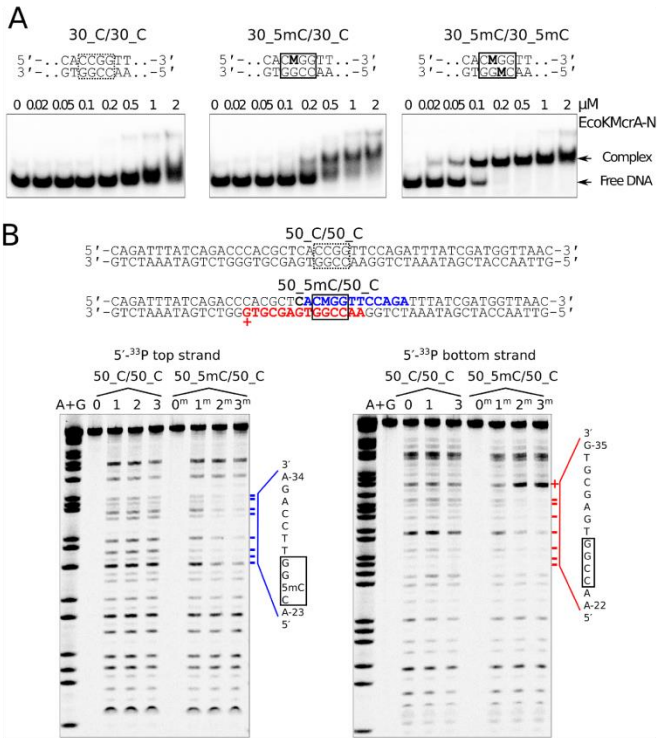


Figure 5. DNA binding by the N-terminal fragment of EcoKMcrA (EcoKMcrA-N). (**A**) Electrophoretic mobility shift assay. DNA was unmodified, hemi- or fully methylated (central sequences shown on top; 'M' denotes 5mC). DNA concentration was 50 nM, concentrations of EcoKMcrA-N are indicated above gel lanes. Positions of free DNA and protein-DNA complexes are indicated. (**B**) DNase I footprint. Top: the region of the DNA protected from DNase I cleavage by EcoKMcrA-N binding is shown in bold font. The bottom strand G-35 position that becomes more susceptible to DNase I treatment upon EcoKMcrA-N binding is shown by '+'. Bottom: DNA protection by EcoKMcrA-N. Gel lanes '0' contained untreated unmodified DNA, '1' – the DNA treated with DNase I in the absence of EcoKMcrA-N, lanes '2' and '3' – the DNA treated with DNase I in the presence of 0.5 and 1 M EcoKMcrA-N, respectively. Lanes '0m', '1m', '2m' and '3m' contained analogous samples prepared with hemimethylated DNA. Size markers (lanes 'A+G') were generated using a standard Maxam-Gilbert sequencing reaction. Positions of the DNA protected from DNase I cleavage upon EcoKMcrA-N binding are marked by blue (top strand) or red (bottom strand) dashes. The sequences of the protected regions are shown on the right-hand side of the gels.

charged region was the most conserved (Figure 4D and Supplementary Figure S13).

Modification specificity of the EcoKMcrA N-terminal domain

The domain architecture of EcoKMcrA is reminiscent of other methyl directed endonucleases, which are built as fusions of nuclease and modification binding domains. This makes the N-terminal domain of EcoKMcrA the prime candidate for the recognition of methylated DNA. We carried out electrophoretic mobility shift experiments using the N-terminal fragment of EcoKMcrA and the DNA that was previously used in the activity assay. The data indicated tighter binding of the methylated DNA compared to the unmodified DNA, confirming that EcoKMcrA-N is the methylation specific domain (Figure 5A). Methyl specific binding was also confirmed by the DNase I footprint of EcoKMcrA-N. The protection from DNase I cleavage was observed only for the methylated DNA (Figure 5B). The

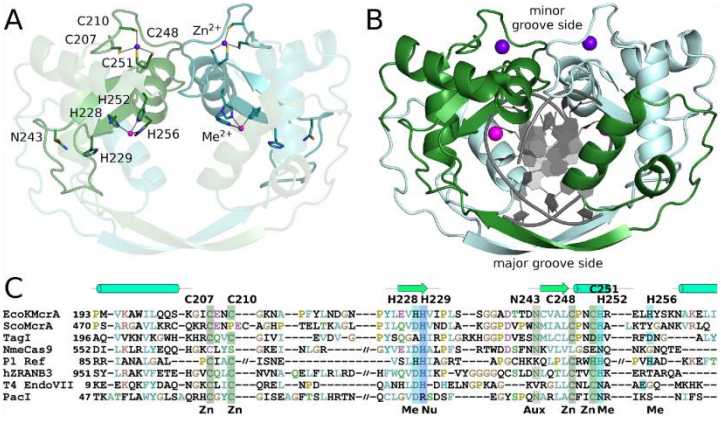


Figure 6. HNH domain dimer of EcoKMcrA. (**A**) Ribbon representation of the structure of the EcoKMcrA HNH domain. The structural Zn^{2+} ions, and the active site metal ions (Zn^{2+} in the crystal structure, and most likely Mn^{2+} in the physiological state), are shown together with their ligands. (**B**) Model of EcoKMcrA DNA complex, created based on the comparison with the Hpy99I restriction endonuclease. The blue and pink balls represent structural Zn^{2+} and active site metal ions, respectively. (**C**) Sequence alignment of the HNH core region of EcoKMcrA and related endonucleases: ScoMcrA (WP_011029780), TagI (WP_084609162), NmeCas9 (PDB: 5VGB), P1 phage Ref protein (PDB: 3PLW), human ZRANB3 endonuclease (PDB: 5MKW), T4 phage endonuclease VII (PDB: 1E7D) and PacI endonuclease (PDB: 3LDY). Zn – Zn^{2+} ligand, Me – active site metal ligand, Nu – nucleophile activator, Aux – auxiliary HNH residue.

protected region around the 5mC was identical to the region covered by full-length EcoKMcrA (Supplementary Figure S14), but for the full-length enzyme, we additionally observed scattered footprint for unmethylated DNA. This suggests that EcoKMcrA-N has higher overall selectivity for the methylated site than the full-length enzyme. The methyl-independent DNA binding observed with EcoKMcrA is therefore likely due to the DNA interaction with the C-terminal HNH domains.

No indication for modified base flipping by EcoKMcrA in fluorescence assay using pyrrolocytosine

In order to check whether nucleotide flipping played a role in the interactions of the EcoKMcrA N-terminal domain with DNA, we used a fluorescence based assay. The method relies on the properties of pyrrolocytosine (pC), which due to quenching exhibits about twice lower fluorescence in double stranded compared to single stranded DNA. Meaningful use of the pC analogue requires that the enzyme accepts it instead of the physiologically relevant cytosine bases. In the case of EcoKMcrA, a hemimodified 12-mer oligoduplex containing pC bound more tightly than unmodified control, but less tightly than the duplex containing 5mC (Supplementary Figure S15A). The dsDNA duplex containing pC had very similar properties in the absence of the protein, and in the presence of either full length EcoKMcrA or its N-terminal fragment (Supplementary Figure S15B). This negative result speaks against nucleotide flipping, but could also be due to quenching of the flipped pC fluorescence by aromatic amino acid residues of a binding pocket. According to the crystal structure, such a pocket is unlikely to be present in the N-terminal fragment of EcoKMcrA. Therefore, the data suggest that the enzyme recognizes modified cytosine bases in the context of dsDNA without nucleotide flipping.

EcoKMcrA HNH domain fold and active site

HNH endonucleases are built around a structural Zn^{2+} ion, and an active site divalent metal ion (Figure 6). The Zn^{2+} ion is typically coordinated by four cysteine residues in two CxxC motifs. Both CxxC motifs are present in the EcoKMcrA sequence, and in the structure indeed coordinate a Zn^{2+} ion, presumably taken from the expression cells and retained through the purification procedure. The HNH endonucleases are also known as NUNL-Me endonucleases (where NUNL denotes three consecutive secondary structure elements and Me stands for the active site metal ion). The NUNL motif is canonically placed after the first CxxC motif, and overlaps with the second CxxC motif (Figure 6A).

All moniker residues of the HNH motif are present in EcoKMcrA in their standard places. The first histidine residue is located near the end of the first NEstrand, and is generally thought to act as the activator for the water that is incorporated into the substrate. In EcoKMcrA, this residue is H229. The nucleophile activator histidine is typically preceded by a ligand for the active site metal ion. In the EcoKMcrA structure, the preceding residue is a histidine

(H228), which is indeed a metal ion ligand and is required for activity (Supplementary Figure S3).

The asparagine residue of the HNH motif is typically located at the N-terminus of the second β-strand, and has a structural role. The residue in EcoKMcrA is N243, which donates hydrogen bonds from its side chain carboxamide group to the carbonyl oxygen atoms of I231 and A238, and accepts a hydrogen bond from the NH of I231. The interactions stabilize the extended loop between the two β-strands of the ββα-Me motif, and judging from the high conservation of this residue, seem important for the integrity of the motif.

The second histidine of the HNH motif is located in the ββα core α-helix and is typically a ligand for the active site metal ion. In EcoKMcrA, this residue is H252. A further metal ion ligand, H256, follows one helical turn downstream. Judging from the *in vitro* activity of the alanine variants, both histidine residues are required for activity (Supplementary Figure S3).

