Molecular basis and modulation of target DNA recognition in the Type IV restriction endonuclease McrBC

A Thesis submitted in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

Indian Institute of Science Education and Research Pune

By

Nevin Korath Zacharia

20151084

Under supervision of

Dr Saikrishnan Kayarat

Associate Professor

Department of Biology

IISER Pune

Thesis Advisor

Dr Jeetender Chugh

Assistant Professor

Department of Chemistry

IISER Pune

Certificate

This is to certify that this dissertation entitled "Molecular basis and modulation of target DNA recognition in the Type IV restriction endonuclease McrBC " towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents work carried out by Nevin Korath Zacharia at IISER Pune under the supervision of Dr Saikrishnan Kayarat, Associate Professor, Department of Biology during the academic year 2019-2020.

Committee:
Supervisor: Dr Saikrishnan Kayarat
Thesis advisor: Dr Jeetender Chugh

Dr Saikrishnan Kayarat
Associate Professor
Department of Biology
IISER Pune

Declaration

I hereby declare that the matter embodied in the report “Molecular basis and modulation of target DNA recognition in the Type IV restriction endonuclease McrBC” are the results of the work carried out by me at the Department of Biology, Indian Institute of Science Education and Research, Pune, under the supervision of Dr Saikrishnan Kayarat and the same has not been submitted elsewhere for any other degree

Date: 10-04-2020

Nevin Korath Zacharia
5th Year BS-MS Student
IISER Pune

Table of Contents

List of Tables

List of Figures

Abstract

Target DNA recognition is a crucial factor that drives the cellular functions carried out by DNA binding proteins. An increasing number of DNA binding proteins that read epigenetic modifications on DNA bases and certain DNA sequences in the genome are found to use a mechanism of DNA base flipping, wherein the target base is rotated out of the DNA helix into a recognition pocket of the enzyme. One feature that is common to these base flipping enzymes is the presence of a protein residue that intercalates into the DNA cavity and fills the space left out by the flipped base. To understand the role of this residue in target DNA recognition and in the base flipping mechanism in DNA reader proteins, a restriction enzyme McrBC which utilizes DNA base flipping for recognizing methylcytosine in DNA, was used as the study system. A series of mutants of this enzyme was studied, which carried amino acids of varying size and hydrophobicity as the DNA intercalating residue. Analysis of the target recognition in these mutants through enzyme activity assays and DNA binding assays indicated that modulation of the base flipping process by specifically varying the size and hydrophobicity of the intercalating residue resulted in generation of enzyme constructs that differ in target recognition and enzymatic efficiencies. These observations indicate that the intercalating residue directly plays a role in mediating target recognition and enzymatic efficiency in DNA reader proteins by affecting base flipping. These insights will have further implications in the field of enzyme engineering and protein design for base flipping proteins.

Acknowledgements

I am extremely thankful to my guide Dr Saikrishnan Kayarat for the guidance and help he provided during every step of this project. I am thankful to him for the time he has dedicated in training me to think and work independently and for motivating me in times of need. Along with this, I would like to thank Dr Gayathri Pananghat for the valuable inputs and suggestions that she gave during the lab meetings.

I would like to express my gratitude to Vishal Adhav, who was a constant support, both personally and professionally during my entire time in the lab. I thank him for all the assistance and help that he provided for my work and for his helpful and supportive attitude towards any troubles that I faced. I am grateful to all the seniors in the lab, especially to Pratima, for the training and help I received from her during my initial days in the lab and to Mahesh for helping me with the anisotropy experiments in spite of his busy schedule. I thank Basila, Manil, Jazleena and all other members of Dr Saikrishnan's and Dr Gayathri's lab for providing a very friendly and pleasant work environment.

I would like to thank my friends at IISER for making the past five years here memorable and enjoyable. Finally, I would like to thank my parents and my brother for their constant love and support.

Chapter 1 **Introduction**

DNA binding proteins play vital roles in various cellular processes such as gene expression, DNA repair and genome defense in all forms of life. Interaction between these proteins with the DNA is crucial for these processes. In general, these interactions can be specific i.e. dependent on particular DNA base sequences or non-specific; independent of DNA sequences. For example, processes like gene expression involve transcription factors that shows very high specificity to a particular target DNA sequence, whereas enzymes involved in DNA repair should have minimal DNA sequence specificity, but an enhanced specificity to DNA containing damaged bases. On the other end, there are proteins such as histones that bind DNA non-specifically to carry out its function (Jakubec et al., 2016).

In general, DNA sequence dependent interactions are made by proteins by a direct DNA base readout (Rohs et al., 2010). X-ray crystallography and other techniques have been used to dissect out the exact mechanisms of base reading. One such mechanism is DNA base flipping, which is a highly conserved mechanism employed by various enzymes that need to access DNA bases buried inside the helix. In the process of DNA base flipping, an intra-helical base in normal B-DNA is rotated completely out of the helix into an extra-helical region, which is generally an active site or binding pocket of enzymes (Roberts et al., 1998). This mechanism allows enzymes to read a DNA base and perform a chemistry on the base, if needed. The first evidence of DNA base flipping was observed in the crystal structure of M.Hha1 methyltransferase bound to its DNA substrate (Klimasauskas et al., 1994, **Figure 1.1**), where a cytosine was flipped out into the active site of the enzyme. This allowed the enzyme to add a methyl group to the cytosine, which could not have been possible if the cytosine was buried inside the DNA helix. Further studies identified that this process is a common mechanism in other methyltransferases and was also observed for DNA glycosylases which also needed access to the DNA base for cleavage of the glycosylic bond (Slupphaug et al., 1996). Recently the crystal structure of human UHRF1 bound to DNA (Arita et al., 2008) showed the first instance of DNA base flipping used solely for the purpose to read

DNA status, where this mechanism was used to recognize the methylation status of cytosine in DNA. Since then, several eukaryotic proteins involved in reading epigenetic marks in DNA have been shown to use base flipping for DNA readout.

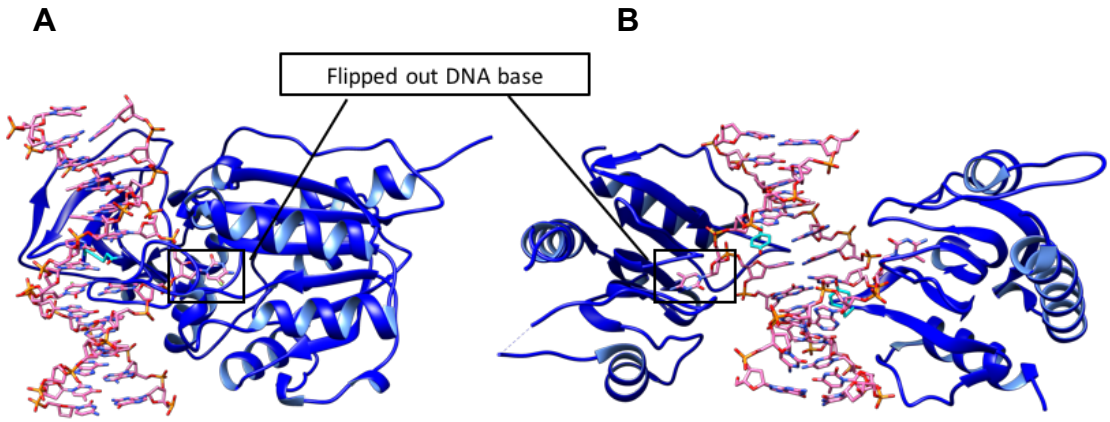


Figure 1.1 Crystal structures of proteins that uses base flipping mechanism (**A**) Hha1 methyltransferase with flipped out cytosine. (**B**) Nter-McrB with flipped out methylcytosine.

Studies on the base flipping mechanism in various proteins over the years, have led to the proposal that the stability of the flipped conformation is the parameter that determines whether the enzyme will recognize the DNA base or DNA sequence as its target or not. In presence of enzyme, the enzyme actively promotes the flipping out of bases form the DNA helix. If the flipped out conformation does not have the required stability, the base will revert back into the intra-helical region, causing the enzyme to dissociate from DNA and prevent enzymatic activity on that DNA. On the other hand, if the flipped out conformation is stable, it will result in a stronger binding to the DNA leading to target recognition and enzymatic activity on the DNA. A study that specifically looked at the temporal coordination of DNA binding and base flipping in EcoR1 Methyltransferase concluded that both processes occur simultaneously (Allan et al., 1999).

Common features that contribute to stability of the base flipping mechanism

Crystal structures have shown several aspects that confer stability to the flipped base, irrespective of the function or class of proteins. First is the presence of

phosphate contact between the residues of the enzyme and the DNA backbone, which induces structural changes in DNA such as alterations in DNA phosphate backbone dihedral angles or widening of DNA major groove (Sukackaite et al., 2012, Slupphaug et al., 1996). These structural changes are proposed to be the earliest steps in the process, which makes it favorable for the flipping of the base out of the helix. Indeed for the EcoR1 DNA methyltransferase, it has been shown that DNA bending by the enzyme increases the rate of base flipping (Allan et al., 1999).

Second is the protrusion of an amino acid residue of the enzyme into the DNA helix, which occupies the void left in the helix by the base that is flipped. Looking at various base flipping enzymes, it is seen that the identity of the intercalating residue is not conserved. It ranges from Gln in Hha1, Leu in Uracil DNA Glycosylase (UDG), Arg in Thymine DNA Glycosylase (TDG) to Tyr in McrBC (Klimasauskas et al., 1994, Slupphaug et al., 1996, Maiti et al., 2009; Sukackaite et al., 2012). Studies aimed at deciphering the temporal coordination of the various steps of the base flipping process has shown that in UDG, the intercalation of this Leu residue precedes the flipping of the base (Wong et al., 2002). These observations led to the proposal that these intercalating residues have an active role in expelling the base out of the DNA helix. Studies using molecular modelling and steady-state kinetics on Hha1 has proposed an active role of the intercalating Gln in the base flipping process (Daujotyt et al., 2004). Here, in the initial stages of the base flipping process, this residue forms steric clashes with the intra-helical base to destabilize the native state of DNA and lowers the energy barrier for base flipping to occur. Subsequently, this residue provides stability to the base flipped protein DNA complexes by forming stabilizing interactions with the DNA bases. These observations has led to various studies looking at the effects of mutating this DNA intercalating residue of various base flipping enzymes. Mutation of intercalating residue to Ala has shown to reduce the catalytic efficiency of Hha1 (Daujotyt et al., 2004), TDG (Maiti et al., 2009) and UDG (Slupphaug et al., 1996), and it has been also shown that this reduced catalytic efficiency in TDG and UDG is due to a reduction in DNA binding affinities of the enzyme. In the case of DNA reader proteins, mutagenesis has shown to completely abolish DNA binding capability of Arabidopsis SUVH5 and DNA binding domain of McrB (Rajakumara et al., 2011, Sukackaite et al., 2012). Taken together, these findings speculate the intercalating residue to play an active role in the base flipping

process, by affecting the stability of the flipped base and contributing to enzymatic efficiency.

The third main component in this process is the protein pocket to which the base is flipped. It has been observed that the structure and amino acid composition of this pocket preferentially stabilizes only the target DNA base in the flipped state, mainly by forming a network of hydrogen bonds and nonpolar contacts with the cognate base and by providing a binding pocket that structurally accommodates the cognate base. Mutagenesis of an Asp residue in the protein pocket of UDG has shown to broaden the base specificity of the enzyme (Kimber et al., 2014), indicating a direct role of the protein pocket in dictating base specificity in DNA modifier proteins. Whereas, in Ecl18kI restriction enzyme, mutation of a Tyr residue that stacks with the flipped base to Ala resulted in a reduction of nuclease activity to target DNA (Tamulaitis et al., 2008). This reduction of nuclease activity has been shown to be due to a decrease in the base flipping stability (Neely et al., 2009). These studies also indicate the protein pocket to be another major factor that affects the base stability and contributes to base specificity and enzymatic efficiency.

