

**Biochemical and crystallographic studies of Type III  
restriction  
modification enzymes: insights into the mechanism of ATP-  
dependent endonuclease**

**A thesis submitted for the partial fulfillment of the  
requirement**

**for the degree of**

**DOCTOR OF PHILOSOPHY IN BIOLOGY**

**by**

**Ishtiyaq Ahmed**

**20123154**



**INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE**

**May, 2018**

## **CERTIFICATION**

I certify that the thesis entitled Biochemical and crystallographic studies of Type III restriction modification enzymes: insights into the mechanism of ATP-dependent endonuclease, presented by Mr. Ishtiyaq Ahmed represents his original work which was carried out by him at IISER Pune, under my guidance and supervision during the period from 02/01/2012 to 24/05/2018. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution. I further certify that the above statements made by him in regard to his thesis are correct to the best of my knowledge

Date:

Dr.Saikrishnan Kayarat

## DECLARATION

I declare that this written submission represents my idea in my own words and where other ideas have been included I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date:

Ishtiyaq Ahmed  
(Registration No. 20123154)

## **Acknowledgements**

Ph.D. has been a long journey with lots of ups and downs throughout. Writing the thesis was one of the difficult part.

I feel privileged in expressing my profound sense of gratitude, high esteem and indebtedness to my supervisor, Dr. Saikrishnan Kayarat, Associate Professor, IISER Pune for his precious guidance, keen interest, discussions, undivided attention, critical opinion and valuable suggestions, during the course of my work. His guidance not only has helped me in my research but also have allowed me to grown as a research scientist.

I am thankful to my research advisory committee members, Dr. Radha Chauhan (NCCS Pune), Dr Nagaraj Balsubramanian (IISER Pune) and Dr. Sudha Rajamani for their inputs and suggestion during the course of my Ph.D. I am also thankful to Dr. Gayathri Pananghat for her critical insight and suggestion on my research work during our lab meetings. I am very grateful to the IISER Pune for providing me all the necessary facilities.

I also thank all the members of SK lab and G3 lab for their suggestions during lab meetings. Best wishes and thanks to senior and junior members of SK lab and G3 lab, particularly Dr. Manasi Kulkarni, Dr. Neha Nirwan, Mahesh Chand, Sujata Sharma, Jyoti Baranwal, Vishal Adhav, Vinayak Sadasivam, Pratima Singh, Sutirtha bandyopadhyay for their help in the various stages of this work. I would especially thank Aathira Gopinath and Karishma Bhagat, BS-MS students of IISER Pune, which whom I worked on some aspects of my thesis.

To my friends Dr. Imtiyaz Bhat, Shivik Garg, Saleem Yusuf, Wasim Mir, Zahid Bhat, Javeed Rashid, Manzoor Bhat for their advice support during the entire process. Finally and importantly I would like to thank my parents Mr. Gulam Nabi Khan, Shafiq Bano and my brother Mr. Basharat Ahmad Khan for their love, support and encouragement to follow my dreams.

Ishtiyag Ahmed

May 2018



## List of Abbreviations

|                     |                                       |
|---------------------|---------------------------------------|
| °C                  | Degree Celsius                        |
| μL                  | Microliters                           |
| μm                  | Micrometers                           |
| μM                  | Micromolar                            |
| 1D                  | 1 Dimensional                         |
| 3D                  | 3 Dimensional                         |
| Å                   | Angstrom                              |
| ABD                 | AdoMet binding domain                 |
| AdoMet              | S-adenosyl-L-methionine               |
| ADP                 | Adenosine diphosphate                 |
| ADP-VO <sub>4</sub> | Adenosine diphosphate vanadate        |
| AHJR                | Archeal Holliday Junction Resolvase   |
| AMP                 | Adenosine monophosphate               |
| AMPPP               | Adenylyl-imidodiphosphate             |
| ATP                 | Adenosine triphosphate                |
| bp                  | Base Pair                             |
| cm                  | Centimeters                           |
| CTD                 | C-terminal domain                     |
| DNA                 | Deoxyribonucleic acid                 |
| DNase               | Deoxyribonuclease                     |
| dsDNA               | Double stranded deoxyribonucleic acid |
| DTT                 | Dithiothreitol                        |
| <i>E. coli</i>      | <i>Escherichia coli</i>               |

|           |                                                                   |
|-----------|-------------------------------------------------------------------|
| EDTA      | Ethylenediaminetetraacetic acid                                   |
| EMSA      | Electrophoretic mobility shift assay                              |
| g/L       | grams per litre                                                   |
| GC        | Guanine and cytosine                                              |
| GTP       | Guanosine triphosphate                                            |
| HEPES     | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid                |
| HsdM      | Host specificity for DNA modification                             |
| HsdR      | Host specificity for DNA restriction                              |
| HsdS      | Host specificity for DNA specificity                              |
| H-to-H    | Head to Head                                                      |
| H-to-T    | Head to Tail                                                      |
| IDT       | Integrated DNA technologies                                       |
| Pi        | Inorganic phosphate                                               |
| IPTG      | Isopropyl $\beta$ -D-1-thiogalactopyranoside                      |
| ISP       | Type I Single Polypeptide                                         |
| K         | Kelvin                                                            |
| Kb        | Kilobase                                                          |
| KD        | Dissociation constant                                             |
| kDa       | Kilodalton                                                        |
| L         | Litre                                                             |
| LB        | Lysogeny broth                                                    |
| LC/ESI-MS | Liquid Chromatography / Electrospray ionization mass spectroscopy |
| Maldi-MS  | Matrix Assisted Laser Desorption mass spectrometry                |
| MDR       | Modification Dependent Restriction                                |

|        |                                    |
|--------|------------------------------------|
| mg     | Milligrams                         |
| mL     | Milliliters                        |
| mM     | Millimolar                         |
| Mod    | Modification                       |
| MPD    | 2-Methyl-2,4-pentanediol           |
| MW     | Molecular weight                   |
| MWCO   | Molecular weight cut-off           |
| NA     | Nucleic acid                       |
| NEB    | New England Biolabs                |
| Ni-NTA | Nickel nitrilotriacetic acid       |
| nm     | Nanometers                         |
| nt     | Nucleotide                         |
| NTD    | Nucleotide triphosphate            |
| OD     | Optical density                    |
| PacBio | Pacific BioSciences                |
| PAGE   | Polyacrylamide gel electrophoresis |
| PC     | Parent condition                   |
| PCR    | Polymerase chain reaction          |
| PEG    | Polyethylene glycol                |
| Py     | Pyrimidine                         |
| REBASE | Restriction enzyme database        |
| RecA   | Recombinase A                      |
| RecB   | Recombinase B                      |
| RF     | Restriction free                   |
| SDS    | Sodium dodecyl sulphate            |

|           |                                                              |
|-----------|--------------------------------------------------------------|
| SEC       | Size exclusion chromatography                                |
| SEC- MALS | Size exclusion chromatography - multi angle light scattering |
| Se-Met    | Selenomethionine                                             |
| SNF       | Sinefungin                                                   |
| SF 2      | Superfamily 2                                                |
| SMRT      | Single Molecule Real Time                                    |
| ssRNA     | Single-stranded RNA                                          |
| TBE       | Tris borate EDTA                                             |
| TIRF      | Total internal reflection fluorescence                       |
| TLC       | Translocation Looping & Collision                            |
| TRD       | Target recognition domain                                    |
| T-to-H    | Tail to Head                                                 |
| T-to-T    | Tail to Tail                                                 |
| UV-Vis    | Ultra violet- visible                                        |

## List of Tables

|                   |                                                                                    |     |
|-------------------|------------------------------------------------------------------------------------|-----|
| <b>Table 1.1:</b> | Essential features of NTP-dependent RM enzymes                                     | 4   |
| <b>Table 2.1:</b> | Primer used for the cloning and mutagenesis of EcoP15I and EcoP1I                  | 43  |
| <b>Table 2.2:</b> | Primer used for the complete sequencing of EcoP15I                                 | 44  |
| <b>Table 2.3:</b> | Oligomers used in biochemical assays for EcoP15I and EcoP1I                        | 46  |
| <b>Table 3.1:</b> | List of oligomers used for the crystallization of EcoP15I and EcoP1I               | 70  |
| <b>Table 3.2:</b> | Primers used for generating methionine mutants of EcoP1I                           | 71  |
| <b>Table 3.3:</b> | Composition of initial condition where crystal were obtained                       | 75  |
| <b>Table 3.4:</b> | Composition of initial condition where crystal were obtained                       | 76  |
| <b>Table 3.5:</b> | Composition of condition where EcoP1I-DNA complex was crystallized                 | 78  |
| <b>Table 3.6:</b> | Composition of initial condition that gave initial crystal                         | 79  |
| <b>Table 3.7:</b> | Data collection and processing statistics collected at ESRF ID29                   | 82  |
| <b>Table 3.8:</b> | Composition of initial conditions collections where initial crystals were obtained | 84  |
| <b>Table 4.1:</b> | Oligomers used in cloning and mutagenesis of MboIII                                | 96  |
| <b>Table 4.2:</b> | Oligomers used to form duplex substrates for nuclease assay                        | 98  |
| <b>Table 4.3:</b> | Primers used for complete sequencing of MboIII                                     | 98  |
| <b>Table 4.4:</b> | A list of Type III RM enzymes with SSR                                             | 120 |
| <b>Table 5.1:</b> | Primers used for generation of methionine mutants of Mod subunit                   | 134 |
| <b>Table 5.2:</b> | Composition of initial condition where initial crystals were observed              | 136 |

|                   |                                                                                  |     |
|-------------------|----------------------------------------------------------------------------------|-----|
| <b>Table 5.3:</b> | Data collection and processing statistics                                        | 138 |
| <b>Table 5.4:</b> | Data collection and refinement statistics of Mod crystallized under condition 2. | 139 |

## List of figures

|                     |                                                                                                                                                                                                  |    |
|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figure 1.1.</b>  | The Bertani and Weigle experiment.                                                                                                                                                               | 6  |
| <b>Figure 1.2.</b>  | Organization of <i>mod</i> and <i>res</i> gene in EcoP1I and EcoP15I.                                                                                                                            | 9  |
| <b>Figure 1.3.</b>  | Domain organization of Mod subunit from Type III RM enzyme.                                                                                                                                      | 12 |
| <b>Figure 1.4.</b>  | Domain organization of Res subunit from Type III RM enzyme.                                                                                                                                      | 14 |
| <b>Figure 1.5.</b>  | Recognition sequences and site orientation of EcoP1I and EcoP15I.                                                                                                                                | 16 |
| <b>Figure 1.6.</b>  | Schematic representation of the translocation looping collision model.                                                                                                                           | 19 |
| <b>Figure 1.7.</b>  | End reversal model.                                                                                                                                                                              | 20 |
| <b>Figure 1.8.</b>  | Transient looping collision model.                                                                                                                                                               | 22 |
| <b>Figure 1.9.</b>  | 1D diffusion model.                                                                                                                                                                              | 24 |
| <b>Figure 2.1.</b>  | Cartoon illustrating the possible models of single-site cleavage                                                                                                                                 | 39 |
| <b>Figure 2.2.</b>  | A schematic illustrating the products of a hypothetical single-site cleavage as single-strands that are observed when analyzed on a denaturing urea-formamide PAGE using ethidium bromide stain. | 40 |
| <b>Figure 2.3.</b>  | A heterologous cooperation assay of EcoP1I and EcoP15I.                                                                                                                                          | 41 |
| <b>Figure 2.4.</b>  | Cartoon representation of oligos used for biochemical assays                                                                                                                                     | 47 |
| <b>Figure 2.5.</b>  | Expression and purification of EcoP15I                                                                                                                                                           | 49 |
| <b>Figure 2.6.</b>  | DNA binding assay of EcoP15I with short DNA substrate                                                                                                                                            | 50 |
| <b>Figure 2.7.</b>  | Single-site cleavage by EcoP15I                                                                                                                                                                  | 51 |
| <b>Figure 2.8.</b>  | Effect of upstream end of recognition sequence on single-site DNA cleavage                                                                                                                       | 52 |
| <b>Figure 2.9.</b>  | Single-site cleavage is the result of cooperation between a <i>cis</i> - bound and a <i>trans</i> -acting enzyme                                                                                 | 54 |
| <b>Figure 2.10.</b> | <i>Trans</i> enzyme mediated DNA cleavage in single enzyme system                                                                                                                                | 55 |

|                     |                                                                                                                                                     |     |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 2.11.</b> | Cleavage assay demonstrating that the <i>cis</i> -bound enzyme cleaved the top strand and the <i>trans</i> -acting enzyme cleaved the bottom strand | 56  |
| <b>Figure 2.12.</b> | Role of ATP on single-site DNA cleavage                                                                                                             | 57  |
| <b>Figure 2.13.</b> | ATP hydrolysis by both enzymes is essential for two-site cleavage.                                                                                  | 58  |
| <b>Figure 2.14.</b> | Denaturing 18% urea-formamide PAGE showing that ATP hydrolysis is essential for nicking or dsDNA break in two site DNA substrate.                   | 59  |
| <b>Figure 2.15.</b> | Model for DNA cleavage by Type III RM enzymes EcoP1I and EcoP15I                                                                                    | 62  |
| <b>Figure 3.1.</b>  | Crystal of EcoP15I-DNA complex                                                                                                                      | 76  |
| <b>Figure 3.2.</b>  | Optimization of the condition for the co-crystallization of EcoP15I- DNA-ADPVO <sub>4</sub> .                                                       | 77  |
| <b>Figure 3.3.</b>  | Optimization of condition for co-crystallization of EcoP1I-DNA- AMPPNP.                                                                             | 80  |
| <b>Figure 3.4.</b>  | Position and number of methionine in EcoP1I                                                                                                         | 81  |
| <b>Figure 3.5.</b>  | Position of incorporations of additional methionine residues in the EcoP1I endonuclease domain.                                                     | 82  |
| <b>Figure 3.6.</b>  | Partial structure of EcoP1I-DNA-AMPPNP complex.                                                                                                     | 83  |
| <b>Figure 3.7.</b>  | Optimization of crystal for co-crystallization of EcoP15I-DNA-ADP VO <sub>4</sub> .                                                                 | 84  |
| <b>Figure 4.1.</b>  | Sequence alignment of the Mod and Res subunits of EcoP1I, EcoP15I and MboIII                                                                        | 93  |
| <b>Figure 4.2.</b>  | Gene organization of <i>MboIII</i> , <i>mod</i> and <i>res</i>                                                                                      | 96  |
| <b>Figure 4.3.</b>  | Gene organization of MboIII in <i>Mycoplasma bovis</i>                                                                                              | 104 |
| <b>Figure 4.4.</b>  | Sequence alignment of Mod1, Mod2 and Mod3 subunits present in the genomic DNA of <i>Mycoplasma bovis</i> Donetta 45 PG 45                           | 105 |



|                     |                                                                              |     |
|---------------------|------------------------------------------------------------------------------|-----|
| <b>Figure 4.5.</b>  | Purification and subunit stoichiometry of MbolIII.                           | 106 |
| <b>Figure 4.6.</b>  | Plasmid library assay of MbolIII                                             | 108 |
| <b>Figure 4.7.</b>  | Identification of MbolIII recognition sequence                               | 108 |
| <b>Figure 4.8.</b>  | Identification of the target base for methylation                            | 110 |
| <b>Figure 4.9.</b>  | Run-off sequencing to determine the cleavage loci of MbolIII                 | 111 |
| <b>Figure 4.10.</b> | Site requirement for DNA cleavage by MbolIII                                 | 112 |
| <b>Figure 4.11.</b> | Methylation activity of MbolIII                                              | 113 |
| <b>Figure 4.12.</b> | Competition assay between methylation and nuclease activities of MbolIII     | 114 |
| <b>Figure 4.13.</b> | Incorporation of 24 AG repeats does not affect Mod and Res complex formation | 115 |
| <b>Figure 4.14.</b> | Incorporation of EGFP into MbolIII                                           | 116 |
| <b>Figure 4.15.</b> | Effect of SSR and EGFP on the methylation and nuclease activity of MbolIII   | 117 |
| <b>Figure 4.16.</b> | Position of SSRs in the Mod subunit Type III RM enzymes                      | 120 |
| <b>Figure 4.17.</b> | Location of the SSR sites mapped on the structure of EcoP15I.                | 121 |
| <b>Figure 5.1.</b>  | Domain organization of MbolIII.                                              | 130 |
| <b>Figure 5.2.</b>  | Ribbon diagram of Rsrl and MbolIII.                                          | 131 |
| <b>Figure 5.3.</b>  | Crystals obtained after optimization of initial condition                    | 137 |
| <b>Figure 5.4.</b>  | Overall structure of Mod subunit of MbolIII with sinefungin                  | 139 |
| <b>Figure 5.5.</b>  | Secondary structure present in Mod dimer                                     | 140 |
| <b>Figure 5.6.</b>  | Asymmetry at the N-terminal domain of Mod protomers                          | 141 |
| <b>Figure 5.7.</b>  | AdoMet binding pocket in MbolIII Mod subunit                                 | 142 |
| <b>Figure 5.8.</b>  | Interaction of helical region (residues 128-155) from Mod-A with AdoMet      | 143 |

**Figure 5.9.** Superimposition of Mod-A of MbolIII with Mod-A of EcoP15I highlighting the location of TRD

144

## Table of Contents

|                       |      |
|-----------------------|------|
| Acknowledgement       | i    |
| List of Abbreviations | iii  |
| List of Table         | viii |
| List of Figures       | x    |

### Chapter 1: Introduction

|      |                                                               |    |
|------|---------------------------------------------------------------|----|
| 1.1  | ATP-dependent molecular switches                              | 2  |
| 1.2  | General introduction                                          | 3  |
| 1.3  | Background literature                                         | 5  |
| 1.4  | Identification of Type III RM enzymes                         | 7  |
| 1.5  | Properties of Type III RM enzymes                             | 8  |
| 1.6  | Models of DNA cleavage of Type III RM enzymes                 | 16 |
| 1.7  | Role of Type III RM enzymes in phase variation                | 24 |
| 1.8  | EcoP1I, EcoP15I and MboIII: Prototypes of Type III RM enzymes | 24 |
| 1.9  | Discussion                                                    | 25 |
| 1.10 | Scope of thesis                                               | 25 |

**Chapter 2: Single-site DNA cleavage by Type III RM enzymes need a site-bound enzyme and *trans* acting enzyme that are ATPase activated**

|     |                       |    |
|-----|-----------------------|----|
| 2.1 | Introduction          | 37 |
| 2.2 | Materials and methods | 42 |
| 2.3 | Results               | 49 |
| 2.4 | Discussion            | 60 |

**Chapter 3: Crystallographic studies of EcoP15I and EcoP1I**

|     |                              |    |
|-----|------------------------------|----|
| 3.1 | Introduction                 | 68 |
| 3.2 | Materials and methods        | 70 |
| 3.3 | Results                      | 74 |
| 3.4 | Structural studies on EcoP1I | 78 |
| 3.5 | Discussion                   | 84 |

**Chapter 4: Identification and biochemical characterization of Type III RM enzyme MbolIII from *Mycoplasma bovis***

|     |                       |     |
|-----|-----------------------|-----|
| 4.1 | Introduction          | 90  |
| 4.2 | Materials and methods | 92  |
| 4.3 | Results               | 101 |
| 4.4 | Discussion            | 116 |

**Chapter 5: Structure of Mod subunit of MbolIII**

|     |                       |     |
|-----|-----------------------|-----|
| 5.1 | Introduction          | 129 |
| 5.2 | Materials and methods | 133 |
| 5.3 | Results               | 135 |

|     |                                  |     |
|-----|----------------------------------|-----|
| 5.4 | Discussion                       | 145 |
| 5.5 | Conclusion and future directions | 146 |

# **CHAPTER 1**

## **Introduction**

## Chapter 1

### Introduction

#### 1.1. ATP-dependent molecular switches

Helicases are macromolecular proteins that utilize energy obtained from ATP hydrolysis to separate two complementary strands of a double-stranded dsDNA (1, 2, and 3). They are ubiquitous in nature and participate in a wide range of biological processes essential for the survival of an organism (4, 5, and 6). The classical view of helicases is that the energy obtained from the continuous cycle of ATP hydrolysis is used in the unwinding of dsDNA or dsRNA substrate, along with directed movement on the unwound single-strand ssDNA or ssRNA. In contrast, translocases are ATPase motors that use the chemical energy for movement along dsDNA or dsRNA substrates without unwinding them (7, 8). Apart from helicases and translocases, there is another class of ATPases that act as a molecular switch by carrying out a single round of ATP hydrolysis that possibly leads to large conformation change in the protein, which allows the enzyme to perform its task (9). In case of helicases or translocase, the motor continuously hydrolyses ATP, while a molecular switch uses only a single or a small number of hydrolysis cycles.

A clear-cut difference between a helicase/translocase with that of a molecular switch is that at the end of one ATP hydrolysis cycle the former has moved one-step forward while the latter returns to its original conformation (10). Examples of the existence of such molecular switches have been found in MMR (mismatch repair) protein MutS, MutL and Type III restriction-modification (RM) enzymes (9). In MMR system binding of the ATP to MutS results in the conformational change in MutS to form sliding clamp, after which it starts randomly diffusing on the DNA in search of DNA lesion. For the duration of diffusion, MutS maintain periodic contact with the DNA and acts as a platform to recruit MutL. The MutS-MutL complex than searches for the DNA lesion by 1D diffusion while remaining in contact with DNA. Further binding of MutH to MutL slows the 1D diffusion of MutS-MutL complex to locate the hemimethylated Dam GATC DNA strand (11, 12).

Type III restriction-modification (RM) enzymes are bacterial defense systems against bacteriophage attacks. They are composed of two subunits, a modification (Mod) subunit having methyltransferase activity and a restriction (Res) subunit having

endonucleolytic activity (13). The Res subunit contains an ATPase or helicase domain and an endonuclease domain (13). To cleave DNA, Type III RM enzymes need two inversely oriented recognition sites and the presence of ATP (14, 15). The role of helicase motif is to facilitate a long range communication between two RM enzymes bound on the foreign DNA. Earlier studies proposed that Type III RM enzymes communicate between two recognition sites by the DNA looping mechanism (16, 17), however later studies using magnetic tweezers assay and single-molecule fluorescence microscopy and magnetic tweezers proposed a 1D diffusion model (18,19). According to 1D diffusion model, binding of ATP to the enzyme bound to recognition site causes a rapid burst of ATP hydrolysis. The ATP hydrolysis acts as a molecular switch which results in a drastic conformational change in the enzyme following which the enzyme undergoes 1D diffusion on the DNA. Whether there is further need of ATP during diffusion or to maintain the conformational change in the enzyme is still unknown. When a diffusive enzyme encounters a stationary enzyme bound to target site, a dsDNA break happens (19). In the MutS-MutL system, ATP binding acts as a molecular switch for DNA sliding conformation to happen, while in Type III RM enzymes ATP hydrolysis acts as a molecular switch to trigger the enzyme to undergo 1D diffusion state.

## **1.2. General introduction of RM enzymes**

Bacteria under continuous attack of bacteriophages, have evolved various defense and offense strategies for their protection. Restriction Modification (RM) enzymes is one of the defense systems present in bacteria to prevent attack of invading bacteriophage (20, 21, 22, 23, 24, 25, 26, 27, 28). RM enzymes were discovered more than 60 years ago and the genes encoding them have been found in most of the sequenced bacterial genomes. RM enzymes are composed of two subunits with opposing activities – (i) a restriction or endonuclease subunit that cleaves foreign DNA at specific DNA sequences called the recognition sequence; (ii) a methyltransferase subunit (Mod) which prevents host DNA cleavage by methylating the DNA within the same sequence (13). Depending upon the requirement of nucleotide for DNA cleavage, RM enzymes are further classified into NTP-dependent and NTP-independent RM enzymes (29). In addition, based on the



subunit assembly, recognition site orientation, cofactor requirement and position of DNA cleavage RM enzymes are further classified into four types - Type I, II, III and IV (29). Type II RM enzymes belong to the NTP-independent class (30). Type IV RM enzymes, also referred to as modification-dependent RM enzymes that cleave DNA having recognition sequences that have modified bases (31). Table 1.1 represents the essential features of Type I and Type III RM enzymes.

**Table 1.1:** Essential features of NTP-dependent RM enzyme

| Features                              | Type I                              | Type III                          |
|---------------------------------------|-------------------------------------|-----------------------------------|
| NTP-dependence                        | ATP                                 | ATP                               |
| Oligomeric State of complex           | Pentamer                            | Heterotrimer                      |
| Subunit composition                   | HsdR, HsdM, HsdS                    | Res and Mod                       |
| Position of DNA cleavage              | Variable, far from recognition site | Close to any one recognition site |
| Cofactor Requirement for methylation  | AdoMet (S-adenosyl methionine)      | AdoMet (S-adenosyl methionine)    |
| Cofactor Requirement for DNA cleavage | Mg <sup>2+</sup> , ATP              | Mg <sup>2+</sup> , ATP            |

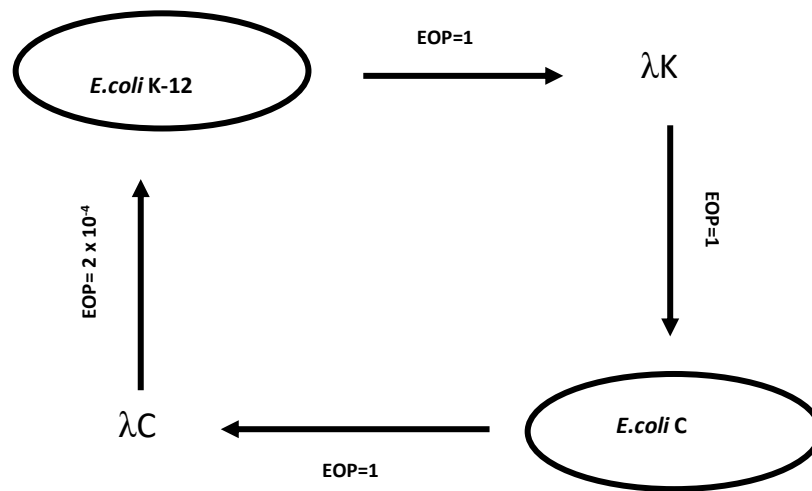
Type III RM enzymes are ATP-dependent multifunctional and multisubunit protein systems involved in bacterial defense. They are heterotrimeric complexes composed of a dimeric modification (Mod) subunit and a single restriction (Res) subunit (13). The dimeric Mod subunit catalyses site-specific methylation of host DNA, while Res subunit on forming a complex with Mod cleaves the unmethylated DNA in the presence of ATP. Type III RM enzymes recognize asymmetric recognition sequences that can be 5-6 bp long (13). To cleave DNA, Type III RM enzymes need at least two such recognition sites

in an inverted orientation. These sites can be separated upto 3.5 kb with the DNA cleavage happening, 25-27 bp downstream to any of one the recognition sites (32, 33). Type III RM enzymes were identified more than 40 years ago. Our understanding of the mechanism of these enzymes has improved a lot mainly from biochemical and biophysical studies carried out on EcoP1I and EcoP15I, the two prototypes of Type III RM enzymes. However, because of the lack of atomic structure, the molecular mechanism of action of these enzymes is still incomplete. In this thesis, I have tried to understand how these enzymes nucleolytically cleave DNA using biochemical and structural studies. To understand the structural basis of how EcoP1I, EcoP15I function, we carried out crystallographic studies of these enzymes not only in the presence of specific DNA, but also with non-hydrolysable ATP analogues. Apart from working on EcoP1I and EcoP15I, I also isolated and purified a new Type III RM enzyme MboIII from *Mycoplasma bovis*. I used this enzyme as an alternate model system to study the mechanism of Type III RM enzymes.

In the subsequent sections of this chapter a brief history and overview of RM enzymes, and discussion about the discovery and characteristic features of Type III RM enzymes are described. The outstanding questions in the field of Type III RM enzymes are discussed towards the end of this chapter.

### **1.3. Background literature:**

The phenomenon of restriction-modification was first observed in 1950's by Luria and Human as a barrier that exists against bacteriophage  $\lambda$  infection in its natural host *Escherichia coli* K-12 strain (34, 35). They found that bacteriophage  $\lambda$  previously propagated on *E. coli* C strain were not able to infect *E. coli* K-12 its natural host. However, few phages that survived on *E. coli* K-12 and their future progeny were now able to infect *E. coli* K-12. From these observations, it was proposed that bacteria strain applied an identification mark to the phages residing in it and same phages when propagated on different bacteria strain lost the identification mark of previous bacterial strain (33, 34). After one generation of growth on a previous host, the phage was restored to infect its original host (Figure 1.1).



**Figure 1.1. The Bertani and Weigle experiment.** Phages ( $\lambda$ K) isolated from *E. coli* K-12 infected *E. coli* K-12 with efficiency of plating (EOP) of 1 (EOP = the relative number of plaques which the phage is capable of producing).  $\lambda$ K phage infected *E. coli* C with EOP of 1. Phages  $\lambda$ C propagated on *E. coli* C now faced a barrier and were restricted if propagated on *E. coli* K-12 resulting in low EOP. However, after single round of replication in *E. coli* K-12, the phages were able to overcome this barrier and were able to infect with an EOP of 1.

In 1962, Werner Arber and his colleagues while working on a radiation-resistant *E. coli* B/r strain observed a drastic degradation of bacteriophage DNA on an invasion-resistant strain of bacteria. They tried to understand how certain bacterial strains were resistant to bacteriophage infection, and how some phages adapt to a new host. Based on their results, they were successful in showing that bacterial strains may express an enzyme that puts a self-mark on its own DNA, and that the invading phages lacked the mark on its DNA (21, 22). Phages that were successful on propagating on *E. coli* were imparted with the same mark on its DNA (21, 22). On the basis of these findings, Arber proposed that there may be two proteins in the bacterium, a restriction enzyme that would recognize and cleave the phage DNA lacking self-mark, while a modification enzyme would recognize the host DNA and would methylate it. It was also proposed that both proteins would act on the same sequence of DNA (36). These findings led to the proposal of a bacterial self-defense system against foreign DNA, where a host enzyme would not only protect the bacteria against bacteriophage DNA, but would also prevent self-DNA

from degradation by methylating it. A few years later, they were able to experimentally validate their hypothesis by purifying the modification and the restriction enzymes from the resistant *E. coli* strain. Subsequently, it was shown that the modification subunit adds a methyl group to the DNA, while the restriction enzyme cleaved the unmethylated DNA.

EcoK and EcoB were the first restriction enzymes to be identified and were later classified as Type I RM enzymes (36). It was found that the activities of EcoK and EcoB required magnesium, S-adenosyl methionine (AdoMet) and ATP as cofactors (37). It was later shown that EcoB cleaved the DNA away from the recognition site at a random position (38). In 1970, Hamilton was successful in purifying a restriction enzyme from *Haemophilus influenza* (HindIII), which, in contrast to EcoK and EcoB, could cut DNA within the recognition sequence (39). HindIII represented the first example of a Type II restriction enzyme, which led to the revolution of genetic engineering.

#### 1.4. Identification of Type III RM enzymes

Until 1960, all the RM enzymes discovered were Type I and were encoded by the chromosomal region of *E. coli*. The first experimental evidence for the presence of non-chromosomal RM enzymes came from genetic experiments carried out on four strains of *E. coli* - K12, B, 15T<sup>-</sup> and K12(P1) lysogen. Phage  $\lambda$ , which were isolated from *E. coli* K12 after propagation were restricted when propagated on *E. coli* K12(P1) lysogen with an efficiency of plating (E.O.P) of  $10^{-4}$ . However phages which survived and propagated on *E. coli* K12(P1) lysogen when isolated were successful in propagating in *E. coli* K12(P1) lysogen with an (E.O.P) of 1 (21, 40). The results indicate that  $\lambda$  phages when propagated on *E. coli* K12(P1) lysogen incorporated a modification which made them immune to restriction. Based on the finding, it was proposed that there are two RM enzymes in *E. coli* K12(P1) independent of each other, and only one RM enzyme in *E. coli* K12 (21, 40). Further experiment by Arber and co-workers provided conclusive proof that the second RM enzyme in *E. coli* K12(P1) was encoded by prophage P1 (24, 41). This was the first RM enzyme identified to be non-chromosomal in nature. Further using conjugation studies between *E. coli* 15T<sup>-</sup> and *E. coli* K12, Stacey et al. (1965) showed that strain *E. coli* 15T has a RM enzyme which is different compared to the RM enzyme present in *E.*

*coli* K12, which later led to the identification of a new RM enzyme encoded by the plasmid P15B in *E. coli* 15T<sup>-</sup> (27, 42). It was also shown that P15B plasmid can recombine with the P1 lysogen for stable inheritance (43). Mutational studies on *E. coli* 15T which carried both EcoA and EcoP15I indicated that the presence of both RM enzymes had an additive effect on restriction (43). Similarly, *E. coli* K12(P1) lysogen containing EcoP1I and EcoK also had additive effect on. EcoA<sup>-</sup> *E. coli* 15T strain was generated upon treated with N-methyl-N-nitroso-N'-nitroguanidine. This mutant strain was then transduced with phage P1 having EcoP15I genes. Next, both the wild type and mutant strains were tested for restriction upon treatment with Phage  $\lambda$ . It was noticed that maximum restriction occurred when both EcoA and EcoP15I were present (43). This observation reconfirmed that *E. coli* 15T strain have two RM enzymes EcoA and EcoP15I that have different specificity.

As  $\lambda$  phage experiments in *E. coli* led to the identification of RM system, similar experiments by Piekarowicz and Kauc using phage HP1c1 and *Haemophilus influenza* led to the identification of HinfIII RM enzyme (44, 45). After their identification, it was assumed that EcoP1I and EcoP15I belonged to the same class as EcoA and EcoB. However, AdoMet depletion experiment with *E. coli* 15T and *E. coli* K12 (P1) lysogen strain showed that while EcoK and EcoB needed AdoMet for DNA restriction, EcoP15I and EcoP1I did not depend on AdoMet for restriction (46). An important finding was the lack of a huge requirement of ATP by EcoP1I and EcoP15I for DNA cleavage, which was unlike Type I RM enzymes. These observations led to the classification of EcoP1I, EcoP15I and HinfIII into a separate class of RM enzymes named Type III RM systems (47).

### 1.5. Properties of Type III RM enzymes

Even though >15000 putative Type III R-M enzymes has been identified till date (<http://rebase.neb.com/cgi-bin/azlist?re3>), only a couple of them have been well studied, with most of the work being carried out on EcoP1I and EcoP15I (9, 15, 48). This section will describe in detail the characteristic properties of Type III RM enzymes including genetic makeup, subunit assembly, domain organization and proposed models explaining mode of DNA cleavage.

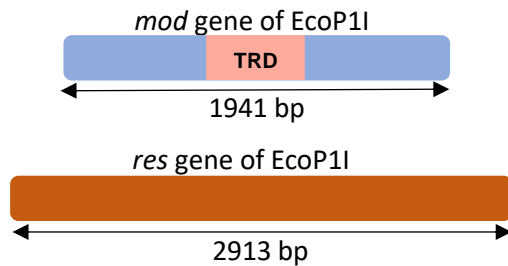
### 1.5.1 Genetic make up

Detailed information on gene organization and size of EcoP1I and EcoP15I came from experiments performed by Mural et al. (1979) (49). They generated a series of DNA fragments by digesting the phage P1 DNA with EcoRI and BamHI and ligated it into vectors. A 9.2 kb fragment obtained by the digestion when inserted in a recombinant vector expressed both the modification and restriction subunits of EcoP1I (49). Insertion mutations using transposons showed that the gene product of Res subunit is 111 kDa in size (50). Lida et al. in 1983 performed restriction fragment analysis of the 9.2 kb DNA fragment of phage P1 and P1-P15 hybrid phage. P1-P15 hybrid phage having the restriction specificity of EcoP15I (51). The restriction maps constructed of P1 phage and P1-P15 hybrid showed the presence of different cleavage sites in P1 phage compared to P1-P15 hybrid. Restriction map analysis together with insertional and deletion mutant of phage P1 and P1-P15 hybrid demonstrated that the forward region of 2.2 kb length encoded Mod subunit, while the adjacent region of 2.8 kb encoded Res subunit. Heteroduplex analysis of the restriction fragment containing genes of EcoP1I and EcoP15I revealed a large degree of homology between them. The *res* gene in EcoP1I and EcoP15I was highly identical to each other (51). The *mod* gene of EcoP1I and EcoP15I were homologous to each other at their start and end region with a large non-homologous region in between the two ends (Figure 1.2). As EcoP1I and EcoP15I recognize different recognition sequences, it was hypothesized that the non-homologous region confers to their different specificities. This was later confirmed by DNA sequencing of the genes encoding EcoP1I and EcoP15I. Using *in vitro* transcription it was shown that both genes are transcribed from separate promoters (51).

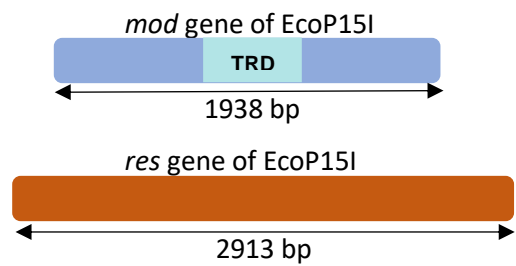
Type III RM enzymes are hetero-oligomeric complexes composed of Mod subunits, which catalyse site-specific methylation of host DNA, and a Res subunit that in complex with Mod cleaves unmodified DNA utilizing ATP (13). Based on results obtained from analytical centrifugation and size exclusion chromatography, it was widely accepted that Type III RM enzymes are hetero-tetrameric complexes composed of Mod and Res subunit in the ratio of (Mod<sub>2</sub>Res<sub>2</sub>) (52, 53). However, the study from Wyszomirski et al. using similar techniques showed that EcoP15I exists in two oligomeric states (Mod<sub>2</sub>Res<sub>2</sub>) and (Mod<sub>2</sub>Res<sub>1</sub>) with the heterotrimeric complex showing more specific activity (54).

Gupta et al. (2012) using analytical ultracentrifugation and SAXS techniques proposed that EcoP15I has an elongated crescent shape in solution with Mod dimer occupying the center of the complex and a Res subunit is present at each end of the dimeric Mod (55).


### Gene organization of EcoP1I



### Gene organization of EcoP15I



 Target recognition domain (TRD) in EcoP1I

 Target recognition domain (TRD) in EcoP15I

**Figure 1.2. Organization of *mod* and *res* gene in EcoP1I and EcoP15I.** The square box in *mod* gene represents the TRD region with lower sequence identity. The blue region in Mod subunits of EcoP1I and EcoP15I have higher sequence identity. The brown region represent the *res* gene in EcoP1I and EcoP15I.

In 2014, Butterer et al. carried a series of experiments on EcoP1I, EcoP15I and PstI using native mass spectrometry and size exclusion chromatography coupled to multiple angle light scattering to determine their stoichiometric composition. The data showed that Type III RM enzymes are heterotrimeric complexes composed of two Mod subunits and one Res subunit (56). Finally, putting all speculations to rest, Gupta et al. in 2015, published a partial structure of EcoP15I bound to its DNA (57). The structure reconfirmed SEC-MALS data from Butterer et al that EcoP15I is a heterotrimeric complex of Mod<sub>2</sub> Res<sub>1</sub> (56, 57).

#### 1.5.3 Domain organization of Type III RM enzymes

The first insight into the structural domains present in a Type III RM came from the study by Wagenfuehr et al. 2007. Proteolytic studies showed that Mod subunit degraded rapidly, indicating absence of any stable domains (58). However on addition of specific DNA, proteolytic degradation of Mod subunit was inhibited. The Restriction subunit gave two stable products of 77 kDa and 27 kDa. However, with addition of specific DNA and ATP

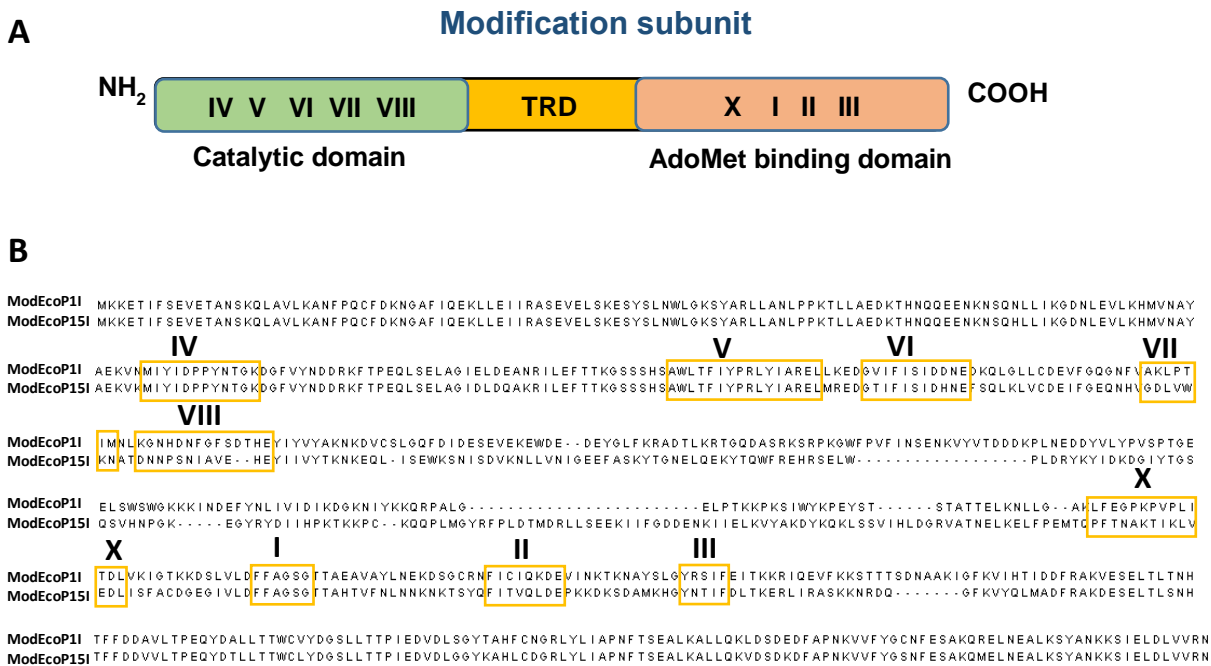
the Res subunit was completely protected against proteolytic degradation. This showed that upon binding to DNA and in presence ATP the enzyme is protected against proteolytic degradation. The products of limited proteolysis of EcoP15I were analyzed using liquid chromatography/electron spray ionization mass spectroscopy (LC/ESI-MS) and matrix-assisted laser desorption mass spectroscopy (MALDI-MS). It was observed that the 77 kDa stably folded domain of EcoP15I is an N-terminal region of Res subunit containing helicase and ATPase motifs. The smaller 27 kDa stably folded domain was from the C-terminal end and contained the conserved catalytic motif of the endonuclease. The stably folded domains of Res subunits are connected to each other by a 23 amino acids linker (58).

#### *1.5.3.1 Modification subunit*

Methylation of host DNA is essential for protection against self-suicide due to the restriction activity. This task is achieved by methylation of host DNA at specific adenine or cytosine in the recognition sequence. In Type III RM enzymes EcoP1I and EcoP15I, the Mod subunit prevents self-suicide by methylating second adenine in their respective recognition sequences ( $AG^{m6}ACC$  &  $CAGC^{m6}AG$ )(13). Based on the nature of the base and position at which the base is methylated, MTases are classified into two types. The C5-methyltransferases methylates the cytosine base in the recognition site at the C5 position yielding C5-methyl cytosine (5mC). The second group, N-Mtases, methylates exocyclic amino group of adenine or cytosine at N6 and N4 positions, respectively, forming N6-methyladenine (N6mA) and N4-methylcytosine (N4mC). Results from the sequence alignment of 5-methyl cytosine transferase revealed the presence of 10 motifs arranged in a linear order, with the variable or TRD region present at the C-terminal end (59, 60, 61, 62). However, it was found that N-MTases have only nine conserved motifs, (I-VIII and X) (63). Depending upon the linear arrangement of the nine motifs, N-MTases are further classified into three different types -  $\alpha$ ,  $\beta$ , and  $\gamma$  types. EcoP1I, EcoP15I and MboIII belong to the  $\beta$ -group of adenine methyltransferases. Furthermore, based on the classification by Timinkas et al. (1995), EcoP1I and EcoP15I belong to the D21 class where D is a conserved aspartate in the catalytic motif IV (DPPY) (63). Also in the D21 class of N-MTases AdoMet binding motif (FXGXG) is present after the motif IV. Malone



et al performed a structure-guided analysis of MTases. From the analysis, the MTases were divided into three functional domains: a catalytic domain, a target recognition domain (TRD) and an AdoMet binding domain. EcoP1I, EcoP15I and MboIII have the catalytic domain containing the motifs IV, V, VI and VIII present at the N-terminal end, the AdoMet binding domain containing the motifs X, I, II and III is present at the C-terminal end, while the TRD is present in between the catalytic and AdoMet domain (64) (Figure 1.3 A & B).



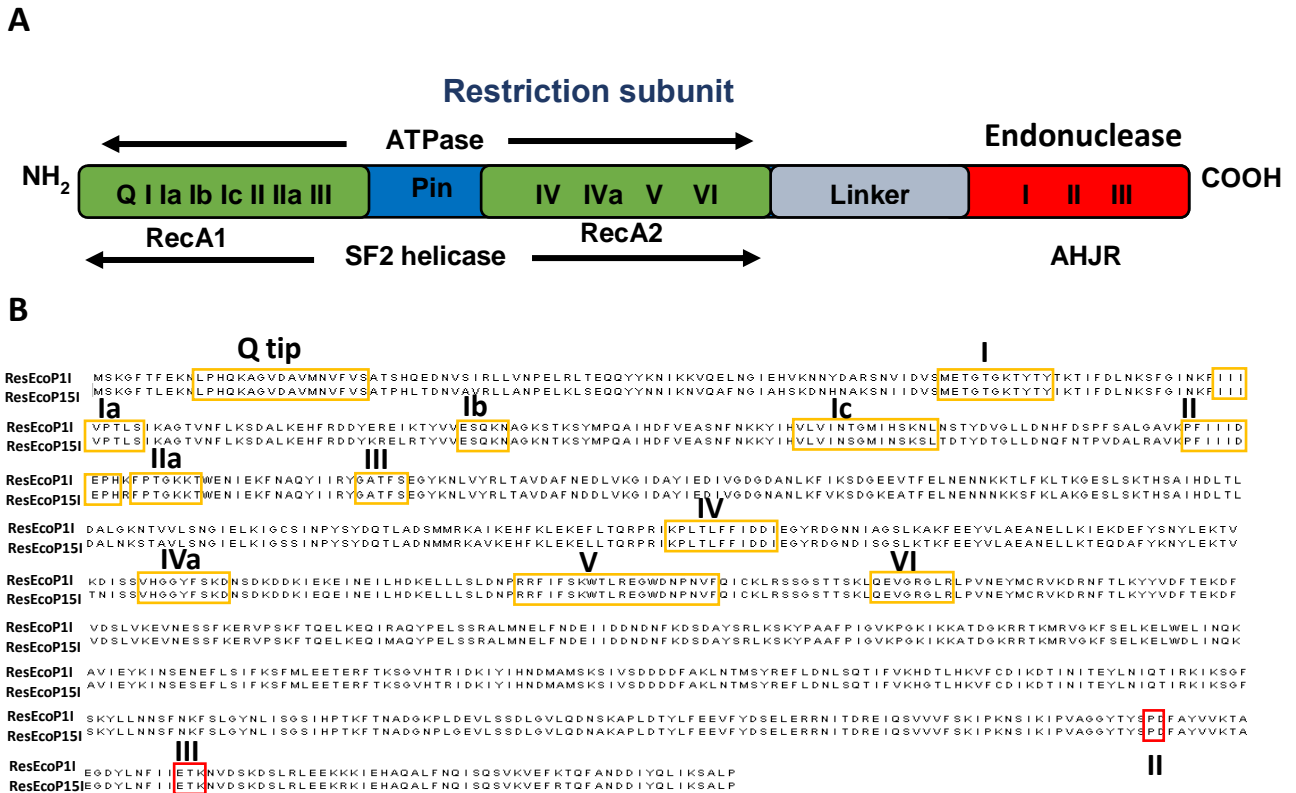
**Figure 1.3 Domain organization of Mod subunit from Type III RM enzyme.** (A) The modification (Mod) subunit contains an N-terminal catalytic domain with a target recognition domain (TRD) in the middle and an AdoMet binding domain at the C-terminal end. (B) Sequence alignment of the Mod subunits of EcoP1I and EcoP15I highlighting the methyltransferase motifs (Yellow boxes). The Roman numbers above the yellow boxes indicate the motif name.

### 1.5.3.2. Restriction subunit

The Res subunits of EcoP1I and EcoP15I catalyze the cleavage of foreign DNA in a complex with the corresponding Mod subunit and in the presence of ATP. The Res subunit is close to 111 kDa in molecular weight (65). After the amino acid sequence of the Res subunit of Type III RM enzymes were determined, Gorbalenya and Koonin

performed the sequence alignment of EcoP1I and with EcoK. Based on the sequence alignment, it was found that the N-terminal region of Res subunit contains the signature motifs belonging to the Superfamily 2 (SF2) helicases (66). Bona fide helicases unwind duplex DNA utilizing energy derived from ATP hydrolysis. As Type III RM enzymes contain the same SF2 helicase motifs, it was initially suggested that these enzymes also unwind dsDNA duplex (67). However, biochemical and single molecule studies have shown that these proteins lack strand separation activity and instead utilize ATP hydrolysis to communicate between two recognition sites for DNA cleavage (19). The N-terminal part of the Res subunit containing the ATPase domain has two subdomains having a similar fold as the Recombinase A (RecA). These two subdomains are present in tandem and are called RecA1 and RecA2 (Figure 1.4 A) (67). The two subdomains contain motifs which coordinate binding and hydrolysis of ATP to nucleic acid remodeling (3). With the availability of the structure of EcoP15I bound to specific DNA, motifs Ib, Ic and IIa were identified (Figure 1.4 B). From the structure of EcoP15I, the authors reported an additional domain called as the Pin domain formed by the insertion of 77 amino acids into the RecA2 subdomain.

The endonuclease domain in Type III RM enzymes is positioned at the C-terminal end of the Res subunit. The endonuclease domain is joined to N-terminal ATPase domain by a long linker (Figure 1.4 A). The C-terminal endonuclease domain contains motifs belonging to the Archeal Holliday Junction Resolvase (AHJR) family of nucleases (68). Members belonging to this family have a fold similar to that present in  $\lambda$  exonucleases and other endonucleases like EcoRV (68, 69, and 70). A characteristic feature of the AHJR family of endonucleases is the presence of three conserved motifs (motif I, motif II and motif III) (68). These three motifs are involved in the formation of a catalytic triad, with the conserved aspartate from motif II (PD) and a conserved glutamate, glutamine or aspartate in motif III (Q/E/DxK) involved in coordination with  $Mg^{2+}$  ion as well as in the scission of the phosphodiester bond of DNA (Figure 1.4 B). Motif I is characterized by the presence of conserved acidic amino acid residue (D/E), however the consensus sequence of motif I is still undetermined (71, 72, 73). The possible role of the acidic side chain of motif I is in stabilizing the binding to  $Mg^{2+}$  or the conformational changes accompanied in the catalytic triad.