The active site metal ion in EcoKMcrA is thus coordinated by H228, H252 and H256. In the crystal structure, this metal ion is likely Zn^{2+} judging from the anomalous diffraction signal. According to the biochemical data, it is a Mn^{2+} ion in the productive EcoKMcrA complex. Commonly observed hexa-coordination of the Mn^{2+} ion provides more than sufficient ligand acceptor sites for interactions with non-bridging oxygen of the scissile phosphate and the leaving group oxygen that would be expected in a productive complex.

EcoKMcrA HNH domain dimerization mode

The HNH domains form a two-fold symmetric dimer, held together by an interface area of ∼3250 Å^2 according to the PISA web server (38). At the center of the dimer, there is a very prominent channel, which could accommodate the bound DNA. Comparison of the HNH domains of EcoKMcrA and Hpy99I (39) or PacI (40) confirms the predicted DNA binding mode, and also defines major and minor groove sides of the DNA. In the 'crossover' region (residues 179–191), the polypeptide chain adopts an extended, β-strand conformation, which engages in main chain hydrogen bonding with its counterpart from the other protomer in a canonical antiparallel mode. The elNEMO server (41), which predicts normal modes of protein motion in solution based on molecular dynamics calculations, shows that the contact region between subunits on the major groove side is very flexible indeed. However, for EcoKMcrA to open up on this side to latch onto DNA, the entire 'crossover' interaction would have to be undone, and thus the crossover region is more likely to act as a hinge.

On the minor groove side of the modelled DNA, the protomers also interact, but the interface is quite different. The contacts involve a long loop preceding the ββα core motif (residues 215–224) and a fragment comprising the α-helix of the motif (residues 249–260). The helix is expected from the structures of other HNH endonucleases to wedge into the DNA minor groove. The two regions approach each other in such a manner that the region 249–260 of one protomer interfaces predominantly with the 215–224 region of the other protomer. Their interactions are mostly mediated by amino acid side chains, and contribute less than one third of the complete interface area. As they involve the catalytic core regions, these interactions are likely to fix the stagger for concerted cleavages of both DNA strands.

Two binding sites for the modified DNA in the EcoKMcrA dimer

The structural data suggest that EcoKMcrA should have one joint DNA binding site in the nuclease domain, and two separate modification dependent DNA binding domains. In order to verify this model, we first tested wild type EcoKMcrA for preferential binding of methylated over unmethylated DNA in an EMSA assay, using either 30mer or 12mer duplexes. As observed for the N-terminal domain alone, binding was promoted by DNA methylation. However, in agreement with the DNase I footprinting results, for 30 nt long substrates the dependence on methylation was more pronounced for EcoKMcrA-N than for the full length enzyme (Supplementary Figure S16A). We interpret these data as evidence that the nuclease domain contributes some unspecific affinity to DNA for the longer substrates, but not the short oligoduplexes that are largely covered by the N-terminal domain. EMSA experiments with full-length enzyme, but not EcoKMcrA-N, showed the presence of two bands of retarded electrophoretic mobility (Figure 5A and Supplementary Figure S16A), as reported earlier, and attributed to a monomer dimer equilibrium (12). In the light of the biochemical and structural data reported here, the two shifted bands more plausibly correspond to EcoKMcrA with either one or two bound oligoduplexes. To test this model, we carried out EMSA experiments with modified oligoduplexes of two different lengths mixed in various ratios (Supplementary Figure S16B). We observed the presence of three methyl-dependent complexes of distinct mobilities, which is consistent with the two independent binding sites for methylated DNA in the EcoKMcrA dimer. Next, we verified whether cleavage could be made more efficient when the nuclease domain was positioned by two closely spaced methylation sites. We have tested various combinations of hemi- and fully methylated substrates with two different distances between the modification sites. Consistent with the weak digestion of densely modified phage DNA, we observed at most a mild increase in cleavage rates with the new substrates (Supplementary Figure S17).

A synthetic cytosine modification-dependent restriction endonuclease

Wild-type EcoKMcrA has very moderate cleavage activity and nuclease specificity. We fused an attenuated variant of the 76 amino acid nuclease domain of the N. Gamma nickase of *Bacillus anthracis* (42) to wt EcoKMcrA, or its inactive H252A variant. The N. Gamma nickase is sequence specific and nicks the top strand of YC↓GGT ('↓' indicates top strand cleavage) or bottom strand of the ACC↑GR ('↑' indicates bottom strand cleavage) (42). The resulting enzyme nicked M.HpaII methylated, but not unmodified pBR322, at sites matching its nuclease target sequence (Supplementary Figure S18).

DISCUSSION

The N-terminal domain of EcoKMcrA binds to DNA in a methylation dependent way. EMSA assays indicate that the affinity is highest for fully-methylated DNA, lower for hemi-methylated DNA, and lowest for unmethylated DNA. The structure of the N-terminal domain of EcoKMcrA is surprisingly dissimilar from other modification specific DNA binding domains. In particular, the domain is not structurally similar to SRA domains and unrelated to the McrB-N domain responsible for the modification dependence of McrBC (9). The DALI server (43) suggested marginal similarity of the EcoKMcrA N-terminal domain to I-DreI, a dimeric meganuclease composed of homing endonucleases I-DmoI and I-CreI (44). PDBeFOLD (45) picked up fold similarities to MotA, a transcription factor from T4 phage (46) (Supplementary Figure S19). As T4 phage can infect *E. coli*, a common evolutionary origin of MotA and EcoKMcrA is plausible, in spite of the very low sequence similarity and different specificity (MotA binds to g5hmC-containing T4 phage DNA, EcoKMcrA restricts 5mC and 5hmC, but not g5hmC-containing DNA). Fold comparisons show that the connectivity of the β-sheet and two α-helices is largely preserved in the three proteins. The major difference is an insertion of residues 64–123 in EcoKMcrA that covers structural elements α2-4 and β5 (Figure 4 and Supplementary Figure S20).

The DNA binding mode of the N-terminal fragment of EcoKMcrA is still unknown. The pyrrolocytosine assay does not provide evidence for nucleotide flipping, but alone is not conclusive, because the fluorescence of the probe could be similarly quenched in a protein pocket as in the context of double stranded DNA. A structural model for DNA binding can be tentatively deduced from similarity to I-DmoI and MotA. The inferred binding mode places the DNA on the positively charged patch of the surface of the EcoKMcrA N-terminal domain, in the region of highest amino acid conservation. It implies that, as in the templates, the protein faces the major groove of the DNA, where – near the outer edges– methyl or hydroxymethyl groups of modified cytosine bases would be located. The model predicts a footprint of about one full DNA turn, in agreement with biochemical observations (Figure 5). The elongated symmetrized EcoKMcrA dimer with DNA bound to the conserved patch of the N-terminal domain surface agrees equally well with the experimental scattering for the EcoKMcrA-DNA complex, as the model for the elongated EcoKMcrA alone with its corresponding SAXS data (Supplementary Figure S20).

The C-terminal domain of EcoKMcrA is responsible for the nuclease activity. The domain has a typical HNH fold, with the expected active site residues, which are conserved in the EcoKMcrA family. It is therefore surprising that demonstration of *in vitro* endonucleolytic activity has been so difficult. In our hands, EcoKMcrA has weak, Mn^{2+} ion dependent, endonuclease activity on double- and single-stranded DNA. The specific requirement for Mn^{2+} ions was unexpected, because many other HNH endonucleases have less stringent preference for the divalent metal cations in their active sites. Mn^{2+} ions may be available physiologically, thanks to mechanisms for active uptake, presumably from the host (47). EcoKMcrA generated nicks in both strands of the oligoduplex. The nicking of two DNA strands is consistent with concerted action resulting in double strand breaks (DSB). The most prominent nicks indicate 1 nucleotide 3′ cleavage stagger (Figure 1).

The overall organization of EcoKMcrA with N-terminal modification specific, and C-terminal nuclease domain is reminiscent of the two-domain organization of the SRA-HNH (48), PD-(D/E)XK-SRA (PvuRts1I) (49) and SRA-PD-(D/E)XK (MspJI) (6) modification specific restriction endonucleases. In all these enzymes, the largely unspecific nuclease domains mediate dimerization, and expose a single joint DNA binding site, whereas the modification specific domains expose separate binding sites for modified DNA. Our EcoKMcrA crystal structure indicates that the nuclease domains of the enzyme indeed form a tight dimer, the conformation of which is compatible with concerted cleavage of two DNA strands. The results of the SAXS experiments confirm that EcoKMcrA is dimeric also in solution. Finally, the biochemical data indicate that each N-terminal domain binds double stranded DNA in modification dependent manner. However, the *in vitro* cleavage assay does not show a clear increase of the enzyme activity in the presence of two closely spaced modification sites.