McrBC as a system to study base flipping

McrBC is an enzyme belonging to the class of DNA readers that utilizes base flipping. McrBC is a GTP dependent prokaryotic restriction enzyme that recognizes and cleaves DNA containing two R^mC (Methylcytosines preceded by Adenine or Guanine) sites separated by 30 – 2000 base pairs apart (Stewart et al., 1998). It is composed of two protein subunits, McrB (54 kDa) and McrC (40 kDa) that oligomerizes to form an active tetradecameric complex which is capable of DNA cleavage (Nirwan et al., 2019). McrB harbors the DNA binding and GTP hydrolysis domains (Gast et al., 1997), whereas McrC contains the nuclease domain (Pieper et al., 2002). Crystal structure of the DNA binding domain of McrB in complex with target DNA (**Figure 1.1**) revealed that it utilizes base flipping to read methylcytosine in DNA. The structure shows that the protein-DNA complex exhibits all conserved features of the base flipping process. The enzyme induces DNA helix distortion involving a minor groove widening and DNA bending by 30° towards the major

groove. Furthermore, a Tyr residue intercalates into the DNA cavity and the flipped out methylcytosine was stabilized in the binding pocket by various hydrogen bonds and van der Waals interactions that helps the pocket discriminate it from other DNA bases (**Figure 1.2**). This makes this enzyme a good system for studying the base flipping process in DNA reader enzymes.

As evident from crystal structures (**Figure 1.2 A and B**), discrimination between the extra-helical $5^{m}C$ and C bases in the binding pocket is achieved by van der Waals interactions between the 5-methyl group of the cytosine and the side chain of Leu68 residue. Accordingly, it has been shown that the mutation Leu68 to Phe decreases the interaction distance with the non-methylated cytosine and results in the enzyme being able to stabilize and recognize a non-methylated cytosine (Gyana Mishra, 2017), unlike the wild type enzyme. Co-crystal structure of the mutant with DNA indicated that the distance between this Phe68 residue and the nearest carbon atom of a non-methylated cytosine was reduced to 4.6 Å (**Figure 1.2 C**), which could result in a van der Waals contact with a non-methylated cytosine (Vishal Adhav, unpublished data).

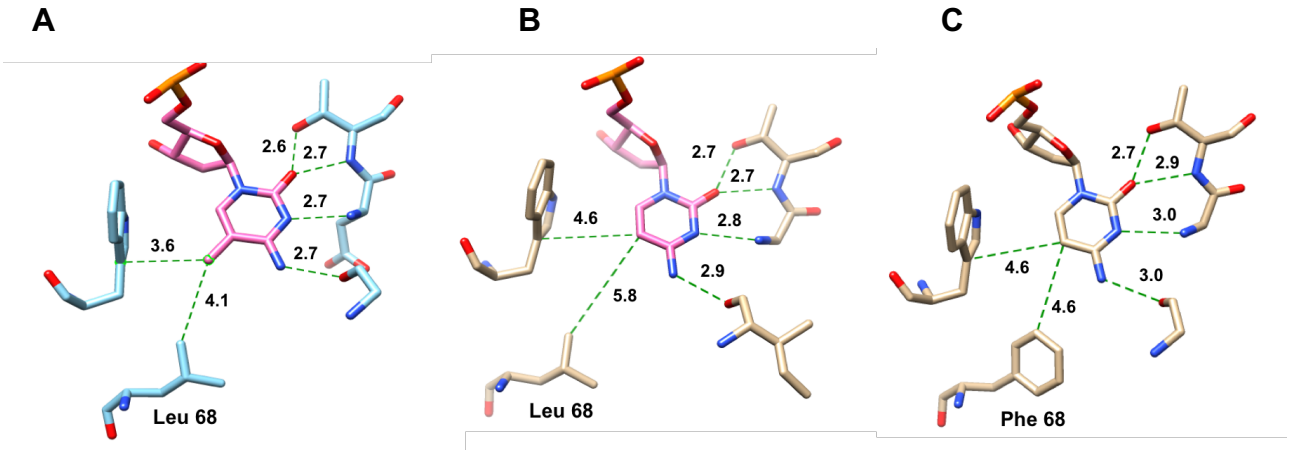


Figure 1.2 Interactions in the cytosine binding pocket for Nter-McrB (**A**) Nter-McrB with methylated cytosine. (**B**) Nter-McrB with non-methylated cytosine. (**C**) Nter-$McrB^{L68F}$ with non-methylated cytosine.

Aim and Scope of the thesis

As an increasing number of DNA reader proteins are being found to utilize DNA base flipping for recognition of their target DNA, it becomes important to ask whether this process can dictate target specificity in these enzymes. It is well established in DNA modifying enzymes such as DNA methyltransferases and DNA glycosylases that the target specificity is achieved by the protein pocket whereas, the intercalating residue mainly contributes to the catalytic efficiency of the enzyme rather than the target specificity. In contrast, the role of DNA intercalating residue in DNA reader proteins is not established. The complete loss of DNA binding observed in SUVH5 and Nter-McrB on Ala mutagenesis of this residue points to a hypothesis that this residue contributes to target DNA recognition in these systems. The main focus of this thesis work will be to study how the intercalating residue contributes to target DNA recognition in McrBC by its contribution to the base flipping mechanism. This will be studied by mutagenesis of the intercalating residue, Tyr41, to a series of amino acids based on size and hydrophobicity. The DNA binding and nuclease activity of these mutant enzymes could then be used as a primary indicator to study whether base stability and target recognition is altered because of the Tyr41 mutations. The $McrB^{L68F}$ mutant will be used for these studies as it could be used to study how the intercalating residue can affect recognition of both modified and unmodified cytosine, as this mutant recognizes both.

Chapter 2 **Materials and Methods**

Cloning of McrB Full length & N-terminus mutants

We employed site directed mutagenesis protocol to clone all mutants. For this purpose, we used a forward primer carrying the desired mutation (primer sequences given in Table 2.1) and T7 terminator primer. Using these primers, the full length gene with the respective mutation was amplified from pHIS vector containing the respective McrB full length gene (Nirwan et al., 2019). The amplified gene with the mutation was subsequently cloned into pHIS vector using restriction free (RF) PCR cloning protocol as used previously (Ent et al., 2006). The cloned vectors were amplified by transformation into NEB electrocompetent cells and the plasmid was purified from the cell lysate by DNA purification spin columns (Qiagen). The clones were confirmed by restriction digestion with Nde1 and BamH1 followed by sequencing using a T7 promoter primer.

For cloning the N-terminal constructs of McrB (Nter-McrB) that code for the 1-161 amino acid residues of McrB, a T7 promoter primer and a gene specific reverse primer (mbnrev) were used to amplify the N-terminal gene fragment from pHIS vectors containing the respective full length clones. This amplified fragment was utilized for the extension (RF) PCR and proceeded similarly as followed for the full length clones. The clones obtained were confirmed by restriction digestion with Nde1 followed by sequencing using a T7 promoter primer.

Table 2.1 List of Primers used for cloning

Primer name	**Sequence (5' – 3')**
T7 Promoter	TAATACGACTCACTATAGGG
T7 Terminator	GCTAGTTATTGCTCAGCGG
mbnrev	ATGATGATGATGATGATGACCTTCGAGAGGTGGAATAACACTC
mbY41F	GTAAAATTGAGTTTCGGTTTTGGTAATTTTACGTCTATT

Protein Purification

Plasmids of McrB full length constructs or full length McrC were transformed into *E. coli* C3013I cells (NEB) and grown in 1L cultures at 37°C until OD reached ~0.6. Cultures were induced with 0.5 mM IPTG at OD = 0.6 and transferred to 18°C for 12-16 hours and then pelleted (5000 rpm, 20 minutes, 4°C). Further, the pellets were suspended in 50 ml lysis buffer (500 mM NaCl, 50 mM Tris-Cl pH 8.0, 25 mM imidazole, 5 mM $MgCl_2$, 10% glycerol) and was then lysed by sonication as per standard protocol. After sonication, the lysate was ultra-centrifuged at 4°C at 37000 rpm for 45 minutes. The supernatant obtained was loaded onto a 5 ml Ni-NTA column equilibrated with Buffer A (500 mM NaCl, 50 mM Tris-Cl pH 8.0, 25 mM imidazole). After complete loading, protein was eluted using a step gradient (5%, 10%, 20%, 40%, 60%, 100%) of Buffer B (500 mM NaCl, 50 mM Tris-Cl pH 8.0, 500 mM imidazole). The eluted fractions were loaded on a 12% SDS PAGE gel for checking purity.

For the McrB full length constructs, the purest Ni-NTA fractions were dialyzed against 2 Liter of B50 buffer (50 mM NaCl, 50 mM Tris-Cl pH 8.0, 1 mM DTT). Dialyzed protein was then centrifuged at 18000 rpm at 4°C for 20 mins, and the supernatant was loaded onto an 8 ml MonoQ10/100 (anion exchange) column equilibrated with Buffer B50. Fractions of 1.5 ml were collected in 20 column volumes over a linear gradient of 0 to 50% of Buffer B1000 (1000 mM NaCl, 50 mM Tris-Cl pH 8.0, 1 mM DTT). The pure fractions were then pooled and concentrated using a 10 kDa concentrator. The protein solution was then washed with B100 (50 mM Tris-Cl pH 8.0, 100 mM NaCl, 1 mM DTT) and stored at -80°C.

For purifying McrC, the Ni-NTA fractions were dialysed against 2 litre of B50 and was loaded onto an 8 ml MonoS10/100 (cation exchange) column equilibrated with Buffer B50. Fractions of 1.5 ml were collected in 20 column volumes over a linear gradient of 0 to 50% of Buffer B1000. The pure fractions were then pooled and concentrated using a 10 kDa concentrator. The protein solution was then washed with B100 and stored in -80°C.

For purification of the Nter-McrB constructs, the plasmids were transformed into BL21(AI) cells and induced with 0.02% (w/v) arabinose at OD=0.6. After a similar Ni-NTA purification step as the full length constructs, the purest Ni-NTA fractions were dialyzed against 2 litre of Nter-McrB buffer (10 mM Tris–Cl pH 8.0, 0.2 M KCl, 0.1 mM EDTA and 1 mM DTT). Dialyzed protein was then centrifuged at 4°C for 20 mins, and the supernatant was loaded onto a size-exclusion column (Superdex 75 - GE Life sciences) equilibrated with Nter-McrB buffer. Fractions of 1 ml were collected in a single column volume, and the fractions that eluted at the expected volume were concentrated and stored at -80°C.

Substrates for DNA cleavage assays

1.2 kb substrate with multiple R^mC and RC sites:

A 1.2 Kb DNA fragment was amplified by using PCR with T7 Promoter and T7 Terminator as primers and pHIS vector containing MBdN gene (Nirwan et al., 2019) as template. A dNTP mix containing d(5^mC)TP instead of dCTP was used for generating the methylated version of the substrate. A dNTP mix containing dCTP was used for generating the non-methylated version of the substrate. The PCR amplification reactions were carried out in a 50 µL solution containing 0.4 µM of each primer, 100 ng of pHIS template and 0.1 mM dNTP's. A PCR cycle consisting of annealing at 60°C and 35 cycles of amplification was used.

114 bp McrBC specific substrates:

The substrate having two R^mC sites (both containing 5-methylcytosine) separated by 54 bp (2X 5^mC 114 bp substrate) was generated by overlap extension PCR using forward primer MB60MSPI-1F & reverse primer MB60MSPI-2R (sequences given in Table 2.2). Both of these primers contained one R^mC site each. The 114 bp substrate having only one R^mC site (1X 5^mC 114 bp substrate) was generated by a similar PCR with forward primer MB60MSPI-1F and reverse primer MB60MSPINM-2R which had RC at the place of R^mC. The PCR extension reactions were carried out in a 50 µL solution containing 0.6 µM of each primer and 0.1 mM dNTP's. A PCR cycle consisting of annealing at 65.1°C and 35 cycles of extension was used.

