

---

**Assembly and architecture of the  
modification-dependent restriction enzyme  
McrBC**

---

*A thesis*

*Submitted in partial fulfilment of the requirements*

*of the degree of*

**DOCTOR OF PHILOSOPHY**

**By**

**NEHA NIRWAN**

**20113115**



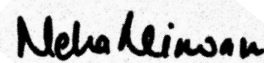
**INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH  
PUNE**

**APRIL 2018**

## DECLARATION

I declare that this written submission represents my research work in my own words and where others ideas or works 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: 30-12-2017

A handwritten signature in black ink that reads "Neha Nirwan". The signature is written in a cursive style and is positioned above a horizontal line.

Neha Nirwan

Reg. No: 20113115

## Acknowledgements

First and foremost I would like to thank my PhD supervisor Dr. Saikrishnan Kayarat for his constant guidance and support. Being his second PhD student I got the opportunity to learn all the skills first hand. I would like to thank him for never losing patience and always being available to discuss any problem. Throughout my PhD he encouraged me to think out of the box and tackle problems as independently as possible. He had been thoroughly involved with the project and that always kept me focused. There were lots of challenges in this project but I never felt helpless and for that I am extremely grateful to him.

I would like to thank Dr. Gayathri Pananghat who had been a great RAC member. We share lab space and lab meetings with Gayathri's lab. I had numerous discussions with her about protein purification, biochemical assays, crystallisation etc. Her invaluable feedback helped me in improving the design of my experiments and analysis of data.

I would like to thank Professor Szczelkun and his lab at University of Bristol, UK where I carried out biophysical studies using stopped-flow spectroscopy. I learnt a lot of biochemistry and biophysical analysis from him and thoroughly enjoyed my stay in his lab.

I would like to thank Dr. Vinothkumar Kutti at LMB, MRC, Cambridge, UK for collecting cryo-EM data for McrB and McrBC and helping us in unravelling the architecture of the complex.

I would like to thank Dr. Shekhar Mande for providing critical reviews on the work during my RAC meetings and allowing me to carry out some experiments in his lab.

I would like to thank my lab mates - both SK Lab and G3 Lab- for maintaining a very amicable work and support system. It was great sharing lab space and equipment with you guys. I will miss our lab a lot.

I would like to thank my friends at IISER pune- Darshika, Srishti and Ayantika for helping me settle in IISER when I first moved away from home. I would like to thank

Darshika for being there when I went through many depressing patches during my PhD. She is one of the people who helped me get through everything and I could not have asked for a better friend. Srishti, in her own way, helped me learn many life lessons and despite our numerous fights, I will always cherish the time we spent together.

I would like to thank the entire biology department and IISER Pune, for providing an ambient research environment. Many thanks to the biology staff that work tirelessly to provide us with working equipment and reagents.

I would like to thank the funding agencies- CSIR and Newton Bhabha Fellowship for providing money to finance work described in this thesis.

I would like to thank my parents, sister and brother for always believing in me when I doubted myself. I would like to thank Anoop who always provided an ear to all my random rants about life and work. Their constant support and love made the hardships of PhD quite easy to endure.

Last but not the least, I would like to thank McrBC for being spoilt, ill-behaved and challenging enough that I got amazing opportunities to learn many techniques along with patience and perseverance.

*“Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.” - Marie Curie*

# CERTIFICATE

Certified that the work incorporated in the thesis entitled ("**Assembly and architecture of the modification-dependent restriction enzyme McrBC**") submitted by Neha Nirwan was carried out by the candidate, under my supervision. 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.

Date: 11-04-2018

Dr.Saikrishnan Kayarat

(Supervisor)

## TABLE OF CONTENTS

<b>LIST OF FIGURES</b>	<b>x</b>
<b>LIST OF TABLES</b>	<b>xiii</b>
<b>ABBREVIATIONS</b>	<b>xiv</b>
<b>ABSTRACT</b>	<b>xvii</b>
<b>SYNOPSIS</b>	<b>xviii</b>
<b>1. Chapter 1: Modification dependent restriction enzyme McrBC</b>	<b>1</b>
1.1 General Introduction	1
1.2 The Restriction-Modification (RM) System	2
1.3 Classification of RM Systems	3
1.4 McrBC: A Type IV MDR enzyme	7
1.5 The AAA+ Superfamily	17
1.6 Future perspective and Scope of the Thesis	30
<b>2. Chapter 2: Purification, assembly and biochemical characterization of McrBC complex</b>	<b>49</b>
2.1 Introduction	49
2.2 Materials and Methods	51
2.2.1 Cloning of mcrB, mcrB $\Delta$ N, mcrC	51
2.2.2 Restriction Free Cloning:	52
2.2.3 Purification	52
2.2.4 Analytical Size Exclusion Chromatography (SEC)	55
2.2.5 Size Exclusion Chromatography coupled Multi-Angle Light Scattering (SEC-MALS)	55
2.2.6 GTP hydrolysis assay	56
2.2.7 DNA binding studies	57
2.2.8 DNA cleavage assay	58
2.2.9 Electron cryomicroscopy (Cryo-EM)	58
2.3 Results	59
2.3.1 Purification of McrB, McrB $\Delta$ N and McrC	59

2.3.2	Oligomeric states of McrB in presence of nucleotides	61
2.3.3	Functional characterization of McrBC	68
2.3.4	Oligomeric structure of McrB from Cryo-EM studies	73
2.4	Discussion	76
2.4.1	Determinants for oligomerization of McrB and McrBC	76
2.4.2	McrB forms a hexameric ring-like structure in presence of GTP	77
2.4.3	A model of McrBC complex	78
2.4.4	Functional implications of the hexameric ring-like structure	79
<b>3.</b>	<b>Chapter 3: Mutational analysis of the AAA+ McrB protein</b>	<b>86</b>
3.1	Introduction	86
3.2	Materials and method	87
3.2.1	Multiple Sequence alignment	87
3.2.2	Mutagenesis and cloning	88
3.2.3	Protein Purification	89
3.2.4	Size Exclusion Chromatography (SEC)	89
3.2.5	Functional Characterisation	89
3.2.6	Dynamic Light Scattering (DLS)	89
3.2.7	Nucleotide binding assay using Mant-GDP	90
3.2.8	Circular-dichroism spectroscopy	90
3.3	Results	91
3.3.1	Multiple Sequence alignment	91
3.3.2	Cloning, mutagenesis, and purification of different mutants	94
3.3.3	Characterisation of Walker B mutants	94
3.3.4	Characterisation of pore loop mutants	102
3.3.5	Characterisation of Sensor II mutants	103
3.4	Discussion	104
<b>4.</b>	<b>Chapter 4: Crystallographic and cryo-EM studies of McrBC</b>	<b>111</b>
4.1	Introduction	111
4.2	Materials and Methods	112
4.2.1	Cloning of different proteins	112
4.2.2	Purification of McrB, McrB $\Delta$ N, McrB $\Delta$ N <sup>wt</sup> and McrC	112
4.2.3	Purification of selenomethionine labeled McrB $\Delta$ N, McrB $\Delta$ N <sup>WT</sup>	114

and McrC	
4.2.4 Purification of McrBC, McrB $\Delta$ NC and McrB $\Delta$ N <sup>WT</sup> C complex	114
4.2.5 Crystallisation	115
4.2.6 X-ray data collection and processing	115
4.2.7 Electron cryomicroscopy (cryo-EM) and image processing of McrB $\Delta$ N <sup>WT</sup> C	116
4.3 Results	119
4.3.1 Purification of McrB $\Delta$ N <sup>WT</sup>	119
4.3.2 Crystallization trials with the full-length McrBC complex	119
4.3.3 Crystallization trials with the truncated protein-McrB $\Delta$ N and McrB $\Delta$ NC complex	120
4.3.4 Self-rotation map	122
4.3.5 Crystallisation trials with truncated protein McrB $\Delta$ N <sup>WT</sup> and McrB $\Delta$ N <sup>WT</sup> C complex	123
4.3.6 The dumbbell-shaped density of McrB $\Delta$ N <sup>WT</sup> C	125
4.3.7 Modeling the structure of McrB $\Delta$ N hexamers	127
4.3.8 Architecture of the McrB $\Delta$ N hexamer	130
4.3.9 Model of McrC	133
4.3.10 Architecture of McrC	135
4.3.11 Interaction of McrC with McrB $\Delta$ N	135
4.4 Discussion	137
4.4.1 A model for McrC-stimulated GTPase activity of McrB	137
4.4.2 The architecture of McrB $\Delta$ NC shows lack of putative sensor II	139
4.4.3 Implication of McrB $\Delta$ NC architecture on probable modes of DNA loading	140
4.4.4 Structure determination using crystal diffraction data	141
<b>5. Chapter 5:Pre-steady-state kinetic studies on McrB and McrBC</b>	
5.1 Introduction	146
5.2 Materials and Methods	148
5.2.1 Protein purification	148
5.2.2 Oligonucleotide	149



5.2.3	Steady-state tryptophan fluorescence measurements	149
5.2.4	Stopped-flow intrinsic tryptophan fluorescence kinetic study	150
5.2.5	Stopped-flow pre-steady-state GTPase rate measurements	150
5.2.6	Stopped-flow DNA anisotropy measurements	152
5.2.7	Steady-state DNA anisotropy measurements	152
5.3	Results	152
5.3.1	McrB shows change in Trp fluorescence in presence of nucleotide	152
5.3.2	Complexation of McrB with McrC does not affect Trp fluorescence	155
5.3.3	GTP hydrolysis follows oligomerization of McrB	156
5.3.4	GTP hydrolysis by McrBC is a cooperative process	158
5.3.5	DNA binding affects the kinetics of GTPase activity of McrBC	160
5.3.6	The effect of DNA binding on the McrB oligomers	162
5.3.7	DNA binding exerts an inhibitory effect on the assembly of McrB oligomers	164
5.3.8	McrC affects DNA binding by McrB	166
5.4	Discussion	167
5.4.1	DNA binding affects the coupling of GTPase reaction rather than its kinetics	167
5.4.2	McrBC does not disassemble and reassemble around the substrate	168
5.4.3	Possible mode of assembly of McrBC complex	169
5.5	Conclusion and Future Directions	170

## LIST OF FIGURES

### Chapter 1:

- Figure 1.1 The NTP independent and NTP dependent RM enzymes
- Figure 1.2 The *mcrbc* locus in *E. coli* genome
- Figure 1.3 The domain diagram of McrBC subunits
- Figure 1.4 McrBC DNA cleavage sites
- Figure 1.5 Schematic of Model proposed for DNA cleavage by McrBC
- Figure 1.6 The substrate diversity of AAA+ family
- Figure 1.7 ATPase site architecture of AAA+ proteins
- Figure 1.8 Topology of different AAA+ Clades
- Figure 1.9 Models for NTP hydrolysis by AAA+ motors
- Figure 1.10 The pore loop of AAA+ proteins.

### Chapter 2:

- Figure 2.1 Purification of McrB, McrB $\Delta$ N and McrC
- Figure 2.2 Nucleotide driven oligomeric assembly of McrBC/McrB $\Delta$ NC
- Figure 2.3 Nucleotide independent-concentration dependent oligomerisation of McrB/McrB $\Delta$ N.
- Figure 2.4 Binding of McrB/McrB $\Delta$ N and McrC in absence of GTP
- Figure 2.5 SEC-MALS chromatogram of McrB
- Figure 2.6 SEC-MALS chromatogram of McrB and McrBC
- Figure 2.7 Concentration dependent GTPase activity of McrB and McrBC
- Figure 2.8 GTPase activity of McrB and McrBC
- Figure 2.9 DNA binding studies with McrB
- Figure 2.10 EMSA showing specific binding of McrB
- Figure 2.11 DNA cleavage assay

- Figure 2.12 Cryo-EM images of McrB and McrB $\Delta$ N in presence of GDPNP
- Figure 2.13 Representative rotational power spectrum for class averages
- Figure 2.14 A Model proposing assembly of the functional McrBC complex

### Chapter 3:

- Figure 3.1 Alignment of McrB with bEBP proteins
- Figure 3.2 Structural and sequence alignment of McrB to identify characteristic AAA+ motifs
- Figure 3.3 Size exclusion chromatography with the McrB Walker B aspartate mutants
- Figure 3.4 Dynamic Light Scattering assay to assess the oligomerization state
- Figure 3.5 Biochemical analysis of Walker B aspartate mutants
- Figure 3.6 Size exclusion chromatography with the McrB Walker B glutamate mutants
- Figure 3.7 SEC-MALS study of McrB E280 mutants
- Figure 3.8 Biochemical analysis of Walker B glutamate mutants
- Figure 3.9 Mant-GDP binding by McrB and its mutants
- Figure 3.10 Circular Dichroism spectroscopy of McrB and its mutants
- Figure 3.11 Size exclusion chromatography with the McrBloop mutants
- Figure 3.12 DNA cleavage activity of McrBK255A mutant.
- Figure 3.13 Biochemical activity of putative Sensor II R404

### Chapter 4:

- Figure 4.1 Cryo-EM images of McrB $\Delta$ NC in presence of GDPNP
- Figure 4.2 Purification of McrB $\Delta$ N<sup>WT</sup>
- Figure 4.3 Images of McrB $\Delta$ N (histidine-tagged) crystals

- Figure 4.4 Self-rotation function map showing stereographic projection of McrB $\Delta$ N+GDPNP in space group P2<sub>1</sub>
- Figure 4.5 Images of McrB $\Delta$ N<sup>WT</sup>C (without histidine-tagged) crystals
- Figure 4.6 Architecture of McrB $\Delta$ NC complex
- Figure 4.7 Superimposition of two McrB $\Delta$ NC maps
- Figure 4.8 Building the model for McrB $\Delta$ N subunit
- Figure 4.9 Goodness of fit between model and electron density map
- Figure 4.10 The asymmetric interface of McrB $\Delta$ N hexamer.
- Figure 4.11 Building a model for McrC
- Figure 4.12 Interaction of McrC with different subunits of McrB $\Delta$ N
- Figure 4.13 Interaction of McrC with different subunits of McrB $\Delta$ N
- Figure 4.14 Model for sequential GTP hydrolysis by McrB $\Delta$ NC.
- Figure 4.15 Position of proposed Sensor II residue R404

## **Chapter 5:**

- Figure 5.1 Possible modes for a close ring molecule to load onto DNA
- Figure 5.2 Steady-state tryptophan fluorescence emission scan
- Figure 5.3 Real-time observation of McrB oligomerization
- Figure 5.4 Relation between GTP hydrolysis and oligomerization
- Figure 5.5 GTP hydrolysis in McrBC show cooperativity
- Figure 5.6 DNA affects the oligomerisation of McrBC
- Figure 5.7 Presence of McrC does not affect DNA induced Trp signal
- Figure 5.8 Effect of DNA on oligomerisation of McrB
- Figure 5.9 Effect of McrC on DNA binding by McrB
- Figure 5.10 Model for DNA loading by McrB/McrBC

## LIST OF TABLES

### Chapter 2:

Table 2.1 List of primers used for cloning McrB, McrC and McrB $\Delta$ N genes

Table 2.2 List of oligonucleotides used DNA binding studies

Table 2.3 Comparison of calculated mass and mass from SEC-MALS

### Chapter 3:

Table 3.1 List of primers for generating McrB mutants

Table 3.2 The Radius of hydration of McrB and its mutants estimated from DLS study

Table 3.3 Table summarising the Mutational analysis of McrB

### Chapter 4:

Table 4.1 Diffraction data-collection statistics of McrB $\Delta$ N+GDPNP crystals

Table 4.2 Diffraction data-collection statistics of McrB $\Delta$ NC+GDPNP crystals

## ABBREVIATIONS

AAA+	<u>A</u> T <u>P</u> ases <u>a</u> ssociated with diverse cellular <u>a</u> ctivity
AdoMet	S-adenosylmethionine
ADP	adenosine diphosphate
ASCE	additional strand conserved E
ATP	adenosine triphosphate
bEBPs	bacterial enhancer binding proteins
CD	circular dichroism
Cryo-EM	electron cryomicroscopy
CTF	contrast transfer function
CTD	C-terminal domain
dsDNA	double-stranded DNA
DLS	dynamic light scattering
DQE	detective quantum efficiency
EMSA	Electrophoretic mobility shift assay
GAP	GTPase activating protein
GDP	guanosine diphosphate
GDPNP	guanosine 5'-[ $\beta$ , $\gamma$ -imido]triphosphate
GEF	guanine nucleotide exchange factor
GTP	guanosine triphosphate
GTP $\gamma$ S	guanosine 5'-[ $\gamma$ -thio]triphosphate
h2i	helix 2 insert
HsdM	host specificity determinant modification

HsdR	host specificity determinant restriction
HsdS	host specificity determinant specificity
ICM	initiator-specific motif
ITP	inositol triphosphate
ICR	immigration control region
LUCA	last universal common ancestor
Mant-GDP	2'/3'-O-(N-methyl-anthraniloyl)-guanosine-5'-diphosphate
Mcr	modified cytosine restriction
MDCC	[2-(1-maleimidyl)ethyl]-7-(diethylamino)coumarin-3-carboxamide
mDNA	methylated DNA
MDR	modification-dependent restriction
7-MG	7-methyl guanosine
Mod	modification subunit of Type III restriction enzymes
Mrr	methylated adenine recognition and restriction
Mtase	methyl transferase
nmDNA	non-methylated DNA
NTD	N-terminal domain
NTP	nucleotide triphosphate
PBP	phosphate binding protein
PCR	polymerase chain reaction
P-loop	phosphate binding loop
PNPase	purine nucleoside phosphorylase
PSI $\beta$ hp	pre-Sensor I $\beta$ -hairpin

Res	restriction subunit of Type III restriction enzyme
RM	restriction-modification
SEC-MALS scattering	size exclusion chromatography coupled multiple angle light scattering
SET	Su(var), Enhancer of zeste, Trithorax
SF3	superfamily 3
SF6	superfamily 6
SRA	SET and Ring finger associated
SRH	second region of homology
ssDNA	single-stranded DNA
STEM	scanning transmission electron microscopy
TRD	target recognition domain
XTP	xanthine triphosphate
UDPG-PPase	uridine diphosphoglucose pyrophosphorylase



## ABSTRACT

Conversion of chemical energy released upon hydrolysis of nucleotides to mechanical action by specialized motors drives many cellular processes. AAA+ NTPase is a prominent family of such motors that power diverse functions, including DNA replication, transcriptional activation, protein degradation or refolding, membrane fusion, microtubule dynamics etc. As a component of the restriction enzyme McrBC, AAA+ motor hydrolyses GTP to drive nucleolytic cleavage of foreign DNA that invade *Escherichia coli*. McrBC recognizes a 5-methylcytosine preceded by a purine (R<sup>m</sup>C) as target site, and cleave DNA containing at least two target sites separated by 40 to 3000 bp. It is a multi-subunit complex of two proteins - McrB and McrC. McrB has a DNA-binding and the AAA+ GTPase domain, while McrC is an endonuclease. As part of my PhD project, I carried out biochemical, biophysical and structural studies of the enzyme complex. As a first step, the subunits and their active complex were purified. Size exclusion chromatography coupled to multi-angle light scattering revealed that in presence of guanine nucleotides, McrB oligomerized to a hexamer. On addition of McrC, a 720 KDa tetradecamer of twelve subunits of McrB and two subunits of McrC was formed. The purified tetradecamer was competent to hydrolyse GTP and cleave DNA. The oligomers of a GTPase-competent but DNA-binding deficient mutant of McrB, and its complex with McrC were crystallized. X-ray diffraction studies were carried out and experimental phases determined. In combination with a low-resolution map obtained using electron cryomicroscopy, the architecture of McrBC was deduced allowing us to understand the structural basis of GTPase stimulation of McrB by McrC. Using stopped-flow spectroscopy, I measured apparent binding affinities of the protein for a minimal DNA substrate and observed the effect of the substrate on the kinetics of GTP hydrolysis. Also, the intrinsic tryptophan fluorescence of the protein was probed to gain insights into the assembly of the complex, and the mode of DNA substrate binding. The thesis presents the results from the study of the assembly and architecture of McrBC.

## **SYNOPSIS**

### **Assembly and architecture of the modification-dependent restriction enzyme McrBC**

Name: Neha Nirwan

Roll Number: 20113115

Name of Supervisor: Dr. Saikrishnan Kayarat

Department: Biology

Date of Registration: August 01, 2011

Indian Institute of Science Education and Research (IISER), Dr. Homi Bhabha Road, Pashan, Pune 411008, Maharashtra, India

#### **Modification dependent restriction enzyme McrBC (Chapter 1)**

McrBC is a modification-dependent restriction enzyme that cleaves DNA containing at least two of target sites  $G/A^mC$ . These sites can be separated by 40 bp to 3000 bp (Sutherland et al, 1992). Endonucleolytic cleavage of DNA is catalyzed by the nuclease McrC coupled to a GTP-dependent AAA+ motor McrB. McrB subunit of McrBC complex is also responsible for recognition of the target sequence. Previous studies showed that McrBC is a complex of fourteen subunits of McrB and two subunits of McrC. Based on biochemical studies it has been proposed when two molecules of McrBC translocating complexes collide, the endonuclease McrC cleaves DNA (Panne

et al, 1999). Along with being the nuclease, McrC also stimulates the GTPase activity of McrB by 30 fold (Panne et al, 1999; Pieper et al, 1997).

As a restriction enzyme, McrBC is unique in employing a AAA+ motor domain to hydrolyze GTP and translocate on DNA. I initiated biochemical, biophysical and structural studies on McrBC to understand the mechanism of action of this multi-subunit AAA+ motor that translocates on dsDNA and is coupled to an endonuclease for its GTPase activity. Chapter 1 in this thesis reviews the brief historical events that lead to the discovery of McrBC followed by an introduction to other well-studied restriction enzymes to highlight the unique features of McrBC. The known biochemical properties of this complex are then reviewed in detail. Next, AAA+ proteins are discussed describing their structural features and categorization to understand the properties of McrB as a AAA+ motor. The information thus discussed lead to open questions in the field and that aided in formulating specific aims of the thesis. Finally, chapter 1 discusses the scope of this thesis.

## **Purification, assembly and biochemical characterization of McrBC complex (Chapter 2)**

The McrB subunit of McrBC complex has been shown to form heptameric rings (Panne et al, 2001). The functional relevance of heptameric state of AAA+ protein has been under debate due to the observation of both heptameric and hexameric states of many AAA+ proteins (Bush & Dixon, 2012). As part of my endeavor to understand the mechanism of AAA+ motor in McrBC, knowing the exact stoichiometry and its functionality was essential. Thus, the stoichiometry of McrB and McrBC complex was re-visited and further biochemical and structural studies were carried out.

For this, I standardized the large-scale purification of the individual subunits-McrB and McrC and assembly of McrBC complex. Additionally, the N-terminal DNA binding domain deletion mutant of McrB (McrB $\Delta$ N) was also cloned, purified and McrB $\Delta$ NC complex was assembled. The purified proteins were tested for their functional activities. The exact stoichiometry of the McrB and McrBC complexes was determined using SEC-MALS. A preliminary investigation of the oligomeric structure of McrB using cryo-EM was also carried out. Chapter 2 describes the detailed methodology for protein purification, complex assembly and biochemical

characterization. The chapter then discusses the results obtained from biochemical characterization and implications of the observations.

### **Mutational analysis of the AAA+ McrB protein: Chapter 3**

McrBC has been characterized under the Clade VI of AAA+ family along with the NtrC1 family. A very detailed functional mutagenesis has been reported in earlier studies on McrBC (Pieper et al, 1997; Pieper et al, 1999a; Pieper et al, 1999b) but most of the characterization was focussed on McrB as a possible canonical GTPase or G-protein. Amongst the five AAA+ motifs, the Sensor II for McrB is not yet known. This motif was not identified in a previous sequence alignment carried out by Iyer et al, 2004. Also, since McrB is proposed to be a dsDNA translocase the loops analogous to pore loops of helicases and translocases (Gai et al, 2016; Pape et al, 2003) have not been identified.

Along with carrying out structural studies, a simultaneous sequence alignment based identification followed by mutagenesis studies was carried out to identify the putative Sensor II and pore loop of McrB. The Walker B residues of McrB were also revisited to confirm the role of the conserved acidic residues in nucleotide binding and hydrolysis. Chapter 3 discusses the results of these mutational characterisation.

### **Crystallographic and cryo-EM studies of McrBC (Chapter 4)**

To understand the mechanism of a multi-subunit complex, it is important to understand the organization and interactions between the subunits of the complex. To gain structural insights into McrBC complex, I carried out crystallization and cryo-EM studies in parallel to the biochemical and biophysical characterization. These studies were carried out with N-terminal DNA binding domain deletion mutant of McrB, McrB $\Delta$ N. Diffraction data of McrB $\Delta$ N+GDPNP at 4.5 Å and McrB $\Delta$ NC-GDPNP diffraction data at 4.4 Å was collected. Simultaneously cryo-EM studies were carried out which yielded a 7.4 Å resolution 3-D map of McrB $\Delta$ NC complex. Chapter 4 discusses the details of my efforts to unravel the McrB and McrBC structure followed by discussion on architecture of McrB $\Delta$ NC complex highlighting the architecture of McrB $\Delta$ N hexamer and its interaction with the endonuclease McrC. This first snapshot of the molecule provided useful insights into the possible mechanism of its GTP hydrolysis.

## **Pre-steady-state kinetic studies of McrB and McrBC: Chapter 5**

Being a AAA+ translocase, assembly of this complex is essential for its function (Joly et al, 2012; Wendler et al, 2012). Following the observations made from the architecture of McrB $\Delta$ NC complex, I carried out pre-steady-state kinetic studies of oligomerization of McrB and McrBC complexes to gain further insights into its assembly. For this millisecond time-resolved stopped-flow spectroscopy was used. Intrinsic tryptophan (Trp) fluorescence of the protein mixed with nucleotide as a signal for assembly and GTPase activity of the protein was measured under similar mixing regimes.

DNA loading by hexameric helicases is an essential and highly regulated process. Many studies have been characterised on these systems to unravel their interactions with DNA but not much is known about dsDNA translocases in terms of their DNA interactions and loading. McrBC has been proposed to be a dsDNA translocase. I also carried out pre-steady-state studies with McrBC to gain insights on DNA binding. GTPase studies combined with intrinsic Trp fluorescence measurements and DNA anisotropy changes upon protein DNA association were carried out. The observations from these studies were combined to decipher plausible pathways for a ring-like molecule (McrB/McrBC) to load onto DNA and understand its dynamics with DNA.

In summary, biophysical and structural studies of the oligomeric structure of McrBC revealed that McrB forms a hexameric ring, while McrBC is a tetradecamer of two McrB rings bridged by a dimer of McrC. The architecture of the assembly provides insights into the mechanism of McrC-stimulated GTP hydrolysis, and a model for the mode of GTP hydrolysis by the AAA+ ring. A combination of Trp fluorescence, pre-steady-state GTPase activity measurements and DNA binding studies using anisotropy fluorescence suggests that loading of DNA onto the McrBC complex could involve either direct binding or McrB ring opening followed by binding.

## References

- Bush M, Dixon R (2012) The role of bacterial enhancer binding proteins as specialized activators of sigma54-dependent transcription. *Microbiology and molecular biology reviews : MMBR* **76**: 497-529
- Gai D, Wang D, Li SX, Chen XS (2016) The structure of SV40 large T hexameric helicase in complex with AT-rich origin DNA. *eLife* **5**
- Joly N, Zhang N, Buck M (2012) ATPase site architecture is required for self-assembly and remodeling activity of a hexameric AAA+ transcriptional activator. *Molecular cell* **47**: 484-490
- Panne D, Muller SA, Wirtz S, Engel A, Bickle TA (2001) The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *The EMBO journal* **20**: 3210-3217
- Panne D, Raleigh EA, Bickle TA (1999) The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. *Journal of molecular biology* **290**: 49-60
- Pape T, Meka H, Chen S, Vicentini G, van Heel M, Onesti S (2003) Hexameric ring structure of the full-length archaeal MCM protein complex. *EMBO reports* **4**: 1079-1083
- Pieper U, Brinkmann T, Kruger T, Noyer-Weidner M, Pingoud A (1997) Characterization of the interaction between the restriction endonuclease McrBC from *E. coli* and its cofactor GTP. *Journal of molecular biology* **272**: 190-199
- Pieper U, Schweitzer T, Groll DH, Gast FU, Pingoud A (1999a) The GTP-binding domain of McrB: more than just a variation on a common theme? *Journal of molecular biology* **292**: 547-556
- Pieper U, Schweitzer T, Groll DH, Pingoud A (1999b) Defining the location and function of domains of McrB by deletion mutagenesis. *Biological chemistry* **380**: 1225-1230
- Sutherland E, Coe L, Raleigh EA (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *Journal of molecular biology* **225**: 327-348
- Wendler P, Ciniawsky S, Kock M, Kube S (2012) Structure and function of the AAA+ nucleotide binding pocket. *Biochimica et biophysica acta* **1823**: 2-14

# Chapter 1

## Modification dependent restriction enzyme McrBC

### 1.1 General Introduction

Bacteriophages are the most abundant organisms in the biosphere, posing a constant threat to their bacterial preys (Clokier et al, 2011). To counter phage attacks, bacteria employ a number of different strategies, such as blocking the attachment and entry of phage DNA, cleaving phage nucleic acid or using abortive infection systems which cause the death of infected cells to protect the bacterial population (Labrie et al, 2010). Amongst these strategies, the phage restriction through nucleic acid cleavage is carried out by specialized restriction-modification (RM) systems. As the name suggests, the RM systems comprise of two components- restriction and modification.

The task of an RM system involves recognition of DNA as foreign or self, followed by activation of endonuclease in case of a foreign DNA or methyltransferase in case of self-DNA. Due to the continuous evolutionary race, phages acquired methods to methylate their genome (in order to evade bacterial restriction) and bacteria acquired another type of restriction system to cleave methylated DNA called Modification Dependent Restriction (MDR) system. In comparison to the classical RM systems, MDR enzymes lack an associated methyltransferase activity (Roberts et al, 2003). Based on their complexity, cofactor requirement and cleavage position with respect to the recognition site, RM enzymes were classified into three classes - Type I, Type II and Type III (Boyer, 1971). Later, new endonucleases were discovered which were difficult to classify under existing classes and thus were classified as Type IV RM enzymes (Bickle & Kruger, 1993; Janulaitis et al, 1992).

Amongst these different types, Type II RM systems are extensively studied and are an integral part of all laboratories using the techniques of molecular biology. The Type II RM systems are different from the rest in that the restriction endonuclease and the cognate methyltransferase function mutually exclusively. Also in contrast to other types, the Type II endonucleases do not require nucleotide triphosphates (NTP) for their DNA cleavage activity and usually exist as homodimers. They specifically recognize 4-8 bp

palindromic sequences and cleave within or close to the recognition site (Orlowski & Bujnicki, 2008) (Figure 1.1 A). The other three classes of restriction enzymes- Type I, Type III and Type IV are also classified together as NTP dependent restriction enzymes (Bourniquel & Bickle, 2002). The work in this thesis describes studies on one of such NTP-dependent restriction enzyme McrBC which belongs to the MDR family of Type IV restriction systems.

The McrBC restriction system is encoded by the Immigration Control Region (ICR) of *Escherichia Coli* K12 (Dila et al, 1990). It is a multi-subunit complex of two proteins, McrB and McrC, which specifically cleaves DNA containing methylated cytosine (Raleigh & Wilson, 1986; Stewart et al, 2000). In this complex, McrB subunit recognizes and binds to specific target (Pieper & Pingoud, 2002) sites while McrC subunit carries out DNA cleavage in GTP hydrolysis-dependent manner (Panne et al, 1999; Raleigh, 1992; Stewart et al, 2000; Stewart & Raleigh, 1998; Sutherland et al, 1992). The McrBC system shows many peculiar features, which set it apart from other known RM systems. The Type I and Type III NTP dependent restriction systems show stimulation of nucleotide hydrolysis in presence of DNA while the McrB subunit of this complex shows no stimulation in GTPase activity in presence of DNA (Pieper et al, 1997). Instead, the GTPase activity is stimulated by 30 fold in presence of the endonuclease subunit McrC (Panne et al, 1999; Pieper et al, 1997). Also, this complex is the only known restriction enzyme, which utilizes GTP instead of ATP for DNA translocation and harbors an ATPase domain for GTP hydrolysis. We carried out structural, biochemical and biophysical studies on McrBC to gain mechanistic insights into the functioning of this restriction enzyme.

In this chapter, I will discuss the history and classification of restriction enzymes followed by a detailed review of the McrBC system describing its biochemical and biophysical properties.

## **1.2 The Restriction-Modification (RM) Systems**

The phenomenon of restriction modifications was first discovered when Luria and Human in 1952, and later, Bertani and Weigle in 1953 observed that some mutants of *E. coli* strain B/4 were able to induce modification in T even phages such that the modified phages were unable to subsequently infect the B strain and its mutants (Bertani & Weigle,



1953; Luria & Human, 1952). This defect in the ability to infect was later explained as the consequence of lack of glucosylation of phage DNA that replicated in these strains, as the *E. coli* B strain and its mutants were deficient in the enzyme uridine diphosphoglucose pyrophosphorylase (UDPG-PPase) (Hattman & Fukasawa, 1963). These modifications on phage DNA were found to be host specific and non-hereditary as pathogenicity was regained upon replicating the phage in a non-modifying host. When the non-glucosylated nucleic acid of T-even phages was injected in *E. coli* K12 or B4, it was found to be cleaved into small fragments (Arber & Dussoix, 1962; Fukasawa, 1964; Hattman, 1964).

Based on their work, Arber and colleagues provided the molecular explanation for host specificity in bacteria (Arber, 1965b; Arber & Dussoix, 1962; Dussoix & Arber, 1962) showing that, in a bacterium, there are enzymes that cut foreign DNA at specific sites. They also showed that host DNA is protected from such degradation by the action of methyltransferase that modifies the host DNA. Hamilton Smith purified one of the restriction enzymes and showed its sequence-specific DNA cleavage activity, thus corroborating Arber's studies (Kelly & Smith, 1970; Smith & Wilcox, 1970). Following these works, Daniel Nathans demonstrated how these restriction enzymes could be used as molecular scissors and can be used to generate genetic maps (Danna & Nathans, 1971; Danna & Nathans, 1972; Lebowitz et al, 1974). The wondrous nature of discovery and the subsequent application of bacterial restriction systems was highlighted when, Werner Arber, Hamilton Smith and Daniel Nathans were jointly awarded the Nobel Prize in Physiology or Medicine for 1978 for the discovery of "restriction enzymes and their application to problems of molecular genetics".

### **1.3 Classification of RM Systems**

#### **1.3.1 Type I RM enzymes:**

In 1968 Meselson and Yuan successfully purified the first Type I restriction enzyme EcoKI (Meselson & Yuan, 1968). They showed that purified complex cleaved non-methylated DNA in presence of ATP, Mg<sup>2+</sup> and S-adenosylmethionine (AdoMet) and methylated the same DNA in absence of ATP thus protecting the DNA from endonucleolytic cleavage.

This corroborated Arber's theory about the restriction-modification systems (Arber, 1965a).

The Type I restriction enzymes are large, multi-subunit complexes comprising three Types of subunits- HsdS, HsdM, HsdR (Loenen, 2003; Murray, 2000). The HsdS (S) subunit harbors two target recognition domains (TRDs) for DNA specificity (Calisto et al, 2005; Kennaway et al, 2009), HsdM (M) subunit contains methyltransferase domain while the HsdR (R) subunit contains the ATPase and endonuclease domain (Davies et al, 1999). The methyltransferase (Mtase) core of these enzymes is a complex of one HsdS and two HsdM subunits ( $M_2S$ ) (Taylor et al, 1992) while active endonuclease complex is a pentamer of one HsdS, two HsdM and two HsdR subunits ( $R_2M_2S_1$ ) (Loenen et al, 2014; Murray, 2000). These enzymes bind to a bipartite recognition sequence through two TRDs of the HsdS subunit. In presence of hemimethylated DNA and AdoMet, two HsdM subunits (methyltransferase), present on either side of the HsdS subunit methylate the DNA. In case of a non-methylated DNA, the HsdR (restriction endonuclease) subunit is activated to cleave the DNA. For DNA cleavage, once activated, the HsdR subunit starts DNA translocation at a speed of 1000 bp/s, utilizing about 1 ATP/bp (Garcia & Molineux, 1999; Kennaway et al, 2012; McClelland et al, 2005; Seidel et al, 2008; Seidel et al, 2004). When two such enzymes translocating in opposite direction collide, dsDNA break occurs midway between the two sites (Davies et al, 1999; Studier & Bandyopadhyay, 1988) (Figure 1.1 B).

### **1.3.2 Type III RM enzymes:**

Type III restriction enzymes are hetero-oligomeric enzymes containing two Types of subunits- Mod subunit and the Res Subunit. The Mod subunit has a target recognition domain (TRD) and a methyltransferase domain (Hadi et al, 1983). Initially, it was understood that the Type III RM complex consists of two Mod and two Res ( $Mod_2Res_2$ ) subunits (Ahmad et al, 1995; Brockes, 1973; Janscak et al, 2001). But, later stoichiometry studies (Wyszomirski et al, 2012) and a recent crystal structure of a Type III RM enzyme EcoP15I showed that the functional complex consists of two Mod and one Res subunit ( $Mod_2Res_1$ ) (Gupta et al, 2015). These enzymes recognize a 5-6 bp long asymmetric recognition site. DNA containing two inversely oriented sites (head-to-head or tail-to-tail)

are cleaved (Meisel et al, 1992) in presence of ATP and  $Mg^{2+}$  (Yuan et al, 1980). Contrary to the Type I RM enzymes, where either methylation or restriction of DNA occurs, these enzymes are capable of carrying out the two activities simultaneously. This leads to a conundrum as the methylation and restriction activities are antagonistic to each other (Rao et al, 2014). The Type III enzymes use much less (almost 1%) ATP than the Type I enzymes but still cleave DNA containing sites thousands of base pairs away (van Aelst et al, 2010). In Type III enzymes DNA cleavage occurs 25-27 bp downstream of one of the recognition site (Bachi et al, 1979; Hadi et al, 1979; Piekarowicz & Brzezinski, 1980).

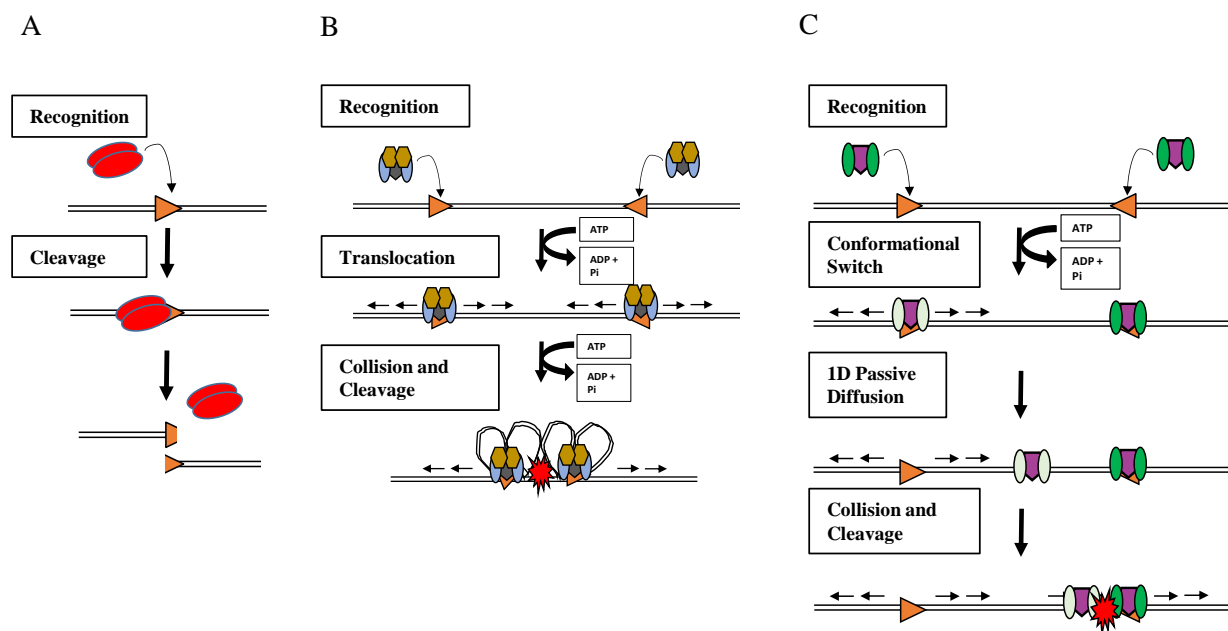


Figure 1.1. **The NTP independent and NTP dependent RM enzymes.** A schematic diagram to show the mechanism and mode of DNA cleavage by different classes of RM enzymes. A) Schematic for the action of Type II NTP independent restriction enzymes where a dimer of endonuclease binds specific sequence on DNA and cleaves in the middle of target site. B) Schematic of the model for action of Type I RM enzymes proposed by Studier and Bandyopadhyay, 1988. According to this model, the pentameric ( $R_2M_2S$ ) Type I RM complex binds to specific sequences and starts DNA translocation leading to loop formation. When two such enzymes collide (usually at the middle of two target sites), DNA cleavage takes place. C) Schematic for Type III RM complex where, when the trimeric complex ( $Mod_2Res_1$ ) binds to specific recognition site ATP dependent conformational changes occur in the complex. This facilitates 1 dimensional passive diffusion of the complex on DNA. Upon encountering a target site bound enzyme, these two complexes cleave DNA close to the target site.

These enzyme complexes recognize and bind to specific sequences via their mod subunit

following which the complex undergoes a conformational change allowing it to passively diffuse on DNA one-dimensionally. When such a passively diffusing enzyme encounters a second complex bound to target site, DNA cleavage occurs (Figure 1.1 C) (Schwarz et al, 2013).

### **1.3.3 The MDR Systems:**

The classical restriction modification system recognizes and cleaves foreign DNA based on the absence of specific methylated sequences. The MDR enzymes, on the other hand, recognize and restrict methylated DNA. These enzymes lack an associated methyltransferase. It has been postulated that these endonucleases evolved to restrict phages which have methylated DNA and hence could easily escape the classical restriction systems (Loenen & Raleigh, 2014). The *E. coli* K12 chromosome encodes four RM enzymes. Amongst these, EcoKI is a Type I RM enzyme while the other three - McrA, McrBC and Mrr - are MDR enzymes (Raleigh, 1992; Raleigh & Wilson, 1986; Tran-Betcke et al, 1986).

The Mrr (methylated adenine recognition and restriction) enzyme was first discovered by Heitman and Model (Heitman & Model, 1987). It was observed that it cleaves DNA carrying methylated adenine (methyl group at the exocyclic N6 position; m6A). The exact specificity of this enzyme is yet not determined as restriction of DNA having methylated cytosine (5-methylcytosine; m5C) has also been reported (Kelleher & Raleigh, 1991; Waite-Rees et al, 1991). The other two MDR systems- McrA and McrBC restrict DNA containing methylated (m5C) or hydroxymethylated cytosine (hm5C). McrA is encoded by an accessory determinant, e14 element while McrBC is chromosomally encoded from the *E. coli* immigration control region. The McrA and McrBC restriction is completely independent of each other.