Surprisingly, the EcoKMcrA *in vitro* activity is weak, and like the activity of the SRA-HNH endonuclease Sco5333 (ScoA3IV) (48), only partially methylation dependent. The activity of EcoKMcrA on unmethylated DNA and the lack of dependence on two modification sites are probably a consequence of high protein concentrations required to achieve any DNA cleavage at all. Why such high protein concentrations and long incubation times are required remains puzzling, especially in the light of sequence and structural data that point to the conservation of the active site residues. The low *in vitro* activity of EcoKMcrA also has to be reconciled with the efficacy of the enzyme as a restricting factor for phages or plasmids with cytosine modifications. It is possible that an optimal substrate for *in vitro* experiments has not been identified. Such a substrate might only be created in cells, for example as a side product of DNA replication or transcription. Alternatively, EcoKMcrA could require an activator or an effector nuclease, from the pool of host proteins. Our finding that the nuclease activity of EcoKMcrA is required for efficient phage or plasmid restriction makes it unlikely that the activity of the enzyme depends on the recruitment of another nuclease. The other possible explanations for the efficacy of EcoKMcrA as a restricting factor remain to be explored.

DATA AVAILABILITY

The atomic coordinates and corresponding structure factors were deposited at PDB under the 6GHC accession code.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The SAXS data were collected at the P12 beamline operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). Crystal diffraction data were collected at the P14 beamline of the PETRA III storage ring (DESY, Hamburg, Germany), and at MX14.2 beamline of the BESSY storage ring (Berlin, Germany). We would like to thank Dr A. Kikhney for the assistance with SAXS data collection, and Drs Uwe Mueller, Manfred Weiss, and Gleb Bourenkov for assistance with crystallographic data collection. We also thank Geoff Wilson and Peter Weigele for providing phage DNA, and Kiersten Flodman for help with enzyme purification.

FUNDING

Ministry of Science and Higher Education [0295/B/PO1/2008/34 to M.B., N N301 425038 to H.C.]; Polish National Science Centre (NCN) [UMO-2011/02/A/NZ1/00052, UMO-2014/13/B/NZ1/03991, UMO-2014/14/M/NZ5/00558 to M.B.]; Research Council of Lithuania [MIP-027/2012 to G.S.]; Part of this work was performed using Centre for Preclinical Research and Technology (CePT) infrastructure [European Union POIG.02.02.00-14-024/08-00 project]. Funding for open access charge: NCN [UMO-2014/13/B/NZ1/03991].

Conflict of interest statement. Dr Shuang-yong Xu is an employee of New England Biolabs.

REFERENCES

1. Raleigh,E.A. and Wilson,G. (1986) Escherichia coli K-12 restricts DNA containing 5-methylcytosine. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 9070–9074.
2. Revel,H.R. (1967) Restriction of nonglucosylated T-even bacteriophage: properties of permissive mutants of Escherichia coli B and K12. *Virology*, **31**, 688–701.
3. Kelleher,J.E. and Raleigh,E.A. (1991) A novel activity in Escherichia coli K-12 that directs restriction of DNA modified at CG dinucleotides. *J. Bacteriol.*, **173**, 5220–5223.
4. Zheng,Y., Cohen-Karni,D., Xu,D., Chin,H.G., Wilson,G., Pradhan,S. and Roberts,R.J. (2010) A unique family of Mrr-like modification-dependent restriction endonucleases. *Nucleic Acids Res.*, **38**, 5527–5534.
5. Cohen-Karni,D., Xu,D., Apone,L., Fomenkov,A., Sun,Z., Davis,P.J., Kinney,S.R., Yamada-Mabuchi,M., Xu,S.Y., Davis,T. *et al.* (2011) The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11040–11045.
6. Horton,J.R., Mabuchi,M.Y., Cohen-Karni,D., Zhang,X., Griggs,R.M., Samaranayake,M., Roberts,R.J., Zheng,Y. and Cheng,X. (2012) Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease. *Nucleic Acids Res.*, **40**, 9763–9773.
7. Horton,J.R., Wang,H., Mabuchi,M.Y., Zhang,X., Roberts,R.J., Zheng,Y., Wilson,G.G. and Cheng,X. (2014) Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. *Nucleic Acids Res.*, **42**, 12092–12101.
8. Neuwald,A.F., Aravind,L., Spouge,J.L. and Koonin,E.V. (1999) AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.*, **9**, 27–43.
9. Sukackaite,R., Grazulis,S., Tamulaitis,G. and Siksnys,V. (2012) The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. *Nucleic Acids Res.*, **40**, 7552–7562.
10. Bujnicki,J.M. and Rychlewski,L. (2001) Grouping together highly diverged PD-(D/E)XK nucleases and identification of novel superfamily members using structure-guided alignment of sequence profiles. *J. Mol. Microbiol. Biotechnol.*, **3**, 69–72.
11. Mulligan,E.A. and Dunn,J.J. (2008) Cloning, purification and initial characterization of E. coli McrA, a putative 5-methylcytosine-specific nuclease. *Protein Expr. Purif.*, **62**, 98–103.
12. Mulligan,E.A., Hatchwell,E., McCorkle,S.R. and Dunn,J.J. (2010) Differential binding of Escherichia coli McrA protein to DNA sequences that contain the dinucleotide m5CpG. *Nucleic Acids Res.*, **38**, 1997–2005.
13. Card,C.O., Wilson,G.G., Weule,K., Hasapes,J., Kiss,A. and Roberts,R.J. (1990) Cloning and characterization of the HpaII methylase gene. *Nucleic Acids Res.*, **18**, 1377–1383.
14. Anton,B.P. and Raleigh,E.A. (2004) Transposon-mediated linker insertion scanning mutagenesis of the Escherichia coli McrA endonuclease. *J. Bacteriol.*, **186**, 5699–5707.
15. Bujnicki,J.M., Radlinska,M. and Rychlewski,L. (2000) Atomic model of the 5-methylcytosine-specific restriction enzyme McrA reveals an atypical zinc finger and structural similarity to betabetaalphaMe endonucleases. *Mol. Microbiol.*, **37**, 1280–1281.
16. Liu,G., Ou,H.Y., Wang,T., Li,L., Tan,H., Zhou,X., Rajakumar,K., Deng,Z. and He,X. (2010) Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. *PLos Genet.*, **6**, e1001253.
17. Zheng,L., Baumann,U. and Reymond,J.L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.*, **32**, e115.
18. Blanchet,C.E., Spilotros,A., Schwemmer,F., Graewert,M.A., Kikhney,A., Jeffries,C.M., Franke,D., Mark,D., Zengerle,R., Cipriani,F. *et al.* (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.*, **48**, 431–443.
19. Franke,D., Petoukhov,M.V., Konarev,P.V., Panjkovich,A., Tuukkanen,A., Mertens,H.D.T., Kikhney,A.G., Hajizadeh,N.R., Franklin,J.M., Jeffries,C.M. *et al.* (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.*, **50**, 1212–1225.
20. Svergun,D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr*, **25**, 495–503.
21. Svergun,D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.*, **76**, 2879–2886.
22. Franke,D. and Svergun,D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **42**, 342–346.
23. Volkov,V.V. and Svergun,D.I. (2003) Uniqueness of ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.*, **36**, 860–864.
24. Kozin,M. and Svergun,D.I. (2001) Automated matching and high- and low-resolution structural models. *J. Appl. Cryst.*, **34**, 33–41.
25. Svergun,D.I., Barberato,C. and Koch,M.H.J. (1995) CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.*, **28**, 768–773.
26. Fischer,H., Oliveira Neto,M., Napolitano,H.B., Polikarpov,I. and Craievich,A.F. (2010) Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Cryst.*, **43**, 101–109.
27. Rambo,R.P. and Tainer,J.A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, **496**, 477–481.
28. Kabsch,W. (2010) Xds. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 125–132.
29. Vonrhein,C., Blanc,E., Roversi,P. and Bricogne,G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.*, **364**, 215–230.
30. Sheldrick,G.M. (2008) A short history of SHELX. *Acta Crystallogr. A*, **64**, 112–122.
31. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.*, **67**, 235–242.