Table 2.2 List of primers used for generating the 114 bp substrate

Primer name	Sequence (5' – 3')
MB60MSPI-1F	GCCGGGTAACCGGGTAAGTCCGGGTAAGA^{m}CCGGTAGTTCGGATCGAGGGGTAG GCCGC
MB60MSPI-2R	AGTCAAATTGCATATGCTGGTCTTTCAGCG^{m}CCGGTAATCGTCTTGTGAAGGATCC GCGGC
MB60MSPINM-2R	AGTCAAATTGCATATGCTGGTCTTTCAGCGCCGGTAATCGTCTTGTGAAGGATCCG CGGC

DNA cleavage assays & their quantification

The assays were carried out in a 10 µL reaction mixture in 1x NEB cutsmart buffer (50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 100 µg/ml BSA, pH 7.9). 75 nM of the 114 bp substrate was incubated with 75 nM of the McrBC complex or 75 ng of the 1.2 kb substrate was incubated with 50 nM of the McrBC complex, in the presence of 1 mM GTP (Jena Bioscience) at 37°C for 1 hour. The reaction was stopped by adding 2 µL of 6X STES buffer (40% (w/v) sucrose, 0.2 M Tris–Cl pH 7.5, 40 mM EDTA, 1% (w/v) SDS) followed by denaturation by heating at 65°C for 15 minutes. The DNA cleavage reactions with 114 bp substrates were loaded on a pre-electrophoresed 10% native PAGE gel. The gels were run at 150 V in 1X TBE buffer (0.089 M Tris-Cl, 0.089 M Boric Acid, and 2 mM EDTA pH 8.0) and stained with a solution containing 2 µg/ml ethidium bromide for 5 min. The DNA cleavage reactions with the 1.2 kb substrate was loaded on a 0.8% agarose gel and was run at 110 V in 1X TAE buffer (40 mM Tris, 20 mM Acetate and 1 mM EDTA pH 8.0). The gels were imaged on an E-Gel imager (Thermo Fisher Scientific).

Quantification of nuclease activity was done using the ImageJ program, where the fraction of uncut DNA substrate (1X 5^{m}C 114 bp substrate) left in each reaction was used to calculate the nuclease efficiency (denotes percentage of DNA substrate cleaved) of the McrB mutants. Four separate and independent reactions for each mutant were used for calculation of their respective nuclease efficiencies. The error estimate represents standard error for the four replicates of each mutant.

Electrophoretic mobility shift assays (EMSA)

A 60 bp specific substrate (with single R^mC site) was used for the binding assay. For generating the substrate, two complementary 60 bp oligo's, MB60MSPINM-2F and MB60MSP1-2R (sequence given in Table 2.3) was mixed in equimolar ratios (1.2 µM) and heated at 95°C for 5 minutes. The oligo's were then allowed to anneal by cooling the mix to 4°C at the rate of 1°C/min. The purity of the annealed DNA was checked by loading it on a 10% native PAGE gel. The substrate was directly used for the DNA binding assays. The binding reactions were carried out in a binding buffer (50 mM KCl, 10 mM Tris-Cl pH 8.0, 15 mM $MgCl_2$, 1mM DTT, 10% glycerol). 250 nM of the DNA was incubated with varying concentration of McrB in a 10 µL reaction mix containing 1 mM GTP (Jena Bioscience) and 0.4 mg/mL BSA. The reaction was incubated at room temperature for 10 minutes. 2 µl of ST buffer (40 % Sucrose, 0.2 M Tris-Cl pH 7.5) was added to the reaction mix and was loaded on to a pre-electrophoresed 5% native PAGE gel. The gel was run at 150 V in 1X TBE buffer at 4°C. The gels were stained with a solution containing 2 µg/ml ethidium bromide for 5 minutes and imaged on an E-Gel imager (Thermo Fisher Scientific).

Table 2.3 List of primers used for generating the 60 bp Substrate

Primer name	Sequence (5' – 3')
MB60MSPINM-2F	GCCGCGGATCCTTCACAAGACGATTACCGGCGCTGAA AGACCAGCATATGCAATTTGACT
MB60MSPI-2R	AGTCAAATTGCATATGCTGGTCTTTCAGCG^{m}CCGGTAAT CGTCTTGTGAAGGATCCGCGGC

Fluorescence anisotropy binding experiments

Fluorescence anisotropy experiments were carried out in a HORIBA FluoroMax-4 spectrophotometer. A 13 bp non-methylated DNA with RC sites on both strands was generated by annealing two oligo's, mbflur1 and mbflur2 (sequences given in table 2.4) to generate an oligoduplex with one nucleotide 5' extrusions (**Figure 2.1**). The oligo mbflur1 was labeled at the 5' end with a 5-IAF tag prior to the annealing reaction (Vishal Adhav, unpublished data). The N terminal constructs of McrB

mutants (Nter-McrB) were used for the anisotropy experiments. 50 nM of this DNA was used in each anisotropy experiment and readings were acquired with an excitation wavelength 492 nm and emission wavelength 515 nm. Anisotropy readings were taken in a reaction volume of 150 µL at 25°C in the anisotropy buffer (20 mM Mes-KOH pH 6.0, 100 mM KCl, 1mM DTT). Protein concentration was increased from 62.5 nM to 15 µM depending on the saturation kinetics of the binding. Protein concentrations were increased by serial addition of ~0.5 µL of the protein stock (concentration estimated in Guanidine Hydrochloride with an extension coefficient of Abs 0.1 % = 1.902) to the reaction such that the increase in reaction volume is negligible. After each addition, there was 2 minute delay before taking readings so that the protein-DNA binding reaches equilibrium. Anisotropy value at a particular protein concentration was taken as an average of 10 readings, with each reading taken with an integration time of 1 second.

The binding curve was plotted using the software GraphPad Prism 8.0 by using the equation $Y = Bmax*X/(K_D + X)$, where Y is the anisotropy value at each protein concentration after subtracting from it, the anisotropy of free double stranded DNA. X is protein concentration and Bmax is the estimated saturation value of anisotropy increase during each experiment. The curve was plotted by averaging data points from three independent experiments. The error for the K_D value was estimated by the standard error of independent K_D values of the triplicates for each mutant. Throughout the calculations, a 1:1 ratio of protein-DNA binding is assumed.

Table 2.4 List of primers used for generating the 13 bp non-methylated DNA

Primer name	**Sequence (5' – 3')**
mbflur1	TGAGACCGGTAGC
mbflur2	AGCTACCGGTCTC

5' T G A G A C C G G T A G C 3'
3' C T C T G G C C A T C G A 5'

Figure 2.1 Sequence of the 13 bp DNA used for fluorescence anisotropy and crystallization (potential RC binding sites for McrB is highlighted in red).

Crystallization of Nter-McrB constructs with DNA

A 13 bp non-methylated DNA containing RC sites, with the same sequence that was used for the anisotropy experiments (**Figure 2.1**) was used for crystallization trials. Protein-DNA complex was formed by mixing Nter-McrB and DNA at a final concentration of 187 μM and 106 μM, respectively, in a 20 μL solution in the Nter-McrB buffer (See purification). Crystallization conditions contained Bis-Tris pH 5.5 (0.1 M – 0.2 M) and PEG 4000 (12% - 22%). Crystallization was set up with the hanging drop vapour diffusion method in 24 well plates. The reservoir contained 400 μL of the respective condition and a 2 μL drop was made by mixing 1μL each of the centrifuged protein-DNA complex and the respective reservoir conditions. The drop was left at 18°C for crystal formation.

Chapter 3 **Results**

Cloning and purification of McrB mutants

As discussed previously, base flipping enzymes employ a variety of amino acid residues for DNA intercalation, ranging from Gln in Hha1 to Tyr in McrBC. To study the role of this residue in the base flipping mechanism and target DNA recognition in McrBC system, a series of McrB mutants replacing Tyr 41 to amino acids of varying size and hydrophobicity were chosen. Tyr 41 residue was replaced by Ala, Gln, Leu, Met, Phe, and Trp. These mutants were cloned in pHIS vector prior to this work. The N-Terminal (DNA binding domain) constructs of the McrB full length mutants were cloned in the same vector as mentioned in the methods section. The gene amplification PCR produced an amplified gene of expected size 600 bp (**Figure 3.1 A**). The cloned plasmids after digestion with Nde1 produced a linearized plasmid of expected size 3.2 kb (600 bp gene + 2600 bp vector) (**Figure 3.1 B**). All these clones were confirmed positive by sequencing the full gene.

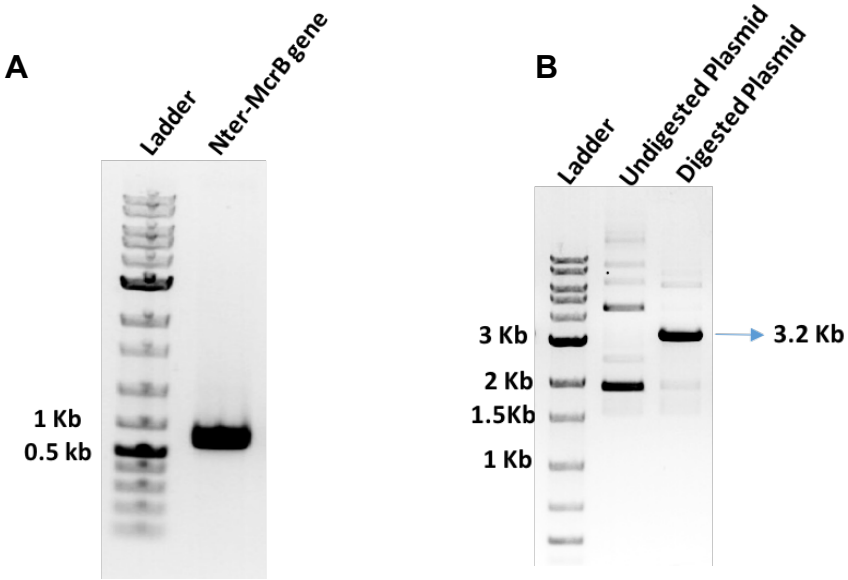


Figure 3.1 Representative gels for cloning of Nter-McrB constructs (**A**) Products of the gene amplification PCR loaded on an agarose gel. (**B**) The cloned vector digested with Nde1 enzyme loaded on an agarose gel.

The McrB and McrC full length constructs were purified by series of affinity and ion exchange chromatography columns as mention in method section (**Figure 3.2**). The McrB N terminal constructs (Nter-McrB) was purified by a series of affinity and size exclusion chromatography columns (**Figure 3.3**). **Figure 3.2 and Figure 3.3 A** shows the representative SDS denaturing PAGE gels for the purification profiles of these proteins. The size exclusion profile of N-terminal McrB constructs (referred to as Nter-McrB here onwards) showed that it eluted in the monomeric form at an elution volume (12-13 ml) corresponding to its molecular mass 19 kDa (**Figure 3.3 B**). All these purifications yielded a large amount of protein, which were used for biochemical studies and crystallization trials.