**Figure 1.4. Domain organization of Res subunit of Type III RM enzyme.** (A) The restriction (Res) subunit contains an N-terminal ATPase domain with a linker in the middle and C-terminal endonuclease domain. (B) Sequence alignment of the Res subunits of EcoP11 and EcoP15I highlighting the ATPase motifs (Yellow boxes) and endonuclease motifs (Red boxes). The Roman numbers above the yellow and red boxes indicate the motif name.

#### 1.5.4. Determination of Target site

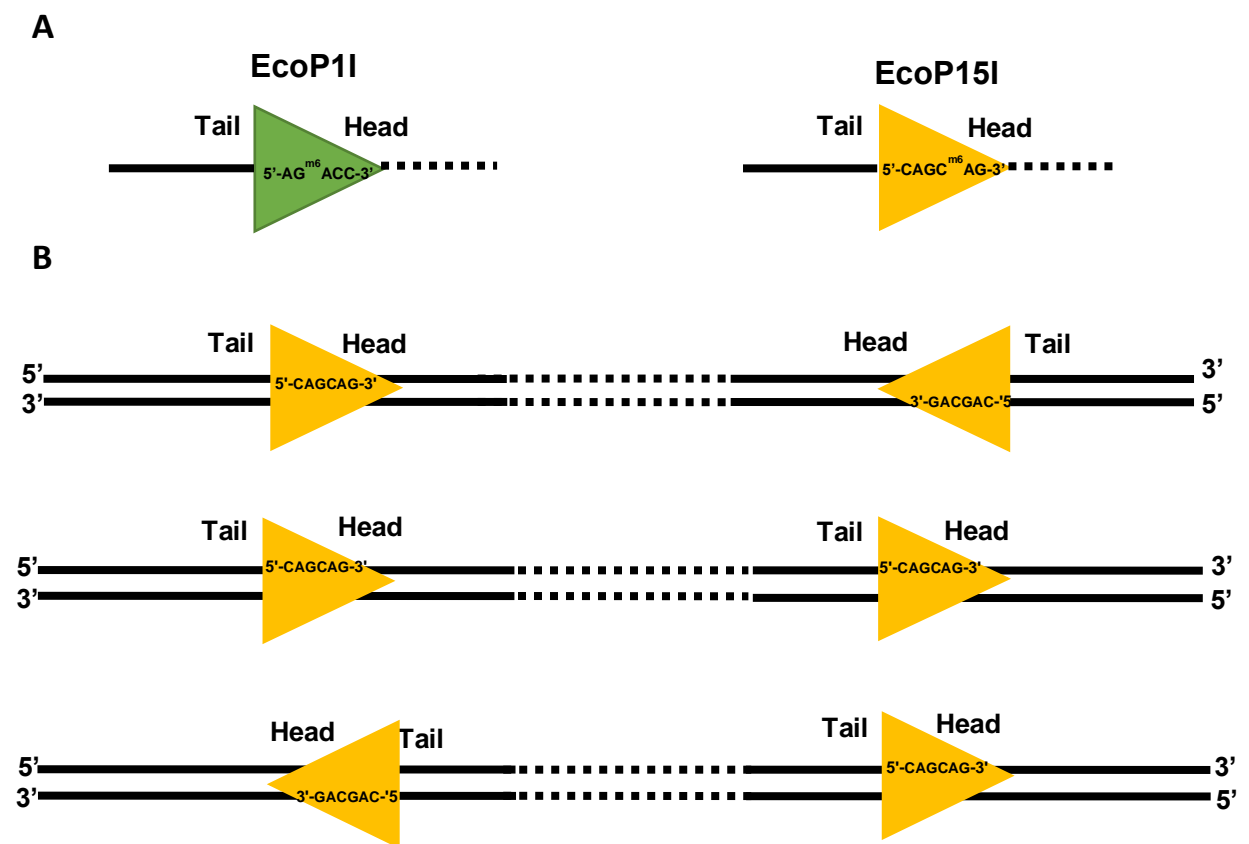
The experiments towards determination of the recognition sequence and methylated base of Type III RM enzymes were based on the assumption that restriction modification enzyme methylates one of the bases in their recognition sequence. In 1973 Brokes et al. were first to show using radioactive AdoMet that EcoP11 methylates adenine base in the DNA substrate (74). Treatment of  $\lambda$  phage DNA with EcoP11 Mod subunit in presence of radiolabelled AdoMet, showed that EcoP11 recognizes the pentameric sequence 5'-AGACY-3' (Y= C/T) with second adenine being methylated (75). Bachii et al.(1979) did *in-vitro* methylation of SV40 DNA with EcoP11 in the presence of radioactive AdoMet (15).

The methylated DNA was digested with the restriction enzymes (HaeIII, AluI, HinfI) to identify the methylated products. Two such products were sequenced. From the sequencing results, the authors were able to map the methylation sequence as well as cleavage loci. Again, it was found that EcoP1I methylates second adenine in the target site (AG<sup>m6</sup>ACC) (15). Using *in-vitro* DNA methylation assay, the recognition sequence of EcoP15I was identified as CAGC<sup>m6</sup>AG with the second adenine in the recognition site methylated (14). A careful look at the recognition sequence of EcoP1I and EcoP15I indicates that there are no adenines in the complementary strands of the recognition site which means that these enzymes methylate only the top strand. The recognition site of another Type III RM enzyme HinfIII was identified to be 5'-CGAAT-3' with enzyme methylating second adenine in the recognition site. Subsequently, another Type III RM enzyme, HineI, was discovered that recognizes the same recognition site as HinfIII making them the first isoschizomers (76).

#### 1.5.5. Requirement of two recognition sites for DNA cleavage

Early indications for the requirement of two recognition sites for successful DNA cleavage came from the studies on the nucleolytic activity of HinfIII using ColE1 DNA by Piekarowicz and co-workers (47, 48). A DNA substrate with a single recognition site of HinfIII was not cleaved indicating the requirement of at least two recognition sites for DNA cleavage. Using restriction fragment analysis of the linearized ColE1 DNA obtained on treating with HinfIII, the authors were able to map five potential recognition sites of HinfIII (48). As single-site DNA substrates were refractive to cleavage, the authors postulated that Type III RM enzymes need multiple sites for a successful nucleolytic cleavage (48). Type III RM enzymes mentioned so far recognize asymmetric recognition sites with the enzymes methylating only one of the adenine at the top strand and none at the bottom strand. It was noticed that such asymmetric methylation would be problematic when the modified DNA replicate. As DNA replication is semi-conservative, one daughter DNA molecule will inherit the methylated DNA strand while the other daughter DNA molecule will be unmodified, which will be the substrate for endonuclease cleavage. This would be lethal for the cell. Studies on the nucleolytic activity of EcoP15I on T7 and T3 phage DNA provided an explanation for how the two newly replicated double-stranded DNA is

protected. Although genomic DNA of both phages contains recognition sites of EcoP15I, T3 phage was cleaved by EcoP15I while, surprisingly, T7 phage was refractive to cleavage. Although T7 phage contains 36 recognition sites of EcoP15I, all of them are in same orientation, while in T3 phage the orientation of many of the recognition sites are inverted. Later, a series of cleavage assay with M13 DNA, where recognition sites of EcoP15I were placed in inverted, but same orientation, were performed. Only those DNA having the recognition sites in inverted orientation were cleaved (32). From these studies, a strand bias model was proposed for Type III RM enzymes, which explained prevention



**Figure 1.5. Recognition sequences of EcoP1I and EcoP15I (Green and Yellow triangles respectively).** (A) The 3' end of recognition sequence denotes the head and the 5' end of the recognition sequence denotes tail of recognition site. (B) Orientation of recognition sites on a dsDNA.

of a suicide cleavage of host DNA, following DNA replication. According to this model, Type III RM enzyme needs two recognition sites in an inverted orientation for the cleavage to happen. Following DNA replication, one site of the new DNA will be methylated while

another site will be unmethylated, and such a combination of sites will be refractive to cleavage (32). Now, because of the polarity associated with the asymmetric recognition sequence, we can have the recognition site in three different orientations i.e. head-to-head, head-to-tail and tail-to-tail orientation (Figure 1.5).

### **1.6. Models of DNA cleavage by Type III RM enzymes**

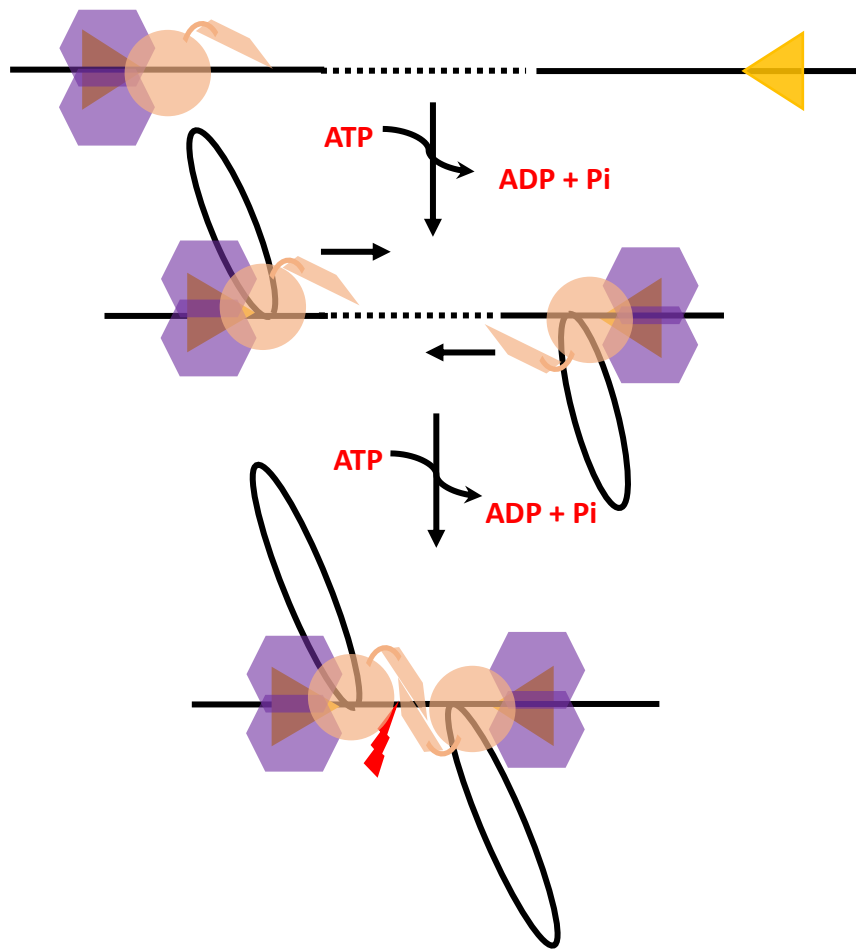
DNA cleavage by Type III RM enzymes need a long-range communication between two enzymes bound to inversely oriented recognition sites, which can be thousands base pair apart (13). The DNA is always cleaved downstream to any one of the recognition sites (13). The long-range communication by Type III RM enzymes between two inverted recognition sites requires 10-12 ATP molecules irrespective of the distance between them (19, 78). This is in contrast to the Type I RM enzymes where one ATP molecule is consumed for every base pairs that separates the two sites (79, 80). It is not completely understood as to how two enzyme molecules bound to their recognition sites that can be thousand base pair apart communicate/translocate on DNA resulting in the nucleolytic cleavage. Recognition sites in head-to-head orientation are favored for DNA cleavage as the collision between two enzymes is head on, with the endonuclease domains pointing towards each other. However, in tail-to-tail orientated recognition sites, the collision between two enzymes is not head on, as the two endonuclease domains are pointing away from each other. Also, the diffusing enzymes in tail-tail oriented sites will fall off on reaching the free ends of DNA (32, 80, and 81). The efficiency of DNA cleavage on tail-to-tail substrates can be increased by blocking the ends of the DNA. The finding indicates that the blocking of the DNA ends increased the lifetime of the enzyme on the DNA by virtue of which tail-to-tail substrates were cleaved with an equal efficiency. Further, communication between two enzyme molecules is essential for the DNA cleavage, as placing a roadblock between two inverted recognition sites prevents the DNA cleavage. The effect of the roadblock was more pronounced on a linear DNA substrate compared to circular DNA substrates (83).

Apart from canonical DNA substrates of Type III RM enzymes with recognition sites present in head-to-head or tail-to-tail orientation, there are reports of DNA substrates

containing single recognition sequences being cleaved (56, 84, 85, 86, 87, 88, 89). Such activities of Type III RM enzymes were observed to happen only at a high protein to DNA concentration, in presence of particular cations and were thus considered as promiscuous cleavage events. For example, with EcoP1I it was found that under permissive conditions (i.e in presence of  $K^+$  ions) DNA cleavage happened at both the recognition sites on a two-site inverted DNA substrate. Under the same condition, a DNA substrate having a single recognition site was also cleaved. However, under non-permissive conditions (in presence of  $Na^+$  ions) such promiscuous DNA cleavage was not noticed (89). Similarly, with EcoP15I it was shown that in the presence of AdoMet analogue sinefungin, EcoP15I was found to cleave DNA substrate at every recognition sequence (87). This non-promiscuous DNA cleavage is also called secondary DNA cleavage event and its efficiency is slower compared to primary or permissive DNA cleavage. To explain about how these enzymes communicate between two recognition sites, various models have been proposed as mentioned below.

#### *1.6.1 Translocation, looping and collision model*

Type III R-M enzyme contains the Res subunit containing SF2 helicase motifs. Because of the presence of a similar SF2 helicase motif present in Type I R-M enzymes, a similar model of translocation, loop extrusion and collision (TLC) which operates in Type I RM enzymes was proposed for Type III RM enzymes (90) (Figure 1.6). According to this model, after the enzymes bind to their recognition site, they start actively pulling DNA towards them. The pulling of DNA by two enzymes results in the extrusion of all DNA between them in the form of loops. The extrusion of the DNA will eventually result in the collision of two enzyme complexes. The collision of two enzyme complexes will result in the dsDNA break. For every base pair of DNA which is translocated by the enzyme one ATP is consumed (6, 80). However, an interesting thing observed in Type III R-M enzymes was that they consume a fraction of the ATP compared to Type I R-M enzyme but perform a similar physiological function. Thus, TLC model failed to explain how Type III RM perform their biological task while consuming little ATP. Furthermore, the TLC model also failed to give explanation of how single-site circular or linear DNA substrates and substrates having two recognition sites close to one another, gets cleaved.



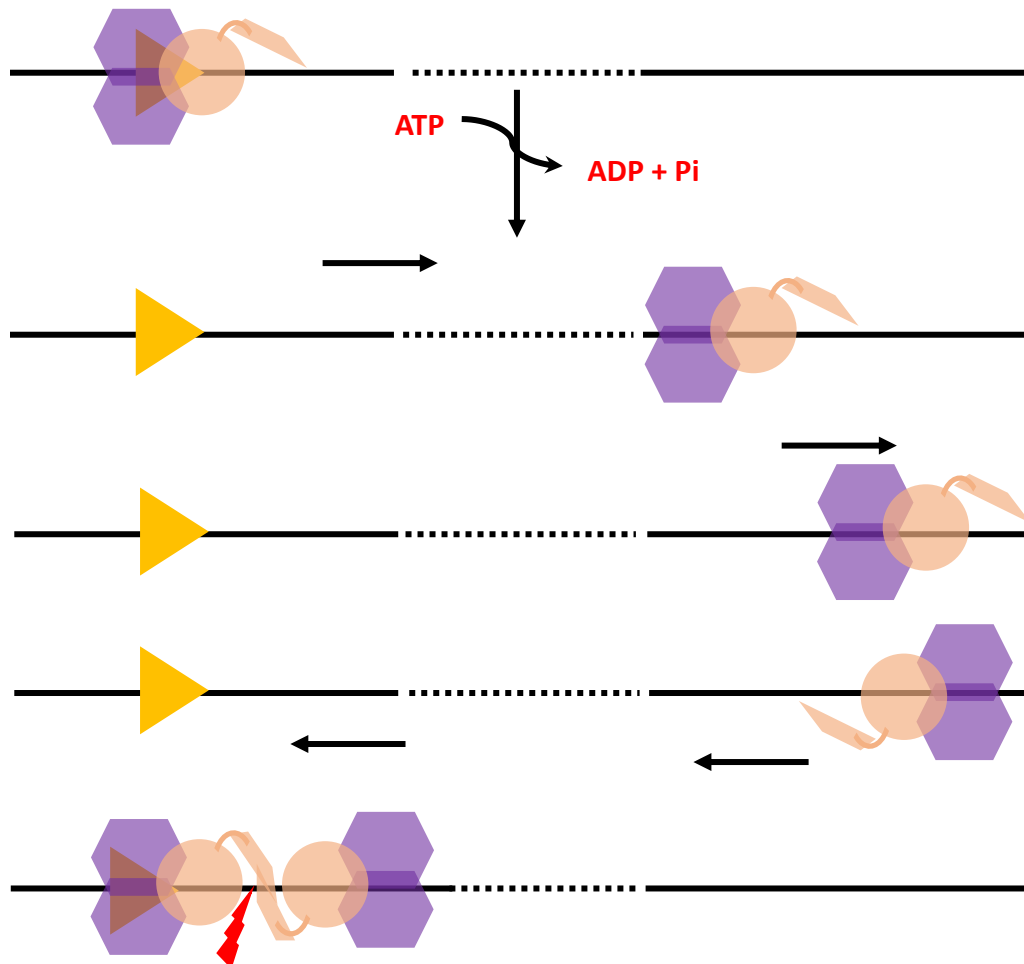
**Figure 1.6. Schematic representation of the translocation looping collision model.** On binding of the enzyme (Mod colored purple and Res colored beige) to the recognition site (yellow triangle), they actively start to loop DNA towards them using ATP hydrolysis. The DNA between two enzymes is looped out, which results in the collision of two enzymes. When enzymes collide, a dsDNA break happens.

### 1.6.2. End Reversal model

A two-site substrate (either linear or circular) with sites in head-to-head orientation is a canonical substrate for Type III RM enzymes (32, 80, and 81). It has been noticed that apart from cleaving canonical substrate Type III RM enzymes can cleave single site circular or linear DNA substrates (56, 84, 85, 86, 87, 88, and 89). The TLC model also failed to explain why the efficiency of cleavage of tail-to-tail substrates was enhanced when ends were blocked with streptavidin. To explain how single-site linear DNA



substrates get cleaved, Raghavendra and Rao et al. in 2004 proposed end reversal model (Figure 1.6).



**Figure 1.6: End reversal model.** After the binding of the enzyme to the recognition sequence (Yellow triangle), ATP hydrolysis occurs. The ATP hydrolysis drives the enzyme to vacate and translocate downstream of the recognition site. On reaching the 3' end of the DNA, the translocating enzyme reverses its direction. In the meanwhile, the recognition site is occupied by another enzyme. When the enzyme translocating in reverse direction collides with the site bound enzyme a dsDNA break happens.

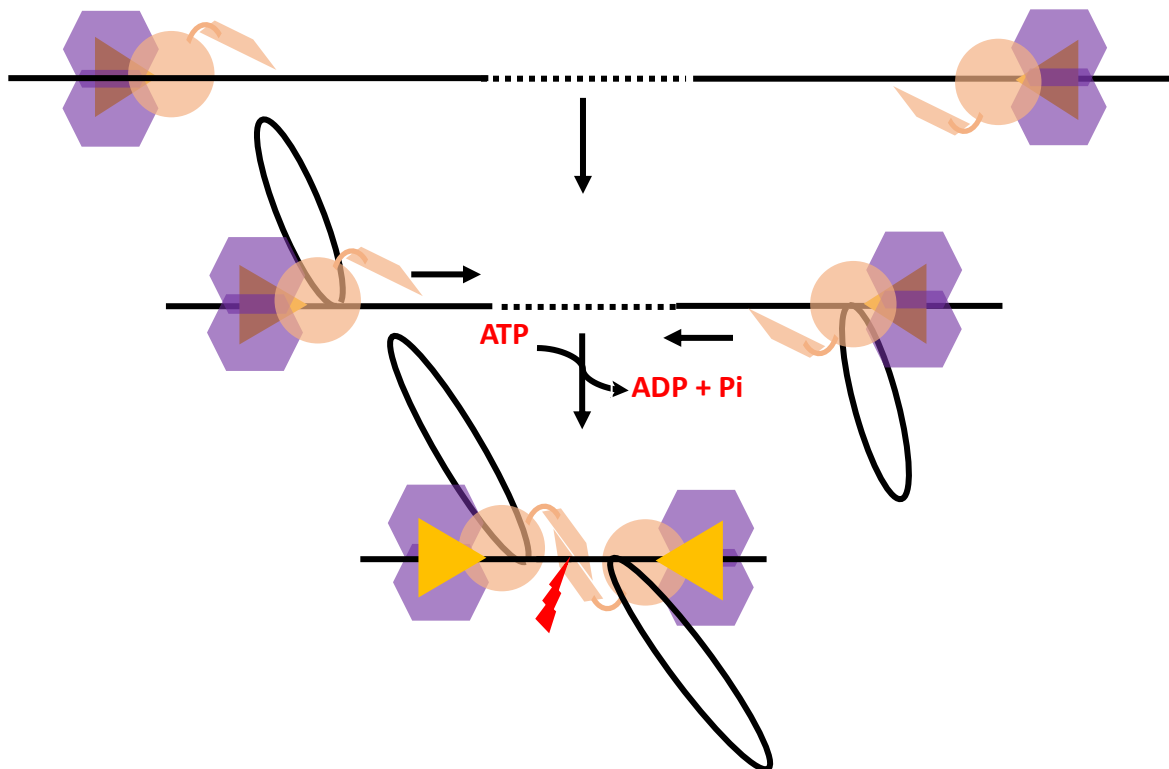
According to the model, after the enzyme binds to their recognition site, ATP hydrolysis occurs. The ATP hydrolysis drives the enzyme to vacate the recognition site and start tracking the DNA in the 3' direction downstream to the target site. When the translocating enzymes collide with the other enzyme bound to recognition site in proper orientation, a dsDNA break will happen. However, if the translocating enzyme will reach

the free end of the DNA substrate, either the enzyme will fall off or perform 180° reversal of the translocation polarity (86). The model when proposed was based on the assumption that the stoichiometry of the Type III RM enzymes is Mod<sub>2</sub>Res<sub>2</sub>, such that on reaching DNA end, the ATPase of the other Res subunit will get activated and the enzyme will start translocating backwards. When two translocating collides with each other at random position on DNA, no cleavage will happen, however if the reverse translocating enzyme collides with the site bound enzyme, the DNA cleavage will happen (86). Although end reversal model is able to explain how DNA substrates with sites in head-to-head, tail-to-tail site and linear DNA substrate containing single site can get cleaved, it fails to explain the cleavage of single-site circular DNA substrate. In case of circular single site DNA substrates, there is no free end so that the translocating enzyme cannot reverse the direction of translocation. The translocating enzyme would simply walk off the recognition sequence. Therefore, the end reversal model fails to explain the cleavage of single-site circular substrates. Moreover as the model assumes that when the translocating enzyme reaches the DNA end, it reverses its direction, which will mean that even head-to-tail substrates will be cleaved efficiently that is not the case. More importantly from biophysical and structural studies we now know that Type III RM enzymes are hetrotrimeric composed of two Mod subunit and a single Res subunit (56, 57). The assumption of end reversal model that two Res subunit can propel enzyme in forward or reverse direction is no longer valid.

### *1.6.3. Transient looping and translocation model*

On a DNA substrate containing single recognition site, it was shown using AFM that EcoP15I bound to its recognition site tend to form DNA loop. The DNA loops were only formed in the presence of specific DNA with no such loops observed with non-specific DNA. In some cases, it was observed that looped DNA did not contain the target site indicating that the enzyme can move past the recognition site after looping out the DNA. The enzyme was observed to hold the DNA adjacent to the recognition site. Based on this observation, a model of DNA cleavage was proposed (Figure 1.7). According to this model, after the enzyme binds to its DNA, it starts to loop DNA by diffusion. The looping of DNA by diffusion will brings the two enzyme complexes in close proximity to each other

without ATP hydrolysis. A limited ATP driven DNA looping will bring the two enzyme-DNA complexes together resulting in the dsDNA break. The model assumes ATP is needed at the end stages, thus explaining the need of low ATP consumption. Although DNA loops were observed on a single site substrate and it was proposed that enzyme after DNA looping would go past the recognition site, allowing second enzyme molecule to bind to recognition site. However, no enzyme dimer on a single DNA molecule was observed in AFM studies on a single site substrate (16, 17). Thus, the model failed to explain cleavage on a single site substrate or a substrate with two closely packed recognition sites.



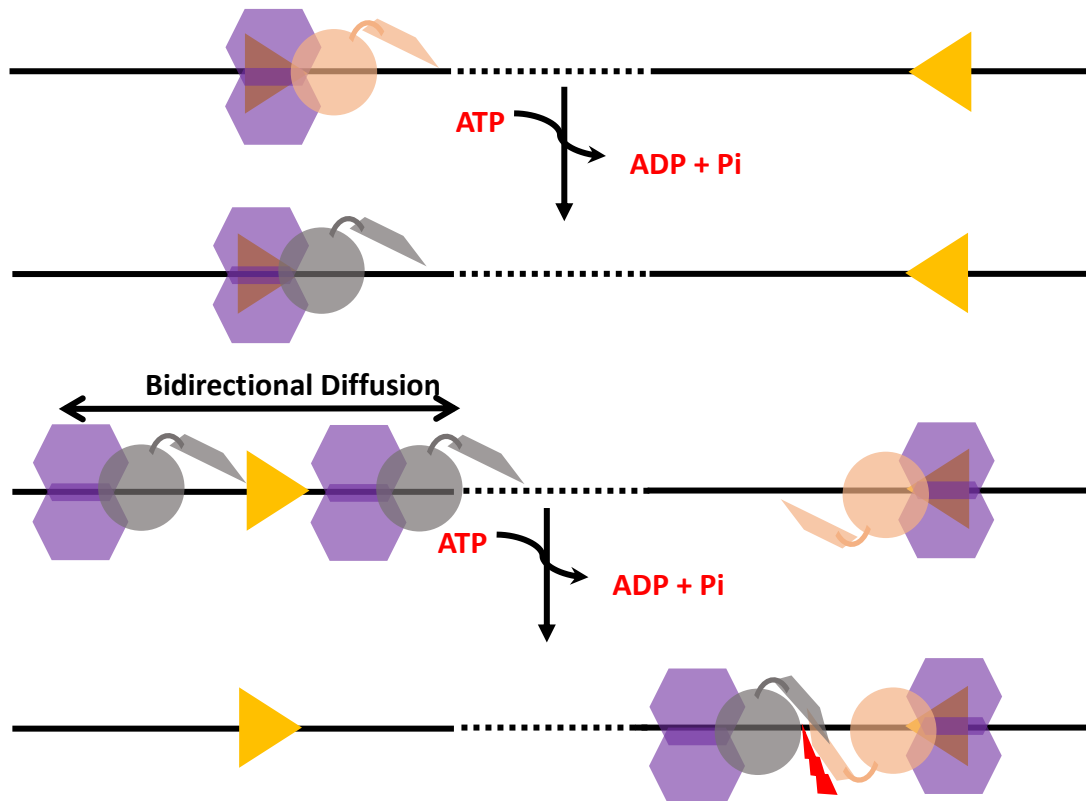
**Figure 1.7. Transient looping collision model.** On binding of enzyme to the DNA the enzyme passively loop the DNA between the two recognition sites *via* 3D looping, which shortens the separation between them. A limited ATP hydrolysis leads to bringing two enzyme close to one another which leads to the collision and a dsDNA break.

#### 1.6.4. 1D diffusion model

First experimental evidence suggesting Type III RM enzymes executed 1D diffusion came from studies carried out by Ramanathan et al. in 2009. Using magnetic tweezers experiments in which the DNA at one end was labelled with digoxigenin, while the other end was labelled with a magnetic bead (18). The DNA was attached to the biotinylated glass surface by strong binding of digoxigenin to biotin. On applying magnetic field across the other end, the DNA is held in upright position. On addition of enzyme and ATP, if the two enzymes communicate with each other by looping out the DNA in between, then it will result in the shortening of the end-to-end length of DNA. However, with EcoP1I and EcoP15I, no such shortening of DNA length was observed, instead after some time the magnetic bead was lost from the view indicating that DNA cleavage had happened (18). This observation clearly indicated that Type III R-M enzymes cleaved the DNA between two recognition sites without looping the DNA in between them. Based on this observation and other studies, a 1D diffusion model was proposed for Type III RM enzymes. According to this model, after the enzyme binds to its recognition site, ATP hydrolysis occurs after which the enzyme vacates the recognition site and starts bidirectional diffusion on the DNA. When one such diffusing enzyme collides with the other enzyme bound to its recognition site, DNA cleavage occurs. The model assumes that ATP is required only for the enzyme to vacate the recognition site and to enter into diffusing state (18).

Using the single molecule fluorescence microscopy Schwarz et al in 2013 refined the 1D diffusion model of EcoP15I and proposed the DNA sliding mechanism used by EcoP15I (19). According to the model, following binding of the enzyme to its recognition sequence, there is a rapid hydrolysis of about 10-12 ATP molecules that results in a large conformational change in the enzyme as indicated by change in tryptophan fluorescence. The initial conformational change can be accomplished only in the presence of ATP and no conformational change was observed with non-hydrolysable ATP analogues. After initial conformation change, the enzyme starts a bidirectional movement of the DNA with negligible or no ATP consumed during the sliding mode. During the sliding mode, the enzyme can dissociate from the internal DNA sites, or fall off from the DNA. After dissociation, the enzyme returns back to its original conformation and can rebind to the DNA and start another round of DNA binding-diffusion cycle. Encounter between a

diffusing enzyme and a stationary enzyme bound to its target site, triggers a dsDNA break (Figure 1.8) (19). The 1D bidirectional diffusion model explains the low ATP requirement by Type III RM enzymes, however, it fails to explain how a single-site cleavage by Type III RM enzymes. The 1D diffusion model is not limited to Type III RM enzymes, similar NTP driven 1D diffusion has also been reported in mismatch repair family of proteins (11, 12).



**Figure 1.8. 1D diffusion model.** After specific binding of enzyme to DNA, ATP hydrolysis occurs. The ATP hydrolysis leads to a large conformational change in the protein (represented by change in color of enzyme) from stationary to sliding state. The sliding enzyme can undergo a bidirectional diffusion on DNA. When a diffusive enzyme encounters a stationary enzyme bound to the recognition site, a dsDNA break happens.

### 1.7. Role of Type III RM enzymes in Phase variation

Host-adapted pathogens have developed various mechanism to evade the host immune response. One such mechanism is to keep generating diverse variety of cell surface markers (90, 91). Phase-variation, random switching on/off of gene expression is one of

the mechanisms used by host-adapted pathogens to produce genetic diversity and enhanced virulence (92). Sequence analysis of a number pathogenic bacteria revealed the presence phase variable Type III RM enzyme (93). It is believed that phase-variable Type III RM enzymes in host-adapted pathogens, apart from acting as a defense strategy to foreign DNA attack, also plays a role in phase-variation. In Type III RM enzymes, phase-variation is mediated by the presence of simple sequence repeats (SSRs) in the *mod* gene (94). Increase or decrease in the length of the SSR happens during DNA replication due to slipped-strand mispairing (95). Variation in repeat number in a coding region can lead to frame shift mutations resulting in alternative protein coding or a truncated non-functional polypeptide. This short term loss of the Type III RM enzyme removes a barrier for bacteria to uptake beneficial foreign DNA from the environment (91, 96). Although the effect of SSRs resulting in phase-variation is studied in detail in prokaryotes, whether the length of SSR has any role in the functionalities of a protein is not known. Our biochemical studies led to the identification and characterization of a new Type III RM enzyme from MbolIII from *Mycoplasma bovis*. A unique feature of Type III RM from *Mycoplasma sp.* is the presence of SSRs in their *mod* genes. Using MbolIII as a model, I tried to address whether increase or decrease in the length of SSR has any role on the activity of protein.

### **1.8. EcoP1I, EcoP15I, MbolIII: Prototypes of Type III RM enzymes**

Although a lot of biochemical and biophysical studies carried out over the years have shed a light on the working mechanism of these enzymes, a complete understanding of these enzymes is still lacking. There are still questions that remain unanswered.

1. How are single-site DNA substrates endonucleolytically cleaved? How does ATP hydrolysis activate the enzyme to perform a nucleolytic cleavage?
2. What conformational changes happen upon ATP hydrolysis which activates the enzyme?
3. Does the length of SSR have any affect the activities of Type III RM enzymes?

## 5. Mechanistic insights into the Mod subunit of Type III RM enzymes.

Towards answering these questions, we started working on EcoP1I and EcoP15I. Apart from working on well-studied RM enzymes, we identified and characterized a new Type III RM enzyme, MboIII, from *Mycoplasma bovis*. EcoP1I and EcoP15I are close homologues with almost 70% sequence identity in their Mod subunit and about 90% sequence identity in their Res subunit. EcoP1I and EcoP15I are heterotrimeric complexes with a MW ~260 kDa. MboIII has a MW of 227 kDa and is composed to two Mod subunits and a single Res subunit. The recognition site of EcoP1I, EcoP15I and MboIII is AGACC, CAGCAG and YAATC (Y= T/C), respectively, with the enzymes methylating second adenine in the recognition sites. For all the enzymes, two pairs of unmethylated inversely oriented recognition sites are must for the DNA cleavage to happen. The DNA cleavage needs ATP, Mg<sup>2+</sup> and always happen 25-27 bp downstream of any one recognition sequence and generates a 2 bp 5' overhangs.

To completely understand the molecular mechanism of Type III RM enzymes, we have carried out biochemical and structural studies on EcoP1, EcoP15I and MboIII. The insight from the thesis will help us better understand how these enzymes function.

## 1.9. Discussion

Type III RM enzymes are a member NTP-dependent RM enzymes and act as a barrier in bacteria against bacteriophage infection. Members of this class are characterized by an oligomeric state of the active complex, DNA cleavage loci, and the low amount of ATP consumption. EcoP1I and EcoP15I are the two prototypes of Type III RM enzymes. EcoP1I and EcoP15I are heterotrimeric composed of a Mod dimer and a Res subunit. The Mod subunit contains an N-terminal catalytic domain, a TRD in the middle and a C-terminal S-adenosyl methionine binding domain. The Mod subunit belongs to  $\beta$  class of methyltransferase and methylates the second adenine in one of the strands of the recognition site. The Res subunit contains an N-terminal ATPase domain belonging to SF2 helicases and a C-terminal endonuclease domain belonging to the AHJR family. The

Res subunit on forming a complex with dimeric Mod can search for unmethylated DNA in a cell.

Endonucleolytic cleavage of unmethylated DNA requires two inversely oriented sites that can be up to 3.5 kb apart with cleavage happening 25-27 bp downstream to any one target site. Although, a pair of an inverted oriented site is a prerequisite for DNA cleavage, there are reports in literature where DNA substrate containing a single-site as well closely spaced sites are cleaved. One of the fascinating aspects of the enzyme, which has remained poorly understood is the low ATP consumption by these enzymes compared to Type I RM enzymes. Recently, an ATP-dependent molecular switch mechanism similar to that observed in mismatch repair protein has been proposed. The model was derived from magnetic tweezers assay combined with TIRF microscopy. According to the model, on binding of the enzymes to their recognition sites, 10-12 ATP molecules are hydrolysed. The enzyme then undergoes a drastic conformational change, by virtue of which it undergoes 1D diffusion along the DNA. The DNA cleavage happens, when a diffusing enzyme collides with a stationary enzyme bound to its recognition sequence. It is not clear whether ATP is required during 1D diffusion of the enzyme.

A partial structure of EcoP15I with DNA gave some insight about the mechanism of DNA translocation by these enzymes. The structure gave critical insights how the division of labor is established between two Mod subunits. From the structure, it was noticed that one Mod subunit catalyzed reading of the recognition sequence, while the other Mod subunit catalyzes transfer of methyl group from AdoMet to the second adenine in the recognition sequence. However, the structure does not provide explanation how ATP will activate these enzyme to undergo 1D diffusion along the DNA. Furthermore, in the present structure, the electron density of the linker region and endonuclease domain is missing, so we don't know how ATP hydrolysis activates the enzymes as well the endonuclease domain. All these questions have remained unanswered because of the lack of complete atomic structure.

## **1.10. Scope of the Thesis**



Although our understanding about the working of these enzymes has increased a lot because of series of biochemical and biophysical studies, a complete understanding about the working of these enzymes is still unclear. In this thesis, we have tried to address question using the biochemical and structural studies.

Chapter 2: In this chapter we address and propose the mechanism by which Type III RM cleave the single-site DNA substrate, and also study in details the role of ATP at various stages of the nucleolytic cleavage.

Chapter 3: To understand the structural basis of the working of these enzymes, we in this chapter describe the co-crystallization studies of EcoP1I and EcoP15I with their specific DNA substrates. Also, described are the co-crystallization experiments of EcoP1I/EcoP15I with specific DNA in the presence of non-hydrolysable ATP analogues to trap the conformational states, which are accompanied with ATP hydrolysis.

Chapter 4: Type III RM enzymes are large molecular weight proteins, it is challenging to crystallize these proteins and obtain their high-resolution structures. As an alternate strategy, we isolated and characterized a new Type III RM enzyme from *Mycoplasma bovis*, which we called MboIII. MboIII is almost 30 kDa smaller in size compared to EcoP1I and EcoP15I but works functionally in a similar way. This chapter describes in details the biochemical and biophysical characterization of the MboIII. One of the interesting features of MboIII is the presence of a Simple Sequence Repeat (SSR) in the *mod* gene. We also studied the effect of the length of SSR on the activity of MboIII, which is described in this chapter.

Chapter 5: While we were carrying out the biochemical studies on MboIII, we initiated the crystallization studies of Mod, Res and MboIII. We were successful in crystallizing Mod subunit of MboIII with AdoMet analogue sinefungin. The structure of the dimeric Mod gave unprecedented insights about how a dimeric Mod functions.

**References:**

1. Wu CG, Spies M. Overview: what are DNA helicases? DNA helicases and DNA motor proteins. 2012.
2. Caruthers, J.M., McKay, D.B. (2002) Helicase structure and mechanism. *Curr Opin Struct Biol.* 12(1):123–33.
3. Singleton, M.R., Dillingham, M.S., Wigley, D.B (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem*, 76:23-50
4. Cordin. O., Banroques. J., Tanner. N.K., Linder. P (2006) The DEAD-box protein family of RNA helicases. *Gene*, 367:17-37.
5. Jankowsky, E., Fairman, M. (2007) RNA helicases — one fold for many functions. *Curr Opin Struct Biol*, 17:316-324.
6. Kadare, G., Haenni, A.L. (1997) Virus-encoded RNA helicases. *J Virol* : 2583-2590.
7. Stanley, L.K., Seidel, R., van der Scheer, C., Dekker, N.H., Szczelkun, M.D., Dekker, C (2006) When a helicase is not a helicase: dsDNA tracking by the motor protein EcoR124I. *EMBOJ.*, 25(10):2230–9.
8. Ryan, D.P., Owen-Hughes, T (2011) Snf2-family proteins: chromatin remodelers for any occasion. *Curr Opin Chem Biol.* 15(5):649–56.
9. Szczelkun, M.D. (2011). Translocation, switching and gating: potential roles for ATP in long- range communication on DNA by type III restriction endonucleases. *Biochem Soc. Trans.*, 39, 589-594.
10. Szczelkun, M.D. (2013) Roles for helicases as ATP-dependent molecular switches. *Adv. Exp. Med. Biol.*, **767**, 225–244.
11. Gorman, J., Chowdhury, A., Surtees, J.A., Shimada, J., Reichman, D.R., Alani, E., and Greene, E.C. (2007) Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. *Mol. Cell*, 28, 359-370.
12. Liu, J., Hanne, J., Brooke, M.B., Bennett, J., Kim, D., Lee, J.B., and Fishel, R. (2016) Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. *Nature*, 539, 583-589.

13. Rao, D.N., Dryden, D.T.F. and Bheemanaik, S. (2014) Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.*, **42**, 45-55.
14. Hadi, S.M., Baechi, B., Shepherd, J.C.W., Yuan, R., Ineichen, K. and Bickle, T.A. (1979) DNA Recognition and Cleavage by the EcoPI5 Restriction Endonuclease. *J. Mol. Biol.*, **134**, 655–666.
15. Baechi, B., Reiser, J. and Pirrota, V. (1979) Methylation and Cleavage Sequences of the EcoPI Enzyme. *J. Mol. Biol.*, **128**, 143–163.
16. Crampton, N., Roes, S., Dryden, D.T.F., Rao, D.N., Edwardson, J.M. and Henderson, R.M. (2007) DNA looping and translocation provide an optimal cleavage mechanism for the type III restriction enzymes. *EMBO J.*, **26**, 3815-3825.
17. Crampton, N., Yokokawa, M., Dryden, D.T.F., Edwardson, J.M., Rao, D.N., Takeyasu, K., Yoshiimura, S.H. and Henderson, R.M. (2007) Fast-scan atomic force microscopy reveals that the type III restriction enzyme EcoPP15I is capable of DNA translocation and looping. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 12775-12760.
18. Ramanathan, S.P., van Aelst, K., Sears, A., Peakman, L.J., Diffin, F.M., Szczelkun, M.D., and Seidel, R. (2009) Type III restriction enzymes communicate in 1D without looping between their target sites. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 1748-1753.
19. Schwarz, F.W., Toth, J., van Aelst, K., Cui, G., Clausing, S., Szczelkun, M.D. and Seidel, R. (2013) The Helicase-Like Domains of Type III Restriction Enzymes Trigger Long-Range Diffusion Along DNA. *Science*, **340**, 353–356.
20. Luria, S.E. and Human, M.L. (1952) A Nonhereditary, Host-Induced Variation of Bacterial Viruses. *J. Bacteriol.*, **64**, 557–569.
21. Bertani, G. and Weigle, J.J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.*, **65**, 113–121.
22. Arber, W. and Dussoix, D. (1962) Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J. Mol. Biol.*, **5**, 18-36.
23. Dussoix, D. and Arber, W. (1962) Host specificity of DNA produced by *Escherichia coli*. II. Control over acceptance of DNA from infecting phage lambda. *J. Mol. Biol.*, **5**, 37–49.

24. Hattman, S. (1964) The Control of Host-Induced Modification by phage P1. *Virology*, **23**, 270–271.
25. Arber, W. and Morse, M.L. (1965) Host Specificity of DNA Produced by *Escherichia coli*. VI. Effects on Bacterial Conjugation. *Genetics*, **51**, 137–148.
26. Dussoix, D. and Arber, W. (1965) Host Specificity of DNA Produced by *Escherichia coli* IV. Host Specificity of Infectious DNA from Bacteriophage Lambda. *J. Mol. Biol.*, **11**, 238–246.
27. Stacey, K.A. (1965) Intracellular modification of nucleic acids. *Br. Med. Bull.*, **21**, 211–216.
28. Boyer, H. (1971) DNA restriction and modification mechanisms in bacteria. *Annu. Rev. Microbiol.*
29. Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.Kh., Dryden, D.T., Dybvig, K. et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
30. Wilson, G.G. and Murray, N.E. (1991) Restriction and modification systems. *Annu. Rev. Genet.*, **25**, 585–627.
31. Loenen, W.A.M. and Raleigh, E.A. (2014) The other face of restriction: modification dependent enzymes. *Nucleic Acids Res.*, **42**, 56–69.
32. Meisel, A., Bickle, T.A., Kruger, D.H. and Schroeder, C (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature*. **355**, 467–469.
33. Miesel, A., Mackeldanz, P., Bickle, T.A., Kruger, D.H. and Schroeder, C (1995) Type III restriction enzymes translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. *EMBO J.*, **14**, 2958–2966.
34. Luria, S.E. and Human, M.L. (1952) A Non-hereditary, Host-Induced Variation of Bacterial Viruses. *J. Bacteriol.*, **64**, 557–569.
35. Bertani, G. and Weigle, J.J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.*, **65**, 113–121.
36. Linn, S. and Arber, W. (1968) Host Specificity of DNA produced by *Escherichia coli*, X. In vitro restriction of phage ED replicative form. *Biochemistry*, **59**, 1300–1306.

37. Meselson, M. and Yuan, R. (1968) DNA Restriction Enzyme from *E. coli*. *Nature*, **217**, 1110–1114.
38. Horiuchi, K. and Zinder, D.N. (1972) Cleavage of Bacteriophage f1 DNA by the Restriction Enzyme of *Escherichia coli* B. *PNAS*, **69**, 3220–3224.
39. Smith, H.O. and Wilcox, K.W. (1970) A restriction enzyme from *Haemophilus influenzae*. I. Purification and general properties. *J. Mol. Biol.*, **51**, 379–391.
40. Willey, Joanne M, Linda. S., Christopher. J., Woolverton, Lansing M. Prescott, and J.M.W. (2011) Prescott's Microbiology New York: McGraw-Hill.
41. Arber, W. (1965) Host-controlled modification of bacteriophage. *Annu. Rev. Microbiol.*, **19**, 365–378.
42. Arber, W. and Linn, S. (1969) DNA modification and restriction. *Annu. Rev. Biochem.*, **38**, 467–500.
43. Arber, W. and Wauters-Willems, D. (1970) Host specificity of DNA produced by *Escherichia coli*. XII. The two restriction and modification systems of strain 15T-. *Mol. Gen. Genet.*, **108**, 203–217.
44. Glover, S.W. and Piekarowicz, A. (1972) Host specificity of DNA in *Haemophilus influenzae*: restriction and modification in strain Rd. *Biochem. Biophys. Res. Commun.*, **46**, 1610–1617.
45. Piekarowicz, A. and Kalinowska, J. (1974) Host specificity of DNA in *Haemophilus influenzae*: similarity between host-specificity types of *Haemophilus influenzae* Re and Rf. *J. Gen. Microbiol.*, **81**, 405–411.
46. Lark, C. and Arber, W. (1970) Host specificity of DNA produced by *Escherichia coli*. 13. Breakdown of cellular DNA upon growth in ethionine of strains with r<sup>+</sup><sub>15</sub>, r<sup>+</sup><sub>P1</sub> or r<sup>+</sup><sub>N3</sub> restriction phenotypes. *J. Mol. Biol.*, **52**, 337–348.
47. Kauc, L. and Piekarowicz, A. (1978) Purification and properties of a new restriction endonuclease from *Haemophilus influenzae* Rf. *Eur. J. Biochem.*, **92**, 417–426.
48. Piekarowicz, A. and Brzeziński, R. (1980) Cleavage and methylation of DNA by the restriction endonuclease HinIII isolated from *Haemophilus influenzae* Rf. *J. Mol. Biol.*, **144**, 415–429.

49. Mural, R.J., Chesney, R.H., Vapnek, D., Kropf, M.M. and Scott, J.R. (1979) Isolation and characterization of cloned fragments of bacteriophage P1 DNA. *Virology*, **93**, 387–397.
50. Heilmann, H., Burkhardt, H.J., Puhler, A. and Reeve, J.N. (1980) Transposon mutagenesis of the gene encoding the bacteriophage P1 restriction endonuclease. Co-linearity of the gene and gene product. *J. Mol. Biol.*, **144**, 387–396.
51. Lida, S., Meyer, J., Baechi, B., Stalhammar-carlemalm, M., Schrickel, S., Bickle, T.A. and Arber, W. (1983) DNA Restriction-Modification Genes of Phage P1 and Plasmid p15B. *J. Mol. Biol.*, **165**, 1–18.
52. Ahmad, I., Krishnamurthy, V. and Rao, D.N. (1995) DNA recognition by the EcoP15I and EcoPII modification methyltransferases. *Gene*, **157**, 143–147.
53. Janscak, P., Sandmeier, U., Szczelkun, M.D. and Bickle, T.A. (2001) Subunit Assembly and Mode of DNA Cleavage of the type III restriction endonucleases EcoP1I and EcoP15I. *J. Mol. Biol.*, **306**, 417–431.
54. Wyszomirski, K.H., Curth, U., Mackeldanz, P., Schutkowski, M., Kru, D.H. and Mo, E. (2011) Type III restriction endonuclease EcoP15I is a heterotrimeric complex containing one Res subunit with several DNA-binding regions and ATPase activity. *Nucleic Acids Res.*, **40**, 3610-3622.
55. Gupta, Y.K., Yang, L., Chan, S., Samuelson, J.C., Xu, S. and Aggarwal, A.K. (2012) Structural Insights into the Assembly and Shape of type III restriction modification (R-M) EcoP15I complex by Small angle X-ray scattering. *J. Mol. Biol.*, **420**, 261-268.
56. Butterer, A., Pernstich, C., Smith, R.M., Sobott, F., Szczelkun, M.D. and Toth, J. (2014) Type III restriction endonucleases are heterotrimeric: comprising one helicase subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Res.*, **42**, 5139–50.
57. Gupta, Y.K., Chan, S.H., Xu, S.Y., and Aggarwal, A.K. (2015) Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat Commun.*, **6**, 7363.
58. Wagenfuhr, K., Pieper, S., Mackeldanz, P., Linscheid, M., Kruger, D.H. and Reuter, M. (2007) Structural domains in the type III restriction endonuclease

- EcoP15I: characterization by limited proteolysis, mass spectrometry and insertional mutagenesis. *J. Mol. Biol.*, **366**, 93–102.
59. Cheng, X., Kumar, S., Klimasauskas, S. and Roberts, R.J. (1993) Crystal structure of the HhaI DNA methyltransferase. *Cold Spring Harbor Symp. Quant. Biol.*, **58**, 331–338.
60. Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) Predictive motifs derived from cytosine methyltransferase. *Nucleic Acids Res.*, **17**, 2421–2435.
61. Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J. and Wilson, G.G. (1994) The DNA (cytosine-5) methyltransferase. *Nucleic Acids Res.*, **22**, 1–10.
62. Lauster, R., Trautner, T.A. and Noyer-Weidner, M. (1989) Cytosine-specific type II DNA methyltransferase. A conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.*, **206**, 305–312.
63. Timinskas, A., Butkus, V. and Janulaitis, A. (1995) Sequence motifs characteristic for DNA [ cytosine-N4 ] and DNA [ adenine-N6 ] methyltransferases . Classification of all DNA methyltransferases. *Gene*, **157**, 3–11.
64. Malone, T., Blumenthal, R.M. and Cheng, X. (1995) Structure- guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferase, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.*, **253**, 618–632.
65. Hadi, S.M., Bachi, B., Iida, S. and Bickle, T.A. (1983) DNA restriction-modification enzymes of phage P1 and plasmid p15B. Subunit functions and structural homologies. *J. Mol. Biol.*, **165**, 19–34.
66. Gorbalenya, A.E. and Koonin, E. V (1991) Endonuclease (R) subunits of type-I and type III restriction-modification enzymes contain a helicase-like domain. *FEBS Lett.*, **291**, 277–281.
67. Fairman-williams, M.E., Guenther, U. and Jankowsky, E. (2010) SF1 and SF2 helicases :family matters. *Curr. Opin. Struct. Biol.*, **20**, 313–324.
68. Aravind, L., Makarova, K.S. and Koonin, E. V (2000) Holliday junction resolvases and related nucleases : Identification of new families , phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.*, **28**, 3417–3432.