32. Cowtan,K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D. Biol. Crystallogr.*, **62**, 1002–1011.
33. Cowtan,K. (2010) Recent developments in classical density modification. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 470–478.
34. Murshudov,G.N., Skubak,P., Lebedev,A.A., Pannu,N.S., Steiner,R.A., Nicholls,R.A., Winn,M.D., Long,F. and Vagin,A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.*, **67**, 355–367.
35. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.*, **60**, 2126–2132.
36. Taylor,D., Cawley,G. and Hayward,S. (2014) Quantitative method for the assignment of hinge and shear mechanism in protein domain movements. *Bioinformatics*, **30**, 3189–3196.
37. Ashkenazy,H., Erez,E., Martz,E., Pupko,T. and Ben-Tal,N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, **38**, W529–W533.
38. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
39. Sokolowska,M., Czapinska,H. and Bochtler,M. (2009) Crystal structure of the beta beta alpha-Me type II restriction endonuclease Hpy99I with target DNA. *Nucleic Acids Res.*, **37**, 3799–3810.
40. Shen,B.W., Heiter,D.F., Chan,S.H., Wang,H., Xu,S.Y., Morgan,R.D., Wilson,G.G. and Stoddard,B.L. (2010) Unusual target site disruption by the rare-cutting HNH restriction endonuclease PacI. *Structure*, **18**, 734–743.
41. Suhre,K. and Sanejouand,Y.H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.*, **32**, W610–W614.
42. Gutjahr,A. and Xu,S.Y. (2014) Engineering nicking enzymes that preferentially nick 5-methylcytosine-modified DNA. *Nucleic Acids Res.*, **42**, e77.
43. Holm,L. and Laakso,L.M. (2016) Dali server update. *Nucleic Acids Res.*, **44**, W351–W355.
44. Chevalier,B.S., Kortemme,T., Chadsey,M.S., Baker,D., Monnat,R.J. and Stoddard,B.L. (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell*, **10**, 895–905.
45. Krissinel,E. and Henrick,K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D. Biol. Crystallogr.*, **60**, 2256–2268.
46. Cuypers,M.G., Robertson,R.M., Knipling,L., Waddell,M.B., Moon,K., Hinton,D.M. and White,S.W. (2018) The phage T4 MotA transcription factor contains a novel DNA binding motif that specifically recognizes modified DNA. *Nucleic Acids Res.*, **46**, 5308–5318.
47. Diaz-Ochoa,V.E., Jellbauer,S., Klaus,S. and Raffatellu,M. (2014) Transition metal ions at the crossroads of mucosal immunity and microbial pathogenesis. *Front. Cell Infect. Microbiol.*, **4**, 2.
48. Han,T., Yamada-Mabuchi,M., Zhao,G., Li,L., Liu,G., Ou,H.Y., Deng,Z., Zheng,Y. and He,X. (2015) Recognition and cleavage of 5-methylcytosine DNA by bacterial SRA-HNH proteins. *Nucleic Acids Res.*, **43**, 1147–1159.
49. Kazrani,A.A., Kowalska,M., Czapinska,H. and Bochtler,M. (2014) Crystal structure of the 5hmC specific endonuclease PvuRts1I. *Nucleic Acids Res.*, **42**, 5929–5936.
50. Stivala,A., Wybrow,M., Wirth,A., Whisstock,J.C. and Stuckey,P.J. (2011) Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics*, **27**, 3315–3316.

APPENDIX 4

Recognition of modified cytosine variants by the DNA-binding domain of methyl-directed endonuclease McrBC

Zagorskaitė E, Manakova E and Sasnauskas G

FEBS Letters, 2018, volume 592, pp. 3335-3345
https://doi.org/10.1002/1873-3468.13244

FEBS Letters

Recognition of modified cytosine variants by the DNA-binding domain of methyl-directed endonuclease McrBC

Evelina Zagorskaitė, Elena Manakova and Giedrius Sasnauskas

Institute of Biotechnology, Vilnius University, Vilnius, Lithuania

Correspondence
G. Sasnauskas, Institute of Biotechnology, Vilnius University, Saulėtekio avenue 7, LT-10257 Vilnius, Lithuania
Fax:+370-5-2234367
Tel: +370-5-2234443
E-mail: gsasnaus@ibt.lt

(Received 9 July 2018, revised 10 August 2018, accepted 5 September 2018)

doi:10.1002/1873-3468.13244

Edited by Christian Griesinger

Cytosine modifications expand the information content of genomic DNA in both eukaryotes and prokaryotes, providing means for epigenetic regulation and self versus nonself discrimination. For example, the methyl-directed restriction endonuclease, McrBC, recognizes and cuts invading bacteriophage DNA containing 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), and N4-methylcytosine (4mC), leaving the unmodified host DNA intact. Here, we present cocrystal structures of McrB-N bound to DNA oligoduplexes containing 5hmC, 5-formylcytosine (5fC), and 4mC, and characterize the relative affinity of McrB-N to various cytosine variants. We find that McrB-N flips out modified bases into a protein pocket and binds cytosine derivatives in the order of descending affinity: 4mC > 5mC > 5hmC ≫ 5fC. We also show that pocket mutations alter the relative preference of McrB-N to 5mC, 5hmC, and 4mC.

Keywords: 5-formylcytosine; 5-hydroxymethylcytosine; 5-methylcytosine; McrBC; methylome profiling; N4-methylcytosine

5-methylcytosine (5mC) is the primary epigenetic mark in vertebrates and plants. It is related to a repressed chromatin state and inhibition of transcription, explaining why establishment, maintenance, readout, and erasure of cytosine methylation patterns is of key importance for all cellular functions, in particular differentiation [1]. The less abundant oxidized derivatives of 5mC, 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC) are not only cytosine demethylation intermediates but may also act as independent epigenetic marks [2]. In prokaryotes, cytosine methylation is primarily related to restriction modification (RM) systems, which protect host cells against bacteriophages and other foreign DNA [3]. In a typical scenario, DNA methyltransferase methylates cytosines (at either the 5th or 4th positions, forming 5mC or N4-methylcytosine 4mC) in host DNA at specific positions of short (4–8 bp) recognition sites, rendering it insusceptible to cleavage by the second component of the RM system, restriction endonuclease (REase). In contrast, any foreign, for example, bacteriophage, DNA lacking these modifications is digested by the REase immediately upon entry into the cell. One of the mechanisms that allow bacteriophages to bypass such RM systems is to introduce their own DNA modifications, for example, by cytosine methylation, hydroxymethylation, or glucosylation. As a response, some bacteria have evolved methyl-directed restriction endonucleases, which recognize and cut modified phage DNA [4].

A key step in epigenetic regulation is the ability of dedicated proteins to faithfully discriminate between the nonmodified cytosine and its modified derivatives. Structural studies of cytosine state 'readers' in

Abbreviations
4mC, N4-methylcytosine; 5caC, 5-carboxycytosine; 5fC, 5-formylcytosine; 5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine; EMSA, electrophoretic mobility shift assay; K_D, dissociation constant; RM, restriction-modification; vdW, van der Waals.

eukaryotes and prokaryotes revealed two different strategies for the modified base recognition. Eukaryotic proteins that share the methyl-CpG–binding domain, including MBD1, MBD2, MBD4, and MeCP2, and also the zinc-finger protein Kaiso, recognize modified cytosine in the context of a Watson–Crick base pair [5–9]. In contrast, the SRA (SET- and RING-associated) domains of eukaryotic proteins UHRF1, UHRF2, and SUVH5 flip out the modified base and place it in a protein pocket [10–13]. Structurally related DNA recognition domains are also found in methyl-directed restriction endonucleases of MspJI (specific for 5mC and 5hmC) and PvuRts1I (specific for 5hmC and glucosylated 5hmC) families, which also flip out the modified base [14,15]. Another methyl-directed endonuclease McrBC, despite an unrelated structure of its DNA recognition domain McrB-N, also employs the base-flipping mechanism [16].

Since methylation levels in mammalian cells are dynamic and vary in different types of cells during development, differentiation, aging, and disease, methylation profiling of genomic DNA is among the key problems in current biomedical research. Due to high specificity for methylated cytosines, modification-dependent restriction endonucleases achieved recognition as valuable tools for DNA methylome studies. Experiments were first performed with McrBC from E. coli K12 strain, the first characterized methylcytosine-directed enzyme capable of DNA cleavage [17]. Pioneering qualitative DNA methylation studies [18–20] were succeeded by microarray-based methods [21,22]; subsequently, by using methyl-directed enzymes of the MspJI and PvuRts1I families, the resolution of 5mC/5hmC detection was brought down to a single base pair [23–25].

The E. coli enzyme, McrBC, is still used as a sensitive molecular tool to detect methylated cytosines in DNA, although its relative ability to recognize different cytosine modifications, including 5mC, 5hmC, 5fC, 5caC, and 4mC, has not been reported yet. In this study, we present cocrystal structures of the McrBC specificity domain McrB-N bound to DNA oligoduplexes containing 5hmC, 5fC, and 4mC modifications and characterize the relative affinities of WT McrB-N and mutants to different cytosine variants.

Experimental procedures

DNA oligonucleotides

Oligoduplex substrates used in this study are listed in Table S1. Oligonucleotides with 5fC, 5caC, and 4mC modifications were purchased from IBA Lifesciences, all other oligonucleotides were from Metabion. Radioactive labeling of oligonucleotides used for electrophoretic mobility shift assay was performed with [c-^{33}P]ATP (PerkinElmer) and T4 polynucleotide kinase (Thermo Fisher Scientific). Oligoduplexes were assembled by annealing complementary modified and unmodified strands.

Site-directed mutagenesis

The McrB-N mutants were generated by the QuickChange Mutagenesis (QCM) protocol [26] using the pBAD_McrB-N_C-HisTag plasmid, which encodes the N-terminal domain of E. coli K12 McrB protein with a C-terminal His_6-tag (Uniprot ID: P15005, the sequence of the construct: $McrB_{1\text{-}161}$-LEGHHHHHH) [16] as a template for PCR reactions. The E. coli strain, ER2267, was used as a transformation host, and sequences of the mutant genes were confirmed by Sanger sequencing.

Protein expression and purification

The McrB-N wt and mutant variants were expressed in E. coli DH10B (ara-) cells. Cells were grown in LB medium at 37 °C to A_{600} 0.5–0.8, and protein expression was then induced by adding arabinose to a final 0.2% (w/v) concentration. After 4 h, cells were collected by centrifugation and stored at −20 °C. For protein purification, the frozen cells were resuspended in the sonication buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 7 mM 2-ME, 2 mM phenylmethylsulfonyl fluoride, 25 mM imidazole) and disrupted by sonication. Cell debris was removed by centrifugation, and the protein was purified by chromatography through HisTrap-chelating column (GE Healthcare) and gel-filtration using a 120 mL HiLoad 16/600 Superdex 200 column (GE Healthcare). All purification steps were monitored by SDS/PAGE. The fractions containing purified protein were pooled, dialyzed against the storage buffer (10 mM Tris-HCl (pH 8.0), 200 mM KCl, 0.1 mM EDTA, 1 mM DTT, 50% v/v glycerol), and stored at −20 °C. Protein concentrations were determined from A_{280} measurements using the theoretical extinction coefficients calculated with the ProtParam tool available at http://web.expasy.org/protparam/. Protein concentrations are expressed in terms of monomer.