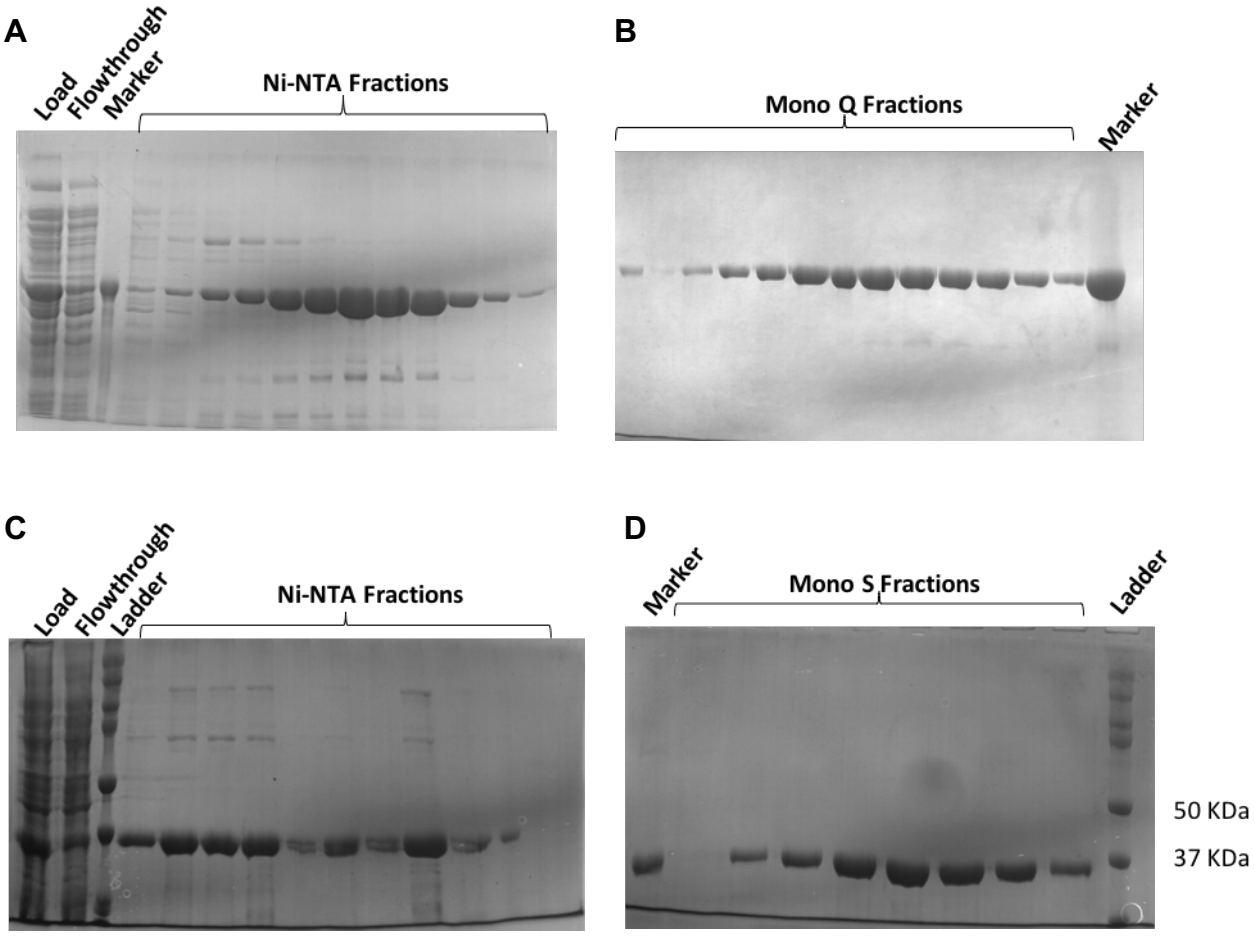


Figure 3.2 Representative protein purification profiles of full length protein constructs in 12% SDS PAGE gel (**A**) Ni-NTA elution gel and (**B**) Mono Q elution gel of McrB full length mutant constructs. (**C**) Ni-NTA elution gel and (**D**) Mono S elution gel for McrC full length construct.

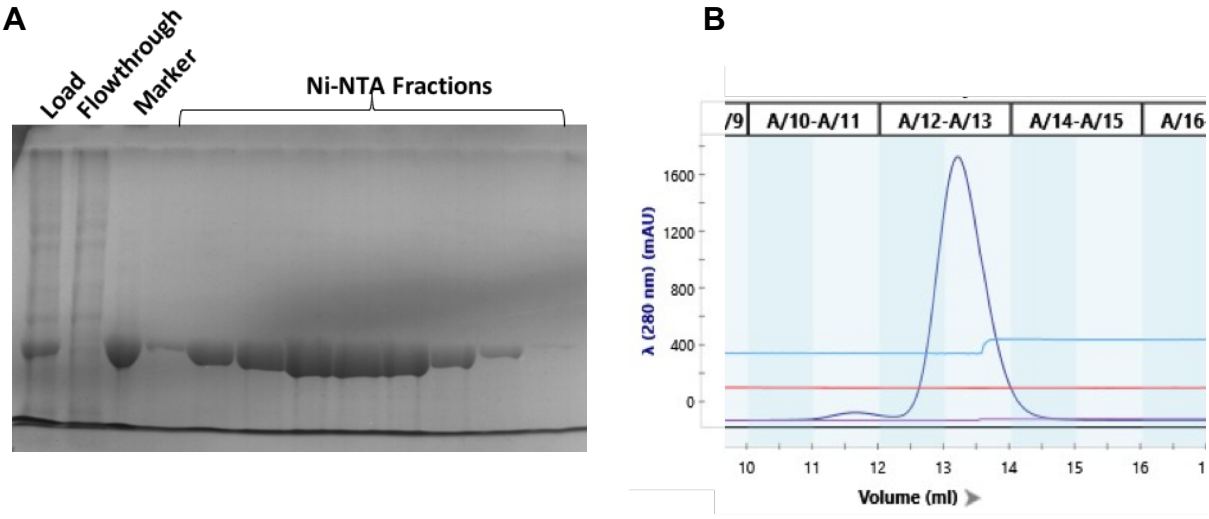


Figure 3.3 Representative protein purification profiles of Nter-McrB constructs (**A**) Ni-NTA elution gel and (**B**) Superdex 75 elution chromatogram of Nter-McrB constructs.

Correlation between DNA cleavage and flipped base stability in McrB^{L68F}

The DNA cleavage activity of McrB^{L68F} was checked using the 1.2 kb substrate (**Figure 3.4 A**). As established previously (Gyana Mishra, 2017), the mutant was able to cleave methylated and non-methylated DNA i.e. containing R^{m}C and RC site. Whereas, it is known that wild type protein only cleaves DNA having R^{m}C sites (Gast et al., 1997, **Figure 3.4 A**). In case of McrB^{L68F} mutant it was proposed that an increased stability of the flipped out base (non-methylated cytosine) in the protein pocket will result in an increased DNA binding affinity for non-methylated DNA, which in turn manifest as cleavage of the DNA. Hence, we wanted to verify that the newly gained activity for RC sites for McrB^{L68F} mutant was due to an increased stability of non-methylated cytosine in the recognition pocket. Towards this, a truncated construct of McrB (Nter-McrB) containing the DNA binding domain of the enzyme was chosen for binding studies with a 13 bp non-methylated DNA with RC sites using fluorescence anisotropy experiments. The binding studies clearly indicated that the Nter-McrB^{L68F} mutant had ~7 fold higher binding affinity (**Figure 3.4 B and C**) when compared to wild type enzyme Nter-McrB. This confirmed that the mutant was

able to stabilize the base (non-methylated cytosine) in the pocket and provides direct evidence that DNA cleavage was a result of cytosine recognition by McrB^{L68F} enzyme.

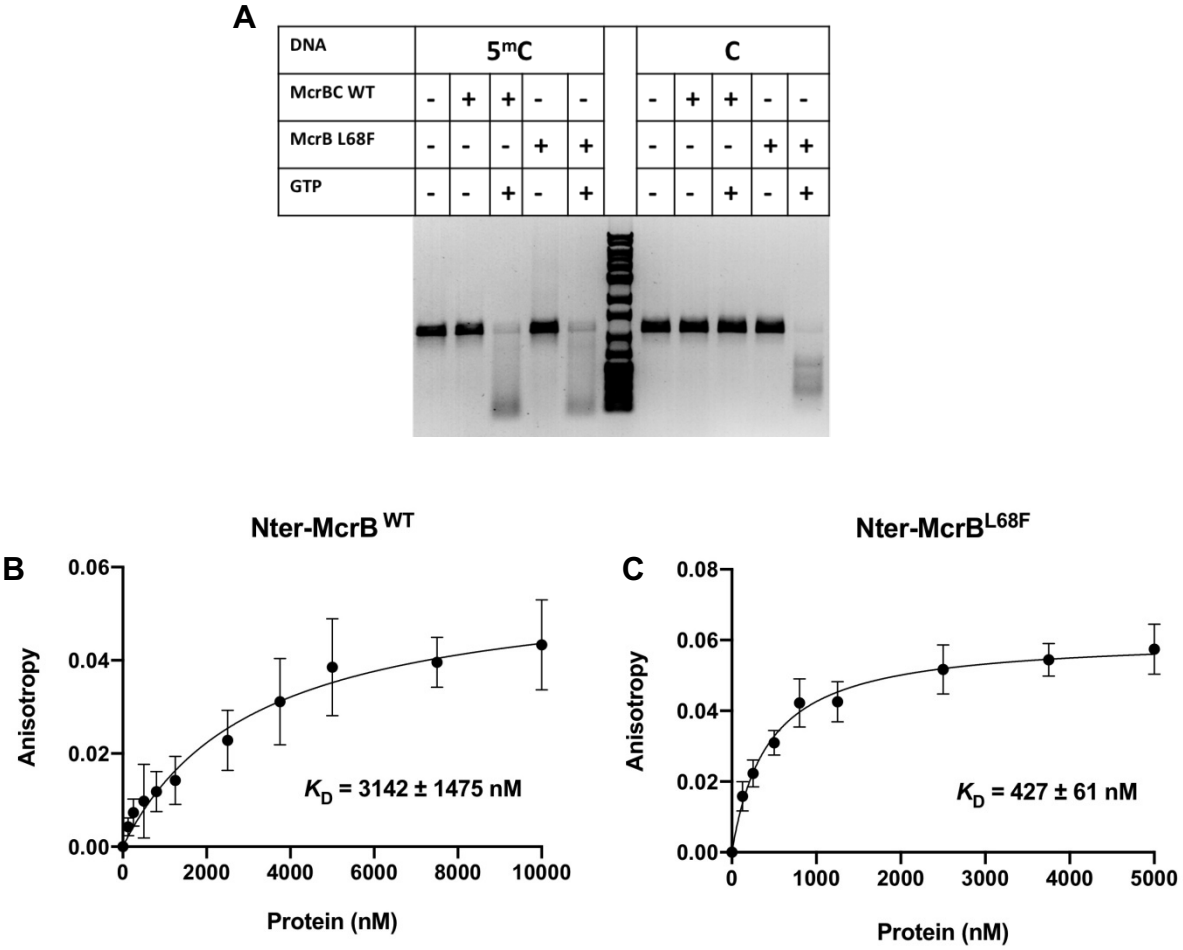


Figure 3.4. Recognition of non-methylated DNA by McrB^{L68F} (**A**) DNA cleavage assay of McrBC^{L68F} with 1.2 kb substrate. (**B**) Anisotropy binding curve of Nter-McrB WT. (**C**) Anisotropy binding curve of Nter-McrB^{L68F}.

Intercalating residue has prominent role in recognition of R^{m}C sites

As discussed previously, McrB^{L68F} had specificity for both methylated and non-methylated DNA (R^{m}C and RC sites). Next, we wanted to understand the role of intercalating residue in target DNA recognition and cleavage activity of this mutant towards substrates with R^{m}C sites. For this purpose, we selected a 114 bp DNA substrate optimized previously for wild type McrBC enzyme (Nirwan et al., 2019) that contained two R^{m}C sites (referred to as 2X 5^{m}C substrate). We selected a series of

double mutants of wild type McrB enzyme already cloned in the pHIS vector that had the intercalating residue (Tyr41) mutated to amino acids with varying size and hydrophobicity ie 1. McrB^{L68F_Y41A}, 2. McrB^{L68F_Y41Q}, 3. McrB^{L68F_Y41L}, 4. McrB^{L68F_Y41M}, 5. McrB^{L68F_Y41F} and 6. McrB^{L68F_Y41W} (Nevin Zacharia, Vishal Adhav, Pratima Singh, unpublished data).