69. Kovall, R.A. and Matthews, B.W. (1998) Structural functional and evolutionary relationships between lambda-exonuclease and the type II restriction endonuclease. *PNAS*, **95**, 7893–7897.
70. Kovall, R. and Matthews, B.W. (1997) Toroidal structure of lambda-exonuclease. *Science*, **277**, 1824–1827.
71. Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. and Wilson, Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. and Wilson, K.S. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.*, **12**, 1781–1795.
72. Galburt, E.A. and Stoddard, B.L. (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry*, **41**, 13851–13860.
73. Horton, N.C. and Perona, J.J. (2004) DNA cleavage by EcoRV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry*, **43**, 6841–685.
74. Brockes, J.P. (1973) The Deoxyribonucleic Acid-Modification Enzyme of Bacteriophage P1. *Biochem. J.*, **133**, 629–633.
75. Hattman, S., Brooks, J.E. and Masurekar, M. (1978) Sequence specificity of the P1 modification methylase (M.Eco P1) and the DNA methylase (M.Eco dam) controlled by the *Escherichia coli* dam gene. *J. Mol. Biol.*, **126**, 367–380.
76. Piekarowicz, A., Bickle, T.A., Shepherd, J.C.W. and Ineichen, K. (1981) The DNA sequence recognised by the HinfIII restriction endonuclease. *J. Mol. Biol.*, **146**, 167–172.
77. Adler, S.P. and Nathans, D. (1973) Studies of SV 40 DNA. V. Conversion of circular to linear SV 40 DNA by restriction endonuclease from *Escherichia coli* B. *Biochim. Biophys. Acta*, **299**, 177–188.78.
78. Toth, J., Bollins, J., and Szczelkun, M.D. (2015) Re-evaluating the kinetics of ATP hydrolysis during initiation of DNA sliding by type III restriction enzymes. *Nucleic Acids Res.*, **43**, 10870-10881.
79. Rao, D.N., Saha, S. and Krishnamurthy, V. (2000) ATP-dependent restriction enzymes. *Nucleic Acid Res., Mol. Biol.*, **64**, 1–63.



80. Dryden, D.T.F., Murray, N.E. and Rao, D.N. (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res.*, **29**, 3728–3741.
81. Saha, S. and Rao, D.N. (1995) ATP Hydrolysis is Required for DNA Cleavage by EcoPI restriction enzyme. *J. Mol. Biol.*, **247**, 559–567.
82. Sears, A., Peakman, L.J., Wilson, G.G. and Szczelkun, M.D. (2005) Characterization of the Type III restriction endonuclease PstII from *Providencia stuartii*. *Nucleic Acids Res.*, **33**, 4775–4787.
83. Aelst, K. Van, Toth, J., Ramanathan, S.P., Schwarz, F.W., Seidel, R. and D, M. (2010) Type III restriction enzymes cleave DNA by long-range interaction between sites in both head-to-head and tail-to-tail inverted repeat. *PNAS*, **107**, 9123–9128.
84. Mucke, M., Reich, S., Moencke-Buchner, E., Reuter, M. and Kruger, D.H. (2001) DNA cleavage by type III restriction-modification enzyme EcoP15I is independent of spacer distance between two head to head oriented recognition sites. *J. Mol. Biol.*, **312**, 687-698.
85. Peakman, L.J., Antognozzi, M., Bickle, T.A., Janscak, P. and Szczelkun, M.D. (2003) S-adenosyl methionine prevents promiscuous DNA cleavage by the EcoP1I type III restriction enzyme. *J. Mol. Biol.*, **333**, 321-335.
86. Raghavendra, N.K. and Rao, D.N. (2004) Unidirectional translocation from recognition site and a necessary interaction with DNA end for cleavage by type III restriction enzyme. *Nucleic Acids Res.*, **32**, 5703-5711.
87. Raghavendra, N.K. and Rao, D.N. (2005). Exogenous AdoMet and its analogue sinefungin differentially influence DNA cleavage by R.EcoP15I- Usefulness in SAGE. *Biochem. Biophys. Res. Commun.*, **334**, 803-811.
88. Moencke-Buchner, E., Rothenberg, M., Reich, S., Wagenfuhr, K., Matsumura, H., Terauchi, R., Kruger, D.H., and Reuter, M. (2009) Functional characterization and modulation of DNA cleavage efficiency of type III restriction endonuclease EcoP15I in its interaction with two sites in the DNA target. *J. Mol. Biol.*, **387**, 1309-1319.
89. Peakman, L.J. and Szczelkun, M.D. (2009). S-Adenosyl homocysteine and DNA ends stimulate promiscuous nuclease activities in the type III restriction endonuclease EcoP1. *Nucleic Acids Res.*, **37**, 3934-3945.

90. van Ham, S. M., van Alphen, L., Mooi, F. R. and vanPutten, J. P. Phase variation of *Haemophilus influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. *Cell* **73**, 1187–1196 (1993).
91. Weiser, J. N., Williams, A. & Moxon, E. R. Phasevariable lipopolysaccharide structures enhance the invasive capacity of *Haemophilus influenzae*. *Infect.Immun.* **58**, 3455–3457 (1990).
92. Ando, T., Xu, Q., Torres, M., Kusugami, K., Israel, D. A., and Blaser, M. J. (2000) Restriction-modification system differences in *Helicobacter pylori* are a barrier to interstrain plasmid transfer. *Mol. Microbiology.*, **37**, 1052-1065.
93. Hallet, B. (2001) Playing Dr Jekyll and Mr Hyde: combined mechanisms of phase variation in bacteria. *Curr Opin Microbiology.*, **4**, 570–581.
94. Srikhanta, Y. N., Fox, K. L. & Jennings, M. P. (2010). The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nat. Rev. Microbiology.*, **8**, 196–206
95. De Bolle, X., Bayliss, C. D., Field, D., Van de Ven, T., Saunders, N. J., Hood, D. W. and Moxon, E. R. (2000) The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. *Mol. Microbiol.*, **35**, 211–222 (2000).
96. Levinson, G. and Gutman, G. A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Bio. Evol.*, **4**(3), 203-221
97. Donahue, J. P., Israel, D. A., Peek, R. M., Blaser, M. J. & Miller, G. G. (2000). Overcoming the restriction barrier to plasmid transformation of *Helicobacter pylori*. *Mol. Microbiology.*, **37**, 1066–1074.

## CHAPTER 2

**Single-site DNA cleavage by Type III RM enzymes need a site bound and a *trans* acting enzyme that are ATPase activated**

## Chapter 2

### **Single-site DNA cleavage by Type III restriction endonuclease requires a site-bound and a trans-acting enzyme that are ATPase-activated**

#### **2.1. Introduction**

Long-range communication between two distal DNA sites, in *cis*, regulates fundamental cellular functions of transcription regulation and DNA metabolism. The communication is mediated by active and directional movement (translocation) or 1D diffusion along DNA, or looping of DNA. ATPase motors power translocation of proteins along DNA, hydrolyzing one nucleotide per base pair. These motors can be categorized as helicases that unwind double-stranded DNA and actively translocate on single-stranded DNA, and translocases that translocate on double-stranded DNA. An emerging theme is the presence of ATPase motors that function as molecular-switches, where a single or a few rounds of ATP hydrolysis is sufficient to promote one dimensional (1D) diffusion of protein along the DNA. (1, 2, 3, 4).

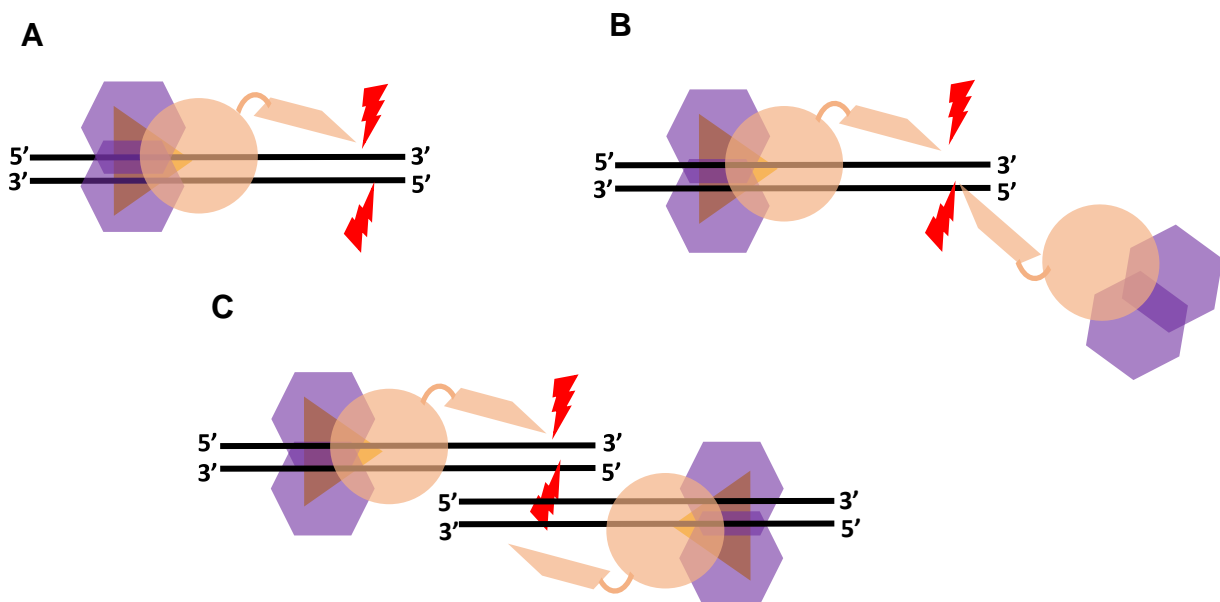
The phenomenon of switch mediated 1D diffusion has been proposed through single molecule studies on the MutS-MutL family of mismatch repair proteins (3, 4). A mechanism involving 1D diffusion has also been proposed for the endonucleolytic activity of Type III restriction-modification (RM) enzymes EcoP1I and EcoP15I (2). Both these enzyme systems use diffusion along DNA to allow communication between distant sites of action. MutS bound to ATP diffuses along the DNA between the site of mismatch and the site of nicking of the unmethylated strand (5). ATP-dependent enzyme diffusion facilitates DNA cleavage by Type III RM enzymes between distant recognition sites. In both these systems, though the presence of ATP is essential, its role is not fully understood. Here we report a study that unravels the role of ATP at various stages of the nucleolytic activity of Type III RM enzymes.

Type III RM enzymes are bacterial defense systems that protect the host from invading foreign DNA by nucleolytically cleaving them at specific recognition sites. The RM enzyme is composed of two subunits of methyltransferase (Mod) and a restriction subunit (Res) having an ATPase and an endonuclease domain. The dimeric Mod or the

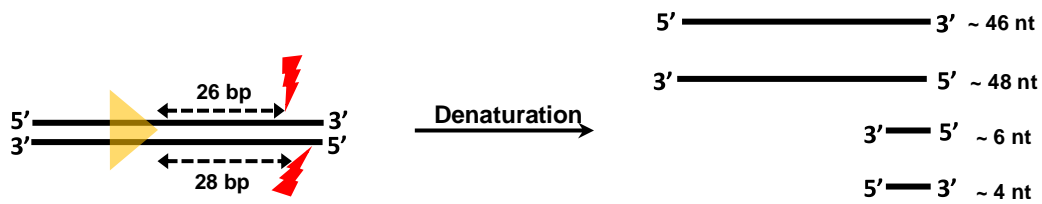
trimeric Res-Mod complex catalyzes methylation of the recognition site. Endonucleolytic cleavage of DNA by Type III RM enzyme requires the presence of at least two unmethylated recognition sites in inverted orientation (6), which will henceforth be referred to as two-site cleavage. The recognition site is asymmetric and as a consequence only one strand is methylated. Methylation of genomic DNA protects it from the restriction activity of the enzyme. A battery of biophysical, biochemical and single-molecule studies carried out over the last decade on the Type III RM enzymes EcoP1I and EcoP15I has led to different models of DNA cleavage (2, 7). According to a recent model derived from magnetic tweezers combined with TIRF microscopy, a Type III RM enzyme, on binding to its recognition site and in presence of ATP, undergoes a change in its conformational state, which promotes its 1D diffusion along the DNA (2). An alternate model based on atomic force microscopy, proposes the convergence of the two enzymes through DNA looping and limited ATP hydrolysis for translocation (7, 8). DNA cleavage without looping has been observed in magnetic tweezers assay (9). Consequently, the requirement and role of DNA looping for cleavage needs further studies.

According to the 1D diffusion model, encounter of the diffusing enzyme with another enzyme bound to a recognition site brings together a nuclease domain each from the respective enzymes. The two nucleases catalyze the nicking of alternate single-strands resulting in a double-strand break at a fixed distance downstream of the recognition site of the stationary enzyme. Independent of the spacing between the recognition sites, which can be as large as 3000 bp, the nucleolytic reaction requires the hydrolysis of about ten ATP molecules (10, 11). The hydrolysis of ATP by Type III RM enzyme alters its conformational state to diffusion-competent (2). However, the requirement of the nucleotide for DNA cleavage *per se* is not fully clear. For example, it is not known if the ATPase of the stationary enzyme, like that of the diffusing enzyme, has to hydrolyse ATP for DNA cleavage. Although the canonical substrates for Type III RM enzymes have at least two recognition sites in inverted orientation, there have been reports of Type III RM enzymes cleaving linear and circular DNA having only a single recognition site in presence of ATP (12, 13, 14, 15, 16, 17, and 18). Earlier reports indicated a circular plasmid having a single recognition site was cleaved better than the linearized plasmid (17). Cleavage of the linearized plasmid could be improved on addition

of sinefungin (17). The site of this cleavage has been mapped to 25-27 bp downstream of the recognition site of EcoP15I and 26-28 bp downstream of EcoP1I recognition site (19). This is identical to that mapped for cleaved canonical substrates (20). While carrying out biochemical studies we also noticed that with EcoP1I and EcoP15I were able to cleave a DNA substrate containing a single recognition site (21, and this study). However, the mechanism of the non-canonical ATP-dependent single-site cleavage is unknown. An endonucleolytic reaction entails hydrolysis of a phosphodiester bond on both strands. A Type III RM enzyme, which has only one nuclease domain, can catalyse single-site cleavage by binding to the recognition site and i) nicking the two strands, akin to the monomeric restriction endonuclease BfiI (Figure 2.1A) (22) or ii) nicking one strand while recruiting another enzyme to nick the other strand, like the endonuclease FokI (Figure 2.1B) (23) and the third possibility that two protein-DNA complex can cooperate with each other to perform dsDNA break (Figure 2.1C).



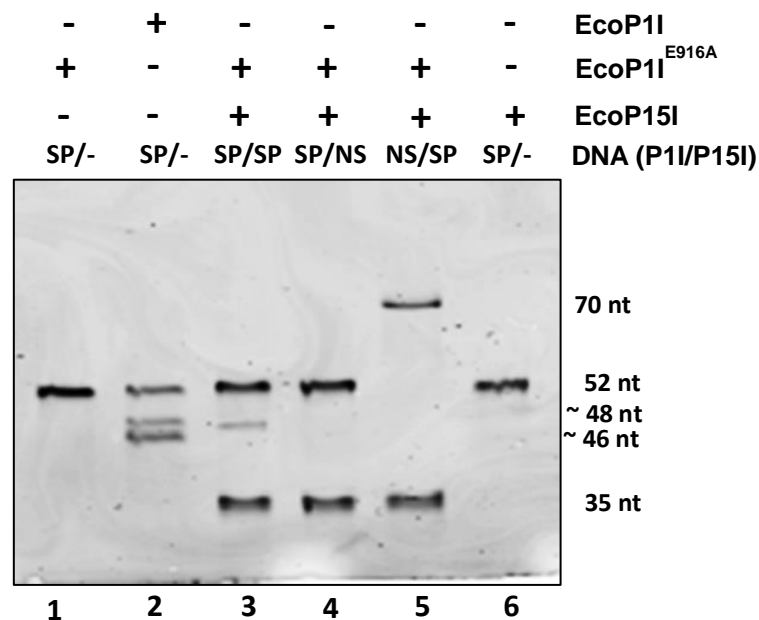
**Figure 2.1. Cartoons illustrating the possible models for single-site cleavage.** (A) Model 1: a single Type III RM enzyme (Mod colored purple and Res colored beige) bound to its recognition site (yellow arrowhead) performs a double-strand DNA break by nicking both the strands. (B) Model 2: an enzyme bound to its recognition site cooperates with a free enzyme from solution to cleave the DNA. (C) Model 3: two enzyme-DNA complexes bound to their recognition site cooperate with each other to perform dsDNA break.



**Figure 2.2.** A schematic diagram illustrating the products of a hypothetical single-site cleavage as single-strands that are observed when analyzed on a denaturing urea-formamide PAGE using ethidium bromide stain. The position of the cleavage sites from the upstream end is mentioned. These positions are based on previous mapping of single-site cleavage.

EcoP1I and EcoP15I have been shown to cooperate to cleave DNA having one each of their respective recognition sites oriented head-to-head (24). Using this knowledge a “heterologous cooperation assay” was designed to find if the single-site cleavage is a *cis* or a *trans* activity. The heterologous cooperation assay involved use of two separate short pieces of DNA each with a single recognition site for EcoP1I or EcoP15I, respectively. Reaction was initiated by addition of ATP to a mix containing EcoP15I and EcoP1I<sup>E916A</sup> (a nuclease dead protein where a glutamate amino acid in the endonuclease motif PD(x)<sub>17</sub>ExK is mutated to alanine) bound to their respective single-site DNA. EcoP1I cleaved 15/32\_P1 (see Figure 2.4 for details) resulting in two single strands of 46 and 48 bases (Figure 2.2). As above, the shorter fragments could not be visualized on the denaturing gel using ethidium bromide. Despite the use of the nuclease-dead EcoP1I<sup>E916A</sup>, the DNA containing its recognition site was nicked, but not cleaved, in presence of EcoP15I bound to its 35 bp long specific DNA 5/24\_P15 (Figure 2.3, Lane 3). However, EcoP15I did not nick EcoP1I specific DNA in the absence of the specific DNA of EcoP15I (Figure 2.3, Lanes 4). Also, EcoP15I in presence of 5/24\_P15 and EcoP1I nicked only the specific DNA of EcoP1I and not a non-specific one (Figure 2.3, Lanes 3 & 5). Consequently, we concluded that the nicking of EcoP1I specific oligomer was a result of cooperation between EcoP1I<sup>E916A</sup> and EcoP15I bound to their specific DNA. It should be noted that the experiment does not rule out the possibility that the DNA bound state is only required for activation of the enzyme to a reaction competent

conformational state, and not essential during the reaction. However, the fixed distance of cleavage from the recognition site of the *cis*-bound enzyme, i.e. EcoP1I<sup>E916A</sup> in this case, indicates that this enzyme is site bound during catalysis. In comparison to the efficiency of single-site cleavage by a single enzyme, EcoP1I or EcoP15I (Figure 2.3 Lane 3), the efficiency of nicking noticed in case of heterologous cooperation involving both EcoP1I and EcoP15I was much less (Figure 2.3 Lane 3), suggesting that the heterologous cooperation is not as optimized as homologous cooperation (21).



**Figure 2.3. A heterologous cooperation assay of EcoP1I and EcoP15I.** The DNA in the gel were unlabeled and stained using ethidium bromide for visualization. SP/SP represents presence of specific DNA of EcoP1I and EcoP15I, SP/- represents presence of EcoP1I specific DNA only, SP/NS represents the presence of EcoP1I specific DNA and non-specific DNA of EcoP15I, and so on. Lanes 1) 15/32\_P1 + EcoP1I<sup>E916A</sup>, 2) 15/32\_P1 + EcoP1I, 3) 15/32\_P1 + EcoP1I<sup>E916A</sup> and 5/24\_P15 + EcoP15I, 4) 15/32\_P1 + EcoP1I<sup>E916A</sup> and NS\_P15 + EcoP15I, 5) NS\_P1 + EcoP1I<sup>E916A</sup> and 5/24\_P15 + EcoP15I, 6) 15/32\_P1 + EcoP15I. 300 nM of DNA and 250 nM of enzymes were used.

Our heterologous cooperation assay indicates that single site DNA cleavage happens by the cooperation of two protein DNA complexes in *trans*, but there are still important questions which remained unanswered.



1. Can a single enzyme bound to a recognition site also perform a dsDNA break?
2. What is the role of ATP in such a DNA cleavage?
3. If single site DNA cleavage happens by the cooperation of two protein-DNA complexes in *trans*, which strand does the *cis* enzyme cleave and which strand does *trans* enzyme cleave?

In this chapter using biochemical assays I tried to address these questions which revealed a quite interesting mechanism how these enzymes cleave the single site DNA cleavage.

## 2.2 Materials and Methods

### 2.2.1 Cloning

*EcoP15I* operon in BZZ vector was provided as a kind gift from Prof. D.N. Rao, IISC, Bangalore, India. *EcoP15I* operon, was sub cloned into pHis17 with C-terminal 6xHistag using For1 5'-GTTTAACTTTAAGAAGGAGATATACATATGAAAAAAGAAACGATTTTT TCCG-3' and Rev1 5'-CTTTTAATGATGATGATGATGATGGGATCCTTATGGTAATG CGCTCTTGATTAAC-3'. It was noticed that there is an NdeI restriction site at 321bp downstream of the *EcoP15I* gene. A reverse primer (Rev2) was used to silence the NdeI site without changing the open reading frame. The resulting operon was double digested with NdeI and BamHI restriction enzymes and ligated into pHis17 (25) vector using T4 DNA ligase. The ligated mixture was electroporated into *E.coli* NEB turbo Electrocompetant cells using Gene Pulsar Xcell. The positive clones were confirmed only after complete sequencing of the gene. *EcoP15I/pRSF* was generated with an N-terminal 6xHistag using, For2 5' AACTTTAAGAAGGAGATATACCATGGATGAAAAAAGAAAC GATTTTTCCG-3' and Rev 4 5'-CGCAGCAGCGGTTTCTTTACCGAGCTCTTAGTTC CTTACCACTAAATCTAAC-3'. The amplified insert was double digested with NcoI and XhoI and ligated into pRSF vector using T4 DNA ligase. The clones were confirmed only after the complete sequencing of the gene. Table 2.1 lists the primers used for the cloning and mutagenesis for *EcoP15I*.

**Table 2.1:** Primer used for the cloning and mutagenesis of EcoP15I and EcoP1I

| Name            | Sequence (5'- >3')                                           |
|-----------------|--------------------------------------------------------------|
| For 1           | GTTTAACTTTAAGAAGGAGATATACATATGAAAAAGAAACGAT<br>TTTT TCCG     |
| Rev 1           | CTTTTAATGATGATGATGATGATGGGATCCTTATGGTAATGCG<br>CTCTTGATTAAC  |
| Rev 2           | CATAAGCATTAAACCATGTGCTTCAATACTTC                             |
| For 2           | AACTTTAAGAAGGAGATATACCATGGATGAAAAAGAAACGAT<br>TTTTTCCG       |
| Rev 4           | CGCAGCAGCGGTTTCTTTACCGAGCTCTTAGTTCCTTACCACT<br>AAATCTAAC     |
| EcP15FD898<br>A | GCTGGTGGATACACTTACTCACCCGCTTTTGCTTATGTTGTAA<br>AAACAGCGGAAGG |
| EcP15RD898<br>A | CCTTCCGCTGTTTTTACAACATAAGCAAAGCGGGTGAGTAAG<br>TGTATCCACCAGC  |
| ForK90R         | GATGTTTCTATGGAGACAGGTACAGGCAGAACGTACACC                      |
| RevK90R         | GGTGTACGTTCTGCCTGTACCTGTCTCCATAGAAACATC                      |
| P1FE227A        | CATTATCATTGATGCTCCACATAAATTCCTACTGG                          |
| P1RE227A        | CCAGTAGGAAATTTATGTGGAGCATCAATGATAATG                         |
| P15FE227A       | CATTTATCATCATTGACGCTCCACATAGATTCCTACTG                       |
| P1RE227A        | CAGTAGGAAATCTATGTGGAGCGTCAATGATGATAAATG                      |

### 2.2.2 Sequencing of the EcoP15I operon

The positive clones were sequenced using Sanger sequencing method. One of the limitations of sequencing using Sanger's method is that the read out from one reaction was 800-900 bp. The size of EcoP15I is 4853 bp making it impossible to sequence the

complete operon using T7 promoter and T7 terminator primers. To sequence the complete operon five other primers were used which has sequence complementary to the internal sequence of *EcoP15I*. The sequence of internal primers is as listed in the Table 2.2.

**Table 2.2:** Primer used for the complete sequencing of *EcoP15I*

| Sr. No | Name          | Sequence (5'- >3')    |
|--------|---------------|-----------------------|
| 1      | T7 Promoter   | TAATACGACTCACTATAGGG  |
| 2      | T7 Terminator | GCTAGTTATTGCTCAGCGG   |
| 3      | F2            | CAAGCAGTCATAGTGCATGG  |
| 4      | F3            | GACTTTTTTGGCTGGCTCTGG |
| 5      | F4            | GAAAGCAGGTGTCGATGCGG  |
| 6      | F5            | GTAAAGGTATTGATGCG     |

### 2.2.3 Mutagenesis of *EcoP1I* and *EcoP15I*

Nuclease dead mutants *EcoP15I*<sup>D898A</sup> and Walker A (Motif I) and Walker B (Motif II) mutants *EcoP1I*<sup>K90R</sup> *EcoP1I*<sup>E227A</sup> *EcoP15I*<sup>E227A</sup> were generated using quick-change site directed mutagenesis. Positive clones were confirmed by sequencing the entire gene. The list of primers used for the generation for mutagenesis is listed in Table 2.1.

### 2.2.4 Protein purification of *EcoP1I* and *EcoP15I*

The recombinant clones were first checked for expression in 5 mL 1X LB culture before starting a large scale purification. After confirming the expression of desired protein a large scale purification was started. *E. coli* BL21(AI) cells were transformed with wild type

EcoP15I operon in recombinant vector pHis17 (25) with a 6xHis tag at the C-terminus of the Res subunit. A 6 L culture was grown at 37°C till OD<sub>600</sub> reached 0.4, after which culture was shifted to 18°C. The culture was induced with 2 g/L L-arabinose and kept further for 10 h with constant shaking. The culture was pelleted by centrifugation using Avanti J-26XP (Beckman Coulter Life Sciences) for 20 min at 4°C. The pelleted culture was resuspended in lysis buffer (50 mM Tris-HCl pH 8, 500 mM NaCl, 10% glycerol, 10 mM MgCl<sub>2</sub>). The re-suspended cell pellet were lysed by sonication. A sonication cycle was carried out for 3 mins with 1 second on and 3 seconds off. For complete lysis of cells the sonication cycle was repeated twice. To prevent generation of heat during sonication, the lysate was kept on ice during the entire process. The lysate was spun at 37000 rpm for 1 h at 4°C using Optima XE ultracentrifuge (Beckman Coulter). As the first purification step, EcoP15I was purified by affinity chromatography using a 5 mL Ni-NTA column (GE Healthcare). Fractions containing EcoP15I were pooled and dialyzed against B-50 buffer (50 mM Tris-Cl pH 8, 50 mM NaCl, 1 mM EDTA, 1 mM DTT) for 6 h at 4°C. The dialyzed protein was further purified using an 8 mL Mono Q 10/100 GL column (GE Healthcare). To check the homogeneity of EcoP1I, size exclusion chromatography (SEC) using Superdex 200 10/300 (GE Healthcare) was carried out as a last purification step. EcoP15I obtained after SEC was pooled, concentrated and stored in buffer 10 mM Tris-HCl pH 8, 100 mM KCl, and 1 mM DTT at -80°C until further use.

*Escherichia coli* BL21(DE3) cells were transformed with wild type EcoP1I operon in recombinant vector pRSF with a 6xHis tag at the N-terminus of the Mod subunit. An 8 L culture was grown at 37°C till the OD<sub>600</sub> reached 0.4 after which the culture was shifted to a lower temperature of 18°C. EcoP1I was purified using the protocol mentioned previously (21).

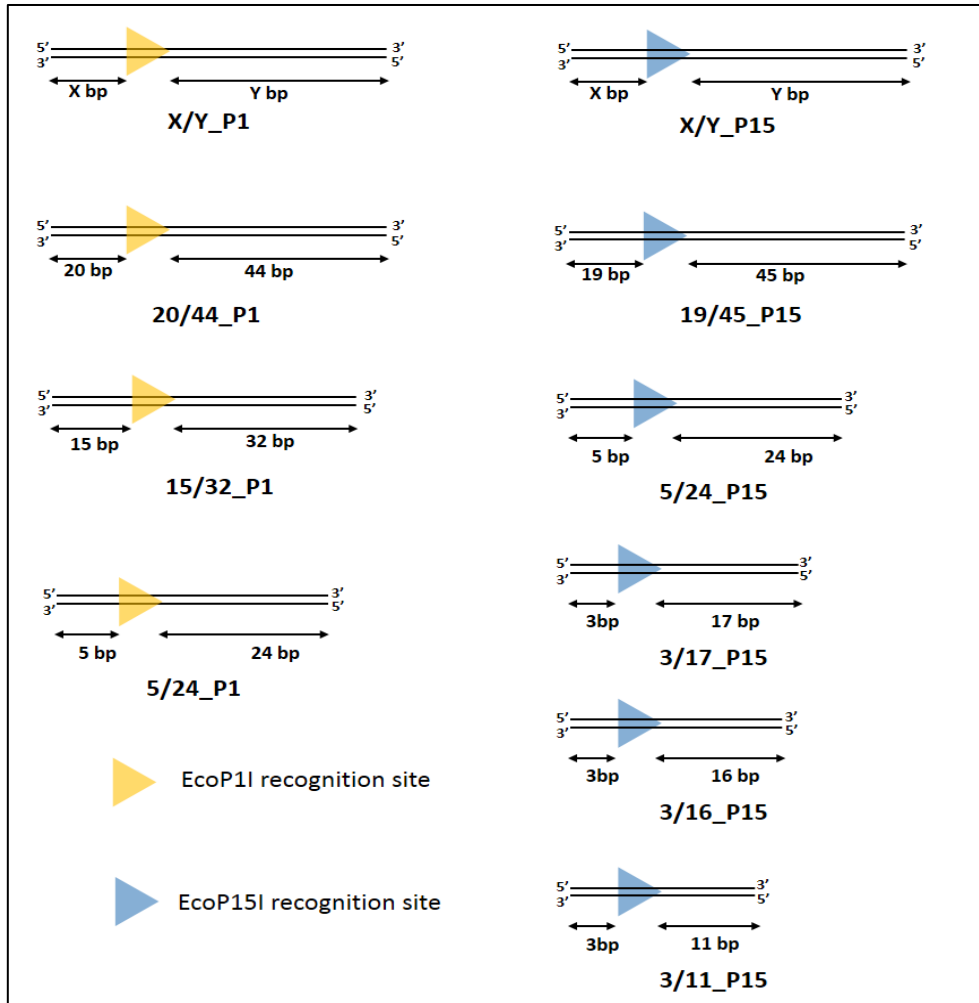
### 2.2.5 DNA substrates

DNA substrates were purchased from Integrated DNA Technologies and Sigma Aldrich. The two complementary DNA strands were annealed using temperature gradient from 95°C - 25°C. The annealed DNA substrates were further purified using an 8 mL Mono Q 10/100 GL column. The resulting duplex was washed thoroughly with Milli-Q water, and concentrated using Vivaspin concentrator (MWCO 3 kDa; GE Healthcare). The

concentrated DNA was stored at  $-30^{\circ}\text{C}$  until further use. The various oligonucleotides used in biochemical studies are mentioned in Supplementary Table 2.3.

**Table 2.3:** Oligomers used in biochemical assays for EcoP15I and EcoP1I

| Name        | Sequence (5' - >3')                                                                 |
|-------------|-------------------------------------------------------------------------------------|
| 15/32_P1_T  | GATCGTAGACGTA <u>CTAGACCT</u> ATCCTGTATGCTACGTATTC<br>GTATCGTGAGC                   |
| 15/32_P1_B  | GCTCACGATACGAATACGTAGCATA <u>CAGGATAGGTCT</u> AGTA<br>CGTCTACGATC                   |
| 19/45_P15_T | TGACCATGATTACGCCAAT <u>CAGCAGCT</u> CCAGGTCGTACCTC<br>CAGCTACCAATCCCCGGGTACCGATCTCG |
| 19/45_P15_B | CGAGATCGGTACCCGGGGATTGGTAGCTGGAGGTACGACC<br>TGGAG <u>CTGCTG</u> ATTGGCGTAATCATGGTCA |
| 2/32_P15_T  | CT <u>CAGCAGT</u> ATCCTGTATGCTACGTATTCGTATCGTGAGC                                   |
| 2/32_P15_B  | GCTCACGATACGAATACGTAGCATA <u>CAGGATACTGCTG</u> AG                                   |
| 0/32_P15_T  | <u>CAGCAGT</u> ATCCTGTATGCTACGTATTCGTATCGTGAGC                                      |
| 0/32_P15_B  | GCTCACGATACGAATACGTAGCATA <u>CAGGATACTGCTG</u>                                      |
| 5/25_P15_T  | GTA <u>CTCAGCAGT</u> ATCCTGTATGCTACGTATTCGTA                                        |
| 5/25_P15_B  | TACGAATACGTAGCATA <u>CAGGATACTGCTG</u> AGTAC                                        |



**Figure 2.4: Cartoon representation of oligo's used for biochemical assay.** X represents the length of the DNA upstream to the recognition site and Y represents the length of the DNA downstream to recognition site. The recognition sequences of EcoP1I and EcoP15I are marked as yellow and blue triangles respectively.

### 2.2.6 Electromobility Shift Assay

Protein and DNA was mixed in Buffer R (50 mM Tris-HCl pH 8, 50 mM KCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT) at 25°C for 15 mins. After incubation 1 volume of native Stop dye (10 mM Tris-HCl, pH8, 60 mM EDTA, 60% glycerol, 0.03% bromophenol blue, 0.03% xylene cyanol ) was added to the samples. The samples were immediately loaded on a 5% Native PAGE gel pre run at 4°C in 1X TBE. The gel was stained with ethidium bromide and scanned using GeneSnap Scanner.

### 2.2.7 DNA Cleavage Assays

DNA cleavage assays were carried out in Buffer R (50 mM Tris-HCl pH 8, 50 mM KCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT) at 25°C. Protein and DNA were incubated in Buffer R supplemented with 20 µM sinefungin for 45 mins after which the reaction was started by the addition of 1 mM ATP. The reaction was carried out for 45 mins after which the reaction was stopped by the addition of 0.5x stop buffer (10 mM Tris-HCl pH 8, 60 mM EDTA, 60% glycerol, 0.025% SDS, 0.03% xylene cyanol). The samples were kept at 25°C for 20 minutes and then loaded on 18% Native PAGE [18% acrylamide:bisacrylamide (19:1), 1XTBE]. In case of denaturing urea-formamide gel runs, the reaction was stopped with formamide stop dye (95% formamide, 0.025% SDS, 0.5 mM EDTA and 0.03% bromophenol blue). For complete denaturation of DNA, the samples were heated at 99°C for 10 minutes before loading on 18% urea-formamide PAGE gel (18% acrylamide: bisacrylamide, 7 M Urea, and 20% formamide, 1X TBE). The gels were stained with a water solution containing ethidium bromide and scanned using Typhoon TRIO+ variable mode imager.

#### *2.2.8 Labelling of DNA substrate with <sup>32</sup>P at 5'-end*

dsDNA substrate 100 nM was labeled at 5'-end with T4 polynucleotide kinase (New England Biolabs) in presence of (gamma-<sup>32</sup>P ATP) at 37°C for 30 mins. Polynucleotide kinase was heat inactivated at 65°C for 20 mins. The DNA was purified using MicroSpin column (GE Healthcare) and stored at -30°C until further use.

#### *2.2.9 Heterologous cooperation assays*

EcoP1I and EcoP15I were incubated with respective DNAs separately in presence of 20 µM sinefungin at 25°C for 45 mins. After which the sample was mixed and reaction was started by addition 1 mM of ATP. The reaction was further carried out for 45 minutes at 25°C after which the reaction was stopped with 1 volume of formamide stop dye. The samples were heated for 10 mins at 99°C and loaded on 18% urea-formamide PAGE gel. The gel was run at 220 V for 90 mins stained with ethidium bromide and scanned using Typhoon TRIO+ variable mode imager (GE Healthcare)

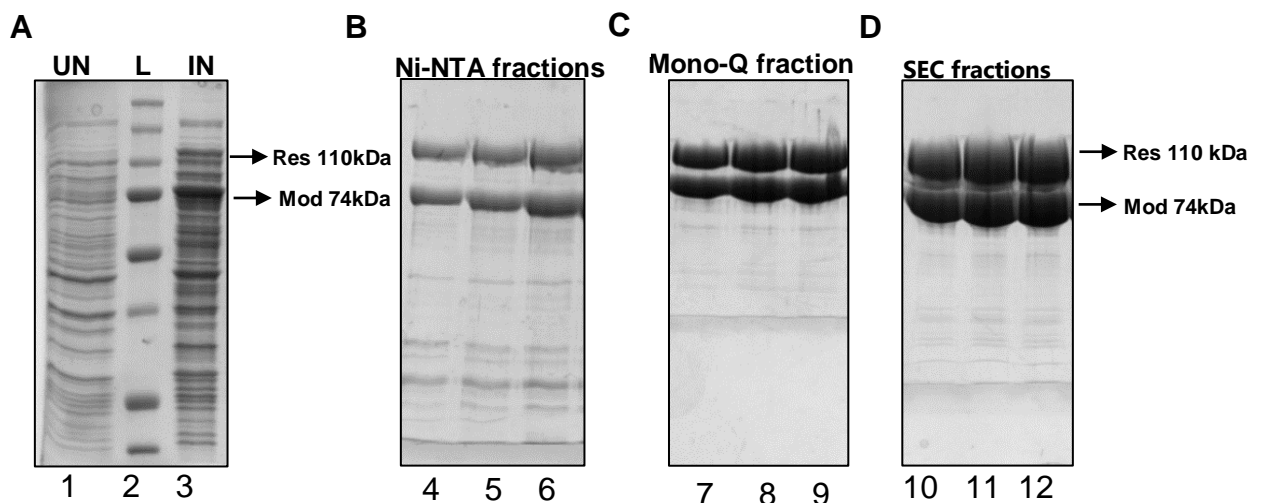
For reactions visualized using radioactive labels, 5 nM of <sup>32</sup>P DNA was incubated with EcoP1I in presence of 20 µM sinefungin for 45 mins at 25°C. Similarly 300 nM of

unlabeled 5/24\_P15 DNA was incubated with 250nM of EcoP15I in presence of 20  $\mu$ M sinefungin. After 45 mins the two protein-DNA complexes were mixed and a reaction was started with 1mM of ATP. The reaction was incubated for 45 mins at 25°C after which the reaction was stopped with equal volume of formamide gel loading dye. The samples were heated at 99°C for 10 mins, and immediately loaded on 18% urea-formamide PAGE gel. The gel was run at 10 W for 90 minutes. The gel was imaged by placing on phosphor imager plate and kept at -80°C for 3 h. The phosphor imager plate was scanned using Typhoon TRIO+ variable mode imager.

## 2.3 Results

### 2.3.1 Purification of EcoP15I

After confirming the expression of protein in soluble fraction (Figure 2.5 A). A 6L culture of EcoP15I<sup>C-His</sup> was also purified first using affinity chromatography (Figure 2.4B). After affinity purification the protein was dialyzed and then loaded on ion-exchange chromatography. After ion-exchange chromatography the fraction containing pure protein was concentrated and loaded onto size exclusion chromatography SEC (Figure 2.5 C). After SEC the protein was >95% pure (Figure 2.5 D).

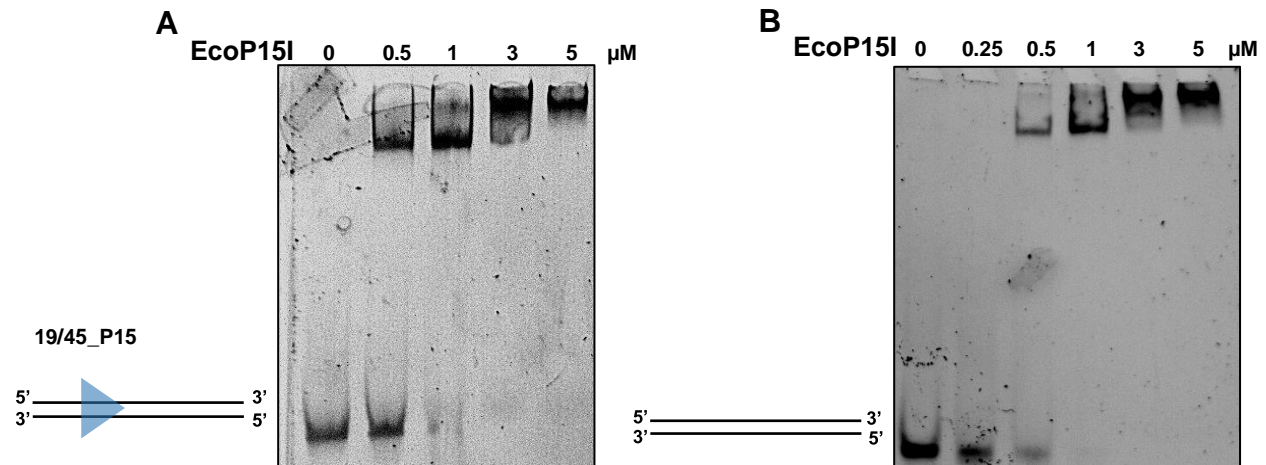


**Figure 2.5. Expression and purification of EcoP15I.** (A) Expression gel of EcoP15I, lane 1 Uninduced (UN), lane 2 ladder (L), lane 3 induced (IN) culture. (B) Ni-NTA fraction of EcoP15I (Lane 4- 6). (C) Mono-Q fraction of EcoP15I (Lane 7-9). (D) Size exclusion chromatography (SEC) fractions of EcoP15I (Lane 10-12).



### 2.3.2. DNA binding assay of *EcoP15I*<sup>D898A</sup> with short DNA substrates

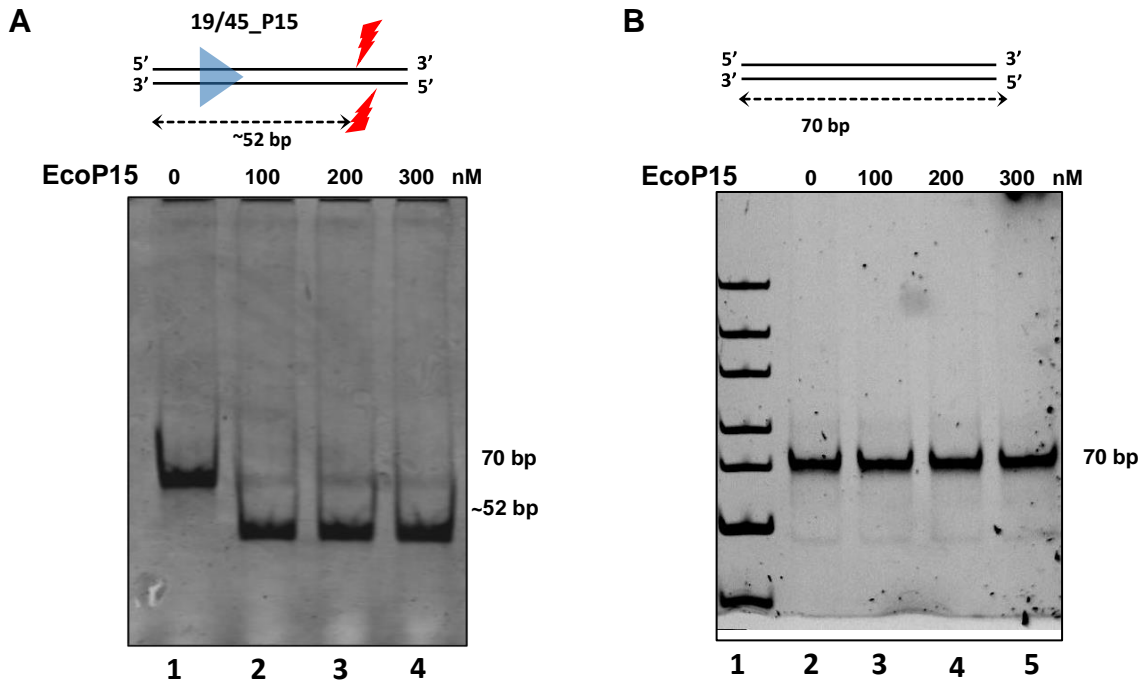
We carried out DNA-binding assays with short DNA substrate. A 19/45\_P15 DNA with a single recognition site of *EcoP15I*<sup>D898A</sup> acts as a specific DNA, while a similar length DNA lacking recognition site of *EcoP15I* acts as a non-specific DNA (Figure 2.6 A & B). In presence of sinefungin and ATP, *EcoP15I*<sup>D898A</sup> was not able to differentiate specific DNA from non-specific DNA, with enzyme binding to both the two DNA's with almost equal affinities.



**Figure 2.6. DNA binding assay of *EcoP15I* with short DNA substrate.** (A) DNA binding assays were carried out with 300nM of 19/45\_P15 DNA with increasing concentration of *EcoP15I*. (B) DNA binding assay with non-specific DNA with increasing concentration of *EcoP15I*.

### 2.3.3 Single-site DNA cleavage by *EcoP15I*

From our DNA binding assays, *EcoP15I* was binding to specific and non-specific DNA with equal efficiency. However, when we used wild type *EcoP15I* instead of *EcoP15I*<sup>D898A</sup>, and resolved the products on Native SDS-PAGE gel, we noticed that *EcoP15I* cleaved the DNA 19/45\_P15 (a 70 bp long DNA with the 6 bp long *EcoP15I* recognition site, CAGCAG, flanked by 19 bp upstream and 45 bp downstream, into ~52 bp and ~18 bp fragments (Figure 2.7 A) (19). The ~18 bp fragment could not be seen as it had run out of the gel. A non-specific DNA of similar length lacking the recognition sequence of *EcoP15I* was not cleaved (Figure 2.7 B). We also noticed that *EcoP11*, a close homologue of *EcoP15I* was also able to cleave the DNA containing a single recognition site (21).

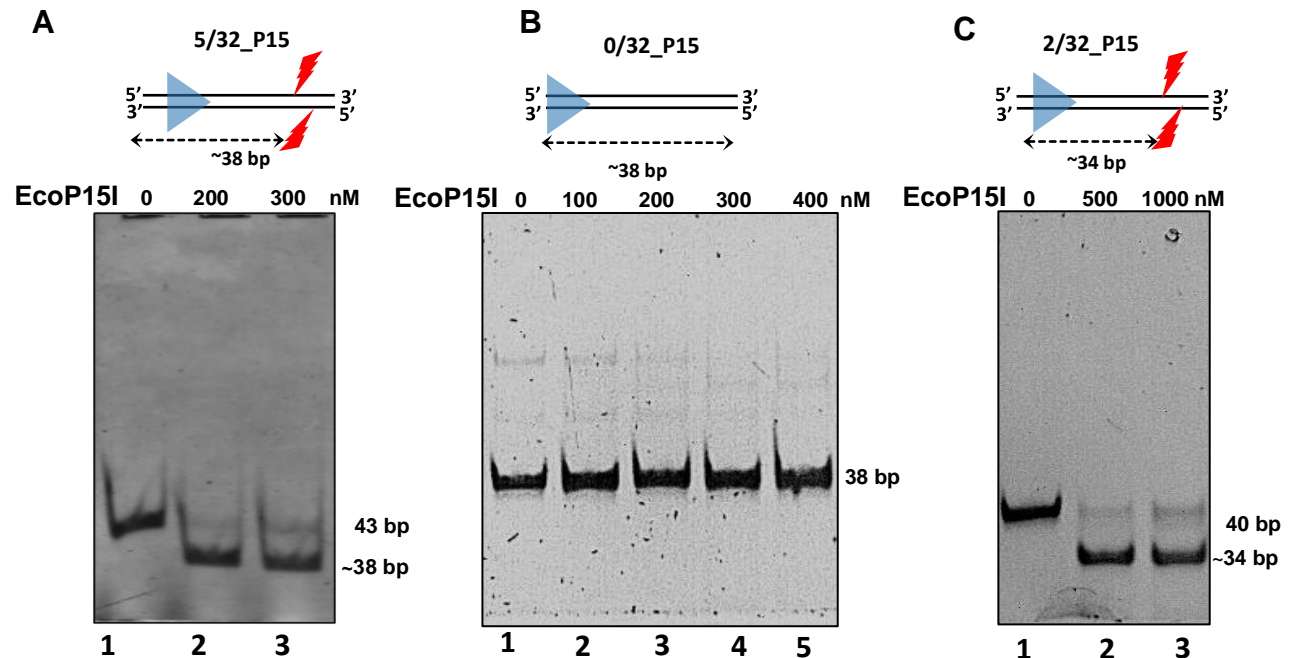


**Figure 2.7. Single-site cleavage by EcoP15I.** (A) 300 nM 19/45\_P15 having a single recognition site of EcoP15I (blue arrowhead) was incubated with increasing concentration of their respective enzymes (the reaction was performed twice and similar results obtained). The reaction was started by the addition of ATP. Red arrows mark the site of nicks, and the position of cut of the top strand from the upstream end is mentioned. Length of DNA substrate and cleaved product are given on the right hand side of the gels respectively. (B) DNA cleavage assay with a DNA substrate lacking the recognition sequence of EcoP15I.

#### 2.3.4 Significance of upstream DNA length flanking recognition site for single-site cleavage

To test the effect of the minimum length of DNA upstream to recognition site for the single site DNA cleavage to happen, we generated three dsDNA substrates, (5/31\_P15, 2/25\_P15, 0/25\_P15) containing 5 , 2 and zero base pairs upstream to the recognition site (Figure 2.7 A, B & C) (See Figure 2.4 for details). We simply incubated EcoP15I with DNA having 0 bp (0/32\_P15) 2 bp (2/32\_P15) and 5 bp (5/31\_P15) upstream of the recognition sequence and monitored their cleavage in presence of ATP. We found that

lack of DNA upstream of the recognition site stopped the cleavage (Figure 2.8 B). Cleavage occurred on addition of two or more base pairs upstream (Figure 2.8 A, C).