The use of MDR systems in epigenetic studies (Cohen-Karni et al, 2011; Schumacher et al, 2006a) and genetic engineering in non-modal bacterial strains has rekindled the interest in the MDR enzymes leading to the recent identification of many MDR enzymes (Loenen & Raleigh, 2014).

## 1.4 McrBC: A Type IV MDR enzyme

### 1.4.1 *Historical Background*

The first restriction phenomenon that was reported by Luria and Human as host controlled phage variation (Bertani & Weigle, 1953; Luria & Human, 1952) was caused by the activity of the McrBC restriction enzyme. In 1967, Revel and Luria identified two independent restriction systems that were causing the restriction of T-even phages. They called them RglA and RglB for the restriction against glucoseless phage (Revel, 1967). Almost 30 years later, these restriction systems were rediscovered due to continuous problems faced in cloning methyltransferase genes associated with the Type II restriction systems (Blumenthal et al, 1985; Kiss et al, 1985; Noyer-Weidner et al, 1986; Raleigh et al, 1989; Raleigh & Wilson, 1986; Tran-Betcke et al, 1986; Trautner et al, 1988). It was observed that these endonucleases specifically cleaved 5'-methyl-cytosine containing DNA. Since then the Mcr nomenclature for Modified Cytosine Restriction was adopted. The RglA restriction system corresponds to McrA and RglB corresponds to McrBC. Thus McrBC is perhaps the longest known restriction enzyme but due to the spur in the identification of other types of restriction endonucleases, this system was not explored completely upon its discovery. Currently, we understand the biochemical properties of this enzyme but details of its mechanism of action remain to be determined.

### 1.4.2 *The McrBC operon*

The *E. coli* K12 genome consists of a 14 kbp Immigration Control Region (ICR) near 99 min (Raleigh et al, 1989). This region is densely packed with genes encoding different restriction systems- Type I RM enzyme EcoKI flanked by the MDR enzymes McrBC and Mrr, displaying an example of a very thorough checkpoint for any foreign DNA invasion (Figure 1.2). McrB locus contains two genes *mcrB* and *mcrC* where the termination of McrB and initiation codon of McrC overlaps by one nucleotide (Ross et al, 1989b). *In vivo* experiments showed that the *mcrB* gene produced two polypeptides of molecular weight 54 kDa and 33 kDa. It was discovered that the smaller of the two peptides (33 kDa) was produced due to the presence of an in-frame translation initiation site in the *mcrB* gene and not because of post-translational cleavage of the larger peptide (54 kDa). This

smaller protein is referred to as McrBs and lacks the N-terminal 161 amino acid residues (Panne et al, 1998; Raleigh et al, 1989; Ross et al, 1987; Ross et al, 1989a; Ross et al, 1989b).

A very noticeable feature of the McrBC operon is its low G+C content (40 %) in comparison to the neighboring *E. coli* genome (50%). The *mcrB* gene has a G+C content of 40 % while *mcrC* gene has even lower G+C content of 37.5 % (Dila et al, 1990). The region is flanked by potential transcription termination dyads with high G+C content. This pattern of G+C content variation leads to the understanding that the McrBC system might have been acquired from a foreign organism in the recent evolutionary history (Dila et al, 1990; Ross et al, 1987; Ross et al, 1989b).

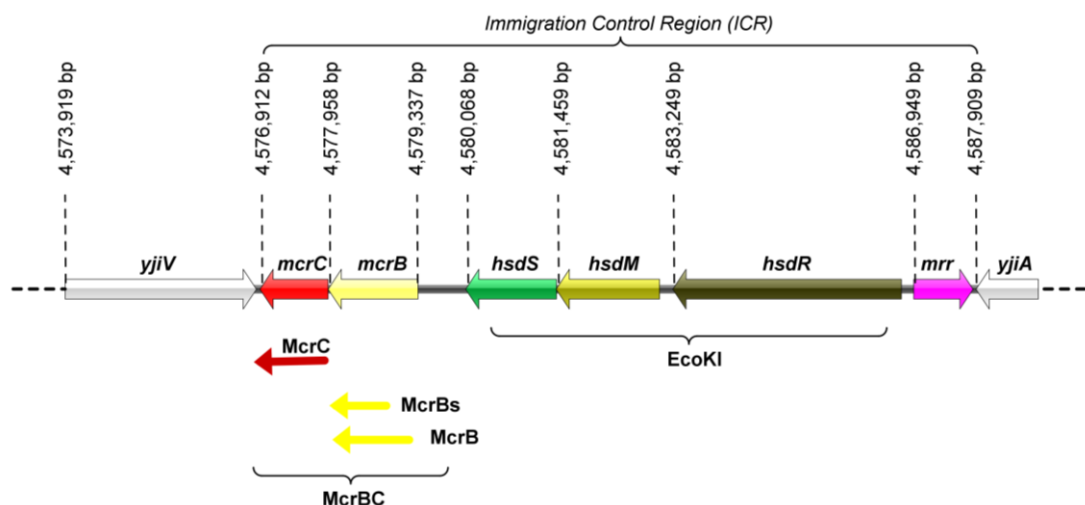


Figure 1.2. **The *mcrbc* locus in *E. coli* genome.** A schematic showing the section of *E. coli* chromosome (4573919 bp – 4587909 bp). The length and distances are scaled to the true length. Direction of the arrows indicate the reading direction oriented in the *E. coli* gene map. The genetic map information was obtained from GenBank entry NC\_00913. The figure was prepared using Illustrator for biological sciences (IBS) (Liu et al, 2015).

### 1.4.3 Constituents of McrBC complex

The McrBC complex consists of two subunits- McrB and McrC. The McrB subunit contains an N-terminal DNA binding and C-terminal GTP hydrolyzing domain (Pieper et al, 1999b; Raleigh, 1992; Sukackaite et al, 2012; Sutherland et al, 1992) while the McrC subunit

contains the nuclease domain (Pieper & Pingoud, 2002) (Figure 1.3). As mentioned in the previous section, the McrBC operon transcribes three proteins- McrB, McrBs and McrC. Maxicell analysis of translated proteins showed that both the full-length McrB and the truncated McrBs were produced in equimolar amounts. From the maxicell experiment, the relative amounts of McrB, McrBs and McrC were determined to be 3:3:1 respectively, (Dila et al, 1990).

#### **1.4.3.1 McrB**

The McrB subunit is a 54 kDa, 465 amino acid long polypeptide consisting of two functional domains - an N-terminal DNA binding domain (amino acid residues 1-161) (Gast et al, 1997; Kruger et al, 1995; Sukackaite et al, 2012) and a C-terminal GTPase domain (Pieper et al, 1999a; Pieper et al, 1999b) (Figure1.3)

##### ***N-terminal DNA binding domain of McrB:***

McrBC system specifically cleaves DNA containing two or more methylated or hydroxyl-methylated cytosine preceded by a purine ( $R^mC$ ) sites (Gast et al, 1997; Kruger et al, 1995; Stewart et al, 2000; Sutherland et al, 1992). The DNA binding and recognition domain of McrBC system lies in the N-terminal 161 amino acid residues of McrB subunit (Gast et al, 1997). The mechanism of recognition by the DNA binding domain was shown through its crystal structure bound to DNA (Sukackaite et al, 2012). The structure showed that the domain binds and recognizes DNA by flipping out methylated cytosine of  $R^mC$  site into a binding pocket sized only to accommodate a pyrimidine. The specificity for cytosine over thymine is added by direct base read out through donor-acceptor atoms on the Watson-Crick edge of the flipped base that is unique to cytosine. Finally the methylation status of cytosine is sensed via van der Waals interaction between the Y64 and L68 of the binding pocket and the methyl group of the flipped base.

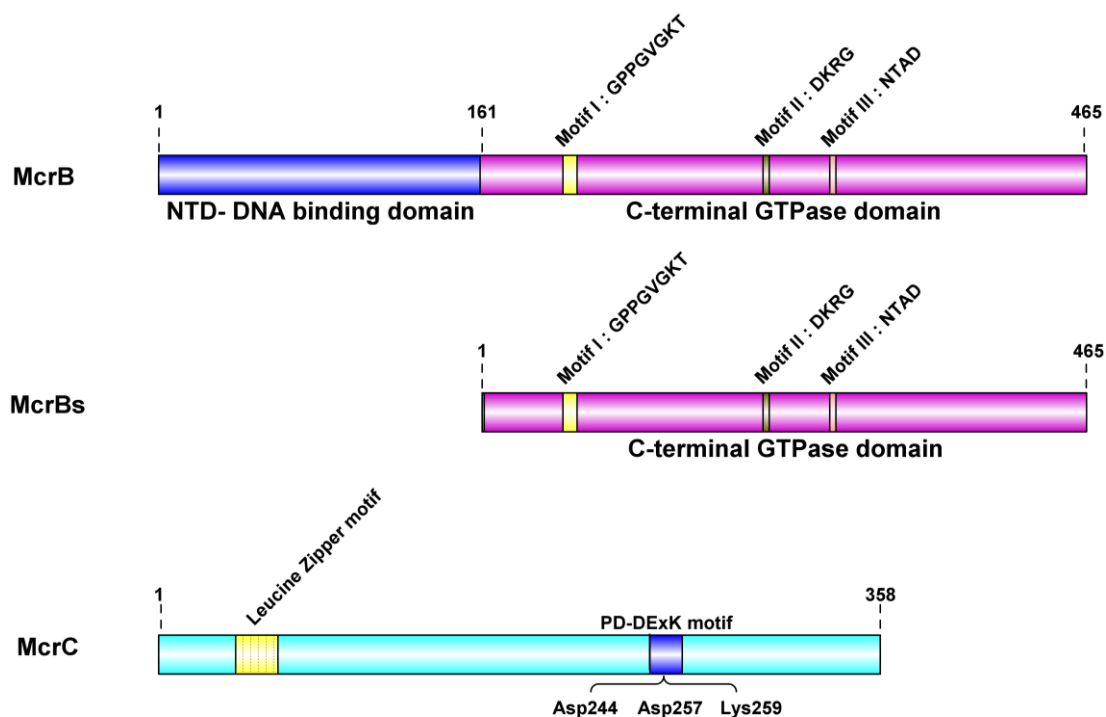


Figure 1.3. **The domain diagram of McrBC subunits.** A schematic showing different domains in the McrB, McrBs and McrC subunits of the McrBC complex. The figure was prepared using Illustrator for biological sciences (IBS) (Liu et al, 2015).

The base flipping mechanism for recognition and DNA binding has been observed for eukaryotic methyl CpG binding proteins containing the SRA (SET and Ring finger associated) domain (Hashimoto et al, 2010; Rajakumara et al, 2011; Sharif & Koseki, 2011). Even though the mode of recognition appears very similar for both SRA domain proteins McrB DNA binding domain, the fold of the latter is distinct with no eukaryotic homolog known.

#### ***C-terminal GTPase domain of McrB:***

C-terminus of McrB subunit harbors the GTPase domain. McrBC is the only known restriction system that uses GTP for its activity. But apart from GTP, though less efficiently, McrBC has been shown to be capable of using XTP or ITP but not ATP for its activity (Pieper et al, 1999a; Sutherland et al, 1992).

Since McrBC is a GTPase, initial amino acid sequence characterization was carried out to identify the core GTP binding and hydrolysis motifs in the protein. The GTP



binding proteins like EF-Tu or Ras-p21, harbor three well conserved characteristic motifs- i) Motif I - the Walker A or P- loop (Phosphate binding loop) motif with consensus amino acid sequence GxxxxGK(S/T) (Walker et al, 1982); ii) Motif II –this motif, also called Switch II (DxxG), is shown to be involved in interaction with magnesium ion and undergoes conformational changes during GTP hydrolysis cycle; iii) Motif III- this is a guanine nucleotide binding motif with consensus sequence NKxD (Bourne et al, 1991; Dever et al, 1987; Kjeldgaard et al, 1996).

Through sequence alignment and mutational analysis Motif I of McrB was identified as GPPGVGKT (201-208), Motif II as DKRG (300-303) and Motif III as NTAD (333-336) (Figure 1.3) (Pieper et al, 1997; Pieper et al, 1999a). As was expected, mutations in Motif I and Motif II affected the GTP hydrolysis-dependent DNA cleavage activity of the enzyme. But a mutation in Motif III carried out as an attempt to change substrate specificity from GTP to XTP (wild-type can use XTP but inefficiently) lead to complete loss of protein activity. In subsequent studies, the GTPase domain of McrB was identified and classified as a AAA+ domain (Neuwald et al, 1999), having additional features and characteristic motifs. These will be discussed in detail later in this chapter. Until now, McrB sequence has not been explored based on the features associated with the AAA+ domain.

The McrBC complex has been shown to hydrolyze GTP in a multiple-turnover reaction. McrB being the GTP binding subunit has a low intrinsic GTPase activity, which shows a Michaelis-Menten dependence on GTP giving a  $K_{cat}$  of  $0.4 \text{ min}^{-1}$  and  $K_m$  of  $1 \mu\text{M}$  (Pieper et al, 1997). The binding of GTP or GDP to McrB improves the thermal stability of the protein, a property similar to G-protein like EF-Tu (Pingoud et al, 1977; Wittinghofer & Leberman, 1976). The G-proteins bind to GDP more strongly than to GTP due to which they require a GEF (Guanine nucleotide Exchange Factor) to bind to GTP. In case of McrB, the affinity for GTP was found to be 50 fold higher than GDP and therefore it has been proposed that it does not require a GEF for its activity (Pieper et al, 1997).

The GTPase activity of McrB is stimulated 30 fold in presence of McrC, while DNA has been shown to produce a negligible stimulatory effect on the GTPase activity of both McrB and McrBC (Panne et al, 1999; Pieper et al, 1997). It is known that GTP hydrolysis

is necessary for DNA cleavage. However, it is not clear if GTP hydrolysis powers active DNA translocation as in the case of Type I RM enzymes, or facilitates bidirectional diffusion as in the case of Type III RM enzymes, or switches the enzyme between active and inactive states, like in the case of G-proteins.

#### **1.4.3.2 McrBs**

The smaller protein produced from the *mcrb* gene- McrBs lacks N-terminal 161 amino acids and consists only the C-terminal 304 amino acids constituting the GTPase domain. This 33 kDa protein has been shown to play a modulatory role in the activity of McrBC system (Panne et al, 1998). Although this protein has the GTPase domain similar to the full-length protein (Figure 1.2), it is not able to cleave DNA on its own or in presence of McrC (mostly because of the lack of N-terminal DNA binding domain). But both overproduction and underproduction of this protein lead to inhibition of DNA cleavage by full-length McrBC complex (Beary et al, 1997). In their study, Panne et al, 1998, showed that the optimal ratio of 3-5 McrB for 1 McrC was required for efficient DNA cleavage. A DNA cleavage reaction with an equimolar ratio of McrB and McrC was more efficient when McrBs was added. Also, McrBs shows GTPase and oligomeric properties with McrC similar to full length McrB. Since McrBs is produced in equimolar amounts *in vivo* (Dila et al, 1990; Ross et al, 1987; Ross et al, 1989a) a regulatory role of sequestering excess McrC has been proposed for this protein (Panne et al, 1998) such that optimum ratio of McrB and McrC can be attained.

#### **1.4.3.3 McrC**

McrC is the 39 kDa second subunit of McrBC complex consisting of 358 amino acid residues. Due to the basic nature of this protein (Ross et al, 1989b), it was initially thought to be the DNA binding subunit (Dila et al, 1990). But later investigations showed that this subunit was deficient in DNA binding (Kruger et al, 1995) or nucleotide binding but harbored the nucleolytic cleavage center (Pieper & Pingoud, 2002). DNA binding studies showed that along with stimulating the GTPase activity, McrC, in presence of GTP, increases the DNA affinity of McrB and promotes the formation of high molecular mass complexes (Kruger et al, 1995; Stewart et al, 2000).

Sequence analysis of McrC indicated a leucine zipper motif (Lx6Lx6Lx6L) spanning amino acid position 39 to 60 (Figure 1.2) (Dila et al, 1990). The leucine zipper motif has been shown to be involved in protein dimerization or protein-protein interactions (Landschulz et al, 1988). Such function could be attributed to McrC via this motif as it stimulates the McrB GTPase activity. Another motif “PD-X(10-30)....-DEXK” was identified (Pieper & Pingoud, 2002) which is a characteristic feature of the (PD-D/E)XK superfamily of Mg<sup>2+</sup> dependent endonucleases. Initially, this motif was observed in Type II restriction enzymes (Aggarwal, 1995; Pingoud & Jeltsch, 1997) but later found conserved amongst many DNA metabolizing enzymes like the endonuclease domain of Type I and Type III restriction systems (Venclovas et al, 1994) and DNA repair enzymes like RecB endonuclease (Davies et al, 1999; Singleton et al, 2004). This superfamily shares a common structural fold comprising of a 4 stranded mixed  $\beta$ -sheet flanked by  $\alpha$ -helices. The nucleolytic active site includes two to three acidic residues (Asp or Glu) and one Lys (Venclovas et al, 1994). Mutagenesis of the catalytic residues of PD-DEXK motif identified in McrC- Asp244, Asp257 and Lys259- lead to the abolishment of DNA cleavage activity even though the stimulation of McrB GTPase activity was not affected. Variation in the acidic residues of this motif (Asp to Asn and Glu to Gln) have been reported (Bujnicki & Rychlewski, 2001; Skirgaila et al, 1998), due to which identifying this motif solely based on sequence comparison is difficult and requires three-dimensional structural information for the correct assignment. Thus these mutagenesis results require to be corroborated with further structural and mechanistic studies for a confirmed PD-DExK motif assignment.

#### **1.4.4 The structural assembly of McrBC complex**

As mentioned earlier, McrBC is an enzyme complex of two subunits- McrB and McrC. During DNA binding studies, it was observed that McrB formed higher molecular mass species when GTP was present (Stewart et al, 2000). Through gel filtration studies and electron microscopy images, it has been proposed that both McrB and McrBs form a seven-membered close ring in presence of GTP. The McrBC complex has been proposed to be a tetradecamer consisting of two McrB heptameric rings and two McrC (Panne et

al, 2001). This heptameric state of McrB is similar to that of the oligomeric state observed for the bacterial enhancer binding protein (bEBP) NtrC1, both of which (McrB and NtrC1) are classified under the same Clade of AAA+ enzymes. But recent studies on bEBPs also reported hexameric assemblies of the bEBPs thus bringing forth the debate of functional relevance of a heptameric vs hexameric assembly.

In the study reported by Panne et al, 2001, although the ring like assembly is clear in the electron microscopy images, the resolution of the images is not good enough to reliably estimate the exact number of subunits. Thus the stoichiometry was obtained through mass calculations using scanning transmission electron microscopy (STEM). The obtained masses of the complex were less precise as the error was more than the mass of an entire subunit. For example, the mass of McrBs complex in presence of GTP was calculated as 239 ( $\pm 57$ ) kDa. Similarly, the mass estimated for McrB (361 ( $\pm 71$ ) kDa) and McrBC (759 ( $\pm 100$ ) kDa) was less accurate. Thus the oligomer size and stoichiometry need to be revisited with a more precise technique. Furthermore, if the heptameric assembly is indeed the true state of this protein, such non-canonical AAA+ oligomeric form (hexamers being the most predominant) need to be studied to understand their functional relevance.

#### **1.4.5 DNA Cleavage by McrBC complex**

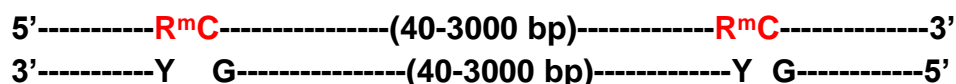
McrBC binds and cleaves DNA in a GTP hydrolysis-dependent manner. It is capable of recognizing 5-methylcytosine, 5-hydroxymethylcytosine and N4-methylcytosine. The only essential prerequisite is that the modified cytosine must be preceded by a purine base i.e. R<sup>m</sup>C (R= purine base and <sup>m</sup>C is modified cytosine) (Noyer-Weidner et al, 1986; Raleigh et al, 1989; Raleigh & Wilson, 1986; Sutherland et al, 1992).

For efficient cleavage of a linear DNA, a minimum of two 5'-R<sup>m</sup>C-3' sites optimally separated at least by 40 bp or more is required. Two sites with distance less than 40 bp are not cleaved (Stewart & Raleigh, 1998; Sutherland et al, 1992). The reason for this can be understood based on the footprint of McrBC complex. DNA foot-printing assay for McrB showed that it protects 19 bases around the recognition sequence. In accordance with these results, DNA cleavage always occurs at about 21-26 bases downstream of the recognition sequence.

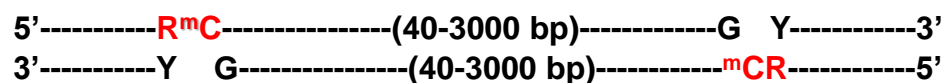
DNA cleavage is not affected by the presence of hemimethylated (only one strand methylated) or fully methylated (both strands methylated) sites (Sutherland et al, 1992) (Figure 1.4 ). Also, the presence of the two hemimethylated sites in *cis* or *trans* (Figure 1.4) does not affect the DNA cleavage. But, the local sequence surrounding the recognition site seems to affect both DNA binding and cleavage activity of McrBC (Gast et al, 1997; Sutherland et al, 1992).

McrBC has also been shown to cleave circular single site substrates and linear single site substrates having a *lac* repressor bound at a distance (Panne et al, 1999). These observations evoked the concept of translocational block inducing the cleavage competent form of McrBC. A collision of two enzymes translocating on DNA in opposite direction can also provide similar translocational block culminating in nuclease activation and DNA cleavage.

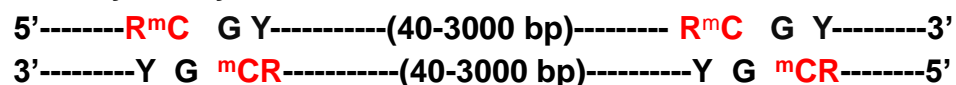
**A two site hemi-methylated substrate with sites in 'cis'**



**A two site hemi-methylated substrate with sites in 'trans'**



**A two site fully methylated substrate**



**A two site fully methylated substrate not amenable for cleavage**

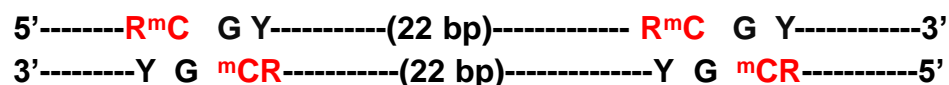
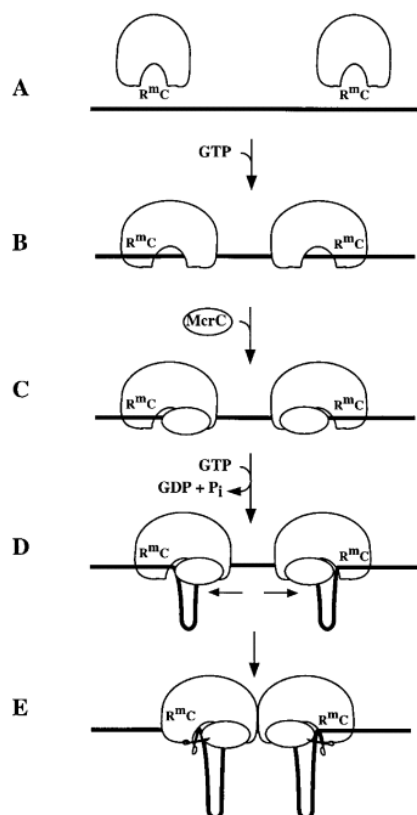


Figure 1.4. **McrBC DNA cleavage sites.** A schematic to show different types of sites amenable for McrBC cleavage. Amongst these the two site substrate with only 22 bp distance between two sites is not a substrate for McrBC even though it is fully methylated. Such a substrate is also considered a single site substrate.

### ***Model proposed for the mechanism of DNA cleavage by McrBC***

The proposed mechanism of DNA cleavage by McrBC is very similar to that proposed for Type I enzymes (Studier & Bandyopadhyay, 1988): 1) McrBC complex loads onto two distant recognition sites ( $R^mC$ ); 2) the two complexes then translocate towards each other. This translocation is powered by GTP hydrolysis. 3) When the two translocating complexes collide, McrC (nuclease) cleaves DNA downstream of one of the recognition sequence (Figure 1.5). However, unlike Type I RM enzymes, McrBC cleaves DNA close to either of the two target sites, a feature similar to Type III RM enzymes. McrBC is unique amongst restriction enzymes not only because it cleaves modified DNA, but also because it is the only known restriction enzyme, which uses GTP as its energy source and harbors a AAA+ domain to do so. Another peculiarity of this enzyme is that the GTP hydrolysis domain of this enzyme is stimulated by another protein (McrC) instead of the substrate DNA. This is in contrast with other NTP dependent restriction systems which show a spur in nucleotide hydrolysis in presence of specific DNA. Thus it's imperative to discuss the GTPase domain of McrB in detail before going to the specific aims of this study.



**Figure 1.5. Schematic of Model proposed for DNA cleavage by McrBC.** Image directly adapted from Panne et al, 1999. A) McrB binds to GTP which increases its affinity for DNA leading to B) binding to DNA at specific recognition sites. This followed by C) binding of McrC to GTP bound McrB which D) stimulates GTP hydrolysis causing DNA translocation. E) When two translocating enzymes collide, DNA cleavage occur close to either of recognition sites (indicated by scissors).

## 1.5 The AAA+ Superfamily

Utilizing chemical energy from nucleoside triphosphate (NTP) hydrolysis, especially that of ATP and GTP, for various cellular activities is a key feature of NTP dependent motor proteins. These molecular machines are ubiquitous and indispensable in cellular functioning. Through comparative genomic analysis, a large number of motor proteins have been identified and characterized into a monophyletic group called the P-loop NTPases (Leipe et al, 2003; Leipe et al, 2002) which is further classified into 7 major lineages (Figure 1.6) (Leipe et al, 2003).

One of the superfamilies classified under P-loop NTPases is the AAA+ family of proteins. The AAA+ family of proteins are ATPases Associated with diverse cellular Activities” (Confalonieri & Duguet, 1995; Erdmann et al, 1991; Kunau et al, 1993; Patel & Latterich, 1998), where the ‘+’ sign was added to the abbreviation signifying further extension of this family to include a large number of diverse proteins showing deviation from the classical AAA domain (Koonin, 1993; Koonin, 1997; Neuwald et al, 1999). Hence “AAA+” rightfully stands for “extended ATPases Associated with various cellular Activities” (Neuwald et al, 1999). Through phylogenetic sequence analysis, the AAA+ ATPases have been traced way back to LUCA (Last Universal Common Ancestor) (Iyer et al, 2004a) thus demonstrating their conservation throughout the evolutionary history. These proteins are indispensable for cell survival as they carry out diverse tasks ranging from DNA replication (helicases), recombination and repair (translocases) to RNA transcription (bEBP), protein translation and protein folding (Ogura & Wilkinson, 2001; Patel & Latterich, 1998). The AAA+ motor has been observed to function as a regulatory subunit of the proteasome, thus assisting in degradation of a variety of ubiquitinated substrates (Gottesman, 1996; Langer, 2000). As chaperones, they help in protein folding (Neuwald et al, 1999) and as motors, they help in cellular transport (Babst et al, 1998; Schmidt et al, 2012; Tomoyasu et al, 1993). Apart from this, AAA+ proteins play important roles in organelle biogenesis (Grimm et al, 2012), protein complex recycling (Tagaya et al, 1993), chromosomal segregation during cell division (Aussel et al, 2002; Bigot & Marians, 2010; Bigot et al, 2007; Iyer et al, 2004b), DNA replication (Erzberger & Berger,

2006; Lee & Bell, 2000), recombination and DNA repair (Putnam et al, 2001). In viruses, the AAA+ motors help in DNA packaging during viral replication (Schwartz et al, 2012).

The common feature that brings these functionally diverse proteins together is the employment of a similar motor to accomplish their respective tasks. Numerous studies carried out on different AAA+ proteins show that these proteins utilize different adapter domains to accomplish substrate variability (polypeptides vs oligonucleotides) (Dougan et al, 2002), but a detailed mechanistic understanding of how the motor is used under different functional scenarios is lacking.

### 1.5.1 AAA+ ATPase domain architecture:

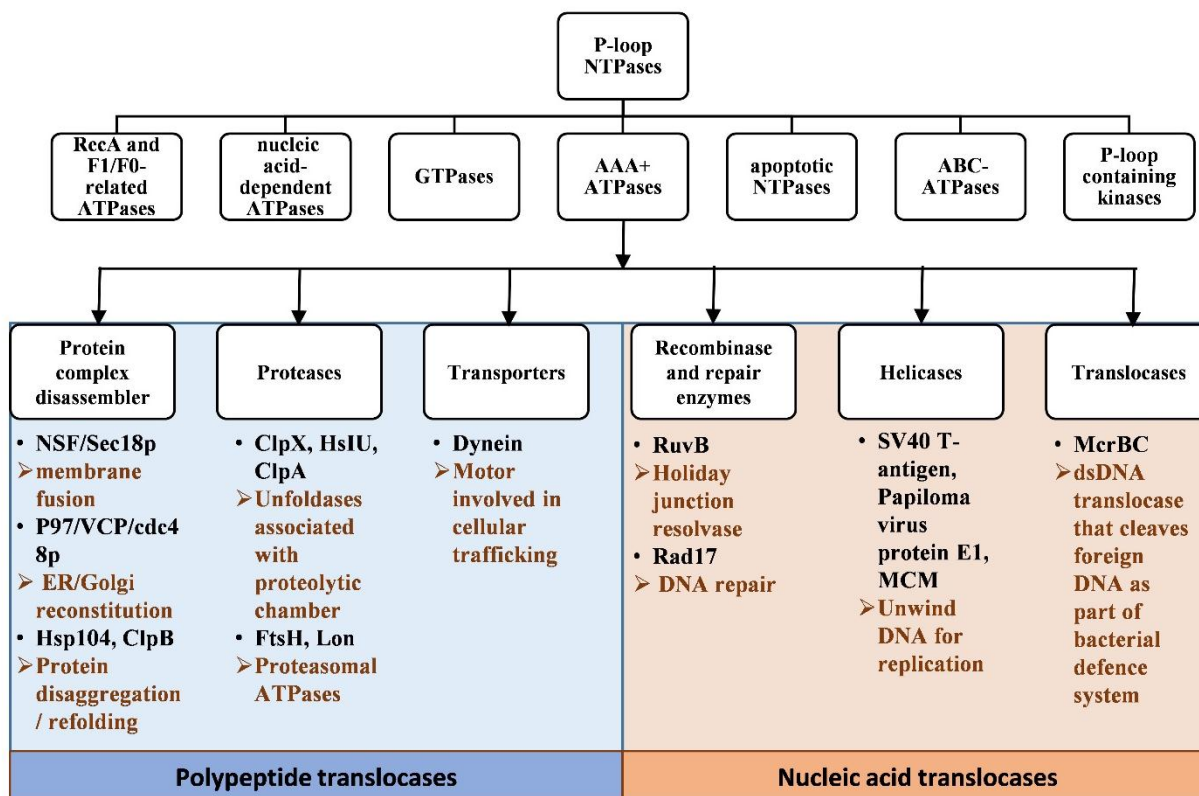


Figure 1.6. **The substrate diversity of AAA+ family.** Figure showing the classification of P-loop-NTPases into 7 major lineages and highlighting the diversity and omnipresence of AAA+ proteins inside cell by mentioning examples of motors functioning with either nucleic acids or polypeptides.

The highly conserved classical AAA ATPase domain consists of about 200-250 amino acids. This domain is further divided into two subdomains- a large N-terminal  $\alpha$ - $\beta$ - $\alpha$



subdomain and a small C-terminal helical bundle which looks like a lid over the  $\alpha$ - $\beta$ - $\alpha$  subdomain. The  $\alpha$ - $\beta$ - $\alpha$  subdomain consists of a  $\beta$ -sheet, with strands arranged in  $\beta 5$ - $\beta 1$ - $\beta 4$ - $\beta 3$ - $\beta 2$  order, sandwiched by  $\alpha$  helices on both sides (Figure 1.9 A) (Erzberger & Berger, 2006; Hanson & Whiteheart, 2005; Iyer et al, 2004a). The AAA+ domain harbors several characteristic motifs essential for ATP hydrolysis (Figure 1.7):

### **i) The Walker motifs:**

The two characteristic P-loop motifs were described for the first time by Walker in 1982 (Walker et al, 1982). The first motif- Walker A (GXXGXGKT/S) is observed at the top of  $\beta$ -strand 1, and the Walker B motif (hhhhDD/EXX) present at the apex of  $\beta$ -strand 3. From the structure of adenylate kinase, it was shown that glycine-rich Walker A motif is responsible for binding to the phosphate of the nucleotide. Hence this motif is also referred to as P-loop (Phosphate binding loop) (Schulz et al, 1974; Schulz & Schirmer, 1974). The second motif called Walker B consists of an acidic residue (an aspartate or a glutamate). This residue coordinates the  $Mg^{2+}$  ion which is necessary for NTP hydrolysis (Saraste et al, 1990). In AAA+ proteins, there is another conserved glutamate residue in Walker B motif, which primes a water molecule for nucleophilic attack on  $\gamma$ -phosphate (Mogk et al, 2003; Seybert & Wigley, 2004). Due to the presence of this AAA+ superfamily is also classified under ASCE (Additional Strand Conserved E) group of NTPases. This residue lies at the top of  $\beta$ -strand 2, adjacent to the Walker B motif. Mutational studies of this residue show that it is essential for nucleotide hydrolysis and not for nucleotide binding (Weibezahn et al, 2003; Wendler et al, 2012).

From the structural comparison of ATP bound and ADP bound structures of the AAA+ domain of bacterial transcription activator PspF (Rappas et al, 2006), Zhang and Wigley observed that although the two structures align well, there is sufficient deviation in orientation of the conserved glutamate residue. As compared to ADP bound state, in ATP bound state the glutamate residue is moved by about  $100^\circ$  and forming a hydrogen bond with an asparagine residue, which is present on an adjacent  $\beta$  strand. This engagement of glutamate residue prevents it from priming incoming water molecule. This implicates that the conserved glutamate residue acts as a switch (glutamate Switch) which converts the inactive enzyme from low energy state to high energy state (Zhang & Wigley, 2008).

It is possible that interaction with substrates releases this glutamate from asparagine and hence stimulates ATPase activity. Such switching via the glutamate residue has been observed in other AAA+ proteins as well (Chiraniya et al, 2013; Mogni et al, 2009).

## **ii) Sensor motifs:**

The Sensor motifs of AAA+ proteins were first identified and described in  $\delta'$  subunit structure by Guenther et al, 1997, describing two interactions of the protein with nucleotides in addition to Walker motif interactions. These Sensors appeared spatially analogous to the Switch II of Ras GTPases and lid segment of adenylate kinases. In AAA+ proteins the insertion of  $\beta$ -strand 4 between strand 1 and 3 (between Walker A and Walker B) positions a polar residue (asparagine, aspartic acid, serine or threonine) at its tip that acts in concert with Walker B residues to activate the incoming water molecule for ATP hydrolysis (Gai et al, 2004). This polar residue, called Sensor 1 (Guenther et al, 1997; Hattendorf & Lindquist, 2002; Karata et al, 1999; Steel et al, 2000), interacts with the  $\gamma$ -phosphate via hydrogen bond aiding in its hydrolysis (Liu et al, 2000).

The C-terminal  $\alpha$ -helical bundle of AAA+ protein shows considerable motion with respect to the N-terminal  $\alpha$ - $\beta$ - $\alpha$  subdomain. The C-terminal lid segment of adenylate kinases shows a conformational change in presence of nucleotide such that the active site is closed and sequestered from water (Muller & Schulz, 1992). This lid segment contains arginine residues which directly interact with phosphates of bound nucleotide and mediate the movement of lid segment over the catalytic pocket. With respect to the N-terminal  $\alpha$ - $\beta$ - $\alpha$  subdomain, AAA+  $\alpha$ -helical bundle is spatially analogous to the lid domain of adenylate kinases. A conserved arginine residue in this subdomain, at the tip of helix  $\alpha 7$ , has been observed to interact with bound nucleotide and undergo conformational changes based on the type of nucleotide bound (Zeymer et al, 2014). This arginine is part of the Sensor 2 motif with consensus sequence G/PX $\phi$ RX $\phi$ , where  $\phi$  is a hydrophobic residue (Ammelburg et al, 2006; Beyer, 1997; Erzberger & Berger, 2006; Neuwald et al, 1999). Like adenylate kinases, the Sensor 2 in AAA+ has been observed to mediate conformational changes leading to closing and opening of ATPase pocket at the interface of two protomers in a AAA+ ring (Gai et al, 2004; Guenther et al, 1997).

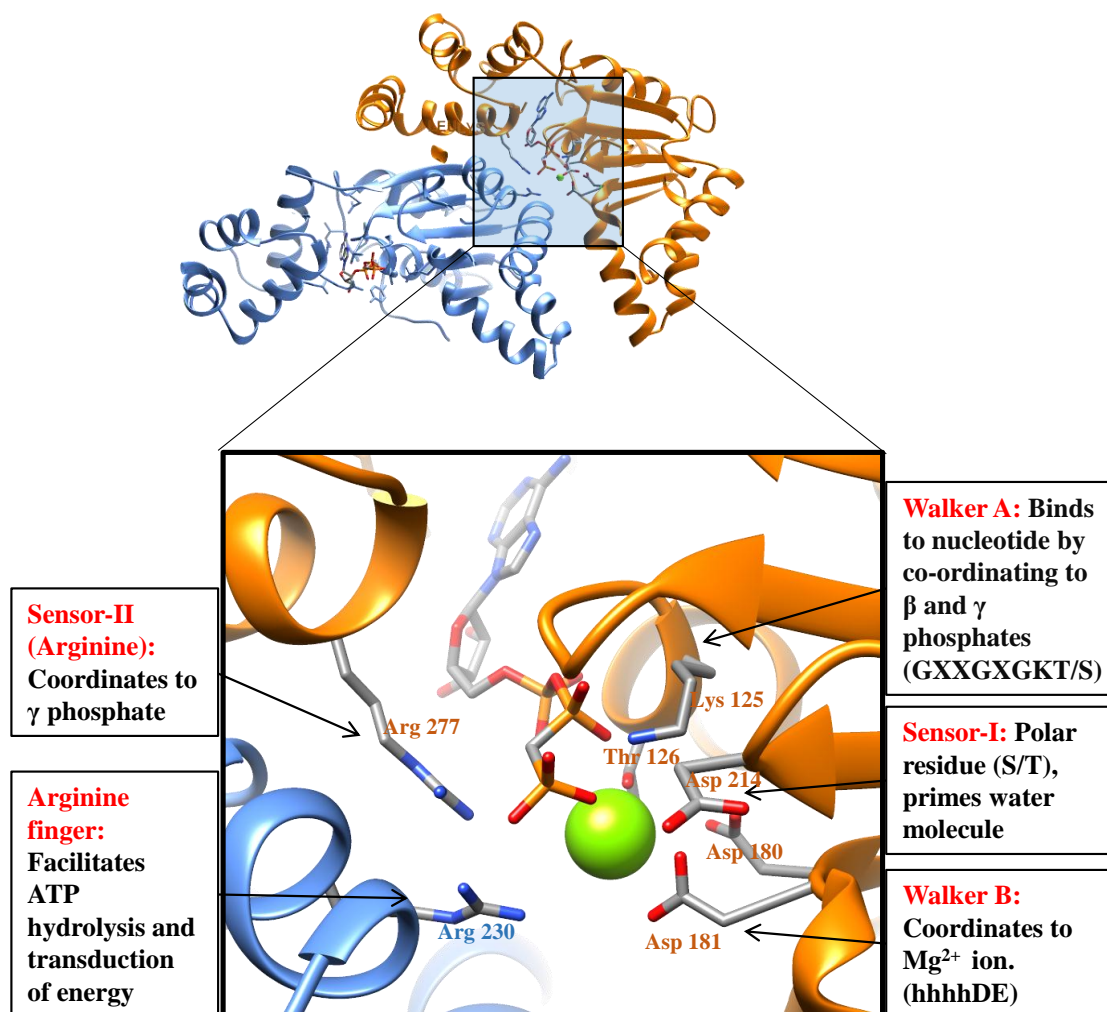


Figure 1.7. **ATPase site architecture of AAA+ proteins.** Nucleotide binds at the interface of two protomers in AAA+ proteins. This figure shows the bound nucleotide at the nucleotide binding site. The zoomed-in version of the binding site shows the conserved AAA+ motifs in the site. The figure was prepared from the DnaA structure with PDB ID: 2HCB. Molecular graphic images were made using the UCSF Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081) (Pettersen et al, 2004).

### iii) R-Finger motif:

**R-Finger motif:** Along with above mentioned features, AAA+ proteins also contain an arginine residue that plays a catalytic activation role similar to that of arginine finger observed in case of GTPase activating proteins (GAP) (Wittinghofer et al, 1997). Detailed mutational studies and structural position of this residue in AAA+ protein p97/ VCP have

led to the belief that this residue is not only essential for ATP hydrolysis but also for oligomer stability (Karata et al, 1999; Wang et al, 2005). The position of the arginine finger along with the Sensor 1 motif is a characteristic feature of only AAA+ proteins and does not exist in other P-loop ASCE families. This region lies c-terminal to the Walker B motif and spans part of  $\beta_4$ ,  $\alpha_4$  and the loop connecting  $\beta_4$  and  $\alpha_4$ . It is called Box VII, the Second Region of Homology (SRH) (Ammelburg et al, 2006; Erzberger & Berger, 2006; Neuwald et al, 1999). The Location of various conserved motifs of AAA+ proteins is depicted in Figure 1.7. Together the entire SRH plays an important role in nucleotide hydrolysis and transduction of hydrolysis energy to mechanical energy by undergoing and propagating conformational changes.

Classification of AAA+ proteins has been difficult due to their diverse functions and close structural relationship with other P- loop ATPases. Based on structural alignments and topology of AAA+ N-terminal  $\alpha$ - $\beta$ - $\alpha$  subdomain and C-Terminal helical bundle, the AAA+ proteins have been further classified into Clades and Superclades (Erzberger & Berger, 2006; Iyer et al, 2004a). The description of each Clade is as follows:

**Clade I:** The Clamp-loaders of replication process are classified under this category. These proteins have minimal AAA+ domain features (i.e. without any modification) - N-terminal  $\alpha$ - $\beta$ - $\alpha$  subdomain consists of a  $\beta$ -sheet, with strands arranged in  $\beta_5$ - $\beta_1$ - $\beta_4$ - $\beta_3$ - $\beta_2$  order, sandwiched by  $\alpha$  helices on both side and a C-terminal  $\alpha$ -helical bundle (lid). These proteins exist as pentameric oligomers. Examples of this class are Replication Factor C (RFC) and  $\gamma/\delta$  DNA polymerase III subunits (Figure 1.8 A).