For McrB^{L68F_Y41A} mutant, there is only a methyl group side chain that fills the cavity generated in the DNA upon base flipping. DNA cleavage assay with this mutant shows that this mutation leads to the enzyme being nuclease inactive towards target DNA (**Figure 3.5**). This suggested that the mutation of the intercalating residue has possibly affected base flipping.

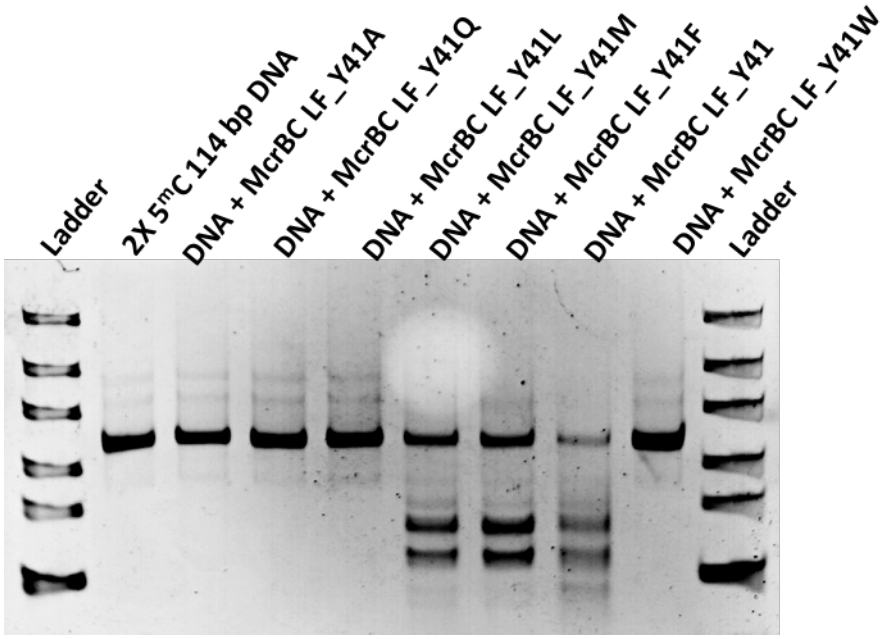


Figure 3.5 DNA cleavage assay with the various McrBC^{L68F_Y41} mutants with the 2X 5^{m}C 114 bp substrate.

When the intercalating residue is a Gln, the enzyme is nuclease inactive, similar to the Ala mutant (**Figure 3.5**). As the hydrophobicity of the intercalating residue increased from Gln, a polar residue to Met, a nonpolar residue, it is observed that the enzyme gains activity against this 2X 5^{m}C 114 bp substrate (**Figure 3.5**). This observation holds true even if we replace a hydrophobic and longer Phe at that position (**Figure 3.5**), suggesting that an intercalating Met and Phe can stabilize a

flipped $5^{m}C$ base. When the intercalating residue is Leu, which is more hydrophobic than a methionine (Wimley et al., 1996), but shorter in length than a methionine, the enzyme is nuclease inactive and was not able to provide stability to the flipped $5^{m}C$ for DNA cleavage. The inability of a more hydrophobic and longer Trp mutant to show activity could be because of the steric clashes that prevent the large side chain from intercalating into the DNA helix (**Figure 3.5**). This reasoning is supported by a previous study in Hha1 methyltransferase, where mutating the intercalating residue to Trp prevented the enzyme from crystallizing with DNA (Dong et al., 2004).

To verify whether the observed nuclease activities are indeed caused due to a reduction in the ability to stabilize a $5^{m}C$ base, DNA binding assays were performed with $McrB^{L68F\_Y41L}$ and $McrB^{L68F\_Y41M}$ with a 60 bp DNA containing one $R^{m}C$ site. Results from this assay showed that Leu mutant does not show any binding to the DNA, even at 30 μM protein concentration (**Figure 3.6 A**). Whereas, the Met mutant was starting to show a binding at 3 μM of protein and shows a higher molecular weight protein-DNA complex at 30 μM (**Figure 3.6 B**). This provides further evidence that the observed enzymatic activities are due to the mutations of intercalating residue affecting the base flipping mechanism.

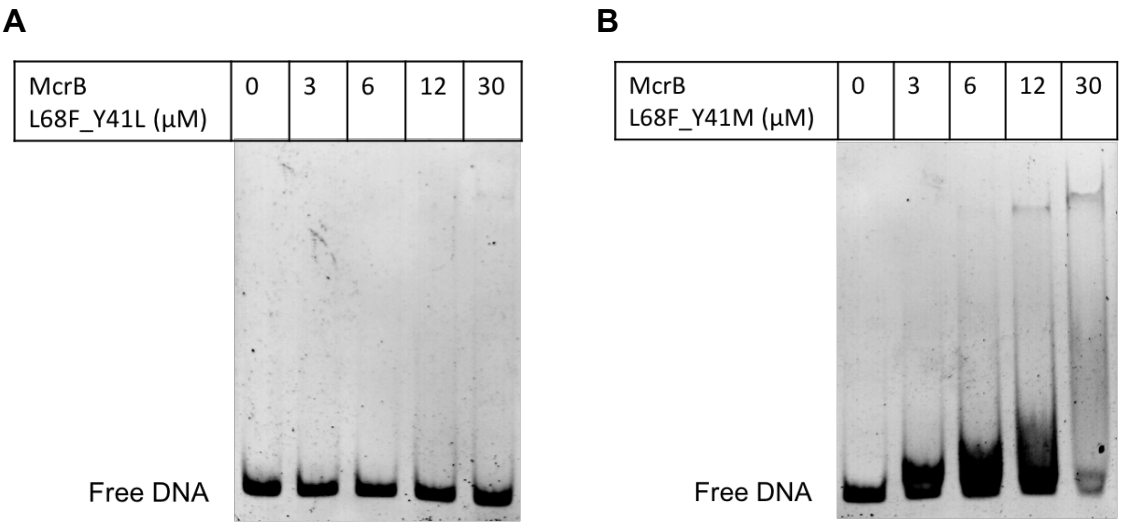


Figure 3.6 DNA binding assay with 60 bp DNA containing one $R^{m}C$ site for (**A**) $McrB^{L68F\_Y41L}$ (**B**) $McrB^{L68F\_Y41M}$.

Intercalating residue also has prominent role in recognition of RC sites

As noted previously, McrB^{L68F} has an affinity for non-methylated cytosine (RC sites). To check the activity of the Tyr41 mutants towards RC sites, we used a 114 bp McrBC specific substrate having only one R^{m}C site (1X 5^{m}C substrate). This substrate was chosen for studying RC site recognition, because McrBC requires at least two target sites for DNA cleavage and it will cleave the substrate with single R^{m}C site only if it can recognize atleast one other RC site in the substrate. As expected, we observed that McrB^{L68F} cleaved this 114 bp substrate (**Figure 3.7**). This further indicates that McrB^{L68F} is able to stabilize a non-methylated cytosine (RC sites). However, the cleavage pattern and cleavage efficiency (quantification not shown) was slightly different from that observed with the 2X 5^{m}C 114 bp substrate (**Figure 3.5** and **Figure 3.7**). This is possibly because of the presence of multiple RC sites that could be recognized by the enzyme in this 114 bp substrate and because of a weaker binding affinity of McrB^{L68F} to RC sites when compared to R^{m}C sites. Accordingly, we used this 1X 5^{m}C 114 bp substrate to check the ability of Tyr41 mutants of McrB^{L68F} for their nuclease activity. As observed from the DNA cleavage assay (**Figure 3.7**), none of the other amino acids at that position except Tyr was able to show activity towards this substrate, indicating that an intercalating Ala, Gln, Leu and Met did not provide the required stability to a flipped non-methylated cytosine. Surprisingly, when the intercalating Tyr41 residue in McrB^{L68F} was mutated to Phe, we found loss of activity towards the 1X 5^{m}C 114 bp DNA (**Figure 3.7**), although the only difference between Tyr and Phe is the presence of –OH group at para positon in Tyr. Also, from crystal structure of Nter-McrB it was seen that this –OH group is exposed to solvent and does not make any direct interaction with DNA (**Figure 1.1**).

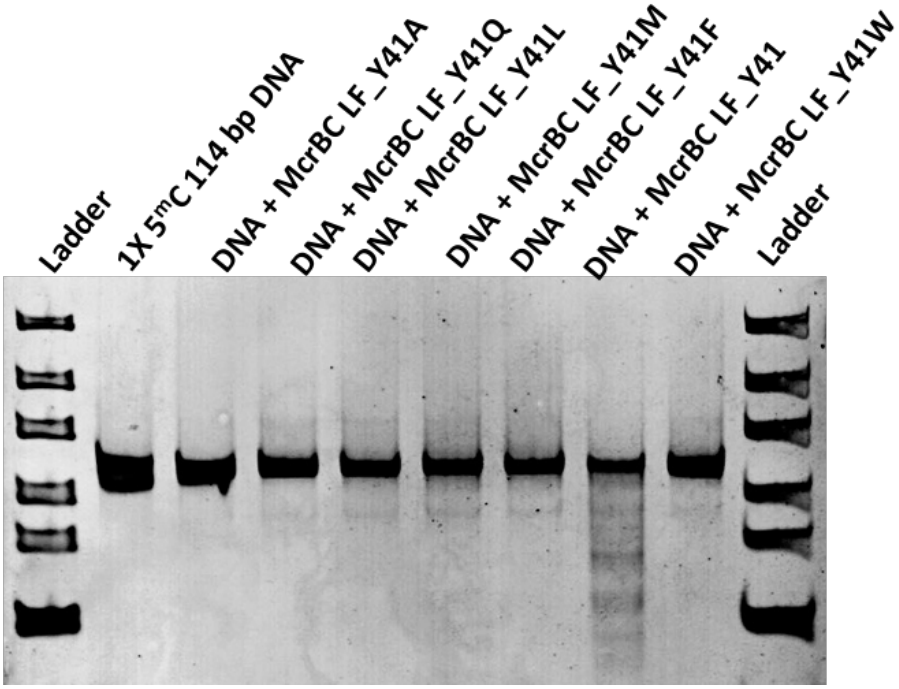


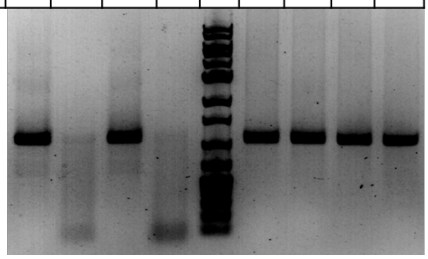
Figure 3.7 DNA cleavage assay with the various McrBC^{L68F_Y41} mutants with the 1X 5^{m}C 114 bp DNA.

McrB^{L68F_Y41F} lost its activity even for RC DNA with large number of RC sites and at high enzyme concentration

As we noted that McrB^{L68F} and McrB^{L68F_Y41F} mutants showed starkly different activity towards the RC sites (1X 5^{m}C 114 bp DNA), although both residues are very similar, we asked if McrB^{L68F_Y41F} can show activity for non-methylated DNA in presence of multiple RC sites. For this purpose, we used the 1.2 kb DNA substrate containing multiple RC sites. Interestingly, we do not find any activity for the 1.2 kb non-methylated substrate by McrB^{L68F_Y41F} enzyme (**Figure 3.8 A**). However, McrB^{L68F_Y41F} enzyme was found to be nuclease active as it cleaved the 1.2 kb DNA containing R^{m}C sites (**Figure 3.8 A**). This eliminated the possibility that the Y41F mutation has rendered the protein misfolded or inactive.