Protein crystallization

Protein crystallization was performed by sitting drop vapor diffusion method at 293 K. The McrB-N protein in the storage buffer was transferred into a glycerol-free buffer (200 mM KCl, 10 mM Tris-HCl pH 7.5, 0.1 mM EDTA and 0.02% NaN_3) using NAP5 columns (GE Healthcare). The protein was concentrated to 9 mg mL^{-1} (~400 lM) and mixed with appropriate DNA (the protein:DNA ratio, as

in the previous study [16], was kept from 2 : 1 to 2 : 1.1). Crystals with 5hmC/5hmC DNA were obtained by mixing 0.5 l L of the complex (200 l M protein and 100 l M DNA) with an equal volume of the crystallization buffer (10 mM Bis-Tris, pH 5.5, 0.2 M ammonium acetate and 14% PEG4000); crystals with 5fC/C and 4mC/C DNAs were obtained by mixing 1 l L of the appropriate complex (400 l M protein, 220 l M DNA) with an equal volume of the crystallization buffer (0.2 M LiCl and 22% PEG4000 for the 5fC/C DNA, 17% PEG4000 for the 4mC/C DNA). In all cases, crystals appeared within 2 weeks and reached the maximum size within 2 months.

Data collection and structure refinement

Diffraction data for the McrB-N complex with 5hmC/5hmC DNA were collected using an in-house X-ray source (Rigaku-MM-007HF) and a Raxis-IV++ detector. For cryoprotection, the crystals were transferred into a buffer containing 7 mM BisTris (pH 5.5 at 25 °C), 0.15 M ammonium acetate, 12.5% PEG4000, and 27% PEG400.

Diffraction data for the complex with 5fC/C DNA were collected on the I911-3 beamline (MAX-Lab, Lund), and data for the complex with 4mC/C DNA were collected on the P13 beamline (PETRAIII, DESY, Hamburg) without any additional cryoprotection.

Data were processed with XDS [27], SCALA, and TRUNCATE [28]. The structure was solved by molecular replacement with MOLREP [29]. The structure of McrB-N bound to unmodified DNA (PDB ID: 3ssc) was used as the search model. The DNA in each case was built anew using COOT [30]. In all cases, the DNA in the structure did not have a unique orientation, therefore, two alternative DNA orientations were built, each with an occupancy of 0.5. Structure refinement was performed with PHENIX [31]. Data collection and refinement statistics are summarized in Table 1. Coordinates and structure factors of McrB-N structures with 5hmC/5hmC, 5fC/C, and 4mC/C DNAs were deposited into the RCSB protein data bank under PDB IDs 6GCD, 6GCE, and 6GCF, respectively.

Electrophoretic mobility shift assay

DNA-binding was analyzed by the electrophoretic mobility shift assay (EMSA) using ^{33}P or ^{32}P-labeled oligoduplexes with various cytosine variants (Table S1). Radiolabeled DNA (final concentration 500 nM) was incubated with the protein (final concentrations varied from 500 to 5000 nM) for 15 min in 20 l L of the binding buffer (30 mM MES/30 mM histidine, pH 6.0 at 25 °C, 10% v/v glycerol, 0.1 mg mL^{-1} BSA, and 1 mM DTT) at 25 °C. Free DNA and protein–DNA complexes were separated by electrophoresis through 8% w/v acrylamide gels (29 : 1 acrylamide:bisacrylamide in 30 mM MES/30 mM histidine, pH 6.0 at 25 °C, gel length:width:thickness 22 : 15 : 0.1 cm). Samples were separated at room temperature for 60 min at 5 V/cm. Low power consumption during electrophoresis runs (~1.8 W or 110 V 9 ~16 mA per electrophoretic unit containing two gels and 1 L of the MES/histidine electrophoretic buffer) ensured that the gels remained at room temperature (~22 °C). Radiolabeled DNA was detected using Cyclone phosphorimager (Packard Instrument).

EMSA competition experiments

The EMSA samples contained McrB-N protein (final concentration 0.5 l M), radiolabeled methylated 5mC/C DNA (0.5 l M), and variable amounts (final concentrations 0.5–5.0 l M) of unlabeled competitor DNA (carrying unmodified C or various cytosine modifications, Table S1). Electrophoresis was run as described above. The amounts of the radiolabeled protein–DNA complexes (i.e., the complex with the 5mC/C DNA) at different competitor concentrations were determined by densitometric analysis of EMSA gel images using OptiQuant software (Packard Instrument). Competition data were analyzed as described in ref. [32], providing dissociation constant for the protein—radiolabeled 5mC/C DNA complex ($K_{D(5mC)}$, in l M), and dissociation constant for protein—unlabeled competitor DNA interaction (K_D, in l M). All competition experiments were repeated several times (raw data provided in Fig. S1–6 and Table S2), and the results are reported as the average binding constants ($1/K_{D(5mC)}$ and $1/K_D$, in l M^{-1}) 1 SD.

Fluorescence anisotropy measurements

Anisotropy measurements were performed using oligoduplex substrates carrying a HEX (hexachloro-fluorescein, Table S1) label at the 5′-terminus of the unmodified DNA strand. The measurements were performed on a Fluoromax-3 spectrofluorimeter equipped with polarizers. Excitation wavelength was 530 nm (2 nm slit), emission wavelength was 570 nm (16 nm slit), measurements were performed at 25 °C in the measurement buffer (100 mM KCl, 20 mM Mes-KOH pH 6.0). Protein was titrated into the 160-l L cuvette containing the DNA and buffer (final protein concentrations were from 20 to 1600 nM, DNA concentration was 4 nM). The measured anisotropy values were corrected for the background anisotropy signal measured with single-stranded HEX-labeled DNA and plotted against protein concentration. The resultant dependencies were described by the Eqn (1)

$$A([p]) = A_0 + A_1 \cdot [p]/([p] + K_D) \quad (1)$$

where A([p]) is the observed anisotropy value at protein concentration [p], A_0 is anisotropy observed with free DNA, A_1 is the amplitude of anisotropy increase, and K_D is the dissociation constant of protein–DNA

Table 1. Data collection and refinement statistics.

DNA	5hmC/5hmC	5fC/C	4mC/C
Data collection[a]			
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1	P 2_1 2_1 2_1
a (Å)	36.07	36.18	36.28
b (Å)	68.93	69.99	69.68
c (Å)	143.71	144.27	145.05
a/b/c (°)	90/90/90	90/90/90	90/90/90
Wavelength (Å)	1.54	1.001	1.0332
X-ray source	Rigaku R-AXIS IV	MAX-Lab/I911-3	DESY/PETRAIII/P13
Total reflections	624757 (66917)	308176 (30022)	354537 (33117)
Unique reflections	34155 (3347)	49273 (4859)	54386 (5309)
Resolution range (Å)	23.88–1.80 (1.864–1.80)	35.09–1.600 (1.657–1.600)	72.535–1.55 (1.605–1.55)
Completeness (%)	99.83 (100)	99.62 (99.92)	99.88 (99.16)
Multiplicity	18.3 (20.0)	6.3 (6.2)	6.5 (6.2)
Mean I/r (I) (%)	48.73 (5.48)	17.15 (2.39)	28.36 (5.38)
R (merge) (%)	6.6 (45.1)	5.40 (44.8)	4.035 (40.23)
B(iso) from Wilson (Å^2)	22.80	29.46	21.29
Refinement			
Resolution range (Å)	23.88–1.80	35.09–1.600	62.81–1.55
Reflections work (nonanomalous)/test	34121/3435	49205/5002	54378/5509
Macromolecule/solvent atoms	3586/218	3485/370	3402/264
Average B factors: Macromolecule/solvent/all atoms	26.1/33.2/28.6	26.0/35.5/29.5	27.8/39.5/31.6
R-factor/R-free	0.166/0.214	0.168/0.206	0.139/0.177
R.m.s.d. bond lengths (Å)/angles (deg)	0.018/1.53	0.010/1.06	0.016/1.45
Ramachandran plot (%) favored/allowed/outliers	95/5/0	97/3/0	97/3/0
PDB ID	6GCD	6GCE	6GCF

[a] Data in parenthesis—outer shell.

complex. Nonlinear fitting was performed using KyPlot software (Version 2.0 beta 14) [33] with A_0, A_1, and K_D as fit parameters. In cases where parameters A_1 and K_D were poorly defined by experimental data (weak protein interactions with certain DNA variants), amplitude A_1 was set to a fixed value determined with methylated (5mC/C) DNA. The results are reported as the average binding constants ($1/K_D$, in l_M ^{1}) determined from several separate titrations ± 1 SD (raw data provided in Fig. S7–11 and Table S3). Different batches of McrB-N mutant V66N in this assay produced erroneous results (no detectable increase in the anisotropy signal).