**Figure 2.8. Effect of upstream end of recognition sequence on single-site DNA cleavage.** (A) Cleavage assay of 300 nM 5/32\_P15 DNA with increasing concentration of EcoP15I resolved on a native SDS-PAGE gel. (B) DNA cleavage assay of 300 nM 0/32\_P15 DNA with increasing concentration of EcoP15I resolved on a native PAGE gel. (C) DNA cleavage assay of 300 nM 2/32\_P15 DNA with increasing concentration of EcoP15I resolved on a native PAGE gel.

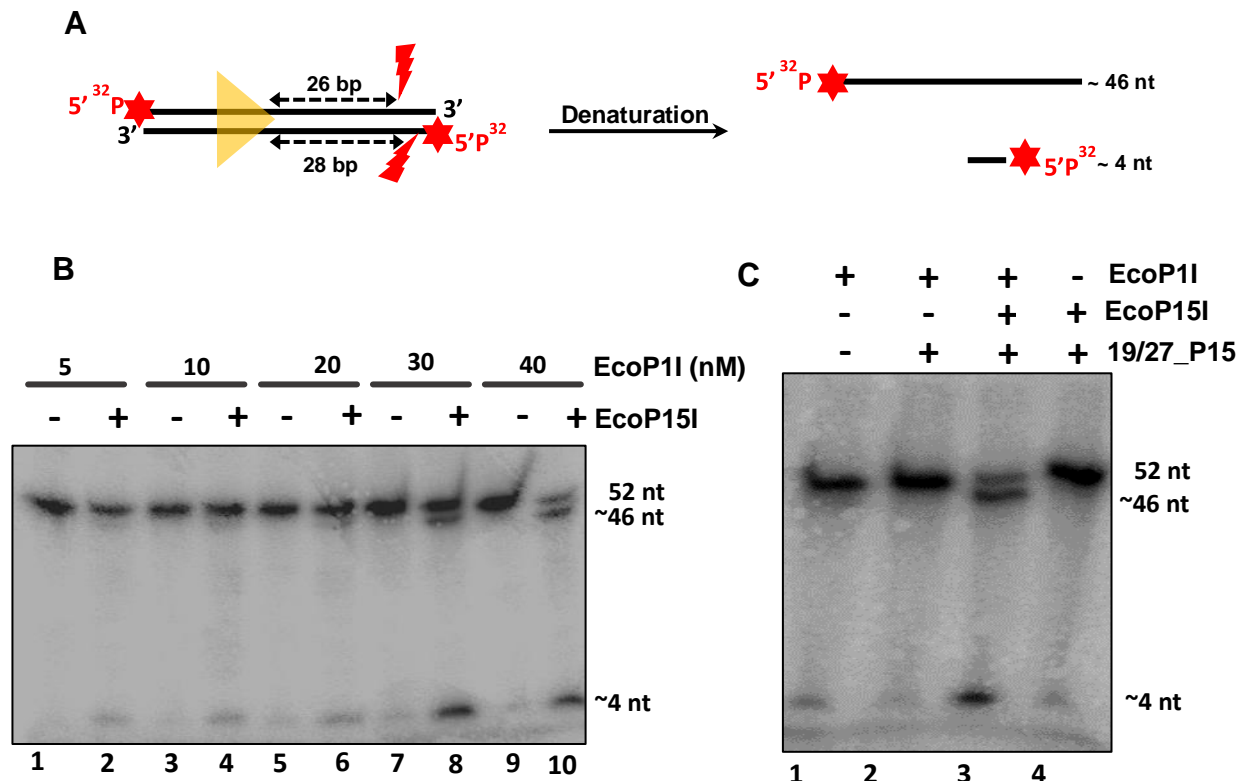
### 2.3.5 Heterologous cooperation Assay

Although our earlier heterologous cooperation assay revealed that single-site cleavage happened by the cooperation of two enzymes in *trans* (21), it did not rule out the possibility of a single enzyme bound to its recognition site also performing double-strand break on its own in *cis*. Due to the high concentration of enzyme-DNA complex used in the above assay, it was hard to delineate cleavage due to a single enzyme and cleavage due to cooperation between two enzymes encountering each other in solution (Figure 2.1 A, 2.1 C). To overcome the ambiguity, we studied single-site cleavage reaction at low concentrations of enzyme-DNA complex. Our reasoning was that if single-site cleavage required the enzymes to cooperate in *trans*, then the reaction would not occur at very low

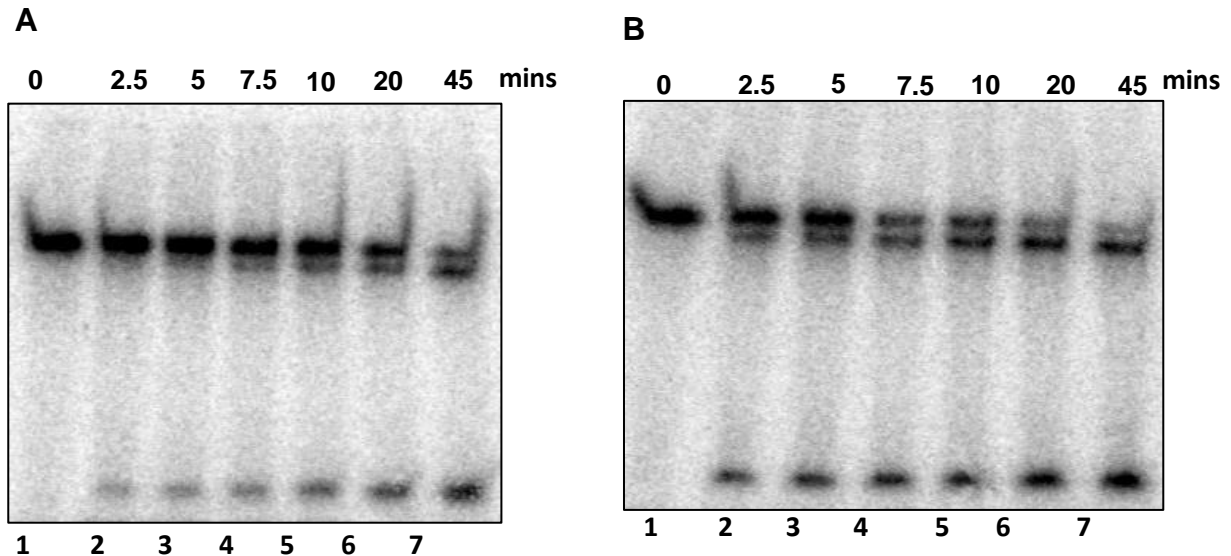
concentrations of enzyme-substrate complex due to a reduced frequency of two of the complexes coming together.

In this set of reactions, as the amount of DNA used was much less than what could be visualized using ethidium bromide stain, we used radiolabeled EcoP1I DNA substrates instead (Figure 2.9 A). Reaction mixes with varying concentrations of EcoP1I (5 to 40 nM) was incubated with 5 nM substrate DNA. Under such low concentrations of enzyme-DNA complex, amount of either single-strand nick or double-strand break observed was negligible (Figure 2.9 B, Lanes 1, 3, 5, 7 & 9). This suggested that an enzyme bound to its recognition site could not catalyze a nick or double-strand break on its own. However, a significant increase in DNA cleavage was noticed on addition of 250 nM of EcoP15I and 300 nM of its substrate DNA (Figure 2.9 B, Lanes 8 & 10). This led us to conclude that the single-site cleavage of a DNA required cooperation between an enzyme bound to the DNA in *cis* and an enzyme with its specific DNA in *trans*.

We next asked the question if the DNA product of a single-site cleavage reaction could in complex with the enzyme participate in cleaving other DNA substrates? Towards this, we purified cleaved DNA product of a single-site cleavage assay by EcoP15I using a native PAGE (Figure 2.8 A). The purified DNA was then used as the EcoP15I specific DNA in a heterologous cooperation assay. The purified DNA alone cannot produce nucleolytic cleavage of 15/32\_P1 (Figure 2.9 C, Lane 2). However, the EcoP1I specific substrate DNA (15/32\_P1) was cleaved in presence of the above-mentioned purified product DNA, EcoP1I, EcoP15I and ATP (Figure 2.9 C, Lane 3). To verify that the *trans* enzyme activity occurs in single enzyme system too, we carried out two experiments – (i) 5 nM of radiolabeled 15/32\_P1 was incubated with 250 nM of EcoP1I for different time periods (Figure 2.10 A); (ii) The same experiment as in (i) performed in presence of 250 nM of unlabeled 5/26\_P1 (Figure 2.10 B). We found cleavage under condition (ii), where unlabeled DNA was supplemented, to be much more efficient than under condition (i) (Figure 2.10 A and B). This observation is consistent with the model of the requirement of a *trans* acting enzyme to catalyze single-site cleavage. The DNA cleavage observed at longer periods of incubation under condition (i) could be the result of an activated enzyme that has dissociated from the DNA but still proficient to cooperate with the *cis*-bound enzyme to catalyze cleavage.



**Figure 2.9. Single-site cleavage is the result of cooperation between a *cis*-bound and a *trans*-acting enzyme.** (A) The cleavage products observed when a DNA with its 5'-ends labeled with  $^{32}\text{P}$  is analyzed on a denaturing urea-formamide PAGE gel. (B) Cleavage assay with 5 nM 15/32\_P1 DNA labeled at 5'-end with increasing concentration of EcoP1I. dsDNA break happened only when a higher concentration of EcoP15I/DNA complex was added indicating that single-site cleavage happened as a result of cooperation between *cis*-bound and *trans*-acting enzymes. (C) Heterologous cooperation assay demonstrating that the DNA product of a single-site cleavage can participate in further rounds of single site cleavage. Lane 1: 5 nM 15/32\_P1 DNA labeled at the 5'-end and 40 nM EcoP1I. Lane 2: 5 nM 15/32\_P1 DNA labeled at the 5'-end, 40 nM EcoP1I and 300 nM of purified product of the single-site cleaved 19/27\_P15 DNA. Lane 3: 5 nM 15/32\_P1 DNA labeled at the 5'-end, 40 nM EcoP1I, 250 nM of EcoP15I and 300 nM of the purified cleaved product of 19/27\_P15 DNA. Lane 4: 5 nM 15/32\_P1 DNA labeled at the 5'-end, 250 nM of EcoP15I and 300 nM of the purified cleaved product of 19/27\_P15 DNA.

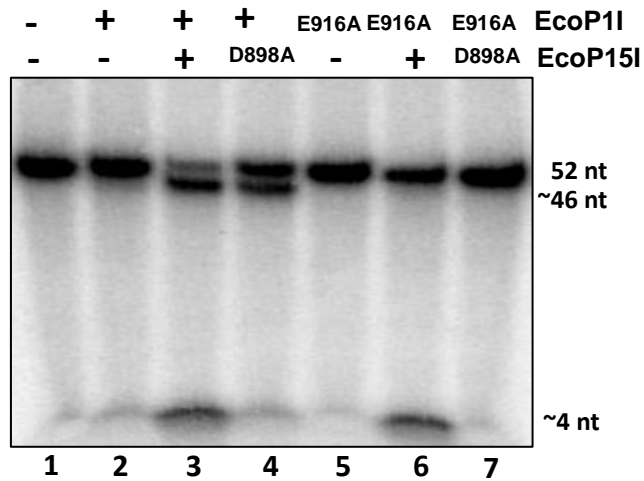


**Figure 2.10. *Trans* enzyme mediated DNA cleavage in single enzyme system.** (A) 5 nM of radiolabelled 15/32\_P1 (52 bp) was incubated with 250 nM of EcoP1I for different time periods in presence of ATP and checked for cleavage. At lower incubation periods, the cleavage of DNA was negligible. However, about 50% of DNA was cleaved in 45 mins. (B) The same experiment as in panel A performed in presence of 250 nM of unlabeled 5/26\_P1. Again 50% of DNA was cleaved in 7.5 mins. Lane 1 is free 15/32\_P1 DNA. In both the gels, the cleaved products are ~46, and ~4 nt long. Both the gels were 18% urea-formamide PAGE.

### 2.3.6 *The cis-bound enzyme nicks the top strand and the trans-acting enzyme nicks the bottom strand*

In the previous section we had shown that inactivating the nuclease of one of the enzymes resulted in only a nick, indicating that the double-strand break was the result of two nicks each catalyzed by one of the two cooperating enzymes. To identify which of the two enzymes nick the top and the bottom strands, we carried out another heterologous assay. In this assay, we used a combination of wild type and nuclease-dead enzymes, and radiolabelled the EcoP1I DNA substrate. As the 5'-ends of the EcoP1I DNA substrate were labeled, difference in length of the cleaved fragments generated, will help us to identify which of the two strands were cleaved by the *cis*-bound and *trans*-acting

enzymes, by keeping one of the enzymes active and the other nuclease-dead (Figure 2.9 A).



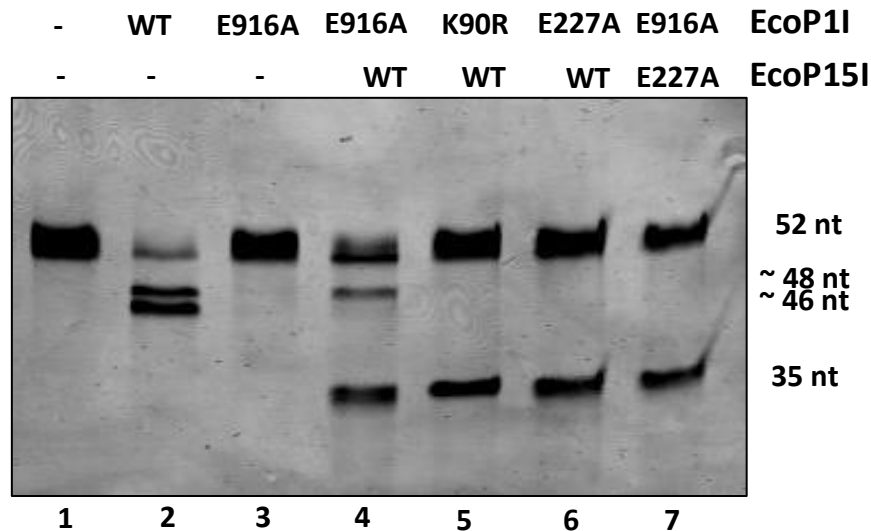
**Figure 2.11.** Cleavage assay demonstrating that the *cis*-bound enzyme cleaved the top strand and the *trans*-acting enzyme cleaved the bottom strand. 5 nM 15/32\_P1 and 300 nM 5/24\_P15 were used for the assay. See text for details.

As in the previous experiment, we maintained the EcoP1I or EcoP1I<sup>E916A</sup> concentration at 40 nM and its substrate DNA at 5 nM to reduce the frequency of two EcoP1I-DNA complexes coming together and causing DNA cleavage. To this reaction mix, we added 250 nM of EcoP15I or its nuclease-dead mutant EcoP15I<sup>D898A</sup> and 300 nM of its substrate DNA. Cleavage reaction was initiated by addition of ATP. A reaction mix with EcoP1I and EcoP15I yielded two fragments corresponding to nicks on both the top and bottom strands (Figure 2.11, Lane 3). A reaction involving wild type EcoP1I and EcoP15I<sup>D898A</sup> resulted in the nick of the top strand (Figure 2.11, Lane 4), while a reaction involving EcoP1I<sup>E916A</sup> and EcoP15I resulted in the nick of the bottom strand (Figure 2.11, Lane 6). These results revealed that the *cis*-bound enzyme cleaved the top strand, while the *trans*-acting enzyme cleaved the bottom strand.

### 2.3.7 Role of ATP in Single-Site Cleavage

In the preceding sections we demonstrated that single-site cleavage requires two enzymes to cooperate in *trans* and hydrolyze ATP. This led us to the question if ATP binding and/or hydrolysis are required by one or both of the cooperating enzymes. To address this question, we carried out heterologous cooperation assays with one of the

enzymes deficient in either ATP binding or hydrolysis. Mutation of Walker A and Walker B motif of the ATPase of EcoP15I has been shown to affect ATP binding and hydrolysis, respectively (26, 27). Accordingly, Walker A mutant EcoP1I<sup>K90R</sup> deficient in ATP binding, and the Walker B mutants EcoP1I<sup>E227A</sup> and EcoP15I<sup>E227A</sup> deficient in ATP hydrolysis were used for this experiment.



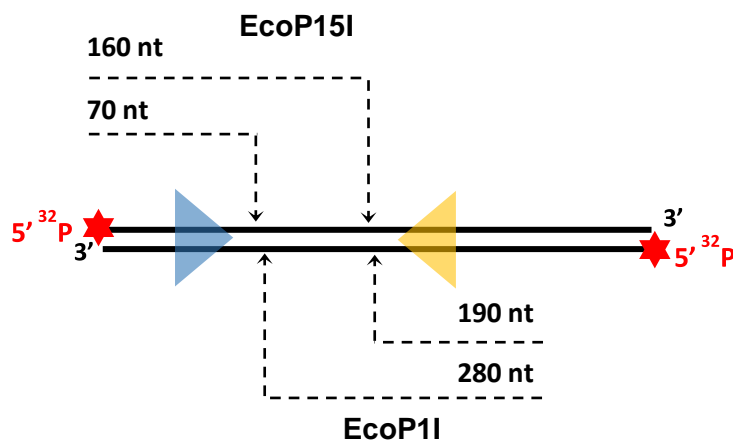
**Figure 2.12. Role of ATP on single-site DNA cleavage.** A denaturing urea-formamide PAGE gel stained with ethidium bromide to visualize the cleaved product of a heterologous assay involving the Walker mutants. Lanes 1) 15/32\_P1, 2) 15/32\_P1 + EcoP1I, 3) 15/32\_P1 + EcoP1I<sup>E916A</sup>, 4) 15/32\_P1 + EcoP1I<sup>E916A</sup> and 5/24\_P15 + EcoP15I, 5) 15/32\_P1 + EcoP1I<sup>K90R</sup> and 5/24\_P15 + EcoP15I, 6) 15/32\_P1 + EcoP1I<sup>E227A</sup> and 5/24\_P15 + EcoP15I, 7) 15/32\_P1 + EcoP1I and 5/24\_P15 + EcoP15I<sup>E227A</sup>. 300 nM of DNA and 250 nM of enzymes were used of EcoP1I specific DNA and its cleavage products, and the 35 bp long DNA is 5/24\_P15.

As shown earlier, heterologous cooperation assay with nuclease-dead EcoP1I<sup>E916A</sup> and wild type EcoP15I resulted in a DNA nick (21, Figure 2.12, Lane 4). However, replacement of the nuclease-dead enzyme by the Walker mutants EcoP1I<sup>K90R</sup> or EcoP1I<sup>E227A</sup> inhibited any form of nucleolytic activity from either of the enzymes (Figure 2.12, Lanes 5 & 6). A heterologous assay with EcoP1I<sup>E916A</sup> and ATP hydrolysis mutant EcoP15I<sup>E227A</sup> also showed a complete lack of either nick or double-strand break (Figure 2.12, Lane 7). These observations suggested that both the cooperating enzymes have to be proficient in not only ATP binding but also ATP hydrolysis for single-site cleavage to happen.

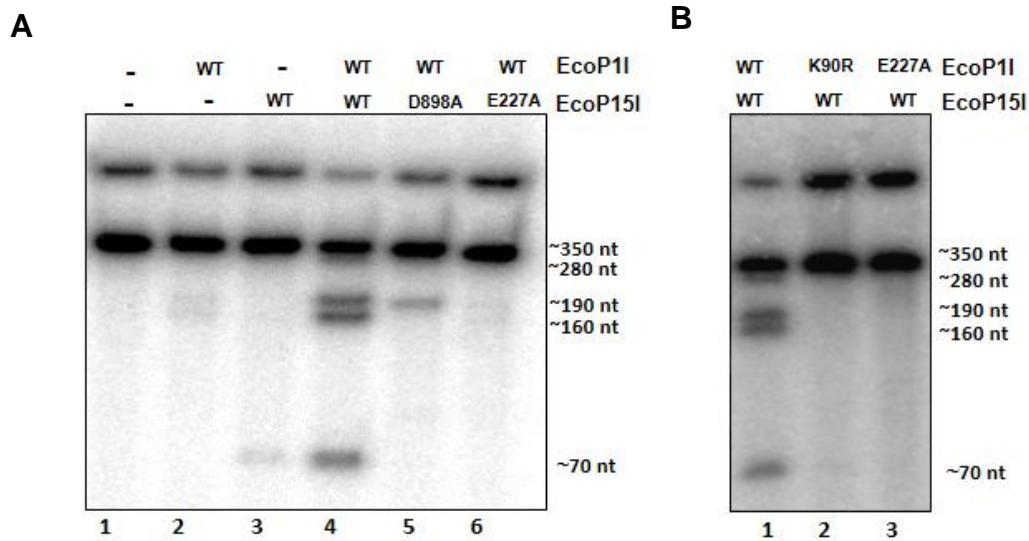


### 2.3.8 Effect of Walker mutants on cleavage of two-site substrates

We showed above that single-site cleavage requires ATP binding and hydrolysis by both the *trans*-acting enzymes. We, thence, proceeded to see if this conclusion extends to cleavage of the canonical two-site substrate. A previous study using a DNA substrate having an EcoP1I and an EcoP15I recognition site showed that ATP binding mutant EcoP1I<sup>R394A</sup> in cooperation with active EcoP15I could not cleave DNA while the wild type enzyme could (20). In similar lines, we designed an assay in which a 350 bp long DNA substrate with two recognition sites, one of EcoP1I and the other of EcoP15I, in head-to-head orientation was used. The expected products of cleavage are shown in Figure 2.13. The radiolabelled DNA incubated with 20 nM of EcoP1I or EcoP15I showed the occurrence of very faint single-site cleavage (Figure 2.14 A, Lanes 2 & 3). Incubation with an enzyme mix of 20 nM each of EcoP1I and EcoP15I resulted in DNA cleavage pattern that would be expected from cleavage of a canonical two-site substrate (Figure 2.14 A, Lane 4 and 2.14 B Lane 1). We observed four fragments corresponding to ~190 nt and ~160 nt resulting from cleavage next to the EcoP1I recognition site, and ~70 nt and ~280 nt resulting from cleavage next to the EcoP15I site (Figure 2.14 A, 2.14 B Lane 4 and Lane 1).



**Figure 2.13. ATP hydrolysis by both enzymes is essential for two-site cleavage.** (A) A schematic illustrating that the cleavage product of a 350 bp DNA labeled at 5'-end with <sup>32</sup>P having a single recognition site of both EcoP1I and EcoP15I in head-to-head orientation will result in four single-strands of length 70, 160, 190 and 280 nt when resolved on a denaturing urea-formamide PAGE gel.



**Figure 2.14. Denaturing 18% urea-formamide PAGE showing that ATP hydrolysis is essential for nicking or dsDNA break in two site DNA substrate.** (A) Lanes 1) 350 nt DNA, 2) 350 nt DNA + EcoP1I, 3) 350 nt DNA + EcoP15I, 4) 350 nt DNA + EcoP1I + EcoP15I, 5) 350 nt DNA + EcoP1I + EcoP15I<sup>D898A</sup>, 6) 350 nt DNA + EcoP1I + EcoP15I<sup>E227A</sup>. 3 nM of DNA and 20 nM of enzymes were used. (B) Denaturing 18% urea-formamide PAGE showing that ATP hydrolysis by EcoP1I is cleavage. Lanes 1) 350 nt DNA + EcoP1I + EcoP15I, 2) 350 nt DNA + EcoP1I<sup>K90R</sup> + EcoP15I, 3) 350 nt DNA + EcoP1I<sup>E227A</sup> + EcoP15I. 3 nM of DNA and 15 nM of enzymes were used. The gels in panel B and C demonstrate conclusively the ATP hydrolysis by both the stationary and diffusing enzymes are required for DNA nicking/cleavage. The top most bands in the gels B and C possibly are the result of incomplete denaturation of the 350 bp DNA.

A similar experiment with the nuclease-dead EcoP15I<sup>D898A</sup> resulted in only the ~280 nt and ~190 nt fragments (Figure 2.14 A, Lane 5). We interpreted the appearance of the faint ~280 nt fragment to be the result of a nick of the bottom strand catalyzed by a 1D diffusing EcoP1I in cooperation with a site bound EcoP15I<sup>D898A</sup>. The ~190 nt fragment to be the result of a nick of the top strand catalyzed by site-bound EcoP1I in cooperation with a 1D diffusing EcoP15I<sup>D898A</sup>, in *cis*. Replacement of nuclease-dead EcoP15I<sup>D898A</sup> by Walker B mutant EcoP15I<sup>E227A</sup> abolished nicking of the DNA (Figure 2.14 A, Lane 6). A similar experiment in which Walker mutants of EcoP1I were used in combination with active EcoP15I did not result either in nick or double-strand break (Figure 2.14 B, Lane 2

& 3). These results indicated that ATP binding and hydrolysis by the stationary as well as the diffusing enzymes were required for cleavage of a two-site substrate.

## 2.4. Discussion

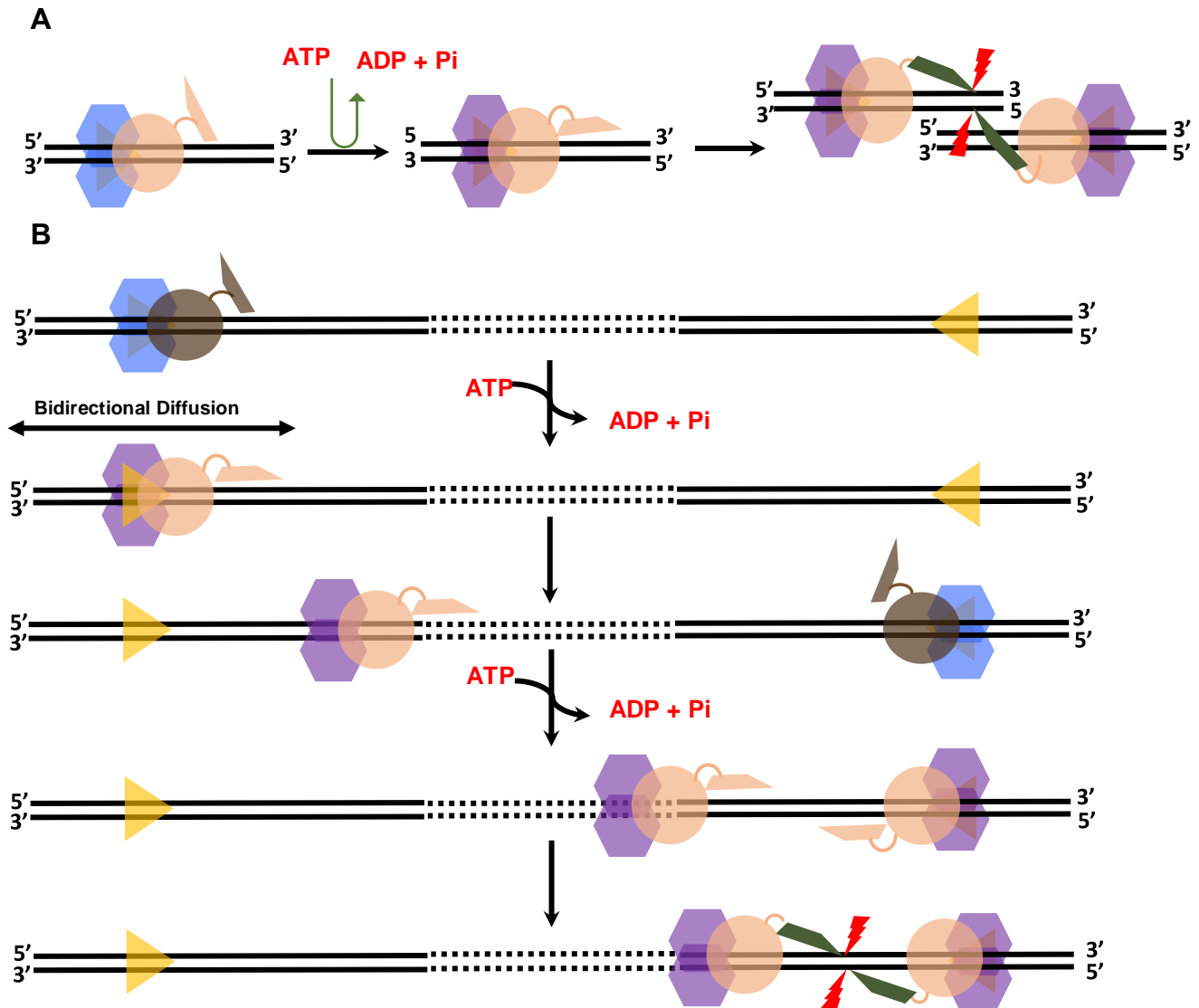
The canonical substrates of EcoP1I and EcoP15I have at least two sites in inverted orientation. Although observation of ATP-dependent endonucleolytic cleavage of DNA substrates with single-site has previously been mentioned in a few reports, its mechanism, hitherto, remained unknown (12, 13, 14, 15, 16, 17 and 18). This study demonstrates that single-site cleavage is solely a result of the interaction between two independent enzymes - a *cis*-acting enzyme which is bound to its recognition site in the DNA and a *trans*-acting enzyme either bound to or activated by its substrate in ATP-dependent manner. The two cooperating enzymes in solution can come together via diffusion.

For cleavage to happen, the interaction between the two nucleases should occur within a time period shorter than the time required by the enzyme in *cis* to leave its recognition site. In case of EcoP15I, this time is measured to be about 6-17 seconds (2). Furthermore, based on stopped-flow tryptophan fluorescence measurements, it has been proposed that the ATP-activated conformation of the enzyme retains for ~25 seconds even after it has dissociated from a short linear DNA substrate (2). It is possible that the activated *trans*-acting enzyme dissociates from its specific DNA but can still cooperate and catalyze DNA cleavage for a short period. As mentioned earlier, our experiments do not address this possibility. However, it has been shown previously that supercoiled plasmids that lack free ends are substrates for single-site cleavage by EcoP15I (17). This suggests that the activated enzyme in *trans* does not have to dissociate from the DNA to catalyze the cleavage. This also rules out the potential requirement of the *trans* enzyme to thread on to the DNA, as supercoiled plasmids do not have free ends.

Single-site cleavage shares two striking similarities with two-site cleavage – 1) in both the modes, the nucleolytic cut maps to identical positions with respect to the recognition site (19, 20); 2) the site-bound enzyme nicks the top strand and the

cooperating enzyme nicks the bottom strand (this study for single-site cleavage and Janscak *et al* (20) for two-site cleavage). Based on these similarities, we conclude that the configurations of two nucleases required to catalyze double-strand break are very similar in the two modes. The difference lies in how this configuration is achieved. In two-site cleavage, the nucleases converge by 1D diffusion and/or looping, while in single-site cleavage they come in proximity via 3D diffusion. Therefore, the two-site cleavage is a special form of single-site cleavage, where the motion of the enzyme is constrained to the length of the DNA. To achieve similar configuration in both the modes of cleavage, one would expect the two nucleases to be structurally flexible. This requirement is consistent with the reported crystal structure of EcoP15I in which the electron density for the C-terminal nuclease domain is poorly defined, possibly because of its mobility within the crystal lattice, preventing determination of the structure of the domain (28). A long linker domain connects the nuclease to the rest of the enzyme, which could contribute to the flexibility of the nuclease (28).

Previous studies have characterized in great detail the importance of ATP hydrolysis promoting 1D diffusion. It has been shown that about ten ATPs are hydrolyzed to initiate sliding of the enzyme along DNA (12). However, it remains, as yet, unaddressed if ATP hydrolysis is required only to initiate 1D diffusion, or if it also regulates the subsequent nucleolytic activity. In this study, we used Walker mutants to demonstrate that in both single-site and two-site cleavage ATP hydrolysis is required to activate the nucleases. Both the enzymes that cooperate to catalyze the cleavage have to be ATP hydrolysis-competent (Figure 2.11). Therefore, we conclude that ATP hydrolysis not only changes the conformational state of the enzyme to initiate sliding along DNA, but also activates the nuclease. This commonality in the two modes of cleavage is consistent with our above conclusion that the two have the same mechanism, and that they only differ in the manner in which the nucleases come together. Although ATP hydrolysis is essential for nuclease activation, an activated nuclease bound to its recognition site on its own is not proficient for strand nicking. We came to this conclusion from the results of single-site cleavage carried out at low concentrations of the enzyme, which, despite satisfying all the prerequisites, showed that the enzyme bound to its recognition site did not nick the DNA in presence of ATP. Instead, the nick catalyzed only in presence of another hydrolysis



**Figure 2.15. Model for DNA cleavage by Type III RM enzymes EcoP11 and EcoP15I.**

(A) A cartoon illustrating a model for single-site cleavage emphasizing that a target bound enzyme upon ATP hydrolysis changes to a nuclease activated state (represented as a change in color of the enzyme). However, nucleolytic activity manifests only when the enzyme encounters another activated nuclease (represented as change in color of the nuclease to green). (B) The model for two-site cleavage is similar to that of single-site cleavage except that the encounter between two activated nuclease is facilitated by 1D diffusion along the DNA and/or looping. For the sake of simplicity and clarity only 1D diffusion is shown.

competent enzyme (Figure 2.12 Lane 4 and 24). Based on these results we conclude that strand nicking and the subsequent endonucleolytic cleavage requires interaction between

an ATPase-activated nuclease bound to its recognition site with another activated nuclease. The same holds true in case of two-site cleavage (Figure 2.14)

In conclusion, the study reported here leads to a model (Figure 2.15 A) of DNA cleavage in which a recognition site bound Type III RM enzyme would undergo a conformational change induced by the hydrolysis of ATP making it both diffusion-competent and nucleolytically active. However, the activated nuclease can catalyze single-strand scission only in cooperation with another ATP-activated nuclease. We propose that this model is valid for both single-site and two-site cleavage, except that in single-site cleavage the cooperating nucleases come together by 3D diffusion, and in two-site cleavage they converge by 1D diffusion and/or looping (Figure 2.15 B). Our data suggests that single-site cleavage requires higher amount of enzyme than two-site cleavage. Due to decrease in dimension of the search space, convergence of enzymes by 1D diffusion and/or looping is expected to be more efficient than 3D diffusion due to higher local concentration of the enzymes at the site of action. A similar suggestion has been made by Butterer et al. 2014. (19). Hence, the role of single-site cleavage is likely to be less pronounced, and thus alleviating its toxic affects, in particular in a cell with newly replicated genomic DNA having unmodified daughter strand. However, under right enzyme, DNA and cofactor concentrations, the single-site cleavage could very well supplement two-site cleavage *in vitro* and perhaps *in vivo* (18). In particular, the study highlights that the products of cleavage by EcoP1I or EcoP15I can be substrates for and activators of single-site cleavage.

## References

1. Szczelkun,M.D. (2011). Translocation, switching and gating: potential roles for ATP in long- range communication on DNA by type III restriction endonucleases. *Biochem Soc. Trans.*, **39**, 589-594.
2. Schwarz,F.W., Toth,J., van Aeslt,K., Cui,G., Clausing,S., Szczelkun,M.D. and Seidel, R. (2013) The helicase-like domains of type III restriction enzymes trigger long-range diffusion along DNA. *Science*, **340**, 353-356.

3. Gorman,J., Chowdhury,A., Surtees,J.A., Shimada,J., Reichman, D.R., Alani, E., and Greene, E.C. (2007) Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. *Mol. Cell*, **28**, 359-370.
4. Liu,J., Hanne,J., Brooke,M.B., Bennett,J., Kim,D., Lee,J.B., and Fishel,R. (2016) Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. *Nature*, **539**, 583-589.
5. Fishel,R. (2015) Mismatch repair. *J. Biol. Chem.*, **290**, 26395-26403.
6. Rao,D.N., Dryden,D.T.F. and Bheemanaik,S.(2014) Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.*, **42**, 45-55.
7. Crampton,N., Roes,S., Dryden,D.T.F., Rao,D.N., Edwardson,J.M. and Henderson,R.M. (2007) DNA looping and translocation provide an optimal cleavage mechanism for the type III restriction enzymes. *EMBO J.*, **26**, 3815-3825.
8. Crampton,N., Yokokawa,M., Dryden,D.T.F., Edwardson,J.M., Rao,D.N., Takeyasu.K., Yoshimura.S.H. and Henderson,R.M. (2007) Fast-scan atomic force microscopy reveals that the type III restriction enzyme EcoPP15I is capable of DNA translocation and looping. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 12775-12760.
9. Ramanathan,S.P., van Aelst,K., Sears,A., Peakman,L.J., Diffin,F.M., Szczelkun,M.D., and Seidel, R.(2009) Type III restriction enzymes communicate in 1D without looping between their target sites. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 1748-1753.
10. Meisel,A., Bickle,T.A., Kruger,D.H. and Schroeder.C (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature*. **355**, 467-469.
11. Miesel.A., Mackeldanz.P., Bickle.T.A., Kruger,D.H. and Schroeder.C (1995) Type III restriction enzymes translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. *EMBO J.*, **14**, 2958-2966.
12. Toth,J., Bollins,J., and Szczelkun,M.D. (2015) Re-evaluating the kinetics of ATP hydrolysis during initiation of DNA sliding by type III restriction enzymes. *Nucleic Acids Res.*, **43**, 10870-10881.

13. Mucke,M., Reich,S., Moencke-Buchner,E., Reuter,M. and Kruger, D.H. (2001) DNA cleavage by type III restriction-modification enzyme EcoP15I is independent of spacer distance between two head to head oriented recognition sites. *J. Mol. Biol.*, **312**, 687-698.
14. Peakman,L.J., Antognozzi,M., Bickle,T.A., Janscak,P. and Szczelkun,M.D. (2003) S-adenosyl methionine prevents promiscuous DNA cleavage by the EcoP1I type III restriction enzyme. *J. Mol. Biol.*, **333**, 321-335.
15. Raghavendra,N.K. and Rao,D.N. (2004) Unidirectional translocation from recognition site and a necessary interaction with DNA end for cleavage by type III restriction enzyme. *Nucleic Acids Res.*, **32**, 5703-5711.
16. Raghavendra,N.K. and Rao,D.N. (2005). Exogenous AdoMet and its analogue sinefungin differentially influence DNA cleavage by R.EcoP15I- Usefulness in SAGE. *Biochem. Biophys. Res. Commun.*, **334**, 803-811.
17. Moncke-Buchner,E., Rothenberg,M., Reich,S., Wagenfuhr,K., Matsumura,H., Terauchi,R., Kruger,D.H., and Reuter,M. (2009) Functional characterization and modulation of DNA cleavage efficiency of type III restriction endonuclease EcoP15I in its interaction with two sites in the DNA target. *J. Mol. Biol.*, **387**, 1309-1319.
18. Peakman,L.J. and Szczelkun,M.D. (2009). S-Adenosyl homocysteine and DNA ends stimulate promiscuous nuclease activities in the type III restriction endonuclease EcoP1. *Nucleic Acids Res.*, **37**, 3934-3945.
19. Butterer,A., Pernstich,C., Smith,R.M., Sobott,F., Szczelkun,M.D. and Toth,J. (2014). Type III restriction endonuclease are heterotrimeric: comprising one helicase-nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Res.*, **42**, 5139–5150.
20. Janscak,P., Sandmeier,U., Szczelkun,M.D. and Bickle,T.A. (2001) Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I. *J. Mol. Biol.*, **306**, 417-431.
21. Kulkarni, M (2017) Biochemical, bioinformatics and crystallographic analysis of type III restriction-modification enzyme EcoP1I (Doctoral dissertation)



22. Sasnauskas,G., Halford,S.E. and 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.
23. Bitinaite,J., Wah,D.A., Aggarwal,A.K. and Schildkraut, I. (1998) *FokI* dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 10570-10575.
24. Kunz,A., Mackeldanz,P., Mucke,M., Meisel,A., Reuter,M., Schroeder,C. and Kruger,D.H (1998) Mutual activation of two restriction endonuclease: interaction of EcoP1 and EcoP15. *Biol. Chem.*, **379**, 617-620.
25. Miroux, B. and Walker, J.E. (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.*, **260**, 289-298.
26. Saha,S. and Rao, D.N. (1997) Mutations in the Res subunit of the EcoP1 restriction enzyme that affect ATP-dependent reactions. *J. Mol. Biol.*, **269**, 342-354.
27. Mackeldanz,P., Alves,J., Moncke-Buchner,E., Wyszomirski,K.H., Kruger,D.H., Reuter,M. (2013) Functional consequences of mutating conserved SF2 helicase motifs in the type III restriction endonuclease EcoP15I translocase domain. *Biochemi.*, **95**, 817-823.
28. Gupta,Y.K., Chan,S.H., Xu,S.Y., and Aggarwal,A.K. (2015) Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat Commun.*, **6**, 7363.

# **CHAPTER 3**

## **Crystallographic studies of EcoP15I and EcoP1I**

## Chapter 3

### Crystallographic studies of EcoP15I and EcoP1I

#### 3.1 Introduction

Type III R-M systems are complex heterooligomeric molecules exerting both restriction and modification activities. They are composed of two subunits, Mod subunit encoded by the *mod* gene and Res subunit encoded by *res* gene (1). The dimeric Mod subunit (Mod<sub>2</sub>) causes site-specific binding and acts as methyltransferase. Res subunit on binding with Mod forms an active complex Mod<sub>2</sub>Res<sub>1</sub> that is proficient to move on DNA and nucleolytically cleave it using ATP hydrolysis (1). The Mod subunit contains an N-terminal catalytic domain and a C-terminal S-adenosyl methionine binding domain with the target-recognition domain (TRD) in between the two (Fig 1.3 A). The Res subunit contains an N-terminal ATPase domain belonging to SF2 helicase family connected to the C-terminal endonuclease domain by a long linker (Fig 1.4 A). The two well-studied Type III RM enzymes are EcoP1I and EcoP15I encoded by the P1 bacteriophage and plasmid p15B, respectively. Both EcoP1I and EcoP15I are homologous enzymes with almost 75% sequence identity between the Mod subunits and 95% sequence identity between the Res subunits (Fig 1.3 B & 1.4 B). The recognition sequence of EcoP1I and EcoP15I is AGACC and CAGCAG respectively with enzymes methylating second adenine in their recognition site (1).

To cleave un-methylated foreign DNA, EcoP1I and EcoP15I need two such unmodified recognition sites in an inverted orientation that can be up to 3.5 kb base pair, apart with cleavage happening 25-27 bp downstream to any one of the sites (2, 3). 1D diffusion model is the mostly accepted model explaining how, two enzymes that can be thousands of base pair apart communicate with each other (4). According to the model, after the binding of enzymes to their recognition site, a rapid burst of 10-12 ATP molecules occurs (5, 6). The ATP hydrolysis causes a large conformational change in the protein after which it vacates the recognition site and enters into sliding or diffusive state. Whether, the diffusion of the enzyme on DNA need further ATP is not known (5, 6). When one such diffusive enzyme encounters another enzyme bound to its recognition site dsDNA break is triggered (5, 6). The role of ATP hydrolysis in Type III RM enzymes is to

act as a molecular switch, changing the enzyme from a stationary state to an activated state (5, 6).

Apart from cleaving two site DNA substrate, Type III RM enzymes can also cleave single-site DNA substrate (7, 8, 9, 10, 11, 12, and 13). From our results in Chapter 2, we know for single-site DNA cleavage, two enzyme-DNA complexes come together by 3D diffusion. Furthermore single-site DNA cleavage happens only when ATP hydrolysis occurs, indicating it is essential to activate these enzyme for DNA cleavage (Chapter 2). Our result from single-site DNA cleavage assay shows that in presence of ATP, a nuclease dead mutant enzyme although ATPase active, can cooperate with a wild type enzyme to perform single-site cleavage (Chapter 2). However an enzyme deficient in ATP binding or hydrolysis, even though nuclease active, cannot cooperate with the wild type enzyme to perform single site DNA cleavage. The result indicates that ATP hydrolysis causes a conformational change not only in the ATPase domain, but also activates the endonuclease domain of these enzyme to endonucleolytically cleave DNA. In single-site DNA cleavage, two ATP activated enzymes still bound to their DNA can cooperate with each other to perform dsDNA break without entering diffusion. Our single-site assays with Walker mutants of EcoP1I and EoP15I, indicate that mere ATP binding cannot activate these enzymes to perform single-site DNA cleavage, but hydrolysis of ATP is necessary for DNA cleavage to happen (Chapter 2). To understand the structural basis of how ATP hydrolysis activates the enzyme, I aim to crystallize a ternary complex of Type III RM enzymes with specific DNA and in presence of non-hydrolysable ATP analogues. A crystal structure of enzyme-DNA complex in presence of AMPPNP will give structural insights into the enzyme before ATP hydrolysis, while a crystal of enzyme-DNA complex with ADP-vanadate will provide structural insights accompanied when the enzyme get activated by ATP hydrolysis. Recently Gupta et al. in 2015 published a partial structure of EcoP15I bound to specific DNA (14). The partial structure sheds new light about how the division of labour is accomplished in case of dimeric methyltransferases, where one Mod subunit undertakes the role of site-specific recognition, while the other Mod carries transfer of methyl group to target adenine in recognition site (14). However, there are still many questions, which remain unaddressed.

To dissect the mechanistic details of how ATP hydrolysis activates these enzymes, we started to carry out structural studies on EcoP15I and EcoP1I. Our biochemical studies, led us to unravel the mechanism of how EcoP15I and EcoP1I cleave a DNA substrate containing a single recognition site (Chapter 2). The DNA binding and nucleolytic cleavage studies gave us the idea in designing DNA substrates for co-crystallization with EcoP15I/EcoP1I. As crystallographic studies needed relative high quantity and quality of protein, EcoP15I and EcoP1I were first cloned in high expression vectors and purified to near homogeneity (Chapter 2). To start off, we first started with co-crystallization studies with EcoP15I followed by EcoP1I. The following section details the work carried out in crystallization of binary and ternary complexes of EcoP15I and EcoP1I.

## 3.2 Materials and Methods

### 3.2.1 DNA substrates

DNA substrates were purchased from Integrated DNA Technologies and Sigma Aldrich as single strand oligos. The two complementary DNA strands were annealed using temperature gradient from 95°C - 25°C. The annealed DNA substrates were further purified using an 8 mL Mono Q 10/100 GL column. The resulting duplex was washed thoroughly with Milli-Q water, and concentrated using Vivaspin concentrator (MWCO 3 kDa; GE Healthcare). The concentrated DNA was stored at -30°C until further use. The list of dsDNA used in our crystallographic studies is listed in the Table 3.1.

**Table 3.1:** List of oligomers used for the crystallization of EcoP15I and EcoP1I

| Name       | Sequence (5'->3')                             | Length |
|------------|-----------------------------------------------|--------|
| 5/25_P15_T | GTACT <u>CAGCAGT</u> ATCCTGTATGCTACGTATTCGTAC | 36     |
| 5/25_P15_B | GTACGAATACGTAGCATA <u>CAGGATACTGCTG</u> AGTAC | 36     |
| 5/24_P15_T | GTACT <u>CAGCAGT</u> ATCCTGTATGCTACGTATTCGTC  | 35     |
| 5/24_P15_B | GTACT <u>CAGCAGT</u> ATCCTGTATGCTACGTATTCGTC  | 35     |
| 5/26_P1_T  | GTACT <u>AGACCT</u> ATCCTGTATGCTACGTATTCGTATC | 36     |
| 5/26_P1_B  | GATACGAATACGTAGCATA <u>CAGGATAGGTCT</u> AGTAC | 36     |

### 3.2.2 Protein Purification

Wild type EcoP15I and nuclease dead EcoP15I<sup>D898A</sup> were purified as described in Chapter 2. Similarly EcoP1I and selenomethionine (Se-Met) derivative of EcoP1I (EcoP1I<sup>Se-Met</sup>) was purified as described in Chapter 2. Based on SDS-Page analysis only the purest fractions were used for crystallization.

### 3.2.3 Generation of Methionine mutant of EcoP1I

Single methionine mutants EcoP1I<sup>K953M</sup>, EcoP1I<sup>S947M</sup>, double methionine mutants EcoP1I<sup>I866M,K953M</sup> and triple methionine mutants EcoP1I<sup>I815M,I866M,K953M</sup> were generated using overlap PCR. The primers used for incorporation of methionine residues in the endonuclease domain of EcoP1I are as listed in the Table 3.2. The amplified inserts were double digested with NcoI and XhoI restriction enzymes and ligated into pRSF vector using T4 DNA ligase (New England Biolabs). The positive clones obtained were sent for sequencing and confirmed only after complete sequencing of the gene. The methionine mutants of EcoP1I were purified using the same protocol as wild type EcoP1I.

**Table 3.2.** Primers used for generating methionine mutants of EcoP1I

| Name      | Sequence (5'->3')             | Length |
|-----------|-------------------------------|--------|
| For Pri 1 | GTTAAAGTTGAATTCATGACGCAGTTTGC | 30     |
| For Pri 2 | CAATCAAATAAGTATGTCTGTAAAG     | 26     |
| For Pri 3 | GAAAGGCGTAATATGACAGATAGAGAG   | 28     |
| For Pri 4 | CGGGTTCAATGCATCCACAAAGTTC     | 27     |

### 3.2.4 Crystallization of EcoP15 with DNA

A co-crystallization complex of purified EcoP15I and DNA was made by mixing the two in a molar ratio of 1:1.3 in presence of 1 mM sinefungin (SNF). The final volume of the crystallization sample was made with crystallization buffer (10 mM Tris-HCl, pH 8, 100 mM KCl, 1 mM DTT). The complex was kept on ice for 15 mins, after which the samples were centrifuged at 15000 rpm at 4°C for 15 mins. A total of 1440 commercially available conditions from Hampton, Molecular dimension were screened for initial crystallization

hits. Protein/DNA complex of 100 nL was mixed with 100 nL reservoir buffer with the help of TTP LABTECH LTD mosquito crystal robotic liquid handler. The plates were set up by sitting drop method and kept at 18°C using MRC 2 Well Crystallization Plate UVP. Out of 1440 conditions, only two conditions gave possible crystal hits (See section 3.3.1 for further details). The initial crystal hits were further optimized by setting up the additive screen (Hampton) in 1:10 ratio with the parent condition. Using additive screen we got crystals with various additives. The crystal obtained in 96 well plates were small for diffraction studies. To increase the size of the crystal, big 2 µL drops (1 µL of Protein/DNA complex + 1 µL reservoir solution) were set up in 48 well crystallization plates (MRC 2 Well Crystallization Plate UVP). The volume of reservoir solution was kept at 80 µL.