**Clade II:** The Clade II consists of replication initiator proteins, which are involved, in origin recognition and helicase loading. The characteristic feature of this group is an insertion of  $\alpha$ -helix between  $\alpha_2$  and  $\beta_2$  in the  $\alpha$ - $\beta$ - $\alpha$  core. This insertion harbors an Initiator-Specific motif (ISM), which is necessary for binding of initiator to origin DNA and helicase loading. These proteins exist as open spiral hexameric assemblies. Examples of this category are DnaA/DnaC and Orc families (Figure 1.8 B).

**Clade III:** Also called the classic Clade, this family of proteins forms closed hexameric ring structures with a small  $\alpha$ -helical insertion between  $\alpha 2$  and  $\beta 2$  (Figure 1.8 C). This insert projects into the central pore of the ring and has been shown to be involved in substrate translocation through the pore. Structure based alignment show that this class does not have a Sensor 2 residue at the base of  $\alpha 7$ . Examples of this Clade are NSF family, ClpABC NTD (N-terminal domain) family and FtsH family.

**The Pre-Sensor 1  $\beta$ -hairpin (PS1 $\beta$ hp) Superclade:** The rest of the four AAA+ Clades (Clade 4-7) are grouped together in this Superclade due to a common insertion of a  $\beta$ -hairpin between  $\alpha 3$  and  $\alpha 4$ . The role of this insertion has been observed to be Clade specific and will be discussed under different Clade categories.

**Clade IV:** This Clade comprises of the superfamily 3 (SF3) helicases. These proteins lack a canonical C-terminal  $\alpha$ -helical lid subdomain. Instead, a helical subdomain is formed by element both from N-terminal and C-terminal of the  $\alpha$ - $\beta$ - $\alpha$  subdomain (Figure 1.8 D). The PS1 $\beta$ hp plays an important role in substrate (nucleic acid) translocation in this Clade. The Sensor 2 of these proteins is defined based on structural positioning and not by sequence alignment. Also, in these proteins, the Sensor 2 is a trans-acting residue in contrast to the cis-acting Sensor 2 in other AAA+ proteins. Examples of Clade IV are helicases like SV40 Large T-antigen helicase, Papillomavirus E1 helicase and adeno associated virus Rep40 protein.

**Clade V:** Amongst the PS1 $\beta$ hp proteins, this group has the minimal AAA+ domain like the Clade I, only differing from Clade I due to the presence of PS1 $\beta$ hp (Figure 1.8 E). This Clade comprises of four major families- HslU/ClpX, ClpABC-CTD (C-terminal domain), Lon and RuvB- and hence also referred to as the HCLR Clade. In this Clade the PS1 $\beta$ hp is not involved in substrate translocation. Instead, the loop connecting the  $\alpha 2$  and  $\beta 2$  (like Clade III) projects into the central pore and is involved in substrate translocation. This loop (also called Loop 1) has a characteristic aromatic-hydrophobic dipeptide motif. In case of dsDNA translocase RuvB, the PS1 $\beta$ hp insertion is involved in the protein-protein interaction (i.e. interaction with RuvA) rather than for interaction with DNA.

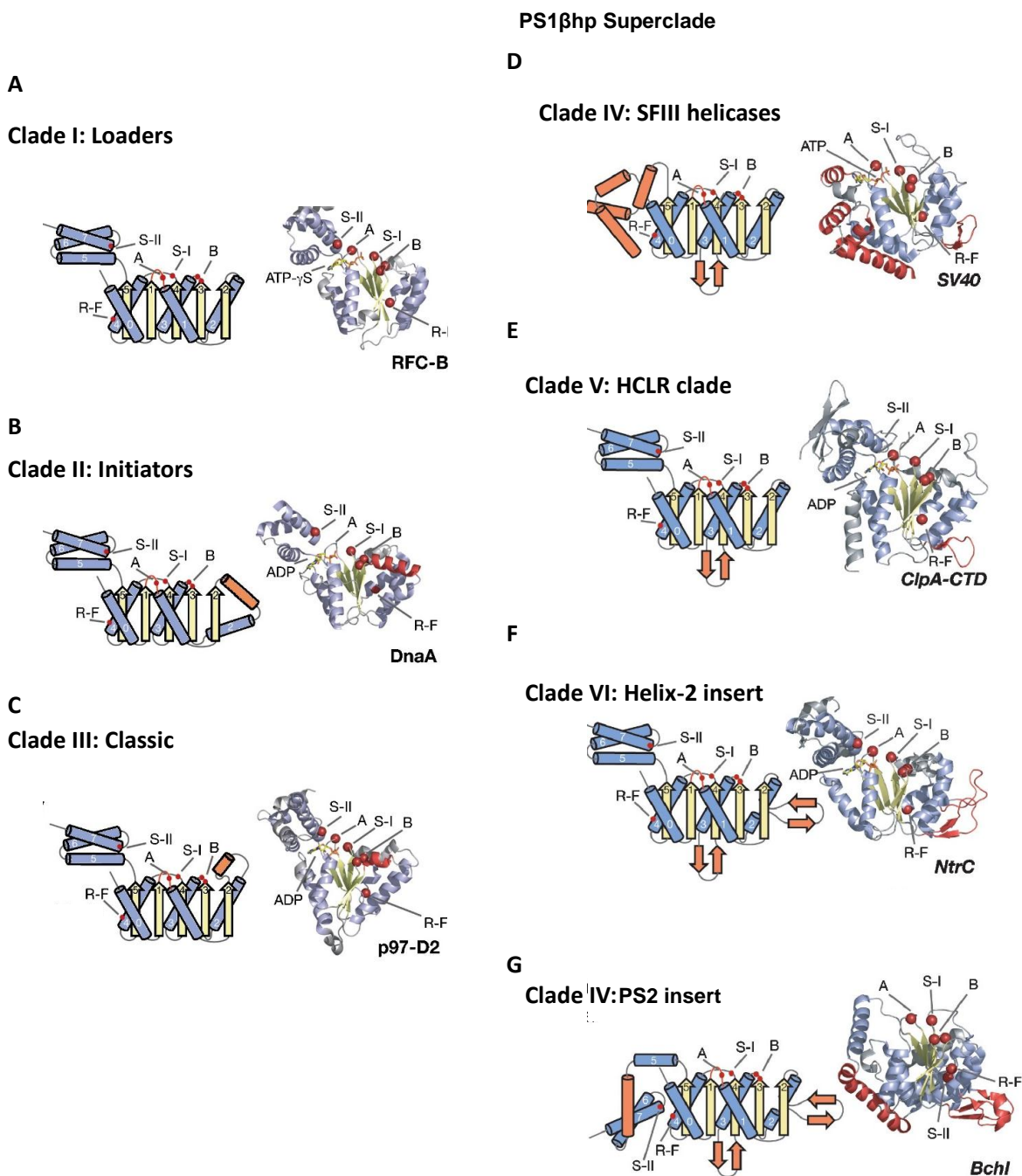


Figure 1.8. **Topology of different AAA+ Clades** (modified from Erzberger and Berger, 2006). Topological diagrams along with cartoon diagram of one representative proteins from different Clades of AAA+ superfamily are depicted to highlight the small insertions and deletions (colored in orange) in an otherwise conserved AAA+ core (colored in blue and yellow).

**Clade VI:** This Clade contains an additional  $\beta$  hairpin insertion in the  $\alpha 2$  helix and hence also referred to as the helix 2 insert (h2i) Clade (Figure 1.8 F). The major member of this Clade are the NtrC family and McrB family. Like PS1 $\beta$ hp in RuvB, the h2i has been shown to be involved in protein-protein interaction. Mutation in this loop breaks the interaction between bEBP and its target  $\sigma^{54}$  RNA polymerase.

**Clade VII:** Also called the Pre-Sensor 2 insert (PS2i) Clade, this group contains PS2 $\beta$ hp, h2i and an extra  $\alpha$ -helical insertion after  $\alpha 5$  (Figure 1.8 G). Due to the PS2 insertion, the C-terminal helical lid domain is positioned at the back of the  $\alpha$ - $\beta$ - $\alpha$  in comparison to other AAA+ lid containing Clades. This positions the Sensor 2 residue close to the neighboring domain thus making the Sensor 2 residue trans-acting. Examples of this family are MCM family and dynein Family.

### ***Characteristic features of AAA+ proteins***

Along with having a common structural fold, the proteins belonging to the AAA+ superfamily show three characteristic features:

#### **A. Oligomerization:**

AAA+ proteins have been shown to form predominantly ring like hexameric structure in their active state with the clamp loaders (being pentameric) and bEBPs (being heptameric) as exception. The oligomeric state of these proteins is essential for the ATPase activity as the nucleotide binding site is formed at the interface of two promoters. Though pentameric status of clamp loader proteins is well established, both heptameric and hexameric state of many AAA+ proteins have been observed (Batchelor et al, 2009; Chen et al, 2007; De Carlo et al, 2006; Lee et al, 2003; Miyata et al, 2000; Pape et al, 2003; Sallai & Tucker, 2005; Schumacher et al, 2006b; Yu et al, 2002). This lead to the debate of functional relevance of one oligomeric form over the other. For ClpB, it has been suggested that under heat shock, when the protein concentration is high, the protein might interchange into heptamers and hexamers during its ATPase cycle. This will cause partial opening and closing of the ring which might be a mechanism for the “prying apart”

aggregated proteins. In case of ring helicases like T7gp4 and MCM, it is proposed that these proteins exist as heptamers before interaction with DNA (Toth et al, 2003; Yu et al, 2002) and convert into a hexamer through loss of a subunit thus creating a gap through which the substrate can slip into the ring. (Crampton et al, 2006b; O'Shea & Berger, 2014).

### **B. The nucleotide binding pocket:**

In AAA+ proteins, the nucleotide binding and hydrolysis takes place at the interface of two monomers where functional residues are provided by both promoters. Figure 1.7 shows the architecture of nucleotide binding site and spatial positioning of different functional residues.

The nucleotide binding site exists at the interface of two monomers. An important feature of AAA+ proteins is that although they oligomerize in presence of nucleotide, not all of the six sites are similar in affinity for nucleotide. This leads to heterogeneity in both occupancies of binding sites as well as the kinetics of nucleotide hydrolysis at different sites within a single oligomer. Based on studies with different AAA+ proteins, several modes of such heterogeneous nucleotide hydrolysis processes have been described, which are discussed below.

### **Sequential Rotary model:**

This model draws an analogy from F1 ATPase wherein three ATPase sites are in different inter-convertible states (Abrahams et al, 1994). According to this model, at a given time, three consecutive catalytic sites are in either of three states- ATP bound state (T), ADP bound state (D) or empty state E. The three sites coordinate nucleotide hydrolysis such that in one cycle of ATP hydrolysis, the ATP bound site becomes ADP binding site i.e. D, ADP bound site releases the bound ADP and thus becomes empty i.e. E, and empty site binds to ATP and becomes the T site. Thus with every cycle, the trio or a block of these states move like a treadmill (Singleton et al, 2007). This cycle continues so that every catalytic site passes through these three states. This model was proposed for T7 DNA helicase called gp4 (Singleton et al, 2000), which is a hexameric ring helicase that binds to ssDNA through central hole of the ring (Hingorani et al, 1997), and Rho, a AAA+ ring



helicase (Thomsen & Berger, 2009) (Figure 1.9 A). In such a mechanism, if there is any mutation or hindrance to ATP hydrolysis at one site, it will stop the entire ATPase cycle.

Such a situation has been shown for T7 gp4, where, reduction in DNA dependent dTTP hydrolysis was observed when catalytically inactive monomers were incorporated into the hexamer (Crampton et al, 2006a). A sequential nucleotide hydrolysis model has also been put forth for the bEBP PspF. Since the protein showed maximal catalytic activity under heterogeneous nucleotide occupancy, a concerted model (discussed below) was ruled out. Also, the data indicated coordination between these states heterogeneous sites making the stochastic hydrolysis model less favorable (Joly et al, 2006). This model has also been proposed for the papillomavirus protein E1 helicase (Enemark & Joshua-Tor, 2006).

#### **Concerted model:**

Unlike F1- ATPase, T7 gp4 and Rho protein, the SV 40 Large tumor Antigen shows an all or none ATP binding (Gai et al, 2004). This binding is compatible with the concerted or simultaneous ATP hydrolysis, in which each site is ready to hydrolyze ATP equally and simultaneously thus going through the stages of all T, all D and all E (Figure 1.9 B).

#### **Stochastic or probabilistic model:**

Studies on the ClpX AAA+ protease led to the proposal of this model (Martin et al, 2005). It has been shown that ClpX, with the different geometric arrangement of covalently linked active and inactive monomers, hydrolyzed ATP and transported denatured polypeptide into proteolytic chamber. This showed that ATP hydrolysis instead of being sequential or concerted depends on the catalytic activity of any one subunit thus avoiding stalling of translocation even if any subunit is inactive (Figure 1.9 C).

Though different models for ATP hydrolysis have been proposed for several AAA+ proteins, some of the models have been refuted due to contradictory observations made in further studies. For example, though an all or none concerted model for SV40 LTag has been proposed based on crystal structures (Gai et al, 2004), computational studies

indicate that such a model might not support a directional motion of the motor on DNA (Yoshimoto et al, 2010). Also, the stochastic model for ClpX was favored because of no effect of inactive protomers in the hexamer. But later studies indicated that certain combinations of inactive protomers did influence the activity (Martin et al, 2005) suggesting a possible sequential model for ClpX. It is proposed that the sequential mode in ClpX is probably flexible enough to by-pass effects of subunit inactivation under some conditions. These findings indicate the sequential firing mechanism might be a prevalent mode of nucleotide hydrolysis in these motors. To develop a consensus of such kind, more studies on the mode of hydrolysis by diverse AAA+ proteins need to be carried out.

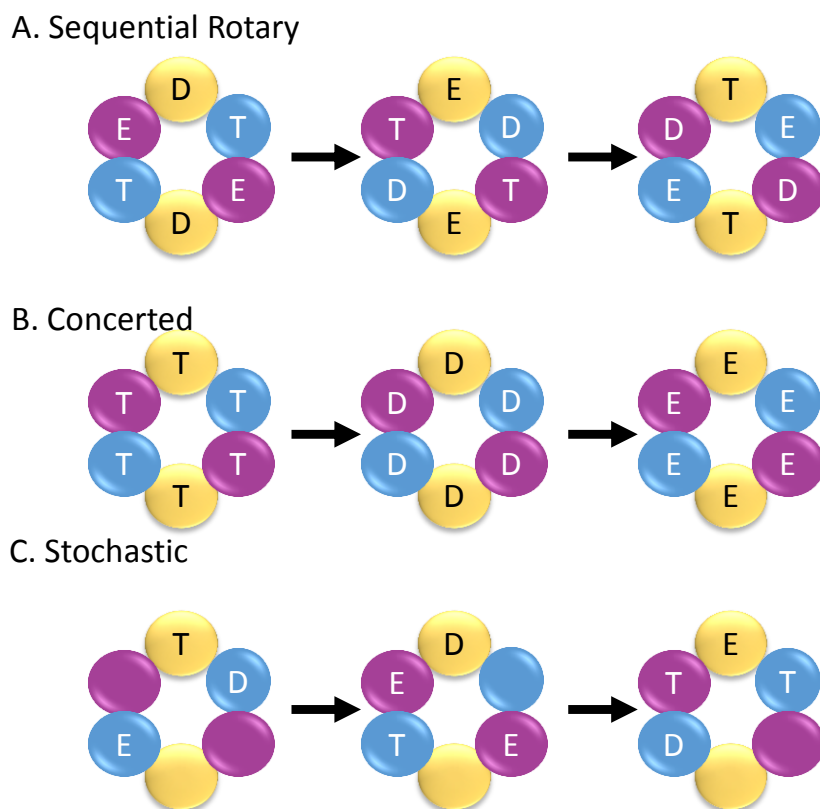


Figure 1.9. **Models for NTP hydrolysis by AAA+ motors.** A. Sequential rotary model of NTP hydrolysis in which events proceed sequentially through bound (T) hydrolysis (D) and release or empty (E) states like a rotary motor. B. Concerted model showing NTPase events occurring simultaneously at all sites. C. Stochastic model showing NTP hydrolysis occurring in a random manner without any influence from adjacent subunits.

### C. Pore loops for substrate translocation:

A prominent feature of AAA+ proteins is the presence of axial loops protruding into the central pore of the ring assembly (Figure 1.10). The AAA+ motors translocate upon different substrates (peptides and nucleic acids Figure 1.6). In this process, the substrate is passed through the central pore of the ring via these pore loops. Though there are several different insertions found in different AAA+ families, the loop insertion projecting axially into the central pore is invariantly present in all AAA+ proteins. The position of the pore loop insertion varies from Clade to Clade.

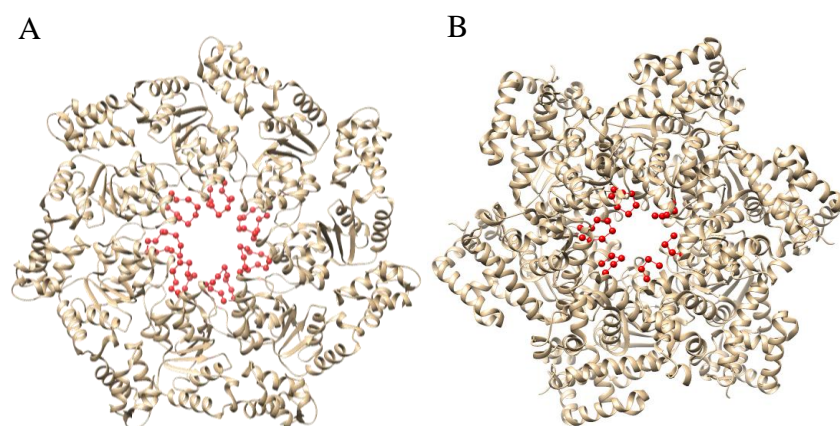


Figure 1.10. **The pore loop of AAA+ proteins.** Crystal structure of A) bEBP NtrC1 (PDB ID 3MOE) and B) helicase SV40 large T antigen showing the loops involved in substrate translocation projecting into the pore (red spheres). Molecular graphic images were produced using the UCSF Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081) (Pettersen et al, 2004).

Amongst proteases, the characteristic feature of pore loop is the presence of a dipeptide hydrophobic-aromatic motif. This loop has been shown to move as the ATP cycle progresses through the ring. Mutating this loop in proteases like ClpX (Siddiqui et al, 2004), HslU (Song et al, 2000), ClpA and ClpB abolished the unfoldase activity (Martin et al, 2008). In bEBPs, the pore loop, with consensus sequence GAFTGA, is essential for interaction and activation of the transcription factor  $\sigma^{54}$  (Bordes et al, 2003; Chen et al, 2007; De Carlo et al, 2006; Sysoeva et al, 2013). Helicases like SV40 LTag and Papillomavirus protein E1 have a  $\beta$ -hairpin insert with positively charged residues for interaction with DNA (Enemark & Joshua-Tor, 2006; Gai et al, 2004). Mutating these

hairpin residues affected the DNA binding activity while ATPase activity remained intact (Shen et al, 2005).

Though it appears that these loops might be involved in substrate binding, their primary function is to couple the nucleotide hydrolysis and substrate translocation. For substrate specificity and binding, the AAA+ proteins employ additional adapter domains. In the crystal structure of nucleotide bound and free states, the pore loops show an up and down motion coupled to the state of nucleotide bound to that subunit (Gai et al, 2004). This motion causes the substrate to move through the pore as nucleotide hydrolysis progresses.

There are many crystal structures now available that show a staircase-like arrangement of these pore loops in the hexamer suggesting a path to pass the substrate along the six protomers. This observation amongst different classes of AAA+ is suggestive of a common basic mechanism of substrate translocation, which might require slight modifications depending on the function of the protein.

## **1.6 Future perspective and Scope of the Thesis**

There are several examples of hexameric helicases that have been extensively characterized to date but very little is understood about dsDNA translocases. Although the crystal structure of RuvB has been solved (Miyata et al, 2000; Putnam et al, 2001), its dsDNA translocation mechanism is not yet known. I initiated studies on McrB, which can serve as another model for studying dsDNA translocation by AAA+ motors.

McrBC is a unique AAA+ protein, which hydrolyzes GTP instead of ATP. This peculiarity is unique amongst restriction enzymes as well and is suggestive of an evolution to use an ATPase fold for GTPase activity. Motion along the DNA is required by nucleotide-dependent restriction enzymes to communicate or collide with another such enzyme bound at far away target sequence to catalyze nucleolytic cleavage. McrBC has been proposed to translocate upon dsDNA using energy derived from GTP hydrolysis. However, unlike most other motors that act on DNA, the stimulation of the McrB GTPase has been shown to be dependent on complexation with the nuclease McrC rather than DNA. The role of McrC as a GEF, GAP or an effector can add more information to the regulation of McrBC as no such regulatory mechanism is known for this enzyme.

As a first step towards understanding the functioning of this complex, I have carried out biochemical, biophysical and structural studies on McrBC complex. These studies were aimed at understanding the molecular architecture of McrBC, the inter-subunit interactions in the complex and possible modes of GTP hydrolysis and assembly on substrate DNA. Thus, the specific aims of the thesis were:

## **1. Purification, assembly and biochemical characterization of McrBC complex:**

### **Chapter 2**

To carry out biochemical and structural studies with McrBC, I standardized the large-scale purification of the individual subunits. Following which, the McrBC complex was assembled. The purified proteins were tested for their oligomerization, DNA binding, GTP hydrolysis and DNA cleavage activity. Next, I determined the exact stoichiometry of the McrB and McrBC complexes using size exclusion coupled multiple angle light scattering (SEC-MALS). I carried out a preliminary investigation of the oligomeric structure of McrB using electron cryomicroscopy (cryo-EM).

## **2. Mutational analysis of the AAA+ McrB protein: Chapter 3**

McrBC has been characterized under the Clade VI of AAA+ family along with the NtrC1 family. A very detailed functional mutagenesis has been reported in earlier studies on McrBC (Pieper et al, 1997; Pieper et al, 1999a; Pieper et al, 1999b) but the protein was explored as a GTPase with focus only on three GTP binding motifs. I carried out functional mutagenesis of McrB to identify some of the motifs which were not identified through multiple sequence alignment (Iyer et al, 2004a). Also, some of the previously described mutations of Walker B motif were revisited in light of observations made about similar mutants in other families of AAA+ proteins. The mutants were characterized by their DNA cleavage, nucleotide-dependent oligomerization, GTP binding and hydrolysis activities as per relevance.

## **3. Crystallographic and cryo-EM studies of McrBC: Chapter 4**

In parallel to the biochemical and biophysical characterization of McrBC, I carried out crystallization and cryo-EM studies of McrB and McrBC complexed with non-hydrolysable

GTP analog- GDPNP. After screening a large number of crystals, I was able to achieve success in crystallization with the N-terminal DNA binding domain deletion mutant of McrB; McrB $\Delta$ N. Finally, diffraction data of McrB $\Delta$ N+GDPNP at 4.5 Å and McrB $\Delta$ NC+GDPNP diffraction data at 4.4 Å was collected. Simultaneously, through cryo-EM, I was able to reconstruct the 3-D map of McrB $\Delta$ NC complex at 7.4 Å resolution. The density map obtained from cryo-EM showed the architecture and arrangement of McrB $\Delta$ N and McrC in the complex along with their interactions.

#### **4. Pre-steady-state kinetic studies on McrB and McrBC: Chapter 5**

To gain further insights into assembly and GTPase activity of McrB and McrBC, I carried out pre-steady-state kinetic studies of oligomerization of McrB and McrBC complexes in presence of nucleotide. For this, I used intrinsic tryptophan (Trp) fluorescence of the protein as a signal. In conjunction GTPase activity of the protein was also measured under different mixing regimes to understand the structure-function relation. Furthermore, intrinsic Trp fluorescence, GTPase studies and DNA anisotropy changes were measured to understand protein DNA association kinetics under different mixing regimes. Following this, different pathways possible for a ring-like molecule (McrB/McrBC) to load onto DNA were explored and a more plausible model of assembly is proposed.

#### **References**

Abrahams JP, Leslie AG, Lutter R, Walker JE (1994) Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**: 621-628

Aggarwal AK (1995) Structure and function of restriction endonucleases. *Current opinion in structural biology* **5**: 11-19

Ahmad I, Krishnamurthy V, Rao DN (1995) DNA recognition by the EcoP15I and EcoPI modification methyltransferases. *Gene* **157**: 143-147

Ammelburg M, Frickey T, Lupas AN (2006) Classification of AAA+ proteins. *J Struct Biol* **156**: 2-11

Arber W (1965a) Host-controlled modification of bacteriophage. *Annu Rev Microbiol* **19**: 365-378

Arber W (1965b) Host Specificity of DNA Produced by Escherichia Coli V . The Role of Methionine in the Production of Host Specificity. *J Mol Biol* **11**: 247-256

Arber W, Dussoix D (1962) Host specificity of DNA produced by Escherichia coli. I. Host controlled modification of bacteriophage lambda. *J Mol Biol* **5**: 18-36

Aussel L, Barre FX, Aroyo M, Stasiak A, Stasiak AZ, Sherratt D (2002) FtsK Is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell* **108**: 195-205

Babst M, Wendland B, Estepa EJ, Emr SD (1998) The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. *EMBO J* **17**: 2982-2993

Bachi B, Reiser J, Pirrotta V (1979) Methylation and cleavage sequences of the EcoP1 restriction-modification enzyme. *Journal of molecular biology* **128**: 143-163

Batchelor JD, Sterling HJ, Hong E, Williams ER, Wemmer DE (2009) Receiver domains control the active-state stoichiometry of Aquifex aeolicus sigma54 activator NtrC4, as revealed by electrospray ionization mass spectrometry. *Journal of molecular biology* **393**: 634-643

Beary TP, Braymer HD, Achberger EC (1997) Evidence of participation of McrB(S) in McrBC restriction in Escherichia coli K-12. *Journal of bacteriology* **179**: 7768-7775

Bertani G, Weigle JJ (1953) Host controlled variation in bacterial viruses. *J Bacteriol* **65**: 113-121

Beyer A (1997) Sequence analysis of the AAA protein family. *Protein Sci* **6**: 2043-2058

Bickle TA, Kruger DH (1993) Biology of DNA restriction. *Microbiological reviews* **57**: 434-450

Bigot S, Mariani KJ (2010) DNA chirality-dependent stimulation of topoisomerase IV activity by the C-terminal AAA+ domain of FtsK. *Nucleic Acids Res* **38**: 3031-3040

Bigot S, Sivanathan V, Possoz C, Barre FX, Cornet F (2007) FtsK, a literate chromosome segregation machine. *Mol Microbiol* **64**: 1434-1441

Blumenthal RM, Gregory SA, Cooperider JS (1985) Cloning of a restriction-modification system from *Proteus vulgaris* and its use in analyzing a methylase-sensitive phenotype in *Escherichia coli*. *Journal of bacteriology* **164**: 501-509

Bordes P, Wigneshweraraj SR, Schumacher J, Zhang X, Chaney M, Buck M (2003) The ATP hydrolyzing transcription activator phage shock protein F of *Escherichia coli*: identifying a surface that binds sigma 54. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 2278-2283

Bourne HR, Sanders DA, McCormick F (1991) The GTPase superfamily: conserved structure and molecular mechanism. *Nature* **349**: 117-127

Bourniquel AA, Bickle TA (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie* **84**: 1047-1059

Boyer HW (1971) DNA restriction and modification mechanisms in bacteria. *Annu Rev Microbiol* **25**: 153-176

Brockes JP (1973) The deoxyribonucleic acid-modification enzyme of bacteriophage P1. Subunit structure. *The Biochemical journal* **133**: 629-633

Bujnicki JM, Rychlewski L (2001) Identification of a PD-(D/E)XK-like domain with a novel configuration of the endonuclease active site in the methyl-directed restriction enzyme Mrr and its homologs. *Gene* **267**: 183-191

Calisto BM, Pich OQ, Pinol J, Fita I, Querol E, Carpena X (2005) Crystal structure of a putative type I restriction-modification S subunit from *Mycoplasma genitalium*. *Journal of molecular biology* **351**: 749-762

Chen B, Doucleff M, Wemmer DE, De Carlo S, Huang HH, Nogales E, Hoover TR, Kondrashkina E, Guo L, Nixon BT (2007) ATP ground- and transition states of bacterial enhancer binding AAA+ ATPases support complex formation with their target protein, sigma54. *Structure* **15**: 429-440

Chiraniya A, Finkelstein J, O'Donnell M, Bloom LB (2013) A novel function for the conserved glutamate residue in the walker B motif of replication factor C. *Genes (Basel)* **4**: 134-151



Clokie MR, Millard AD, Letarov AV, Heaphy S (2011) Phages in nature. *Bacteriophage* **1**: 31-45

Cohen-Karni D, Xu D, Apone L, Fomenkov A, Sun Z, Davis PJ, Kinney SR, Yamada-Mabuchi M, Xu SY, Davis T, Pradhan S, Roberts RJ, Zheng Y (2011) The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 11040-11045

Confalonieri F, Duguet M (1995) A 200-amino acid ATPase module in search of a basic function. *Bioessays* **17**: 639-650

Crampton DJ, Mukherjee S, Richardson CC (2006a) DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. *Molecular cell* **21**: 165-174

Crampton DJ, Ohi M, Qimron U, Walz T, Richardson CC (2006b) Oligomeric states of bacteriophage T7 gene 4 primase/helicase. *Journal of molecular biology* **360**: 667-677

Danna K, Nathans D (1971) Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proc Natl Acad Sci U S A* **68**: 2913-2917

Danna KJ, Nathans D (1972) Bidirectional replication of Simian Virus 40 DNA. *Proc Natl Acad Sci U S A* **69**: 3097-3100

Davies GP, Martin I, Sturrock SS, Cronshaw A, Murray NE, Dryden DT (1999) On the structure and operation of type I DNA restriction enzymes. *Journal of molecular biology* **290**: 565-579

De Carlo S, Chen B, Hoover TR, Kondrashkina E, Nogales E, Nixon BT (2006) The structural basis for regulated assembly and function of the transcriptional activator NtrC. *Genes & development* **20**: 1485-1495

Dever TE, Glynias MJ, Merrick WC (1987) GTP-binding domain: three consensus sequence elements with distinct spacing. *Proceedings of the National Academy of Sciences of the United States of America* **84**: 1814-1818

Dila D, Sutherland E, Moran L, Slatko B, Raleigh EA (1990) Genetic and sequence organization of the mcrBC locus of *Escherichia coli* K-12. *Journal of bacteriology* **172**: 4888-4900

Dougan DA, Mogk A, Zeth K, Turgay K, Bukau B (2002) AAA+ proteins and substrate recognition, it all depends on their partner in crime. *FEBS letters* **529**: 6-10

Dussoix D, 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

Enemark EJ, Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**: 270-275

Erdmann R, Wiebel FF, Flessau A, Rytka J, Beyer A, Frohlich KU, Kunau WH (1991) PAS1, a yeast gene required for peroxisome biogenesis, encodes a member of a novel family of putative ATPases. *Cell* **64**: 499-510

Erzberger JP, Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annual review of biophysics and biomolecular structure* **35**: 93-114

Fukasawa T (1964) The Course of Infection with Abnormal Bacteriophage T4 Containing Non-Glucosylated DNA on Escherichia Coli Strains. *J Mol Biol* **9**: 525-536

Gai D, Zhao R, Li D, Finkielstein CV, Chen XS (2004) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**: 47-60

Garcia LR, Molineux IJ (1999) Translocation and specific cleavage of bacteriophage T7 DNA in vivo by EcoKI. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 12430-12435

Gast FU, Brinkmann T, Pieper U, Kruger T, Noyer-Weidner M, Pingoud A (1997) The recognition of methylated DNA by the GTP-dependent restriction endonuclease McrBC resides in the N-terminal domain of McrB. *Biological chemistry* **378**: 975-982

Gottesman S (1996) Proteases and their targets in Escherichia coli. *Annu Rev Genet* **30**: 465-506

Grimm I, Saffian D, Platta HW, Erdmann R (2012) The AAA-type ATPases Pex1p and Pex6p and their role in peroxisomal matrix protein import in Saccharomyces cerevisiae. *Biochim Biophys Acta* **1823**: 150-158

Guenther B, Onrust R, Sali A, O'Donnell M, Kuriyan J (1997) Crystal structure of the delta' subunit of the clamp-loader complex of E. coli DNA polymerase III. *Cell* **91**: 335-345

Gupta YK, Chan SH, Xu SY, Aggarwal AK (2015) Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nature communications* **6**: 7363

Hadi SM, Bachi B, Iida S, Bickle TA (1983) DNA restriction--modification enzymes of phage P1 and plasmid p15B. Subunit functions and structural homologies. *Journal of molecular biology* **165**: 19-34

Hadi SM, Bachi B, Shepherd JC, Yuan R, Ineichen K, Bickle TA (1979) DNA recognition and cleavage by the EcoP15 restriction endonuclease. *Journal of molecular biology* **134**: 655-666

Hanson PI, Whiteheart SW (2005) AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* **6**: 519-529

Hashimoto H, Vertino PM, Cheng X (2010) Molecular coupling of DNA methylation and histone methylation. *Epigenomics* **2**: 657-669

Hattendorf DA, Lindquist SL (2002) Cooperative kinetics of both Hsp104 ATPase domains and interdomain communication revealed by AAA sensor-1 mutants. *EMBO J* **21**: 12-21

Hattman S (1964) The Functioning of T-Even Phages with Unglycosylated DNA in Restricting Escherichia Coli Host Cells. *Virology* **24**: 333-348

Hattman S, Fukasawa T (1963) Host-Induced Modification of T-Even Phages Due to Defective Glycosylation of Their DNA. *Proc Natl Acad Sci U S A* **50**: 297-300

Heitman J, Model P (1987) Site-specific methylases induce the SOS DNA repair response in Escherichia coli. *Journal of bacteriology* **169**: 3243-3250

Hingorani MM, Washington MT, Moore KC, Patel SS (1997) The dTTPase mechanism of T7 DNA helicase resembles the binding change mechanism of the F1-ATPase. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 5012-5017

Iyer LM, Leipe DD, Koonin EV, Aravind L (2004a) Evolutionary history and higher order classification of AAA+ ATPases. *Journal of structural biology* **146**: 11-31

Iyer LM, Makarova KS, Koonin EV, Aravind L (2004b) Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic acids research* **32**: 5260-5279

Janscak P, Sandmeier U, Szczelkun MD, Bickle TA (2001) Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I. *Journal of molecular biology* **306**: 417-431

Janulaitis A, Petrusyte M, Maneliene Z, Klimasauskas S, Butkus V (1992) Purification and properties of the Eco57I restriction endonuclease and methylase--prototypes of a new class (type IV). *Nucleic Acids Res* **20**: 6043-6049

Joly N, Schumacher J, Buck M (2006) Heterogeneous nucleotide occupancy stimulates functionality of phage shock protein F, an AAA+ transcriptional activator. *The Journal of biological chemistry* **281**: 34997-35007

Karata K, Inagawa T, Wilkinson AJ, Tatsuta T, Ogura T (1999) Dissecting the role of a conserved motif (the second region of homology) in the AAA family of ATPases. Site-directed mutagenesis of the ATP-dependent protease FtsH. *The Journal of biological chemistry* **274**: 26225-26232

Kelleher JE, Raleigh EA (1991) A novel activity in Escherichia coli K-12 that directs restriction of DNA modified at CG dinucleotides. *Journal of bacteriology* **173**: 5220-5223

Kelly TJ, Jr., Smith HO (1970) A restriction enzyme from Hemophilus influenzae. II. *J Mol Biol* **51**: 393-409

Kennaway CK, Obarska-Kosinska A, White JH, Tuszynska I, Cooper LP, Bujnicki JM, Trinick J, Dryden DT (2009) The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic acids research* **37**: 762-770

Kennaway CK, Taylor JE, Song CF, Potrzebowski W, Nicholson W, White JH, Swiderska A, Obarska-Kosinska A, Callow P, Cooper LP, Roberts GA, Artero JB, Bujnicki JM, Trinick J, Kneale GG, Dryden DT (2012) Structure and operation of the DNA-translocating type I DNA restriction enzymes. *Genes & development* **26**: 92-104

Kiss A, Posfai G, Keller CC, Venetianer P, Roberts RJ (1985) Nucleotide sequence of the BsuRI restriction-modification system. *Nucleic acids research* **13**: 6403-6421

Kjeldgaard M, Nyborg J, Clark BF (1996) The GTP binding motif: variations on a theme. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **10**: 1347-1368

Koonin EV (1993) A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Res* **21**: 2541-2547

Koonin EV (1997) Evidence for a family of archaeal ATPases. *Science* **275**: 1489-1490

Kruger T, Wild C, Noyer-Weidner M (1995) McrB: a prokaryotic protein specifically recognizing DNA containing modified cytosine residues. *The EMBO journal* **14**: 2661-2669

Kunau WH, Beyer A, Franken T, Gotte K, Marzioch M, Saidowsky J, Skaletz-Rorowski A, Wiebel FF (1993) Two complementary approaches to study peroxisome biogenesis in *Saccharomyces cerevisiae*: forward and reversed genetics. *Biochimie* **75**: 209-224

Labrie SJ, Samson JE, Moineau S (2010) Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **8**: 317-327

Landschulz WH, Johnson PF, McKnight SL (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**: 1759-1764

Langer T (2000) AAA proteases: cellular machines for degrading membrane proteins. *Trends Biochem Sci* **25**: 247-251

Lebowitz P, Kelly TJ, Jr., Nathans D, Lee TN, Lewis AM, Jr. (1974) A colinear map relating the simian virus 40 (SV40) DNA segments of six adenovirus-SV40 hybrids to the DNA fragments produced by restriction endonuclease cleavage of SV40 DNA. *Proc Natl Acad Sci U S A* **71**: 441-445

Lee DG, Bell SP (2000) ATPase switches controlling DNA replication initiation. *Curr Opin Cell Biol* **12**: 280-285

Lee SY, De La Torre A, Yan D, Kustu S, Nixon BT, Wemmer DE (2003) Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. *Genes & development* **17**: 2552-2563

Leipe DD, Koonin EV, Aravind L (2003) Evolution and classification of P-loop kinases and related proteins. *Journal of molecular biology* **333**: 781-815

Leipe DD, Wolf YI, Koonin EV, Aravind L (2002) Classification and evolution of P-loop GTPases and related ATPases. *Journal of molecular biology* **317**: 41-72

Liu J, Smith CL, DeRyckere D, DeAngelis K, Martin GS, Berger JM (2000) Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. *Molecular cell* **6**: 637-648

Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, Lahrmann U, Zhao Q, Zheng Y, Zhao Y, Xue Y, Ren J (2015) IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* **31**: 3359-3361

Loenen WA (2003) Tracking EcoKI and DNA fifty years on: a golden story full of surprises. *Nucleic acids research* **31**: 7059-7069

Loenen WA, Dryden DT, Raleigh EA, Wilson GG (2014) Type I restriction enzymes and their relatives. *Nucleic acids research* **42**: 20-44

Loenen WA, Raleigh EA (2014) The other face of restriction: modification-dependent enzymes. *Nucleic acids research* **42**: 56-69

Luria SE, Human ML (1952) A nonhereditary, host-induced variation of bacterial viruses. *Journal of bacteriology* **64**: 557-569

Martin A, Baker TA, Sauer RT (2005) Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* **437**: 1115-1120

Martin A, Baker TA, Sauer RT (2008) Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Molecular cell* **29**: 441-450

McClelland SE, Dryden DT, Szczelkun MD (2005) Continuous assays for DNA translocation using fluorescent triplex dissociation: application to type I restriction endonucleases. *Journal of molecular biology* **348**: 895-915

Meisel A, Bickle TA, Kruger DH, Schroeder C (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* **355**: 467-469

Meselson M, Yuan R (1968) DNA restriction enzyme from *E. coli*. *Nature* **217**: 1110-1114

Miyata T, Yamada K, Iwasaki H, Shinagawa H, Morikawa K, Mayanagi K (2000) Two different oligomeric states of the RuvB branch migration motor protein as revealed by electron microscopy. *Journal of structural biology* **131**: 83-89

Mogk A, Schlieker C, Strub C, Rist W, Weibezahn J, Bukau B (2003) Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity. *J Biol Chem* **278**: 17615-17624

Mogni ME, Costa A, Ioannou C, Bell SD (2009) The glutamate switch is present in all seven clades of AAA+ protein. *Biochemistry* **48**: 8774-8775

Muller CW, Schulz GE (1992) Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 Å resolution. A model for a catalytic transition state. *Journal of molecular biology* **224**: 159-177

Murray NE (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiology and molecular biology reviews : MMBR* **64**: 412-434

Neuwald AF, Aravind L, Spouge JL, Koonin EV (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome research* **9**: 27-43

Noyer-Weidner M, Diaz R, Reiners L (1986) Cytosine-specific DNA modification interferes with plasmid establishment in *Escherichia coli* K12: involvement of rglB. *Molecular & general genetics : MGG* **205**: 469-475

O'Shea VL, Berger JM (2014) Loading strategies of ring-shaped nucleic acid translocases and helicases. *Current opinion in structural biology* **25**: 16-24

Ogura T, Wilkinson AJ (2001) AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells* **6**: 575-597

Orlowski J, Bujnicki JM (2008) Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res* **36**: 3552-3569

Panne D, Muller SA, Wirtz S, Engel A, Bickle TA (2001) The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *The EMBO journal* **20**: 3210-3217

Panne D, Raleigh EA, Bickle TA (1998) McrBs, a modulator peptide for McrBC activity. *The EMBO journal* **17**: 5477-5483

Panne D, Raleigh EA, Bickle TA (1999) The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. *Journal of molecular biology* **290**: 49-60

Pape T, Meka H, Chen S, Vicentini G, van Heel M, Onesti S (2003) Hexameric ring structure of the full-length archaeal MCM protein complex. *EMBO reports* **4**: 1079-1083

Patel S, Latterich M (1998) The AAA team: related ATPases with diverse functions. *Trends Cell Biol* **8**: 65-71

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**: 1605-1612

Piekarowicz A, Brzezinski R (1980) Cleavage and methylation of DNA by the restriction endonuclease HinfIII isolated from *Haemophilus influenzae* Rf. *Journal of molecular biology* **144**: 415-429

Pieper U, Brinkmann T, Kruger T, Noyer-Weidner M, Pingoud A (1997) Characterization of the interaction between the restriction endonuclease McrBC from *E. coli* and its cofactor GTP. *Journal of molecular biology* **272**: 190-199

Pieper U, Pingoud A (2002) A mutational analysis of the PD...D/EXK motif suggests that McrC harbors the catalytic center for DNA cleavage by the GTP-dependent restriction enzyme McrBC from *Escherichia coli*. *Biochemistry* **41**: 5236-5244

Pieper U, Schweitzer T, Groll DH, Gast FU, Pingoud A (1999a) The GTP-binding domain of McrB: more than just a variation on a common theme? *Journal of molecular biology* **292**: 547-556

Pieper U, Schweitzer T, Groll DH, Pingoud A (1999b) Defining the location and function of domains of McrB by deletion mutagenesis. *Biological chemistry* **380**: 1225-1230



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

Pingoud A, Urbanke C, Krauss G, Peters F, Maass G (1977) Ternary complex formation between elongation factor Tu, GTP and aminoacyl-tRNA: an equilibrium study. *European journal of biochemistry / FEBS* **78**: 403-409