Results

Novel structures of McrB-N

We have determined X-ray structures of McrB-N bound to DNA oligoduplexes containing 5-hydroxymethylcytosine, 5-formylcytosine, and N4-methylcytosine DNA modifications. The crystals were grown under similar conditions as previously characterized crystals of McrB-N bound to C and 5mC DNA [16]. Due to successful optimization of crystallization conditions, the resolution of the newly solved structures is considerably higher (1.55–1.80 Å, Table 1, vs 2.1–2.7 Å, ref. [16]).

We find that McrB-N flips-out 5hmC, 5fC, and 4mC bases (whether the base extrusion is an active or a passive process is still unknown) in a similar fashion to 5mC (Fig. 1). In all cases the flipped bases occupy identical positions in the protein pocket, with the O2, N3, and N4 atoms forming H-bonds to backbone and side chain atoms of McrB-N residues T85, D84, and I82 (Fig. 1).

The 5-methyl groups of 5hmC and 5fC make vdW contacts to the surrounding W49, Y64, L68, and Y117 residues. Due to a shifted location, the N4-methyl group of 4mC also makes vdW interactions to V66 and S120, but not to W49 (Fig. 1). The hydroxyl group oxygen of 5hmC is located outside the plane of the aromatic ring, and makes vdW contacts to V66 and the hydroxyl oxygen of Y64. Formation of an H-bond between 5hmC and Y64 hydroxyl groups is unlikely due to unfavorable geometry. The formyl oxygen of 5fC is located in the plane of the aromatic ring, and in addition to the intramolecular H-bond to 5fC 4-aminogroup, forms vdW interactions with Y117, L68, Y64, and V66 residues (Fig. 1).

Pocket mutants

Distinct types of interactions made by alternative cytosine modifications prompted us to engineer pocket mutations that would alter the relative preference of McrB-N to different cytosine variants. We have made three groups of point mutations at conserved positions of McrB-N pocket:

#1 Y117L, Y64F—mutations that increase the size of the flipped base–binding pocket;
#2 V66L, V66I—mutations that reduce the pocket size;
#3 V66D, V66N, V66Q, L68N, L68Q, S120K, S120E, S120Q—mutations that introduce additional H-bond donors and/or acceptors on various faces of the pocket surface.

Some of the #3 group mutants (V66D, L68Q, S120E, S120Q, S120K) could not be purified due to negligible yields; the yields of the #1 group mutant Y117L and #3 group mutants V66N, V66Q, and L68N were also considerably reduced in comparison to WT McrB-N, but we were able to purify sufficient amounts of proteins for biochemical experiments.

The initial characterization of the mutant domains was performed using EMSA assay (conditions as in ref. [16]) on DNAs oligoduplexes containing 5mC, 5hmC, 5fC, and 5caC modifications, or an unmodified C (Fig. 2). We find that WT McrB-N readily interacts with 5mC, 5hmC, and 4mC DNAs, but not with C, 5fC, or 5caC DNAs (Fig. 2). The group #1 mutant Y117L failed to form complexes with any of the tested DNAs, implying that large aromatic pocket residues that make stacking interaction with the flipped base are an absolute requirement for DNA-binding by McrB-N. No binding was also observed with the V66Q mutant, suggesting that extra pocket space occupied by the long glutamine side chain may prevent base flipping with all cytosine variants. On the other hand, we find that #1 group mutant, Y64F, the #2 group mutants, V66L and V66I, and the #3 mutants, V66N and L68N, formed variable amounts of complexes with 5mC, 5hmC, and 4mC DNAs, but none interacted with C, 5fC, or 5caC DNAs (Fig. 2).

Selectivity of WT McrB-N and pocket mutants to modified cytosine variants

More thorough examination of DNA-binding affinities and relative preferences was performed with the WT domain and McrB-N mutants Y64F, V66I, V66L, and V66N, employing either an EMSA-based competition assay (Fig. 3B), or fluorescence anisotropy measurements (Fig. 3C; raw data are provided in Figs. S1–S11 and Tables S2–S3). Since the V66N mutant produced erroneous results in the fluorescence anisotropy assay (no increase in the anisotropy signal upon titration), we have also characterized the double mutant Y64F_V66N.

Competition experiments indicate that the preferred modification for WT McrB-N is N4-methylcytosine (dissociation constant K_D 0.1 l M, $1/K_D$ 10 l M ^{1} Fig. 3B). Binding is 2.5–5-fold weaker to 5mC and 5hmC DNAs (K_Ds ~0.25 and 0.5 l M, respectively), and negligible to 5fC and 5caC (K_D > 4 l M, or $1/K_D$ < 0.25 l M ^{1}). Qualitatively and quantitatively similar results, despite different buffer conditions, were also obtained using the anisotropy measurements (Fig. 3C).

Mutation V66I, which reduces the flipped base-binding pocket, decreased the affinity of McrB-N to 5mC, 5hmC and 4mC modifications to a similar extent (~4-fold decrease in $1/K_D$ values, Fig. 3B,C), leaving the relative preferences for distinct cytosine variants unchanged. A similar decrease in affinity without well-defined changes in the relative preferences was also observed for the V66L mutant.

Mutation Y64F, which removes a single hydroxyl group from the pocket, had virtually no effect on the McrB-N interaction with 5mC and 5hmC DNA, but reduced threefold the affinity to 4mC DNA (Fig. 3B,C). Even more pronounced discrimination against 4mC was observed for the V66N and Y64F_V66N mutants (binding to 4mC, K_D 2–4 l M, is comparable to 5fC and 5caC, Fig. 3B,C). The V66N mutation had a limited negative effect on 5mC DNA-binding (the K_D increased ~2.5-fold to 0.6 l M, Fig. 2B), but had a positive effect on 5hmC binding (0.4–0.6 l M K_Ds, as determined in different assays, are either unchanged or up to 1.5-fold lower in comparison to WT McrB-N, Fig. 3B). This makes 5hmC the preferred DNA modification for the V66N and Y64F_V66N McrB-N mutants.

Discussion

Relative affinity of McrB-N to different DNA modifications

Cocrystal structures of WT McrB-N were obtained with DNA oligoduplexes containing unmodified cytosine, 5mC (ref. [16]), 5hmC, 5fC, and 4mC (this study). In all cases, the (modified) cytosine base is placed into the protein pocket where it occupies an identical position, which allows formation of cytosine-specific hydrogen bonds between the Watson–Crick edge of the cytosine and the McrB-N pocket

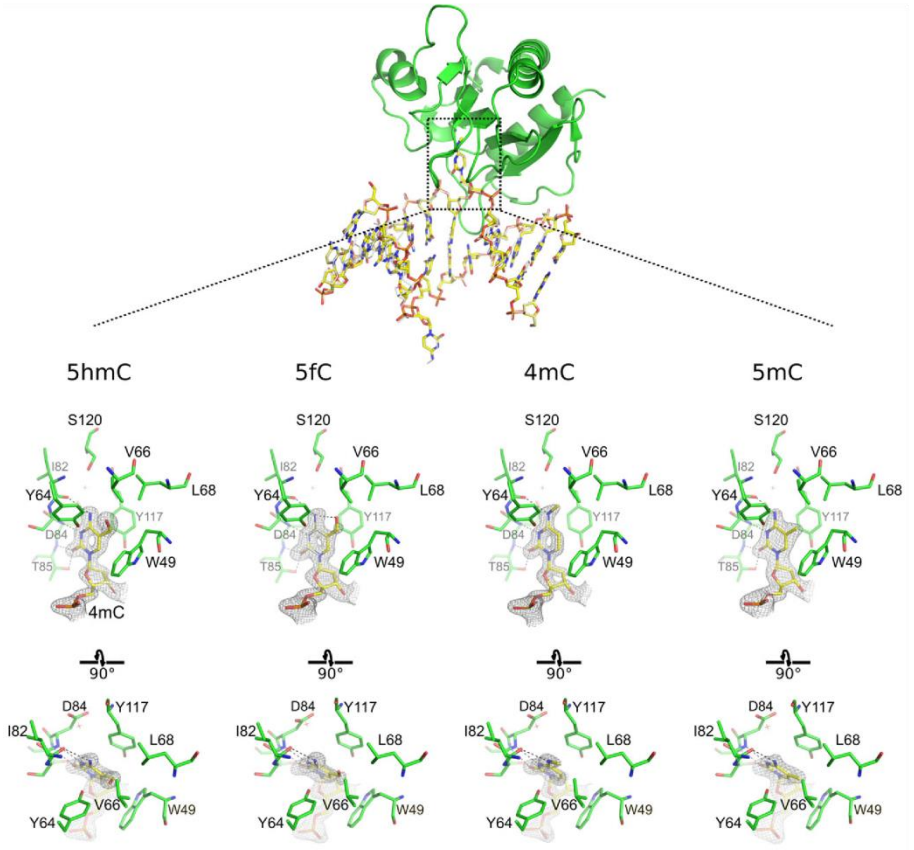


Fig. 1. Structures of McrB-N with 5hmC, 5fC, 4mC, and 5mC DNA. Top: the overall structure of McrB-N bound to 4mC DNA (PDB ID 6gcf). Only one protein subunit (chain A) is depicted; the DNA is shown in stick representation (yellow; the model of an alternative DNA orientation is shown as white sticks). Bottom: the flipped out base-binding pocket in the structures with 5hmC, 5fC, and 4mC DNA (PDB IDs: 6gcd, 6gce, and 6gcf, respectively); the structure with 5mC DNA (PDB ID: 3ssc, ref. [16]) is also shown. In all panels, the 2mFo-DFc electron density map contoured at a 2.0 r (5hmC, 4mC) or 1.3 r (5fC, 5mC) level is shown for the flipped out nucleotides.

residues (Fig. 1A). Here, we show that despite a similar DNA distortion, the affinity of McrB-N to DNA target sequences with various cytosine variants differs dramatically, with the values of equilibrium dissociation constants (K_Ds) ranging from 0.1 lM (4mC) to over 4 lM (C, 5fC, 5caC), Fig. 3B,C. We attribute these differences in affinity to two major factors: (a) different energetic cost for the disruption of the cytosine derivative:guanine base pair in duplex DNA; and (b) different energetic gain for placing the flipped base in a hydrophobic protein pocket, defined by the energy of the newly formed van der Waals interactions and H-bonds, extent of steric clashes, and the energetic cost for the flipped base and protein pocket dehydration.