### *3.2.5 Crystallization of EcoP15I-DNA with ADP-VO<sub>4</sub>*

Apart from crystallizing binary complex of EcoP15I with DNA, we co-crystallized EcoP15I with DNA in presence of non-hydrolysable ATP analogues (ADP-Vanadate). A ternary complex of EcoP15I, DNA and ADP-VO<sub>4</sub> was formed by mixing EcoP15I and DNA in the molar ratio of 1:1.3. The final condition contained 5 mM MgCl<sub>2</sub>, 2 mM ADP, 2 mM ortho-vanadate and 1 mM sinefungin. The final volume of the crystallization sample was made with crystallization buffer (10 mM Tris-HCl, pH 8, 100 mM KCl, 1 mM DTT). The complex was incubated on ice for 15 mins after which the samples were centrifuged at 15000 rpm, 4°C for 15 mins. A total of 1440 commercially available conditions from Hampton, Molecular dimension were screened for initial crystallization hits. EcoP15I-DNA + ADP-VO<sub>4</sub> complex of 100 nL was mixed with 100 nL reservoir buffer with the help of Mosquito crystal robotic liquid handler. The plates were set up using sitting drop method and kept at 18°C. Out of 1440 conditions, two conditions gave possible crystal hits, and among the two only one condition gave reproducible crystal hits. (See section 3.3.2 for further details). To increase the size of the crystal, 2 µL drops (1 µL of Protein/DNA complex + 1 µL reservoir solution) were set up in 48 well MRC 2 Well Crystallization Plate UVP. The volume of reservoir solution was kept at 80 µL.

### *3.2.6 Crystallization of EcoP1/DNA complex with AMPPNP*

Purified EcoP1I and DNA were mixed in the molar ratio of 1:1.3 in presence of 5 mM MgCl<sub>2</sub>, 1 mM AMPPNP and 1 mM sinefungin (SNF) at 4°C. The final volume of the

crystallization sample was made with crystallization buffer (10 mM Tris-HCl, pH 8, 100 mM KCl, 1 mM DTT). The mixture was kept on ice for 15 mins to allow the ternary complex to form, after which the samples were centrifuged briefly at 15000 rpm at 4°C for 15 mins. The supernatant was transferred to another tube and crystallization screens were set up. A total of 1440 commercially available conditions from Hampton, Molecular dimension were screened for initial crystallization hits. Protein/DNA complex of 100 nL was mixed with 100 nL of reservoir buffer with the help of TTP LABTECH LTD mosquito crystal robotic liquid handler. The plates were set up in sitting drop manner and kept at 18°C (MRC 2 Well Crystallization Plate UVP). The plates were checked for initial crystal hits after 2 days. From 1440 commercially available conditions, only one condition gave a possible crystal hit (See Section 3.4.1 for details). The initial crystal hit was further optimized by changing the concentration of the components in the parent condition. The crystals obtained in 96 well plate were small for diffraction studies. To increase the size of the crystal, 2 µL drops (1 µL of Protein/DNA complex + 1 µL reservoir solution) were set up in 48 well crystallization plates (MRC 2 Well Crystallization Plate UVP). The volume of reservoir solution was kept at 80 µL.

### *3.2.7 Crystallization of EcoP1/DNA complex with ADP-VO<sub>4</sub>*

A ternary complex of EcoP1I, DNA and ADP-VO<sub>4</sub> was formed by mixing EcoP1I and DNA in the molar ratio of 1:1.3. The final condition contained 5 mM MgCl<sub>2</sub>, 2 mM ADP, 2 mM orthovanadate and 1 mM sinefungin. The components and final volume of the crystallization sample was made with crystallization buffer (10 mM Tris-HCl, pH 8, 100 mM KCl, 1 mM DTT). For proper complex formation the sample mixture was kept on ice for 15 mins after which the samples were centrifuged at 15000 rpm, 4°C for 15 mins. The supernatant was transferred into another tube and crystallization screens were setup. A total of 1440 commercially available conditions from Hampton, Molecular dimension were screened for initial crystallization hits. Out of 1440 conditions, only three conditions gave possible crystal hits. Among the three conditions, one of the conditions were giving reproducible crystal hits. (See section 3.4.4 for further details). The condition was further modified to obtain better quality crystal. To improve the size and quality of the crystals an Additive screen (Hampton) was set up in 1:10 ratio with parent condition. The crystal



appeared using a couple of additives, however single and good quality crystal were obtained with Benzamidine hydrochloride. To further increase the size of the crystal, big drop of 2  $\mu\text{L}$  (1  $\mu\text{L}$  of Protein/DNA complex + 1  $\mu\text{L}$  reservoir solution) was set up in 48 well crystallization plates (MRC 2 Well Crystallization Plate UVP). The volume of reservoir solution was kept at 80  $\mu\text{L}$ .

### *3.2.8 X-ray diffraction data collection and processing*

Single isolated and bigger crystals were picked from the drop and flash frozen with 30% ethylene glycol as cryo-protectant. In house X-ray diffraction was collected using a Rigaku MicroMax 007 X-ray generator and Mar Research 345D detector. Best diffracting crystals at home source were retrieved and taken for diffraction studies at synchrotron facilities, Diamond Light Source (DLS) Oxfordshire, UK and European Synchrotron Research Facility (ESRF), Grenoble. Diffracted data was indexed and processed using XDS (15). Scaling and merging of processed data was carried out using AIMLESS in the CCP4 suite (16, 17). Model building was carried out in COOT (18).

### *3.2.9 Structure solution by molecular replacement and SAD phasing*

Structure solution of EcoP11-DNA-AMPPNP crystal was determined using a combination of molecular replacement and SAD phasing. The crystal structure of EcoP15I (PDB code: 4ZCF) was used as a model for molecular replacement by the program PHASER (19) in the CCP4 suites of programs (17). The different subunits of EcoP15I were used as individual search models to obtain a composite structure. The selenium positions were identified using the program SOLVE-RESOLVE (20) in the PHENIX suite of crystallography programs (21). In all, 18 of the 21 methionines were located in the three subunits. The partial phases obtained from molecular replacement solution was combined with the SAD phases. A simple Fourier map calculated using the SAD phases was used to manually build the structure of EcoP11. Initial electron density maps were visualized and further model building was carried out in COOT (18). Due to limitation in time, I could not proceed with model building and subsequent refinement of the structure. This is now being pursued in the laboratory.

### 3.3 Results:

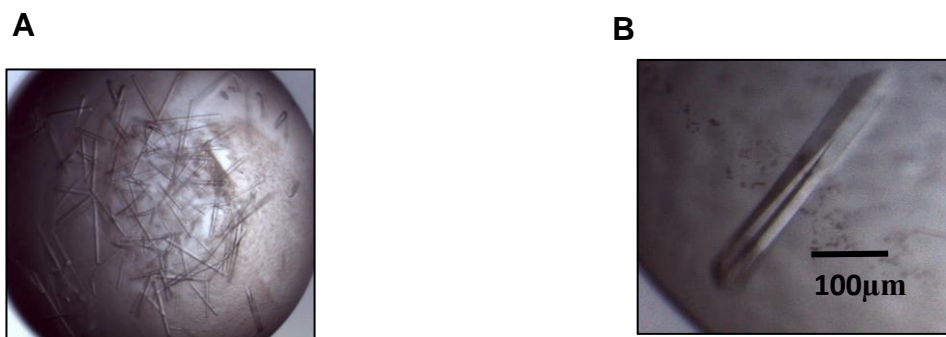
#### 3.3.1 Initial optimization and crystallization of EcoP15I/DNA complex

Our DNA binding studies clearly indicated that EcoP15I can bind to a short oligos containing a single recognition site. The knowledge gained helped us in designing a DNA substrate used for co-crystallization experiments. Purified EcoP15I was mixed with specific DNA and commercially available crystallization screens were screened. I screened about 1440 different commercially available condition from which two conditions gave crystal hits. The composition of the conditions, where initial crystal hits appeared are given in the Table 3.3. Out of the two conditions, only one condition gave reproducible hits.

**Table 3.3:** Composition of initial condition where crystal were obtained

|                 | Condition 1       | Condition 2                      |
|-----------------|-------------------|----------------------------------|
| Precipitant Mix | 10% PEG 6K,5% MPD | 10 % PEG 8K, 10% Ethylene glycol |
| Buffer          | Na-HEPES, 0.1 M   | Na-HEPES, 0.1 M                  |
| pH              | 7.5               | 7.5                              |

Condition 1 (Parent condition) was further optimized with respect to pH of the buffer, PEG 6K concentration and MPD. It was noticed that crystals only appeared with PEG 6K (8%-9%), 3-5% MPD and 0.1M Na-HEPES. To further improve the quality of crystals, we carried out additive screen (Hampton) with the parent condition. In additive screen, I got single big and isolated crystals with various additives, however, best crystals were obtained with 0.01 M Phenol. Further crystallization, was carried out with 0.01 M phenol along with the parent condition. To further increase, the size of crystals the volume of crystallization drop was increased to 2  $\mu$ L (1  $\mu$ L Protein +1  $\mu$ L Reservoir buffer) and crystallization was carried out in 48 well plate. The crystals obtained in 48 well plate were of big size (Figure 3.1).



**Figure 3.1. Crystals of EcoP15I-DNA complex.** (A) Initial crystal hits of EcoP15I-DNA-SNF. (B) Crystals obtained after optimization of the parent condition with the addition of 0.01 M Phenol.

The crystals were flash frozen with different cryo-protectants and tested for diffraction at in-house X-ray diffractometer. The crystal at home source diffracted poorly to 40 Å. Similar quality crystals were taken to synchrotron where most of the crystals diffracted to 25 Å. As crystals diffracted poorly, we also tried room temperature diffraction as well as diffraction studies with different cryo-protectant. However, in all cases, there was no improvement in the diffraction quality. I also crystallized EcoP15I with oligos of different length, but the crystals obtained were of same morphology and in all cases diffraction was poor. Despite trying out all possibilities, we were unsuccessful in improving the resolution of the crystals. As an alternate strategy and to get structural information about ATP-activated enzyme, we started co-crystallizing EcoP15I with DNA in presence of ATP analogues.

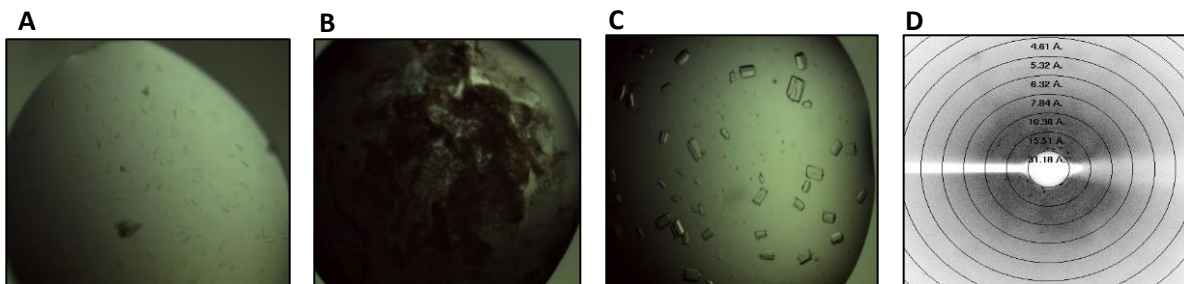
### 3.3.2 Initial optimization and crystallization of EcoP15I/DNA complex in presence of ADP-VO<sub>4</sub>

As crystals of binary complex (EcoP15I-DNA) failed to diffract, I started an alternative strategy of carrying crystallization screen of EcoP15I+DNA in presence of ATP analogues. The idea being that a ternary complex EcoP15I + DNA with ATP analogues will crystallize in a different condition and with different crystal packing, by virtue of which the crystal of ternary complex EcoP15I/DNA + ADP-VO<sub>4</sub> may diffract better. I screened 1440 different commercially available for any crystal hits. Out of 1440 different commercially available conditions, we got crystalline hits only in one condition (Figure 3.2 A). The composition of the condition, where initial crystal hits were observed is as shown

in Table 3.4. I tried to reproduce crystals in the parent condition, however, in cases, all the crystallization drops showed brown precipitation (Figure 3.2 B). I next set up the grid of varying PEG, buffer and salt concentration, but still, in all these cases, the crystallization drops showed brown precipitation.

**Table 3.4:** Composition of the initial condition where crystals were observed

|                 | Condition                                         |
|-----------------|---------------------------------------------------|
| Precipitant Mix | 15% PEG4000                                       |
| Buffer          | 0.2 M Potassium thiocyanate, 0.1 M Sodium acetate |
| pH              | 5.5                                               |



**Figure 3.2. Optimization of the condition for the co-crystallization of EcoP15I-DNA-ADPVO<sub>4</sub>.** (A) Initial crystal hits in 96 well crystallization plate. (B) Brown precipitation obtained while reproducing the initial crystals using the commercially available condition. (C) Picture of EcoP15I-DNA-ADPVO<sub>4</sub> crystal obtained in 48 well crystallization plate. (D) X-ray diffraction picture from one such crystal.

Next, I checked the pH of the commercially available condition with the homemade condition using pH strip. I noticed that the pH of commercially available condition to be 7.5, while on the crystallization screen document it was mentioned to be pH 5.5. I next made home-made condition with pH 7.5. We were able to obtain the crystals of EcoP15I-DNA +ADP-VO<sub>4</sub>, using parent condition with pH 7.5, indicating that pH mentioned in the commercially available condition was wrong (Figure 3.2 C). Although I was able to reproduce the crystals, the size of the crystals was small due to high nucleation in the drops. In some drops, I was able to obtain crystals of 60  $\mu\text{m}$ . In order to improve the size of crystals we decreased the protein concentration to reduce nucleation, but still, the size

of the crystals did not increase. I also tried to add paraffin oil above the reservoir buffer to reduce nucleation but it also did not work. In all cases, we were only able to obtain small crystals. The crystals of about 60  $\mu\text{m}$  size were flash frozen with 30% ethylene glycol as cryo-protectant. The crystals diffracted to 11  $\text{\AA}$  at home source, while the similar size crystals diffracted to 8  $\text{\AA}$  at a synchrotron (Figure 3.2 D). All attempts to increase the size of the crystals to obtain high resolution were in vain, and hence we did not proceed further with the crystallographic studies.

Although from our crystallographic studies on EcoP15I, I could co-crystallize EcoP15I with DNA and with ADP-VO<sub>4</sub>, however I was unsuccessful in obtaining a high-resolution data to get insights into the molecular architecture of the enzyme.

### **3.4. Structural studies of EcoP1I.**

As an alternate strategy, I started to carry crystallization trials on EcoP1I, which is a close homologue of EcoP15I. EcoP1I as a binary complex with specific DNA was already co-crystallized in our lab. The crystals had diffracted to 4.4  $\text{\AA}$ , and structure solution required experimental phase information (22). The composition of the condition where EcoP1I was crystallized is as shown in Table 3.5 (22). I started with the aim to co-crystallize EcoP1I with specific DNA in presence of non-hydrolyzing ATP analogues and to obtain phase information. The aim was also to trap the conformational changes that happen when EcoP1I accomplish ATP hydrolysis, which is required for nucleolytic activity. We carried out co-crystallization screens of EcoP1I/DNA complex with non-hydrolysable ATP analogue i.e. AMPPNP and transition state analogues ADP-VO<sub>4</sub>. The following section will discuss the efforts carried out in crystallization and structure determination of EcoP1I-DNA complex in presence of ATP analogues.

**Table 3.5.** Composition of condition where EcoP11-DNA complex was crystallized

|             |                                |
|-------------|--------------------------------|
|             | Condition                      |
| Precipitant | 10% PEG 6K, 5% MDP             |
| Buffer      | Na-HEPES, 0.1 M                |
| Additive    | Hexamine cobalt (III) chloride |
| pH          | 7.5                            |

#### 3.4.1 Crystallization of EcoP1/DNA complex with AMPPNP

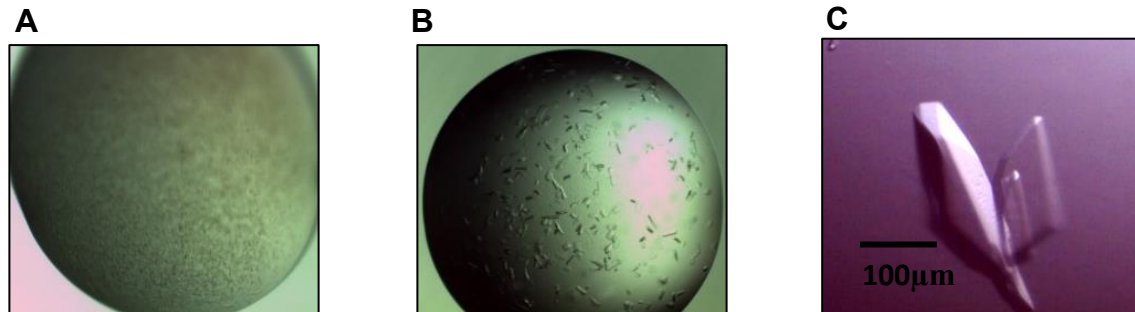
From about 1440 different commercially available conditions screened only one condition gave crystal hits (Figure 3.3 A). The composition of the condition where initial crystal hits appeared are given in the Table 3.6.

**Table 3.6:** Composition of initial condition that gave initial crystals.

|                 |                  |
|-----------------|------------------|
|                 | Parent condition |
| Precipitant Mix | 12% PEG 3350     |
| Buffer          | 0.1 M, Na-HEPES, |
| pH              | 7                |

The initial crystal hits were further optimized by changing the concentration of the components in the parent condition. The crystals obtained in 96 well plate were small for diffraction studies. All further optimization to improve the quality of crystals were performed on the parent condition. It was noticed that excessive nucleation was prevented on increasing the concentration of buffer to 0.125 M. The crystals were also highly sensitive to the concentration of PEG, with best single crystals appearing at around 11.5% - 13% PEG 3350, above this PEG concentration excessive nucleation was noticed. Below the 11% PEG 3350 concentration no crystal growth was noticed. To increase the size of the crystal, 2  $\mu$ L drops (1  $\mu$ L of Protein/DNA complex + 1  $\mu$ L reservoir solution) were set up in 48 well crystallization plates (Figure 3.3 B). Bigger crystals were obtained

by changing drop ratio of protein to reservoir of 1.2:1(Figure 3.3 C). The volume of reservoir solution was kept at 80  $\mu\text{L}$ .



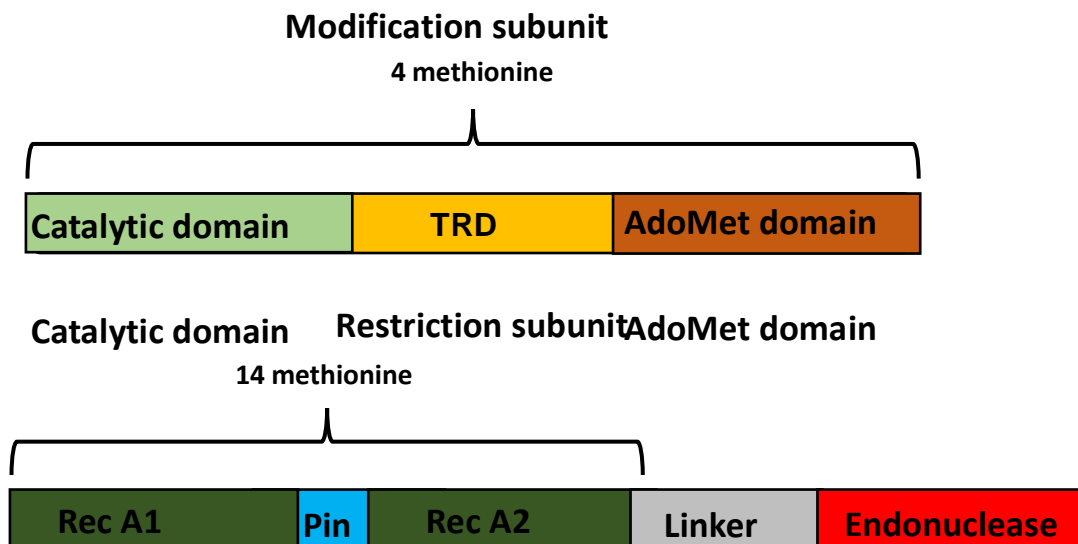
**Figure 3.3. Optimization of condition for co-crystallization of EcoP1I-DNA-AMPPNP.** (A) Initial crystalline hits obtained in 96 well plate. (B) Crystals obtained in the 48 well crystallization plate (C) Isolated and single crystals obtained at Protein/Drop ratio of 1:1.2.

The crystals were flash frozen and it was found that better diffraction was obtained with 30% ethylene glycol as a cryo-protectant. The resolution from the crystals were dependent on the size of the crystal with bigger crystals diffracting better than small crystals. One of the bigger crystal as shown in Fig 3.3 C diffracted to 4.5  $\text{\AA}$  at synchrotron, however the data was highly anisotropic.

#### 3.4.2 Experimental phase determination

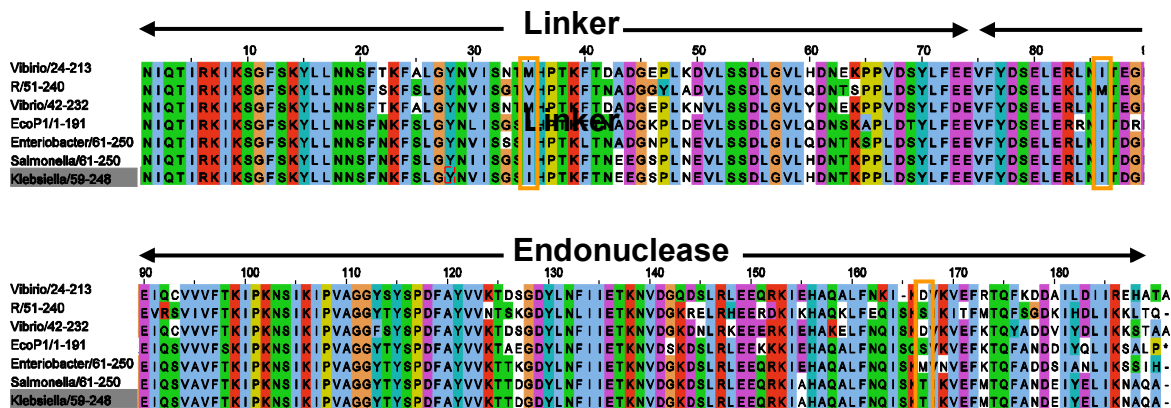
To obtain phase information, we purified the selenomethionine (Se-Met) derivative of the EcoP1I. The Se-Met derivative of the EcoP1I crystallized in the same condition. One of the Se-Met crystals diffracted to 4.5  $\text{\AA}$  at synchrotron source. There are 22 methionines residues in the EcoP1I, with four methionines in the Mod subunit and fourteen methionines in the restriction subunit. Figure 3.4 below shows the position of methionines in EcoP1I. Using single anomalous diffraction we were able to locate the position of 16 methionines, however I could not obtain clear and continuous electron density for the endonuclease domain of the restriction subunit. One of the reasons for lack of clear electron density of the endonuclease domain could be because of poor phasing of this region due to lack of methionine residues in the endonuclease domain. To get phase information of the endonuclease domain, I decided to incorporate methionines residues

in the endonuclease domain of Res subunit. To identify which residues can be changed to methionine I performed a protein blast of endonuclease domain of EcoP1 to find close Type III RM homologues that have methionines residues in their endonuclease domain. Based on sequence alignment we found that there are four possible amino acids in EcoP1I, which can be changed to methionines (Figure 3.5). Using site directed mutagenesis we generated constructs of EcoP1I, with single, double and a triple methionine mutants where methionines were incorporated in the endonuclease domain. The Se-Met derivative protein of double methionine and triple methionine mutant were purified and setup for crystallization. Both double-mutant EcoP1I and triple-mutant EcoP1I crystalized under the same condition. One of the triple mutants diffracted to 4.6 Å. From the triple-mutant diffraction data we were able to pick two additional methionines located in the endonuclease domain. Table 3.7 lists the data collection and processing statistic from the triple methonines mutant crystal.



**Figure 3.4. Position and number of methionines in EcoP1I.** The modification subunit of EcoP1I has four methionines at the N-terminal end. The restriction subunit has fourteen methionines with thirteen methionines are present in the ATPase subunit and one methionine present in the linker.





**Figure 3.5. Positions of incorporation of additional methionines residues in the EcoP1I endonuclease domain.** Sequence alignment of endonuclease domain of EcoP1I with close homologues of other Type III RM enzymes. Yellow boxes showing three positions where methionines can be incorporated. The sequence alignments were performed in Jalview.

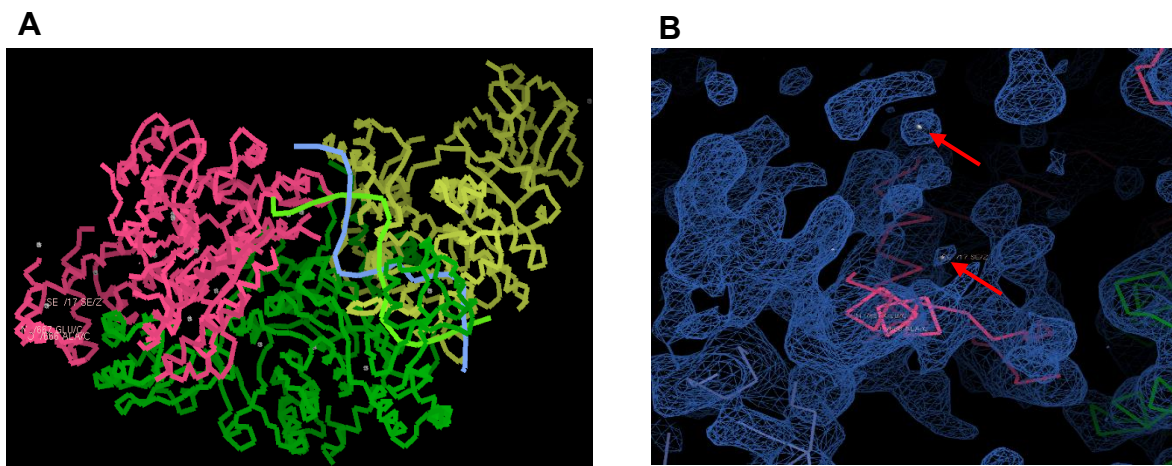
**Table 3.7:** Data collection and processing statistics collected at ESRF ID29

|                                   |                     |
|-----------------------------------|---------------------|
| Space group                       | C2                  |
| Cell dimensions                   |                     |
| a, b, c (Å)                       | 124.1, 100.5, 256.1 |
| $\alpha$ , $\beta$ , $\gamma$ (°) | 90, 93.1, 90        |
| Resolution                        | 50-4.6 (5 - 4.6)    |
| R <sub>merge</sub>                | 0.094 (1.244)       |
| I/ $\sigma$ I                     | 10.6 (2.1)          |
| Completeness (%)                  | 99.7 (99.6)         |
| Anomalous completeness            | 99.6(99.6)          |
| CC <sub>1/2</sub>                 | 0.999 (0.943)       |

### 3.4.3 A partial structure of EcoP1I-DNA-AMPPNP

The selenomethionine crystals of triple methionine mutant of EcoP1I-DNA-AMPPNP were taken for diffraction studies at synchrotron source. One of crystals diffracted to 4.6 Å at ESRF ID29 beam line. Indexing and processing of data was carried out using XDS.

The data was scaled and merged using AIMLESS. The structures of the three subunits of EcoP15I were used as models to obtain molecular replacement solution. The position of selenium in the crystal was located using the SAD data. In all, 18 of the 21 selenium positions were located. The selenium positions and the partial molecular replacement solution were used to obtain SAD phases. The phases yielded a reasonably good quality map with electron density for certain parts of the DNA bound to the enzyme (Figure 3.6 A). The coordinates of the DNA were not present in the partial molecular replacement solution. Of the three selenomethionine sites introduced through mutagenesis in the nuclease domain, two of them could be located (Figure 3.6 B). While the presence of the additional selenium positions allowed demarcation of the electron density corresponding to the nuclease domain, it was not of good enough quality for easy interpretation and model building (Figure 3.6 B). Work is in progress to build the nuclease domain iteratively. At this stage, we have not been able to pick the electron density of AMPPNP. Due to limitation of time, I stopped my efforts on this aspect of the project at this stage.



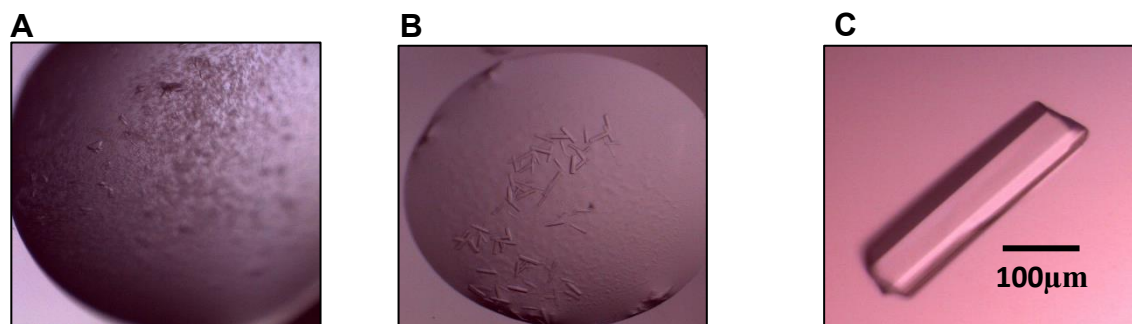
**Figure 3.6. Partial structure of EcoP11-DNA-AMPPNP complex.** (A) A partial structure of EcoP11-DNA complex. Green color represents Mod-A, Pale yellow color represent the Mod-B. The magenta color represents the Res subunit. (B) A simple Fourier map phased using a combination of SAD and partial structure is shown in blue color. Electron density corresponding to the unbuilt nuclease domain is clearly visible. Also highlighted using red arrows are positions of two of three selenomethionines that were incorporated in to the nuclease domain by site directed mutagenesis. The selenium positions were obtained using RESOLVE

### 3.4.4 Crystallization of EcoP1/DNA complex with ADP-VO<sub>4</sub>

A ternary complex of purified EcoP1/DNA with ADP-VO<sub>4</sub> was set up for crystallization by screening 1440 different conditions. Out of 1440 different conditions, three conditions gave possible initial hits. The composition of the conditions is shown in the Table 3.8.

**Table 3.8:** Composition of initial conditions where initial crystals were obtained.

|             | Condition 1           | Condition 2                    | Condition 3                        |
|-------------|-----------------------|--------------------------------|------------------------------------|
| Precipitant | 12% PEG 3350          | 12 % PEG 3350                  | 12% PEG 3350                       |
| Buffer      | 0.1 M Sodium malonate | 0.1 M Ammonium citrate dibasic | 4% Tris(hydroxymethyl)aminomethane |
| pH          | 7                     | 7                              | 6                                  |



**Figure 3.7. Optimization of crystal for co-crystallization of EcoP15I-DNA-ADP VO<sub>4</sub>.** (A) Initial crystal hits obtained in 96 well crystallization plate. (B) Crystals obtained with benzamidine hydrochloride as an additive. Single big crystal obtained with benzamidine hydrochloride as additive and paraffin oil (7%) on top of reservoir buffer.

The best reproducible crystals were obtained with condition three (Figure 3.7 A). To further improve the quality of crystals we carried out additive screen (Hampton) with condition three. With additive screen we got crystals in many conditions. From the additive screen, crystals obtained with benzamidine hydrochloride were single, isolated and bigger in size (Figure 3.7 B). To further increase the size of the crystals, the volume of the crystallization drop was increased to 2  $\mu$ L (1  $\mu$ L protein +1  $\mu$ L reservoir) in 48 well crystallization plates. The crystals appeared in 2-3 days and grew to full size in 7 days. To reduce the nucleation present in the crystallization drop we added paraffin oil (5-10

$\mu\text{L}$ ) on top of the reservoir. The crystals obtained with ADP-VO<sub>4</sub> were bigger than crystals obtained with AMPPNP (Figure 3.7 C). The crystals were flash frozen with ethylene glycol as cryo-protectant, and taken to synchrotron for diffraction studies. Despite the crystals being bigger in size compared to crystals in AMPPNP condition, the crystals diffracted poorly to 5 - 5.5 Å.

### 3.5 Discussion

From our single-site studies we shown that ATP hydrolysis activates these enzyme. To understand the structural basis of how ATP hydrolysis activates these enzymes we carried out the crystallographic studies on EcoP1I and EcoP15I with their specific DNA and in presence of non-hydrolysable ATP analogues.

#### 3.5.1 Crystallization studies of EcoP15I

Our DNA binding assays indicated that EcoP15I can bind to a single-site DNA substrate. Based on these assays we designed a 36 bp DNA substrate containing a single recognition site of EcoP15I. We were able to crystallize EcoP15I with its specific DNA. To improve the size and quality of crystals we carried out additive screen, with best crystal obtained using 0.1 M phenol as an additive. Although the crystals looked promising, they diffracted poorly to 26 Å. As an alternative strategy I crystallized EcoP15I with specific DNA and in presence of AMPPNP and ADP-vanadate. I was able to crystalize EcoP15I-DNA with ADP-vanadate. One of the crystals of EcoP15I with ADP-vanadate diffracted to only 9 Å. The crystals were small in size, and attempts to improve the size of crystals were in vain.

#### 3.5.2 Crystallization of EcoP1I with DNA in presence AMPPNP and ADP-vanadate

EcoP1I and EcoP15I are homologous proteins. EcoP1I as a binary complex was already crystallized in our lab. The binary crystals of EcoP1I with specific DNA had diffracted to 4.4 Å, however because of the lack of phase information we were not able to build the structure (22). To get phase information and to get structural insights which are accompanied on ATP binding and hydrolysis, I crystallized Se-Met derivative of EcoP1I

with DNA in presence of AMPPNP and ADP-vanadate. The crystals of EcoP1I-DNA with AMPPNP diffracted to 4.5 Å. The phase information obtained was used to build the model. To get additional phase information from region of proteins having poor phasing I incorporated additional methionines in the linker and endonuclease domain of EcoP1I. One of the triple methionine mutants of EcoP1I diffracted to 4.6 Å at synchrotron source. We were able to pick electron density of additional two methionines. The additional electron density obtained was used in building the model of EcoP1I. I also crystallized EcoP1I-DNA as a ternary complex with ADP-vanadate. The crystals were optimized using additive screen with best crystals were obtained using benzamidine hydrochloride. Diffraction studies were also carried out, however the crystals diffracted only 5.5 Å at synchrotron source.

In conclusion I carried out crystallographic studies on EcoP1I and EcoP15I the two prototypes of Type III RM enzymes. Our biochemical studies led us to identify the role of ATP at different stages of nucleolytic reaction. I tried to understand the structural basis of how ATP hydrolysis activates these enzyme using tools of crystallography. These enzymes are large molecular weight complex with MW close to 260 kDa. It is challenging to crystallize such large MW complex as such large proteins contain flexible regions and unstructured region. I successfully crystallized EcoP15I with 5/25\_P15I as a binary complexes and in presence of ADP-vanadate. Phase information was obtained by crystallizing Se-Met derivative EcoP1I with 5/26\_P1I in presence of AMPPNP and ADP-Vanadate. The crystals with AMPPNP diffracted to 4.5 Å. The SAD data thus obtained is used in building the structure along with the partial structure of EcoP15I as a molecular replacement model. To build the complete structure a high resolution data is needed. One reason which has prevented us from obtaining high resolution structure is the high anisotropy in our diffraction data.

An alternative strategy to get structural insights about Type III RM enzyme will be using the CryoEM technique. The high MW of these enzymes make them a suitable candidate for CryoEM studies. Our results from single site DNA cleavage suggests, that it happens only in presence of ATP and was abolished on using ATP binding and the hydrolysis-deficient mutant of the enzyme. At this stage, the structural information about

the conformation changes accompanied upon ATP hydrolysis is missing. A CryoEM structure of DNA substrate with the single recognition site of EcoP1I and EcoP15I with a combination nuclease dead, Walker A and B mutants of EcoP1I and EcoP15I will provide crucial mechanistic insights about the conformational changes happening during ATP hydrolysis. Such a mega complex will be an ideal candidate for CryoEM techniques.

Single site DNA cleavage via 3D diffusion or two-site DNA cleavage by 1D diffusion happens only when two enzymes collide. From Chapter 2, we know that a single ATP activated enzyme still bound to DNA on its own cannot perform a nick or dsDNA cleavage. These results indicate that dimerization of endonucleases is essential for DNA cleavage to happen. What conformational changes happen, when two endonucleases dimerize either via 3D diffusion or 1D diffusion, is not known. A CryoEM structure, of EcoP1I and EcoP15I with a DNA substrate with single-site recognition sites of each enzyme, and in presence of non-hydrolysable ATP analogues will give insights how dimerization of endonuclease domain happen. The structure will provide insights into how the dimerization endonuclease domains make the enzymes proficient to perform nick or dsDNA break.

In short, the present study had provided a provided strong platform for the future work, which will provide the crucial insights about the working mechanism of Type III RM enzymes.

## References.

1. Rao,D.N., Dryden,D.T.F. and Bheemanaik,S.(2014) Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.*, **42**, 45-55.
2. Meisel,A., Bickle,T.A., Kruger,D.H. and Schroeder.C (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature*. **355**, 467-469.
3. Miesel.A., Mackeldanz.P., Bickle.T.A., Kruger,D.H. and Schroeder.C (1995) Type III restriction enzymes translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. *EMBO J.*, **14**, 2958-2966.

4. Ramanathan,S.P., van Aelst,K., Sears,A., Peakman,L.J., Diffin,F.M., Szczelkun,M.D., and Seidel, R.(2009) Type III restriction enzymes communicate in 1D without looping between their target sites. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 1748-1753.
5. Schwarz,F.W., Toth,J., van Aelst,K., Cui,G., Clausing,S., Szczelkun,M.D. and Seidel,R.(2013) The Helicase-Like Domains of Type III Restriction Enzymes Trigger Long-Range Diffusion Along DNA. *Science*, **340**, 353–356.
6. Toth,J., Bollins,J., and Szczelkun,M.D. (2015) Re-evaluating the kinetics of ATP hydrolysis during initiation of DNA sliding by type III restriction enzymes. *Nucleic Acids Res.*, **43**, 10870-10881.
7. Mucke,M., Reich,S., Moencke-Buchner,E., Reuter,M. and Kruger, D.H. (2001) DNA cleavage by type III restriction-modification enzyme EcoP15I is independent of spacer distance between two head to head oriented recognition sites. *J. Mol. Biol.*, **312**, 687-698.
8. Peakman,L.J., Antognozzi,M., Bickle,T.A., Janscak,P. and Szczelkun,M.D. (2003) S-adenosyl methionine prevents promiscuous DNA cleavage by the EcoP1I type III restriction enzyme. *J. Mol. Biol.*, **333**, 321-335.
9. Raghavendra,N.K. and Rao,D.N. (2004) Unidirectional translocation from recognition site and a necessary interaction with DNA end for cleavage by type III restriction enzyme. *Nucleic Acids Res.*, **32**, 5703-5711.
10. Raghavendra,N.K. and Rao,D.N. (2005). Exogenous AdoMet and its analogue sinefungin differentially influence DNA cleavage by R.EcoP15I- Usefulness in SAGE. *Biochem. Biophys. Res. Commun.*, **334**, 803-811.
11. Moncke-Buchner,E., Rothenberg,M., Reich,S., Wagenfuhr,K., Matsumura,H., Terauchi,R., Kruger,D.H., and Reuter,M. (2009) Functional characterization and modulation of DNA cleavage efficiency of type III restriction endonuclease EcoP15I in its interaction with two sites in the DNA target. *J. Mol. Biol.*, **387**, 1309-1319.
12. Peakman,L.J. and Szczelkun,M.D. (2009). S-Adenosyl homocysteine and DNA ends stimulate promiscuous nuclease activities in the type III restriction endonuclease EcoP1. *Nucleic Acids Res.*, **37**, 3934-3945.

13. Butterer, A., Pernstich, C., Smith, R.M., Sobott, F., Szczelkun, M.D. and Toth, J. (2014). Type III restriction endonucleases are heterotrimeric: comprising one helicase-nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Res.*, **42**, 5139–5150.
14. Gupta, Y.K., Chan, S.H., Xu, S.Y., and Aggarwal, A.K. (2015) Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat Commun.*, **6**, 7363.
15. Kabsch, W. (2010) XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D66**, 125–132.
16. Evans, P.R. and Murshudov, G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D69**, 1204–1214.
17. Evans, P. (2006) Scaling and assessment of data quality. *Acta Cryst*, **62**, 72–82
18. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 486–501.
19. Read, R.J. and McCoy, A.J. (2011) Using SAD data in Phaser. *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, **67**, 338–344.
20. Terwilliger, T. (2004) SOLVE and RESOLVE: automated structure solution, density modification, and model building. *J. Synchrotron Rad.* (2004). **11**, 49–52.
21. Adams, P.D., Afonine, P. V, Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 213–221.
22. Kulkarni, M (2017) Biochemical, bioinformatics and crystallographic analysis of type III restriction-modification enzyme EcoP1I (Doctoral dissertation)



**CHAPTER 4**  
**Identification and biochemical characterization**  
**of Type III RM enzyme MbolIII from *Mycoplasma***  
***bovis***

## Chapter 4

### Identification, characterization and biochemical studies on MboIII

#### 4.1 Introduction

Restriction-modification (RM) enzymes are bacterial defense against invading foreign DNA, such as bacteriophage DNA. Based on their cofactor requirement, subunit assembly and the mode of endonucleolytic DNA cleavage RM enzymes are classified as Type I, Type ISP, Type II, Type III and the modification-dependent Type IV restriction enzymes (17). Type III RM enzymes are ATP-dependent and are composed of two subunits. The modification (Mod) subunit through its methyltransferase activity methylates the host DNA at specific sites giving a mark of self. The restriction (Res) subunit on forming a complex with Mod cleaves unmodified DNA in the presence of ATP. Although more than >15000 putative Type III R-M enzymes have been identified (<http://rebase.neb.com/cgi-bin/azlist?re3>) only a handful of them have been characterized, with the enzymes EcoP1I and EcoP15I serving as the prototypes (18). Endonucleolytic cleavage of unmethylated DNA requires two invertely oriented sites that can be up to 3.5 kb apart with cleavage happening 25-27 bp downstream to any one target site.

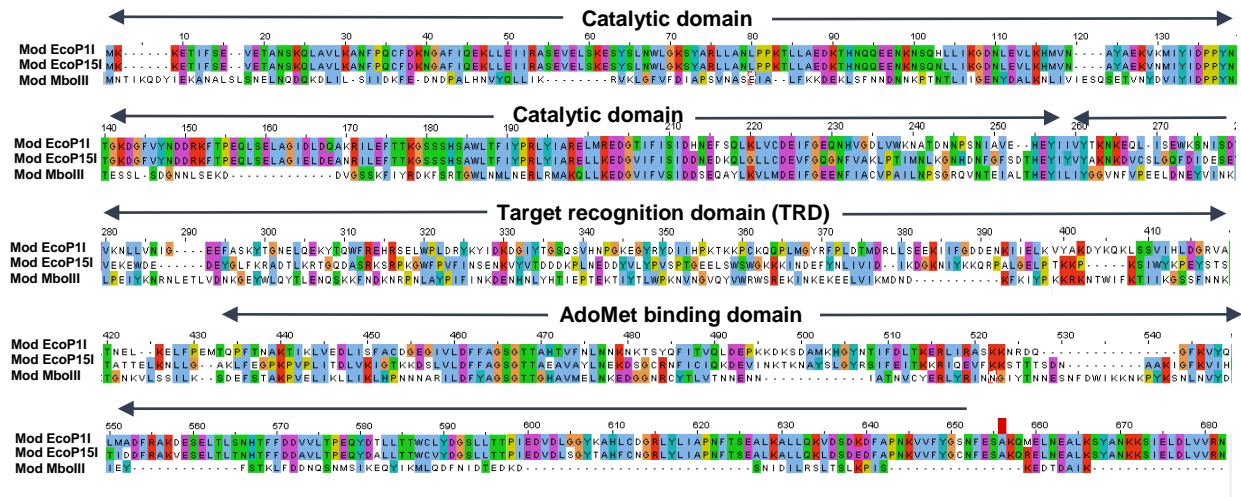
An interesting feature of these enzymes, that remained poorly understood is their low ATP consumption compared to Type I RM enzymes. Based on magnetic tweezers and single molecule studies combined with TIRF microscopy, an ATP-dependent molecular switch mechanism similar to that once observed in mismatch repair protein has been proposed. According to the model, on binding of the enzymes to their recognition sites, 10-12 ATP molecules are hydrolyzed (3, 4, and 5). The enzyme then undergoes a drastic conformational change, by virtue of which it undergoes 1D diffusion along the DNA. The DNA cleavage happens when a diffusing enzyme collides with a stationary enzyme bound to its recognition sequence (3, 4, and 5). Although a DNA substrate with pair of recognition sites in inverted orientation are a necessity for the DNA cleavage, there were reports in a literature about single-site DNA substrates getting cleaved. From our biochemical studies (Chapter 2), we were able to dissect the molecular mechanism how these enzymes perform a single-site DNA cleavage. Our biochemical studies not only

revealed role of the ATP as an activator but also unveiled how ATP hydrolysis activates these enzymes to perform an endonucleolytic cleavage. To understand the structural basis of how ATP activates these enzymes, we carried our crystallographic studies of EcoP1I and EcoP15I with their specific DNA and in the presence of non-hydrolysable ATP analogues. However, our crystallographic studies on EcoP1I and EcoP15I failed to obtain a high resolution structure of these enzymes. One reason why I was abortive could be because of the large molecular weight (MW) of these proteins. Type III RM enzymes are macro molecular proteins with MW ~260 kDa, such huge protein contain a large flexible region and an unstructured region, which may have hindered our attempts to obtain a high resolution diffraction data.

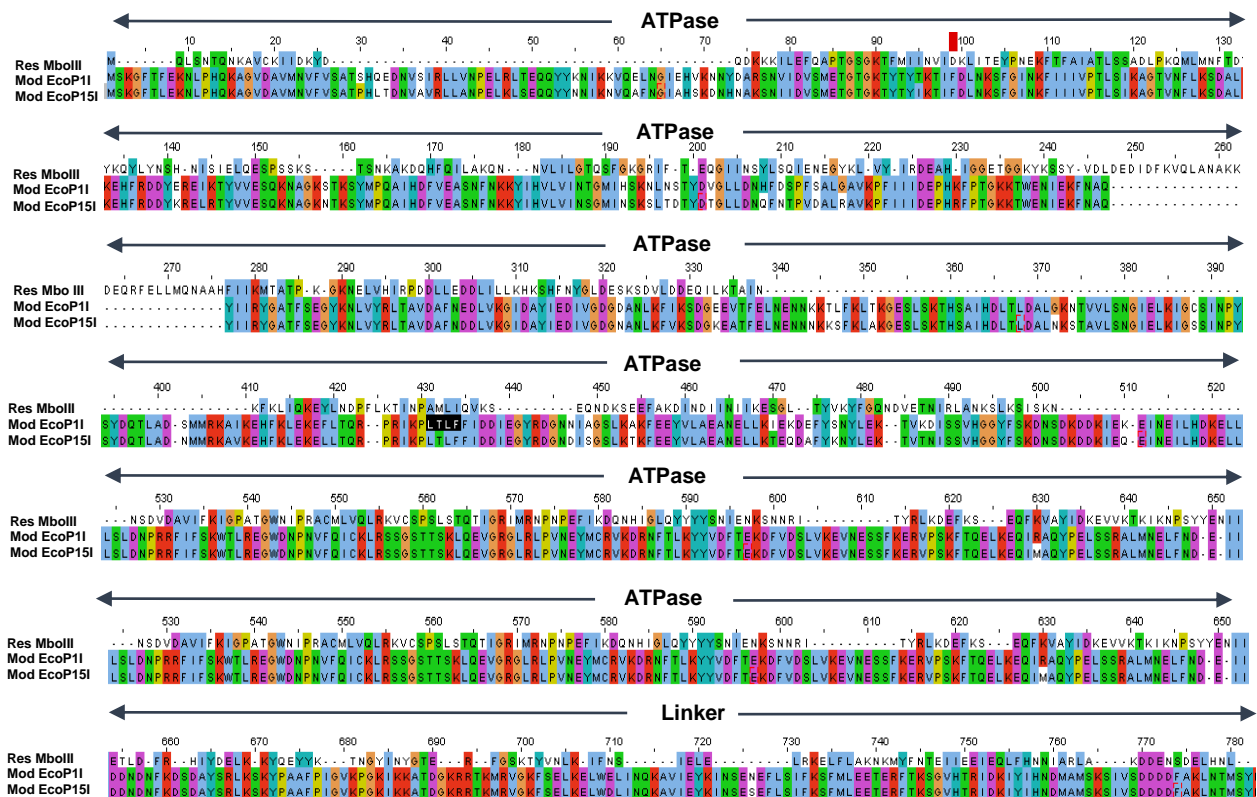
As an alternative strategy, I started to look for another model system of Type III RM enzymes, with an idea to find out the Type III RM with the smaller MW, comparable to EcoP1I and EcoP15I. A careful analysis of the genome of *M. bovis* Donetta PG 45, revealed the presence of Type III RM enzyme which we named as MboIII. On the basis of amino acid composition, the MW of MboIII is ~227kDa (a smallest Type III RM to our knowledge). Figure 4.1 A represents the sequence alignment of Mod subunits of EcoP1I, EcoP15I and MboIII. The sequence alignment indicates that the C-terminal end of the Mod subunit of MboIII is relatively shorter compared to EcoP1I and EcoP15I. Similarly, the sequence alignment of Res subunits between the three enzymes shows deletions in the Res subunit of MboIII (Figure 4.1 B). These differences accounts for a smaller MWs of the Mod and Res subunits of MboIII.

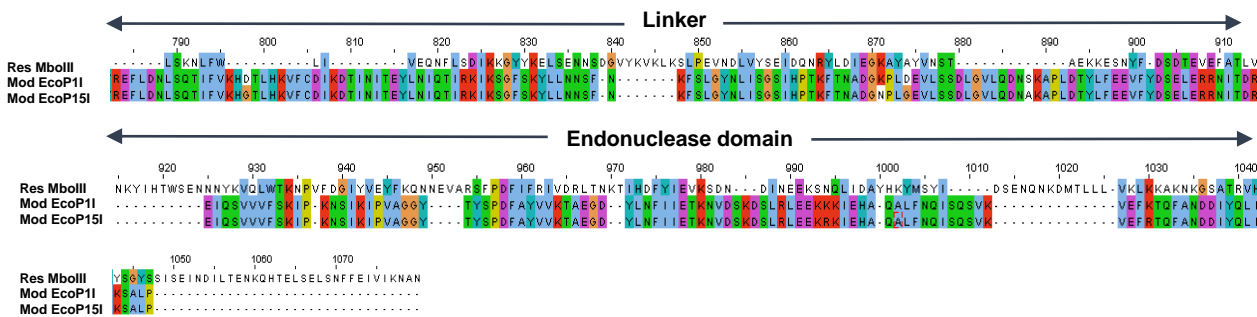
I isolated, purified and characterized the oligomeric assembly of the active MboIII enzyme complex. The DNA recognition sequence of the enzyme was identified, and its methylation and nucleolytic activities were characterized. Based on the gene sequence from GenBank, the gene encoding MboIII Mod has a simple sequence repeats (SSRs) made of 11 AG repeats located close to the catalytic motif DPPY, which results in a frame-shift and hence turning-off of the gene in the sequenced strain clone. The presence of SSRs in Type III RM enzymes results in Phase variation.