Putnam CD, Clancy SB, Tsuruta H, Gonzalez S, Wetmur JG, Tainer JA (2001) Structure and mechanism of the RuvB Holliday junction branch migration motor. *J Mol Biol* **311**: 297-310

Rajakumara E, Law JA, Simanshu DK, Voigt P, Johnson LM, Reinberg D, Patel DJ, Jacobsen SE (2011) A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. *Genes & development* **25**: 137-152

Raleigh EA (1992) Organization and function of the mcrBC genes of Escherichia coli K-12. *Molecular microbiology* **6**: 1079-1086

Raleigh EA, Trimarchi R, Revel H (1989) Genetic and physical mapping of the mcrA (rgIA) and mcrB (rgIB) loci of Escherichia coli K-12. *Genetics* **122**: 279-296

Raleigh EA, Wilson G (1986) Escherichia coli K-12 restricts DNA containing 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America* **83**: 9070-9074

Rao DN, Dryden DT, Bheemanaik S (2014) Type III restriction-modification enzymes: a historical perspective. *Nucleic acids research* **42**: 45-55

Rappas M, Schumacher J, Niwa H, Buck M, Zhang X (2006) Structural basis of the nucleotide driven conformational changes in the AAA+ domain of transcription activator PspF. *J Mol Biol* **357**: 481-492

Revel HR (1967) Restriction of nonglucosylated T-even bacteriophage: properties of permissive mutants of Escherichia coli B and K12. *Virology* **31**: 688-701

Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev S, Dryden DT, Dybvig K, Firman K, Gromova ES, Gumport RI, Halford SE, Hattman S, Heitman J, Hornby DP, Janulaitis A, Jeltsch A, Josephsen J, Kiss A, Klaenhammer TR, Kobayashi I, Kong H, Kruger DH, Lacks S, Marinus MG, Miyahara M, Morgan RD, Murray NE, Nagaraja V, Piekarowicz A, Pingoud A, Raleigh E, Rao DN,

Reich N, Repin VE, Selker EU, Shaw PC, Stein DC, Stoddard BL, Szybalski W, Trautner TA, Van Etten JL, Vitor JM, Wilson GG, Xu SY (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* **31**: 1805-1812

Ross TK, Achberger EC, Braymer HD (1987) Characterization of the Escherichia coli modified cytosine restriction (mcrB) gene. *Gene* **61**: 277-289

Ross TK, Achberger EC, Braymer HD (1989a) Identification of a second polypeptide required for McrB restriction of 5-methylcytosine-containing DNA in Escherichia coli K12. *Molecular & general genetics : MGG* **216**: 402-407

Ross TK, Achberger EC, Braymer HD (1989b) Nucleotide sequence of the McrB region of Escherichia coli K-12 and evidence for two independent translational initiation sites at the mcrB locus. *Journal of bacteriology* **171**: 1974-1981

Sallai L, Tucker PA (2005) Crystal structure of the central and C-terminal domain of the sigma(54)-activator ZraR. *Journal of structural biology* **151**: 160-170

Saraste M, Sibbald PR, Wittinghofer A (1990) The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci* **15**: 430-434

Schmidt H, Gleave ES, Carter AP (2012) Insights into dynein motor domain function from a 3.3-A crystal structure. *Nat Struct Mol Biol* **19**: 492-497, S491

Schulz GE, Elzinga M, Marx F, Schirmer RH (1974) Three dimensional structure of adenyl kinase. *Nature* **250**: 120-123

Schulz GE, Schirmer RH (1974) Topological comparison of adenyl kinase with other proteins. *Nature* **250**: 142-144

Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, Yau P, Virtanen C, Winegarden N, Cheng J, Gingeras T, Petronis A (2006a) Microarray-based DNA methylation profiling: technology and applications. *Nucleic acids research* **34**: 528-542

Schumacher J, Joly N, Rappas M, Zhang X, Buck M (2006b) Structures and organisation of AAA+ enhancer binding proteins in transcriptional activation. *Journal of structural biology* **156**: 190-199

Schwartz C, Fang H, Huang L, Guo P (2012) Sequential action of ATPase, ATP, ADP, Pi and dsDNA in procapsid-free system to enlighten mechanism in viral dsDNA packaging. *Nucleic Acids Res* **40**: 2577-2586

Schwarz FW, Toth J, van Aelst K, Cui G, Clausing S, Szczelkun MD, Seidel R (2013) The helicase-like domains of type III restriction enzymes trigger long-range diffusion along DNA. *Science* **340**: 353-356

Seidel R, Bloom JG, Dekker C, Szczelkun MD (2008) Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I. *The EMBO journal* **27**: 1388-1398

Seidel R, van Noort J, van der Scheer C, Bloom JG, Dekker NH, Dutta CF, Blundell A, Robinson T, Firman K, Dekker C (2004) Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I. *Nature structural & molecular biology* **11**: 838-843

Seybert A, Wigley DB (2004) Distinct roles for ATP binding and hydrolysis at individual subunits of an archaeal clamp loader. *EMBO J* **23**: 1360-1371

Sharif J, Koseki H (2011) Recruitment of Dnmt1 roles of the SRA protein Np95 (Uhrf1) and other factors. *Progress in molecular biology and translational science* **101**: 289-310

Shen J, Gai D, Patrick A, Greenleaf WB, Chen XS (2005) The roles of the residues on the channel beta-hairpin and loop structures of simian virus 40 hexameric helicase. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 11248-11253

Siddiqui SM, Sauer RT, Baker TA (2004) Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes & development* **18**: 369-374

Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB (2004) Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* **432**: 187-193

Singleton MR, Dillingham MS, Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annual review of biochemistry* **76**: 23-50

Singleton MR, Sawaya MR, Ellenberger T, Wigley DB (2000) Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**: 589-600

Skirgaila R, Grazulis S, Bozic D, Huber R, Siksnys V (1998) Structure-based redesign of the catalytic/metal binding site of Cfr10I restriction endonuclease reveals importance of spatial rather than sequence conservation of active centre residues. *Journal of molecular biology* **279**: 473-481

Smith HO, Wilcox KW (1970) A restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties. *J Mol Biol* **51**: 379-391

Song HK, Hartmann C, Ramachandran R, Bochtler M, Behrendt R, Moroder L, Huber R (2000) Mutational studies on HslU and its docking mode with HslV. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 14103-14108

Steel GJ, Harley C, Boyd A, Morgan A (2000) A screen for dominant negative mutants of SEC18 reveals a role for the AAA protein consensus sequence in ATP hydrolysis. *Mol Biol Cell* **11**: 1345-1356

Stewart FJ, Panne D, Bickle TA, Raleigh EA (2000) Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *Journal of molecular biology* **298**: 611-622

Stewart FJ, Raleigh EA (1998) Dependence of McrBC cleavage on distance between recognition elements. *Biological chemistry* **379**: 611-616

Studier FW, Bandyopadhyay PK (1988) Model for how type I restriction enzymes select cleavage sites in DNA. *Proc Natl Acad Sci U S A* **85**: 4677-4681

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

Sutherland E, Coe L, Raleigh EA (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *Journal of molecular biology* **225**: 327-348

Sysoeva TA, Chowdhury S, Guo L, Nixon BT (2013) Nucleotide-induced asymmetry within ATPase activator ring drives sigma54-RNAP interaction and ATP hydrolysis. *Genes & development* **27**: 2500-2511

Tagaya M, Wilson DW, Brunner M, Arango N, Rothman JE (1993) Domain structure of an N-ethylmaleimide-sensitive fusion protein involved in vesicular transport. *J Biol Chem* **268**: 2662-2666

Taylor I, Patel J, Firman K, Kneale G (1992) Purification and biochemical characterisation of the EcoR124 type I modification methylase. *Nucleic acids research* **20**: 179-186

Thomsen ND, Berger JM (2009) Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* **139**: 523-534

Tomoyasu T, Yuki T, Morimura S, Mori H, Yamanaka K, Niki H, Hiraga S, Ogura T (1993) The Escherichia coli FtsH protein is a prokaryotic member of a protein family of putative ATPases involved in membrane functions, cell cycle control, and gene expression. *J Bacteriol* **175**: 1344-1351

Toth EA, Li Y, Sawaya MR, Cheng Y, Ellenberger T (2003) The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Molecular cell* **12**: 1113-1123

Tran-Betcke A, Behrens B, Noyer-Weidner M, Trautner TA (1986) DNA methyltransferase genes of Bacillus subtilis phages: comparison of their nucleotide sequences. *Gene* **42**: 89-96

Trautner TA, Balganesch T, Wilke K, Noyer-Weidner M, Rauhut E, Lauster R, Behrens B, Pawlek B (1988) Organization of target-recognizing domains in the multispecific DNA (cytosine-5)methyltransferases of Bacillus subtilis phages SPR and phi 3T. *Gene* **74**: 267

van Aelst K, Toth J, Ramanathan SP, Schwarz FW, Seidel R, Szczelkun MD (2010) Type III restriction enzymes cleave DNA by long-range interaction between sites in both head-to-head and tail-to-tail inverted repeat. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 9123-9128

Venclovas C, Timinskas A, Siksnyys V (1994) Five-stranded beta-sheet sandwiched with two alpha-helices: a structural link between restriction endonucleases EcoRI and EcoRV. *Proteins* **20**: 279-282

Waite-Rees PA, Keating CJ, Moran LS, Slatko BE, Hornstra LJ, Benner JS (1991) Characterization and expression of the Escherichia coli Mrr restriction system. *Journal of bacteriology* **173**: 5207-5219

Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO journal* **1**: 945-951

Wang Q, Song C, Irizarry L, Dai R, Zhang X, Li CC (2005) Multifunctional roles of the conserved Arg residues in the second region of homology of p97/valosin-containing protein. *The Journal of biological chemistry* **280**: 40515-40523

Weibezahn J, Schlieker C, Bukau B, Mogk A (2003) Characterization of a trap mutant of the AAA+ chaperone ClpB. *J Biol Chem* **278**: 32608-32617

Wendler P, Ciniawsky S, Kock M, Kube S (2012) Structure and function of the AAA+ nucleotide binding pocket. *Biochimica et biophysica acta* **1823**: 2-14

Wittinghofer A, Leberman R (1976) Elongation factor T from *Bacillus stearothermophilus* and *Escherichia coli*. Purification and some properties of EF-Tu and EF-Ts from *Bacillus stearothermophilus*. *European journal of biochemistry / FEBS* **62**: 373-382

Wittinghofer A, Scheffzek K, Ahmadian MR (1997) The interaction of Ras with GTPase-activating proteins. *FEBS letters* **410**: 63-67

Wyszomirski KH, Curth U, Alves J, Mackeldanz P, Moncke-Buchner E, Schutkowski M, Kruger DH, 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 research* **40**: 3610-3622

Yoshimoto K, Arora K, Brooks CL, 3rd (2010) Hexameric helicase deconstructed: interplay of conformational changes and substrate coupling. *Biophys J* **98**: 1449-1457

Yu X, VanLoock MS, Poplawski A, Kelman Z, Xiang T, Tye BK, Egelman EH (2002) The *Methanobacterium thermoautotrophicum* MCM protein can form heptameric rings. *EMBO reports* **3**: 792-797

Yuan R, Hamilton DL, Hadi SM, Bickle TA (1980) Role of ATP in the cleavage mechanism of the EcoP15 restriction endonuclease. *Journal of molecular biology* **144**: 501-519

Zeymer C, Barends TR, Werbeck ND, Schlichting I, Reinstein J (2014) Elements in nucleotide sensing and hydrolysis of the AAA+ disaggregation machine ClpB: a structure-based mechanistic dissection of a molecular motor. *Acta crystallographica Section D, Biological crystallography* **70**: 582-595

Zhang X, Wigley DB (2008) The 'glutamate switch' provides a link between ATPase activity and ligand binding in AAA+ proteins. *Nat Struct Mol Biol* **15**: 1223-1227

## Chapter 2

# Purification, assembly and biochemical characterization of McrBC complex

### 2.1 Introduction

McrBC is one of the three modification-dependent restriction (MDR) enzyme systems encoded by *Escherichia coli* K12 chromosome (Raleigh, 1992), the other two being McrA and Mrr (Loenen & Raleigh, 2014). The *mcrbc* locus in *E. coli* K12 chromosome has two overlapping genes *mcrb* and *mcrc*, which code for the 54 kDa McrB protein and the 40 kDa McrC protein. The N-terminal 161 residues of McrB form the DNA binding domain, while the rest constitute a GTP binding AAA+ domain (Gast et al, 1997; Panne et al, 2001; Pieper et al, 1999a; Pieper et al, 1999b; Sukackaite et al, 2012). The nucleolytic active site of this complex is located in McrC, which belongs to the PD-(D/E)xK family of nucleases (Pieper & Pingoud, 2002). The two proteins in presence of GTP form the McrBC complex (Panne et al, 1999). McrBC recognizes and cleaves DNA having 5-hydroxymethylcytosine, 5- methylcytosine or 4-methylcytosine preceded by a purine (RmC) (Raleigh, 1992; Sutherland et al, 1992), making it a useful tool for studying epigenetic modifications (Fouse et al, 2010). DNA cleavage requires at least two recognition sites, which can be 40 bp to 3000 bp apart (Stewart & Raleigh, 1998; Sutherland et al, 1992). Like Type III RM enzymes, the DNA cleavage happens close to one of the recognition sites and requires hydrolysis of GTP (Panne et al, 1999; Pieper & Pingoud, 2002).

McrBC is unique in employing a GTP hydrolyzing AAA+ motor to catalyze endonucleolytic cleavage. It has been proposed that translocation of double-stranded (ds) DNA by the AAA+ motor of a target bound McrBC culminates in DNA cleavage upon collision with another translocating McrBC (Panne et al, 2001; Panne et al, 1999). Translocation of substrate by AAA+ motors are essential for a number of biological processes such as proteolysis by proteasome- ClpX, Lon, FtsH etc (Erzberger & Berger, 2006); DNA unwinding activity of superfamily 3 (SF3) helicases and SF6 helicases (e.g. replicative helicase MCM) (Singleton et al, 2007); bacterial chromosome segregation by

FtsK and viral capsid packaging (Iyer et al, 2004). Thus, McrBC serves as a model system to understand the mechanism of double-strand DNA translocation and the coupled nucleolytic activity.

The McrB GTPase motor belongs to a Clade of AAA+ proteins characterized by a  $\beta$ -hairpin insertion before the Sensor 1 motif ( pre-Sensor  $\beta$ -hairpin insert) and an additional  $\beta$ -hairpin inserted in helix 2 of the canonical core (helix-2 insert) (Erzberger & Berger, 2006). The role of AAA+ proteins has been documented in diverse functions. In almost all cases, the functions of these proteins are dependent on oligomerization of the AAA+ domains, which is also true for McrBC (Panne et al, 2001; Pieper & Pingoud, 2002). In an earlier study, using size exclusion chromatography (SEC), negative stain electron microscopy (EM) and mass analysis by scanning transmission electron microscopy (STEM), it was proposed that McrB probably forms a heptameric ring-like structure in presence of GTP (Panne et al, 2001). Addition of McrC results in the formation of a tetradecameric McrBC oligomer made of two heptameric rings of McrB and two subunits of McrC. The oligomeric McrBC is the functional restriction enzyme (Sutherland et al, 1992). The oligomerization of McrB appears to be primarily mediated by the AAA+ domain. A variant form of McrB that lacks the N-terminal DNA-binding domain, McrBs, was also reported to form a heptameric ring similar to the full-length protein (Panne et al, 2001), while isolated N-terminal domain was found to be monomeric (Sukackaite et al, 2012).

AAA+ proteins predominantly form hexameric rings, and hence the heptameric assembly of McrB is unusual (Erzberger & Berger, 2006). Although, occurrence of heptameric rings in case of some AAA+ proteins have been reported (Chen et al, 2010; Dey et al, 2015; Hastings et al, 2003; Kim et al, 2000; Miyata et al, 2000; Yu et al, 2002; Zarate-Perez et al, 2012), their functional significance is under debate (Batchelor et al, 2009; De Carlo et al, 2006; Dey et al, 2015; Joly et al, 2012; Pape et al, 2003; Smith et al, 1997; Sysoeva et al, 2013).

As part of our efforts to understand the molecular mechanism of the restriction enzyme McrBC, I first sought to understand the assembly of the McrB and McrBC oligomers. The integrity of the oligomeric structure of an AAA+ motor is essential for its function. For example, structural studies of the hexameric SF3 helicase E1 from



papillomavirus in complex with a DNA substrate mimic revealed the essentiality of the oligomeric structure and cooperation between the six subunits to translocate single-stranded DNA (Enemark & Joshua-Tor, 2006).

In this chapter I will be discussing the results of my efforts to purify and assemble the McrBC complex, determining the accurate molecular masses of oligomeric McrB and McrBC, functional characterisation of the GTP hydrolysis, DNA binding and DNA cleavage activity of purified proteins. This chapter will also discuss results towards efforts to observe the solution state of assembled McrB and McrBC complex by using electron cryomicroscopy (Cryo-EM). Since an N-terminal truncated mutant McrB<sub>s</sub> already exists naturally, I also deleted the N-terminal DNA-binding domain (McrB $\Delta$ N). Biochemical characterisation this mutant will also be discussed in this chapter.

## 2.2 Materials and Methods

### 2.2.1 *Cloning of mcrB, mcrB $\Delta$ N, mcrC*

*mcrB* was PCR amplified by using McrBHisF and McrBHisR primers and *mcrC* was PCR amplified using McrCHisF and McrCHisR primer from genomic DNA of *E. coli* K12. The sequence of the primers is given in Table 2.1.

Table 2.1. List of primers used for cloning McrB, McrC and McrB $\Delta$ N genes

McrBHisF	CTTTAAGAAGGAGATATACATATGGAATCTATTCAACCCTGGATTG
McrBHisR	GATGATGGGATCCCGATGAGTCCCC
McrCHisF	CTTTAAGAAGG AGATATACATATGGAACAGCCCGTGATACC
McrCHisR	GATGATGGGATCCTTATTTG AGATATTC
McrB $\Delta$ NHisF	GTTTAACTTTAAGAAGGAGATATACATATGTCAAAAAGTGAATCATA CG

The amplified products were cloned into pHIS17 vector (Miroux & Walker, 1996) using a restriction-free cloning method (van den Ent & Lowe, 2006).

N-terminal deletion mutant *mcrB* $\Delta$ N, similar to the McrB<sub>S</sub>, was amplified from *mcrB* gene in pHIS17 vector by using McrB $\Delta$ NHisF and McrBHisR primers (Table 2.1). The amplified product was also cloned into pHIS17 vector using restriction-free cloning method (van den Ent & Lowe, 2006). The resulting *mcrB-his6*, *mcrB* $\Delta$ N-*his6* and *mcrC* genes were fully sequenced.

### **2.2.2 Restriction Free Cloning:**

The method of restriction-free cloning, as described by van den Ent and Lowe (van den Ent & Lowe, 2006), is a modification of the site-directed mutagenesis method in which an entire gene is inserted instead of a single site mutation. Primers were designed such that the flanking region of the gene was complementary to the site on the plasmid where the gene is to be inserted. The cloning strategy involved two steps. In the first step, gene product was PCR amplified from the genomic DNA using primers described in section 2.2.1. The gene product was then used as primer and the vector as a template to carry out a PCR reaction. 9  $\mu$ l of the final product was subjected to digestion by 1  $\mu$ l of modification-dependent DpnI restriction enzyme (New England BioLabs<sup>®</sup> Inc (NEB)) at 37°C for one hour to degrade the methylated template DNA. The digested product was electroporated into *E.coli* NEB<sup>®</sup> Turbo cells and plated on LB agar plates containing 100  $\mu$ g/ml ampicillin. Simultaneous negative control- PCR reaction with only template DNA followed by digestion with DpnI was also electroporated in *E.coli* NEB<sup>®</sup> Turbo cells.

### **2.2.3 Purification**

#### **2.2.3.1 Purification of McrB and McrB $\Delta$ N**

McrB and McrB $\Delta$ N were expressed with a tag of six histidines at C-terminus by overexpression of pHISMcrB and pHISMcrB $\Delta$ N plasmids, respectively in *E. coli* BL21 (AI) cells. The tag was preceded by a glycine and a serine. The cultures were grown in 2 L LB media containing 100  $\mu$ g/ml ampicillin in an incubator-shaker at 37°C until OD reached 0.3 at 600 nm. The temperature of incubator-shaker was then reduced to 18°C and cultures were induced with 0.06% w/v L-Arabinose. The cultures were grown further overnight (15-16 hours) at 18°C.

Cells were pelleted by centrifugation at 4°C and 3,315 g for 15 minutes. The pellet was resuspended in 50 ml lysis buffer (50 mM Tris-Cl pH 8, 25 mM imidazole, 500 mM NaCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, 0.04% CHAPS). The cells were lysed by sonication at 4°C using Vibra-Cell™ system set at 60% amplitude, pulsing in 1 sec ON and 3 sec OFF mode for 3 minutes. The cycle was repeated after 10 minutes to allow heat dissipation. The cell lysate was then clarified by ultracentrifugation at 4°C and 159,200 g for 40 minutes. Both McrB and McrB $\Delta$ N were first purified by affinity chromatography using an identical strategy. The clarified supernatant of the cell lysate was loaded onto a 5 ml NiNTA column (GE Life Sciences) equilibrated with Buffer A (50 mM Tris-Cl pH 8, 25 mM imidazole, 500 mM NaCl). The protein was eluted using Buffer A and Buffer B (50 mM Tris-Cl pH 8, 500 mM Imidazole, 500 mM NaCl) by a step gradient from 5% to 100% at intervals of 20%. The purest of the NiNTA fractions were dialyzed against 2 L dialysis buffer (50 mM Tris-Cl pH 8, 50 mM NaCl, 1 mM EDTA, and 1 mM DTT) at 4°C for 2 hours.

Dialysed McrB or McrB $\Delta$ N was loaded onto an 8 ml MonoQ 10/100 GL column (GE Life Sciences) equilibrated with Buffer B50 (50 mM Tris-Cl pH 8, 50 mM NaCl, 1 mM EDTA, 1 mM DTT). 2 ml fractions were collected in 20 column volumes over a linear gradient of 0% to 50% buffer using B50 and B1000 (50 mM Tris-Cl pH 8, 1000 mM NaCl, 1 mM EDTA, 1 mM DTT). The pure fractions were pooled and concentrated using a 2 ml 10 kDa vivaspin2 concentrator (GE Life Sciences). Concentrated sample (500  $\mu$ l) was washed with 5 ml buffer B100 (50 mM Tris-Cl pH 8, 100 mM NaCl and 1 mM DTT) to remove EDTA. The concentrated protein was then incubated with 2.5 mM GTP, 5 mM MgCl<sub>2</sub> for 10 minutes at room temperature followed by centrifugation at 21,000 g before loading onto 24 ml Superdex200 10/300 GL column (GE Life Sciences), equilibrated with buffer B100<sup>+GTP</sup> (50 mM Tris-Cl pH 8, 100 mM NaCl, 0.1 mM GTP, 5 mM MgCl<sub>2</sub>, 1 mM DTT). Pure fractions were pooled and concentrated using a 2 ml 10 kDa Vivaspin2 concentrator (GE Life Sciences). The concentrated protein was washed with storage buffer (100 mM NaCl, 10 mM Tris-Cl pH 7.4 and 1 mM DTT) to remove GTP. Protein concentration was estimated using the Bradford reagent (Bradford, 1976) with BSA as standard. The concentrated proteins were stored in storage buffer at -80°C.

### 2.2.3.2 Purification of McrC

McrC was purified with a method very similar to that of McrB and McrB $\Delta$ N purification. McrC, without a histidine tag, in pHISMcrC, was overexpressed in *E. coli* BL21 (AI) cells. The cultures were grown in 2 L LB media containing 100  $\mu$ g/ml ampicillin in an incubator-shaker at 37°C until OD reached 0.3 at 600 nm. The temperature of incubator-shaker was then reduced to 18°C and cultures were induced with 0.06% w/v L-Arabinose. The cultures were grown further overnight (15-16 hours) at 18°C. Cells were pelleted by centrifugation at 4°C and 3,315 g for 15 minutes. The pellet was resuspended in 50 ml lysis buffer and lysed by sonication at 4°C. The cell lysate was then clarified by ultracentrifugation at 4°C and 159,200 g for 40 minutes. The clarified supernatant was loaded onto a 5 ml NiNTA column (GE Life Sciences) equilibrated with Buffer A. The protein was eluted using Buffer A and Buffer B by a step gradient from 5% to 100% at an interval of 20%. The purest of the NiNTA fractions were dialyzed against 2 L dialysis buffer at room temperature for 45 minutes. Dialysed protein was loaded onto an 8 ml MonoS 10/100 GL column (GE Life Sciences) equilibrated with Buffer B50. 2 ml fractions were collected in 20 column volumes over a linear gradient of 0% to 50% buffer using B50 and B1000. The pure fractions were pooled and concentrated using a 2 ml 10 kDa Vivaspin2 concentrator (GE Life Sciences). The concentrated protein was washed with storage buffer (see above) to remove EDTA and the concentrated pure protein was stored in storage buffer at -80°C.

### 2.2.3.3 Purification of McrBC/McrB $\Delta$ NC complex

After purifying the individual subunits, a complex of McrB or McrB $\Delta$ N with McrC was purified through size exclusion. McrB or McrB $\Delta$ N was mixed with McrC at 4 fold higher molar concentration (i.e 4:1 ratio) and incubated with 2.5 mM GTP and 5 mM MgCl<sub>2</sub> in buffer B100 for 10 minutes at room temperature. Sample was centrifuged at 21,000 g for 15 minutes before loading onto 120 ml Superdex200 10/300 GL column (GE Life Sciences), equilibrated with buffer B100<sup>+GTP</sup>. Pure fractions were pooled and concentrated using a 2 ml 10 kDa Vivaspin2 concentrator (GE Life Sciences). The concentrated protein was washed with storage buffer to remove GTP. Protein

concentration was estimated using the Bradford reagent (Bradford, 1976) with BSA as standard. The concentrated complex was stored in storage buffer at -80°C

#### **2.2.4 Analytical Size Exclusion Chromatography (SEC)**

Analysis of the oligomeric status of McrB and McrBΔN was carried out by SEC using 24 ml Superdex200 10/300 GL either in presence or absence of GTP. For studies without and with nucleotide, the column was equilibrated with buffer B100 and B100<sup>+GTP</sup> respectively. For both McrB and McrBΔN, 500 μl solution containing 18 μM protein (~ 1 mg/ml) in buffer B100 was injected with or without 2.5 mM GTP and 5 mM MgCl<sub>2</sub>. The sample was centrifuged for 15 minutes at 21,000 g, 4°C before loading. The column was calibrated using a set of standard protein solutions. Blue dextran 2000 was used to determine the void volume ( $V_o = 8.4$  ml). The standards used for calibration of molecular masses were β amylase (200 kDa;  $V_e = 12.3$  ml), alcohol dehydrogenase (150 kDa;  $V_e = 13.3$  ml), bovine serum albumin (66 kDa;  $V_e = 14.3$  ml), ovalbumin (43 kDa;  $V_e = 15.7$  ml) and carbonic anhydrase (29 kDa;  $V_e = 16.8$  ml). The  $K_{av}$  was calculated as  $K_{av} = (V_e - V_o) / (V_t - V_o)$ , where  $V_t = 24$  ml (total column volume).

The oligomerisation of McrB in presence of McrC and GTP was studied using a 24 ml Superose6 10/300 GL SEC column (GE Life Sciences). The column was equilibrated with B100<sup>+GTP</sup>. 500 μl solution containing 18 μM McrB or McrBΔN and 4.5 μM McrC (4:1 ratio) in B100, with 2.5 mM GTP and 5 mM MgCl<sub>2</sub> was incubated for 10 minutes at room temperature and centrifuged at 21,000 g before injection.

#### **2.2.5 Size Exclusion Chromatography coupled Multi-Angle Light Scattering (SEC-MALS)**

The molecular mass in solution of McrB, McrBΔN, McrBC and McrBΔNC was determined by SEC-MALS measurements using a Wyatt Heleos II 18 angle light scattering instrument coupled to a Wyatt Optilab rEX online refractive index detector. Detector 12 in the Heleos instrument was replaced with Wyatt's QELS detector for dynamic light scattering measurement. Protein samples (100 μL) were resolved using a Superdex S-200 (McrB) or Superose 6 (McrBC) 10/300 analytical gel filtration column (GE Healthcare) running at 0.5 ml/min in 10 mM Tris pH 7.4, 100 mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM DTT with and without

0.1 mM nucleotide (GTP or GDP) buffer before passing through the light scattering and refractive index detectors in a standard SEC-MALS format.

Protein concentration was determined from the excess differential refractive index (RI) based on 0.186 RI increment for 1 g/ml protein solution. The concentration and the observed scattered intensity at each point in the chromatograms were used to calculate the absolute molecular mass from the intercept of the Debye plot using Zimm's model as implemented in Wyatt's ASTRA software. Autocorrelation analysis of data from the dynamic light scattering detector was also performed using Wyatt's ASTRA software and the translational diffusion coefficients determined were used to calculate the hydrodynamic radius ( $R_h$ ) using the Stokes-Einstein equation and the measured solvent viscosity of  $9.3 \times 10^{-3}$  Poise.

### **2.2.6 GTP hydrolysis assay**

The GTPase activity was qualitatively measured by monitoring release of phosphate ion ( $P_i$ ) using a standard malachite green assay (Baykov et al, 1988; Geladopoulos et al, 1991). Each GTPase assay was performed in a set of triplicates. A master mix containing protein and 1mM GTP (Jena Bioscience) in hydrolysis buffer (10 mM Tris-Cl (pH 8), 50 mM KCl, 5 mM  $MgCl_2$ , 1 mM DTT) was incubated at 37°C. To check the effect of DNA on the GTPase activity a final concentration of 1  $\mu$ M of either the 60 bp specific or the 62 bp non-specific DNA was added (see below for sequence details). 20  $\mu$ l volume of the reaction mix was withdrawn at regular time intervals and 5  $\mu$ l of 0.5 M EDTA was added to stop the reaction at each time point. The samples were then transferred to a 96 well flat bottom plate. 50  $\mu$ l of the freshly prepared malachite green mix (800  $\mu$ l malachite green solution, 200  $\mu$ l of 7.5% ammonium molybdate and 16  $\mu$ l of 11% Tween 20) was added to each reaction and incubated for 10 minutes at room temperature. Absorbance was measured at 630 nm in Varioscan plate reader.

The malachite green solution was prepared by adding 44 mg malachite green carbinol base (Sigma-Aldrich) powder to 36 ml 3 N sulphuric acid solution. A reaction mixture quenched with EDTA at 0 minutes was used as a blank and was included for every set of reactions. This blank reading was comparable to the absorbance measured for hydrolysis buffer containing 1mM GTP but without protein. Blank absorbance reading

was taken at the end of one hour and subtracted from all the absorbance readings to rule out spontaneous GTP hydrolysis at 37°C. To measure the amount of Pi released (in moles), standard phosphate curve was plotted by preparing different dilutions of a 2 M aqueous NaH<sub>2</sub>PO<sub>4</sub> solution. 50 µl of malachite green solution was added to 25 µl of each dilution and the reaction was incubated for 10 minutes at room temperature.

### **2.2.7 DNA binding studies**

Two complementary, 60bp ssDNA MB60MSPI-1F and MB60MSPI-1R (Table 2.2), containing a methylated cytosine, were mixed in equimolar amount and heated at 99°C for 5 minutes. The mix was then allowed to cool to 25°C at a rate of 1°C/s. The DNA was run on a 10% native PAGE to confirm purity. A 62 bp unmethylated dsDNA was used as a non-specific substrate for DNA binding studies. This substrate was obtained by annealing two single-stranded (ss) DNA oligonucleotides MB62-1F and MB62-1R (Table 2.2). The oligonucleotides were obtained from Integrated DNA Technologies, USA. The annealed product was purified using a MonoQ 10/100 GL column.

*Table 2.2. List of oligonucleotides used DNA binding studies*

MB60MSPI-1F	GCCGGGTAACCGGGTAAGTCCGGGTAAGAmCCGGTAGTTCGGA TCGAGGGGT AGGCCGC
MB60MSPI-1R	GCGGCCTACCCCTCGATACCGAACTAmCCGGTCTTACCCGG ACTTACCCG GGTTACCCGGC
MB62-1F	GAGTCAATCGGATCGTAGACGTACTAGACCTATCCTGTATGCTAC GTATTCGTAT CGTGAGC
MB62-1R	GCTCACGATACGAATACGTAGCATACAGGATAGGTCTAGTACG TCTACGATCCGATTGACTC

Binding reactions were carried out at room temperature in a binding buffer containing 10 mM Tris-Cl pH 8, 50 mM KCl, 15 mM MgCl<sub>2</sub>, 1 mM DTT, 10% glycerol. A 10 µl reaction mix containing 250 nM of either specific or the non-specific DNA was incubated with different protein concentrations in presence or absence of 1 mM nucleotide (GTP (Jena Bioscience), GDP (Sigma-Aldrich), GDPNP (Jena Bioscience). 0.4 mg/ml BSA (NEB) was added to each reaction for enzyme stability. Reactions were incubated

at room temperature for 10 minutes and 2  $\mu$ l 6X ST buffer (40% Sucrose, 0.2 M Tris-Cl pH 7.5) was added before loading onto a 5% native polyacrylamide gel. The native gel was pre-electrophoresed for 30 minutes at 4°C, 20 mA in 1XTBE buffer (89 mM Tris-Borate, 2 mM EDTA, pH 8.3). The gel was run at a constant current of 20 mA at 4°C for 20 minutes and stained using a solution containing 2  $\mu$ g/ml ethidium bromide for 5 minutes. The gels were imaged using Typhoon TRIO+ variable mode imager at high sensitivity.

### **2.2.8 DNA cleavage assay**

A 114 bp substrate MB114MspI was generated by overlap PCR using MB60MSPI-1F and MB60MSPI-2R- 5'AGTCAAATTGCATATGCTGGTCTTTTCAGCGmCCGGTAATCGTCTTGTGAAGGATCCGCGGC-3'- as primers. The duplex MB114MspI was purified by gel extraction from a 2% agarose gel. Nucleolytic cleavage of DNA was carried out in 10  $\mu$ l reaction mix of a digestion buffer (10 mM Tris-Cl pH 8, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT) containing 75 nM MB114MspI incubated with protein in presence or absence of 1 mM GTP (Jena Bioscience). The reactions were incubated at 37°C for 30 minutes. 2  $\mu$ l 6X STES buffer (40% Sucrose, 0.2 M Tris-Cl pH 7.5, 40 mM EDTA, 1% SDS) was added before loading on a 10% native polyacrylamide gel (pre-electrophoresed for 30 minutes at room temperature in 1XTBE buffer). The gel was run at 230 V for 40 minutes, and then stained with a solution containing 2  $\mu$ g/ml ethidium bromide for 5 minutes and imaged on Typhoon TRIO+ variable mode imager at high sensitivity.

### **2.2.9 Electron cryomicroscopy (Cryo-EM)**

For Cryo-EM studies, grid preparation and data collection were carried out by Dr. Vinothkumar Kutti at MRC LMB, Cambridge UK. Full length and the N-terminally truncated McrB were assembled by addition of 1 mM GDPNP and 1 mM MgCl<sub>2</sub> at a final concentration of 2.5 mg/ml in 10 mM Tris-Cl pH 7.4 and 0.1 M NaCl. 3  $\mu$ l of the assembled complex was applied to a glow discharged Quantifoil grids R 0.6/1 or 1.2/1.3 and blotted for 11 seconds, then plunge-frozen in liquid ethane using an environmental plunge-freeze apparatus (Bellare et al, 1988). McrB $\Delta$ N were transferred to Krios cartridges and imaged with a FEI Titan Krios electron microscope and Falcon II direct detector at 300 keV with



the specimen temperature at  $-186^{\circ}\text{C}$  at a calibrated magnification of 105263x (nominal magnification is 59,000), corresponding to a  $1.33 \text{ \AA}/\text{pixel}$  with a 2.5 second exposure. The full-length McrB was imaged on the Polara microscope at 300 keV equipped also with a Falcon II detector with the specimen temperature at  $-186^{\circ}\text{C}$  at a calibrated magnification of 104477x (nominal magnification is 78,000), corresponding to a  $1.34 \text{ \AA}/\text{pixel}$  with a 3 second exposure.

All image processing was done with EMAN2 (Tang et al, 2007). Particles from both data sets were picked with e2boxer and extracted with a box size of 160 pixels. Preferred orientation was observed and only top/bottom views were picked. A total of 1568 and 3350 particles from the full length and truncated McrB were subjected to reference-free 2D class averaging. The oligomeric state of both versions of McrB in class averages was further checked with rfitim (Crowther & Amos, 1971).

## **2.3 Results**

### **2.3.1 Purification of McrB, McrB $\Delta$ N and McrC**

The proteins- McrB, McrB $\Delta$ N and McrC were purified to 99% homogeneity. Figure 2.1 shows the SDS PAGE for different purification steps. Although McrC did not have a histidine tag we observed good affinity of the protein towards a NiNTA column. All three proteins- McrB, McrB $\Delta$ N and McrC were first pulled from clarified cell lysate using metal affinity chromatography. The proteins, even McrC, were fairly pure with small impurities. Since these are DNA binding and nucleotide binding proteins, ion exchange chromatography using a MonoQ (McrB and McrB $\Delta$ N) or MonoS (McrC) was employed to get rid of such contaminants. Also, this step facilitated in concentrating the sample at an optimum salt concentration. In our efforts to purify the McrB and McrC, problems like salt sensitivity and temperature sensitivity were observed.

All proteins, especially McrC showed a tendency to precipitate upon concentrating, lowering salt or change in temperature. It was also observed that amongst all three proteins, McrC was not stable if kept on ice but did not precipitate if kept at temperatures above  $8^{\circ}\text{C}$ . McrB and McrB $\Delta$ N were purified by affinity and ion-exchange chromatography to ~95% homogeneity (Figure 2.1). An additional SEC column was used to purify the

samples in presence of GTP. During McrB purification, I observed a persistent protein band at ~35 kDa even after SEC runs. This band was believed to be the native McrBs

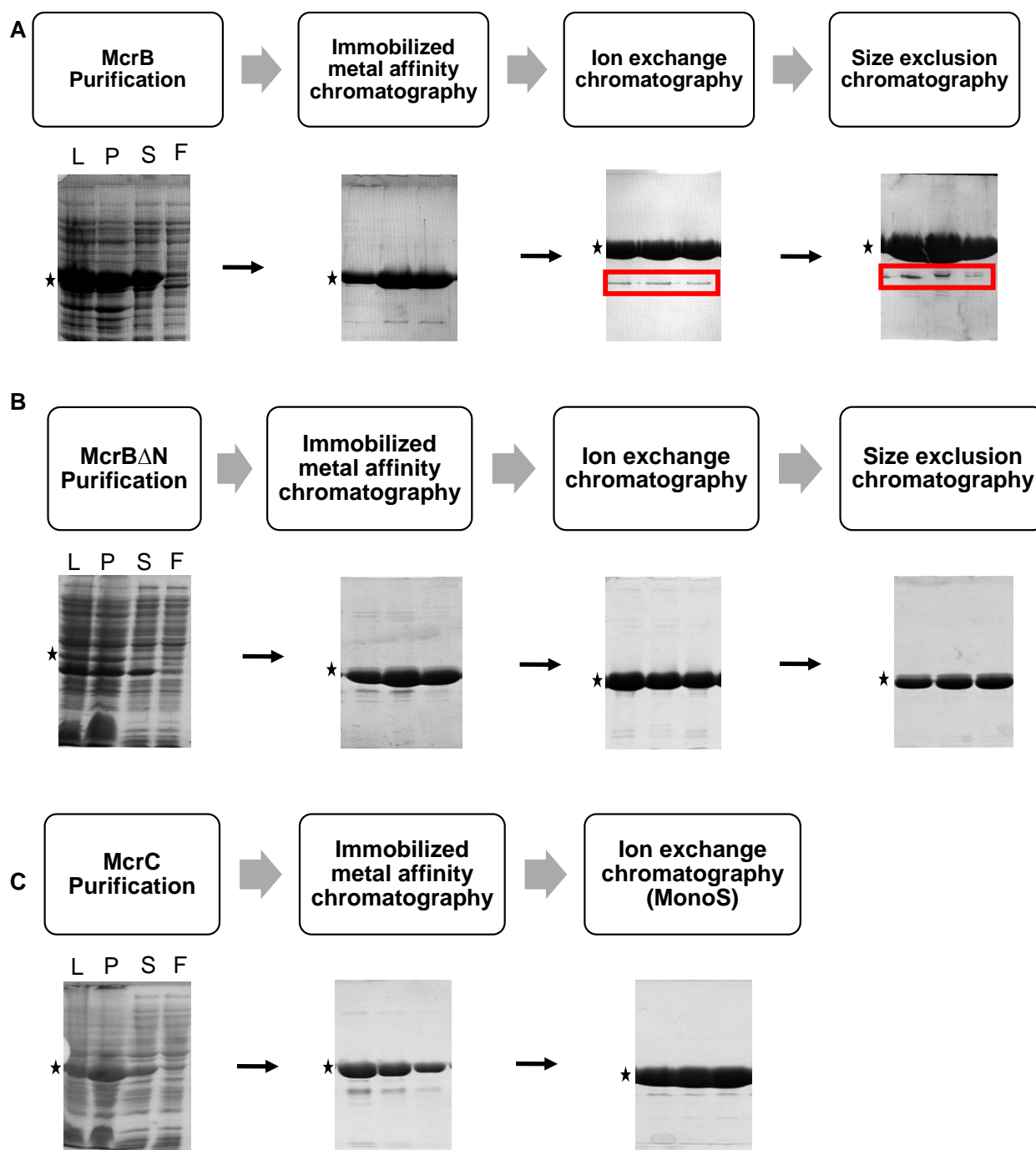


Figure 2.1. **Purification of McrB, McrB $\Delta$ N and McrC.** SDS PAGE gels corresponding to different steps in purification of A) McrB. The red box highlights the ~35 kDa band that was persistent in McrB purification, B) McrB $\Delta$ N and C) McrC. The star indicates the specific protein in each gel. Key- L: Lysate, P: Pellet, S: Supernatant, F: Flow through.

expressed in the cell. Since both full-length McrB and McrBs show very similar oligomeric properties, it is possible that it oligomerized with full-length protein.