The smallest interaction surface between the flipped base and the protein pocket, and, consequently, the lowest gain in vdW interaction energy upon base flipping, is observed for the unmodified cytosine C; this correlates with the lack of interactions with the C

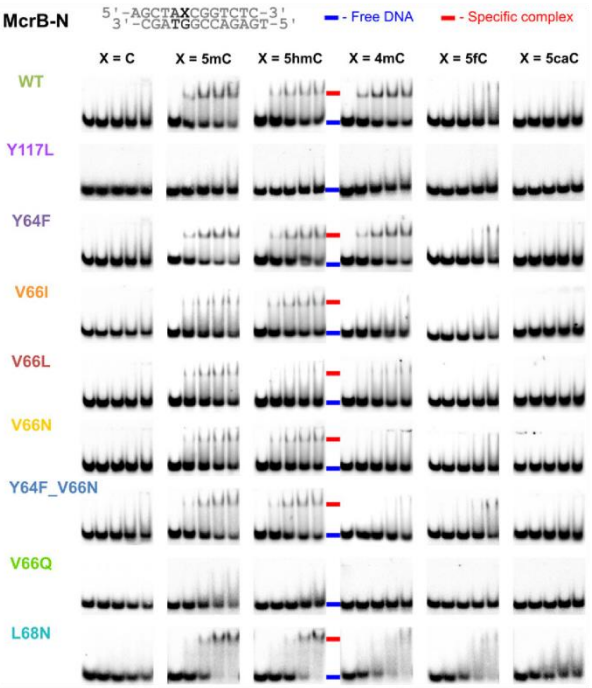


Fig. 2. Characterization of pocket mutants by electrophoretic mobility shift assay. Reactions contained 0, 0.5, 1, 2, 5 l M protein and 0.5 l M DNA carrying various cytosine variants. Experiments were performed as described in Experimental Procedures.

DNA (Fig. 2) and lack of unmodified DNA cleavage by the McrBC complex [17]. The specific-like binding and base flipping observed in the cocrystal structure with the C DNA (PDB ID: 3sse, ref. [16]), therefore, must be a result of high protein/DNA concentrations used in the crystallization experiments and the effects of crystal packing forces.

In contrast, cytosine variants 4mC, 5mC, 5hmC, and 5fC have a larger contact surface with the pocket walls thanks to the extra methyl, hydroxymethyl, and formyl groups (Fig. 1). In particular, the extra methyl groups of 4mC and 5mC make contacts to the pocket residues V66, L68, and S120 (4mC) or W49 (5mC). The ~2.5-fold difference in the K_Ds observed for 4mC and 5mC DNAs corresponds to a 2 kJ mol ^{1} difference in the binding energy DG, a value that does not exceed the energy of a single H-bond. Thus, the observed preference for 4mC must be due to small differences in the extent of vdW interactions, lower stability of the 4mC:G base pair as judged by lower melting temperature of 4mC-containing oligoduplexes [34], and lower dehydration energy of 4mC compared to 5mC (due to the N4-methyl group 4mC is expected to form less H-bonds to water molecules that must be broken upon insertion into the protein pocket).

5hmC, 5fC, and 5caC have larger modifications than 4mC and 5mC, but the extra size, and potentially larger interaction surface with the protein pocket do not translate into tighter binding (Fig. 3). The affinity of McrB-N to 4mC DNA is fivefold higher compared to 5hmC DNA (~4 kJ mol ^{1} difference in DG), and more than 40-fold higher compared to 5fC and 5caC DNA (~10 kJ mol ^{1} difference in DG, Fig. 3B,C). The solved structures with 5hmC and 5fC DNAs did not reveal any steric clashes between the flipped bases and the protein pocket; moreover, the extent of vdW interactions formed by the flipped 5hmC and 5fC bases is similar, neither base is capable of forming additional H-bonds, and none is reported to increase DNA duplex stability as judged by imino proton exchange rates [35]. This implies that slightly lower

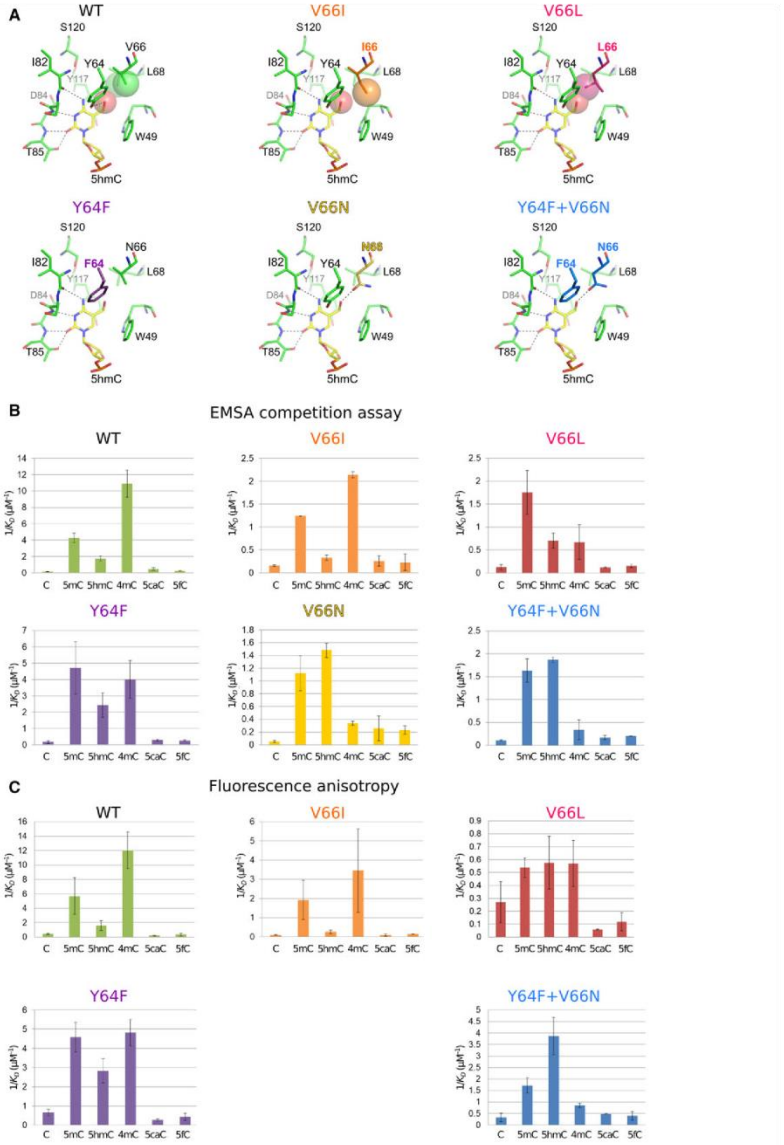


Fig. 3. Selectivity of McrB-N pocket mutants. (A) Models of flipped out 5hmC interactions with various McrB-N mutants. Experimentally determined DNA-binding constants (in l_M^{-1}) of McrB-N mutants with various DNA variants are shown in panels B (EMSA competition assay) and C (fluorescence anisotropy measurements). Error bars correspond to SD; raw data are provided in Fig. S1–11 and Tables S2–S3.

affinity of McrB-N to 5hmC DNA, and very low affinity to 5fC DNA must be due to other factors, primarily higher dehydration energy of polar hydroxymethyl and especially formyl groups, and possibly differences in the energies of vdW interactions due to distinct polarization properties of 5hmC and 5fC.

In contrast, the lack of interactions with 5caC DNA could simply be due to steric clashes (modeling, based on the McrB-N structure with 5fC DNA suggests that an additional 5caC carboxyl oxygen would be as close as 2.6 Å to the W49 residue), and increased stability of DNA duplex containing 5caC:guanine base pairs [35].

Pocket mutations

We show that the relative preference of McrB-N to DNAs containing different cytosine modifications may be altered by replacement of residues forming the walls of the flipped base-binding pocket. The most prominent effect is caused by the V66N mutation, which essentially abrogates McrB-N binding to 4mC DNA, and concomitantly has a positive effect on 5hmC DNA-binding (Fig. 3). The former effect is likely due to the fact that the N4-methyl group of 4mC is unable to form an energetically favorable interaction with the N66 asparagine; in contrast, 5hmC may form an additional H-bond (Fig. 3A).