**A** Sequence alignment between the Mod subunits of EcoP11, EcoP15I and MboIII



**B** Sequence alignment between the Res subunits of EcoP11, EcoP15I and MboIII





**Figure 4.1. Sequence alignment between EcoP11, EcoP15I and MbolII subunits.**

(A) Sequence alignment of the Mod subunits from EcoP11, EcoP15I and MbolII. (B) Sequence alignment between the Res subunits from EcoP11, EcoP15I and MbolII.

Phase variation is the phenomenon of high-frequency reversible on-off switching of a gene, often observed in host-adapted bacterial pathogens (6). It is a mean of introducing phenotypic diversity and increasing pathogenicity. The regulation is often achieved by SSRs present within either the open reading frame or promoter region of a phase variable gene (7, 8). Addition or subtraction of repeats occurs at SSR at a high rate probably due to slipped-strand mispairing during replication (9). While the importance of SSRs in regulating phase-variation in prokaryotes is well documented, the effect of variation on the functional properties of prokaryotic gene products is generally unknown. A prominent set of examples is the variation in repeat numbers within the coding region of certain genes resulting in neurodegenerative diseases, where pathology is caused by aberration in the function of the encoded protein due to an increase in number of repeats (10). Other diverse examples include, variation in the repeat length modulating cell adhesion strength of Flo1 protein (11) and hydrophobicity of Flo11 (12) in *Saccharomyces cerevisiae*; variation in repeat numbers controlling and tuning the circadian clock by modulating the function of the encoded gene product (13,14). SSRs mediated phase variation has been observed in prokaryotic genes involved in production of surface antigens (15, 16), which facilitate colonization of the host, adaptation to host environment and evasion of the host immune system. Another frequently observed class of phase variable genes encode for the methyltransferase component of restriction-modification (RM) systems, in particular the Type III RM enzymes (17-19).

Using MboIII as a model, I tried to address the question whether the presence of SSRs has any effect on the activity of MboIII, we characterized and compared the activities of the full length MboIII having either 9 and 24 repeats. We find that the activities of MboIII are not affected by the shorter or longer SSR, despite being located close to the catalytic region and the interface with Res. Furthermore, I find that the site of the SSR is such that it can accommodate even a large polypeptide, enhanced green fluorescence protein (EGFP), without affecting the activities of the enzyme. An inspection of SSRs in methyltransferases of Type III RM enzymes from other species of *Mycoplasma* and other pathogens, suggests that the repeats are located at sites endowed with structural plasticity. This possibly allows the sites to accommodate the repeats without affecting their activities, as was noted in case of MboIII.

## 4.2. Materials and methods.

### 4.2.1 oligos used for generation of dsDNA substrates

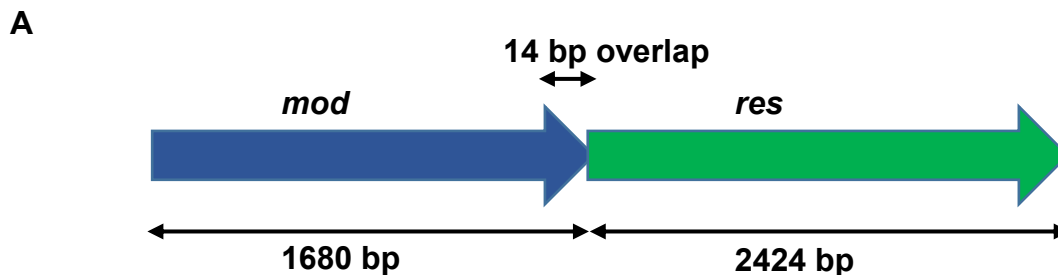
All oligonucleotides were purchased from Sigma-Aldrich or Integrated DNA Technologies. A DNA substrate containing two recognition sites of MboIII oriented head-to-head was generated using primers 5'-GGTGATGTACGAAGAGGAGTTCACCATAATCAACGCC GTTTG-3' and 5'-GTGGAAGTGGTTGGGGAAGTTCCTCCCTAATCAGCTTGGC-3' (MboIII sequence underlined). The resulting 235 bp DNA substrate was PCR purified and concentration was determined using Thermo Scientific NanoDrop spectrophotometer.

A 650 bp DNA containing a single recognition site was used to generate DNA substrates containing head-to-head, head-to-tail and tail-to-tail substrate using primers 5'-GTCTGTATTTGATTACACCTAATTTTAC-3', 5'-CAAAGAATGTATTAATCGTGAGCG TAAG-3', 5'-GTCTGTATTTTAATCCACCTAATTTTAC-3', 5'-CAAAGAATGTATTAATC GTGAGCGTAAG-3', 5'-GTCTGTATTTTAATCCACCTAATTTTAC-3' (MboIII sequence underlined). The list of various primers used in the generation of dsDNA substrates for the nucleolytic assay are shown in Table 4.1.

### 4.2.2 Isolation and cloning of MboIII

I amplified the *mod* and *res* genes as a single operon and as individual genes from the genomic DNA of *M. bovis* Donetta PG 45 obtained from DSMZ (Catalogue no. 22781). *M. bovis* uses TGA as the codon for tryptophan, which is a stop codon in *Escherichia coli*. Heterologous expression in *E. coli* of the *res* and *mod* genes from *M. bovis* required the mutation of all thirteen TGA tryptophan codons to TGG to express the full-length protein. In addition, the SSR was deleted by an overlap PCR. All the primers used for cloning and the mutagenesis are listed in Supplementary Table 2. Subsequent to this, *mod* and *res* as individual genes were cloned in high expression vectors pRSF and pHIS, respectively, as mentioned below.

*mod* gene was amplified using Primer 1: 5'-GATATACCATGGCACATCACCACC ACCATCACATGAATACTATAAAACAG-3' and Primer 2: 5'-CGCAGCAGCGGTTTC TTTACCCTCGAGTTACTTAATTGCATCAGT-3'. The amplified gene was then double digested using the restriction enzymes NcoI and XhoI, and ligated into pRSF vector using NEB T4 DNA ligase. *E. coli* electrocompetent cells were transformed with the ligation mixture using the method of electroporation. The recombinant vector pRSF*mod* was confirmed by DNA sequencing. Similarly, the *res* gene with all internal TGA codons changed to TGG was PCR amplified using Primer 3: 5'-GTTTAACTTTAAGAAG GAGATATACATATGCAATTAAGTAATACAC-3' and Primer 4: 5'-GTGTATTACTTAATT GCATATGTATATCTCCTTCTTAAAGTTAAAC-3'. The amplified *res* gene was double digested with NdeI and BamHI restriction enzyme and ligated into the pHIS vector (20). *E. coli* electrocompetant cells were transformed with the ligation mixture using the method of electroporation. The positive clones were confirmed only after the complete sequencing of the gene. The *MbolIII* operon was cloned into pRSF vector with an N-terminal 6XHistag as using Primer 1: 5'-GATATACCATGGCACATCACCACCACCATCACATGAATACTAT AAAACAG-3' and Primer 5: 5'-CGCAGCAGCGGTTTCTTTACCCTCGAGTTAATTAG CTTTTTTTATTA-3'. The amplified insert was double digested with NcoI and XhoI restriction enzymes and ligated into pRSF vector using the similar method as for *mod* and *res* gene. Figure 4.2 represents the detailed organization of the *MbolIII* operon.

**Gene organisation of MbolIII**

**B**

| <u>Name of insert</u> | <u>Protein</u> | <u>MW</u> |
|-----------------------|----------------|-----------|
| <i>mod</i>            | Mod            | 64 kDa    |
| <i>res</i>            | Res            | 94 kDa    |
| <i>MbolIII</i>        | MbolIII        | 224 kDa   |

**Figure 4.2. Gene organisation of *MbolIII*, *mod* and *res*.** The *mod* gene shown blue in colour is 1680 bp in length. The *res* gene shown in green colour is 2424 bp in length, overlaps with *mod* gene by 14 bp.

**Table 4.1:** Oligomers used in cloning and mutagenesis of *MbolIII*

| Sr.No | Name             | Sequence (5' - >3')                                    |
|-------|------------------|--------------------------------------------------------|
| 1     | Primer 1         | GATATACCATGGCACATCACCACCACCATCACATGAA<br>TACTATAAAACAG |
| 2     | Primer 2         | CGCAGCAGCGGTTTCTTTACCCTCGAGTTACTTAATTG<br>CATCAGT      |
| 3     | Primer 3         | GTTTAACTTTAAGAAGGAGATATACATATGCAATTAAG<br>TAATACAC     |
| 4     | Primer 4         | GTGTACTTAATTGCATATGTATATCTCCTTCTTAAA<br>GTAAAC         |
| 5     | MbolIIIRepeatFor | CTTATAGTGATCGAGTCACAGTCAGAGACTG                        |
| 6     | MbolIIIRepeatRev | CAGTCTCTGACTGTGACTCGATCACTATAAG                        |
| 7     | MbolIII476For    | CAAATTTATCTATCGTGGTAAATTTAGTCGTAAGTGG                  |





**Table 4.2:** Oligomers used to form duplex substrates for nuclease assay

| Sr.No | Name        | Sequence (5'->3')                           |
|-------|-------------|---------------------------------------------|
| 1     | For Pri 235 | GGTGATGTACGAAGAGGAGTTCACCATAATCAACGCCGTTTG  |
| 2     | Rev Pri 235 | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTAATCAGCTTGGC |
| 3     | Rev Pri 1   | GTGGAACTGGTTGGGGAACCTCGTTCTCCCAATCAGCTTGGC  |
| 4     | Rev Pri 2   | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTGATCAGCTTGGC |
| 5     | Rev Pri 3   | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTAGTCAGCTTGGC |
| 6     | Rev Pri 4   | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTAACCAGCTTGGC |
| 7     | Rev Pri 5   | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTCATGAGCTTGGC |
| 8     | Rev Pri 6   | GTGGAACTGGTTGGGGAACCTCGTTCTCCCGAATCAGCTTGGC |
| 9     | Rev Pri 7   | GTGGAACTGGTTGGGGAACCTCGTTCTCCCAATCAGCTTGGC  |
| 10    | Rev Pri 8   | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTTATCAGCTTGGC |
| 11    | Rev Pri 9   | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTATTCAGCTTGGC |
| 12    | Rev Pri 10  | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTACTCAGCTTGGC |
| 13    | Rev Pri 11  | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTAAACAGCTTGGC |
| 14    | Rev Pri 12  | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTAAGCAGCTTGGC |
| 15    | Rev Pri 13  | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTAATAAGCTTGGC |
| 16    | Rev Pri 14  | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTAATGAGCTTGGC |
| 17    | Rev Pri 15  | GTGGAACTGGTTGGGGAACCTCGTTCTCCCTAATTAGCTTGGC |
| 18    | Primer 5    | GTCTGTATTTGATTACACCTAATTTTAC                |

|    |          |                              |
|----|----------|------------------------------|
| 19 | Primer 6 | CAAAGAATGTATTAATCGTGAGCGTAAG |
| 20 | Primer 7 | GTCTGTATTTTAATCCACCTAATTTTAC |
| 21 | Primer 8 | CAAAGAATGTATTAATCGTGAGCGTAAG |
| 22 | Primer 9 | GTCTGTATTTTAATCCACCTAATTTTAC |

#### 4.2.3 Primers used for the complete sequencing of *mod*, *res* and *MbolIII*

On an average, the read out from one sequencing reaction using the Sanger sequencing method is about 800 bp. Modification subunit has a length of 1680 bp and was sequenced using T7 promoter and T7 terminator primers. Restriction subunit is 2424 bp long, while *MbolIII* operon is 4090 bp long. So apart from T7 promoter and T7 terminator, primers which are complementary to internal region of operon were designed. The list of various primers used to get complete sequencing results of *res* subunit and *MbolIII* are listed in Table 4.3.

**Table 4.3:** Primers used for complete sequencing of *MbolIII*

| Sr.No | Name         | Sequence (5'-> 3')                                      |
|-------|--------------|---------------------------------------------------------|
| 1     | T7 Promoter  | TAATACGACTCACTATAGGG                                    |
| 2     | T7Terminator | GCTAGTTATTGCTCAGCGG                                     |
| 3     | Mbo821For    | GATAATAAGGGGGAATATTGGTTGCAATATACTTGAG                   |
| 4     | Mbo1448For   | GAATCTAATTTTGATTGGATTAAGAAAAACAAG                       |
| 5     | Mbo2781Rev   | GCTCTAGGTATGTTCCAGCCAGTAGCTGGCC                         |
| 6     | Mbo3377Rev   | CAAAAACAGGATTTTTTTGTCCAGAGTTGCACTTTATAA<br>TTATTATTTTCG |

#### 4.2.4 Purification of MbolIII

Recombinant vectors pRSF*mod* and pHIS*res* were co transformed into BL21(DE3) cells with both ampicillin (50µg/ml) and kanamycin (50µg/ml) as selection markers. A 6 L of 1xLB was grown at 37°C with constant shaking at 250 rpm. The culture was grown until the OD<sub>600</sub> reached 0.6 after which the culture was shifted to 18°C and induced with 2g/L of L-arabinose and 1 mM IPTG. The culture was harvested after 12 h, pelleted down and re-suspended into 200 mL of lysis buffer [50 mM Tris-Cl pH 8, 500 mM NaCl, 10% glycerol, 10 mM MgCl<sub>2</sub>, and 1 mM DTT]. The re-suspended culture was lysed using sonication at 4°C, with 1 sec ON and 3 sec OFF and the sonication cycle was repeated two twice. The lysate was spun at 37000 rpm at 4°C for 1 h using Optima XE ultracentrifuge (Beckman Coulter). The resulting supernatant was loaded on to the 5mL Ni-NTA column pre-equilibrated with Buffer A [ 50 mM Tris-HCl pH 8, 500 mM NaCl, 15 mM imidazole]. To get rid of impurities bound non-specifically to the column, a 4 column volume (CV) wash with 3% and 8% Buffer B [50 mM Tris-HCl pH 8, 500 mM NaCl, 500 mM imidazole] was performed. The protein was eluted with increasing concentration of imidazole in a step wise gradient using a mix of Buffer A and B. The eluted fraction containing protein were pooled together and dialyzed against B-50 (50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM DTT) for 6 h at 4°C. The dialyzed protein was centrifuged for 20 mins using Avanti J-26XP (Beckman Coulter Life Sciences) at 4°C. The dialyzed protein was further purified using Mono Q 10/100 GL column (GE Healthcare) pre equilibrated with B-50 buffer. MbolIII eluted from the column with an increasing linear gradient of NaCl (50 mM-500 mM). MbolIII eluted from Mono Q 10/100 GL column at ~0.28 M NaCl. Size exclusion chromatography (SEC) using Superdex 200 10/300 (GE Healthcare) was carried out as a last step to check the homogeneity of MbolIII. The fractions containing pure protein was pooled, concentrated using Viva spin column (MWCO 30kDa, GE healthcare) and stored at -80°C until further use. Activity test of the protein was not carried out during the steps of purification. The purified sample was used for all the subsequent enzymatic assays.

#### 4.2.5 DNA Cleavage assay

DNA cleavage assay was carried out in cleavage buffer [10 mM Tris-acetate pH 8, 10mM potassium acetate, 10 mM MgCl<sub>2</sub>, 1 mM DTT] at 37°C. Protein and DNA were mixed together in presence of 20 µM sinefungin and reaction was started with the addition of 4 mM ATP. The reaction was further carried out for 45 mins and stopped with the addition of 0.5X volume of STOP buffer (10 mM Tris-HCl pH8, 100 mM EDTA, 60% w/v glycerol, 0.025% SDS, 0.03% bromophenol blue). The samples were heated at 65°C for 15 mins to inactivate the enzyme and then loaded on to a 12% native PAGE gel.

#### 4.2.6 Identification of the MboIII cleavage loci

To locate the position of the DNA cleavage by MboIII, a 1 kb DNA substrate (50 nM) containing two recognition sites of MboIII in head-to-head orientation was treated with MboIII (300 nM) in cleavage buffer. The reaction was started with the addition of 4 mM ATP and carried for 1 h at 37°C. The products of DNA cleavage were isolated from agarose gel, purified using QIAGEN gel extraction kit and were analyzed by run-off sequencing.

#### 4.2.7 Identification of target adenine in the recognition site of MboIII

Methylation reaction was carried out with 50 nM of plasmid with 300 nM of Mod subunit in the presence of 200 µM S-adenosyl methionine (AdoMet). Reaction was carried out at 37°C for 1 h. The plasmid DNA was purified using Qiagen PCR purification kit. Extent of methylation was confirmed by subjecting the purified DNA to cleavage by using MboIII. Absence of any cleavage suggested full methylation. The methylated plasmid was PacBio sequenced.

Recombinant plasmid pRSF*mod* was transformed into *E. coli* BL21(DE3) cells. A single colony was inoculated into 15 mL 1xLB media containing kanamycin. The culture was grown until an OD<sub>600</sub> of 0.5, after which expression of Mod was induced with 1 mM IPTG and kept overnight at 37°C with constant shaking. The culture was pelleted next morning and genomic DNA isolated by QIAGEN Genomic DNA purification kit. The

purified genomic DNA was PacBio sequenced and methylation analyzed using PacBio software as previously described (21).

#### 4.2.8 Insertion of 9 and 24 AG into *Mod* subunit

MbolIII<sup>AG-9</sup> with 9 AG were incorporated into the *mod* gene by overlap PCR (22). MbolIII<sup>AG-24</sup> with 24 AG repeats were incorporated into the *mod* gene by overlap PCR. Using *mod* gene as a template and Primer 1, RevPri MbolIII<sup>AG-24</sup> (Table ) as a forward and reverse primer respectively a 378 bp amplicon was generated. In the next round of PCR, the 378 bp amplicon was used as a forward primer and Primer 2 as a reverse primer to generate *mod*<sup>AG-24</sup>. The primers used for the overlap PCR are listed in Table 1. The resulting *mod* gene with 9, 24 AG repeats were double digested with NcoI and XhoI restriction enzymes and inserted into pRSF vector by ligation using T4 ligase (New England Biolabs®). Positive clones were confirmed only after sequencing. MbolIII<sup>AG-9</sup>, MbolIII<sup>AG-24</sup> were purified using the same protocol as wild type MbolIII.

#### 4.2.9 Incorporation of EGFP into *Mod* subunit

Enhanced green fluorescent protein (EGFP) was incorporated into the *Mod* subunit by restriction free cloning (23) using two rounds of PCR. In the first PCR, *egfp* gene was amplified using primers having their 5'-end complementary to the flanking regions to the SSR in *mod* gene. Subsequently, the amplified *egfp* gene was used as a mega primer and pRSF*mod* plasmid was used as a template to replace the SSR with the *egfp* gene. The resulting PCR product was treated with DpnI restriction enzyme to digest away the methylated template pRSF*mod*. The resulting reaction mixture was transformed into *E. coli* NEB-Turbo electrocompetent cells by electroporation. Positive clones were confirmed by sequencing the insert.

#### 4.2.10 DNA methylation assay

300 nM of MbolIII, MbolIII<sup>AG-24</sup> or MbolIII<sup>EGFP</sup> were incubated with 50 nM of 235 bp DNA substrate containing two recognition sites in head-to-head orientation at 37°C in presence of 200 µM AdoMet for 1 h. The methylated DNA was purified and extent of DNA methylation was confirmed by carrying out DNA cleavage assay with MbolIII.

#### 4.2.11 Size-exclusion chromatography with multi-angle light scattering

Size-exclusion chromatography in combination with multiple-angle light scattering was used to accurately determine the molecular weight of MboIII. MboIII at 2 mg/mL was loaded on BioRad ENrich™ SEC 650 at the flow rate of 0.3 mL/min. The column was in turn connected to light scattering diode array and differential refractive index detector (Wyatt Technology UK Ltd). Before the start of experiment, BSA was run as a standard marker to check the reproducibility of the results. The chromatographs were analyzed using Graph Pad software.

### 4.3 Results

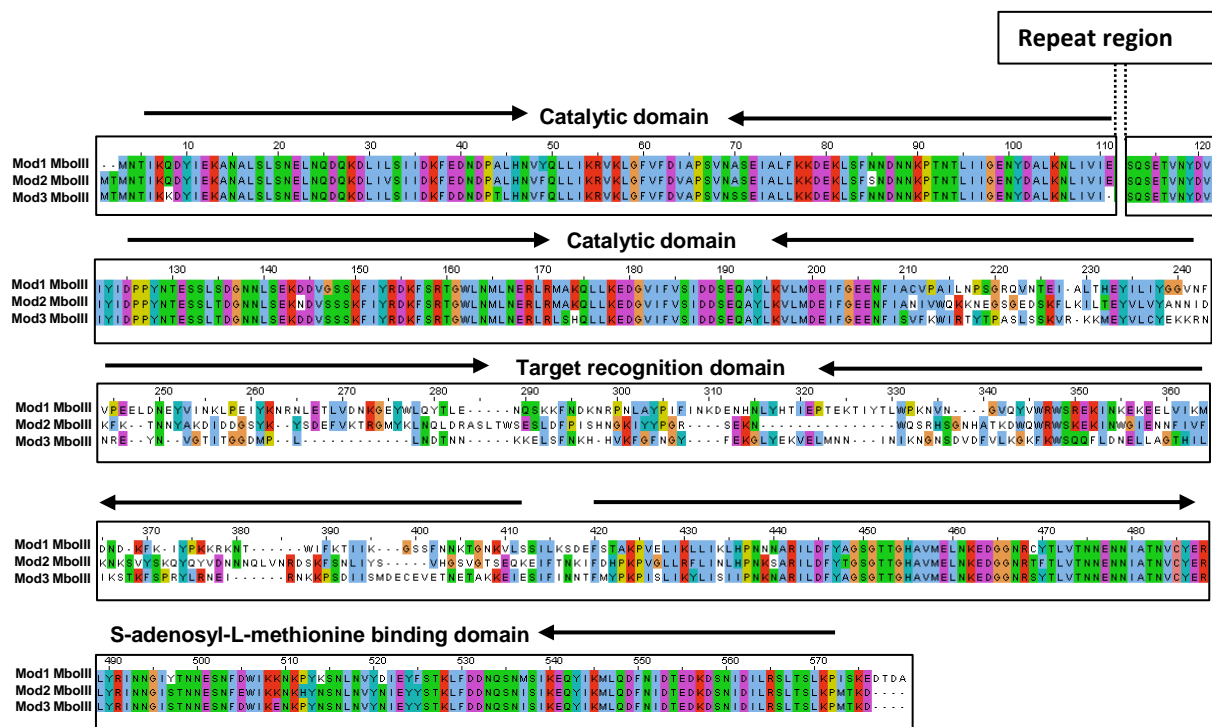
#### 4.3.1 Organization of the genes encoding a Type III RM enzyme in *M. bovis*

Genome analysis of *M. bovis* Donetta PG 45 strain revealed the presence of a *res* gene and three *mod* genes - *mod1*, *mod2* and *mod3* (locus Mbo45ORF167P in the REBASE database) (Figure 4.3 A). The end of the *mod1* gene overlapped with the first 14 bp of the *res* gene. We amplified the *mod1* and *res* genes together as a single operon, and separately as individual genes from the genomic DNA. A feature of the *mod* gene is the presence of a dinucleotide AG repeat, which can vary in length. Genes having repeat number that is a multiple of three results in full-length product (Figure 4.3 B). *mod1* amplified from the genomic DNA had 11 AG repeats, which would result in premature termination of translation. To obtain a full-length gene product, we generated three variants of *mod1* with 0, 9 (22) and 24 AG repeats. A fourth variant having the sequence encoding EGFP was inserted at the site of the repeat (see below). *mod2* and *mod3* have 14 and 24 AG repeats, respectively.



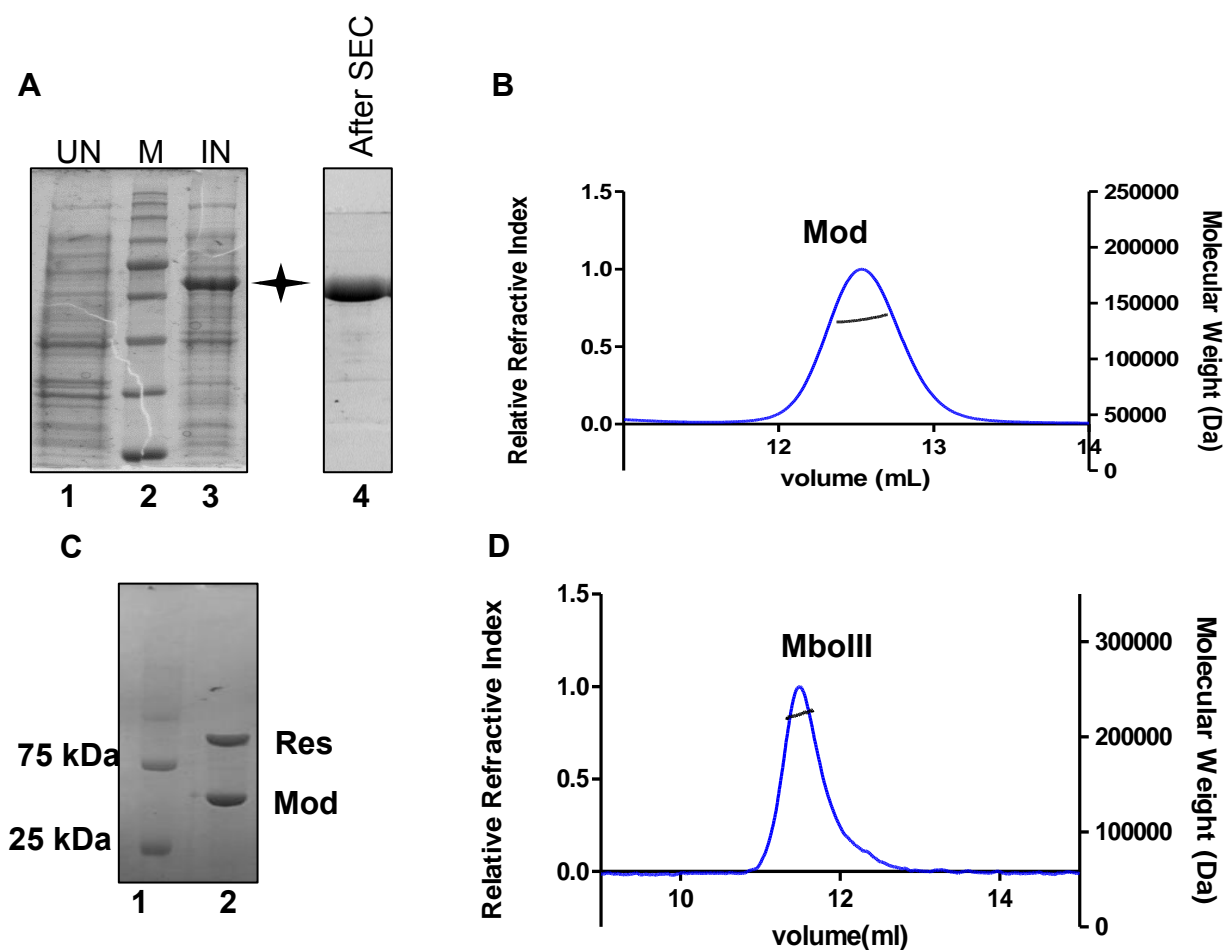


4.4). However, the region corresponding to target recognition domain (TRD) showed significantly less similarity, indicating that the recognition sequence for the three methyltransferases may be different (Figure 4.4).



**Figure 4.4. Sequence alignment of Mod1, Mod2 and Mod3 subunits present in the genomic DNA of *Mycoplasma bovis* Donetta PG 45.** Mod1 and Mod2 share almost 70% sequence identity between them. Similarly, Mod1 and Mod3 have about 68% sequence identity between them.

Mod, the gene product of *mod1* with no SSR (to be referred to as *mod*), could be overexpressed in *E. coli* BL21(DE3) cells and purified to homogeneity (Figure 4.5 A). SEC-MALS analysis of Mod yielded a molecular mass of 131 kDa corresponding to a dimeric protein (Figure 4.5 B). The holoenzyme MboIII, was purified by coexpression of *mod* and *res* (Figure 4.5 C). Though MboIII could be reconstituted *in vitro* by mixing Mod and Res we used the complex purified by coexpression for our studies. SEC-MALS analysis of MboIII yielded a molecular mass of 227 kDa corresponding to a MboIII as heterotrimeric complex composed of two Mod protomers and a single Res subunit (Figure 4.5 D).

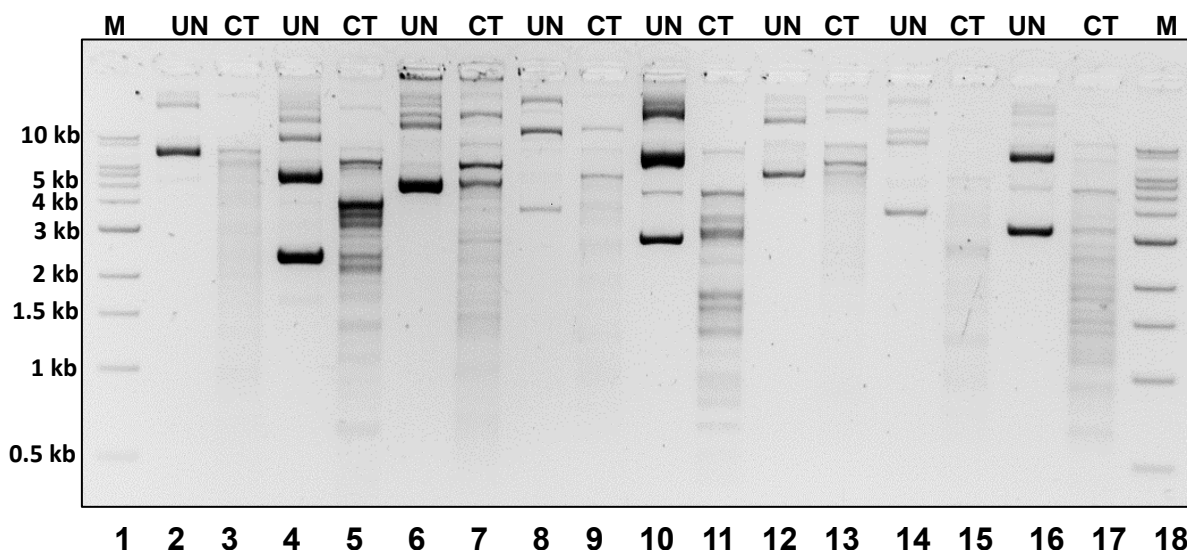


**Figure 4.5. Purification and subunit stoichiometry of MboIII.** (A) Expression and purification Mod subunit of MboIII. Lane 1 represents uninduced (UN), lane 2 represents marker (M), lane 3 represents induced (IN) Mod protein represented as asterick, lane 4 represented purified Mod subunit after size exclusion chromatography (SEC). (B) SEC-MALS elution profile of Mod of MboIII. Mod eluted as a single peak with a molecular weight of 131 kDa indicating that it is a homodimer. The solid line on left (Y-axis) represents the normalized light scattering while the dotted line on the right (Y-axis) represents the calculated molecular weight. (C) SDS-polyacrylamide gel representing the purity of MboIII obtained after size exclusion chromatography (lane 2). The lower band in lane 2 is Mod subunit and the upper band is Res subunit. Lane 1 is a protein ladder. (D) SEC-MALS elution profile of MboIII. MboIII eluted as a single peak with a molecular weight of 225 kDa indicating that MboIII is a heterotrimer composed of two Mod protomers and a single Res subunit. The solid line on left (Y-axis) represents the normalized light scattering while the dotted line on the right (Y-axis) represents the calculated molecular weight.

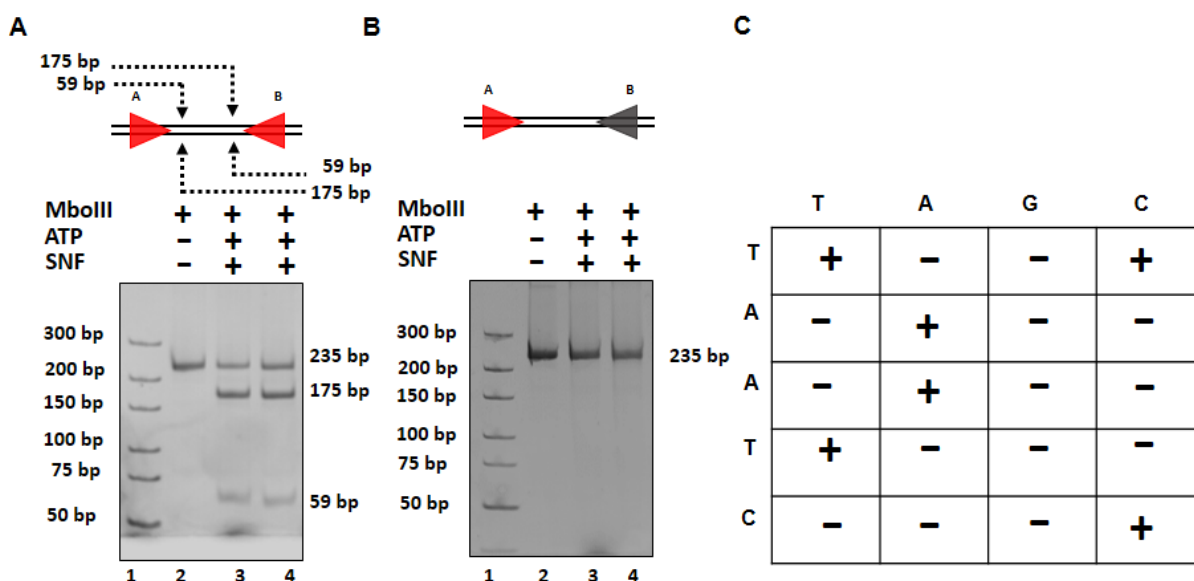
#### 4.3.3 Identification of the recognition sequence of MboIII

To identify the recognition site of MboIII, a plasmid library assay was carried out. A library of different plasmids were treated with MboIII and checked to see if they were nucleolytically cleaved. Assuming that MboIII, like other Type III RM enzyme, is an ATP-dependent endonuclease, nucleolytic cleavage in the presence of ATP but not in its absence would indicate that the plasmid DNA harbors the recognition sites. All the plasmids treated with MboIII were nucleolytically cleaved into a number of fragments in the presence of ATP indicating that they had multiple recognition sites of MboIII (Figure 4.6). One of these plasmids was linearized and subsequently shortened, so that only a single double-strand DNA break was observed. The hallmark of a Type III RM enzyme is that it cuts DNA having at least two recognition sites in inverted orientation. The cut occurs close to one of the recognition sites. Consequently, we assumed that the shortened linearized plasmid, which yielded only two fragments, had only a pair of the recognition site.

To locate the position of the recognition site, a primer-walking assay was carried out by systematically shortening one end of the DNA while keeping the other end intact and *vice-versa*. These DNAs were then tested for cleavage. Absence of cleavage on shortening the DNA suggested loss of a recognition sites. The assay resulted in a 218 bp DNA, which on shortening from either end prevented cleavage by MboIII. A careful examination of the sequence of DNA substrate revealed the presence of two TAATC sites oriented head-to-head. Position of the two sites was consistent with the length of DNA fragments that were produced on cleavage by MboIII. To identify the consensus recognition sequence of MboIII, I generated a 235 bp DNA substrate with two TAATC sites (site A and site B in Figure 4.7 A) in head-to-head orientation and performed DNA cleavage assay. Each base of site B was systematically changed to the other three bases while keeping site A constant to see their affect of DNA cleavage (see for example Figure 4.7 B). In total, 15 different DNA substrates were generated and their affect on DNA cleavage was studied (Figure 4.7 C). Based on this study, the recognition site of MboIII was identified as YAATC, Y=T/C (Figure 4.7 C).



**Figure 4.6. Plasmid library assay of MboIII.** DNA cleavage assay was carried out by incubating 300 nM MboIII and 30 nM of a different plasmid DNA substrate in the presence of 4 Mm of ATP. Lane 1, 18 represents the ladder, lane 2, 4, 6, 8, 10, 12, 14, 16 represents different plasmids (Uncut= UN) in absence of MboIII. Lane 3, 5, 7, 9, 11, 13, 15, 17 represents the same plasmids treated (Cut= CT) with MboIII and ATP. DNA cleavage is noticed by the presence of smaller linear fragments (Lane 3, 5, 7, 9, 11, 13, 15, and 17).



**Figure 4.7. Identification of MboIII recognition sequence.** (A) DNA cleavage assay was carried out by incubating 300 nM MboIII and 50 nM of a 235 bp DNA substrate containing two recognition sites (red arrowheads) of MboIII in head-to-head orientation. Such a substrate was cleaved into two fragments of size ~179 bp and ~59 bp. (B) DNA cleavage was abolished when the recognition site at one end was altered (grey arrow-

-head). (C) The different DNA substrates generated for DNA nuclease assay when recognition site at each base was changed to other three bases. DNA cleavage happened (marked by a tick) only when recognition site was (YAATC, where Y = T/C).

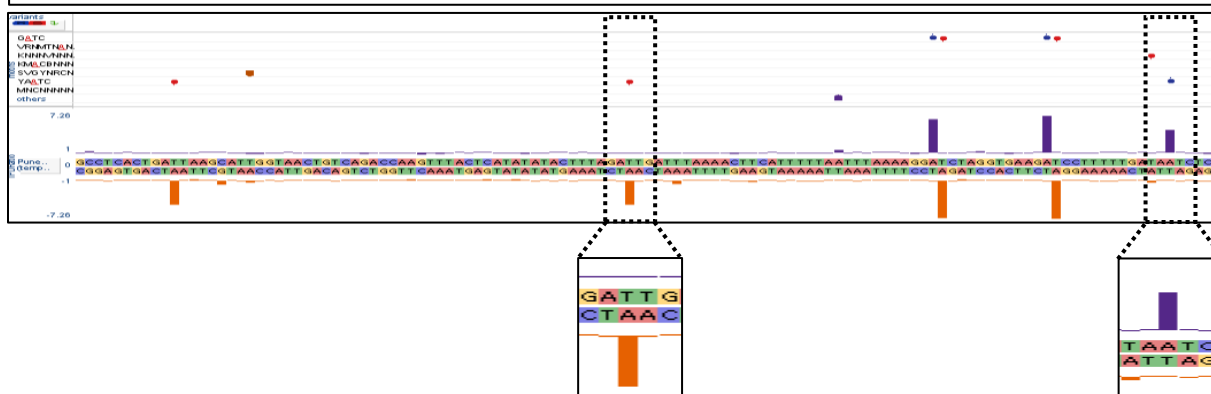
#### 4.3.4 Target base for methylation

Type III RM enzymes can catalyze the methylation of their recognition site in presence of AdoMet. The methyltransferase of MboIII was predicted to belong to  $\beta$ -class of N6-adenine methyltransferases. As the recognition site of the MboIII, YAATC, has two adenine in the top strand and one adenine in the bottom strand, it was important to identify which of the three were target for methylation. I addressed this using PacBio sequencing of DNA methylated by MboIII. An *in vitro* methylation of a 7.5 kb plasmid having 17 MboIII recognition site was carried out using Mod in presence of AdoMet. The extent of methylation of the plasmid was confirmed by DNA cleavage assay using MboIII and ATP.

A

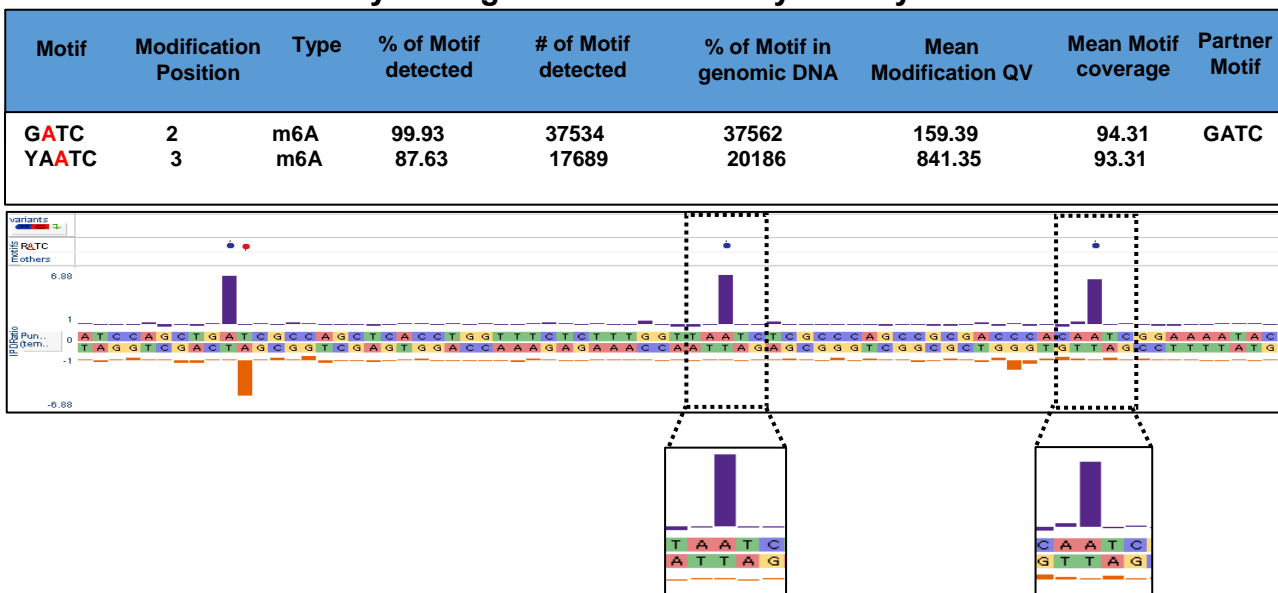
#### SMRT modification analysis of plasmid DNA methylated by MboIII

| Motif | Modification Position | Type | % of Motif detected | # of Motif in plasmid DNA | % of Motif in plasmid DNA | Mean Modification QV | Mean Motif coverage | Partner Motif |
|-------|-----------------------|------|---------------------|---------------------------|---------------------------|----------------------|---------------------|---------------|
| GATC  | 2                     | m6A  | 100                 | 24                        | 24                        | 2284.08              | 2199.00             | GATC          |
| YAATC | 3                     | m6A  | 100                 | 21                        | 21                        | 841.35               | 2353.25             | GATC          |



## B

## SMRT modification analysis of genomic DNA methylated by MbolII



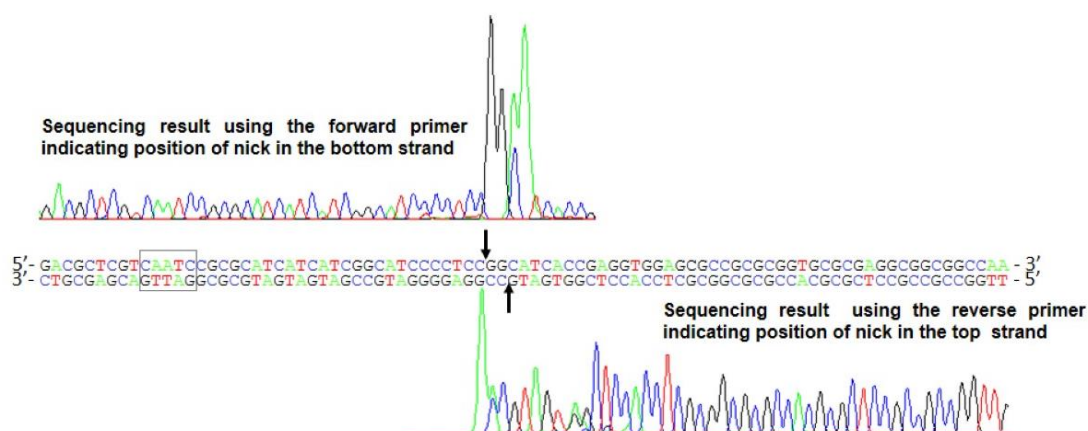
**Figure 4.8. Identification of the target base for methylation.** (A) Methylome analysis of a plasmid DNA methylated with Mod subunit. An Inter Pulse Duration (IPD) plot showing methylation happening in the second adenine of the recognition site ( $YA^{m6}ATC$ ). (B) Methylome analysis of genomic DNA of *E.coli* cells expressing Mod subunit. The IPD plot shows methylation of the second adenine in both the  $TA^{m6}ATC$  and  $CA^{m6}ATC$  recognition sequences.

Absence of DNA cleavage of the plasmid indicated complete methylation. This plasmid was then sequenced using PacBio sequencing. Consistent with the result of the above analysis, PacBio sequencing result led to the identification of YAATC as the recognition site. The sequencing results further showed that the second adenine of the recognition sequence was methylated, i.e.  $YA^{m6}ATC$  (Figure 4.8 A).

I also checked the position of methylation on genomic DNA upon overexpression of Mod subunit, by transforming *E. coli* BL21(DE3) cell with a construct having *mod* gene and inducing with IPTG. The purified genomic DNA was subjected to PacBio sequencing. The site YAATC was found to have a methylated adenine at the second position (Figure 4.8 B). As *E. coli* BL21(DE3) is a *Dam*<sup>+</sup> strain, the site GATC, which is the recognition site of Dam methyltransferase, was also picked as being methylated, in addition to  $YA^{m6}ATC$  (Figure 4.8 B).

### 4.3.5 Cleavage loci of MboIII

Run-off sequencing of the products of DNA cleavage by MboIII was carried out to identify the position of scission of the phosphodiester bonds. The sequencing results revealed that MboIII cuts 26 bp 3' to the recognition site on the YAATC strand and 28 bp 5' to the recognition site on the GATTR strand to generate a 2 base 5'-overhang (Supplementary Figure 4.9). This cleavage pattern is similar to that of the prototypical Type III RM enzymes EcoP1I and EcoP15I (23).



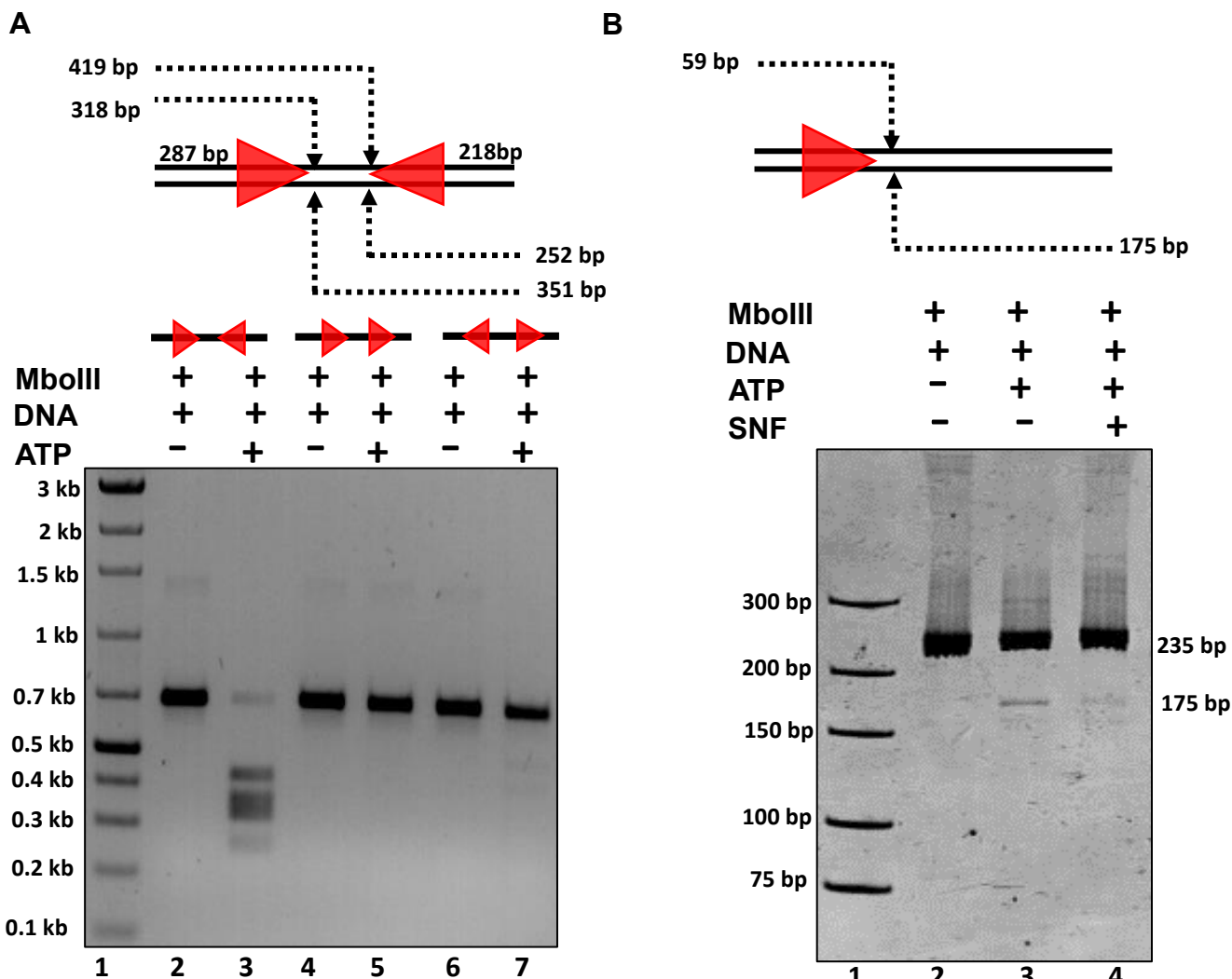
**Figure 4.9. Run-off sequencing to determine the cleavage loci of MboIII.** The recognition sequence is marked by a rectangular box. The arrows in the top and bottom strands indicate the position of the two nicks. MboIII nicks the top strand 26 bp downstream of the recognition sequence and bottom strand is nicked 28 bp downstream of the recognition sequence.

### 4.3.6 Site requirement for DNA cleavage

Previous studies with EcoP1I and EcoP15I have shown that directionality of recognition sites is important for efficiency of DNA cleavage (25). Recognition sites have to be in inverted orientation for DNA cleavage, and that head-to-tail sites are refractory to cleavage. In the context of linear DNA, recognition sites that are in head-to-head orientation are cleaved with better efficiencies than those having tail-to-tail orientation (26). To find if orientation of the recognition sites affect DNA cleavage, I generated three DNA substrates with recognition sites in head-to-head, head-to-tail and tail-to-tail orientation. The assay revealed that substrates with recognition site in head-to-head orientation were cleaved, while that having head-to-tail orientation was refractory to cleavage (Figure 4.10 A, Lane 4 and 5). This short DNA having recognition sites in tail-



to-tail orientation was cleaved weakly (Figure 4.10 A, Lane 6 and 7). As has been reported in case of EcoP1I and EcoP15I, we also noted that MboIII can catalyze nucleolytic cleavage of single-site substrate but only weakly (Figure 4.10 B).