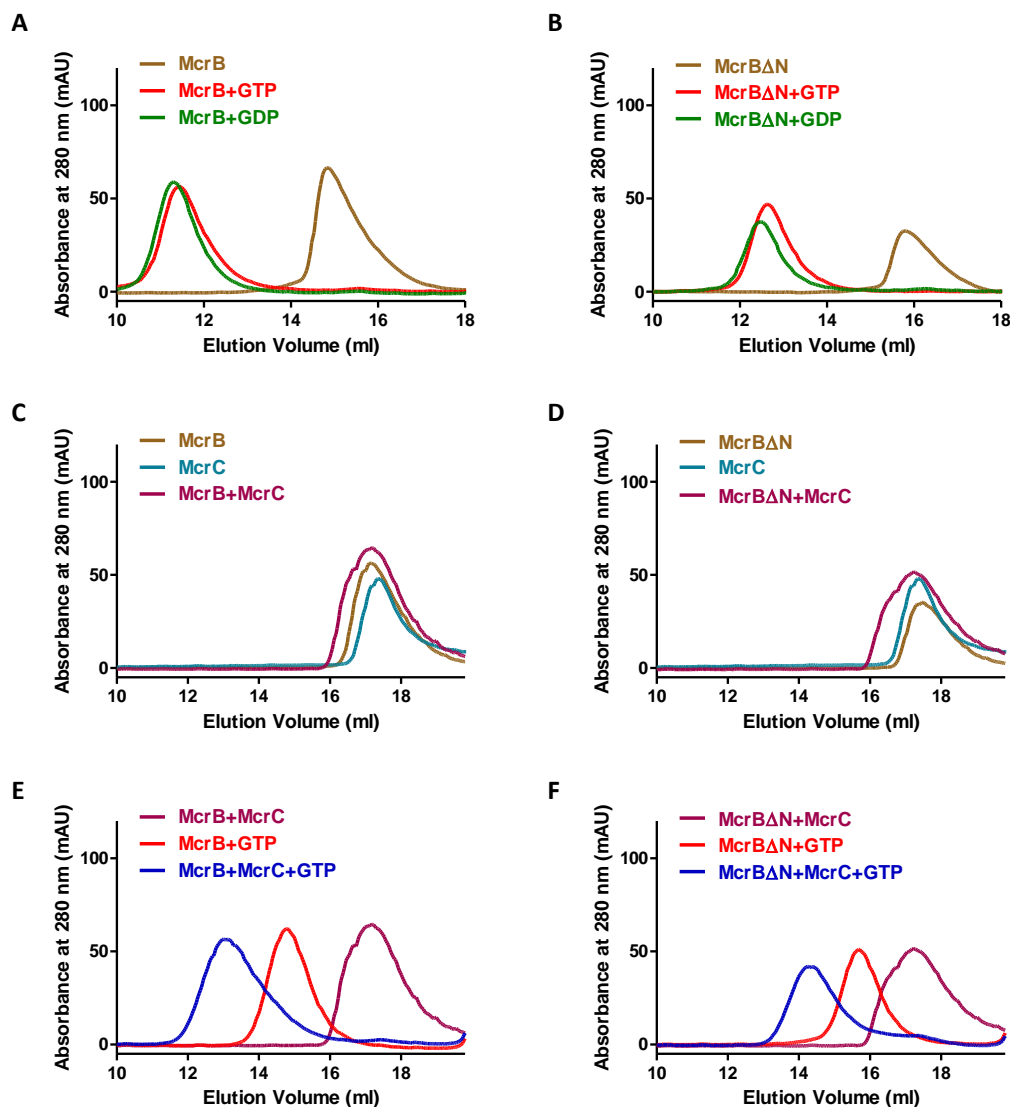
### **2.3.2 Oligomeric states of McrB in presence of nucleotides**

Since McrB and McrB $\Delta$ N both have the AAA+ domain, oligomerization is an extremely important characteristic of these proteins. Before assessing the functional state of the purified state I first carried out oligomerization studies of these proteins. For this I used SEC-MALS.

#### **2.3.2.1 Size Exclusion chromatography**

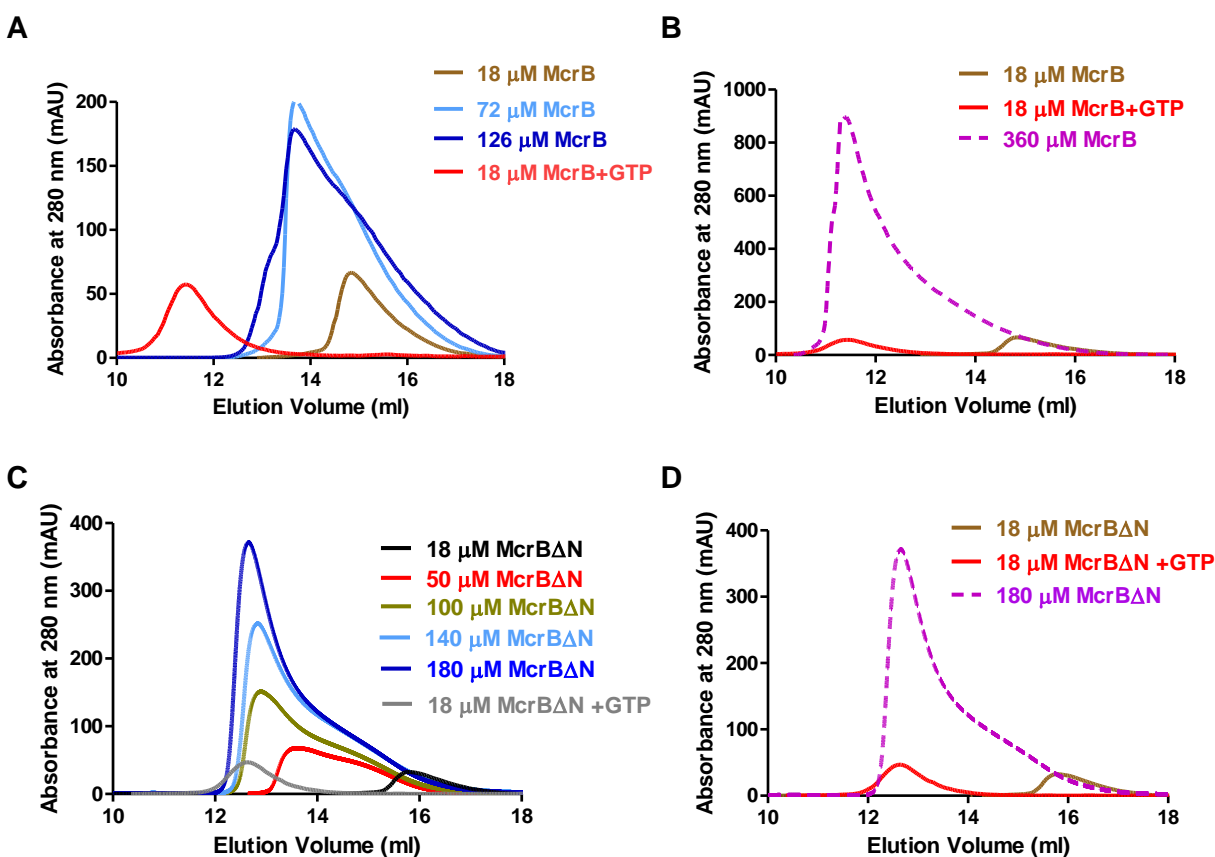
To assess the oligomeric states, size exclusion chromatography was used. In presence of GTP or GDP, McrB and McrB $\Delta$ N eluted as a high molecular mass oligomer (Figure 2.2 A, B). SEC studies indicated an apparent molecular mass of ~380 kDa for McrB, which is consistent with an oligomeric form containing seven subunits (theoretical molecular mass = 372 kDa). However, the apparent molecular mass of McrB $\Delta$ N was ~210 kDa, which corresponded to a six subunit-oligomer (theoretical molecular mass = 208 kDa).

Further, McrB showed interaction with McrC only in presence of nucleotide. In absence of nucleotide, a mixture of McrB and McrC eluted at a volume similar to that of McrB or McrC monomers (Figure 2.2 C). McrB $\Delta$ N and McrC mix, on the other hand, showed a slight shift towards higher molecular mass (Figure 2.2 D). In presence of GTP, both McrB and mcrB $\Delta$ N formed a higher order complex with McrC. McrBC and McrB $\Delta$ NC complex had a molecular mass much larger than the McrB-GTP and McrB $\Delta$ N-GTP oligomer respectively as observed by SEC using a 24 ml Superose6 10/300 GL column (Figure 2.2 E,F). These observations were consistent with the existing model of assembly of McrBC complex made of two rings of McrB and two subunits of McrC (Panne et al, 2001)



**Figure 2.2. Nucleotide driven oligomeric assembly of McrBC/McrB $\Delta$ NC.** A) SEC elution profile of McrB using 24 ml Superdex200 10/300 GL column. In absence of GTP, the protein elutes at 14.85 ml while in presence of nucleotide, the protein elution peak shifts to 11.44 ml. B) Size Exclusion chromatography elution profile of McrB $\Delta$ N using 24 ml Superdex200 10/300 GL column showing an elution peak shift from 15.8 ml (in absence of GTP) to 12.64 ml (in presence of GTP). C) Gel filtration profile of 18  $\mu$ M McrB, 25  $\mu$ M McrC and a mixture of 18  $\mu$ M McrB and 25  $\mu$ M McrC without GTP on a 24 ml Superose6 10/300 GL column. D) Gel filtration profile of 18  $\mu$ M McrB $\Delta$ N, 25  $\mu$ M McrC and a mixture of 18  $\mu$ M McrB $\Delta$ N and 25  $\mu$ M McrC without GTP on a 24 ml Superose6 10/300 GL column, E) Gel filtration profile of McrB with and without McrC in presence of GTP (using 24 ml Superose6 10/300 GL column). F) Gel filtration profile of McrB $\Delta$ N with and without McrC in presence of GTP (using 24 ml Superose6 10/300 GL column). The study shows that in presence of McrC, McrB peak (oligomer observed in presence of GTP) shifts from 14.8 ml to 13 ml and McrB $\Delta$ N (observed in presence of GTP) peak shifts from 15.7 ml to 14.2 ml.

In the absence of GTP or GDP and at a protein concentration of 18  $\mu\text{M}$ , both McrB and McrB $\Delta\text{N}$  eluted as a single peak with an apparent molecular mass of 64 kDa and 39 kDa (Figure 2.2 A, B), which are close to the values expected for their monomeric forms. However, McrB showed a concentration-dependent oligomerization even in the absence of GTP. At a protein concentration of 72  $\mu\text{M}$  and 126  $\mu\text{M}$ , McrB eluted as an inhomogeneous mixture of different molecular weight species, composed possibly of monomers, dimers, trimers etc (Figure 2.3 A). At a much higher concentration (378  $\mu\text{M}$ ),



**Figure 2.3. Nucleotide independent-concentration dependent oligomerization of McrB/McrB $\Delta\text{N}$ .** Gel filtration profile (using 24 ml Superdex200 10/300 GL) of A) different concentrations (without GTP) of McrB, B) 378  $\mu\text{M}$  McrB without GTP (using 24 ml Superdex200 10/300 GL) indicating the formation of a higher-order oligomers, C) different concentrations of McrB $\Delta\text{N}$  without GTP and comparison with McrB $\Delta\text{N}$  + GTP (using 24 ml Superdex200 10/300 GL), D) Gel filtration profile of 180  $\mu\text{M}$  McrB $\Delta\text{N}$  without GTP (using 24 ml Superdex200 10/300 GL). The broad elution peak indicates the heterogeneity in the oligomeric population with different sizes encompassing the monomeric to heptameric (GTP induced oligomeric size- (Panne et al, 2001)) elution volume.

McrB eluted as a higher order oligomer with an apparent molecular mass corresponding to McrB-GTP oligomer (Figure 2.3 B). Concentration-dependent oligomerization was also observed when a similar study was done with McrB $\Delta$ N (Figure 2.3 C). Since McrC showed interaction with oligomeric McrB/ McrB $\Delta$ N (GTP bound oligomers) (Figure 2.2 E, F), we next sought to find whether McrC could interact with oligomers formed as a result of increased concentration. For this, we injected McrB/McrB $\Delta$ N at high concentration incubated with McrC (maintaining the McrB/McrC $\Delta$ N:McrC::4:1 ratio) (Figure 2.4). We observed that both in case of McrB and McrB $\Delta$ N, oligomers formed at high protein concentration indeed interacted with McrC as the elution peak shifted towards a higher molecular weight (Figure 2.4 A, B). Presence of McrC was checked by running the fractions on SDS PAGE. This observation indicates that the role of the nucleotide in McrB and McrC interaction might primarily be for the formation of McrB oligomers.

To investigate whether smaller oligomers (smaller than those eluting at heptameric size) could also show interaction with McrC, we injected McrC with McrB $\Delta$ N (50  $\mu$ M) at a concentration where we could observe oligomeric sizes intermediate to monomer and heptamer (Figure 2.3 C). Again, McrC caused the formation of higher oligomers, although the complex did not elute at the position of the tetradecamer. Whether McrC interacts with McrB in monomeric form could not be ascertained due to the resolution limit of the SEC columns used in this study.

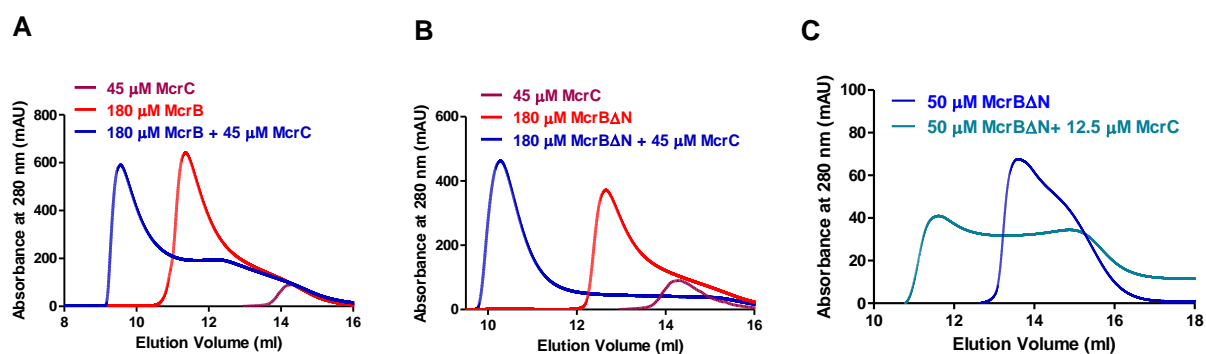


Figure 2.4. **Binding of McrB/McrB $\Delta$ N and McrC in absence of GTP.** Gel filtration profile (using 24 ml Superdex200 10/300 GL) of A) McrC interaction with oligomeric form of McrB at 180  $\mu$ M, B) McrC interaction with oligomeric form of McrB $\Delta$ N at 180  $\mu$ M and C) McrC interaction with oligomeric form of McrB at 50  $\mu$ M.

### 2.3.2.2 Size Exclusion Chromatography Coupled Multi-Angle Light Scattering

In the previous section, I showed that SEC runs with GTP and GDP (Figure 2.2 A, B) showed a slight shift in the elution peak at 11.4 and 10.8 ml respectively. McrB is classified under Clade VI of AAA+ family along with bacterial enhancer binding proteins like NtrC1, NtrC4, PspF and ZraR (Erzberger & Berger, 2006). Crystal structures both heptameric - NtrC1, PspF and NtrC4 (Chen et al, 2007; Chen et al, 2010; Lee et al, 2003; Zhang et al, 2014) and hexameric - NtrC and PspF; (Batchelor et al, 2009; De Carlo et al, 2006; Rappas et al, 2005; Schumacher et al, 2004; Sysoeva et al, 2013) have been determined. As the shape of the macromolecules influences their mobility during

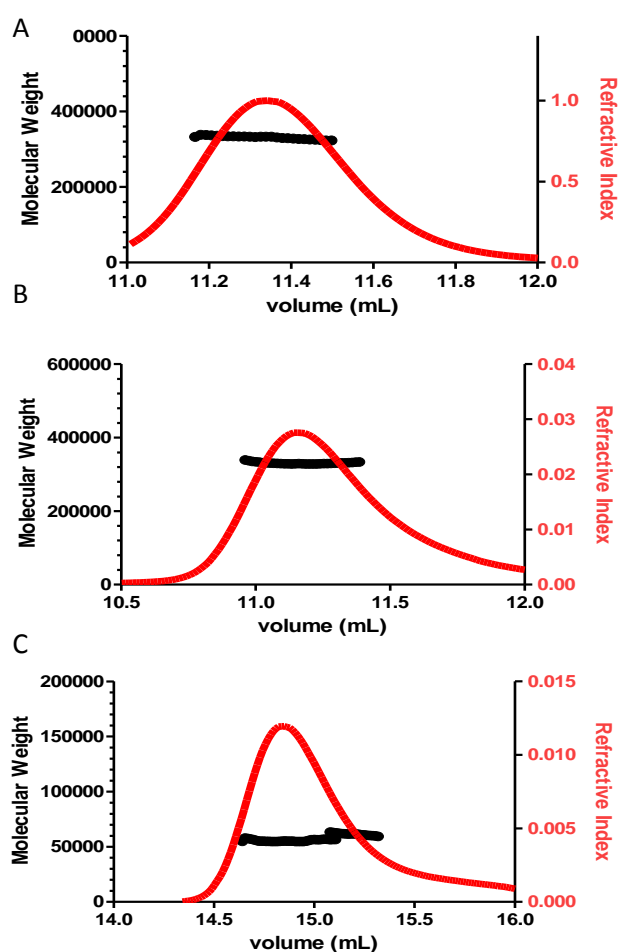


Figure 2.5. **SEC-MALS chromatogram of McrB:** The chromatogram shows the refractive index signal with the derived molar masses indicated by the thicker horizontal lines. A) In presence of GTP, B) in presence of GDP displayed single highly monodisperse peaks with average mass over the indicated regions 320 kDa, C) without nucleotide the protein displayed a peak with average molecular mass of 55 kDa. 24 ml Superdex200 10/300 GL was used for SEC.

SEC, the molecular weights determined based on molecular weight standards are not accurate. This possibly explains the disparity in the mass of oligomeric state of McrB and McrB $\Delta$ N derived by SEC. In order to obtain the absolute molecular weight of the protein oligomers, SEC-MALS experiments were carried out. The SEC-MALS runs in presence of nucleotide (both GTP and GDP) showed a monodisperse peak with molecular weight 331.2 kDa (in presence of GTP) and 330.5 kDa (in presence of GDP) (Figure 2.5 A, B). The protein at similar concentration eluted as a monomer in absence of nucleotide with molecular weight 55.5 kDa (Figure 2.5 C).

The SEC-MALS measurements clearly show that in presence of nucleotide, both McrB and McrB $\Delta$ N form a monodisperse peak at a protein concentration of 18  $\mu$ M (Figure 2.6 A). The observed masses from SEC-MALS match relatively well with the calculated mass from the amino acid sequence for a hexameric McrB and McrB $\Delta$ N (Figure 2.6 A, Table1). The complex of McrBC and McrB $\Delta$ NC also show monodisperse peaks and the molecular mass corresponds to the mass of 12 subunits of McrB plus 2 subunits of McrC. This suggests that McrBC is likely to be a heteromeric tetradecamer of two McrB hexamers and two McrC monomers (Figure 2.6 B, Table 2.3). No self-association of McrB hexamers into higher oligomers was observed in absence of McrC (Figure 2.6 A), as was reported earlier (Panne et al, 2001). The samples were also run at 1/10th of the

*Table 2.3. Comparison of calculated mass and mass from SEC-MALS*

Molecular weight of McrB monomer: 54.2 kDa Molecular weight of McrB $\Delta$ N monomer: 35.7 kDa Molecular weight of McrC monomer: 40.6 kDa			
Measurements	Oligomer mass from amino acid sequence		Mass From SEC-MALS
	Hexamer (6*monomer)	Heptamer (7*monomer)	Experimental
<b>McrB</b>	325.2	379.4	320 kDa
<b>McrB<math>\Delta</math>N</b>	214.2	249.9	211 kDa
	Tetradecamer (12*McrB + 2*McrC)	Tetradecamer (14*McrB + 2*McrC)	Experimental
<b>McrBC</b>	731.6	840	720 kDa
<b>McrB<math>\Delta</math>NC</b>	509.6	581	486 kDa



concentrations shown in the figure (that is  $\sim 0.05$  mg/ml loaded and thus  $0.005$  mg/ml or less on the column) and the results were identical (data not shown).

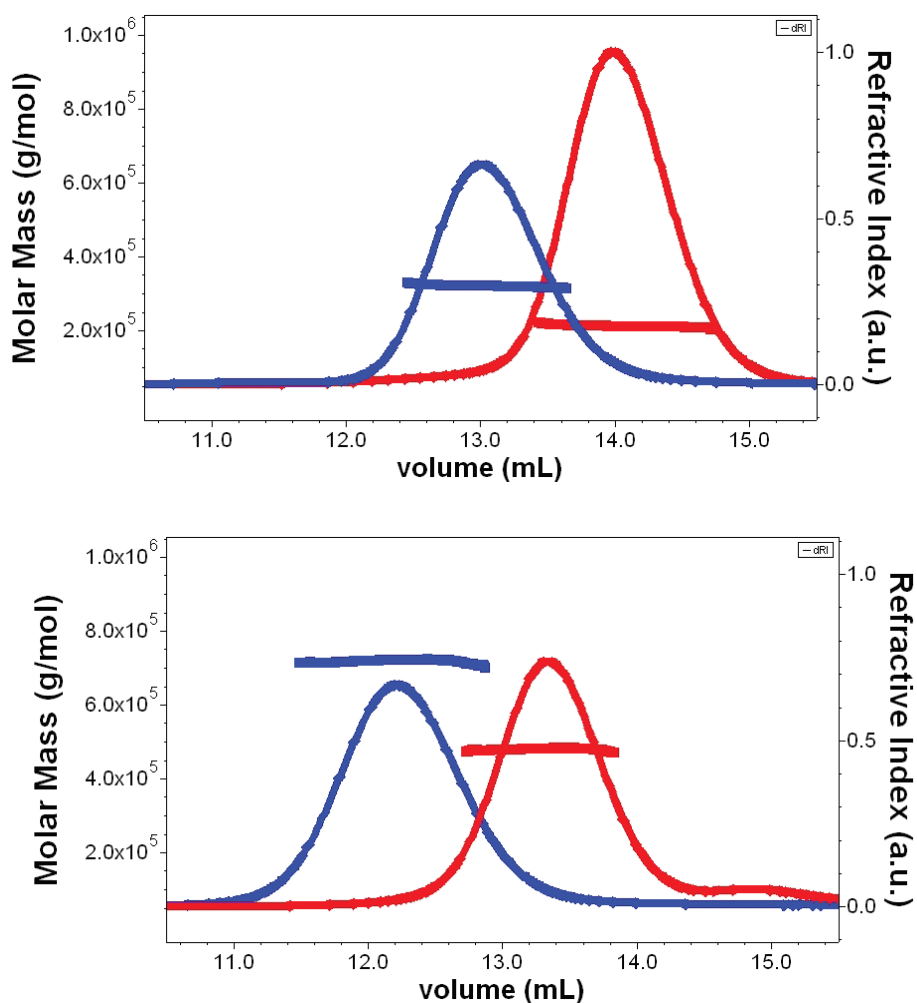


Figure 2.6. **SEC-MALS chromatogram of McrB and McrBC:** The chromatogram shows the refractive index signal with the derived molar masses indicated by the thicker horizontal lines. A) McrB (blue) and McrB $\Delta$ N (red) run at 0.6 and 0.8 mg/ml (loading concentration; column concentration is  $> \times 1/10^{\text{th}}$  than loaded) displayed single highly monodisperse peaks with average mass over the indicated regions 320 kDa and 211 kDa, respectively. The Rh evaluated from DLS data over the same regions was  $6.1 \pm 0.2$  nm and  $5 \pm 0.2$  nm, respectively. B) McrBC (blue) and McrB $\Delta$ NC (red) run at 0.6 mg/ml displayed single highly monodisperse peaks with average mass over the indicated regions 720 kDa and 486 kDa respectively. The Rh evaluated from DLS data over the same regions was  $8.7 \pm 0.2$  nm and  $6.7 \pm 0.2$  nm, respectively. A 24 ml Superdex200 10/300 GL column was used for these studies.

### 2.3.3 Functional characterization of McrBC

As the molecular mass of the oligomers of McrB, McrB $\Delta$ N and McrBC were different from those reported earlier, we sought to find if the purified proteins were active. For this, we qualitatively tested their GTPase, DNA binding, and cleavage activities.

#### 2.3.3.1 GTPase activity:

A time-dependent release of phosphate ion upon hydrolysis of GTP was measured at three different concentrations of McrB and McrB $\Delta$ N. The three concentrations were chosen from a McrB and McrBC concentration-dependent phosphate-release assay (Figure 2.7 A, B).

We noticed only a weak GTPase activity for both McrB and McrB $\Delta$ N (Figure 2.7 A,

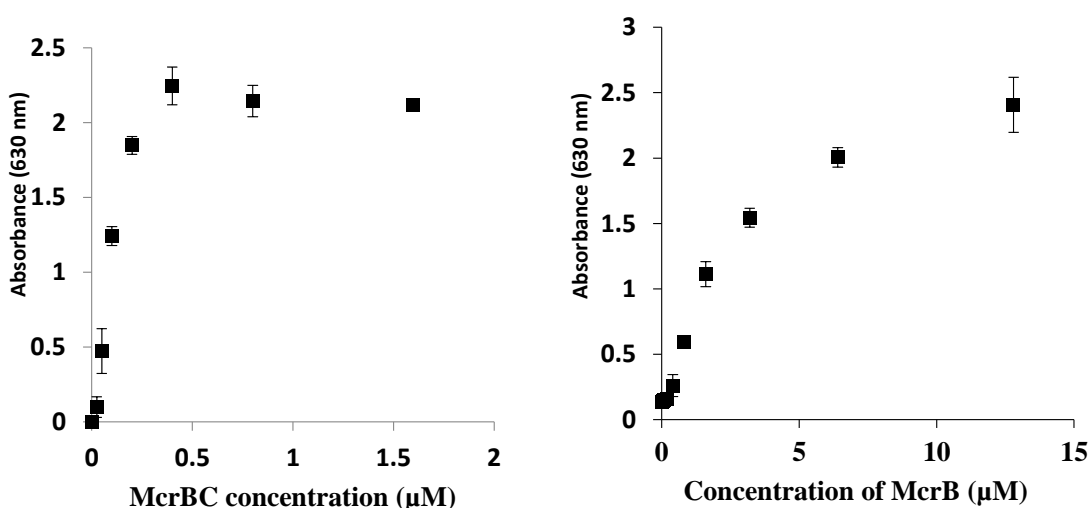


Figure 2.7. **Concentration dependent GTPase activity of McrB and McrBC:** A) Concentration dependent GTPase activity of McrBC in presence of 1mM GTP, B) Concentration dependent GTPase activity of McrB in presence of 1mM GTP

B). There was no significant change in the GTPase activity upon addition of specific DNA (Figure 2.7 C). However, on the addition of McrC to McrB or McrB $\Delta$ N (molar ratio of 4:1 of McrB:McrC or McrB $\Delta$ N: McrC), we noticed more than 30-fold increase in the release of phosphate (Figure 2.7 C, D). Fitting a linear regression line through the time points, we obtained GTPase rate of 0.5 min<sup>-1</sup> for McrB and 15 min<sup>-1</sup> for McrBC which is very well in agreement with previous studies (Pieper et al, 1997).

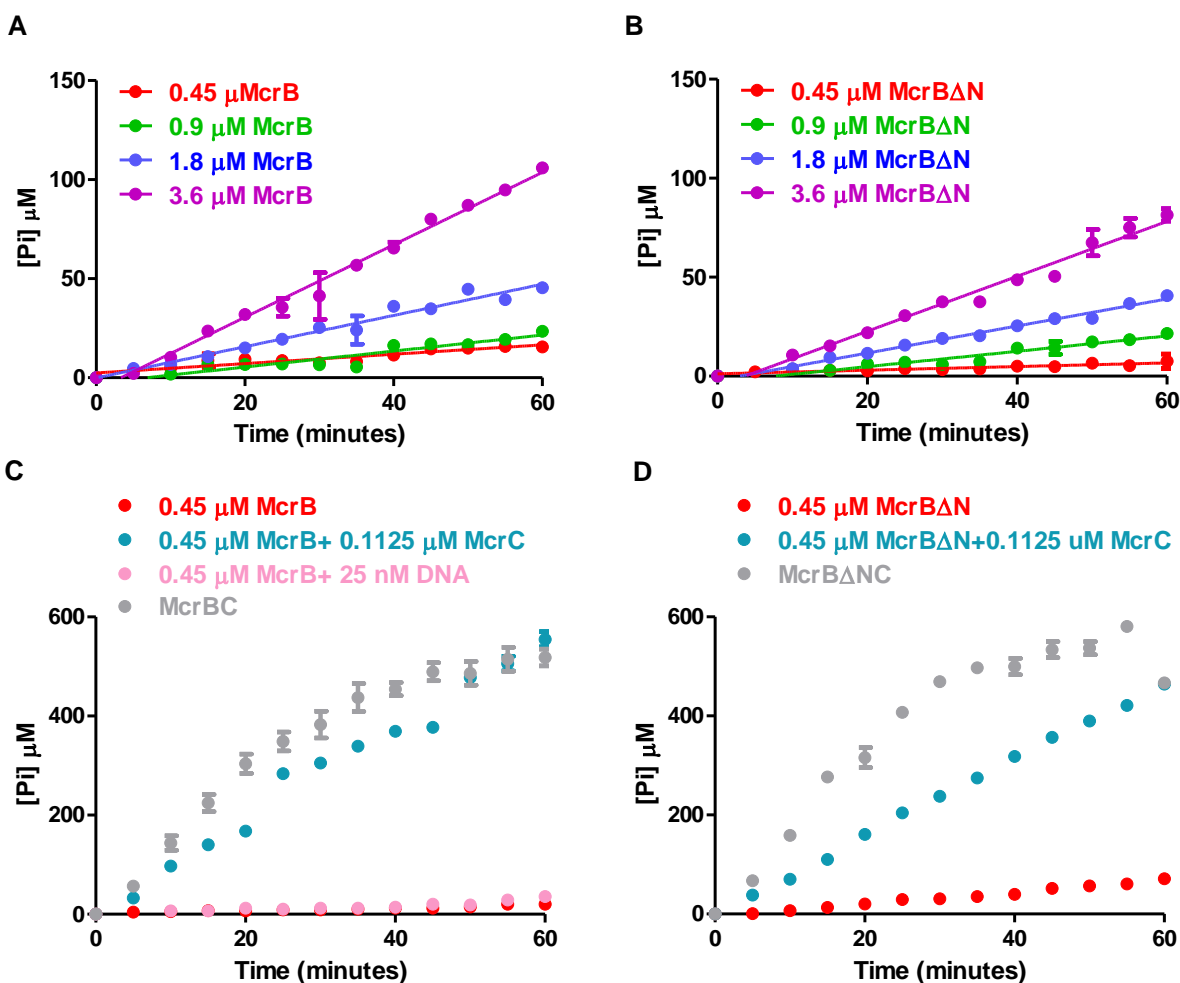


Figure 2.8. **GTPase activity of McrB and McrBC:** A) Time dependent GTPase activity of McrB at different concentrations, B) Time dependent GTPase activity of McrBΔN at different concentrations, C) Comparison of GTPase activity of 450 nM McrB alone, 450 nM McrB in presence of 112.5 nM McrC, 450 nM McrB in presence of 1 μM specific DNA (spDNA) and 37.5 nM McrBC complex. D) Comparison of GTPase activity of 450 nM McrBΔN alone, 450 nM McrBΔN in presence of 112.5 nM McrC and 37.5 nM McrBΔNC complex.

### 2.3.3.2 DNA binding activity:

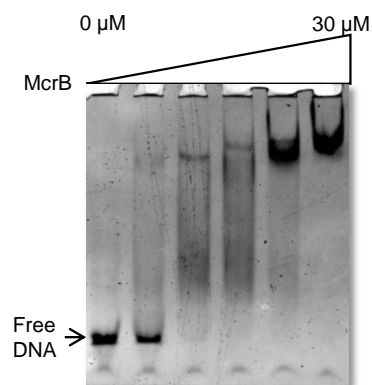
Though DNA did not stimulate the GTPase activity of McrB, we investigated if the protein on itself binds to DNA without the need of McrC. Previous studies had established that McrB in presence of GTP binds specifically to dsDNA containing a target site - methylated

cytosine preceded by a purine (RmC) (Gast et al, 1997; Kruger et al, 1995; Pieper et al, 2002; Stewart et al, 2000).

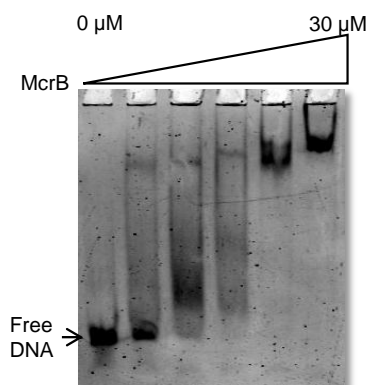
A

5'-GCCGGGTAACCCGGGTAAGTCCGGGTAAGA **C** CG GTAGTTCGGTATCGAGGGGTAGGCCG-3'  
 3'-CGGCC ATTGGGCCA TTCAGGCCA TTCTGGC **C** A TCAAGCCATAGCTCCC CATCCGGCG-5'

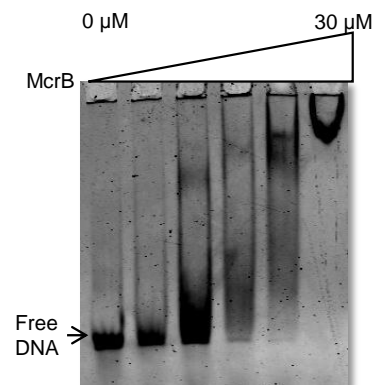
B



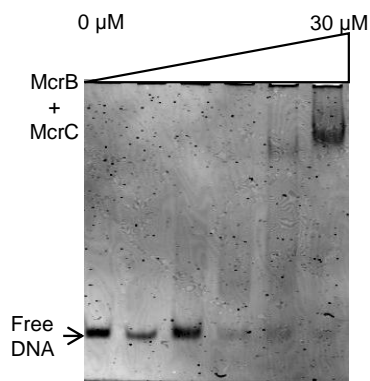
C



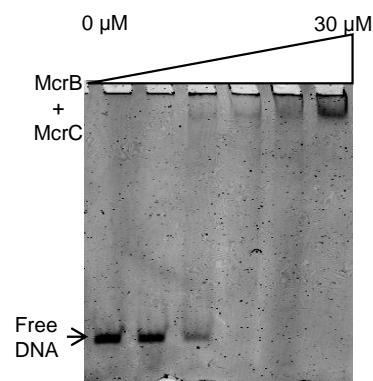
D



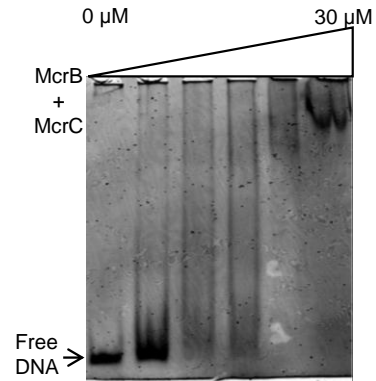
E



F



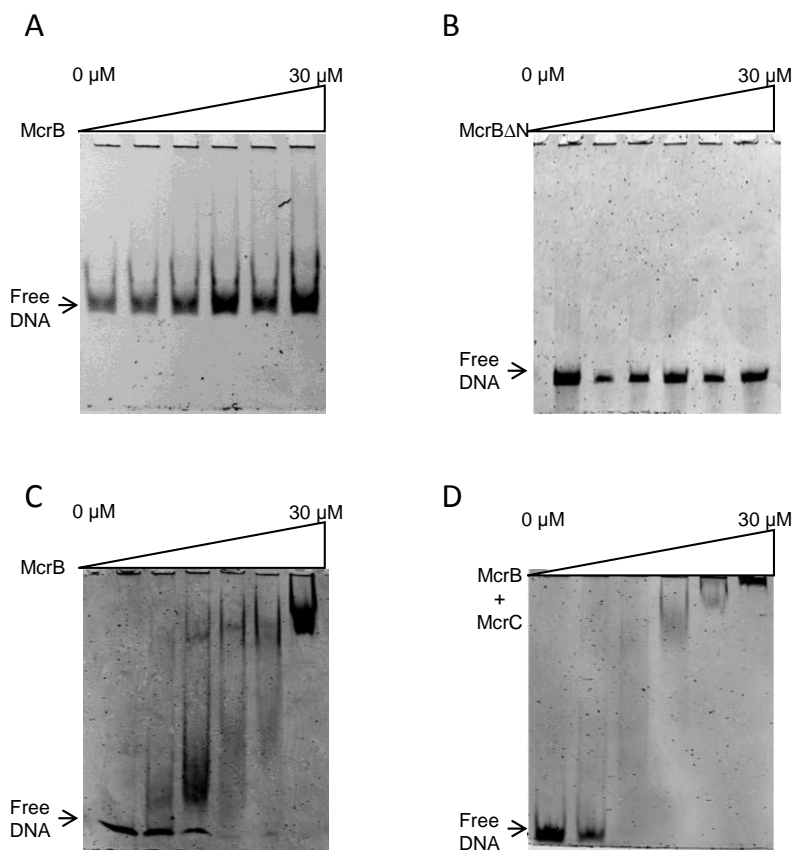
G



**Figure 2.9. DNA binding studies with McrB:** A) Schematic of the 60 bp double strand specific DNA. The red coloured C depict the methylated cytosine preceded by a purine. 5% native PAGE showing McrB binding to 250 nM specific DNA in presence of B) 1 mM GTP, C) 1 mM GDPNP, and D) in absence of nucleotide. 5% native PAGE showing binding of McrB+McrC (McrB:McrC::4:1) to specific DNA in presence of E) 1 mM GTP, F) 1 mM GDPNP, and G) in absence of nucleotide GDPNP, and F) 1 mM GTP. Concentration of protein used in each assay were- 0 uM, 1.5 uM, 3 uM, 6 uM, 12 uM, 30 uM of monomer.

Based on these studies, we designed a 60 bp dsDNA substrate mimic (specific DNA) with the target site 5'-A<sup>m</sup>C-3' at the center (Figure 2.9 A). Electrophoretic mobility assay (EMSA) with this oligonucleotide showed that McrB formed a complex with the specific DNA in the presence of either GTP or its non-hydrolysable analog GDPNP (Figure 2.9 B,C). Interestingly, the absence of GTP only had a marginal effect on the binding of McrB to specific DNA (Figure 2.9 D). The shift in DNA noted in this case likely arose from complexation with nucleotide-independent oligomers of McrB.

McrBC also formed a complex with the specific DNA in presence of GDPNP (Figure 2.9 F). The shift in DNA was higher than that noted in case of McrB in accordance with the larger molecular size of McrBC. In presence of GTP, however, an intermediate



**Figure 2.10. EMSA showing specific binding of McrB.** A). 5% native PAGE, showing McrB binding to 250 nM non-specific DNA in presence of 1mM GTP, B) 5% native PAGE, showing McrBΔN binding to specific DNA in presence of 1 mM GTP, C) 5% native PAGE, showing McrB binding to specific DNA in presence of 1 mM GDP, and D) 5% native PAGE, showing McrBC binding to specific DNA in presence 1 mM GDP. Concentration of protein used in each assay were- 0 uM, 1.5 uM, 3 uM, 6 uM, 12 uM, 30 uM of monomer.

shift was noted (Figure 2.9 E). It is likely that the GTPase-driven translocation by McrBC prevents the formation of a long-lived DNA complex resulting in the lower, diffused shift. In absence of nucleotide (Figure 2.9 G), the DNA binding with McrB and McrB+McrC appeared very similar, indicating that McrC might not play a significant role in protein-DNA recognition and binding. We also found that McrB $\Delta$ N, which lacks the DNA binding domain, failed to bind to specific DNA (Figure 2.10 B). These observations are in consonance with the report that the isolated N-terminal domain binds to target sites specifically (Sukackaite et al, 2012). Also, McrB did not bind to non-specific DNA in presence of GTP (Figure 2.10 A). The binding of both McrB and McrB+McrC to specific DNA was also checked in presence of GDP (Figure 2.10 C, D). The binding looked similar to that observed in presence of GDPNP showing band shift in case of McrB+McrC higher than that observed with only McrB. Thus, it appears that McrB and McrBC complex does not require GTP hydrolysis for its binding activity.

#### **2.3.3.3 DNA cleavage activity:**

A 114 bp long DNA substrate containing two target sites separated by 53 bp was used as a substrate to study the nucleolytic activity of McrBC (Figure 2.11 A). McrB cleaves DNA 26-30 bp upstream of either of the two binding sites. Thus, in the 114 bp DNA, expected fragment sizes were 60 bp and 54 bp if cleavage occurs with respect to the left hand site or 61 bp and 53 bp if cleavage occurs with respect to the right hand site. Since the 1 bp difference could not be resolved on a 10% PAGE gel, we expected two cleavage products at around 60 bp and 55 bp. In presence of GTP, this substrate was readily cleaved by a mixture of McrB and McrC (4:1 molar ratio). Neither McrB nor McrB $\Delta$ N+McrC could cleave specific DNA in presence of GTP (Figure 2.11 B). The McrBC complex purified by SEC retained its ability to cleave the substrate (Figure 2.11 C). As expected, the nucleolytic cleavage yielded two products corresponding to a dsDNA break close to either of the target sites.

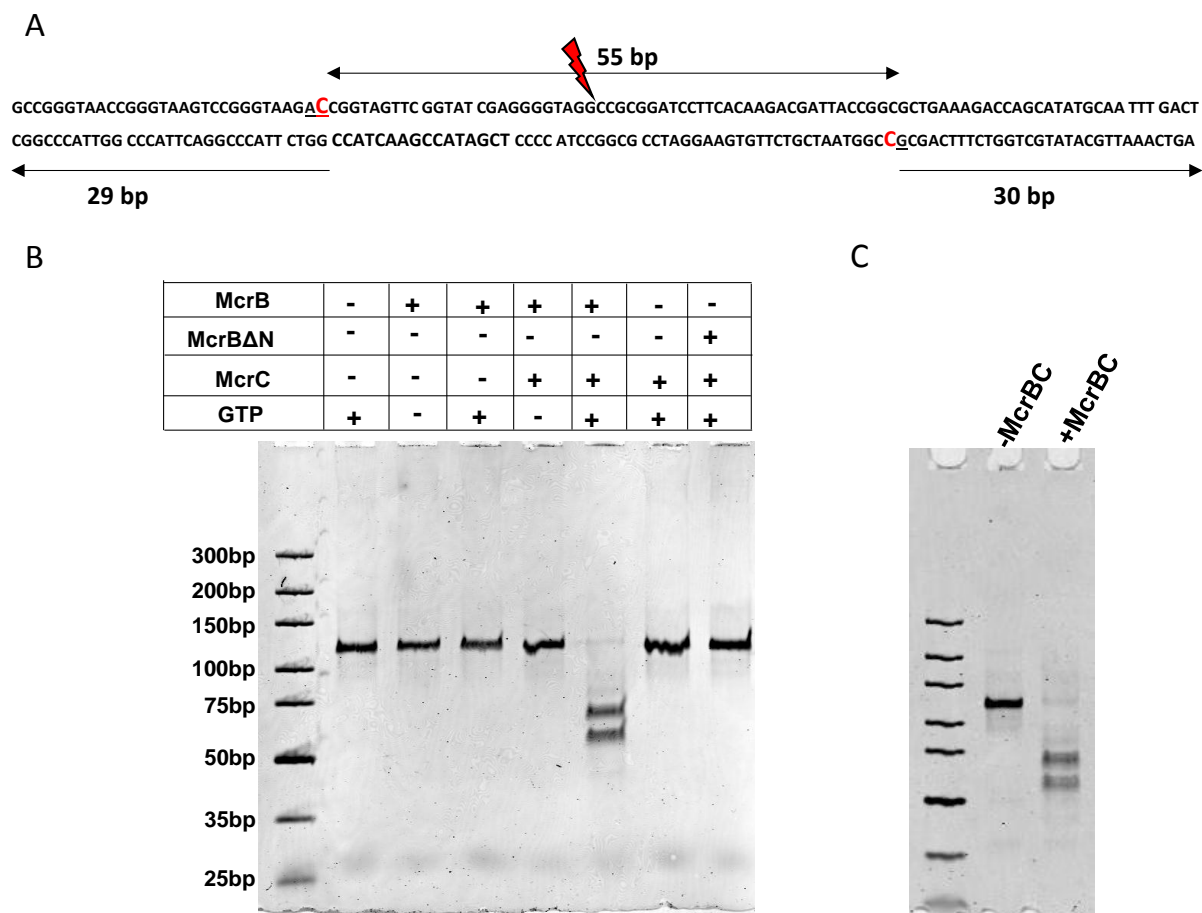


Figure 2.11. **DNA Cleavage assay:** A) Schematic representation of 114 bp substrate DNA used for DNA cleavage assay. The red coloured C depict the methylated cytosine preceded by a purine B) 10% Native PAGE gel showing DNA cleavage activity of 900 nM McrB in presence of 225 nM McrC, 75 nM substrate DNA and 1mM GTP. C) DNA cleavage activity of 75 nM McrBC complex (assembled and purified by SEC) in presence of 75nM of substrate DNA and 1mM GTP.