A

DNA	5^mC				C			
McrBC WT	-	+	-	-	-	+	-	-
McrBC L68F_Y41F	-	-	+	+	-	-	+	+
GTP	-	+	-	+	-	+	-	+



B

DNA	2X 5^mC 114 bp		1X 5^mC 114 bp				
McrBC L68F_Y41F	-	+	-	+	+	+	-
McrBC L68F	-	-	-	-	-	-	+
Protein (nM)	0	75	0	75	150	300	75

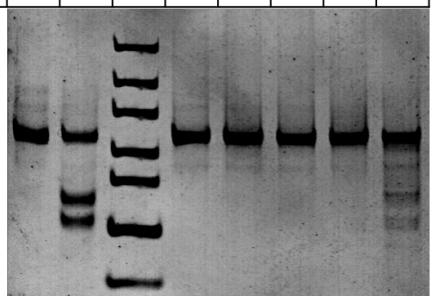


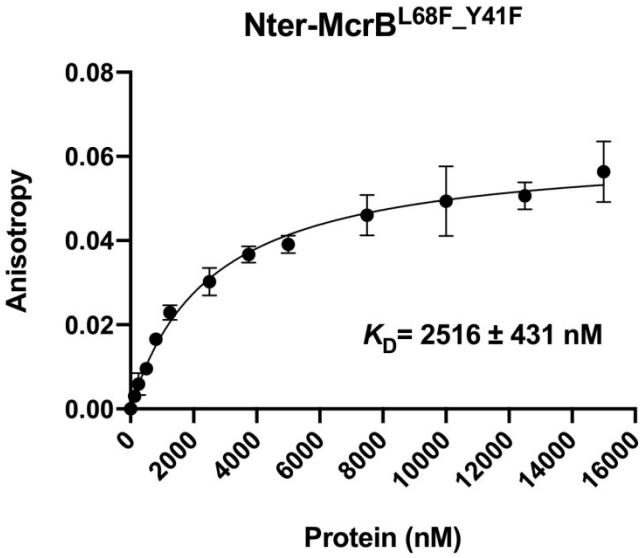
Figure 3.8 McrBC^{L68F_Y41F} does not recognize non-methylated cytosine (**A**) DNA cleavage assay of McrBC^{L68F_Y41F} with the 1.2 kb substrate. (**B**) Concentration dependent DNA cleavage assay of McrBC^{L68F_Y41F} with the 1X 5^mC 114 bp DNA.

We further asked if McrBC^{L68F_Y41F} can cleave the 1X 5^mC 114 bp substrate in a concentration dependent manner. However, even at elevated concentrations of the McrBC^{L68F_Y41F} complex, the enzyme was not recognizing RC sites (**Figure 3.8 B**). This confirms that mutation of Try41 residue to Phe in McrB^{L68F} results in loss of specificity for non-methylated substrate (RC sites).

Binding study of Nter-McrB^{L68F_Y41F}

To verify that the Y41F mutation is indeed resulting in a stability loss for non-methylated cytosine, binding study using fluorescence anisotropy was performed on Nter-McrB^{L68F_Y41F} with the 13 bp non-methylated DNA having RC sites. The assay indicated that this mutant had ~6 fold reduction in binding affinity towards the DNA as compared to Nter-McrB^{L68F} (**Figure 3.9 A**) and a comparable binding affinity to the Wild Type protein (**Figure 3.4 B**) that does not act on non-methylated DNA.

A



B

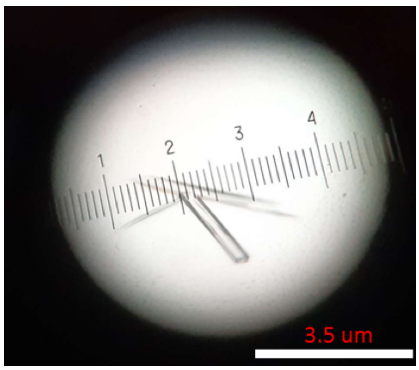


Figure 3.9 Binding studies and Crystallization of Nter-McrB^{L68F_Y41F} (**A**) Anisotropy binding curve for Nter-McrB^{L68F_Y41F} with the 13 bp non-methylated DNA. (**B**) Crystals obtained for Nter-McrB^{L68F_Y41F} in complex with the 13 bp DNA.

Structural basis to understand the lost specificity for RC site by McrB^{L68F_Y41F}

Through binding studies, it was shown that the inability of McrB^{L68F_Y41F} to cleave DNA was a result of its loss of stability to RC sites, but still does not answer as to why a Phe does not give stability to the flipped cytosine, whereas a Tyr can stabilize. A definitive way of probing whether it is due to differences in intercalation and stacking with the DNA bases is by analyzing the structure of the protein-DNA complex. As Nter-McrB was showed to co-crystallize with nonspecific DNA (Sukackaite et al., 2012), we expected Nter-McrB^{L68F_Y41F} also to crystallize with the 13 bp DNA with RC sites, to which it binds poorly. Crystallization trials were set up in similar conditions as reported for the WT complex (Sukackaite et al., 2012). Basic optimization of the crystallization conditions by removing salt content from the crystallization conditions allowed us to get larger crystals for the DNA bound complex (**Figure 3.9 B**). The crystals were diffracted at DLS (Diamond Light Source) UK and the diffraction pattern showed a resolution of 1.4 Å. Structure determination is in progress (Vishal Adhav, unpublished data).

Complementarity of the protein pocket and the DNA cavity

Previous studies on the base flipping process in various enzymes pointed to the DNA cavity and the protein pocket to be two independent parts of the base flipping process that contribute to stability. We wanted to ask if the loss of RC site recognition due to a destabilizing mutation in the DNA cavity (Y41F) can be restored by increasing the stability to cytosine in the protein pocket.

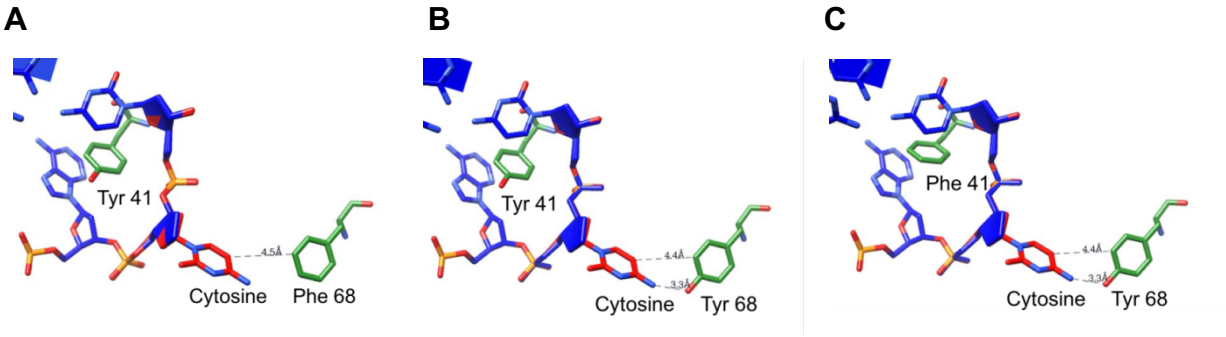


Figure 3.10 Structure modelling of the McrB mutants in UCSF Chimera (**A**) McrB^{L68F} (**B**) McrB^{L68Y} (**C**) McrB^{L68Y_Y41F}.

For this purpose we used McrB^{L68Y} mutant which was cloned previously (Vishal Adhav, unpublished data). From the structure modeled using UCSF Chimera (**Figure 3.10 B**), the Tyr68 residue in the protein pocket was predicted to make an additional hydrogen bond with the cytosine, thereby potentially providing an increased stability for cytosine in the protein pocket when compared to the Phe68 residue in McrB^{L68F} (**Figure 3.10 A**). Cleavage assays of McrBC^{L68Y} with 1.2 kb substrate showed that this mutant had activity for non-methylated DNA (**Figure 3.11 A**). Anisotropy experiments with the 13 bp non-methylated DNA showed that this mutant had a 2.2 fold higher binding affinity for DNA than Nter-McrB^{L68F} and a 16 fold higher affinity compared to the wild type enzyme (**Figure 3.11 B**). To check whether this increased DNA binding affinity results in an increased nuclease activity, DNA cleavage assays with the specific substrate (1X 5^{m}C 114 bp) was performed (**Figure 3.12 A**). Quantification of nuclease activity towards the 1X 5^{m}C 114 bp substrate indicated that McrB^{L68Y} had a 2.3 fold higher nuclease activity than

McrB^{L68F}, whereas the DNA cleavage pattern was similar for McrB^{L68F} and McrB^{L68Y} (**Figure 3.12 A and B)**. These observations confirm the expectation that the additional hydrogen bond in the protein pocket has increased the stability of cytosine in the flipped state and has not changed target specificity of the enzyme.

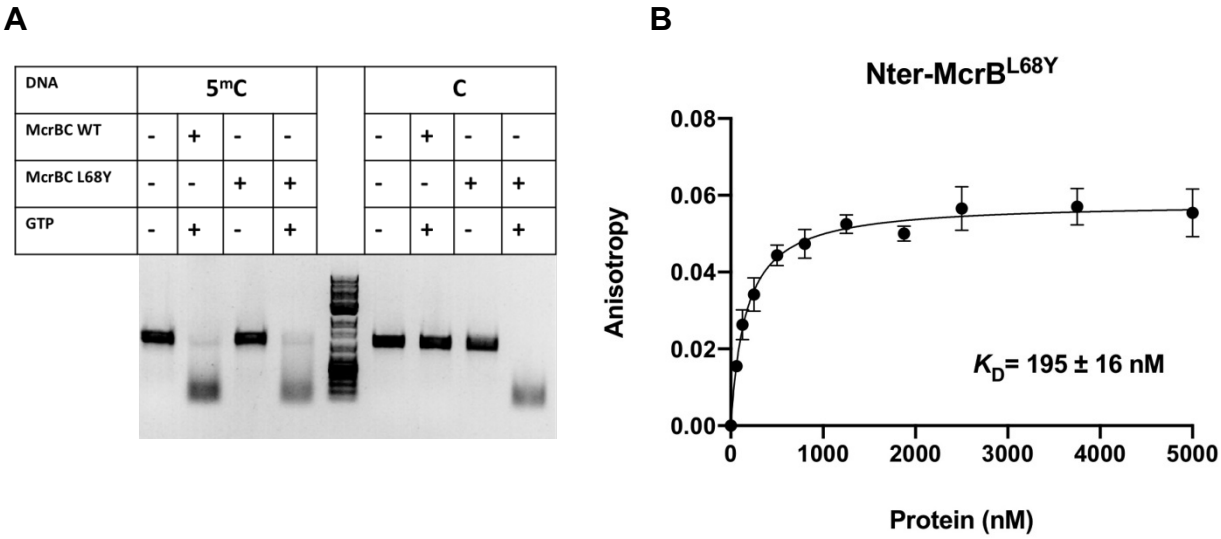
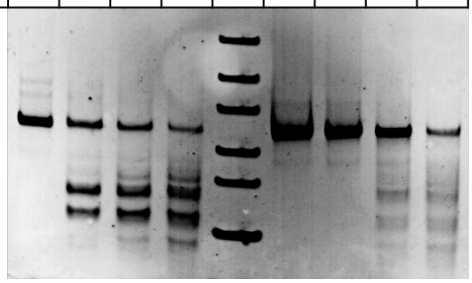


Figure 3.11 McrB^{L68Y} recognizes and binds DNA containing RC sites with a higher affinity than McrB^{L68F} (**A**) DNA cleavage assay of McrBC^{L68Y} with 1.2 kb substrate. (**B**) Anisotropy binding curve of Nter-McrB^{L68Y} with 13 bp non-methylated DNA.