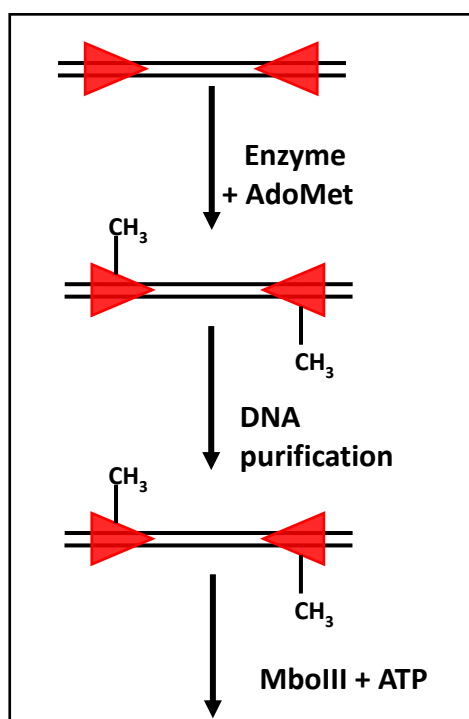
**Figure 4.10. Site requirement for DNA cleavage by MboIII.**(A) DNA cleavage assay with 200 nM MboIII and 30 nM DNA substrate containing two recognition sites in head-to-head orientation (Lane 2 and 3) resulted in generation of four fragments that were 481, 318, 351 and 218 bp long when resolved on 1.5 % agarose gel. DNA substrate with recognition site in head-to-tail (Lane 4 and 5) were refractory to cleavage while the substrate with recognition sites in tail-to-tail orientation was cleaved very poorly (Lane 6 and 7). (B) A DNA substrate having only single recognition site was cleaved but less efficiently either in the absence (Lane 3) or presence (Lane 4) of sinefungin.



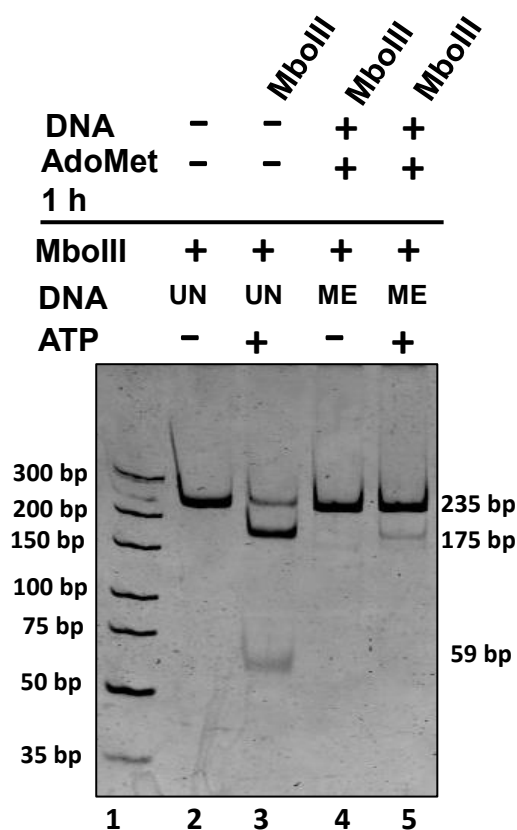
#### 4.3.7 Methylation of DNA by MboIII inhibits its nucleolytic activity

I next proceeded to test the methylation activity of MboIII. I incubated a DNA substrate having two target sites oriented head-to-head with MboIII holoenzyme in presence of AdoMet for 60 mins. Subsequent to this the DNA was purified and subjected to endonucleolytic cleavage by freshly added MboIII in presence of ATP (Figure 4.11 A), and the reaction product analyzed on a native PAGE (Figure 4.11 B, Lane 5). Most of the DNA was refractory to DNA cleavage, suggesting that MboIII methylated the substrate and protected it from the nucleolytic activity.

A

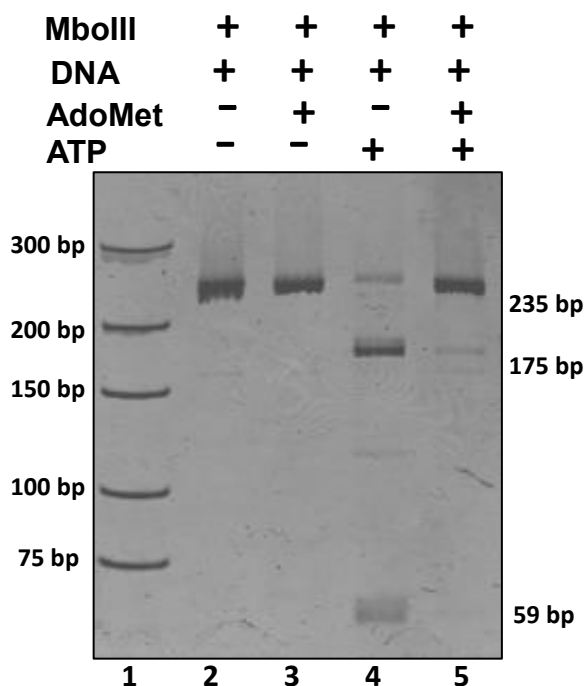


B



**Figure 4.11. Methylation activity of MboIII.** (A) A schematic protocol used for studying the methylation activity of MboIII. (B) DNA cleavage assay was carried out with 50nM DNA substrate and 300 nM of MboIII in presence of 4 mM of ATP. The methylated DNA (ME) was obtained using the protocol as shown in Panel A. The methylated DNA was purified and 50 nM of ME DNA was incubated with 300 nM of MboIII in presences of ATP. Methylated DNA were protected from DNA cleavage (Lane 5). As a control, 50 nM unmethylated DNA was cleaved when treated with 300 nM MboIII (Lane 3).

As shown above, MboIII can catalyze either methylation or cleavage of the DNA substrate depending on the availability of cofactors, i.e. ATP or AdoMet. To find which of the two reactions would be prominent in case both the cofactors are present simultaneously, I performed an assay in which a DNA substrate was incubated with MboIII, 4 mM ATP and 200  $\mu$ M AdoMet, simultaneously. After incubation for 60 minutes, the DNA was analyzed on native PAGE. We found that only a small fraction of DNA was cleaved, while the rest of the fractions were protected (Figure 4.12). This suggested that in the presence of both the cofactors, methylation was catalytically much faster than the nucleolytic cleavage.

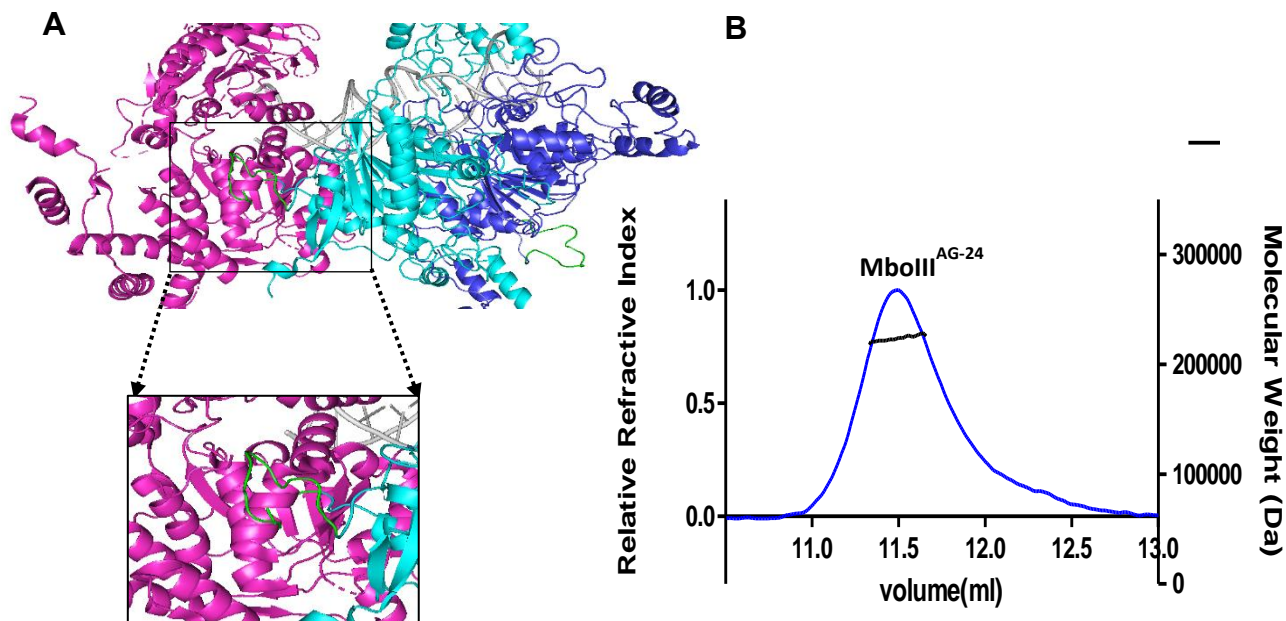


**Figure 4.12. Competition assay between methylation and nuclease activities of MboIII.** 50 nM 235 bp two-site DNA substrate was incubated with 300 nM of the enzyme was incubated in presence of AdoMet (200  $\mu$ M) and ATP (4 mM). The presence of AdoMet diminished DNA cleavage, suggesting that the DNA substrate was methylated before the nucleolytic cleavage could occur.

#### 4.3.8 Effect of repeats on the methylation activity of MboIII

In MboIII, the SSR is present at the N-terminus of Mod subunit 13 amino acids ahead of the catalytic DPPY motif. To check if the repeats have any effect on the activity of MboIII, we purified two variants of MboIII, MboIII<sup>AG-9</sup> having 9 AG repeats resulting in 6 RE amino

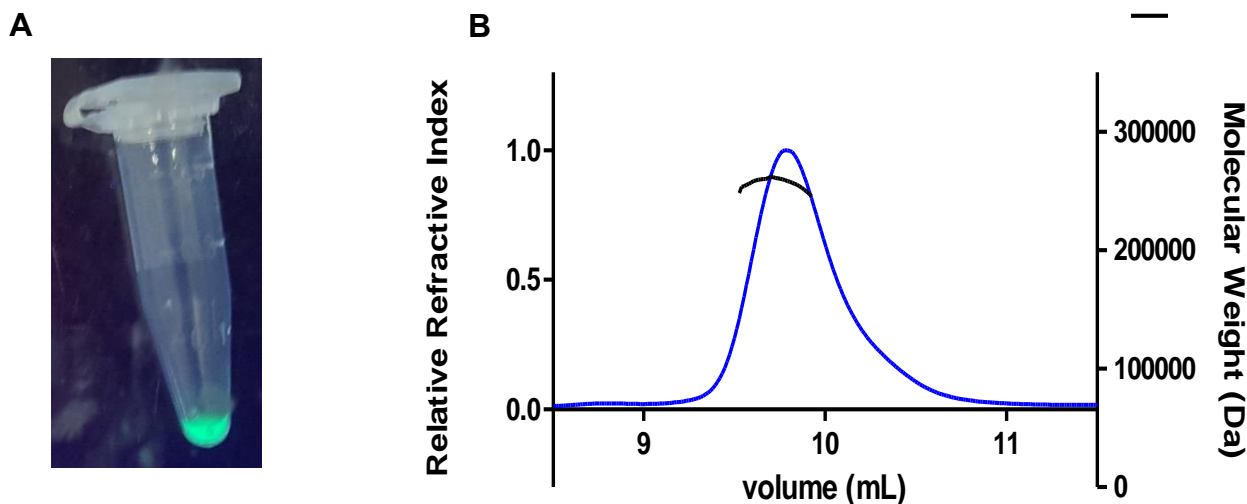
acid repeats (22) and MboIII<sup>AG-24</sup> having 24 AG repeats resulting in 16 RE amino acid repeats. A model of the dimeric Mod generated using the structure of EcoP15I, suggested that the SSR of one of the Mod protomers was located spatially close to the interface with Res where it could potentially act as a steric block to oligomerisation with Res (Figure 4.13 A). The proteins were purified to homogeneity and the oligomeric assemblies of MboIII with and without repeats were identical (Figure 4.13 B).



**Figure 4.13. Incorporation of 24 AG repeats does not affect Mod and Res complex formation.** (A). A predicted structural model of Mod<sup>24-AG</sup> dimer showing the presence of repeats exposed to surface in case of Mod<sub>2</sub> and close to Mod<sub>1</sub>, Res oligomerisation interface. The inset is a zoom of a part of the interface highlighting the possible clash of SSR with Res. (B) SEC-MALS elution profile of MboIII<sup>24-AG</sup>. MboIII<sup>24-AG</sup> eluted as a single peak with a molecular weight of 223 kDa indicating that Mod subunit is a heterotrimer composed of two Mod subunits and a single Res subunit. The solid line on left (Y-axis) represents the normalized light scattering, while the dotted line on the right (Y-axis) represents the calculated molecular weight

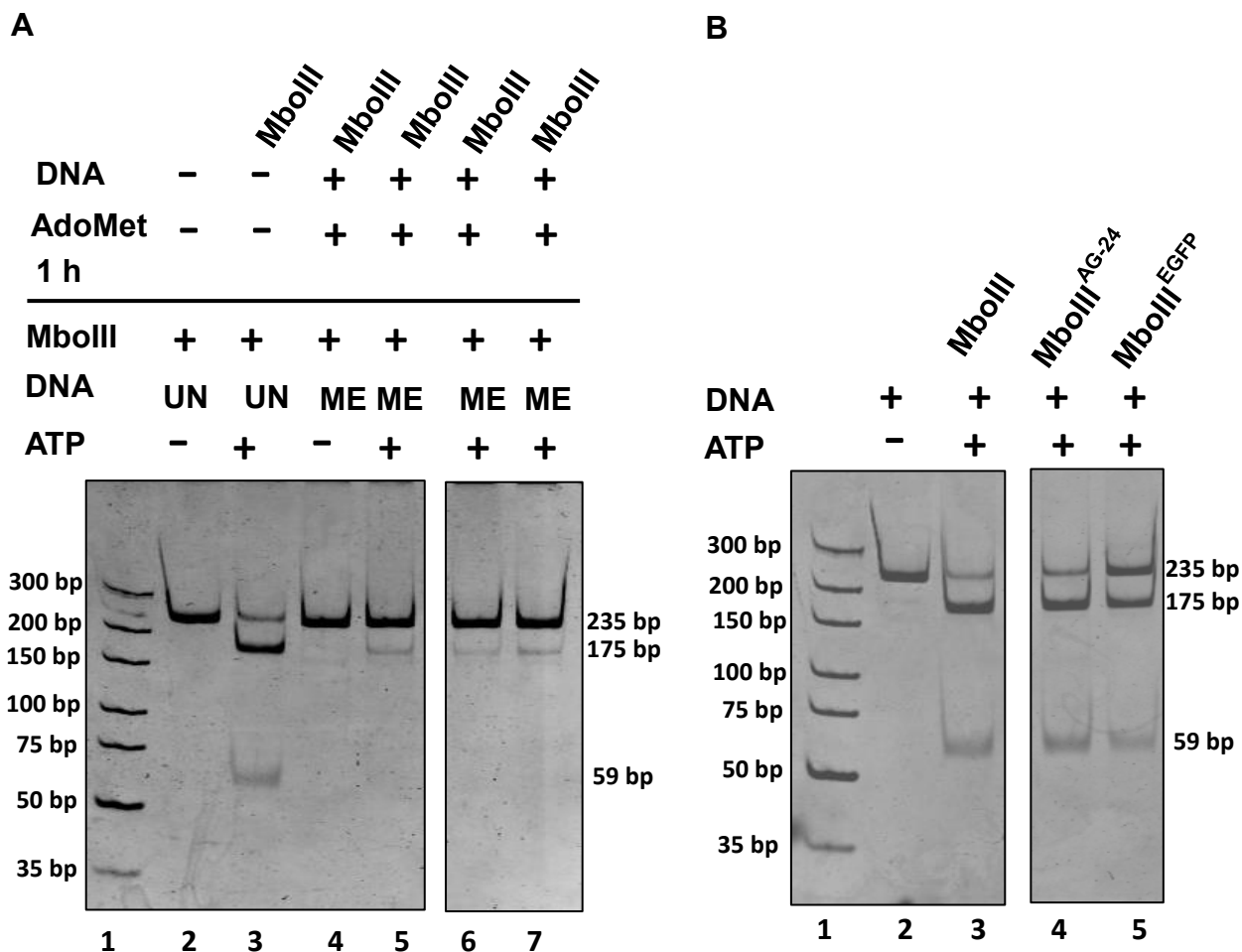
This study revealed that despite the spatial proximity of SSR to the oligomeric interface between Mod and Res, oligomerisation remained unaffected independent of the length of the SSR. To test the plasticity of the site of SSR for longer polypeptides, I replaced the repeats with a sequence encoding the 27 kDa EGFP. MboIII<sup>EGFP</sup> was purified and the protein fluoresced under UV light confirming that the EGFP was folded (Figure

4.14 A). The oligomeric state of MboIII<sup>EGFP</sup> was analyzed using SEC-MALS and the subunit stoichiometry was identical to that of MboIII (Figure 4.14 B).



**Figure 4.14. Incorporation of EGFP into MboIII.** (A) Image of MboIII<sup>EGFP</sup> under UV light indicates the presence of a folded EGFP inserted as replacement to SSR in Mod. (B) SEC-MALS elution profile of MboIII<sup>EGFP</sup>. MboIII<sup>EGFP</sup> eluted as a single peak with a molecular weight of 260 kDa, indicating that the incorporation of EGFP into Mod subunit did not affect the oligomeric structure of the protein. The solid line on left (Y-axis) represents the normalized light scattering, while the dotted line on the right (Y-axis) represents the calculated molecular weight.

Methylation activities of the MboIII, MboIII<sup>24-AG</sup> and MboIII<sup>EGFP</sup> were studied by incubating the enzymes with a DNA substrate having two recognition sites in the presence of AdoMet. After methylation reaction, the DNA was purified (using similar protocol as mentioned in Figure 4.10 A), and the extent of methylation monitored by treating it with MboIII in the presence of ATP (Figure 4.15 A). Analysis of cleavage efficiency was used as an indicator of extent of methylation. Methylation of DNA by MboIII MboIII<sup>24-AG</sup> or its variants including MboIII<sup>EGFP</sup> provided almost complete protection against the restriction activity of MboIII (Figure 4.15 A, Lanes 6, 7). This data clearly indicated that the repeats of the insertion of EGFP did not change the methylation activity of the enzyme. We also compared the methylation activity of MboIII and MboIII<sup>9-AG</sup> and found that MboIII<sup>9-AG</sup> methylated the DNA with a similar activity as that of MboIII (22).



**Figure 4.15. Effect of SSR and EGFP on the methylation and nuclease activity of MbolIII.** (A) DNA substrates were methylated with, MbolIII<sup>AG-24</sup> and MbolIII<sup>EGFP</sup> with AdoMet using the protocol shown in Panel A of Figure 4.10 A. The methylated DNA (ME) was purified and incubated with 300 nM of MbolIII in presences of ATP to check the extent of methylation. Methylated DNA generated on treatment with MbolIII<sup>AG-24</sup> and MbolIII<sup>EGFP</sup> were protected from DNA cleavage (Figure 14 Lane 6, 7). (B) DNA cleavage assay was carried out with 50 nM DNA and 300 nM of either MbolIII, MbolIII<sup>AG-24</sup> or MbolIII<sup>EGFP</sup>. MbolIII, MbolIII<sup>AG-24</sup> (Lane 3-5) cleaved with a similar efficiency, while MbolIII<sup>EGFP</sup> cleaved DNA with a slightly lower efficiency.

#### 4.3.9 Effect of repeats on DNA cleavage by MbolIII

Having found that the inserts do not affect the methylation activity of MbolIII, I proceeded to check its effect on DNA cleavage. DNA cleavage assay indicated that MbolIII with no SSR, and MbolIII<sup>AG-24</sup> cleaved the substrate with an equal efficiency (Figure 4.15 B, Lane 3-4). This indicated that the presence of repeats did not affect the nuclease activity of the

enzyme. Next, I tested the nucleolytic activity of MbolIII<sup>EGFP</sup>, which to our surprise had significant nucleolytic activity that was only slightly lower than that of MbolIII, MbolIII<sup>AG-9</sup> and MbolIII<sup>AG-24</sup> (Figure 4.15 B, Lane 5). We also compared the nuclease activity of MbolIII and MbolIII<sup>9-AG</sup> and found that MbolIII<sup>9-AG</sup> cleaved the DNA with a similar activity as that of MbolIII (22)

#### 4.4 Discussion

While the effect of length of SSR on the properties of eukaryotic proteins is well documented, similar information for many prokaryotic proteins is lacking. Here, I investigated the affect of the length of SSR on the biochemical properties of MbolIII, a Type III RM enzyme from *M. bovis*. Towards this objective, we cloned, purified, and characterized the oligomeric assembly, DNA recognition sequence, nucleolytic and methylation activities of the enzyme. Like the prototypical EcoP1I and EcoP15I (27, 28), MbolIII is an oligomer of two Mod and one Res polypeptides. The enzyme cuts DNA with two recognition sequences in inverted orientation, with substrates having sites in head-to-head orientation being cleaved significantly better than those having sites in tail-to-tail orientation. Consistent with the prototypes, the cleavage occurs ~26-28 bp downstream of the recognition sequence resulting in a 2 nucleotide 5' overhang. The enzyme also displayed weak nucleolytic activity for a single-site substrate.

MbolIII methylates the second adenine of its recognition sequence YAATC, which silences the restriction activity. In the presence of both AdoMet and ATP, the nucleolytic activity of the enzyme was considerably diminished, suggesting that its methylation activity is more prominent than cleavage activity. This observation is akin to what has been noted for EcoP15I (29), but unlike what has been reported for Type ISP enzymes. In case of Type ISP enzyme LlaBIII, it has been noted that the enzyme cleaves DNA in presence of both ATP and AdoMet (30). It must be, however, be noted that the substrate used in this reaction was a short substrate with just 2 sites. It is likely that under natural conditions where the DNA substrate would be long with multiple sites (such as phage genomic DNA) the nucleolytic activity is prominent.

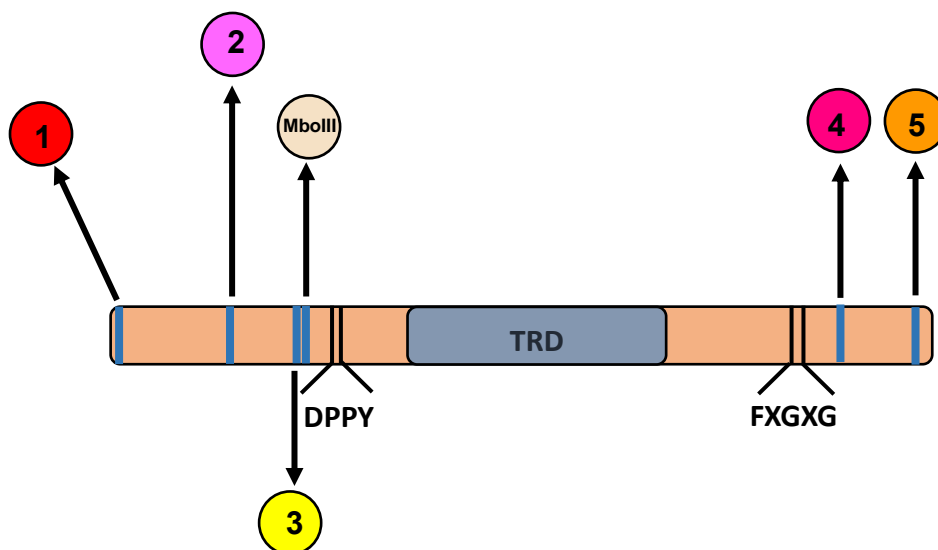
MbolIII has an SSR 13 residues from the DPPY motif towards the N-terminus of Mod subunit. The SSR is a repeat of AG nucleotides, which encodes a repeat of RE

residues. The *mod* gene isolated from the genomic DNA had 11 AG repeats, which suggested that in the clonal population of *M. bovis* from which the genome was isolated, the gene was in an off state. I could turn on the gene by deleting the SSR or by changing the length of the SSR to 9 or 24 AG repeats. All the three products displayed a very similar biochemical properties, demonstrating that the length of the SSR did not affect them. Though the location of SSR is close to the catalytic motif DPPY and the interface of Mod and Res, increase in length of SSR did not appear to either affect the structural integrity of MboIII, or alter its catalytic activities. Insertion of a polypeptide encoding EGFP at the site of SSR did not affect the properties of the enzyme. Our results highlight the structural plasticity of the site at which SSR is located.

I next studied the location of SSR in the amino acid sequences of the Mod subunit of Type III RM enzymes from other *Mycoplasma sp.* Of the 92 sequences that we analyzed, 87 Type III RM enzymes have an SSR consisting of AG repeats located at a position 9 to 14 amino acids N-terminal to the DPPY motif, very similar to that in MboIII, where it was located 13 amino acids away. Two of them had an SSR located 29 amino acids N-terminal to the DPPY motif. Mod from a few species had SSR located at the C-terminal end of the protein. I extended our study to include SSR containing sequences of Mod from other organisms. The Jennings's laboratory had previously catalogued such sequences from a number of organisms (17, 18, 31). I extended this list with the additional sequences from REBASE (Table 4.4). In all 174 sequences of Type III Mod subunits were examined, I located the position of these SSRs in the context of their polypeptide three-dimensional structure, predicted based on the structure of the dimeric Mod of EcoP15I (32).

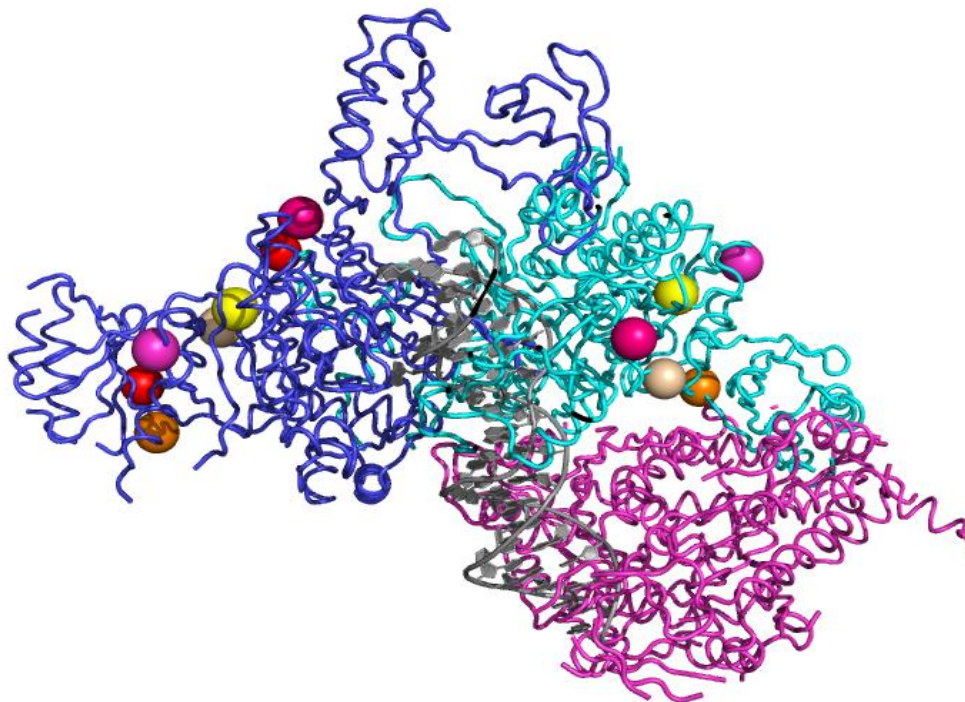
The location of the SSR in Mod from the various organisms populated a few select regions on the primary structure of the protein. In all, we identified 5 such positions (Figure 4.16). Based on the structure of EcoP15I, position 1 has the SSR located upstream of the N-terminal domain of the Mod subunit. An SSR at this position is placed away from core domain of the methyltransferase or the interface between the subunits and thus is unlikely to affect the catalytic activity or the structural integrity of the protein (Figure 4.16, and 4.17). However, this remains to be proved. Position 2 is about 41 amino acids towards the N-terminus from the DPPY motif. This site is located on the periphery of the MTase

domain, away from the core, catalytic and DNA-binding region, and the oligomeric interface protein (Figure 4.16, and 4.17). Position 3 has SSRs located 9 to 29 amino acids N-terminus to the DPPY motif, such as in case of MboIII, which has SSR 13 amino acids away from the DPPY motif protein (Figure 4.16, and 4.17). As discussed above, SSR at this position does not affect the biochemical properties of the enzyme protein (Figure 4.16, and 4.17).



**Figure 4.16. Position of SSRs in the Mod subunit Type III RM enzymes.** The position of SSR in Mod of Type III RM enzymes from various organism are marked as blue lines (numbered 1 to 5). MboIII represents the position of the SSR in this Type III RM enzyme.





**Figure 4.17. Location of the SSR sites mapped on the structure of EcoP15I.** The two Mod subunits are colored blue and cyan, the Res subunit is colored magenta, and the DNA is colored grey. Red spheres represent the position of SSR which are at the N-terminus, magenta spheres represent the position of SSR 41 amino acids upstream of DPPY motif, yellow spheres represent the position of SSR 29 amino acids upstream of DPPY motif, wheat spheres represent the position of SSR in MbolIII (13 amino acids upstream of DPPY motif), pink spheres represent the position of SSR 19 amino acids downstream of FXGXG motif and orange spheres represent the position of SSR at the C-terminus.

Position 4 is 19 amino acids from FXGXG motif towards the C-terminus. FXGXG motif, which is involved in AdoMet binding, is located at the loop connecting the C-terminus of a  $\beta$ -strand and the N-terminus of an  $\alpha$ -helix protein. Position 4 is located in a loop at the other end, i.e. the C-terminus of the  $\alpha$ -helix protein. This places the SSR at this position away from the DNA and AdoMet binding sites as well as the MTase catalytic site and subunit interfaces (Figure 4.16 and 4.17). Position 5 is at the C-terminus and is located away from the core of the MTase (Figure 4.16, and 4.17). In conclusion, the analyses of the sequences and structures of SSR-containing Mod from various organisms revealed that the repeats were located in general at sites exposed to the solvent, and away from the catalytic and DNA binding regions or oligomerisation interface (Figure



|   |                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |       |
|---|---------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|   |                                                                                                                                                         | M.Hin199ORF4730P, M.HinAWORF329P,<br>M.McaST6ORF1990P, M.Hha21621ORF50P,<br>M.HindVI,                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |       |
| 1 | <i>Neisseria meningitidis</i>                                                                                                                           | M.NemG2136ORFAP, M.Nme108ORFLP,<br>M.Nme1114ORF5660P, M.Nme1207ORFAP,<br>M.Nme12176ORF5650P, M.Nme12330ORF5655P<br>, M.Nme210ORF4975P, M.Nme22747ORF9015P,<br>M.Nme22748ORFAP, M.Nme22759ORF6665P,<br>M.Nme22801ORF5640P,<br>M.Nme23413ORF8990P, M.Nme2419ORF1070P,<br>M.Nme24730ORF9915P, M.Nme25073ORF1030P<br>,<br>M.Nme25459ORF10310P, M.Nme25472ORF4510<br>P, M.Nme3686ORF11025P, M.Nme400ORFAP,<br>M.Nme5094ORFAP, M.Nme563ORF5665P,<br>M.Nme600ORF5650P, M.Nme6178ORF5655P,<br>M.Nme7124ORF8770P, M.Nme22160ORF6880P,<br>M.Nme22745ORF460P, M.Nme3682ORF7745P. | AGCC  |
| 2 | <i>Streptococcus thermophiles</i>                                                                                                                       | M.SthCS8ORF535P, M.Sth9ORF655P,<br>M.Sth18311ORF885P, M.Sth1477ORF4810P,<br>M.Sth1066ORF835P.                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | CACAG |
| 3 | <i>Mycoplasma agalactis</i> ,<br><i>Mycoplasma auris</i> ,<br><i>Mycoplasma bovis</i> ,<br><i>Mycoplasma cannis</i> ,<br><i>Mycoplasma columbinum</i> , | M1.Mag5632ORF1570P, M2.Mag5632ORF1570P,<br>M.MagPGORF1530P, M.Mau15026ORF50P,<br>M1.Mbo08MORF855P, M2.Mbo08MORF855P,<br>M3.Mbo08MORF855P, M.Mbo165ORF2390P,<br>M.Mbo2012ORF820P, M1.Mbo4278ORF138P,<br>M2.Mbo4278ORF138P, M3.Mbo4278ORF138P,<br>M1.Mbo45ORF167P, M2.Mbo45ORF167P,<br>M3.Mbo45ORF167P, M1.Mbo70ORF800P,<br>M2.Mbo70ORF800P, M.Mbo801ORF166P,<br>M1.MboH1ORF158P, M2.MboH1ORF158P,                                                                                                                                                                      | AG    |

|  |                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |  |
|--|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|  | <p><i>Mycoplasma hyorhinis</i>,<br/> <i>Mycoplasma dispar</i>,<br/> <i>Mycoplasma flocculare</i>,<br/> <i>Mycoplasma hyopneumoniae</i>,<br/> <i>Mycoplasma hyorhinis</i>,<br/> <i>Mycoplasma mycoides</i>,<br/> <i>Mycoplasma ovipneumoniae</i>,<br/> <i>Mycoplasma synoviae</i></p> | <p>M.MboJF4428ORF498P, M1.MboN1ORF850P,<br/> M2.MboN1ORF850P, M3.MboN1ORF850P,<br/> M.MboP115ORF140P, M.MboP180ORF140P,<br/> M.Mca160ORF435P, M.McaLVORF3590P,<br/> M.McaPG14ORF990P, M.McaUF31ORF1891P,<br/> M.McaUFG1ORF1846P, M.McoHRCORF2220P,<br/> M2.McoSF7ORF2868P, M3.McoSF7ORF2868P,<br/> M.McyMORF1065P, M1.McyMORF1065P,<br/> M2.McyMORF1065P, M.Mdi27140ORF3355P,<br/> M.MdiGS01ORF3135P, M.Mfl27399ORF1440P,<br/> M.Mfl27399ORF1645P, M.Mfl27399ORF3220P,<br/> M1.Mfl27-399ORF1580P, M.MhyORF399P,<br/> M2.Mfl27399ORF1580P, M.MgalPTORF2920P,<br/> M.MhyORF329P, M.MhyORF427P,<br/> M1.Mhy1050ORF200P, M2.Mhy1050ORF200P,<br/> M3.Mhy1050ORF200P, M.Mhy14ORF608P,<br/> M.Mhy14ORF702P, M.Mhy14ORF704P,<br/> M.Mhy14ORF729P, M.Mhy14ORF756P,<br/> M.Mhy168ORF318P, M.Mhy168ORF409P,<br/> M.Mhy168ORF435P, M.Mhy168LORF318P,<br/> M.Mhy168LORF409P, M.Mhy168LORF750P,<br/> M.Mhy7422ORF2629P, M.Mhy7422ORF3257P,<br/> M.Mhy7422ORF3394P, M.Mhy7448ORF316P,<br/> M.Mhy7448ORF388P, M.Mhy7448ORF410P,<br/> M1.Mhy76ORF202P, M2.Mhy76ORF202P,<br/> M3.Mhy76ORF202P, M1.MhyGDL1ORF187P,<br/> M2.MhyGDL1ORF187P, M3.MhyGDL1ORF187P,<br/> M1.MhyIPVORF217P, M3.MhyIPVORF217P,<br/> M.MhyJORF308P, M.MhyJORF382P,<br/> M.MhyJORF383P, M.MhyJORF399P,<br/> M.MhyJORF423, M.MmyCl,</p> |  |
|--|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

|   |                                        |                                                                                                               |    |
|---|----------------------------------------|---------------------------------------------------------------------------------------------------------------|----|
|   |                                        | M.MmyGM12ORF590P, M.MmyGM12SI,<br>M.MovSC01ORFIP, M.MpuCORF4800P,<br>M.Msy25204ORF870P, M.MsyMSHORF6350P.     |    |
| 4 | <i>Helicobacter pylori</i>             | Hpy13409ORFWP, Hpy13609ORFVP,<br>Hpy10109ORFWP, Hpy14309ORFWP,<br>HpyML1ORFDP, Hpy328ORF2850P,<br>HpyB8ORF7P. | G  |
| 5 | <i>Mycoplasma pulmonis/colu mbinum</i> | M1.MpuCI, M2.MpuCI, M.McoSF7ORF3193P.                                                                         | AG |

## References.

1. 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., Firman, K., Gromova, E.S., Gumport, R.I., Halford, S.E., Hattmann, S., Heitmann, J., Hornby, D.P., Janulaitis, A., Jeltsch, A., Josephsen, J., Kiss, A., Klaenhammer, T.R., Kobayashi, I., Kong, H., Kruger, D.H., Lacks, S., Marinus, M.G., Miyahara, M., Morgan, R.D., Murray, N.E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao, D.N., Reich, N., Repin, V.E., Selker, E.U., Shaw, P.C., Stein, D.C., Stoddard, B.L., Szybalski, W., Trautner, T.A., Etten, J.L.V., Victor, J.M.B., Wilson, G.G. and Xu, S.Y. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
2. Rao, D.N., Dryden, D.T.F. and Bheemanaik, S. (2014) Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.*, **42**, 45-55.
3. Ramanathan, S.P., van Aelst, K., Sears, A., Peakman, L.J., Diffin, F.M., Szczelkun, M.D., and Seidel, R. (2009) Type III restriction enzymes communicate

- in 1D without looping between their target sites. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 1748-1753.
4. Schwarz,F.W., Toth,J., van Aelst,K., Cui,G., Clausing,S., Szczelkun,M.D. and Seidel,R.(2013) The Helicase-Like Domains of Type III Restriction Enzymes Trigger Long-Range Diffusion Along DNA. *Science*, **340**, 353–356.
  5. Toth,J., Bollins,J., and Szczelkun,M.D. (2015) Re-evaluating the kinetics of ATP hydrolysis during initiation of DNA sliding by type III restriction enzymes. *Nucleic Acids Res.*, **43**, 10870-10881.
  6. Hallet, B. (2001) Playing Dr Jekyll and Mr Hyde: combined mechanisms of phase variation in bacteria. *Curr Opin Microbiology.*, **4**, 570–581.
  7. van Belkum, A., Scherer, S., van Alphen, L., & Verbrugh, H. (1998). Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev.*, **62**, 275–293.
  8. Bayliss,C.D. and Palmer,M.E.(2012) Evolution of simple sequence repeat-mediated phase variation in bacterial genomes. *Annals of The New York Academy Of Sciences*.1267., 39-44
  9. Levinson,G.and GutmanG.A.(1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol.Bio.Evol.*,**4**(3),203-221.
  - 10.Gatchel,J.R. and Zoghbi,H.Y(2005) Diseases of unstable repeat expansion: mechanisms and common principles. *Nature Reviews Genetics.*, **6**: 743-755.
  - 11.Verstrepen, K.J., Jansen, A., Lewitter, F. and Fink, G.R. (2005) Intragenic tandem repeats generate functional variability. *Nature Genetics.*, **37**,986-990.
  - 12.Fidalgo,M., Barrales,R.R., Ibeas,J.I., and Jimenez,J.(2006) Adaptive evolution by mutations in the FLO11 gene. *Proc. Natl Acad. Sci. USA.*, **103**, 11228–11233.
  - 13.Micheal, T.P., Park, S., Kim, T.S., Booth, J., Byer, A., Sun,Qi., Chory,J. and Kwangwon.L. (2007) Simple sequence repeats provide a substrate for phenotypic variation in the *Neurospora crassa* cardiac clock. *PLoS ONE* **8**: e795. doi : 10.1371/ journal.pone.0000795.

14. Johnsen, A., Fidler, A.E., Kuhn, S., Carter, K.L., Hoffmann, A., Barr, I.R., Biard, C., Charmantier, A., Eens, M., Korsten, P., Siitari, H., Tomiuk, J. and Kempenares, B. (2007) Avian Clock gene polymorphism: evidence for a latitudinal cline in allele frequencies. *Molecular Ecology*, **16**, 4867-4880.
15. Weiser, J.N., Love, J.M., & Moxon, E.R. (1989). The molecular mechanism of phase variation of *Haemophilus Influenzae* lipopolysaccharide. *Cell*, **59**, 657–665.
16. van der Ende, A., Hopman, C.T., Zaat, S., Essink, B.B., Berkhout, B. and Dankert, J. (1995). Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the -10 and -35 regions of the promoter. *J. Bacteriol.*, **177**, 2475–2480.
17. Srikhanta, Y.N., Fox, K.L. & Jennings, M.P. (2010). The phase variation: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nat. Rev. Microbiology*, **8**, 196–206.
18. Seib, K.L., Jen, F.E., Tan, A., Scott, A.L., Kumar, R., Power, P.M., Chen, L.T., Wu, H.J., Wang, A.H., Hill, D.M., Luyten, Y.A., Morgan, R.D., Roberts, R.J., Maiden, M.C., Biotano, M., Clark, T.A., Korlach, J., Rao, D.N. and Jennings, M.P. (2015) Specificity of the ModA11, ModA12 and ModD1 epigenetic regulators N(6)-adenine DNA methyltransferases of *Neisseria meningitidis*. *Nucleic Acids Res.*, **43**, 4150-4162.
19. Blakeway, L.V., Power, P.M., Jen, F.E., Worboys, S.R., Boitano, M., Clark, T.A., Korlach, J., Bakaletz, L.O., Jennings, M.P., Peak, I.R. and Seib, K.L. (2014) ModM DNA methyltransferase methylome analysis reveals a potential role for *Moraxella catarrhalis* phase variations in otitis media. *FASEB J.*, **12**, 5197-5207.
20. Miroux, B. and Walker, J.E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.*, **260**, 289-298.
21. Murray, I.A., Clark, T.A., Morgan, R.D., Biotano, M., Anton, B.P., Luong, K., Fomenkov, A., Turner, S.W., Korlach, J. and Roberts, R.J. (2012) The methylomes of six bacteria. *Nucleic Acids Res.*, **40**, 11450–11462.

22. Bhagat, Karishma (2017) Charactersation of short sequence repeats of Type III restriction modification enzyme MbolIII. Masters dissertation.
23. Van den Ent, F. and Lowe, J. (2006) RF cloning : A restriction-free method for inserting target genes into plasmids. *J. Biochem. Biophys. methodsiochemical Biophys. methods*, **67**, 67–74.
24. Janscak, P., Sandmeier, U., Szczelkun, M.D. and Bickle, T.A. (2001) Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I. *J. Mol. Biol.*, **306**, 417-431.
25. Meisel, A., Bickle, T.A., Krüger, D.H. and Schroeder, C. (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature.*, **355**, 467–469.
26. van Aelst, K., Tóth, J., Ramanathan, S.P., Schwarz, F.W., Seidel, R. & Szczelkun, M.D. (2010). Type III restriction enzymes cleave DNA by long-range interaction between sites in both head-to- head and tail-to-tail inverted repeat. *Proc. Natl Acad. Sci. USA.*, **107**, 9123–9128.
27. Wyszomirski, K.H., Curth, U., Alves, J., Mackeldanz, P., Möncke-Buchner, E., Schutkowski, M., Kruger, D.H. & Reuter, M. (2012). Type III restriction endonuclease EcoP15I is a heterotrimeric complex containing one Res subunit with several DNA-binding regions and ATPase activity. *Nucleic Acids Res.*, **40**, 3610–3622.
28. Butterer, A., Pernstich, C., Smith, R.M., Sobott, F., Szczelkun, M.D. and Toth, J. (2014). Type III restriction endonuclease are heterotrimeric: comprising one helicase-nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Res.*, **42**, 5139–5150.
29. Moncke-Buchner, E., Rothenberg, M., Reich, S., Wagenfuhr, K., Matsumura, H., Terauchi, R., Kruger, D.H., and Reuter, M. (2009) Functional characterization and modulation of DNA cleavage efficiency of type III restriction endonuclease EcoP15I in its interaction with two sites in the DNA target. *J. Mol. Biol.*, **387**, 1309-1319.



30. Chand, M. K., Nirwan, N., Diffin, F.M., van Aelst, K., Kulkarni, M., Pernstich, C., Szczelkun, M. & Saikrishnan, K. (2015). Translocation-coupled DNA cleavage by the Type IIS restriction-modification enzymes. *Nature Chemical Biology.*, **11**, 870-877.
31. Seib, K. L., Peak, I. R. & Jennings, M. P. (2002) Phase variable restriction-modification systems in *Moraxella catarrhalis*. *FEMS Immunol. Med. Microbiol.* **32**, 159–165.
32. Gupta, Y.K., Chan, S.H., Xu, S.Y., and Aggarwal, A.K. (2015) Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat Commun.*, **6**, 7363.

# **CHAPTER 5**

## **Structure of Mod subunit of MboIII**

## Chapter 5

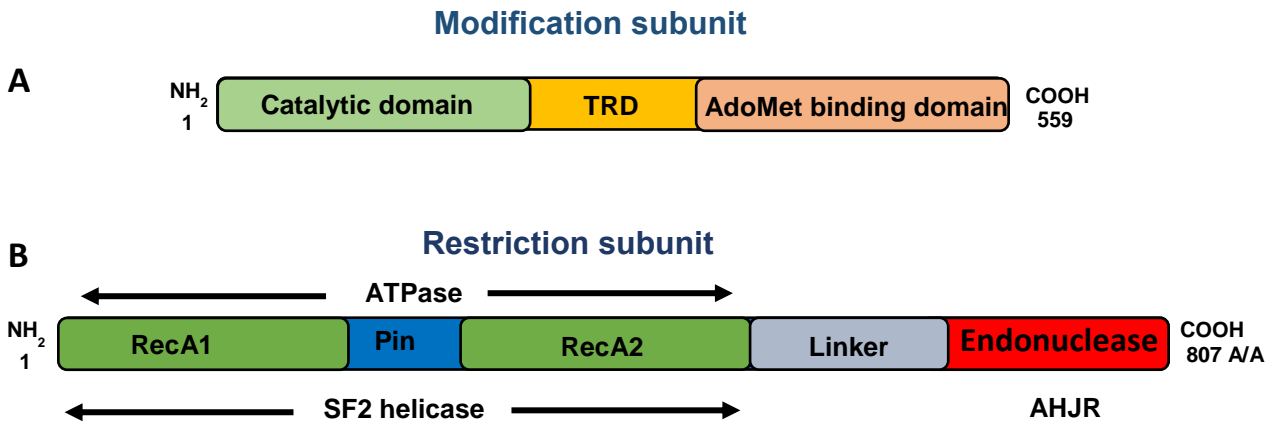
### Structure of Mod subunit of MbolIII.

#### 5.1 Introduction

EcoP1I and EcoP15I are two prototypes of Type III RM enzymes. These are heterotrimeric complexes composed of dimeric Mod and a single Res subunit. EcoP1I and EcoP15I are large MW protein with a molecular weight (MW) of ~260kDa (1). The large size accompanied by flexible and unstructured regions could have hindered our efforts to improve crystal quality to get better diffraction data. While I was carrying out crystallographic studies on EcoP1I and EcoP15I, I simultaneously looked for an alternative model system of Type III RM enzymes. The search was aimed at looking for a smaller Type III RM enzyme. The assumption was that the smaller size means lesser flexible and unstructured regions. This system, I felt, would yield better diffracting crystals.

Analysis of the genomic DNA of *Mycoplasma bovis* revealed the presence of Type III RM enzyme, which had a smaller Mod and Res subunits. Compared to EcoP1I and EcoP15I where the Mod subunit is ~ 74 kDa in size, the Mod subunit in MbolIII ~64 kDa. Similarly, the Res subunit of EcoP1I and EcoP15I has a MW of ~111kDa, while Res subunit is ~94 kDa in size. As described in Chapter 4, I amplified the corresponding *mod* and *res* genes and biochemically and biophysically characterized the holoenzyme, which was named MbolIII. The recognition sequence of MbolIII was identified to be YAATC (Y=T/C), with the enzyme methylating the second adenine in the recognition sequence. SEC-MALS experiment revealed that MbolIII is a heterotrimeric complex composed of two Mod subunits and a single Res subunit with MW close to 227 kDa. The Mod subunits on their own existed as dimer, while the Res subunit on forming a complex with dimeric Mod subunit cleaves foreign DNA in presence of ATP. The Mod subunit of MbolIII belongs to  $\beta$ -class of N6mA methyltransferase. Members of this class are characterized by the presence of nine conserved motifs with IV, V, VI, VII, VIII motifs forming the catalytic domain present at the N-terminal end, target recognition domain (TRD) in the middle, and motifs X, I, II, III forming the C-terminal AdoMet binding domain (Figure 5.1 A)(2, 3). The

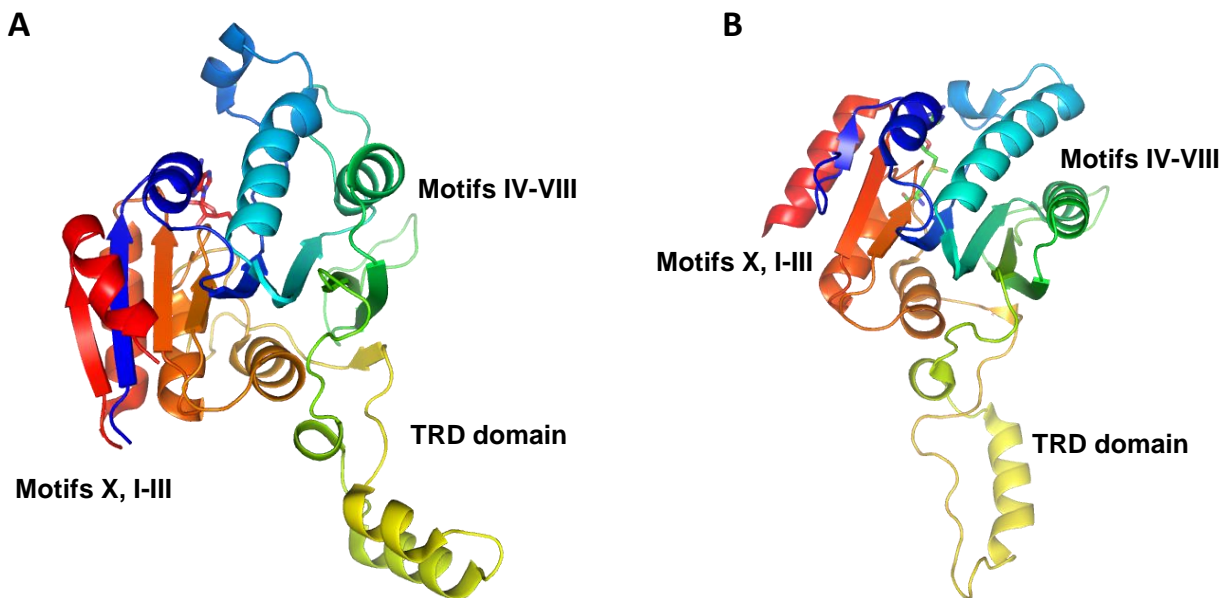
Res subunit of MboIII contains an N-terminal ATPase domain containing SF2 helicase motifs, and a C-terminal endonuclease domain (Figure 5.1 B)( 4, 5).