### 2.3.4 Oligomeric structure of McrB from Cryo-EM studies

Having found that McrB/McrB $\Delta$ N formed a homogenous hexameric oligomer in presence of GTP, we pursued structural characterization of the hexamer. We imaged samples of both McrB/McrB $\Delta$ N in presence of the non-hydrolysable analog GDPNP. Imaging of McrB and McrB $\Delta$ N after the addition of GDPNP as expected showed round particles (Figure 2.12 A, B). With present freezing conditions, particles preferred only certain orientations, predominantly showing the top/bottom views. Due to the use of direct electron detector and higher defocus the hexamers of the protein in raw images could occasionally be

visualized (Figure 2.12 A, B). Reference-free two-dimensional (2D) class averages show clearly the hexameric nature of the both the McrB and McrB $\Delta$ N (Figure 2.12 C, D). The hexamers are ring-shaped having a central pore. The outer diameter of the ring is  $\sim 95$  Å

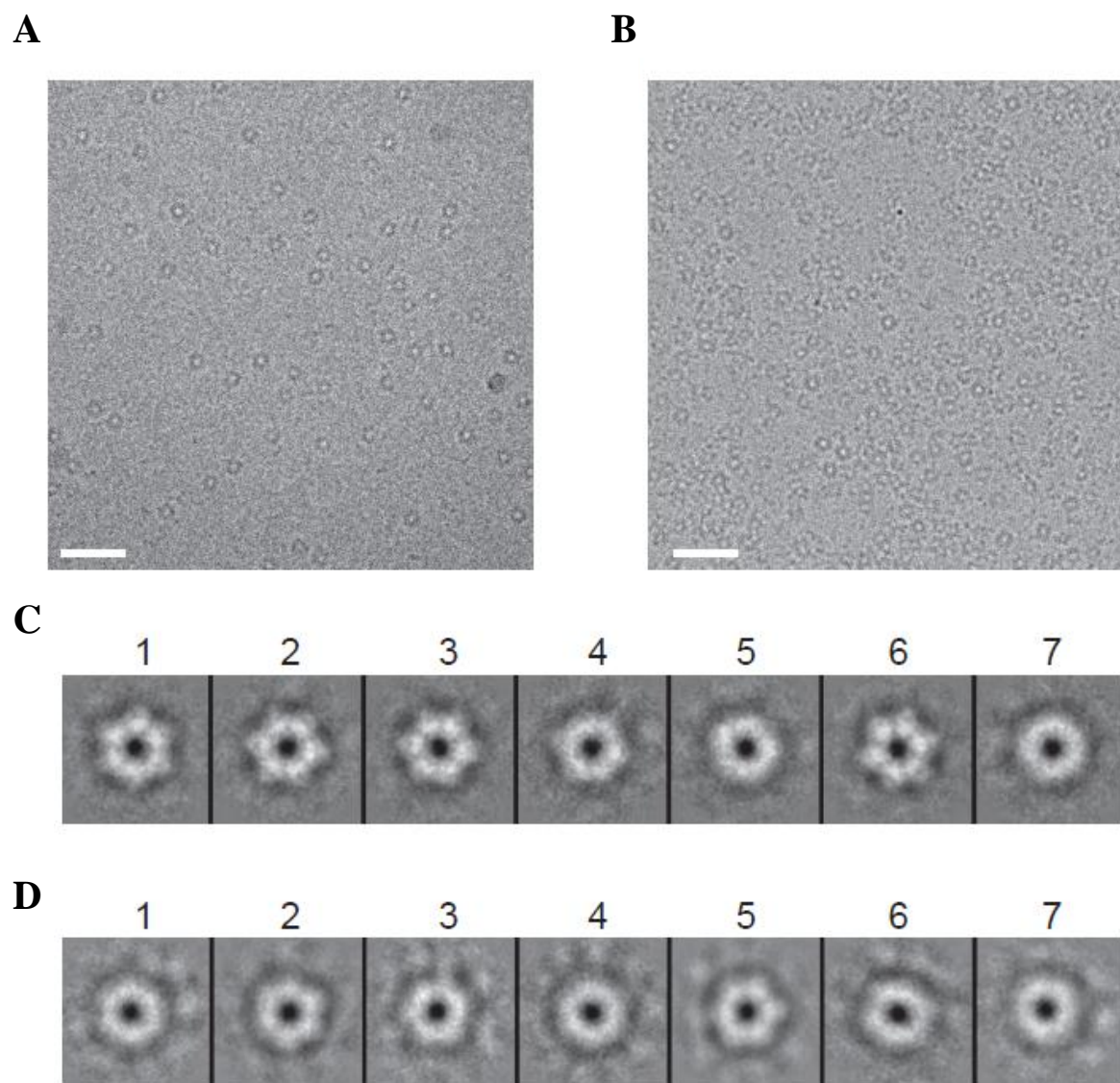


Figure 2.12. **Cryo-EM images of McrB and McrB $\Delta$ N in presence of GDPNP: A)** A small area of micrographs of McrB $\Delta$ N, and **B)** McrB showing round particles on ice. The images were captured with Falcon II CMOS detector. The scale bar is 400 Å. **C)** Reference-free 2D class averages of McrB $\Delta$ N, and **D)** McrB. The box size is 160 pixels. McrB $\Delta$ N was sampled at 1.33 Å/pixels and McrB at 1.34 Å/pixels. Approximate diameter of the particles is  $\sim 95$  Å and the central hole measures  $\sim 26$  Å.



and the central pore diameter is  $\sim 26$  Å. When the rotational power spectrum of the class averages was analyzed using the program rfilm (Crowther & Amos, 1971) a number of

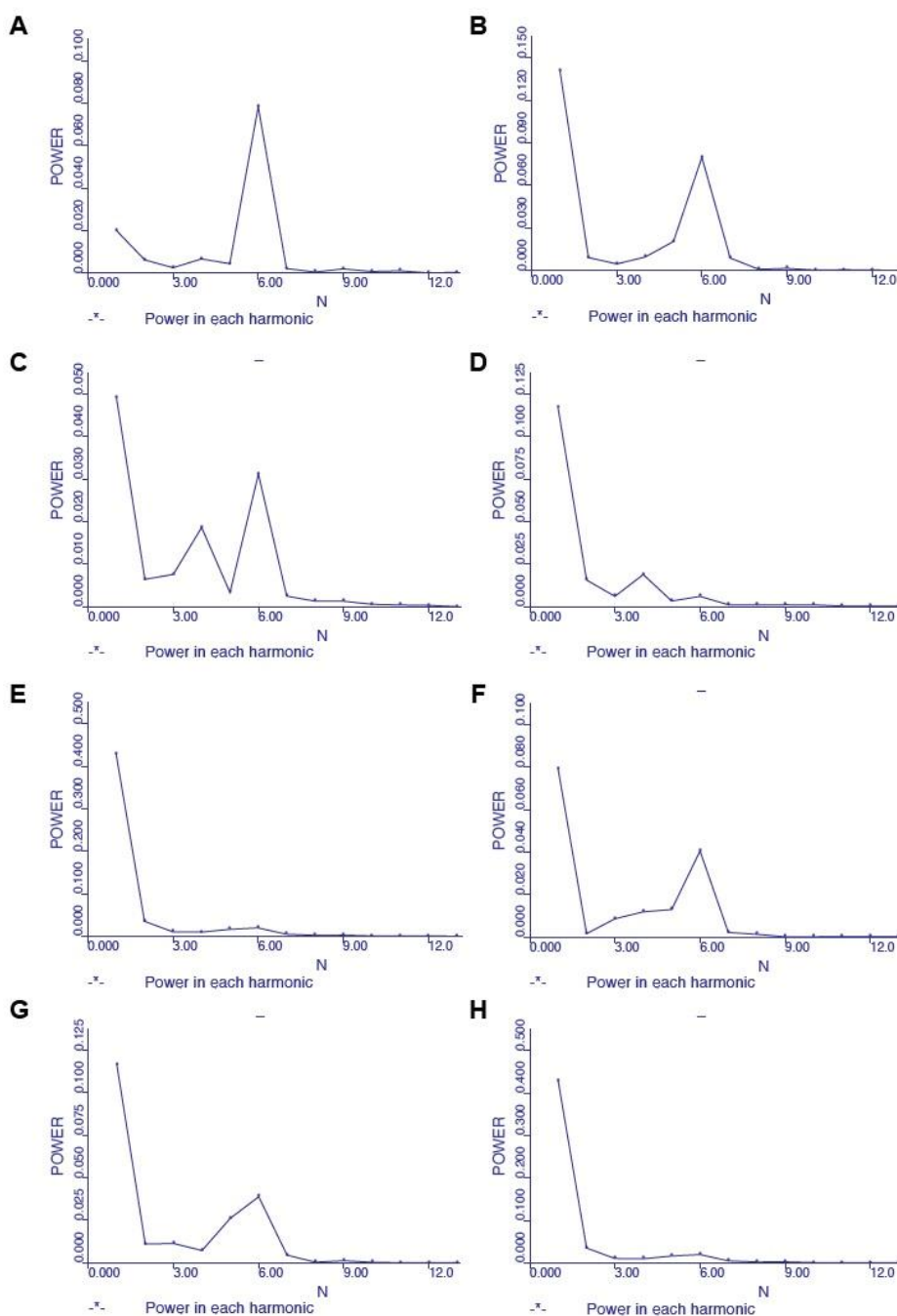


Figure 2.13. **Representative rotational power spectrum for class averages** in A) Fig. 2.12 C---1, B) Fig. 2.12 C---2, C) Fig. 2.12 C---4, D) Fig. 2.12 C---5, E) Fig. 2.12 D---1, F) Fig. 2.12 D---2, G) Fig. 2.12 D---5, H) Fig. 2.12 D---6

them showed a peak at 6- fold (for example class 1, 2 in McrB $\Delta$ N). The more round looking classes (for example class averages 4, 5 or 7 in Fig. 6C) were not as clear and the corresponding rotational power spectrum peaks ranged from 4-fold to 6-fold (Figure 2.13). However, no peaks were observed >6-fold. The SEC-MALS analysis (see above) clearly demonstrated that the GTP dependent oligomeric complex of McrB/McrB $\Delta$ N were homogenous. The more round looking classes could be the result of defocus and thickness of ice.

## **2.4 Discussion**

### **2.4.1 *Determinants for oligomerization of McrB and McrBC***

An integral feature of proteins containing AAA+ domain is their oligomerization into functional units. McrB and a variant form of the protein lacking the N-terminal DNA-binding domain, McrBs, has previously been reported to form a heptameric ring-like structure in presence of GTP, while McrBC was reported to be a complex of two heptameric rings of McrB and two subunits of McrC (Panne et al, 2001). We carried out studies on the assembly of McrB oligomer and McrBC complex, towards understanding the mechanism of this enzyme. In addition, studies with McrB $\Delta$ N, which is an equivalent of McrBs, were also carried out. SEC studies confirmed that monomeric McrB and McrB $\Delta$ N assembled into a higher-order oligomeric structure in presence of GTP, GDP or GDPNP (Figure 2.2 A, B), which could be disassembled by washing away the nucleotides. Similar observations have been made in case of other AAA+ proteins, such as SV40 large tumor antigen (LTag), ClpAP, ClpB and ClpX, where these proteins dissociate into monomers, dimers, trimers or tetramers when nucleotide is absent, but assemble into hexamers once nucleotide is added (Akoev et al, 2004; Dean et al, 1992; Grimaud et al, 1998; Singh & Maurizi, 1994). McrB also showed a concentration-dependent oligomerization in absence of nucleotide. At lower protein concentrations, McrB predominantly existed as monomers (Figure 2.3). But at a very high concentration of the protein, a mixture of intermediate-size and higher-order oligomers was observed (Figure 2.3). Concentration-dependent, nucleotide-independent oligomerization has been reported in case of other AAA+ proteins like ClpB and Rep68 (Akoev et al, 2004; Zarate-

Perez et al, 2012). Finally, a complex of McrBC was assembled from a mix of 4:1 molar ratio of McrB and McrC in presence of GTP.

#### **2.4.2 McrB forms a hexameric ring-like structure in presence of GTP**

Analysis by SEC-MALS showed clearly that in presence of nucleotide both McrB and McrB $\Delta$ N existed as highly monodisperse hexamers in solution. In presence of GTP, McrB and McrB $\Delta$ N assembled into a hexamer, while McrBC and McrB $\Delta$ NC assembled into an oligomer made of twelve subunits of McrB and two subunits of McrC. The hexameric structure of McrB and McrB $\Delta$ N was further confirmed by Cryo-EM and 2D class averages. Based on the dimensions of the toroid shaped hexameric ring obtained from the Cryo-EM studies (outer diameter  $\sim$ 95 Å and central pore diameter  $\sim$ 26 Å) and using the equation given by Huang et al. (Huang et al, 2009), the radius of gyration (Rg) was calculated to be  $\sim$  55 Å. For a solid globular structure, the Rg is expected to be smaller than the hydrodynamic radius (Rh) by a factor of 0.77, while for a toroid the Rg is bigger than Rh by  $\sim$  1.05 (Huang et al., 2009). The Rg value of McrB $\Delta$ N is 1.1 times the Rh value of  $\sim$  50 Å as measured using MALS, while for McrB it is 0.9 times bigger. Consequently, the measured Rh values in solution are consistent with the dimensions of the ring-like structures seen by Cryo-EM. In summary, we conclude that the functional unit of McrB is a hexameric ring and that the McrBC complex is possibly made of the two hexameric rings of McrB bound to two subunits of McrC. Though our results on the oligomeric structure contradict the earlier study, which found McrB and McrB $\Delta$ N to be heptameric (Panne et al, 2001), functional characterization of the proteins yielded results consistent with those reported previously by other researchers. The intrinsic GTPase activities of both McrB and McrB $\Delta$ N were very low (Figure 2.8 A, B), but the addition of McrC stimulated the GTPase activity by almost 30 fold in case of McrB (Figure 2.8 C, D). McrBC and McrB $\Delta$ NC complexes assembled and purified by SEC retained a high GTPase activity (Figure 2.8 C, D). A mix of McrB and McrC in 4:1 ratio or a purified assembly of McrBC complex in presence of GTP displayed a similar nucleolytic activity towards a substrate having two target sites (Figure 2.11). The disagreement on the oligomeric structure of McrB could be due to the limitations of the experimental techniques used in the previous study. In the previous study, a heptameric structure was deduced using SEC,

negative stain EM and mass analysis by STEM (Panne et al, 2001). The accuracy of molecular mass deduced from the retention volume of molecules obtained by SEC analysis is greatly affected by their shape; the non-native staining artefacts and the quality of images obtained by electron microscopy (EM) can contribute to errors in characterisation of the oligomeric structure; the mass analysis by STEM for a molecule in the mass range of McrB and McrBC can have error of ~10%, and is not sensitive enough to determine the number of subunits (Gribun et al, 2008). However, the techniques of SEC-MALS and Cryo-EM that maintain the native state of the oligomer, unequivocally demonstrate that both McrB and McrB $\Delta$ N are hexamers in presence of GTP.

### **2.4.3 A model of McrBC complex**

The previous study of the oligomeric structure of McrB and McrBC proposed a model for the assembly of McrBC in which, in presence of GTP, two heptameric rings of McrB stack together and a McrC monomer sits on each of the two outer surfaces of the tetradecameric McrB (Panne et al, 2001). Our studies show that both McrB and McrB $\Delta$ N can interact with McrC, and also form the higher order oligomeric structure consisting of twelve subunits of McrB and two subunits of McrC, in presence of GTP. This suggests that the N-terminal domain of McrB neither interacts with McrC nor participates in formation of the higher order McrBC structure. Based on these observations, we propose a model for McrBC assembly constituting two hexamers of McrB bridged and stabilized by their interaction with a dimer of McrC (Figure 2.14). It is tempting to speculate that the dimer of McrC would get activated upon collision of two assemblies of McrBC translocating DNA through the pore of the ring. The activated nuclease dimers of the McrBC bound to the target sequence would then nick the two strands of DNA resulting in a dsDNA break.

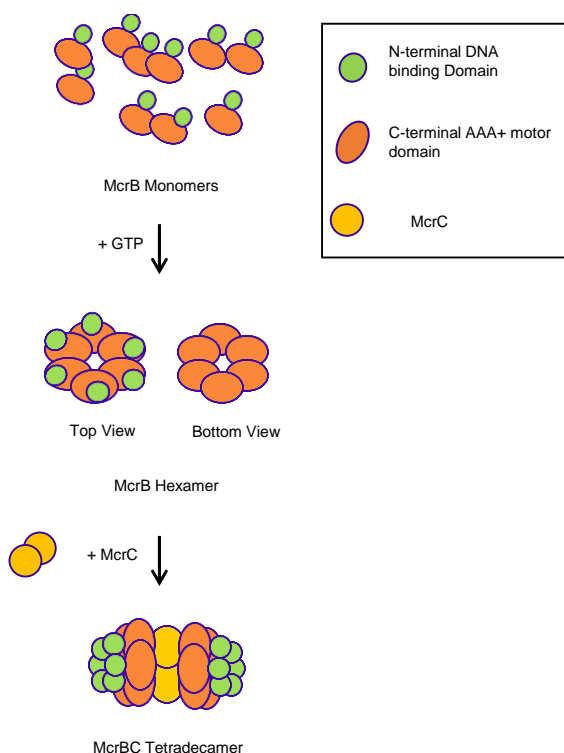


Figure 2.14. **A Model proposing assembly of functional McrBC**

#### **2.4.4 Functional implications of the hexameric ring-like structure**

The hexameric ring of McrB poses a challenge in loading the protein complex on to its substrate DNA, which includes long non-linear DNA. This, in general, is true for preformed hexameric ring helicases and translocases, such as DnaB, SV40 LTag, T7 gp4, MCM and FtsK that have to assemble on a DNA substrate (O'Shea & Berger, 2014). In the case of DnaB, DnaC opens up the ring of DnaB to load the hexamer onto the DNA (Arias-Palomo et al, 2013). T7 gp4 helicase and MCM can exist as heptamers, (Toth et al, 2003; Yu et al, 2002), which converts into a hexamer through loss of a subunit thus creating a gap for the substrate to slip in (Costa et al, 2006; Crampton et al, 2006). In case of FtsK and SV40 LTag, it has been suggested that the subunits assemble around the substrate to form a functional hexameric ring (Ben-Yehuda et al, 2003; Chang et al, 2013; Crozat & Grainge, 2010; Grimaud et al, 1998). The hexameric ring of McrB could open up on complexation with McrC. Alternatively, hexameric rings could disassemble into

monomers and other oligomeric intermediates, which could then reassemble on the DNA substrate to form the functional oligomeric complex with McrC.

To better understand this process and gain further mechanistic insights, I carried out biochemical, biophysical and structural studies of McrBC. The AAA+ motors are complex entities and McrB being a AAA+ motor provides manifold exploring opportunities. In the following chapter, I will discuss functional mutagenesis studies of McrB, carried out to gain further insights into its oligomerisation, GTPase and DNA cleavage activities.

## References

Akoev V, Gogol EP, Barnett ME, Zolkiewski M (2004) Nucleotide-induced switch in oligomerization of the AAA+ ATPase ClpB. *Protein science : a publication of the Protein Society* **13**: 567-574

Arias-Palomo E, O'Shea VL, Hood IV, Berger JM (2013) The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell* **153**: 438-448

Batchelor JD, Sterling HJ, Hong E, Williams ER, Wemmer DE (2009) Receiver domains control the active-state stoichiometry of Aquifex aeolicus sigma54 activator NtrC4, as revealed by electrospray ionization mass spectrometry. *Journal of molecular biology* **393**: 634-643

Baykov AA, Evtushenko OA, Avaeva SM (1988) A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Analytical biochemistry* **171**: 266-270

Bellare JR, Davis HT, Scriven LE, Talmon Y (1988) Controlled environment vitrification system: an improved sample preparation technique. *J Electron Microsc Tech* **10**: 87-111

Ben-Yehuda S, Rudner DZ, Losick R (2003) Assembly of the SpoIIIE DNA translocase depends on chromosome trapping in Bacillus subtilis. *Current biology : CB* **13**: 2196-2200

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72**: 248-254

Chang YP, Xu M, Machado AC, Yu XJ, Rohs R, Chen XS (2013) Mechanism of origin DNA recognition and assembly of an initiator-helicase complex by SV40 large tumor antigen. *Cell reports* **3**: 1117-1127

Chen B, Doucleff M, Wemmer DE, De Carlo S, Huang HH, Nogales E, Hoover TR, Kondrashkina E, Guo L, Nixon BT (2007) ATP ground- and transition states of bacterial enhancer binding AAA+ ATPases support complex formation with their target protein, sigma54. *Structure* **15**: 429-440

Chen B, Sysoeva TA, Chowdhury S, Guo L, De Carlo S, Hanson JA, Yang H, Nixon BT (2010) Engagement of arginine finger to ATP triggers large conformational changes in NtrC1 AAA+ ATPase for remodeling bacterial RNA polymerase. *Structure* **18**: 1420-1430

Costa A, Pape T, van Heel M, Brick P, Patwardhan A, Onesti S (2006) Structural studies of the archaeal MCM complex in different functional states. *Journal of structural biology* **156**: 210-219

Crampton DJ, Ohi M, Qimron U, Walz T, Richardson CC (2006) Oligomeric states of bacteriophage T7 gene 4 primase/helicase. *Journal of molecular biology* **360**: 667-677

Crowther RA, Amos LA (1971) Harmonic analysis of electron microscope images with rotational symmetry. *Journal of molecular biology* **60**: 123-130

Crozat E, Grainge I (2010) FtsK DNA translocase: the fast motor that knows where it's going. *Chembiochem : a European journal of chemical biology* **11**: 2232-2243

De Carlo S, Chen B, Hoover TR, Kondrashkina E, Nogales E, Nixon BT (2006) The structural basis for regulated assembly and function of the transcriptional activator NtrC. *Genes & development* **20**: 1485-1495

Dean FB, Borowiec JA, Eki T, Hurwitz J (1992) The simian virus 40 T antigen double hexamer assembles around the DNA at the replication origin. *The Journal of biological chemistry* **267**: 14129-14137

Dey S, Biswas M, Sen U, Dasgupta J (2015) Unique ATPase site architecture triggers cis-mediated synchronized ATP binding in heptameric AAA+-ATPase domain of flagellar regulatory protein FlrC. *The Journal of biological chemistry* **290**: 8734-8747

Enemark EJ, Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**: 270-275

Erzberger JP, Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annual review of biophysics and biomolecular structure* **35**: 93-114

Fouse SD, Nagarajan RO, Costello JF (2010) Genome-scale DNA methylation analysis. *Epigenomics* **2**: 105-117

Gast FU, Brinkmann T, Pieper U, Kruger T, Noyer-Weidner M, Pingoud A (1997) The recognition of methylated DNA by the GTP-dependent restriction endonuclease McrBC resides in the N-terminal domain of McrB. *Biological chemistry* **378**: 975-982

Geladopoulos TP, Sotiroudis TG, Evangelopoulos AE (1991) A malachite green colorimetric assay for protein phosphatase activity. *Analytical biochemistry* **192**: 112-116

Gribun A, Cheung KL, Huen J, Ortega J, Houry WA (2008) Yeast Rvb1 and Rvb2 are ATP-dependent DNA helicases that form a heterohexameric complex. *Journal of molecular biology* **376**: 1320-1333

Grimaud R, Kessel M, Beuron F, Steven AC, Maurizi MR (1998) Enzymatic and structural similarities between the Escherichia coli ATP-dependent proteases, ClpXP and ClpAP. *The Journal of biological chemistry* **273**: 12476-12481

Hastings CA, Lee SY, Cho HS, Yan D, Kustu S, Wemmer DE (2003) High-resolution solution structure of the beryllofluoride-activated NtrC receiver domain. *Biochemistry* **42**: 9081-9090

Huang H, Chung B, Jung J, Park HW, Chang T (2009) Toroidal micelles of uniform size from diblock copolymers. *Angew Chem Int Ed Engl* **48**: 4594-4597

Iyer LM, Makarova KS, Koonin EV, Aravind L (2004) Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic acids research* **32**: 5260-5279

Joly N, Zhang N, Buck M (2012) ATPase site architecture is required for self-assembly and remodeling activity of a hexameric AAA+ transcriptional activator. *Molecular cell* **47**: 484-490

Kim KI, Cheong GW, Park SC, Ha JS, Woo KM, Choi SJ, Chung CH (2000) Heptameric ring structure of the heat-shock protein ClpB, a protein-activated ATPase in Escherichia coli. *Journal of molecular biology* **303**: 655-666



Kruger T, Wild C, Noyer-Weidner M (1995) McrB: a prokaryotic protein specifically recognizing DNA containing modified cytosine residues. *The EMBO journal* **14**: 2661-2669

Lee SY, De La Torre A, Yan D, Kustu S, Nixon BT, Wemmer DE (2003) Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. *Genes & development* **17**: 2552-2563

Loenen WA, Raleigh EA (2014) The other face of restriction: modification-dependent enzymes. *Nucleic acids research* **42**: 56-69

Miroux B, Walker JE (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of molecular biology* **260**: 289-298

Miyata T, Yamada K, Iwasaki H, Shinagawa H, Morikawa K, Mayanagi K (2000) Two different oligomeric states of the RuvB branch migration motor protein as revealed by electron microscopy. *Journal of structural biology* **131**: 83-89

O'Shea VL, Berger JM (2014) Loading strategies of ring-shaped nucleic acid translocases and helicases. *Current opinion in structural biology* **25**: 16-24

Panne D, Muller SA, Wirtz S, Engel A, Bickle TA (2001) The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *The EMBO journal* **20**: 3210-3217

Panne D, Raleigh EA, Bickle TA (1999) The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. *Journal of molecular biology* **290**: 49-60

Pape T, Meka H, Chen S, Vicentini G, van Heel M, Onesti S (2003) Hexameric ring structure of the full-length archaeal MCM protein complex. *EMBO reports* **4**: 1079-1083

Pieper U, Brinkmann T, Kruger T, Noyer-Weidner M, Pingoud A (1997) Characterization of the interaction between the restriction endonuclease McrBC from E. coli and its cofactor GTP. *Journal of molecular biology* **272**: 190-199

Pieper U, Groll DH, Wunsch S, Gast FU, Speck C, Mucke N, Pingoud A (2002) The GTP-dependent restriction enzyme McrBC from Escherichia coli forms high-molecular mass complexes with DNA and produces a cleavage pattern with a characteristic 10-base pair repeat. *Biochemistry* **41**: 5245-5254

Pieper U, Pingoud A (2002) A mutational analysis of the PD...D/EXK motif suggests that McrC harbors the catalytic center for DNA cleavage by the GTP-dependent restriction enzyme McrBC from *Escherichia coli*. *Biochemistry* **41**: 5236-5244

Pieper U, Schweitzer T, Groll DH, Gast FU, Pingoud A (1999a) The GTP-binding domain of McrB: more than just a variation on a common theme? *Journal of molecular biology* **292**: 547-556

Pieper U, Schweitzer T, Groll DH, Pingoud A (1999b) Defining the location and function of domains of McrB by deletion mutagenesis. *Biological chemistry* **380**: 1225-1230

Raleigh EA (1992) Organization and function of the mcrBC genes of *Escherichia coli* K-12. *Molecular microbiology* **6**: 1079-1086

Rappas M, Schumacher J, Beuron F, Niwa H, Bordes P, Wigneshweraraj S, Keetch CA, Robinson CV, Buck M, Zhang X (2005) Structural insights into the activity of enhancer-binding proteins. *Science* **307**: 1972-1975

Schumacher J, Zhang X, Jones S, Bordes P, Buck M (2004) ATP-dependent transcriptional activation by bacterial PspF AAA+protein. *Journal of molecular biology* **338**: 863-875

Singh SK, Maurizi MR (1994) Mutational analysis demonstrates different functional roles for the two ATP-binding sites in ClpAP protease from *Escherichia coli*. *The Journal of biological chemistry* **269**: 29537-29545

Singleton MR, Dillingham MS, Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annual review of biochemistry* **76**: 23-50

Smith RH, Spano AJ, Kotin RM (1997) The Rep78 gene product of adeno-associated virus (AAV) self-associates to form a hexameric complex in the presence of AAV ori sequences. *Journal of virology* **71**: 4461-4471

Stewart FJ, Panne D, Bickle TA, Raleigh EA (2000) Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *Journal of molecular biology* **298**: 611-622

Stewart FJ, Raleigh EA (1998) Dependence of McrBC cleavage on distance between recognition elements. *Biological chemistry* **379**: 611-616

Sukackaite R, Grazulis S, Tamulaitis G, Siksnyš V (2012) The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. *Nucleic acids research* **40**: 7552-7562

Sutherland E, Coe L, Raleigh EA (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *Journal of molecular biology* **225**: 327-348

Sysoeva TA, Chowdhury S, Guo L, Nixon BT (2013) Nucleotide-induced asymmetry within ATPase activator ring drives sigma54-RNAP interaction and ATP hydrolysis. *Genes & development* **27**: 2500-2511

Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ (2007) EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology* **157**: 38-46

Toth EA, Li Y, Sawaya MR, Cheng Y, Ellenberger T (2003) The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Molecular cell* **12**: 1113-1123

van den Ent F, Lowe J (2006) RF cloning: a restriction-free method for inserting target genes into plasmids. *Journal of biochemical and biophysical methods* **67**: 67-74

Yu X, VanLoock MS, Poplawski A, Kelman Z, Xiang T, Tye BK, Egelman EH (2002) The Methanobacterium thermoautotrophicum MCM protein can form heptameric rings. *EMBO reports* **3**: 792-797

Zarate-Perez F, Bardelli M, Burgner JW, 2nd, Villamil-Jarauta M, Das K, Kekilli D, Mansilla-Soto J, Linden RM, Escalante CR (2012) The interdomain linker of AAV-2 Rep68 is an integral part of its oligomerization domain: role of a conserved SF3 helicase residue in oligomerization. *PLoS pathogens* **8**: e1002764

Zhang N, Gordiyenko Y, Joly N, Lawton E, Robinson CV, Buck M (2014) Subunit dynamics and nucleotide-dependent asymmetry of an AAA(+) transcription complex. *Journal of molecular biology* **426**: 71-83

## Chapter 3

### Mutational analysis of the AAA+ McrB protein

#### 3.1 Introduction

In the previous chapter, I had discussed the purification and assembly of the McrBC complex. Contrary to previous studies, I observed that the AAA+ domain harboring McrB subunit forms a hexameric ring and two such rings are bridged together with McrC dimer. Since both McrB and N-terminal DNA binding domain deletion mutant McrB $\Delta$ N interact in a similar manner with McrC, it appears that the N-terminal DNA binding domain might lie on the side opposite to the McrC interacting side of the ring (model proposed in figure 2.14). To gain insights into working of oligomeric McrB and McrBC, I carried out mutagenesis studies. There have been a few functional studies of McrB and McrC reported earlier (Pieper et al, 1997; Pieper & Pingoud, 2002; Pieper et al, 1999a; Pieper et al, 1999b). Pieper et al studied McrB as a possible canonical GTPase or G-protein, exploring the characteristic GTP binding motifs - GxxxxGK(S/T), DxxG and NKxD – by mutagenesis (Pieper et al, 1997). They also mutated several charged residues identified by sequence alignment with close homologs (Pieper et al, 1999a). At this point, I would like to highlight that at that time McrB was yet not classified as a AAA+ protein.

With the knowledge that McrB is a AAA+ protein, I initiated a structure-function analysis through mutagenesis of certain regions, which I predicted to be functionally important. As described in Chapter 1, AAA+ proteins harbor five characteristic motifs- Walker A (GxxGxGK(T/S)), Walker B (hhhhDE), Sensor 1 (S/T), arginine finger (R) and Sensor 2 (R) (Erzberger & Berger, 2006). Walker A motif is involved in nucleotide binding while Walker B motif is essential for nucleotide hydrolysis but not binding (Wendler et al, 2012). Pieper et al mutated the proline of the Walker A motif of McrB and found it to affect nucleotide binding. However, mutation of Walker B aspartate and glutamate also affected nucleotide binding. In comparison to most other AAA+ studied (Wendler et al, 2012), this feature of Walker B mutation appeared unusual to us. Consequently, I performed mutational analysis of Walker B residues of McrB and studied its affect.

AAA+ family of proteins, in general, have very low sequence similarity amongst its various members (Neuwald et al, 1999). Thus, in a sequence alignment with diverse proteins, some specific features might not be noticed. Iyer et al published an extensive sequence analysis of AAA+ proteins in 2004 (Iyer et al, 2004), which allowed them to classify the AAA+ family of proteins. They classified McrB as belonging to Helix-2 insert clade of AAA+ proteins. NtrC family of bacterial enhancer binding proteins (bEBPs) is the other member of the clade (Iyer et al, 2004). In NtrC family of proteins, the helix-2 insert forms a loop projecting into the oligomeric ring of the protein (Lee et al, 2003). This loop interacts with  $\sigma^{54}$  for transcription initiation (Dago et al, 2007; Zhang et al, 2009). Interestingly, several examples of AAA+ family of helicases and translocases have been shown to or proposed to thread DNA through the pore of the ring (O'Shea & Berger, 2014). McrB, which is proposed to be a DNA translocase (Panne et al, 1999), might also thread DNA through its central pore. I employed mutation of the helix-2 insert to find if this region is involved in DNA translocation.

In addition, I carried out mutational studies to find the location of Sensor 2 motif, a characteristic feature of Helix-2 insert clade. The Sensor 2 motif in McrB was not identified in the sequence alignment carried out by Iyer et al, 2000. This might either be because of low sequence similarity amongst sequences used for alignment or that McrB does not have a canonical Sensor 2, like the SF3 helicases (Clade IV of AAA+ superfamily) where this motif is identified based on its spatial location rather than through sequence alignment. In the absence of structural information, I resorted to exploring the possible Sensor 2 through a sequence alignment of McrB with a closely associated family (NtrC1) and mutagenesis study of predicted Sensor 2.

This chapter describes my efforts in identifying and characterizing these functional regions and discusses the results obtained.

## **3.2 Materials and method**

### ***3.2.1 Multiple Sequence alignment***

Protein sequences were obtained from the non-redundant (NR) protein sequence database (National Center for Biotechnology Information, NIH). Multiple sequence alignment was performed using Clustal X (Thompson et al, 1997). The alignment was further refined and analyzed manually in Jalview V2 (Waterhouse et al, 2009), a multiple sequence alignment editor. The sequence alignment from Iyer et al (Iyer et al,

2004) and Bush and Dixon (Bush & Dixon, 2012) were used as a guide to identify probable pore loop and Sensor 2 motif.

### 3.2.2 Mutagenesis and cloning

All mutations were performed using restriction-free cloning method (see Chapter 2). Table 3.1 lists the sequence of the primers used to introduce the mutations. These PCR amplified fragments were used as primers in a second PCR reaction and a plasmid containing McrB wild-type gene (pHISMcrB) was used as a template. The cloning was performed as described in Chapter 2 (Section 2.2.1.2). All the genes were sequenced to ensure only the desired mutations were incorporated.

Table 3.1: List of primers for generating McrB mutants

Sr No	Clone	Forward primer Sequence (5' -> 3')	Reverse primer Sequence (5' -> 3')
1	McrBLoopA	CCGAATGGCGTCGGCTTCG CAGCTGCAGACGGC	GATGATGGGATCCCGATG AGTCCCC
2	McrBK255A	CCGAATGGCGTCGGCTTCC GACGTGCAGACGGC	GATGATGGGATCCCGATG AGTCCCC
3	McrBD279A	GCCAGAGAAAAAGTATATT TTTATTATAGCTGAAATCAA TCGTGCC	GATGATGGGATCCCGATG AGTCCCC
4	McrBD279N	GCCAGAGAAAAAGTATATT TTTATTATAAATGAAATCAAT CGTGCC	GATGATGGGATCCCGATG AGTCCCC
5	McrBD280A	GCCAGAGAAAAAGTATATT TTTATTATAGATGCAATCAAT CGTGCC	GATGATGGGATCCCGATG AGTCCCC
6	McrBD280Q	GCCAGAGAAAAAGTATATT TTTATTATAGATCAAATCAAT CGTGCC	GATGATGGGATCCCGATG AGTCCCC
7	McrBR404A	CTTTAAGAAGGAGATATAC ATATGGAATCTATTCAACC CTGGATTG	CTATGCCCAATGGCGAAT CCTTTCCCAAGGATAGTG GCC

### **3.2.3 Protein Purification**

All mutant proteins were expressed and purified using a protocol similar to that followed for wild-type McrB (see Chapter 2; Section 2.2.2). In general, mutant proteins were found to be less stable in comparison to the wild-type. As mentioned in the previous chapter, both McrB and McrC proteins precipitate over the course of purification. Similar observations were also made by others (Kruger et al, 1995; Sutherland et al, 1992). Since the N-terminal DNA binding domain deletion mutant McrB $\Delta$ N was much better behaved, corresponding mutants of McrB $\Delta$ N were also generated.

### **3.2.4 Size Exclusion Chromatography (SEC)**

SEC and SEC coupled multi-angle light scattering (SEC-MALS) runs were done using the 24 ml Superdex200 10/300 GL (GE Healthcare Life Sciences). The protocol is described in Chapter 2.

### **3.2.5 Functional Characterisation**

GTPase assay, DNA binding assay, and DNA cleavage assays were done with a method similar to that of the wild-type protein. For details refer to Chapter 2; Section 2.2.5, 2.2.6, 2.2.7.

### **3.2.6 Dynamic Light Scattering (DLS)**

Dynamic light scattering experiments were carried out in the DynaPro NanoStar (Wyatt Technologies) DLS instrument in a 1  $\mu$ l quartz cuvette at 25 °C. For each experiment, the instrument was first tested with filtered degassed buffer with and without nucleotide. Protein samples (25  $\mu$ l) at 5  $\mu$ M concentration (prepared in 100 mM NaCl, 10 mM Tris pH and 1 mM DTT) were incubated at room temperature with and without nucleotide. In case of experiments with nucleotides, nucleotide concentration was 1 mM and 5 mM MgCl<sub>2</sub> was also added. Incubation was followed by centrifugation at 21,000 g. The supernatant was then filtered through a 0.22  $\mu$ m centrifuge tube filter (Corning® Costar® Spin-X®, Sigma-Aldrich). 60 acquisitions per measurement were recorded and each acquisition was average of 10 readings. The DYNAMICS (Wyatt technologies) software was used to calculate cumulants and regularisation analysis data.

### **3.2.7 Nucleotide binding assay using Mant-GDP**

The fluorescent nucleotide analog 2'/3'-O-(N-Methyl-anthraniloyl)-guanosine-5'-diphosphate (Mant-GDP) was obtained from Jena bioscience. Fluorescence emission spectra were recorded on Horiba FluoroMax® 4 spectrophotometer (Jobin Yvon) at 25°C in a 10X10 mm quartz cuvette size. The excitation wavelength was set at 360 nm and single point intensities were measured at 440 nm ( $I_{440}$ ). For the fluorescence measurements, 0.5  $\mu$ M protein in buffer (50 mM KCl, 10 mM Tris-Cl pH 8, 5 mM  $MgCl_2$  and 1 mM DTT) was added to the cuvette and a blank spectrum was taken. Mant-GDP was added to the protein gradually with increasing concentration steps and  $I_{440}$  were recorded for each concentration. Before each measurement, the protein was incubated with Mant-GDP for one minute prior to collection of the spectra.  $I_{440}$  readings of Mant-GDP without protein were taken as control. The difference of  $I_{440}$  in presence and absence of protein were plotted against Mant-GDP concentration. A single site binding model was fit to the data using GraphPad Prism 5 (GraphPad Software, Inc, San Diego, CA) and apparent  $K_d$  of Mant-GDP for McrB and McrBE280 mutants were calculated.

### **3.2.8 Circular-dichroism spectroscopy**

Circular-dichroism spectroscopy experiments were carried out as described in Greenfield, 2006 (Greenfield, 2006). The circular-dichroism spectra were measured on a JASCO apparatus in a 1 mm optical path cuvette for a wavelength range of 185 nm to 260 nm. Since Tris buffer gives a higher background signal in CD measurements, 1.75  $\mu$ M protein solution (wild-type and mutants) was prepared in 10 mM potassium phosphate pH 8 and 100 mM potassium. The McrB mutants were not stable and precipitated during the thawing process, thus, before the start of the experiment, protein concentrations were re-estimated with Bradford reagent using BSA as a standard.



### 3.3 Results

#### 3.3.1 Multiple Sequence alignment

Aligning only the C-terminal AAA+ domain sequence of McrB (McrB 161-460 in figure 3.1) with several bacterial enhancer binding proteins indicated two probable functional regions which have not been characterized yet.

##### A. The Pore loop:

The region corresponding to amino acid position 249-257 (GVGFRRKDG) of McrB was found to be a loop using secondary structure prediction using PSIPRED (Buchan et al, 2013) (Figure 3.2 A). Comparison of McrB $\Delta$ N model (generated from threading the amino acid sequence in THREADER (Jones et al, 1999) and an NtrC1 structure (PDB ID: 1NY5) showed that this region is spatially located in the AAA+ helix-2 insert region of the NtrC family of bacterial enhancer binding proteins (bEBPs) (Figure 3.2 B). In NtrC1, this loop, called the pore loop 1, is conserved and has the sequence motif GAFTGA. This loop has been shown to be involved in target interaction (Chen et al, 2010; Lee et al, 2003). In analogy with other AAA+ DNA helicases, the pore loop of McrB might be involved in DNA binding and its translocation through the central pore. The  $\beta$ -hairpin loop in Simian Virus 40 Large T- Antigen (SV40 LTag) helicase (NLEKKHLNKR) and in papillomavirus E1 protein (PVSIDRKHKAAVQIK) interact with DNA through the positively charged residues (Gai et al, 2004b; Sanders et al, 2007). The pore loop of McrB has three basic residues- RRK. Conservation of these possible pore loop residues was checked amongst McrB homologs from different species (<60 % sequence identity) through multiple sequence alignment (Figure 3.2 C). In the alignment, out of the three positively charged residues of proposed pore loop, the third residue, i.e. lysine, was (McrBK255A).