A

DNA	2X 5mC 114 bp					1X 5mC 114 bp			
McrBC WT	-	+	-	-		-	+	-	-
McrBC L68F	-	-	+	-		-	-	+	-
McrBC L68Y	-	-	-	+		-	-	-	+



B

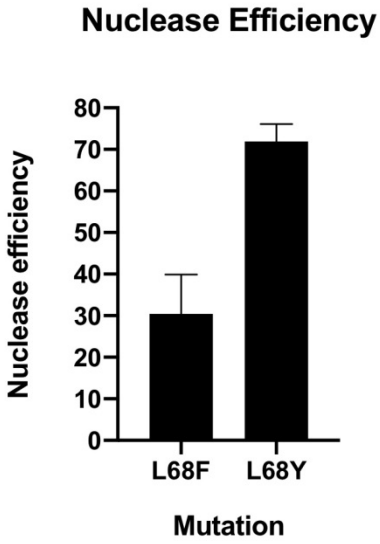


Figure 3.12 McrBC^{L68Y} cleaves DNA containing RC sites with higher efficiency than McrBC^{L68F} (**A**) McrBC^{L68Y} DNA cleavage assay with 114 bp substrates. (**B**) Quantification of DNA cleavage activity of McrBC^{L68Y} with the 1X 5^{m}C 114 bp substrate.

To test whether the McrB^{L68Y} mutant is able to retain recognition towards non-methylated cytosine when the intercalating residue is changed to a Phe, the full length McrB^{L68Y_Y41F} mutant was cloned and purified as mentioned in the methods section. Unlike McrB^{L68F_Y41F}, cleavage assays with 1.2 kb DNA clearly shows that McrB^{L68Y_Y41F} retains its activity of cleaving non-methylated DNA (**Figure 3.13 A**). Binding study with the DNA binding domain of this construct confirms that the observed RC site recognition is a result of a ~6 fold higher DNA binding affinity compared to McrB^{WT} and a similar affinity when compared to McrB^{L68F} (**Figure 3.13 B**). Whereas, the binding affinity is 2.8 fold reduced when compared to McrB^{L68Y}. Cleavage assays with the 1X 5^{m}C 114 bp substrate also showed that this mutant is able to recognize a non-methylated cytosine and the cleavage pattern generated was similar to that of McrB^{L68F} and McrB^{L68Y}, indicating that target specificity is not affected (**Figure 3.13 C and Figure 3.12 A**). Quantification of the nuclease activity shows that its nuclease activity has reduced when compared to that of McrB^{L68Y} but is similar to that of McrB^{L68F} (**Figure 3.13 D and Figure 3.12 B**), which is in good accordance to its DNA binding affinity.

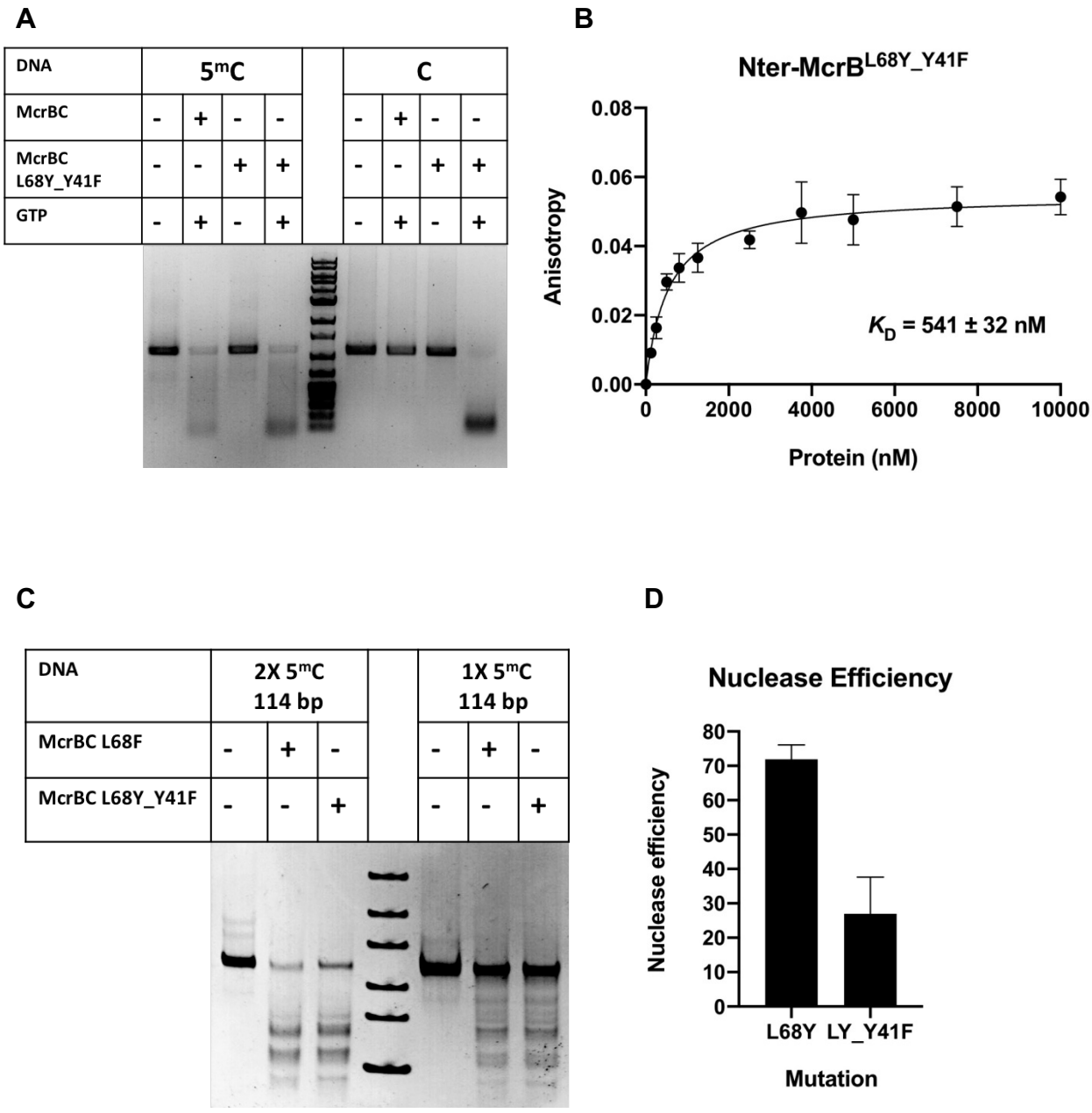


Figure 3.13 McrB^{L68Y_Y41F} retains affinity to recognize non-methylated DNA (**A**) DNA cleavage assay of McrBC^{L68Y_Y41F} with 1.2 kb substrate. (**B**) Anisotropy binding curve of Nter-McrB^{L68Y_Y41F} with 13 bp non-methylated DNA. (**C**) McrBC^{L68Y_Y41F} DNA cleavage assay with 114 bp substrates. (**D**) Quantification of McrBC^{L68Y_Y41F} nuclease activity towards 1X 5^{m}C 114 bp substrate.

Structural studies on McrB^{L68Y_Y41F} and McrB^{L68Y}

As the observed cytosine stabilization in the two mutants (McrB^{L68Y_Y41F} and McrB^{L68Y}) has been proposed to be contributed by a hydrogen bond formed between the OH group of Tyr68 and the cytosine in the protein pocket, getting the structure of these mutants bound to DNA will provide conclusive evidence for this hypothesis. So crystallization was set up for these mutants (Nter-McrB^{L68Y} and Nter-McrB^{L68Y_Y41F}) with the 13 bp non-methylated DNA substrate with RC sites, and in the same conditions as that was set up for Nter-McrB^{L68F_Y41F}. Crystals were obtained for both mutants (**Figure 3.14**) and were diffracted in DLS UK. The diffraction pattern for both crystals reached around 1.4 Å. Structure determination is in progress (Vishal Adhav, unpublished data).

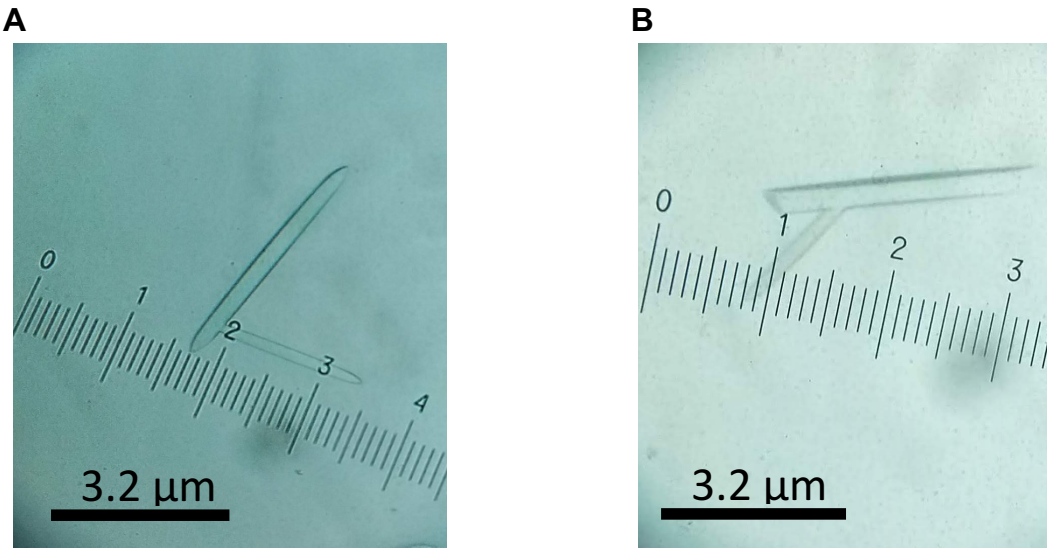


Figure 3.14 Crystallization of the Nter-McrB constructs with the L68Y mutation (**A**) Crystals obtained for Nter-McrB^{L68Y}. (**B**) Crystals obtained for Nter-McrB^{L68Y_Y41F}.

Chapter 4 **Discussion**

Intercalating residue in target base recognition

DNA base flipping is a mechanism for DNA interaction, used mainly by DNA modifying enzymes and DNA repair enzymes to access the base on which it should act. These DNA modifying and repair enzymes are collectively known as DNA writer enzymes. As discussed in the Introduction section, the McrBC restriction enzyme falls into the category of DNA reader proteins that detect specific sequences in DNA or modifications in DNA bases. Recently, co-crystal structures of various DNA readers have indicated that base flipping has been utilized by these enzymes for DNA recognition, mainly for discriminating DNA methylation.

This thesis was aimed at probing the role of the base flipping mechanism and specifically the role played by the intercalating residue in dictating the target specificity of the McrBC enzyme by studying how the DNA cavity affected the process. Towards this, a mutant enzyme McrB^{L68F} that recognized both modified and unmodified DNA was chosen so as to study the enzymatic recognition of both modified and unmodified DNA (R^{m}C and RC sites). The DNA intercalating Tyr41 residue of the enzyme was mutated to an Ala, Gln, Leu, Met, Phe and Trp. The nuclease activities of these mutants to both modified and unmodified substrates were checked. The nuclease dead Ala mutant indicates that in McrBC, a DNA intercalation residue is necessary for the base flipping mechanism, although it remains to be verified whether this intercalating residue is required for active expulsion of cytosine from DNA helix or for the stabilization of the flipped out cytosine. The observation that a polar glutamine could not stabilize a flipped 5^{m}C base for recognition, whereas a methionine which is similar in size to glutamine but is nonpolar was able to provide stability for 5^{m}C base recognition and DNA cleavage to occur, suggests hydrophobicity of the intercalating residue to be one factor that confers stability to the flipped base (5^{m}C). The finding that Leu mutant was nuclease inactive whereas Met, Phe and Tyr mutant cleaves methylated DNA pointed to the size of the intercalating residue to be another property that confers stability to the

flipped base (5^mC), as Leu is higher in hydrophobicity compared to Met (Wimley et al., 1996) but smaller in length (size) when compared to Met, Phe and Tyr. Taken together, this indicates that a higher hydrophobicity and longer residue length (size) of the intercalating residue provides increased base stability, leading to DNA base recognition. The molecular basis of why these properties (hydrophobicity and size) leads to base stability remains to be looked at. These observations propose that the intercalating residue plays a crucial role in R^mC recognition by the enzyme wherein recognition is achieved by stabilization of the flipped 5^mC mediated by the hydrophobicity and size of the intercalating residue in the DNA cavity.