Selectivity of the base-flipping 5(h)mC readers

The observed range of McrB-N affinities to cytosine derivatives, and the changes in the binding preferences due to mutations, is limited to ~40-fold (Fig. 3). However, even such mediocre discriminating power is on par with, or exceeds these reported for eukaryotic SRA domains UHRF1 (a ~3.5-fold preference for 5mC compared to unmodified C [13]) and UHRF2 (a bona fide 5hmC reader in eukaryotes [2], a ~1.5-fold preference for 5hmC compared to C [13]). A much higher discriminating power, as judged from the activity measurements, is expected for the SRA-like domains found in the prokaryotic methyl-directed endonucleases of MspJI and PvuRts1I families [24,25]. For example, MspJI-like enzymes cut DNA containing 5mC and 5hmC modifications in a specific sequence context, and, like the McrBC complex, are inactive on unmodified DNA, while PvuRts1I-like enzymes have a 250-fold or higher preference for 5hmC or glucosylated 5hmC as compared to 5mC and unmodified C [25]. Currently available structures of 5mC-bound MspJI enzyme [15], and the apo-structures of PvuRts1I family [36,37] enzymes, however, provide no obvious explanation for the higher selectivity of their SRA-like domains.

Conclusions

Despite identical positioning of the flipped base observed in the cocrystal structures of McrB-N with DNAs carrying various cytosine modifications, we find that McrB-N has relatively high affinity for 4mC, 5mC, and 5hmC, but low affinity for 5fC. Strong discrimination against 5fC, which is almost isosteric to 5hmC, argues that modified cytosine recognition selectivity of McrB-N, and likely other base-flipping modified cytosine 'readers', in addition to intrapocket contacts, is determined by other factors, such as the stability of the flipped base pair and the flipped base/protein pocket dehydration energies. Low affinity to 5fC and 5caC vs high affinity to 5mC and 5hmC also implies that methylome profiling studies that utilized McrBC as a tool for DNA fragmentation [18–22] reported the distribution of 5mC and 5hmC, but not 5fC and 5caC. It remains to be seen if McrB-N could be tailored for more selective recognition of individual cytosine variants.

Acknowledgements

The authors thank MAX-lab and DESY staff for assistance with crystallographic data collection, and E. Rudokiene for sequencing services. This work was supported by the Research Council of Lithuania (grant number MIP-027/2012 to GS). GS acknowledges the travel support provided by the 'Baltic Science Link' project coordinated by the Swedish Research Council. The funding sources had no role in the study design, data interpretation, and the decision to submit the article for publication.

Author contributions

GS conceived and supervised the study and wrote the manuscript; EZ, EM, and GS designed experiments, performed experiments and analyzed data.

References

1 Smith ZD and Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14, 204–220.

2 Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, Münzel M, Wagner M, Müller M, Khan F et al. (2013) Dynamic readers for 5-

(hydroxy)methylcytosine and its oxidized derivatives. Cell 152, 1146–1159.
3 Pingoud A, Wilson GG and Wende W (2014) Type II restriction endonucleases–a historical perspective and more. Nucleic Acids Res 42, 7489–7527.
4 Labrie SJ, Samson JE and Moineau S (2010) Bacteriophage resistance mechanisms. Nat Rev Microbiol 8, 317–327.
5 Ohki I, Shimotake N, Fujita N, Jee J, Ikegami T, Nakao M and Shirakawa M (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell 105, 487–497.
6 Scarsdale JN, Webb HD, Ginder GD and Williams DC (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. Nucleic Acids Res 39, 6741–6752.
7 Otani J, Arita K, Kato T, Kinoshita M, Kimura H, Suetake I, Tajima S, Ariyoshi M and Shirakawa M (2013) Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. J Biol Chem 288, 6351–6362.
8 Ho KL, McNae IW, Schmiedeberg L, Klose RJ, Bird AP and Walkinshaw MD (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. Mol Cell 29, 525–531.
9 Buck-Koehntop BA, Stanfield RL, Ekiert DC, Martinez-Yamout MA, Dyson HJ, Wilson IA and Wright PE (2012) Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc Natl Acad Sci USA 109, 15229–15234.
10 Arita K, Ariyoshi M, Tochio H, Nakamura Y and Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455, 818–821.
11 Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, Arrowsmith CH and Dhe-Paganon S (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455, 822–825.
12 Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE and Cheng X (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455, 826–829.
13 Zhou T, Xiong J, Wang M, Yang N, Wong J, Zhu B and Xu R-M (2014) Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol Cell 54, 879–886.
14 Zagorskaite E and Sasnauskas G (2014) Chemical display of pyrimidine bases flipped out by modification-dependent restriction endonucleases of MspJI and PvuRts1I families. PLoS ONE 9, e114580.
15 Horton JR, Wang H, Mabuchi MY, Zhang X, Roberts RJ, Zheng Y, Wilson GG and Cheng X (2014) Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. Nucleic Acids Res 42, 12092–12101.
16 Sukackaite R, Grazulis S, Tamulaitis G and Siksnys V (2012) The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. Nucleic Acids Res 40, 7552–7562.
17 Sutherland E, Coe L and Raleigh EA (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. J Mol Biol 225, 327–348.
18 Chotai KA and Payne SJ (1998) A rapid, PCR based test for differential molecular diagnosis of Prader-Willi and Angelman syndromes. J Med Genet 35, 472–475.
19 Lyko F, Ramsahoye BH, Kashevsky H, Tudor M, Mastrangelo M-A, Orr-Weaver TL and Jaenisch R (1999) Mammalian (cytosine-5) methyltransferases cause genomic DNA methylation and lethality in Drosophila. Nat Genet 23, 363–366.
20 Burman RW, Yates PA, Green LD, Jacky PB, Turker MS and Popovich BW (1999) Hypomethylation of an expanded FMR1 allele is not associated with a global DNA methylation defect. Am J Hum Genet 65, 1375–1386.
21 Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, Richard McCombie W, Lavine K, Mittal V, May B, Kasschau KD et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430, 471–476.
22 Lippman Z, Gendrel A-V, Colot V and Martienssen R (2005) Profiling DNA methylation patterns using genomic tiling microarrays. Nat Methods 2, 219–224.
23 Szwagierczak A, Brachmann A, Schmidt CS, Bultmann S, Leonhardt H and Spada F (2011) Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine. Nucleic Acids Res 39, 5149–5156.
24 Zheng Y, Cohen-Karni D, Xu D, Chin HG, Wilson G, Pradhan S and Roberts RJ (2010) A unique family of Mrr-like modification-dependent restriction endonucleases. Nucleic Acids Res 38, 5527–5534.
25 Wang H, Guan S, Quimby A, Cohen-Karni D, Pradhan S, Wilson G, Roberts RJ, Zhu Z and Zheng Y (2011) Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. Nucleic Acids Res 39, 9294–9305.
26 Zheng L, Baumann U and Reymond J-L (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res 32, e115.
27 Kabsch W (2010) XDS. Acta Crystallogr D Biol Crystallogr 66, 125–132.

28 CCP4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50, 760–763.
29 Vagin A and Teplyakov A (2010) Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr 66, 22–25.
30 Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60, 2126–2132.
31 Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH and Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr 68, 352–367.
32 Sasnauskas G, Kauneckaite K and Siksnys V (2018) Structural basis of DNA target recognition by the B3 domain of Arabidopsis epigenome reader VAL1. Nucleic Acids Res 46, 4316–4324.
33 Yoshioka K (2002) KyPlot as a Tool for Graphical Data Analysis. Compstat, pp. 37–46. Physica-Verlag HD, Heidelberg.
34 Butkus V, Klimasauskas S, Petrauskiene L, Maneliene Z, Janulaitis A, Minchenkova LE and Schyolkina AK (1987) Synthesis and physical characterization of DNA fragments containing N4-methylcytosine and 5-methylcytosine. Nucleic Acids Res 15, 8467–8478.
35 Szulik MW, Pallan PS, Nocek B, Voehler M, Banerjee S, Brooks S, Joachimiak A, Egli M, Eichman BF and Stone MP (2015) Differential stabilities and sequence-dependent base pair opening dynamics of Watson-Crick base pairs with 5-hydroxymethylcytosine, 5-formylcytosine, or 5-carboxylcytosine. Biochemistry 54, 1294–1305.
36 Horton JR, Borgaro JG, Griggs RM, Quimby A, Guan S, Zhang X, Wilson GG, Zheng Y, Zhu Z and Cheng X (2014) Structure of 5-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA. Nucleic Acids Res 42, 7947–7959.
37 Kazrani AA, Kowalska M, Czapinska H and Bochtler M (2014) Crystal structure of the 5hmC specific endonuclease PvuRts1I. Nucleic Acids Res 42, 5929–5936.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.
Table S1. Oligonucleotides used in this study.
Tables S2-S3. McrB-N – DNA dissociation constants determined by EMSA competition assay and fluorescence anisotropy.
Figs. S1-S6. EMSA competition assay, raw data.
Figs. S7-S11. Fluorescence anisotropy, raw data.

NOTES

Vilniaus universiteto leidykla
Universiteto g. 1, LT-01513 Vilnius
El. p. info@leidykla.vu.lt,
www.leidykla.vu.lt
Tiražas 12 egz.