##### B. Sensor 2 Motif :

The McrB alignment with the bEBP family proteins showed that there is an arginine R404 in the region aligning with the Sensor 2 motif of bEBP family. An alignment of *E. coli* McrB with its homologs from different species (<60 % sequence identity) showed that the arginine identified through the alignment with bEBPs was not very well conserved amongst the McrB homologs (Figure 3.2). Instead, there is a conserved sequence IGHS in the vicinity. Nevertheless, mutational analysis of R404 was carried out to confirm if it has a functional role.

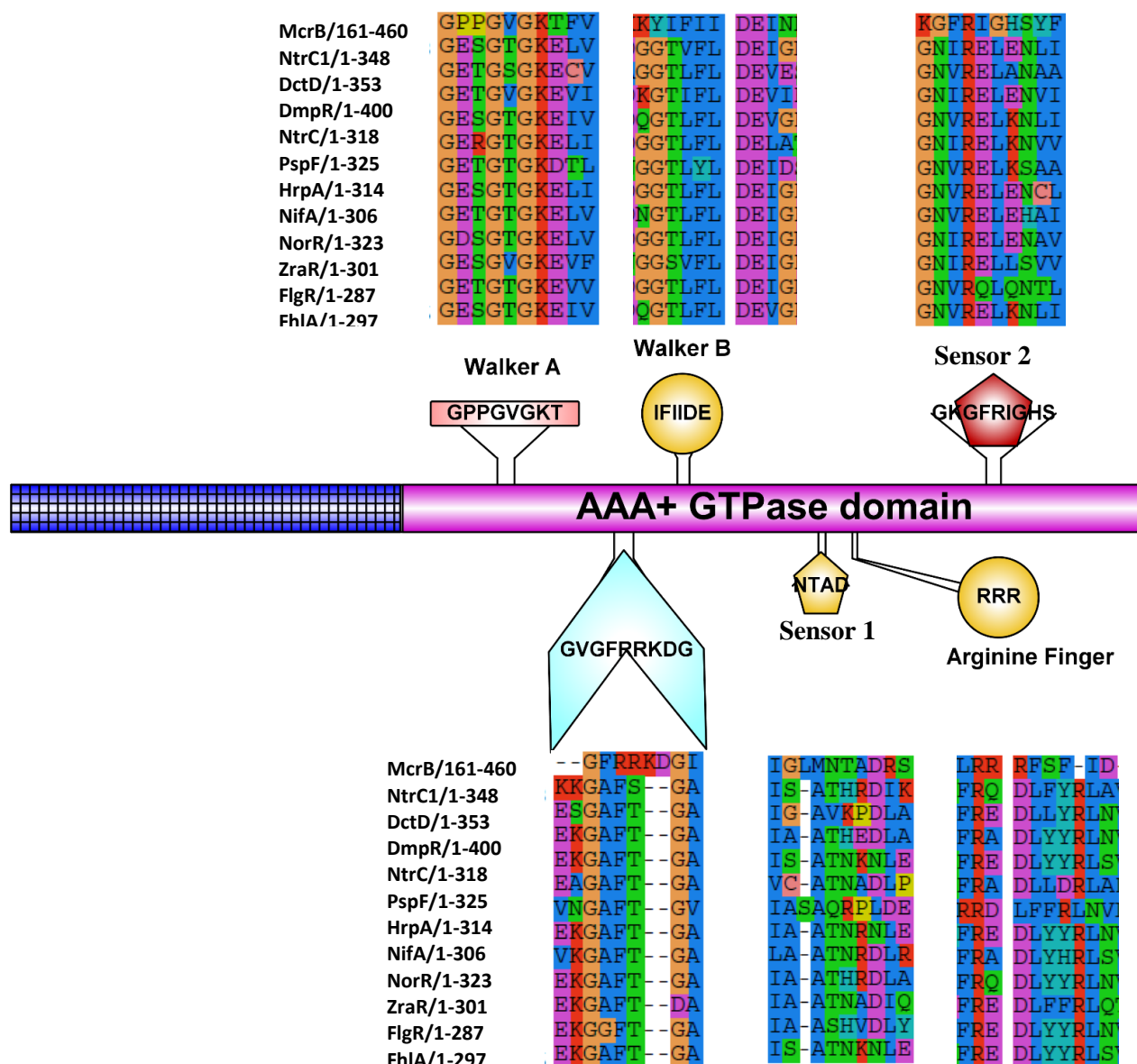


Figure 3.1. **Alignment of McrB with bEBP proteins.** Domain map of McrB and sequence alignment of C-terminal amino acids of McrB with the AAA+ domain of several bEBP to identify AAA+ motifs in McrB. The position of conserved motifs are based on previously reported sequence and structural studies on bEBPs (Bush & Dixon, 2012; Schumacher et al, 2006). The Walker A, Walker B and Sensor 1 were identified in previous studies on McrB (Iyer et al, 2004). This alignment highlighted the insertion in McrB which aligns with the pore loop (GAFTGA) of bEBPs and an arginine aligning with the Sensor 2 of bEBPs. Alignments were done using Clustal X2 (Larkin et al, 2007). Following sequences were used from UniProt/Swiss-Prot as described in (Bush & Dixon, 2012)- McrB (*E. coli*), NtrC1 (*A. aeolicus*), DctD (*S. meliloti*), DmpR (*Pseudomonas sp.*), NtrC (*E. coli*), PspF (*E. coli*), HrpR (*P. syringae*), NifA (*A. vinelandii*), NorR (*E. coli*), ZraR (*E. coli*), FlgR (*H. pylori*), XylR (*P. putida*), FhIA (*E. coli*) and NtrC4 (*A. aeolicus*).

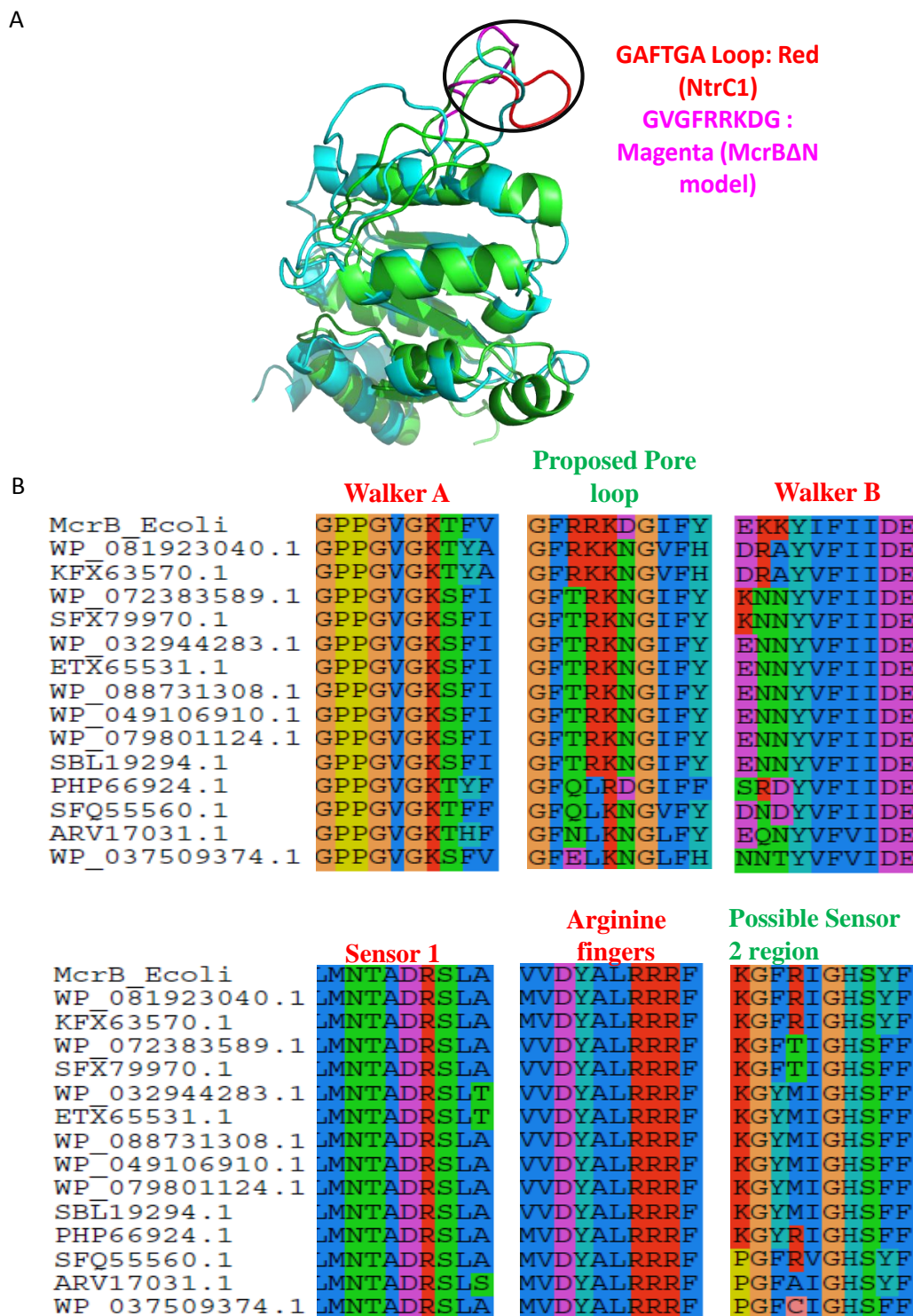


Figure 3.2. **Structural and sequence alignment of McrB to identify characteristic AAA+ motifs.** A. Structural alignment of McrB $\Delta$ N model (obtained from threading the amino acid sequence in THREADER (Jones et al, 1999) with the AAA+ domain of NtrC1 (PDB ID: 1NY5). B. Multiple sequence alignment of McrB (*E. coli*) with its homologs (Obtained from a blast search at NCBI server, sequence identity is <60%) using Clustal X2 (Larkin et al, 2007). The previously identified motifs are labelled in red while the proposed putative motifs explored in this work are labelled in green.

### 3.3.2 Cloning, mutagenesis, and purification of different mutants

The mutations were confirmed by gene sequencing and proteins were purified to 99% homogeneity. Some of the mutants like McrBD279A, McrBLoopA were more unstable and precipitated upon storage at 4°C even for a short time. These two mutants were purified using Ni-NTA and MonoQ columns, while the third step of SEC was avoided. The purified mutants were next tested for their DNA cleavage and oligomerization property in presence of nucleotide.

### 3.3.3 Characterisation of Walker B mutants

#### Walker B Aspartate:

The Walker B motif in McrB has two acidic residues- an aspartate and a glutamate (hhhhDE). Mutation in the aspartate residue to an alanine or an asparagine (D279A and D279N) rendered the protein incapable of forming oligomers in presence of nucleotide (Figure 3.3 A and B).

A problem with the Walker B aspartate mutants was their relatively low stability. Since size exclusion chromatography takes quite long and the proteins tend to precipitate over time, I resorted to re-confirm the SEC observations by a quicker method of size estimation. Dynamic light scattering (DLS) was measured at low protein concentrations to observe the change in hydrodynamic radius as a measure of oligomerization when GTP was added to the proteins (Figure 3.4).

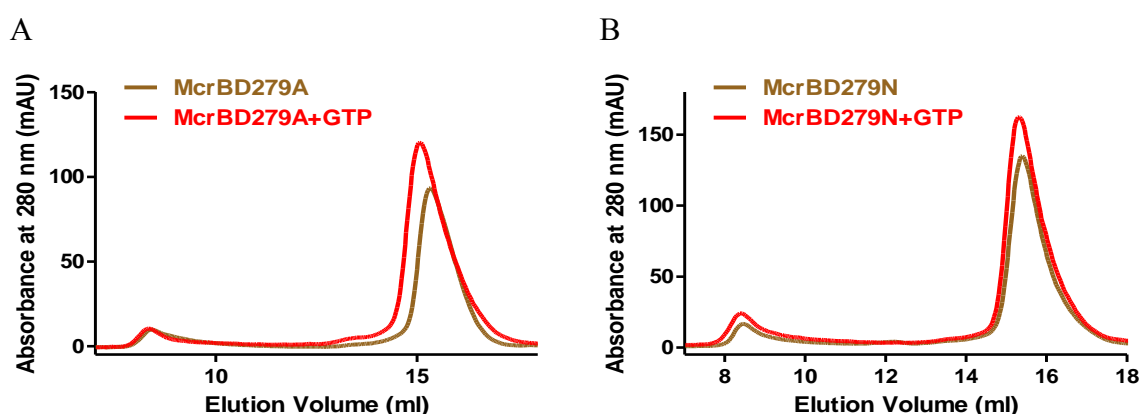


Figure 3.3. **Size exclusion chromatography with the McrB Walker B aspartate mutants.** Gel filtration profile (using 24 ml Superdex200 10/300 GL), both in presence and absence of GTP of A) 18  $\mu$ M McrBD279A and B) 18  $\mu$ M McrBD279N

Table 3.2. The Radius of hydration of McrB and its mutants estimated from DLS study

		-McrC		+McrC	
		-GTP	+GTP	-GTP	+GTP
McrB	Radius (nm)	4.73	9.858	4.982	13.774
	%Polydispersity	21.9	9.9	28.4	8.8
McrB $\Delta$ N	Radius (nm)	2.854	7.244	3.559	8.645
	%Polydispersity	25.5	12.3	11.5	17.9
McrBD279A	Radius (nm)	4.664	3.745	3.745	11.64
	%Polydispersity	24	26.7	26.7	12.4
McrBD279N	Radius (nm)	4.245	3.568	4.434	9.308
	%Polydispersity	31.8	33	12.2	12.4
McrBE280A	Radius (nm)	5.133	9.01	3.831	13.25
	%Polydispersity	15.4	15.3	28.3	3.4
McrBE280Q	Radius (nm)	2.811	6.518	3.995	8.962
	%Polydispersity	27.2	27.9	8.5	25.6

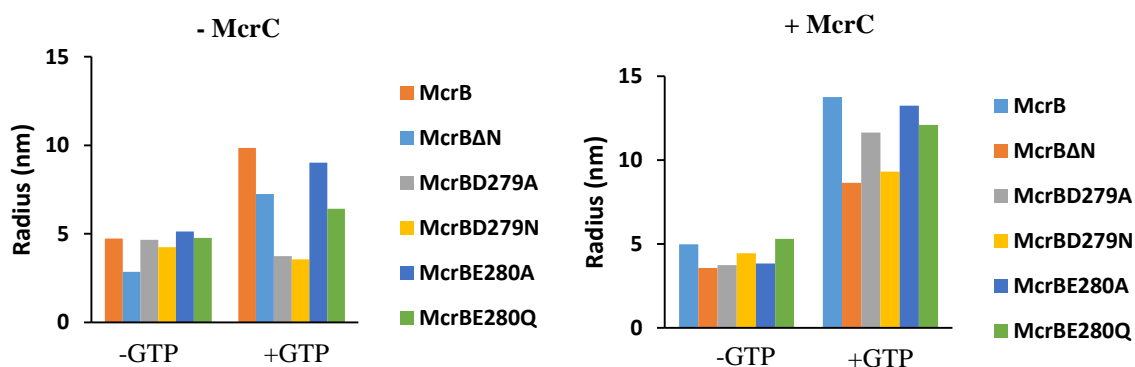


Figure 3.4. **Dynamic Light Scattering assay to assess the oligomerization state.** Radius of hydration determined by dynamic light scattering is shown for 5  $\mu$ M McrB and its mutants at 25 $^{\circ}$ C in presence and absence of 1 mM GTP. The concentration of McrC used was 1.2  $\mu$ M (McrB (and mutants): McrC::1:4).

The wild-type protein McrB was used as a control for the DLS study. Since the shape of the molecule affects the directional translation and scattering properties (Huang et al, 2009), the molar mass estimation might not be accurate. Thus the oligomerization of different proteins was monitored by qualitatively observing the change in the radius of hydration ( $R_h$ ) both in presence and absence of nucleotide. For both McrB and McrB $\Delta$ N, the  $R_h$  increased when nucleotide was added from 4.7 nm

(McrB) and 2.8 nm (McrB $\Delta$ N ) to 9.8 nm (McrB+GTP) and 7.8 nm (McrB $\Delta$ N+GTP). Further increase in  $R_h$  was observed when McrC was added in presence of GTP (13.7 nm for McrBC+GTP and 8.6 nm for McrB $\Delta$ NC+GTP) (Table 3.2). The data obtained from DLS study was highly monodisperse in presence of nucleotide (10% polydispersity index) than in absence of it (20-25 % polydispersity index) indicating that the population of a given mass are relatively more homogenous in presence of nucleotide than in absence of it. The Walker B aspartate mutants- McrBD279A and McrBD279N showed no increase in  $R_h$  upon GTP addition. An increase in  $R_h$  was observed in case of McrBD279N when both GTP and McrC were added (Figure 3.4).

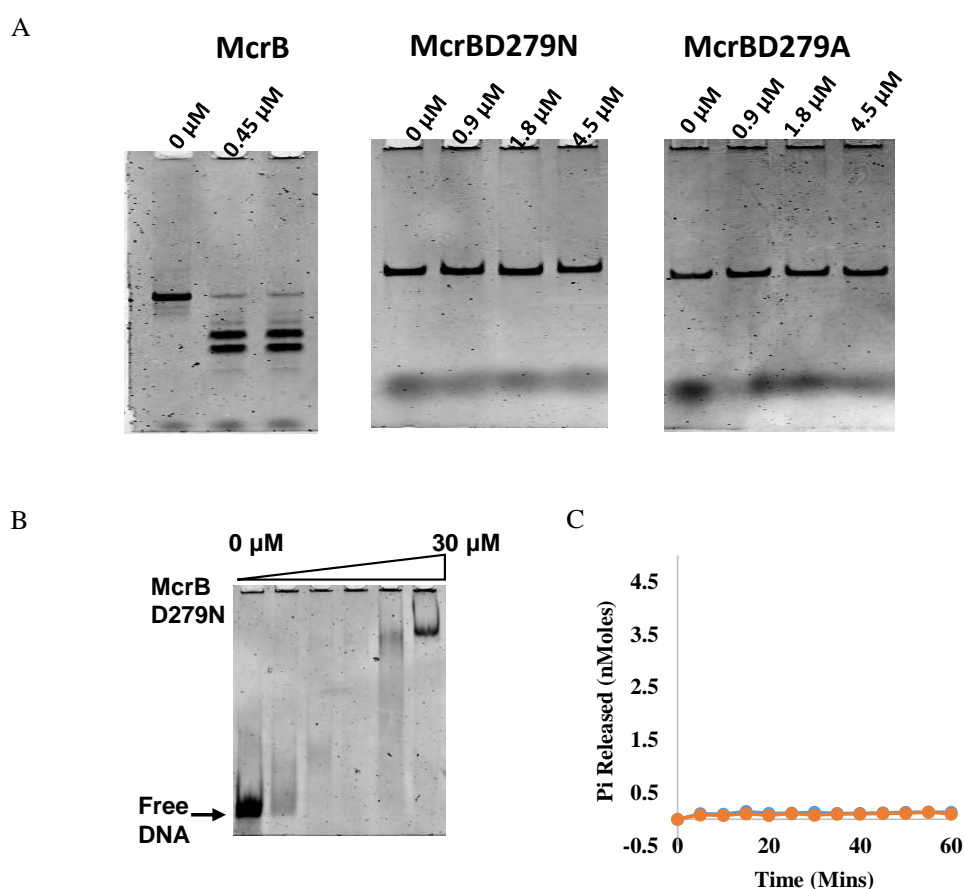


Figure 3.5. **Biochemical analysis of Walker B aspartate mutants.** A) 10 % native PAGE gel showing DNA cleavage activity of McrBD279A and McrBD279N. The Concentration of McrB or mutant is mentioned on each lane. The concentration of McrC was 1/4<sup>th</sup> of the concentration of McrB and its mutants. The reactions were performed with 75 nM 114 bp DNA substrate and 1 mM GTP at 37°C. B) 5% native PAGE gel showing McrBD279N binding to 250 nM specific DNA (described in chapter in presence of 1 mM GTP. Concentrations of McrBD279N used in each lane were- 0  $\mu$ M, 1.5  $\mu$ M, 3  $\mu$ M, 6  $\mu$ M, 12  $\mu$ M, 30  $\mu$ M of monomer. C) Time dependent GTPase activity of McrBD279N with McrC (blue dots) and without McrC (Orange dots). The concentration of McrBD279N was 450 nM, McrC 112.5 nM and GTP 1 mM.

This could have been a result of McrC interacting with McrBD279N oligomers formed due to self-association. A similar observation of McrC interacting with McrB oligomers in absence of nucleotide was discussed in Chapter 2.

Aspartate to alanine mutation caused heavy precipitation of the mutant due to which further biochemical characterization could not be carried out. The Walker B aspartate to asparagine mutant was further characterized for its DNA cleavage activity (Figure 3.5 A). As expected, the mutation abolished the nucleolytic activity of the complex completely. Further DNA binding studies indicated that the protein was capable of binding to DNA (Figure 3.5 B) but its McrC stimulated GTPase activity was severely compromised (Figure 3.5 C). Even though the putative Walker B motif has been shown to only affect nucleotide hydrolysis, lack of oligomerization in our study indicates that the Walker B aspartate might be involved in nucleotide binding as well.

#### Walker B Glutamate:

Mutation in glutamate residue of Walker B of McrB (McrBE280A and McrBE280Q) did not affect nucleotide-dependent oligomerization (Figure 3.6 A and B). Previous studies on McrB and NtrC1 have shown that these proteins form heptameric oligomers (Lee et al, 2003; Panne et al, 2001). Consequently, along with demonstrating oligomerization, it was important to establish that the oligomers obtained with these mutants are similar to the wild-type protein in terms of their stoichiometry. A difference

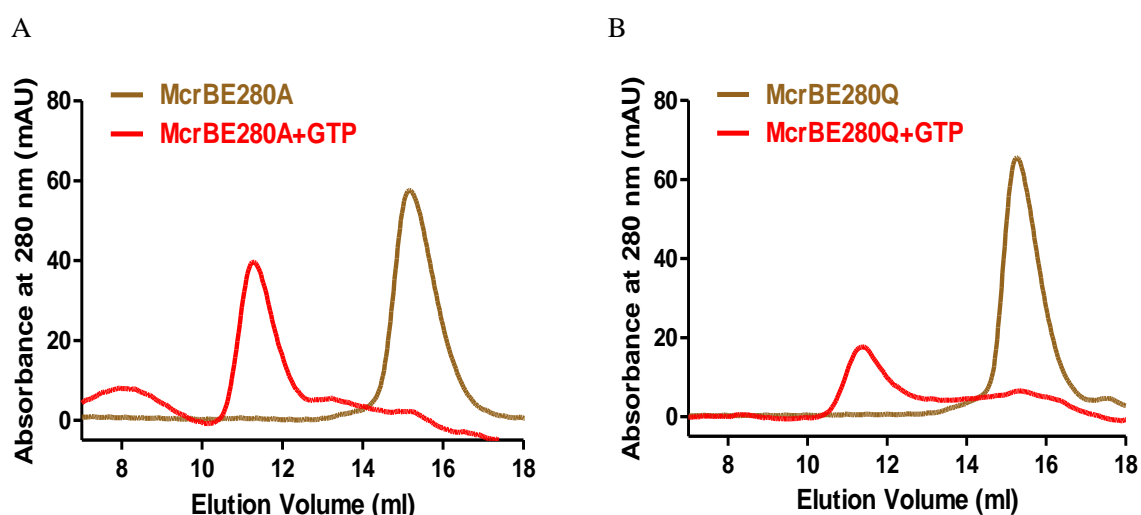


Figure 3.6. **Size exclusion chromatography with the McrB Walker B glutamate mutants.** Gel filtration profile (using 24 ml Superdex200 10/300 GL), both in presence and absence of GTP, of A) 18  $\mu$ M McrBE280A, B) 18  $\mu$ M McrBE280Q

of a hexamer and a heptamer might go undetected due to the low resolution of SEC technique alone.

The molecular mass of McrBE280A and McrBE280Q oligomers in presence of both GTP and GDP was determined using SEC-MALS. The molecular mass estimated for GTP induced oligomers was 332 kDa and 326 kDa for McrBE280A and McrBE280Q respectively (Figure 3.7). This molecular mass is equivalent to the mass of six McrB monomers (molecular mass of monomer is 59 kDa as estimated by SEC-MALS). Size of the GDP induced oligomers for the mutants was also determined (Figure 3.4 B, D). A molecular mass of 353 kDa and 327 kDa was obtained for McrBE280A and McrBE280Q respectively in presence of GDP. Thus confirming that the mutation in the Walker B glutamate did not affect oligomerization or oligomer size of the protein. Combined DLS, SEC, and SEC-MALS showed that the McrBE280A

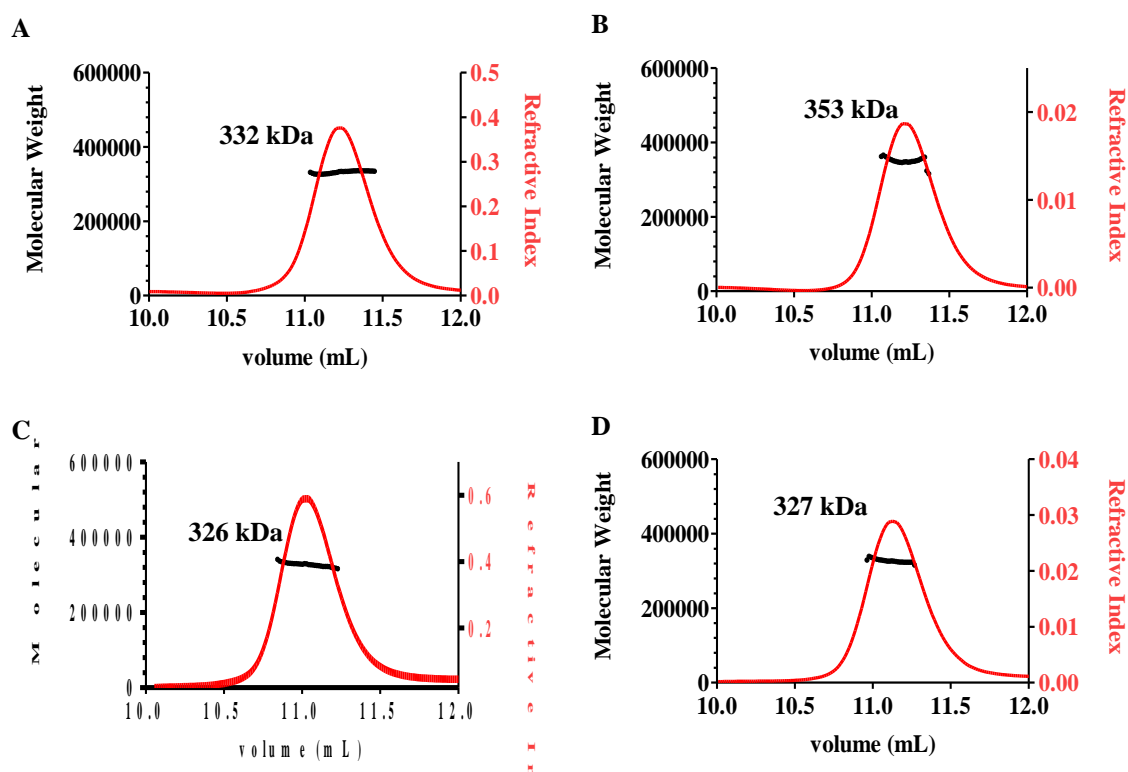


Figure 3.7. **SEC-MALS study of McrB Walker B glutamate mutants.** The chromatogram shows the refractive index signal with the derived molar masses indicated by the thicker horizontal lines for A) McrBE280A in presence of GTP, B) McrBE280A in presence of GDP, C) McrBE280Q, in presence of GTP and D) McrBE280Q in presence of GDP. Protein concentration used in all the runs was 18  $\mu$ M. The elution peaks displayed highly monodisperse population with average mass indicated over the region in each panel.



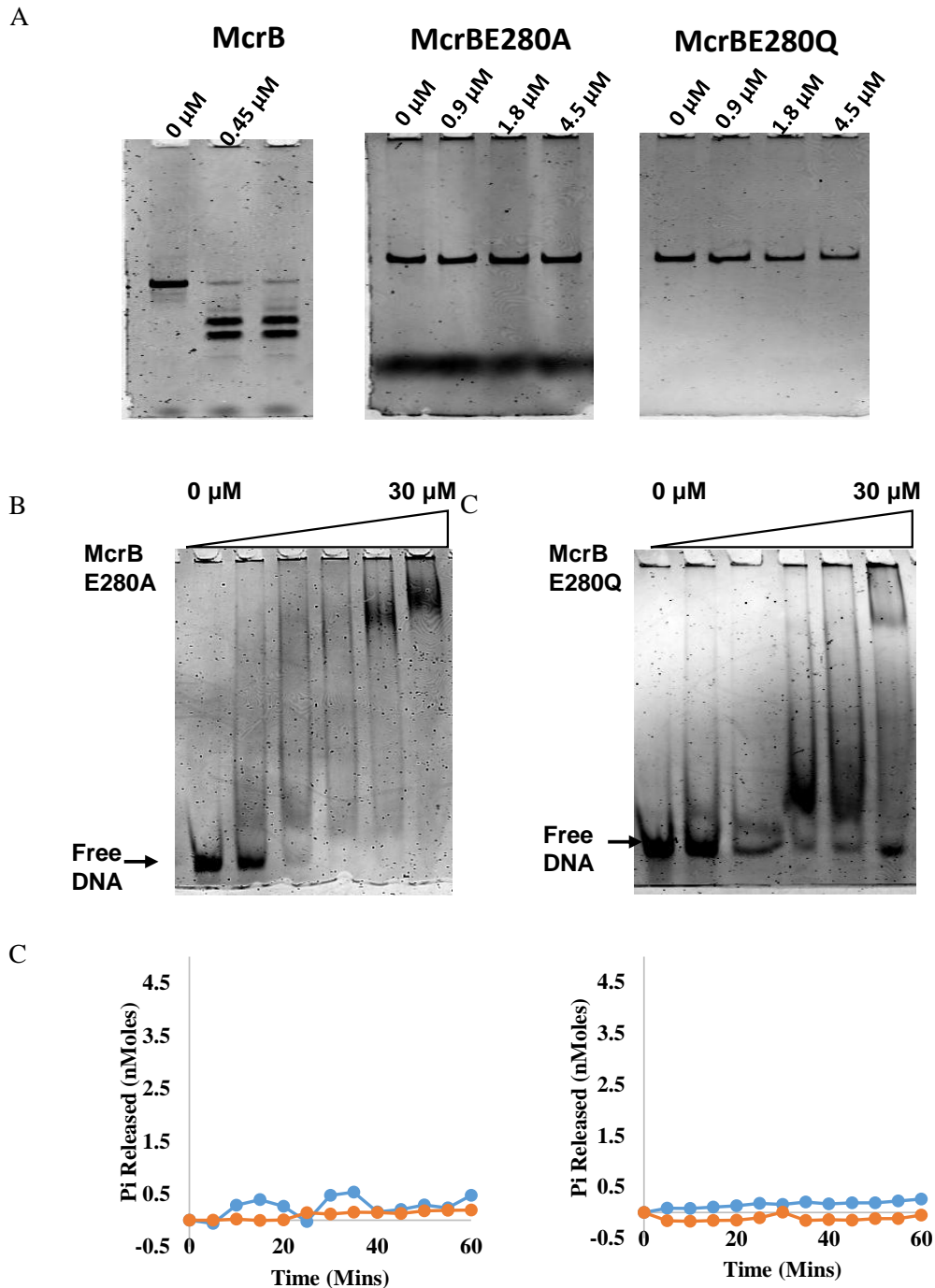


Figure 3.8. **Biochemical analysis of Walker B glutamate mutants.** A) 10 % native PAGE gel showing DNA cleavage activity of McrBD279A and McrBD279N. The Concentration of McrB or mutant is mentioned on each lane. The concentration of McrC was 1/4<sup>th</sup> of the concentration of McrB and its mutants. The reactions were performed with 75 nM 114 bp DNA substrate and 1 mM GTP at 37°C. B) 5% native PAGE gel showing McrBD279N binding to 250 nM specific DNA (described in chapter in presence of 1 mM GTP. Concentrations of McrBD279N used in each lane were- 0  $\mu$ M, 1.5  $\mu$ M, 3  $\mu$ M, 6  $\mu$ M, 12  $\mu$ M, 30  $\mu$ M of monomer. C) Time dependent GTPase activity of McrBD279N with McrC (blue dots) and without McrC (Orange dots). The concentration of McrBD279N was 450 nM, mcrC 112.5 nM and GTP 1 mM.

and McrBE280Q mutants behaved like wild-type showing oligomerisation both in

presence of GTP and in presence of GTP and McrC (Figure 3.7).

Next, the glutamate mutants were tested for their DNA cleavage activity which as expected was lost (Figure 3.8 A). Subsequently, we carried out the DNA binding and GTP hydrolysis characterisation of these mutants. DNA binding activity (Figure 3.8 B) of McrBE280A in presence of GTP was similar to that of the wild-type protein (Chapter 2; Figure 2.9). McrBE280Q, on the other hand, showed a lower DNA binding. Binding studies in presence of McrC, however, was similar to the wild-type protein (data not shown). Being the catalytic residue of Walker B motif, glutamate at the second position has been well characterised amongst AAA+ ATPases. While Walker B aspartate has been shown to bind to  $Mg^{2+}$  ion which is an important co-factor for nucleotide binding and hydrolysis (Gorbalenya & Koonin, 1989; Walker et al, 1982). The glutamate following the aspartate is the catalytic base that primes a water molecule for hydrophilic attack on the  $\beta$ - $\gamma$  phosphodiester bond (Guenther et al, 1997; Karata et al, 1999). Both McrBE280A and McrBE280Q mutants were defective in their GTP hydrolysis activity (Figure 3.8 C).

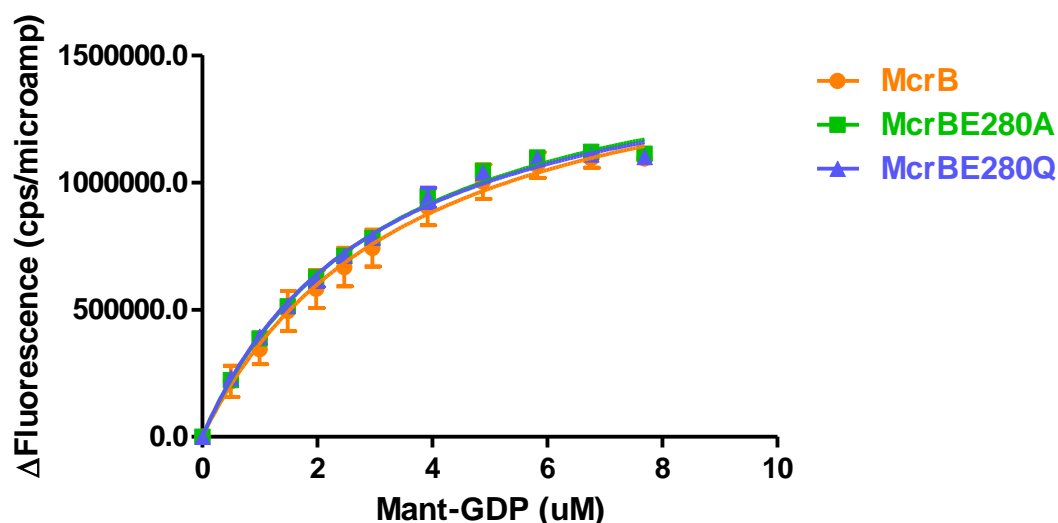


Figure 3.9. **Mant-GDP binding by McrB and its mutants.** The increase in fluorescence induced as a function of binding of 0.5  $\mu$ M protein to Mant-GDP is shown at different concentrations of the Mant-nucleotide.

In a previous report on McrB Walker B residues (Pieper et al, 1999a), it was observed that mutating these residues affected both DNA binding and nucleotide binding. Our mutagenesis studies clearly showed that the GTPase activity of the Walker B glutamate mutants is compromised but they are capable of both oligomerization and DNA binding. Since the mutation in McrBE280 residue showed

both oligomerization and DNA binding, I next sought to see whether the defective GTPase activity is due to altered nucleotide binding or not. For this, I carried out nucleotide binding studies of McrB and its E280 mutants with Mant-GDP (Figure 3.9). For Mant-GDP, the apparent dissociation constant of mutants- McrBE280A ( $K_{d \text{ Apparent}} = 3.2 \mu\text{M}$ ) and McrBE280Q was similar to that of wild-type McrB ( $K_{d \text{ Apparent}} = 3.1 \mu\text{M}$ ) ( $K_{d \text{ Apparent}} = 3.2 \mu\text{M}$ ). Thus clearly, the Walker glutamate in McrB does not participate in nucleotide binding but is crucial for nucleotide hydrolysis.

**Circular Dichroism spectroscopy for ascertaining that the functional defects are not due to protein misfolding:** Circular dichroism (CD) is an excellent tool for rapid determination of the secondary structure and folding properties of proteins that have been obtained using recombinant techniques or purified from tissues. The most widely used application of protein CD spectroscopy is to determine whether an expressed, purified protein is folded, or if a mutation affects its conformation or stability.

I carried out this analysis to determine whether the mutants that were unstable and showed defective oligomerization, DNA binding or GTPase activity were folded properly. Figure 3.10 shows the CD spectra of McrB and Walker B mutants. The  $\alpha$ -helical characteristic indicated by a peak at 222-230 nm,  $\beta$ -sheet structure indicated by minima at 215-218 nm and the conserved relative ratio of  $\alpha$  and  $\beta$  characteristics for different proteins (obtained by calculating the area under the curve) showed that

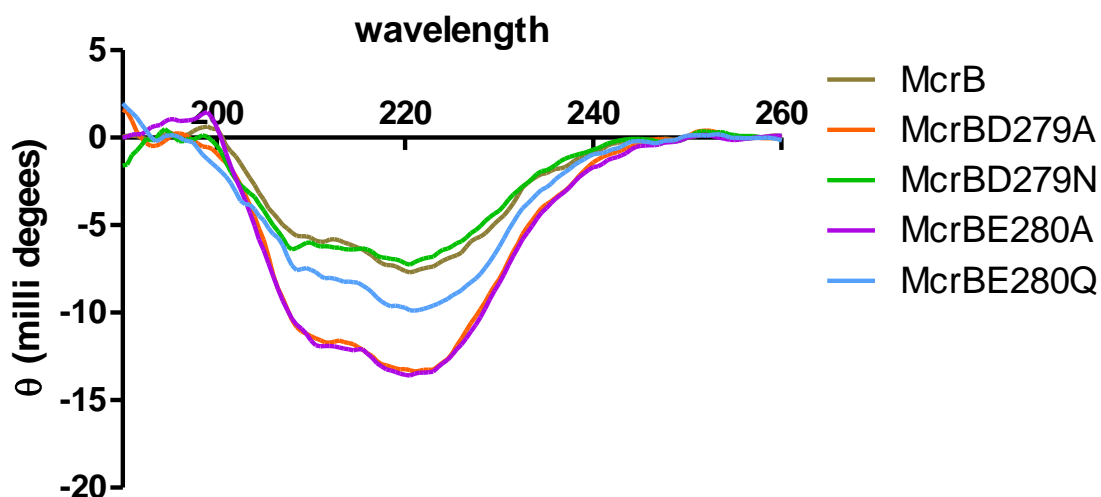


Figure 3.10. **Circular Dichroism spectroscopy of McrB and its mutants.** The circular dichroism spectra of McrB and its Walker B mutants, measured with  $1.7 \mu\text{M}$  protein at  $25^\circ\text{C}$ .

there is no defect in protein folding that would cause a defect in protein activity. Even though a similar concentration of proteins was used, the difference in amplitudes could be due to variations owing to protein precipitation.

### 3.3.4 Characterisation of pore loop mutants

**Oligomerisation:** Mutation in putative pore loop residues also affected protein stability as they tended to precipitate. Replacing the positively charged residues (R253, R254, K255) simultaneously with alanine (McrBLoopA) caused loss of oligomerization (Figure 3.11. A). Changing only one of the positive residue lysine (K255) to alanine (McrBK255A) affected the protein stability. Since during these studies I observed that the N-terminal deletion mutant- McrB $\Delta$ N was more stable than McrB, I mutated the loop of McrB $\Delta$ N as well. The McrB $\Delta$ NLoopA still did not show any improvement in stability or oligomerization. But the McrB $\Delta$ NK255A was relatively well behaved in comparison to the full-length version. Since N-terminal deletion does not affect the oligomerization, I tested the oligomeric property of this mutant. The McrBK255A mutation did not affect oligomerization of the protein (Figure 3.11 B). McrB $\Delta$ NK255A showed oligomer formation in presence of GTP, eluting at 15.8 ml similar to wild-type (15.6 ml) and tetradecamer formation (eluting at 14.1 ml) when McrC was added to the mutant protein in presence of GTP.

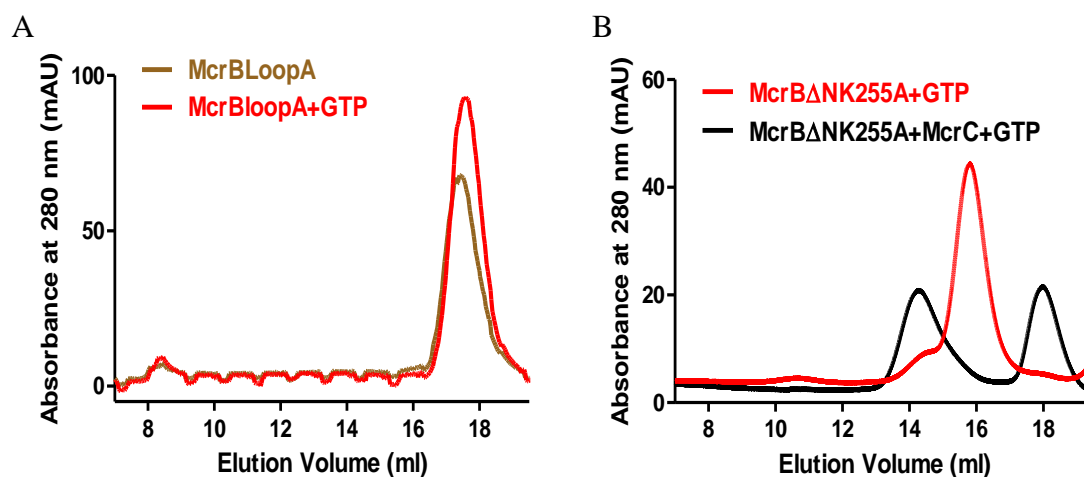
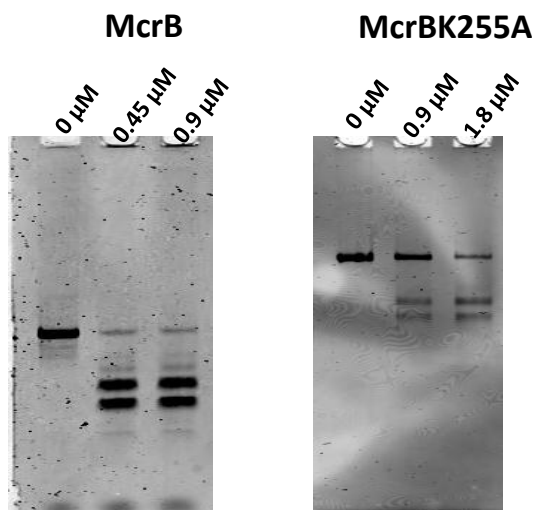


Figure 3.11. **Size exclusion chromatography with the McrBloop mutants.** A) Gel filtration profile (using 24 ml Superose6 10/300 GL) of McrBloopA in presence GTP, B) Gel filtration profile (using 24 ml Superose6 10/300 GL) of McrB $\Delta$ NK255A with and without McrC, showing that the oligomer formed in presence of GTP was capable of interacting with McrC to form the tetradecamer.