In the case of non-methylated cytosine stability, interestingly, only a Tyr residue was able to stabilize a cytosine in the pocket. Ala, Gln, Leu and Met did have the required hydrophobicity or size to provide sufficient stability to a cytosine. Even Phe, which is very similar to Tyr in terms of size and hydrophobicity (Wimley et al., 1996) and is expected to provide very similar base stability, could not stabilize a non-methylated cytosine. This observation suggests that small changes in stability that can be provided by the intercalating residue in the DNA cavity could result in the enzyme recognizing a new base as its target, ie results in RC recognition. This further points to a direct role for the intercalating residue in target recognition.

Here, we propose two basis for the observed difference between the base stabilities in $McrB^{L68F}$ and $McrB^{L68F\_Y41F}$ .1. Solvent interactions mediated by the -OH group of Tyr contributes to the stability, as the –OH group is exposed to the solvent in the crystal structures 2. Stacking interaction made by Tyr residue with DNA bases is stronger than Phe, which inurn provides more stability. It is also possible that both these factors are contributing together for the observation. To verify the last scenario, the protein-DNA complex was crystallized and structure determination is in progress. In any case, a Tyr41 intercalation gives the necessary stability for RC flipping, thereby playing a direct role in RC site recognition by $McrB^{L68F}$. The $McrB^{L68F}$ and $McrB^{L68F\_Y41F}$ mutants shows a scenario in which a modulation in the base flipping due to alteration in the DNA cavity was able to convert an enzyme that recognizes both RC and R^mC sites ($McrB^{L68F}$) to an enzyme that recognizes only R^mC sites ($McrB^{L68F\_Y41F}$). This observation provides direct evidence that the intercalating residue could dictate target specificity in McrBC and that possibly this residue can be altered to modulate target specificity in DNA reader enzymes.

Complementarity of the protein pocket and DNA cavity and its implications in other base flipping enzymes

These findings prompted us to ask whether the loss of stability to RC flipping conferred by the mutation of Tyr to Phe in the DNA cavity could be rescued by compensating with an increased stability to the base from the protein pocket. A mutant that binds and cleaves DNA strongly, McrB^{L68Y} was generated, which is proposed to increase the cytosine stability through formation of an additional hydrogen bond with the cytosine in the protein pocket. To pursue our question, Tyr41 to Phe41 mutation was made in the background of McrB^{L68Y} and the resulting mutant was able to retain its activity towards non-methylated cytosine (RC sites) albeit decreased stability compared to McrB^{L68Y} as indicated by its nuclease activity and DNA binding affinity. These findings propose a complementary dependence of the protein pocket (the flipped base binding pocket) and DNA cavity (resulting from base flipping) in stabilizing the cognate base where, loss of stability in one pocket could be compensated by gain of stability in the other pocket.

Table 4.1 Properties of the four main McrB mutants used in this study

Mutant	Target recognition	Binding affinity to RC sites (nM)	Nuclease efficiency to RC site containing substrate (%)
McrB WT	R^{m}C	3142 ± 1474	No Activity
McrB L68F	R^{m}C and RC	427 ± 61	30.5 ± 4.7
McrB L68F_Y41F	R^{m}C	2516 ± 431	No Activity
McrB L68Y	R^{m}C and RC	195 ± 16	71.9 ± 2.1
McrB L68Y_Y41F	R^{m}C and RC	541 ± 32	27.0 ± 5.3

A comparison between the mutants McrB^{L68F} and McrB^{L68Y_Y41F} (**Figure 3.10 A and C**) indicates that the only difference in these mutants is the position of the -OH group that provides stability to the flipped base. In McrB^{L68F}, the -OH group was in DNA cavity, whereas in McrB^{L68Y_Y41F}, the -OH group is present in the protein pocket.

Looking at the activities of the two mutants we see they are almost identical in terms of DNA recognition, Nuclease activity and DNA binding affinity (**Table 4.1**). This emphasizes the complementary nature of the two cavities where an interchange of a stabilizing functional group between the cavities renders the protein unchanged with respect to its function. This implies that in DNA reader proteins, the DNA cavity has a very similar role as that of the protein pocket, wherein it mediates target recognition by stabilizing the cognate base.

These findings provide a possible explanation as to why the identity of the DNA intercalating residue is not conserved among Base flipping proteins, unlike other motifs like Walker motifs where the catalytic residues remain same across a wide range of non-homologous proteins (Ramakrishnan et al., 2002). For example, in UDG the intercalating residue is a Leu, while in Hha1 it is a Gln, whereas mutation of Tyr in McrBC to either of these residues compromises the enzymes ability to recognize the target sequence. This could be because the protein pocket in those enzymes (UDG and Hha1) are evolved in such a way that the reduced stability conferred by the intercalating residue is compensated by the stabilization in the protein pocket, thereby base recognition is unaffected.

Intercalating residue in enzymatic efficiency

It is well known in DNA writer enzymes that the intercalating residue contributes to its enzymatic efficiency, whereas this aspect is not looked at for DNA reader enzymes. The reduction in nuclease activity and binding affinity of $McrB^{L68Y\_Y41F}$ in comparison with $McrB^{L68Y}$ shows a scenario where a mutation of the intercalating residue results in a reduction of enzymatic efficiency in McrBC. This observation gives direct evidence that the intercalating residue plays a direct role in enzymatic efficiency in DNA reader proteins by affecting base stability. This further points to a possibility that enzymatic efficiency of DNA reader enzymes can be modulated by altering the intercalating residue.

References

Allan, B., Garcia, R., Maegley, K., Mort, J., Wong, D., Lindstrom, W., Beechem, J. and Reich, N. (1999). DNA Bending by EcoRI DNA Methyltransferase Accelerates Base Flipping but Compromises Specificity. Journal of Biological Chemistry, 274(27), pp.19269-19275.

Allan, B., Reich, N. and Beechem, J. (1999). Measurement of the Absolute Temporal Coupling between DNA Binding and Base Flipping†. Biochemistry, 38(17), pp.5308-5314.

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(7214), pp.818-821.

Bochtler, M., Szczepanowski, R., Tamulaitis, G., Grazulis, S., Czapinska, H., Manakova, E. and Siksnys, V. (2006). Nucleotide flips determine the specificity of the Ecl18kI restriction endonuclease. The EMBO Journal, 25(10), pp.2219-2229.

Daujotyt≒, D., Serva, S., Vilkaitis, G., Merkien≒, E., Venclovas, Č. and Klimašauskas, S. (2004). HhaI DNA Methyltransferase Uses the Protruding Gln237 for Active Flipping of Its Target Cytosine. Structure, 12(6), pp.1047-1055.

Dong, A., Zhou, L., Zhang, X., Stickel, S., Roberts, R. and Cheng, X. (2004). Structure of the Q237W mutant of HhaI DNA methyltransferase: an insight into protein-protein interactions. Biological Chemistry, 385(5), pp.373-379.

Ent, F. Van Den, and Lowe, J. (2006). RF cloning: A restriction-free method for inserting target genes into plasmids. J. Biochem. Biophys. Methods.

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.

Gyana Mishra, MS Thesis (2017) IISER Pune .

Horton, J., Wang, H., Mabuchi, M., Zhang, X., Roberts, R., Zheng, Y., Wilson, G. and Cheng, X. (2014). Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. Nucleic Acids Research, 42(19), pp.12092-12101.

Jakubec, D., Laskowski, R. and Vondrasek, J. (2016). Sequence-Specific Recognition of DNA by Proteins: Binding Motifs Discovered Using a Novel Statistical/Computational Analysis. PLOS ONE, 11(7), p.e0158704.

Kimber, S., Brown, T., and Fox, K. (2014). A Mutant of Uracil DNA Glycosylase That Distinguishes between Cytosine and 5-Methylcytosine. Plos ONE 9, e95394.

Klimasauskas, S., Kumar, S., Roberts, R. and Cheng, X. (1994). Hhal methyltransferase flips its target base out of the DNA helix. Cell, 76(2), pp.357-369.

Maiti, A., Morgan, M. and Drohat, A. (2009). Role of Two Strictly Conserved Residues in Nucleotide Flipping andN-Glycosylic Bond Cleavage by Human Thymine DNA Glycosylase. Journal of Biological Chemistry, 284(52), pp.36680-36688.

Neely, R., Tamulaitis, G., Chen, K., Kubala, M., Siksnys, V. and Jones, A. (2009). Time-resolved fluorescence studies of nucleotide flipping by restriction enzymes. Nucleic Acids Research, 37(20), pp.6859-6870.

Nirwan, N., Singh, P., Mishra, G.G., Johnson, C.M., Szczelkun, M.D., Inoue, K., Vinothkumar, K.R., and Saikrishnan, K. (2019). Hexameric assembly of the AAA+ protein McrB is necessary for GTPase activity. Nucleic Acids Res. 47, 868–882.

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.

Rajakumara, E., Law, J., Simanshu, D., Voigt, P., Johnson, L., Reinberg, D., Patel, D. and Jacobsen, S. (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 & Development, 25(2), pp.137-152.

Ramakrishnan, C., Dani, V., and Ramasarma, T. (2002). A conformational analysis of Walker motif A [GXXXXGKT (S)] in nucleotide-binding and other proteins. Protein Engineering, Design And Selection 15, 783-798.

Roberts, R. and Cheng, X. (1998). BASE FLIPPING. Annual Review of Biochemistry, 67(1), pp.181-198.

Rohs, R., Jin, X., West, S., Joshi, R., Honig, B. and Mann, R. (2010). Origins of Specificity in Protein-DNA Recognition. Annual Review of Biochemistry, 79(1), pp.233-269.

Slupphaug, G., Mol, C., Kavli, B., Arvai, A., Krokan, H. and Tainer, J. (1996). A nucleotide-flipping mechanism from the structure of human uracil–DNA glycosylase bound to DNA. Nature, 384(6604), pp.87-92.

Stewart, F.J., and Raleigh, E.A. (1998). Dependence of McrBC cleavage on distance between recognition elements. Biol Chem.

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 Research, 40(15), pp.7552-7562.

Tamulaitis, G., Zaremba, M., Szczepanowski, R., Bochtler, M. and Siksnys, V. (2008). How PspGI, catalytic domain of EcoRII and Ecl18kI acquire specificities for different DNA targets. Nucleic Acids Research, 36(19), pp.6101-6108.

Wimley, W., and White, S. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nature Structural & Molecular Biology *3*, 842-848.

Wimley, W., Creamer, T., and White, S. (1996). Solvation Energies of Amino Acid Side Chains and Backbone in a Family of Host−Guest Pentapeptides†. Biochemistry *35*, 5109-5124.

Wong, I., Lundquist, A., Bernards, A. and Mosbaugh, D. (2002). Presteady-state Analysis of a Single Catalytic Turnover byEscherichia coliUracil-DNA Glycosylase Reveals a “Pinch-Pull-Push” Mechanism. Journal of Biological Chemistry, 277(22), pp.19424-19432.