**Figure 5.1. Domain organization of MboIII.** (A) The modification (Mod) subunit contains an N-terminal catalytic domain with a target recognition domain (TRD) in the middle and an AdoMet binding domain at the C-terminal end. (B) The restriction subunit contains an ATPase domain at its N-terminal end and a C-terminal endonuclease domain with the linker connecting the two.

After successful characterization of MboIII, we started crystallization studies on pure MboIII, Mod and Res. Initial crystallization trials of MboIII and Res were unsuccessful and did not yield any crystals. However, we were successful in obtaining an initial crystallization condition for Mod. Due to limited time, I did not pursue crystallization of MboIII and Res and instead focused on improving the crystals of Mod. MboIII Mod belongs to  $\beta$ -class of N6mA methyltransferases. In literature, there are two other  $\beta$ -class of methyltransferases (Rsrl and Mboll) whose crystal structures have been determined (6, 7). Rsrl is Type II RM enzyme from *Rhodobacter sphaeroides* which recognizes GGATTC and methylates second adenine in the recognition sequence (6). Rsrl functions as a monomer and has a long V-shaped protruding loop forming the DNA binding region. The tip of the loop contains charged and hydrophobic amino acids, which could interact with the DNA through the hydrophobic or electrostatic interaction. The protruding loop containing TRD is located on the opposite side of the protein containing the active site indicating a large conformational change that will be required upon binding to DNA (Figure 5.2 A) (6). Mboll is a Type IIS restriction enzyme from *Morexela bovis* and is the smallest

$\beta$  class of methyltransferase whose structure has been determined (7). MbolI is a dimer with overall architecture similar to RsrI. The core of MbolI is composed of eight  $\beta$ -strands flanked by six  $\alpha$ -helices, three on each side of the protein (7). The two  $\alpha$ -helices are a part of long protruding loop which is analogous to the loop forming the DNA binding region in RsrI (Figure 5.2 B). Similar to RsrI, the loop contains the charged and hydrophobic residues used in binding to DNA. The AdoMet binding region is present at the core of the structure formed by three  $\alpha$ -helices. Again the TRD lies opposite to the catalytic motif indicating a large conformational change in protein on binding to DNA (Figure 5.2 B) (7).



**Figure 5.2. Ribbon diagram of RsrI and MbolI.** (A) Overall structure of the RsrI monomer. The TRD domain contains a long V-shaped protruding loop (Yellow). The catalytic and AdoMet binding motifs are located on the opposite side of TRD. (B) Overall structure of MbolI monomer. The TRD domain in MbolI also contains a long V-shaped protruding loop. Similar to RsrI the catalytic and AdoMet binding motifs are located on the opposite side of TRD.

Type III RM enzymes bind asymmetric DNA sequences with the enzyme methylating a single adenine in the recognition sequence of one strand (8). In Type II RM enzymes, Mod subunit acts as a dimer but recognizes symmetric DNA sequences with each monomer recognizing and methylating target base in both the strands of DNA. Despite having two Mod subunits, Type III RM enzymes bind asymmetric DNA sequence and methylate only a single adenine. How the two monomers crosstalk with each other

to perform specific DNA binding and target base methylation is not known. Some explanation to this question came with the availability EcoP15I structure with DNA (9). The structure gave an unprecedented insight on how dimeric methyltransferases work. From the structure, it was noticed that one of the Mod subunits takes the role of site-specific recognition of DNA, while the other Mod subunit acts as a methyltransferase and catalyzes the transfer of methyl group to target adenine (9). Although the structure of EcoP15I with DNA gave first insights into the working of dimeric methyltransferase, however, there are many questions that remain unanswered. Like, how is it decided which Mod subunit takes the role of site-specific recognition and other Mod subunit takes the role of methyltransferase. What is the molecular mechanism behind such division? The structure of EcoP15I with DNA was solved in the absence of AdoMet. As there is no structure of Type III RM enzyme without DNA, we do not have a structural snapshot of the methyltransferase in absence of DNA.

To address these questions we crystallized Mod subunit of MbolIII in the absence and presence of AdoMet analog sinefungin. We successfully solved the structure of Mod subunit to 3.2 Å. The structure led identification of loop, which gets ordered in presence of sinefungin. This chapter discusses the crystallographic studies and structural analysis of MbolIII Mod.

## 5.2 Materials and Methods

### 5.2.1 Protein Purification

MbolIII *mod* was cloned into high expression vector pRSF with an N-terminal 6XHis tag as described in Chapter 4. *E. coli* BI21(DE3) cells were transformed with recombinant vector pRSF*mod* having kanamycin as a selection marker. A 4 L culture of 1x LB was grown at 37°C until the OD<sub>600</sub> of the culture reached 0.5. The culture was shifted to 18°C and induced at an OD<sub>600</sub> of 0.6 with 1 mM IPTG. The culture was further grown for 12 h after which culture was harvested and pelleted down. The culture was resuspended in 200 mL lysis buffer (50 mM Tris-Cl pH 8, 500 mM NaCl, 10% glycerol, 10 mM MgCl<sub>2</sub>, and 1 mM DTT). The resuspended culture was lysed using sonication in 3 mins cycle with 1

sec on and 3 sec off. To ensure complete lysis of cell the sonication cycle was repeated two times. Also to prevent generation of heat during sonication the entire process was carried out at 4°C. The lysate was spun at 37000 rpm at 4°C for 1 h using Optima XE ultracentrifuge (Beckman Coulter). The resulting supernatant was loaded on to the 5 mL Ni-NTA column pre-equilibrated with Buffer A (50 mM Tris-HCl pH 8, 500 mM NaCl, 15 mM imidazole). The protein was eluted from the Ni-NTA column with stepwise gradient of higher concentration of imidazole Buffer B (50 mM Tris-HCl pH 8, 500 mM NaCl, 500 mM imidazole). The eluted fractions containing protein were pooled and dialyzed against B-50 (50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM DTT) for 8 h at 4°C. After dialysis the protein was centrifuged using Avanti J-26XP (Beckman Coulter Life Sciences) for 30 minutes at 4°C to allow any aggregates to precipitate. The protein was loaded onto the Mono Q 10/100 GL column (GE Healthcare) preequilibrated with B-50. The protein was eluted from Mono Q 10/100 GL with a linear increasing concentration of B-1000 (50 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM DTT). The fractions containing pure protein were pooled together and concentrated using Viva spin column (MWCO 30kDa, GE healthcare). As a last step of purification to remove further impurities and to check the homogeneity of protein size exclusion chromatography (SEC) using Superdex 200 10/300 (GE Healthcare) was carried out. The fractions from SEC were pooled together and concentrated using Viva spin column (MWCO 30kDa, GE healthcare) and stored at -80°C until further use.

### *5.2.2 Purification of selenomethionine (Se-Met) derivative of Mod subunit*

To obtain phase information the selenomethionine (Se-Met) derivative of Mod subunit (Mod<sup>Se-Met</sup>) was purified. Mod<sup>Se-Met</sup> subunit was purified in the similar way as the wild type enzyme, except that during all steps of purification the concentration of DDT was kept at 5 mM. Efforts were made to always obtain the purest protein for crystallization.

### *5.2.3 Generation of methionine mutants of Mod*

Single methionine mutant (Mod<sup>L52M</sup>) and a double methionine mutant (Mod<sup>K36M,L52M</sup>) were generated using overlap PCR. The primers used for incorporation of methionine residues are listed in the Table 5.1. The amplified inserts were double digested with XhoI and NcoI restriction enzymes and were ligated into high expression vector pRSF using T4 DNA

ligase (New England Biolabs). The positive clones were only confirmed after complete sequencing of the gene. The selenomethionine derivatives of the methionine mutants of Mod subunit were purified using a similar protocol as the wild-type enzyme.

**Table 5.1** Primers used for generation of methionine mutants of Mod subunit

| Name    | Sequence (5'->3')          | Length |
|---------|----------------------------|--------|
| RevK36M | GTTGTCCTCCATTTTATCAATGATTG | 26     |
| RevL50M | CTCTTTTGATCATAAGTTGGTAAAC  | 25     |

#### 5.2.4 Crystallization of Mod subunit in presence and absence of sinefungin.

Initial crystallization trials were carried out using 10 mg/ml of Mod subunit in presence of 1 mM sinefungin under crystallization buffer (10 mM Tris-HCl, pH 8, 100 mM KCl, 1 mM DTT). The crystallization mix was kept on ice for 15 minutes, spun briefly at 15000 rpm at 4°C. The supernatant was transferred to a new tube and crystallization screening was carried out. 100 nL of Mod subunit was mixed with 100 nL reservoir buffer with the help of TTP LABTECH LTD mosquito crystal robotic liquid handler. The plates (MRC 2 Well Crystallization Plate UVP) were set up by sitting drop method and kept at 18°C. We screened 15 different commercially available screens from Hampton and Molecular dimensions each screen having 96 different conditions. The crystals obtained using the 96 well plates were small in size (See section 5.3.1 for details). To improve the size of the crystals we increased the size of the drop to 2µL (1 µL Protein + 1 µL reservoir buffer) and set up hanging drop method using Hampton 24 well crystallization plates. The volume of the reservoir was kept at 440 µL. To further increase the size of crystals and to reduce nucleation we decreased the concentration of Mod subunit from 10 mg/mL to 7 mg/mL.

#### 5.2.5 X-ray data collection and processing

Crystals were first screened for diffraction quality in house X-ray diffraction using a Rigaku MicroMax 007 X-ray generator with a Mar Research 345D detector. The crystals were soaked in 30% ethylene glycol. The crystals that diffracted best at home source were retrieved and taken for diffraction studies at synchrotron facilities at Diamond Light Source (DLS) Oxfordshire, UK, and European Synchrotron Research Facility (ESRF), Grenoble.



The diffracted data was indexed and processed using XDS (10). The data was processed scaled and merged using AIMLESS (11, 12). Model building was carried out using COOT (13).

### 5.2.6 SAD phasing, structure solution and refinement

The structure of ModIII Mod subunit was solved using experimentally determined SAD phases to 3.2 Å. The selenium positions were identified using the program SOLVE-RESOLVE (14) in the Phenix suite of crystallography programs (15). In all, 16 methionines were located in the Mod dimer. A simple Fourier map calculated using the SAD phases was used to manually build the structure of Mod. Initial electron density maps were visualized and further model building was carried out in COOT (13). Structure refinement was initially carried out using REFMAC (16) and subsequently by *phenix.refine* (15). Structure visualization and figures were generated using PyMOL (<http://www.pymol.org>)

## 5.3 Results:

### 5.3.1 Initial optimization and crystallization of Mod subunit

Purified Mod subunit was taken and crystallization trials were initiated in the absence and presence of sinefungin. Crystals appeared in only two conditions. The composition of the conditions are mentioned in Table 5.2. Condition 1 was optimized with respect to PEG 8K, pH of buffer and protein concentration. The best crystals were obtained at 0.1 M Tris pH 8 and 8% PEG8K. To reduce nucleation the concentration of protein was reduced from 10 mg/ml to 7 mg/ml. The resulting crystals obtained in the optimized condition were bigger in size. The isolated and bigger crystals were flash frozen with ethylene glycol as cryoprotectant and diffraction data were collected using the in-house X-ray diffractometer. For cryoprotection, we gradually increased the cryoprotectant in the drop from 5% to 30% ethylene glycol over 4 steps at 30 minutes time interval. During this time, the concentration of ethylene glycol in the reservoir buffer was kept constant at 30%. Subsequent to this, the crystals were flash frozen and diffraction experiments carried out. The best diffracting crystals at home source, which diffracted to 6 Å, were retrieved and

taken to synchrotron for data collection. The crystals at synchrotron source diffracted to 4.2 Å. Further efforts to increase the resolution of the crystals were not successful. As a second strategy, we started to crystallize Mod subunit in the second condition. The concentration of PEG 3350, proline, and Na-HEPES in condition 2 was optimized. It was noticed that best crystals were obtained at 10 to 11% PEG 3350, 0.1 to 0.125 M proline and 0.1 M Na-HEPES. The crystals obtained in condition 2 morphologically looked similar to that obtained in condition 1, however these crystals diffracted to better than 3.3Å (Figure 5.3). They also had a different unit cell dimension and space group. Table 5.3 provides the data collection and processing statistics of crystals obtained in condition 1 and condition 2.

**Table 5.2:** Composition of initial condition where initial crystals were observed

|                 | Condition 1 | Condition 2     |
|-----------------|-------------|-----------------|
| Precipitant Mix | 8% PEG 8K   | 10 % PEG 3350,  |
| Buffer          | 0.1 M Tris  | Na-HEPES, 0.1 M |
|                 |             | 0.1 M Proline   |
| pH              | 8.5         | 7.5             |

### 5.3.2 Experimental phase determination

To obtain phase information we purified the Se-Met derivative of Mod subunit (Mod<sup>SeIMet</sup>). Mod<sup>SeIMet</sup> crystallized under the same conditions. One of these crystals obtained using condition 2 diffracted to 3.4 Å. While analyzing the sequence, we noted that a large stretch of the amino acids at the N-terminus did not have methionine. I decided to incorporate two methionines in this region to help experimental phasing and also as a means of obtaining sequence registry during model building. The methionine residues were incorporated by mutating K36 and L50 (Mod<sup>K36M,L50M</sup>). We purified SeI-Met derivative of Mod<sup>L50M</sup> and Mod<sup>K36M,L50M</sup> and crystallized using condition 2. One of the crystals of Mod<sup>L50M</sup> diffracted to 3.2 Å at a synchrotron source. Table 4 represents the data collection and refinement status of the double mutant of Mod subunit



**Figure 5.3. Crystals obtained after optimization of initial condition.** (A) Single crystals obtained in condition 1. (B) Single crystals obtained in condition 2. Morphologically the crystals look identical to each other, however crystals in condition 1 belong to tetragonal space group and diffracted poorly. The crystals in condition 2 belong to orthorhombic space group and diffracted better.

**Table 5.3:** Data collection and processing statistics

| Data collection                | Crystals in condition 1 | Crystals in condition 2                       |
|--------------------------------|-------------------------|-----------------------------------------------|
| Space group                    | P4 <sub>3</sub>         | P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> |
| Cell dimensions<br>a, b, c (Å) | 94.8,94.8,191.2         | 95.4,101.2, 188.7                             |
| α, β, γ (°)                    | 90, 90, 90              | 90, 90, 90                                    |
| Resolution (Å)                 | 95-4.2 (4.7-4.2)        | 70-3.4 (3.63-3.4)                             |
| R <sub>merge</sub>             | 0.054 (0.696)           | 0.094 (0.75)                                  |
| I/sI                           | 6.2 (1.4)               | 7.4 (1.5)                                     |
| Completeness (%)               | 98.4 (98.7)             | 99.7 (99.7)                                   |

**Table 5.4.** Data collection and refinement statistics of Mod crystallized under condition 2.

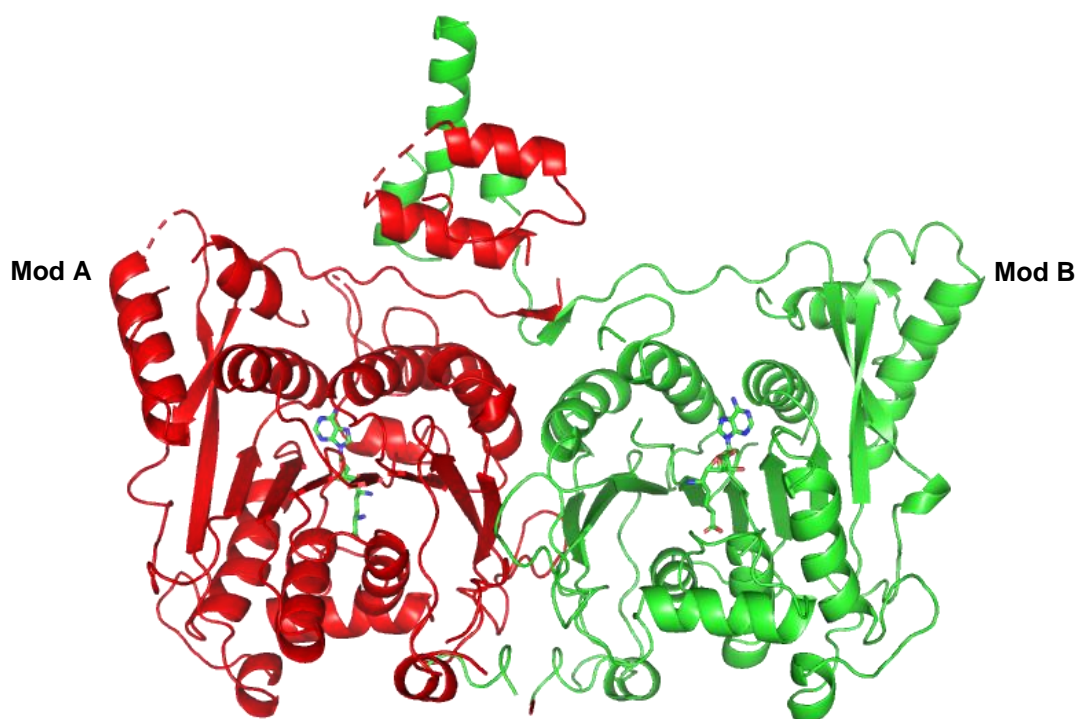
|                 |                                               |
|-----------------|-----------------------------------------------|
| Data collection |                                               |
| Space group     | P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> |
| Cell dimensions |                                               |

|                                    |                        |
|------------------------------------|------------------------|
| a, b, c (Å)                        | 95.3, 100.9, 188.3     |
| $\alpha$ , $\beta$ , $\gamma$ (°)  | 90, 90, 90.            |
| Resolution (Å)                     | 50-3.200 (3.370-3.200) |
| R <sub>merge</sub>                 | 0.115 (1.073)          |
| I/ $\sigma$ (I)                    | 12.8 (2.1)             |
| Completeness (%)                   | 99.5 (99.5)            |
| CC <sub>1/2</sub>                  | 0.997 (0.854)          |
| Refinement statistics              |                        |
| Resolution (Å)                     | 50-3.2(3.258-3.20)     |
| R-work                             | 26.94 (36.3)           |
| R- free                            | 29.61(38.7)            |
| rms deviation                      |                        |
| Bond angle (°)                     | 1.032                  |
| Bond length (Å)                    | 0.004                  |
| Ramachandran plot                  |                        |
| Most favored (%)                   | 93.4%                  |
| Outliers (%)                       | 0.0%                   |
| Average B-factor (Å <sup>2</sup> ) | 112.8                  |

### 5.3.3 Structure of MbolIII Mod.

MbolIII Mod was crystallized both in absence and presence of sinefungin. In the asymmetric unit cell, there are two Mod monomers. The arrangement of the two subunits resembled the arrangement of the subunits in the dimer of EcoP15I and MbolI (9). Hence, we conclude that the subunits in the asymmetric unit constitute the Mod dimer (Figure 5.4). Based on their chain-ids in the coordinates file, the two Mod subunits will be referred to as Mod-A and Mod-B. Figure 5.5 shows the secondary structures present in Mod-A

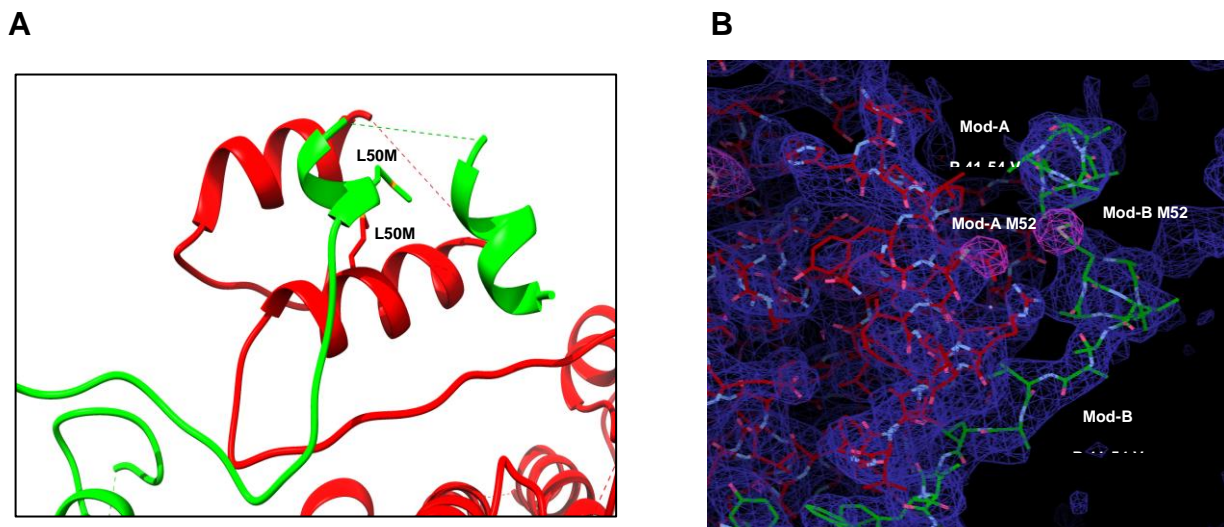
and Mod-B subunit. The overall dimeric interface area, calculated using QT-PISA (17), is about 2243 Å<sup>2</sup>, indicating extensive interaction between the Mod subunits. Dimeric interface at the core is formed by an alpha-6 helix from Mod-A and alpha-3 helix from Mod-B (residues 198-205), two loops (residues 232-242 and 380-396) and a beta-5 strands from Mod-A and Mod-B (residues 217-222). Close to the N-terminus, the region containing residues 55 - 63 from Mod-A and Mod-B also forms extensive interaction with each other. The N-terminal domain (residues 1 to 55) of the Mod subunit could not be built completely because of relatively poor and broken electron densities. However, I could see that the domain was involved in dimerization. The two Mod subunits are related to each other by Pseudo 2-fold symmetry (Figure 5.4). However, the symmetry is broken by the arrangement and the structure of the N-terminal domains of the two subunits (Figure 5.4).



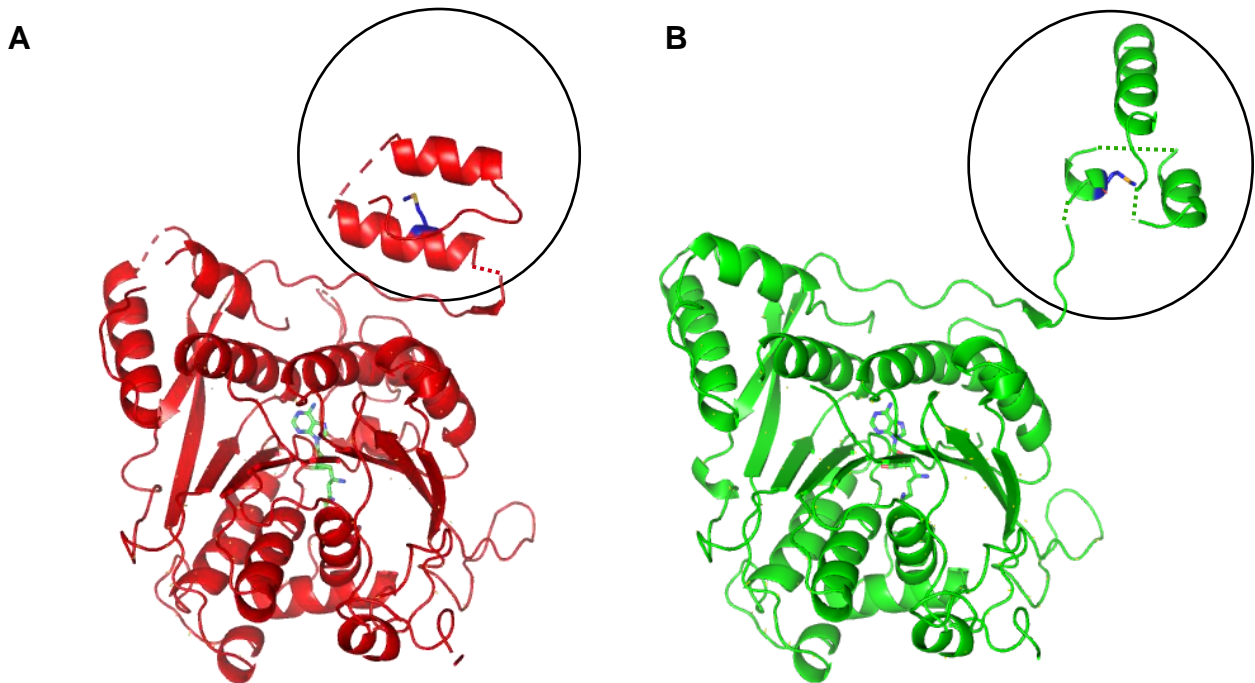
**Figure 5.4. Overall structure of Mod subunit of MbolIII with sinefungin.** (A) The Mod subunit is a dimer composed of two monomers. The two Mod protomers, Mod-A and Mod-B are labelled as red and green respectively. The sinefungin (green) is bound to each protomers. Overall the two protomers are symmetric to each other, however the symmetry between two Mod subunits is broken at their N-terminal domains.



**Figure 5.5. Secondary structures present in Mod dimer.** Secondary structural elements i.e. alpha-helices and  $\beta$ -strands in Mod-A and Mod-B protomers. The secondary structures were identified using 2Struc (Klose et al. 2010).



**Figure 5.6. Asymmetry at the N-terminal domain of Mod protomers.** (A) A ribbon representation of the N-terminal domains of Mod-A (red) and Mod-B (green). The figure highlights the asymmetry in the conformation of two domains (B) Anomalous difference Fourier map (magenta) highlighting selenomethionines at position 50 in the N-terminal domain of two Mod subunits. The  $2F_o - F_c$  map is blue in color. P41- V54 residues of N-terminal domain of Mod-A forms an alpha helix, while the same region in Mod-B forms an extended loop.



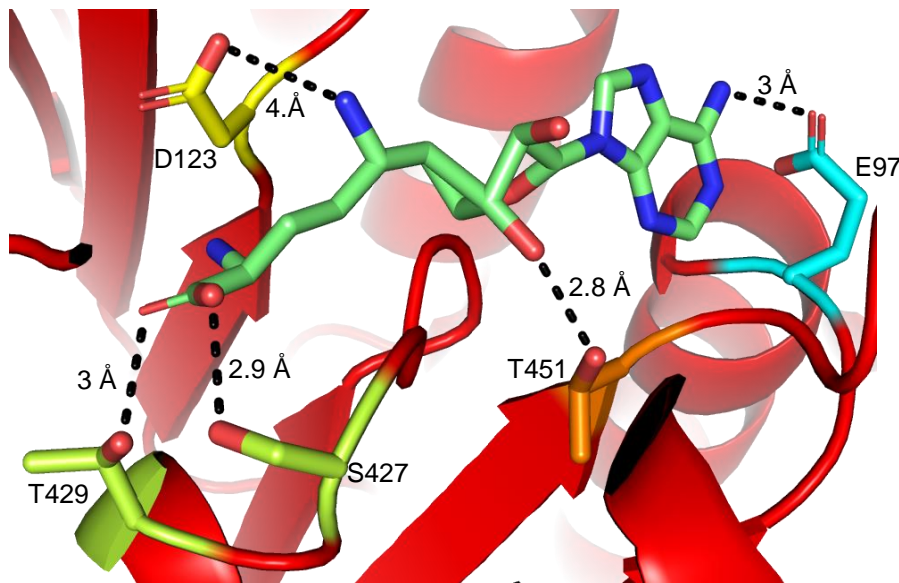
**Figure 5.7. Side view diagram of the Mod-A and Mod-B subunits showing the asymmetry in the N-terminal domain.** (A) Structure of the Mod-A monomer with the circle highlighting the N-terminal domain. The blue color in the N-terminal domain highlights the position of methionine. (B) Structure of the Mod-B monomer with the circle highlighting the N-terminal domain. The blue color in the N-terminal domain shows the position of methionine. The two structures clearly show that the N-terminal domain in Mod-A is part of an alpha helices while in Mod-B it forms an extended loop.

As mentioned earlier, L50 at the N-terminus was mutated to methionine. Using the Se-Met SAD data, we could unambiguously locate the position of this methionine using anomalous Fourier map in both Mod-A and Mod-B. In Mod-A, this methionine is part of one of the two helices, while in Mod-B this residue is part of a loop (Figure 5.6 A and B). Figure 5.7 shows the side view structures of the Mod-A and Mod-B. From the structures it is clear that the N-terminal regions in two monomers have different conformation, and the two monomers are related to each other by Pseudo 2-fold symmetry. The core of the Mod subunit is composed of 9  $\beta$ -strands and 9  $\alpha$ -helices. A loop protruding from one of the  $\beta$ -6 strands from the core and joins back through one of the  $\alpha$ -8 helix. The region in between  $\beta$ -6 and  $\alpha$ -8 could not be modelled in the structure because of no electron density. This region is analogous to that identified in EcoP15I and is involved in site specific DNA binding (9). The catalytic motif DPPY is a part of the loop



emerging from  $\beta$ -3 and the AdoMet binding motif is also part of the loop emerging from  $\beta$ -7. Close inspection of the structure indicates the catalytic and AdoMet binding motif are away from the target recognition domain indicating a large conformation change is required on binding of enzyme to DNA to facilitate DNA binding.

#### 5.3.4. *S*-adenosyl-*L*-methionine binding pocket in Mod subunit.

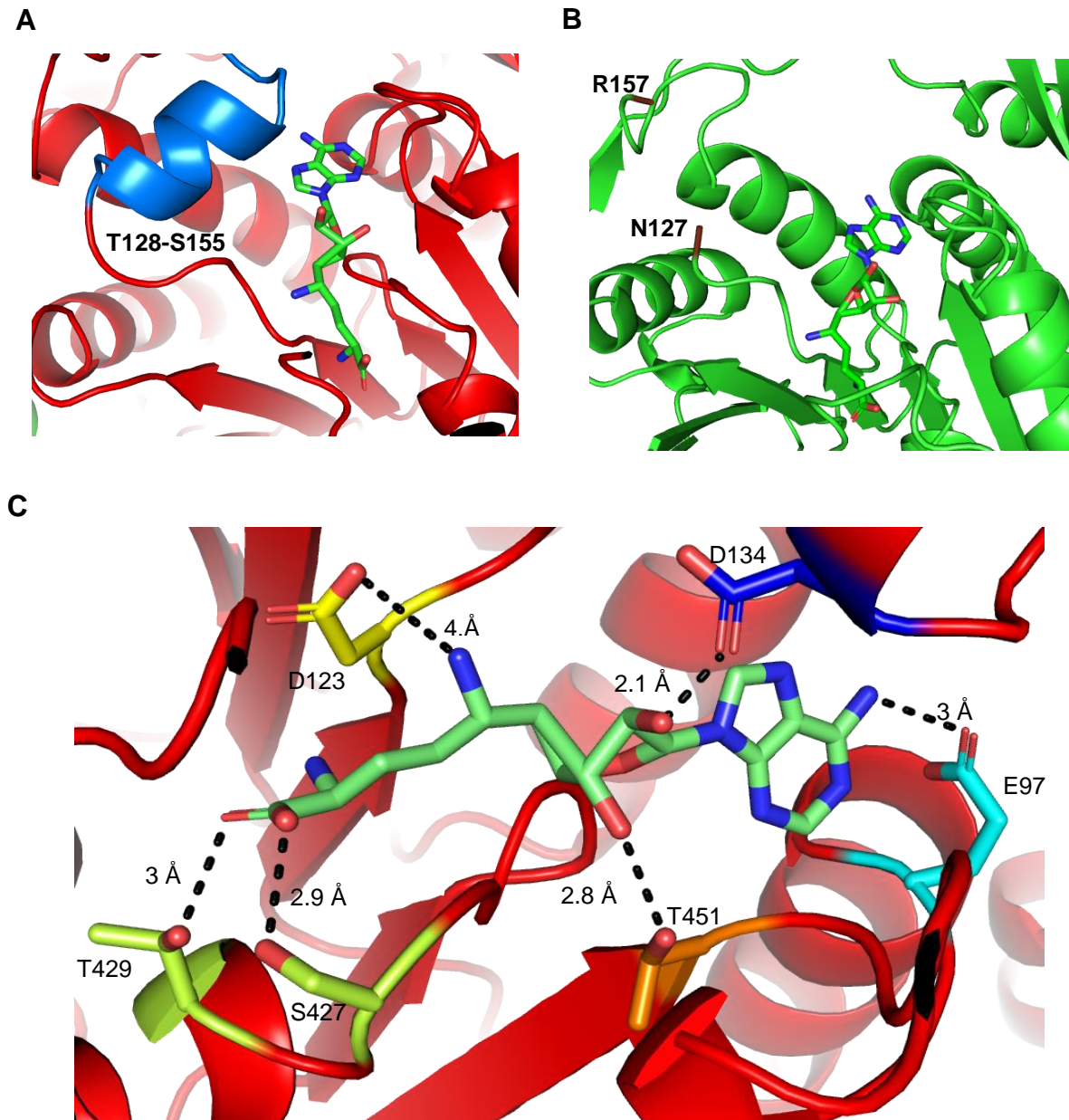


**Figure 5.8. AdoMet binding pocket in MbolIII Mod subunit.** Close up view of the AdoMet binding pocket in Mod-A. The interactions between AdoMet and interacting residues of Mod-A subunit are illustrated. The side chains of interacting residues are shown as sticks. Hydrogen bonds are shown by dashed black lines.

MbolIII Mod structure represents the first structure of the modification unit of a Type III RM enzyme on its own. As mentioned earlier, the structure of both free and sinefungin bound Mod were determined. In the sinefungin-bound structure, I was able to see clear electron density for sinefungin in the two subunits. This structure provided a clear picture of the interactions made by the cofactor with the methyltransferase. The AdoMet in the pocket is surrounded by 16 amino acids. Zooming in to the AdoMet binding pocket, we could see hydrogen bonds formed between  $\beta$ -hydroxyl of S427 and T429 from Motif I with oxygens of the C $\alpha$  of sinefungin. Similarly, we could see hydrogen bond between  $\beta$ -hydroxyl T451 with the 3'-OH group of the ribose sugar of sinefungin. The C $\delta$  hydroxyl



group from E97 forms a hydrogen bond with the NH<sub>2</sub> group of adenine, also the  $\gamma$ -hydroxyl group of D123 of motif IV forms a hydrogen both with the NH<sub>2</sub> group of sinefungin (Figure 5.8). It should be noted that in case of AdoMet the NH<sub>2</sub> group is replaced by methyl group.

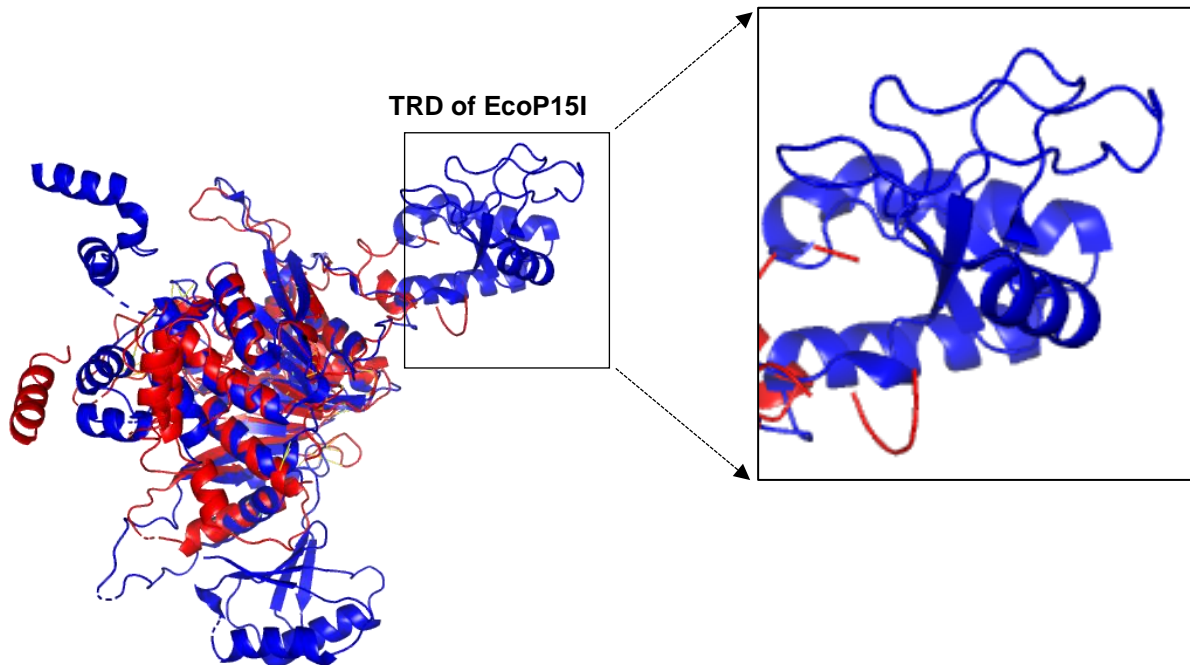


**Figure 5.9. Interaction of helical region (residues 128-155) from Mod-A with AdoMet.** A view of AdoMet binding pocket in Mod-A and Mod-B. (A) In Mod-A the residues 128-155 (in blue color) are ordered and takes a helical conformation (B) In Mod-B this region is disordered. (C) D134 from the helix from Mod-A is within hydrogen bonding distance with the SNF.

There was no significant difference between the free and sinefungin-bound structure of MboIII except for a loop (T128 to S155) located near the AdoMet binding pocket. In the structure without sinefungin, the loop in both Mod-A and Mod-B were disordered (Figure 5.9 B). In the structure with sinefungin, the loop in Mod-A was ordered, while in Mod-B it was disordered. However, the loop was ordered and formed a short  $\alpha$ -helix in Mod-A of the sinefungin-bound structure (Figure 5.9 A). The  $\gamma$ -hydroxyl group of D134 from the structured loop in Mod-A formed a hydrogen bond with the 2'-OH group of ribose ring of sinefungin (Figure 5.9 C). The structure of EcoP15I with DNA has already shown a division of labor between two Mod subunits, where one subunit takes role of methyltransferase and other subunit is used for reading of the DNA recognition sequence (9). However, the molecular basis of how such division between two Mod subunits is established is not known. From our structures we believe that the presence of loop region (T128-S155) in Mod-A provides an additional interaction to firmly hold sinefungin and make Mod-A to take role of methyltransferase. Mod-B will then possibly be used for site-specific binding to DNA.

### 5.3.5 DNA recognition by MboIII

DNA methyltransferase has to perform both site-specific DNA recognition as well as methyltransferase activity. DNA recognition and base flipping of target base is coupled to its methylation. The target recognition domain in MboIII is located in between the N-terminal catalytic domain and the C-terminal AdoMet-binding domain. This region could not be built because of extremely poor electron density. As in EcoP15I, RsrI and MboII where the TRD protrudes from the core of the methyltransferase, we also noticed that a loop emerges from one of the  $\beta$ -6 strands and could form a bilobal TRD domain similar to that in EcoP15I (6, 7, 9) (Figure 5.10). In the MboIII Mod subunit the amino acids containing the region (253-362) is rich in positively charged amino acids similar to that found in TRD region of EcoP15I, RsrI and MboII (6, 7). The positively charged amino acids are likely to be involved in binding to DNA. A careful look at the AdoMet binding pocket and the TRD region of Mod-A indicate that the two lie almost  $180^\circ$  opposite to one another. However the AdoMet binding pocket of Mod-B is placed on the same side of the TRD of Mod-A.



**Figure 5.10. Superimposition of Mod-A of MbolIII with Mod-A of EcoP15I highlighting the location of TRD.** Superimposition of Mod subunit of EcoP15I (Blue) with the Mod subunit of MbolIII (Red) indicates that the core of the two structures superimpose with each other. Zoom in view of the start of TRD region between EcoP15I and MbolIII indicates that the TRD domain in MbolIII will be located at a similar position.

#### 5.4 Discussion.

MbolIII Mod is the first methyltransferase of a Type III RM enzyme whose structure was solved in absence of DNA and presence of sinefungin. Overall the structure is made of two monomers, which are related to each other by a pseudo 2-fold. While the core of the methyltransferase is structurally similar, the N-terminal domain of the two monomers adopt different conformations. The core of the methyltransferase contains the catalytic and AdoMet binding pocket. Compared to MbolI where the AdoMet binding pocket is buried in the core, the AdoMet binding pocket in MbolIII Mod, similar to RsrI and EcoP15I, is relatively open and at the surface of the protein. We noticed that in the AdoMet binding pocket of Mod-A, a part of the region T128-S155 took an alpha-helical conformation, and the residue D134 was within hydrogen bonding distance from the 2'-OH of the ribose of

sinefungin. However the same region in the other Mod-B subunit was completely disordered. This additional hydrogen bond from the loop with the 2'-OH group of ribose may provide additional interaction and hold a tight grip on sinefungin to facilitate methylation of the target adenine. Interestingly, this region was absent in the Mod-A, when we solved the structure in absence of sinefungin. From our structure we believe that additional hydrogen bond from D134 from the loop (T128-S155) in Mod-A will provide an additional interaction to firmly hold sinefungin and make Mod-A to catalyze the transfer of methylgroup from AdoMet to target adenine, while the Mod-B will take the role of site-specific binding to DNA.

To understand whether the region T128-S155 in Mod-A has any effect on the biochemical activities of MboIII, we generated a point mutant (Mod<sup>D134A</sup>) and deletion mutant (Mod<sup>Del</sup>) (17). In Mod<sup>D134A</sup> we mutated D134A which in Mod-A was forming a hydrogen bond with the 2'-OH group of ribose ring of sinefungin. While in Mod<sup>Del</sup> the entire region (T128-S155) was replaced with 4 glycine (17). Comparison of the methylation activities of MboIII, MboIII<sup>D134A</sup> and MboIII<sup>Del</sup> showed that all the three variants methylated the DNA to the same extent (17). This suggested that the loop was not essential for methylation activity. However on performing the nucleolytic assay we found that while MboIII, MboIII<sup>D134A</sup> cleaved the DNA to the same extent, MboIII<sup>Del</sup> cleaved DNA very poorly (17). The DNA cleavage activity of MboIII<sup>Del</sup> was restored when sinefungin was added to the reaction mix (17). It must be noted that the addition of sinefungin did not enhance the nucleolytic activity of either MboIII or MboIII<sup>D134A</sup> (17). Our methylation and nucleolytic experiments on Mod<sup>Del</sup> indicate that the region (T128-S155) do not have any effect on methyltransferase activity but only effects nuclease activity (17). A possible reason of the enhanced nuclease activity of Mod<sup>Del</sup> in presence of sinefungin could be that sinefungin increase the site-specific DNA binding. This also explains why in the presence of AdoMet, Mod<sup>Del</sup> methylates the DNA to the same extent as MboIII and MboIII<sup>D134A</sup>. Another possibility could be that the structured region (T128-S155) in Mod-A allosterically activates the Restriction subunit upon binding to specific DNA. At this stages, we don't have any experimental to prove any of the hypothesis. However in future it would be interesting to see the effect of the deletion T128-S155 on DNA binding affinity using highly sensitive assay (ITC, Anisotropy) and base flipping assay using fluorescent labeled

DNA. This would tell us whether the loop is required in target recognition and/or base flipping.

The TRD domain in MbolIII Mod subunit is formed between residues 236-381, with the region 253-362 residues completely disordered in the structure. Structural alignment of Mod subunit of MbolIII with EcoP15I showed that the two structures overlap very well. Based on structural alignment we noticed that the TRD region of MbolIII will adapt a same conformation as present in EcoP15I Mod subunit. We believe that this region will be used to perform site-specific DNA binding. Although the TRD domain is not modelled completely, we noticed that AdoMet binding pocket of one Mod monomer lies in close proximity to the TRD domain of other Mod monomer.

## **5.5 Conclusion and future directions:**

The work described in the thesis aimed at understanding how ATP hydrolysis activates Type III RM enzymes to endonucleolytically cleave DNA. Type III RM enzymes are heterotrimeric complexes exercising modification and restriction activities. Endonucleolytic cleavage of foreign DNA needs two inversely oriented recognition sites that can be separated by up to 3.5 kb with DNA cleavage happening 25-27 bp downstream of any one of the recognition site. Although two inversely oriented recognition sites are must for DNA cleavage, there were reports in literature about DNA substrates containing a single recognition site being cleaved. The mechanistic details, of how such a DNA cleavage happens were unknown. A previously proposed model invoked looping of the DNA by Type III RM enzymes as means of communication between the enzymes for successful DNA cleavage. However, later studies using magnetic tweezers ruled out the DNA looping mechanism and proposed 1D diffusion as the mechanism of communication. Using biochemical assays described above, we show that these enzymes not only can execute 1D diffusion but can also cleave DNA via 3D diffusion which could involve DNA looping. Single-site DNA cleavage by Type III RM enzymes happen when two enzyme-DNA complexes cooperate with one another via 3D diffusion. Our biochemical assays with Walker A and B mutants of EcoP1I and EcoP15I revealed that ATP hydrolysis is must for these enzymes to get activated to perform dsDNA break. The

study also gave unprecedented details about how ATP hydrolysis activates the endonuclease domain of these enzymes to perform dsDNA break. The study led to the conclusion that long range communication by Type III RM can happen either by 1D diffusion along the DNA or by 3D diffusion in solution or by looping, and that ATP hydrolysis by both the converging enzymes are essential for nucleolytic activity (Chapter 2).

To gain structural insights into the conformational changes on ATP hydrolysis that activates the enzyme, we carried out crystallographic studies of EcoP1I and EcoP15I with their specific DNA substrates and non-hydrolysable ATP analogues. Our crystallographic effort led us in obtaining a partial structure of EcoP1I-DNA-AMPPNP complex. We were able to obtain a SAD data to 4.6 Å, however the quality of phases was not sufficiently good to build the structure of the nuclease domain. At present we have a map with reasonable electron density for the nuclease domain, which is being used for model building (Chapter 3).

One of the reasons that could have hampered obtaining a higher resolution diffraction data of EcoP1I and EcoP15I is the large size of these enzyme complexes. As alternate model system, I identified and characterized MboIII a new Type III RM enzyme from *Mycoplasma bovis*, which is smaller in size compared to the two prototypes. MboIII is the smallest Type III RM enzyme to our knowledge. MboIII recognition sequence was identified to be YAATC (Y= T/C) with enzyme methylating second adenine in recognition site. SEC-MALS experiments revealed that MboIII is a heterotrimeric complex composed of two Mod subunits and a single Res subunits. A characteristic feature of Type III RM enzymes from *Mycoplasma species* is the presence of SSRs in their *mod* gene. Increase or decrease of SSRs in *mod* gene results in phase variation. Using MboIII as a model, I also showed that change in length of SSR did not affect the biochemical activities of MboIII. This result indicates that the SSR region are plastic in nature. Structural modeling of Type III RM with SSR showed that SSRs in the Mod subunit of Type III RM enzymes are located mostly at surfaces away from DNA binding and catalytic centers (Chapter 4).

We also carried out crystallographic studies on MboIII and were successful in crystallizing Mod subunit of MboIII. The Mod subunit was crystallized in presence of the

AdoMet analogue sinefungin. The structure was determined to a resolution of 3.2 Å. The structure gave insights into the mechanism of how Type III methyltransferase function. As the structure of Mod subunits was solved in presence of sinefungin, it gave crucial insights into the AdoMet binding pocket in Type III RM enzymes. Remarkably, the structure reveals that a loop in Mod-A interacts with the cofactor, while the same region in Mod-B is disordered. Further experiments need to be performed to understand the functionality of the loop. MbolIII, will acts as an alternate system in the lab and in the field of Type III RM enzymes to structurally understand the mechanism of how Type III RM enzymes function.

In summary, the work described here provides a platform to carry out crystallographic studies on EcoP1I and MbolIII with a focus to decipher the structural basis for DNA methylation and cleavage by the multifunctional Type III RM enzymes.

## References.

1. Butterer,A., Pernstich,C., Smith,R.M., Sobott,F., Szczelkun,M.D. and Toth,J. (2014)Type III restriction endonucleases are heterotrimeric: comprising one helicase subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Res.*, **42**, 5139–50.
2. Malone,T., Blumenthal,R.M. and Cheng,X. (1995) Structure- guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferase, and suggests a catalytic mechanism for these enzymes.*J. Mol. Biol.*, **253**, 618–632.
3. Timinskas,A., Butkus,V. and Janulaitis,A. (1995) Sequence motifs characteristic for DNA [ cytosine-N4 ] and DNA [ adenine-N6 ] methyltransferases . Classification of all DNA methyltransferases. *Gene*, **157**, 3–11.
4. Gorbalenya,A.E. and Koonin,E. V (1991) Endonuclease (R) subunits of type-I and type III restriction-modification enzymes contain a helicase-like domain. *FEBS Lett.*,**291**, 277–281.

5. Aravind,L., Makarova,K.S. and Koonin,E. V (2000) Holliday junction resolvases and related nucleases : identification of new families , phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.*, **28**, 3417–3432.
6. Scavetta,R.D., Thomas,C.B., Walsh,M.A., Szegedi.S., Joachumiak,A., Gumpert,R.I. and Churchill,M.E.A. (2000) Structure of RsrI methyltransferase, a member of the N6-adenine beta class of DNA methyltransferases. *Nucleic Acids Res.* **28**, 3950–3961 (2000).
7. Osipuk, J., Walsh,M.A and Joachumiak, A (2003) Crystal structure of MbolIA methyltransferase. *Nucleic Acids Res.* **31**, 5440–5448 (2003).
8. Meisel,A., Bickle,T.A., Kruger,D.H. and Schroeder.C (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature.* **355**, 467-469.
9. Gupta,Y.K., Chan,S.H., Xu,S.Y., and Aggarwal,A.K. (2015) Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat Commun.*, **6**, 7363.
10. Kabsch,W. (2010) XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D66**, 125–132.
11. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D69**, 1204–1214.
12. Evans,P. (2006) Scaling and assessment of data quality. *Acta Cryst*, **62**, 72–82.
13. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 486–501.
14. Terwilliger,T. (2004) SOLVE and RESOLVE: automated structure solution, density modification, and model building. *J. Synchrotron Rad.* (2004). **11**, 49-52
15. Adams,P.D., Afonine,P. V, Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 213–221.
16. 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) *REFMAC 5* for the



refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **67**, 355-367.

17. Gopinath, A. (2018). Mutagenesis and biochemical characterization of signature motifs in Type III RM enzyme. Masters dissertation, IISER Pune.