**Figure 3.12. DNA cleavage activity of McrBK255A mutant.** 10% native PAGE gel showing DNA cleavage activity of McrB and its mutant McrBK255A. The Concentration of McrB or mutant is mentioned on each lane. The concentration of McrC was 1/4<sup>th</sup> of the concentration of McrB and its mutants. The reactions were performed with 75 nM DNA and 1 mM GTP at 37°C

**DNA cleavage:** The McrBK255A mutant was next tested for its DNA cleavage activity (Figure 3.12). The mutant displayed nucleolytic activity at higher protein concentration, which could be because the mutation partly compromised the nuclease activity. Alternatively, it could have been because the active fraction of the mutant, which was found to be less stable than wild-type, was low. Consequently, it appears that although the mutation does not affect the protein activity significantly, this residue might play a small role in protein stability.

The McrBLoopA mutant was defective in oligomerization. Also, the protein had a tendency to precipitate at a very fast rate. Thus no further characterization was done with this mutant.

### 3.3.5 Characterisation of Sensor 2 mutants

The R404 residue thought to be Sensor 2 of McrB was mutated to alanine. McrBR404A showed both oligomerization and DNA cleavage activity similar to the wild type (Figure 3.13). Thus this mutant was not characterized any further.

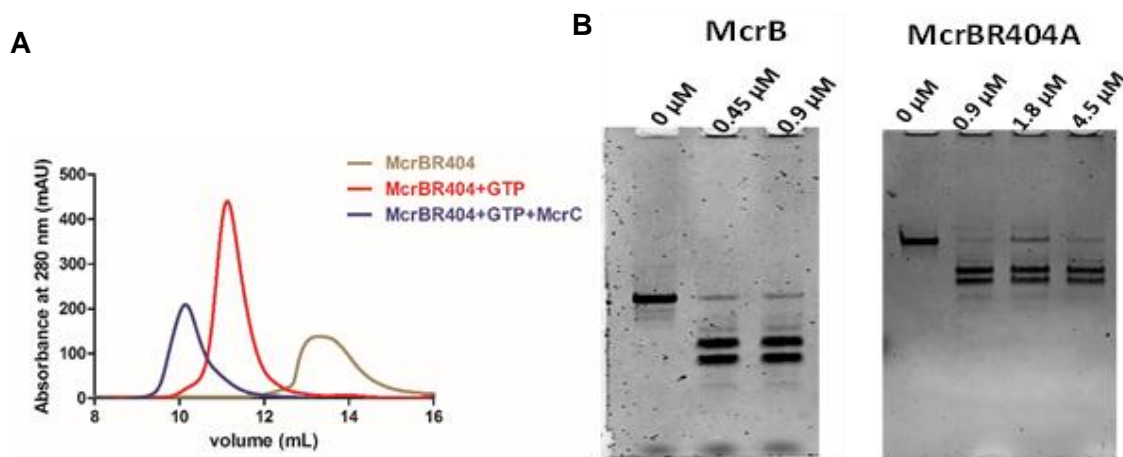


Figure 3.13. **Biochemical activity of putative Sensor 2 R404.** A) Size exclusion chromatography with the 18  $\mu\text{M}$  McrBR404A (using 24 ml Superdex200 10/300 GL), both in presence and absence of 1 mM GTP and 4.5  $\mu\text{M}$  McrC. B) 10% native PAGE gel showing DNA cleavage activity of McrB and its mutant McrBR404A. The Concentration of McrB or mutant is mentioned on each lane. The concentration of McrC was 1/4<sup>th</sup> of the concentration of McrB and its mutants. The reactions were performed with 75 nM DNA and 1 mM GTP at 37°C

### 3.4 Discussion

The mutagenesis studies with McrB were carried out in continuation with studies reported by Pieper et al (Pieper et al, 1999a). The conserved Walker B aspartate in McrB affected the oligomerization of the protein. This can either be due to its role in nucleotide binding, which is integral to oligomerization, and/or because it is part of the interface formed on oligomerization. On the other hand, Walker B glutamate does not appear necessary for oligomerization, but is required for GTP hydrolysis.

I tried to identify a putative pore loop at a position similar to the Helix-2 insert of Clade VI of AAA+ superfamily (Erzberger & Berger, 2006). In the NtrC1 family, which is also classified under Clade VI, this insert consists of a conserved GAFTGA motif and is involved in interaction with its substrate, the transcription factor  $\sigma^{54}$  (Dago et al, 2007; Joly et al, 2012; Zhang et al, 2009). In this study, I observed that a region aligning with the GAFTGA loop of NtrC family proteins has positively charged residues (R253, R254, K255). Since, McrBC is a DNA translocase, these positively charged residues were investigated to see if they are involved in DNA translocation. Two types of mutations were done in this region. In one case, mutating all three residues to alanine affected protein stability and oligomerization property drastically. Since

oligomerization is essential for AAA+ ATPase activity, further studies were not carried out with these mutants.

In the second type of mutation, the McrBK255 was mutated to alanine. Although the sequence alignment of McrB with its homologs showed the K255 residue as well conserved (Figure 3.2 C), mutating this residue did not affect the activity of the protein significantly. Thus, our results indicate that either R253 and R254 or both may be part of the oligomeric interface.

Sensor 2 in AAA+ proteins is involved in both nucleotide binding (Hattendorf & Lindquist, 2002; Lew & Gralla, 2002; Rombel et al, 1999) and hydrolysis (Wendler et al, 2012; Zeymer et al, 2014). The Sensor 2 residue in McrB, based on sequence alignment (Figure 3.1 and 3.2), was postulated to be R404. Mutation of this residue to alanine did not affect the nucleolytic activity of the protein, thus indicating an intact DNA binding and GTPase activity. Since there were no other arginine residues located near  $\alpha$ -7 of McrB, either McrB does not have a Sensor 2 motif like the classical AAA+ clades or the residue acting as a Sensor 2 might not be a canonical arginine as is the case with the SF3 helicases of clade IV which have a lysine as Sensor 2 (Enemark & Joshua-Tor, 2006; Erzberger & Berger, 2006; Gai et al, 2004a; Gai et al, 2004b). In the Rep40 helicase, the Sensor 2 motif has been shown to be replaced by a loop containing LDHDF motif (James et al, 2003). This type of motif has also been observed in other SF3 helicases in the form of LxxxHy (where the Hy is a bulky hydrophobic residue, a tryptophan or a phenylalanine). It is suggested that this loop can compensate for lack of Sensor 2 in Rep40 as it contains hydrophobic residues that can sandwich the base of the bound nucleotide. McrB has the sequence LGKGF preceding R404 (Figure 3.2 A). It will be interesting to see if this stretch of residue acts as a Sensor loop.

The results of mutation of McrB carried out in this work are summarised in Table 3.3. In the next chapter of this thesis, structural studies carried out on McrBC to decipher its mechanism of action, in particular, its GTPase activity, will be discussed.

**Table 3.3.** Table summarising the *Mutational analysis of McrB*. The biochemical properties of McrB and its mutants characterized in this study are summarised in the table. The grey filled boxes indicate that assay was not carried out due to experimental limitations. NA is not applicable to the proteins lacking the DNA binding domain. (+) sign indicate the presence of the activity and (-) sign indicates the absence of activity. The superscript a = shown through SEC, b = SEC-MALS, c = DLS

	Oligomerization						DNA Binding	GTP -ase	Nucle-ase
	w/o McrC			with McrC					
	W/o nucleotide	GTP	GDP	W/o nucleotide	GTP	GDP			
McrB	Monomer (a,b,c)	Hexamer (a,b,c)	Hexamer (a,b,c)	Monomer (a,b,c)	Tetra-decamer (a,b,c)		+	+	+
McrB D279A	Monomer (a,c)	Monomer (a,c)	Monomer (a,c)				+	-	-
McrB D279N	Monomer (a,c)	Monomer (a,c)	Monomer (a,c)				+	-	-
McrB E280A	Monomer (a,b,c)	Hexamer (a,b,c)	Hexamer (a,b,c)				+	-	-
McrB E280Q	Monomer (a,b,c)	Hexamer (a,b,c)	Hexamer (a,b,c)				+	-	-
McrB K255A							+	+	+
MBAN K255A		Hexamer (a)			Tetra-decamer (a)		NA	+	NA
McrB LoopA		Monomer (a)			Monomer (a)			-	
MBAN LoopA	Monomer (a)	Monomer (a)			Monomer (a)		NA	-	NA
McrB R404	Monomer (a)	Hexamer (a)		Monomer (a)	Tetra-decamer (a)		+	+	+

## References

Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic acids research* **41**: W349-357



Bush M, Dixon R (2012) The role of bacterial enhancer binding proteins as specialized activators of sigma54-dependent transcription. *Microbiology and molecular biology reviews : MMBR* **76**: 497-529

Chen B, Sysoeva TA, Chowdhury S, Guo L, De Carlo S, Hanson JA, Yang H, Nixon BT (2010) Engagement of arginine finger to ATP triggers large conformational changes in NtrC1 AAA+ ATPase for remodeling bacterial RNA polymerase. *Structure* **18**: 1420-1430

Dago AE, Wigneshweraraj SR, Buck M, Morett E (2007) A role for the conserved GAFTGA motif of AAA+ transcription activators in sensing promoter DNA conformation. *The Journal of biological chemistry* **282**: 1087-1097

Enemark EJ, Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**: 270-275

Erzberger JP, Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annual review of biophysics and biomolecular structure* **35**: 93-114

Gai D, Li D, Finkielstein CV, Ott RD, Taneja P, Fanning E, Chen XS (2004a) Insights into the oligomeric states, conformational changes, and helicase activities of SV40 large tumor antigen. *J Biol Chem* **279**: 38952-38959

Gai D, Zhao R, Li D, Finkielstein CV, Chen XS (2004b) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**: 47-60

Gorbalenya AE, Koonin EV (1989) Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic acids research* **17**: 8413-8440

Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**: 2876-2890

Guenther B, Onrust R, Sali A, O'Donnell M, Kuriyan J (1997) Crystal structure of the delta' subunit of the clamp-loader complex of E. coli DNA polymerase III. *Cell* **91**: 335-345

Hattendorf DA, Lindquist SL (2002) Cooperative kinetics of both Hsp104 ATPase domains and interdomain communication revealed by AAA sensor-1 mutants. *EMBO J* **21**: 12-21

Huang H, Chung B, Jung J, Park HW, Chang T (2009) Toroidal micelles of uniform size from diblock copolymers. *Angew Chem Int Ed Engl* **48**: 4594-4597

Iyer LM, Leipe DD, Koonin EV, Aravind L (2004) Evolutionary history and higher order classification of AAA+ ATPases. *Journal of structural biology* **146**: 11-31

James JA, Escalante CR, Yoon-Robarts M, Edwards TA, Linden RM, Aggarwal AK (2003) Crystal structure of the SF3 helicase from adeno-associated virus type 2. *Structure* **11**: 1025-1035

Joly N, Zhang N, Buck M, Zhang X (2012) Coupling AAA protein function to regulated gene expression. *Biochim Biophys Acta* **1823**: 108-116

Jones DT, Tress M, Bryson K, Hadley C (1999) Successful recognition of protein folds using threading methods biased by sequence similarity and predicted secondary structure. *Proteins Suppl* **3**: 104-111

Karata K, Inagawa T, Wilkinson AJ, Tatsuta T, Ogura T (1999) Dissecting the role of a conserved motif (the second region of homology) in the AAA family of ATPases. Site-directed mutagenesis of the ATP-dependent protease FtsH. *The Journal of biological chemistry* **274**: 26225-26232

Kruger T, Wild C, Noyer-Weidner M (1995) McrB: a prokaryotic protein specifically recognizing DNA containing modified cytosine residues. *The EMBO journal* **14**: 2661-2669

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948

Lee SY, De La Torre A, Yan D, Kustu S, Nixon BT, Wemmer DE (2003) Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. *Genes & development* **17**: 2552-2563

Lew CM, Gralla JD (2002) New roles for conserved regions within a sigma 54-dependent enhancer-binding protein. *J Biol Chem* **277**: 41517-41524

Neuwald AF, Aravind L, Spouge JL, Koonin EV (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome research* **9**: 27-43

O'Shea VL, Berger JM (2014) Loading strategies of ring-shaped nucleic acid translocases and helicases. *Current opinion in structural biology* **25**: 16-24

Panne D, Muller SA, Wirtz S, Engel A, Bickle TA (2001) The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *The EMBO journal* **20**: 3210-3217

Panne D, Raleigh EA, Bickle TA (1999) The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. *Journal of molecular biology* **290**: 49-60

Pieper U, Brinkmann T, Kruger T, Noyer-Weidner M, Pingoud A (1997) Characterization of the interaction between the restriction endonuclease McrBC from *E. coli* and its cofactor GTP. *Journal of molecular biology* **272**: 190-199

Pieper U, Pingoud A (2002) A mutational analysis of the PD...D/EXK motif suggests that McrC harbors the catalytic center for DNA cleavage by the GTP-dependent restriction enzyme McrBC from *Escherichia coli*. *Biochemistry* **41**: 5236-5244

Pieper U, Schweitzer T, Groll DH, Gast FU, Pingoud A (1999a) The GTP-binding domain of McrB: more than just a variation on a common theme? *Journal of molecular biology* **292**: 547-556

Pieper U, Schweitzer T, Groll DH, Pingoud A (1999b) Defining the location and function of domains of McrB by deletion mutagenesis. *Biological chemistry* **380**: 1225-1230

Rombel I, Peters-Wendisch P, Mesecar A, Thorgeirsson T, Shin YK, Kustu S (1999) MgATP binding and hydrolysis determinants of NtrC, a bacterial enhancer-binding protein. *J Bacteriol* **181**: 4628-4638

Sanders CM, Kovalevskiy OV, Sizov D, Lebedev AA, Isupov MN, Antson AA (2007) Papillomavirus E1 helicase assembly maintains an asymmetric state in the absence of DNA and nucleotide cofactors. *Nucleic acids research* **35**: 6451-6457

Schumacher J, Joly N, Rappas M, Zhang X, Buck M (2006) Structures and organisation of AAA+ enhancer binding proteins in transcriptional activation. *Journal of structural biology* **156**: 190-199

Sutherland E, Coe L, Raleigh EA (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *Journal of molecular biology* **225**: 327-348

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research* **25**: 4876-4882

Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-

requiring enzymes and a common nucleotide binding fold. *The EMBO journal* **1**: 945-951

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191

Wendler P, Ciniawsky S, Kock M, Kube S (2012) Structure and function of the AAA+ nucleotide binding pocket. *Biochim Biophys Acta* **1823**: 2-14

Zeymer C, Barends TR, Werbeck ND, Schlichting I, Reinstein J (2014) Elements in nucleotide sensing and hydrolysis of the AAA+ disaggregation machine ClpB: a structure-based mechanistic dissection of a molecular motor. *Acta crystallographica Section D, Biological crystallography* **70**: 582-595

Zhang N, Joly N, Burrows PC, Jovanovic M, Wigneshweraraj SR, Buck M (2009) The role of the conserved phenylalanine in the sigma54-interacting GAFTGA motif of bacterial enhancer binding proteins. *Nucleic acids research* **37**: 5981-5992

## Chapter 4

### Crystallographic and cryo-EM studies of McrBC

#### 4.1 Introduction

McrBC is a complex of two proteins- McrB, which belongs to the AAA+ family of proteins, and McrC belonging to the PD...D/EXK superfamily of Mg<sup>2+</sup> dependent endonucleases. McrB subunit of this complex is peculiar amongst other AAA+ motors as it hydrolyzes GTP instead of ATP. Although this protein binds to ATP weakly (Pieper et al, 1997) and ATP has been observed to have an inhibitory effect on its DNA cleavage activity (Sutherland et al, 1992), the determinants for GTP specificity over ATP are not yet known. Functional mutagenesis of consensus motifs for GTP binding in McrB (300-DKRG-303, 333-NTAD-336; discussed in Chapter 1 and chapter 3) did not shed much light on GTP specificity (Pieper et al, 1997). The intrinsic GTP hydrolysis rate of McrB is very low but McrC stimulates its GTPase activity by about 30 fold (Pieper et al, 1997). McrBC complex hydrolyzes GTP in a DNA independent manner. While GTPase activity of McrB is not affected by specific DNA, McrBC shows a moderately stimulated GTPase activity in presence of specific DNA (Panne et al, 1999). This is different from other NTP dependent RM systems (Type I and Type III), which show higher ATP hydrolysis in presence of their specific DNA substrates.

McrBC is similar to type III RM enzymes in terms of its DNA cleavage activity, as it cleaves DNA close to one of the recognition sites. This suggests that at least one target site bound enzyme is necessary for DNA cleavage. It was proposed by Panne et al. that nucleolytic cleavage by McrBC is coupled to DNA translocation (Panne et al., 1999). Their studies suggested that GTP binding causes conformational changes in McrB, which lead to higher affinity for DNA. Upon DNA binding, the enzyme translocates on DNA using energy derived from GTP hydrolysis, and DNA cleavage occurs when a translocating enzyme collides with another enzyme complex bound at far away target sequence. Unlike in case of Type I and Type III RM enzymes where convergence of two enzyme complexes are required for double-strand DNA break, collision of an McrBC complex to a non-specific translocational block, in the form of a lac repressor bound to DNA, could also affect DNA cleavage by McrBC (Panne et al, 1999).

Although the McrBC complex is extensively characterized for its biochemical properties, a mechanistic understanding of its activities is still lacking. The foremost unanswered question at the time of my project was the nature of the oligomeric assembly of McrBC. The architecture of the assembly would provide insights into how McrC, the activator of the AAA+ motor, would bind to and stimulate McrB and carry out GTPase-coupled nucleolytic cleavage. Towards this, I carried out X-ray crystallographic and electron cryomicroscopy (cryo-EM) studies to determine the three-dimensional structure of McrB and McrBC. This chapter describes the results of my efforts in this direction.

## 4.2 Materials and Methods

### 4.2.1 *Cloning of different proteins*

Cloning of *mcrB*, *mcrC* and *mcrBΔN* with a hexahistidine tag is described in chapter 2, section 2.2.1. For crystallization, another construct of *mcrBΔN* lacking the C-terminal histidine tag i.e. *mcrBΔN<sup>WT</sup>* (without tag) was also cloned. *mcrBΔN<sup>WT</sup>* was amplified from *mcrBΔN* gene in pHIS17 vector using following primers:

McrBΔNWThis-1F	GTTTAACTTTAAGAAGGAGATATACATATGTCAAAAAGTGA ATCATACTG
McrBΔNWThis-1R	TTAATGATGATGATGATGATGGGATCCCTATGAGTCCCCTA ATAATTTGTTGG

The amplified product was cloned into the pHIS17 vector using restriction-free cloning method as described in chapter 2, section 2.2.1. The resulting *mcrBΔN<sup>WT</sup>* gene was sequenced to ensure the sequence integrity of the gene.

### 4.2.2 *Purification of McrB, McrBΔN, McrBΔN<sup>WT</sup> and McrC*

The purification protocol for histidine-tagged McrB, McrBΔN and McrC proteins is described in chapter 2, section 2.2.2. McrBΔN<sup>WT</sup> was overexpressed using pHISMcrBΔN<sup>WT</sup> plasmid in *E. coli* BL21 (AI) cells. The culture was grown in 2 L LB media containing 100 μg/ml ampicillin in an incubator-shaker at 37°C until OD reached 0.3 at 600 nm. The temperature of incubator-shaker was then reduced to 18°C and cultures were induced with 0.06% w/v L-Arabinose. The cultures were grown further

overnight (15-16 hours) at 18°C. Cells were pelleted by centrifugation at 4°C and 3,315 g for 15 minutes. The pellet was resuspended in 50 ml lysis buffer (50 mM Tris-Cl pH 8, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, 0.04% CHAPS and 1 mM DTT). The cell lysate was then clarified by ultracentrifugation at 4°C and 159,200 g for 40 minutes. The clarified supernatant of the cell lysate was loaded onto three columns connected in series- 5 ml HiTrap™ Heparin column (GE Life Sciences), 5 ml HiTrap™ SP HP (GE Life Sciences), 5 ml HiTrap Q HP (GE Life Sciences). The columns were equilibrated with Buffer B50 (50 mM Tris-Cl pH 8, 50 mM NaCl, 1 mM EDTA, 1 mM DTT) before loading the supernatant. Flow through from this step was collected and 45% ammonium sulfate was added followed by centrifugation in SS34 tubes placed in JA 25.5 rotor (Avanti High-Speed centrifuge) at 39,191 g, 4°C for 20 minutes. The final ammonium sulfate concentration of the supernatant (45% ammonium sulfate) was made to 70% and again centrifuged in SS34 tubes placed in JA 25.5 rotor (Avanti High-Speed centrifuge) at 39,191 g 4°C for 20 minutes. The pellet from 75% ammonium sulphate precipitation was resuspended in 500 ml Buffer B0 (50 mM Tris-Cl pH 8, 1 mM EDTA, 1 mM DTT) and loaded onto a 5 ml HiTrap DEAE FF (GE Life Sciences) column equilibrated with Buffer B50. 4 ml fractions were collected in 20 column volumes over a linear gradient of 0% to 100% of buffer B1000 (50 mM Tris-Cl pH 8, 1000 mM NaCl, 1 mM EDTA and 1 mM DTT). The fractions with the highest purity were pooled and equal amount of buffer B50+2M(NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (50 mM Tris-Cl pH 8, 50 mM NaCl, 1 mM EDTA, 1 mM DTT and 2 M ammonium sulfate) was added. The protein solution was then loaded onto a 5 ml HiTrap Phenyl FF (low substitution) column (GE Life Sciences) equilibrated with buffer B50+2M(NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. 2 ml fractions were collected in 20 column volumes over a linear gradient of 0% to 100% of buffer B50. Pure fractions were dialysed against 2 L dialysis buffer (50 mM Tris-Cl pH 8, 50 mM NaCl, 1 mM EDTA, and 1 mM DTT) overnight. Dialysed McrBΔN<sup>WT</sup> protein solution was loaded onto an 8 ml MonoQ 10/100 GL column (GE Life Sciences) equilibrated with Buffer B50. 2 ml fractions were collected over 20 column volumes using a linear gradient of 0% to 50% of buffer B1000. The pure fractions were pooled and concentrated using a 2 ml 10 kDa vivaspin2 concentrator (GE Life Sciences). Concentrated sample (500 µl) was washed with 5 ml buffer B100 (50 mM Tris-Cl pH 8, 100 mM NaCl and 1 mM DTT) to remove EDTA. The concentrated protein was then incubated with 2.5 mM GTP, 5 mM MgCl<sub>2</sub> for 10 minutes at room temperature. Sample was centrifuged for 15 minutes at 12,000 g, 4°C before loading onto 24 ml

Superdex200 10/300 GL column (GE Life Sciences) equilibrated with buffer B100<sup>+GTP</sup> (50 mM Tris-Cl pH 8, 100 mM NaCl, 0.1 mM GTP, 5 mM MgCl<sub>2</sub>, 1 mM DTT). Pure fractions were pooled and concentrated using a 2 ml 10 kDa Vivaspın2 concentrator (GE Life Sciences). The concentrated protein was washed with storage buffer (100 mM NaCl, 10 mM Tris-Cl pH 7.4 and 1 mM DTT) to remove GTP. The concentration of protein was estimated using Bradford reagent with BSA as standard (Bradford, 1976). The purified protein was stored at -80°C.

#### **4.2.3 Purification of selenomethionine labeled McrB $\Delta$ N, McrB $\Delta$ N<sup>WT</sup> and McrC**

McrB $\Delta$ N has 8 methionines and McrC has 7 methionines. Selenomethionine derivatives of these proteins were generated for experimental phase determination. Selenomethionine labeled McrB $\Delta$ N, McrB $\Delta$ N<sup>WT</sup> and McrC were expressed using the same vector (pHIS17) and in same host strain (*E. coli* BL21 (AI)). The primary culture was grown using LB media containing 100  $\mu$ g/ml ampicillin. 1 ml primary culture was pelleted at 3315 g, 10 minutes, 4°C and resuspended in 1 L LeMaster's minimal media (LeMaster & Richards, 1982). Since the BL21 (AI) strain is not a methionine auxotroph, the cells were first grown in LeMaster's minimal media containing 100  $\mu$ g/ml ampicillin in an incubator-shaker at 37°C in absence of methionine. All the minimal media components were bought from Sigma-Aldrich. At an OD of 0.3 (at 600 nm), the temperature of incubator-shaker was reduced to 18°C and excess of amino acids – 100 mg each of lysine hydrochloride, phenylalanine and threonine, 50 mg each of leucine, isoleucine and valine were added to the culture along with 50 mg L-selenomethionine. The excess of these amino acids is known to inhibit the methionine biosynthesis thus affecting the uptake of selenomethionine from the media (Van Duyne et al, 1993). After 15 minutes, the culture was induced with 0.06% w/v L-arabinose. The culture was grown further overnight (15-16 hours) at 18°C. Cell harvesting and protein purification were similar to that described for unlabelled proteins in Chapter 2, section 2.2.2 and section 5.2.2.

#### **4.2.4 Purification of McrBC, McrB $\Delta$ NC and McrB $\Delta$ N<sup>WT</sup>C complex**

The individual subunits, purified separately, were mixed together to obtain the complexes. Complexation of McrB $\Delta$ N constructs (both tagged and without a tag, labeled or unlabeled with selenomethionine) with McrC (selenomethionine labeled and unlabeled) was carried out using SEC as described in chapter 2, section 2.2.3.3.



## **4.2.5 Crystallization**

### **4.2.5.1 Initial Screening**

For preliminary crystallization trials, sitting drop vapor diffusion method was used. 96 well plates were set with commercial crystallization screens from Hampton, Molecular Dimensions and Jena bioscience. 10 mg/ml protein was incubated with or without nucleotide (GTP, GDP, GDPNP obtained from Jena Bioscience) at 4°C in crystallization buffer (10 mM Tris-Cl, pH 7.4, 100 mM NaCl, 5 mM MgCl<sub>2</sub> and 1 mM DTT). The sample was centrifuged at 15,000 rpm for 10 minutes before setting up crystallization trials. 100 nL protein and 100 nL reservoir solution were mixed to set a crystallization nanodrop of 200 nL. The nanodrops were set using the Mosquito robotic liquid handler. The plates were incubated at 18°C incubator.

### **4.2.5.2 Fine Screening and optimisation**

Following the initial screening, final optimization of crystallization was carried out with larger drop sizes. For this 48-well sitting drop and 24 well hanging drop vapor diffusion method was used. In case of 48-well sitting drops, reservoir wells were filled with 200 µl mother liquor and a drop of 0.5 µl of sample was mixed with 0.5 µl mother liquor. 24 well hanging drop plates were set with reservoir wells containing 500 µl mother liquor. Crystallization drops were set by mixing 1 µl of the sample with 1 µl mother liquor. Coverslips for setting the hanging drops were obtained from Sigma-Aldrich (Grace Bio-Labs HybriSlip™ hybridization cover L × W × thickness 22 mm × 60 mm × 0.25 mm).

### **4.2.6 X-ray data collection and processing**

Initial screening for diffraction quality of crystals was carried out at in-house diffraction facility (Rigaku MicroMax 007 X-ray generator and Mar research 345D detector, IISER Pune). Crystals were cryoprotected by a gradual increment of cryoprotectant in reservoir buffer starting from 5% to a final concentration of 35% ethylene glycol with 5% increment at each step. X-ray diffraction data were collected at the synchrotron facilities at Diamond Light Source (DLS), Oxfordshire, UK and European Synchrotron Radiation Facility (ESRF), Grenoble, France. Indexing and processing of the diffraction data was done using iMosflm (Battye et al, 2011) or XDS (Kabsch, 2010). The processed data were scaled and merged using AIMLESS (Evans & Murshudov,

2013) in CCP4 suite (Winn et al, 2011). 5% of the reflections were set aside to calculate  $R_{\text{Free}}$  during structure refinement.

#### **4.2.7 cryo-EM and image processing of McrB $\Delta$ N<sup>WT</sup>C**

Grid preparation, data collection, 2D classification and 3D reconstructions were carried out by Dr. Vinothkumar Kutti at MRC LMB, Cambridge UK. McrB $\Delta$ N<sup>WT</sup>C (henceforth referred as McrB $\Delta$ NC in cryo-EM studies unless otherwise mentioned) was incubated with final concentrations of 1mM MgCl<sub>2</sub> and 1mM GDPNP for 30 minutes at RT to initiate the assembly of complex and then kept on ice till EM grid preparation. The enzyme at concentrations between 2.5-3.5 mg/ml were applied to Quantifoil holey carbon grids (Au 300 mesh, 1.2/1.3) with blotting and freezing done with a manual plunger. The frozen grids were mounted on Titan Krios Autogrids and imaged with Falcon II detector in integration mode or on Falcon 3 detector in counting mode.

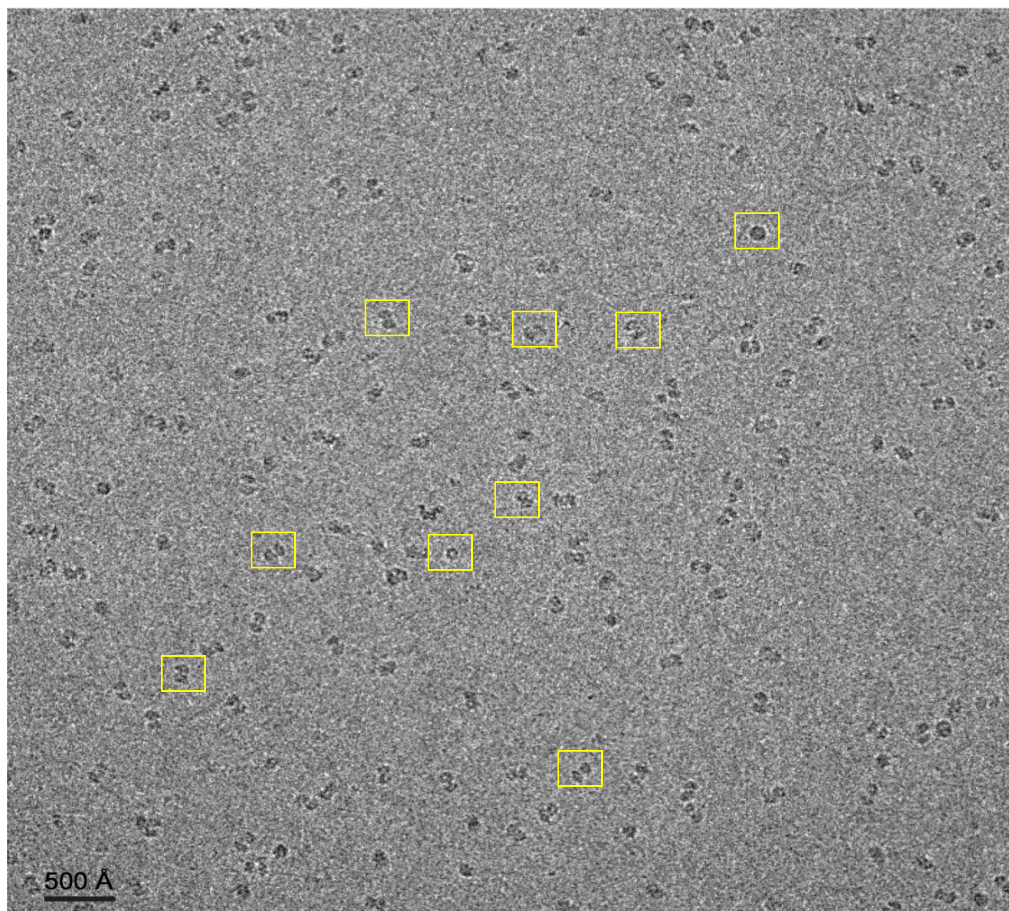
Initially, data was collected at 47,000 x (calibrated magnification 80000 corresponding to 1.75 Å/pixel) with a 4 second exposure and 49 frames were collected using EPU software (Tan et al, 2016). The movie frames were aligned using Unblur (Grant & Grigorieff, 2015) (Grant & Grigorieff, 2015) and frames 1 to 32 were summed and used for further processing. The summed image was used to determine the CTF using CTFFIND4 (Rohou & Grigorieff, 2015). From 492 micrographs, 74318 particles were automatically picked using 2D classes as templates from a previous data set in RELION 1.4 (Scheres, 2012) and extracted with a box size of 180 pixels. Subsequently, 2D classification was performed and classes with good features corresponding to 33905 particles were selected for further refinement. An initial model from the previous refinement determined using the 2D class averages and EMAN2 (Tang et al, 2007) was used as a reference map for further 3D refinement with no symmetry imposed.

The reported resolutions are after post-processing step of RELION with a soft mask applied and the effect of mask checked with phase randomization. With all the particles, the resolution after refinement and masking was 8.5 Å and the map was sharpened with a B-factor of -486. Some regions of the map, in particular, the hexameric region had poor density. The refined particles were subjected to 3D classification into 6 classes with resolution limited to 12 Å. Two classes that showed

good features were individually selected for further 3D refinement. Reconstruction of class1 with 9411 particles resulted in an 8.5 Å map (B-factor -511) and class 2 with 4921 particles resulted in a 9.8 Å map (B-factor of -539). Adding more particles, roughly 3 times more, didn't result in significant improvement in the quality of the map or resolution. This could be simply because of the inherent flexibility of the molecule.

When the new Falcon 3 detector operating in counting mode became available McrB $\Delta$ NC imaging was pursued again with the assumption that increase in DQE (Detective Quantum Efficiency) of the detector might result in better alignment and thus better maps. Data were collected with EPU and Falcon 3 detector in counting mode at 1.33 Å/pixel. Defocus range was set between 2.4 -3.6  $\mu$ m with 0.3  $\mu$ m step with autofocus routine performed every 8  $\mu$ m. The images were exposed for 60 seconds with a total accumulated dose of  $\sim$ 19.5 e<sup>-</sup>/Å<sup>2</sup> and dose fractionated into 75 frames. The movies were grouped into 25 frames resulting in  $\sim$ 0.78 e<sup>-</sup>/frame and Unblur was used for alignment. The summed images were then used for automated particle picking with Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/>) with template derived from previous data collection and contrast transfer function (CTF) was estimated with Gctf (Zhang, 2016). 32579 particles were extracted with a box size of 256 pixels and subjected to two rounds of 2D classification resulting in 15918 good particles (Figure 4.1). Further 3D auto-refinement, 3D classification and post-processing were performed with RELION 2.0. The overall resolution of the map with all the particles was 7.4 Å and the B-factor sharpening of -524.

A



B

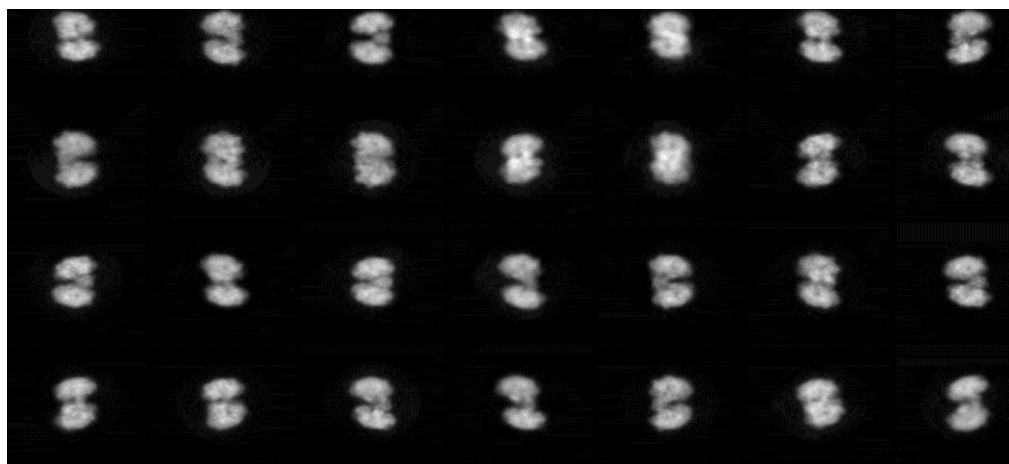


Figure 4.1. **Cryo-EM images of McrB $\Delta$ NC in presence of GDPNP. A)** A small area of a micrograph of McrB $\Delta$ NC, and **B)** Reference-free 2D class averages of McrB $\Delta$ NC

## 4.3 Results

### 4.3.1 Purification of McrB $\Delta$ N<sup>WT</sup>

McrB $\Delta$ N<sup>WT</sup> was purified using a variety of columns and ammonium sulfate precipitation. Figure 4.2 A shows the representative gels corresponding to different steps of purification. The purity of the protein was lower than that of the histidine-tagged protein at the final step of purification. However, I was able to remove many of the impurities when McrB $\Delta$ N<sup>WT</sup>C complex was assembled and purified by size exclusion chromatography (Figure 4.2 B).

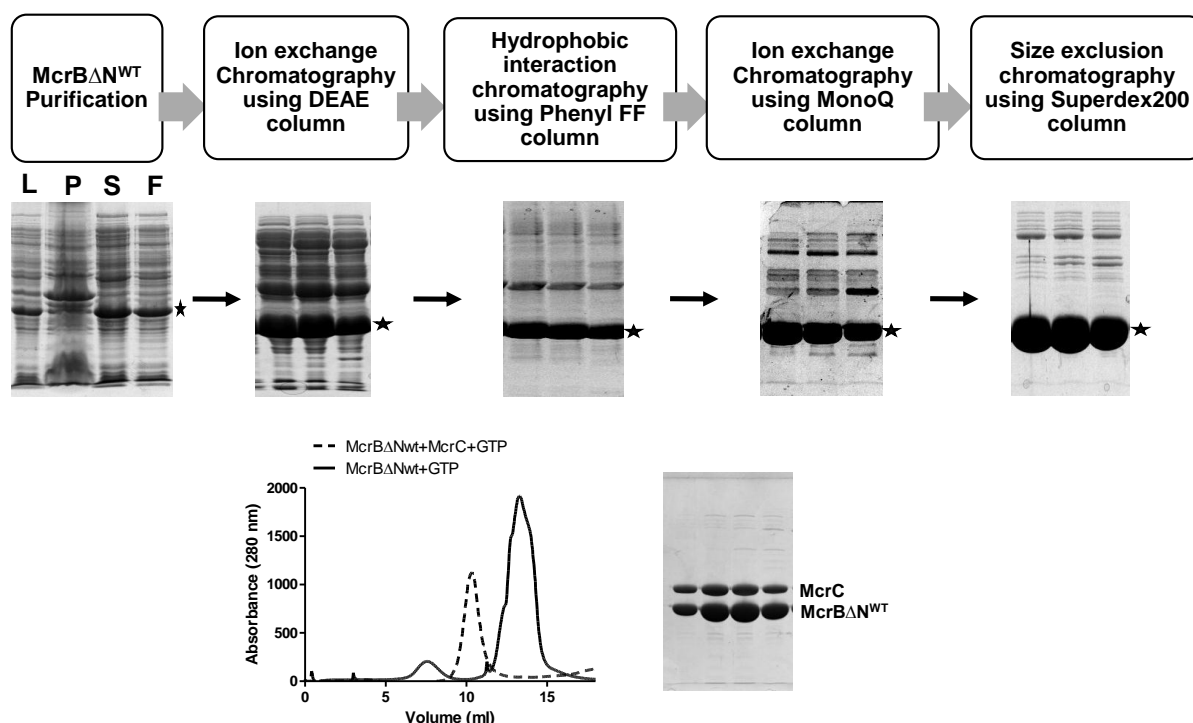


Figure 4.2. **Purification of McrB $\Delta$ N<sup>WT</sup>**. A) SDS PAGE gels corresponding to different steps in purification of McrB $\Delta$ N<sup>WT</sup>. Star indicates McrB $\Delta$ N<sup>WT</sup> in each gel. L=lysate, P=pellet, S=supernatant, F= flow through. B) A chromatogram showing the complexation of McrB $\Delta$ N<sup>WT</sup> and McrC along with the SDS PAGE of samples corresponding to McrB $\Delta$ N<sup>WT</sup>C peak showing both the proteins in the complex.

### 4.3.2 Crystallization trials with the full-length McrBC complex

Initial efforts were made to crystallize the full-length proteins- McrB and McrC individually, as well as the McrBC complex. Amongst these trials, only McrB complexed with non-hydrolysable GTP analog GDPNP crystallized (data not shown). The quality and diffraction of these crystals could not be improved.

### 4.3.3 Crystallization trials with the truncated protein- *McrB* $\Delta$ N and *McrB* $\Delta$ NC complex

Since the full-length protein and its complex did not give any good crystals, crystallization of the truncated protein *McrB* $\Delta$ N was also initiated. This effort yielded crystals both in monomeric and oligomeric form (in presence of GDPNP) of the protein. In absence of nucleotide, the protein crystallized in a condition containing 0.2 M ammonium sulfate, 25% PEG 3350 and 0.1 M Tris-Cl pH 8.5. I tried to improve the size of these crystals using additive screens and changing protein concentration. Unfortunately, the size of these crystals could not be improved and they diffracted poorly despite much optimization efforts (Figure 4.3 A).

The oligomer of *McrB* $\Delta$ N complexed with GDPNP, on the other hand, gave us multiple crystallization hits. Amongst these conditions, best crystals in terms of nucleation, size and reproducibility were obtained in presence of 0.2 M potassium chloride (KCl), 16% PEG 4000 and 0.1 M Tris pH 7.4 (Figure 4.3 B). These crystals diffracted to about 8 Å in-house. The native crystals diffracted to about 4.5 Å at the synchrotron. Crystals of selenomethionine derivative of *McrB* $\Delta$ N complexed with GDPNP diffracted more poorly to about 5.4 Å. Multi-wavelength anomalous dispersion (MAD) data were collected for these crystals (Table 4.1). Calculation of unit cell dimensions and Mathew's coefficient predicted two hexamers in the asymmetric unit (ASU). Each monomer of *McrB* $\Delta$ N contains 8 methionines, giving rise to 48 selenium

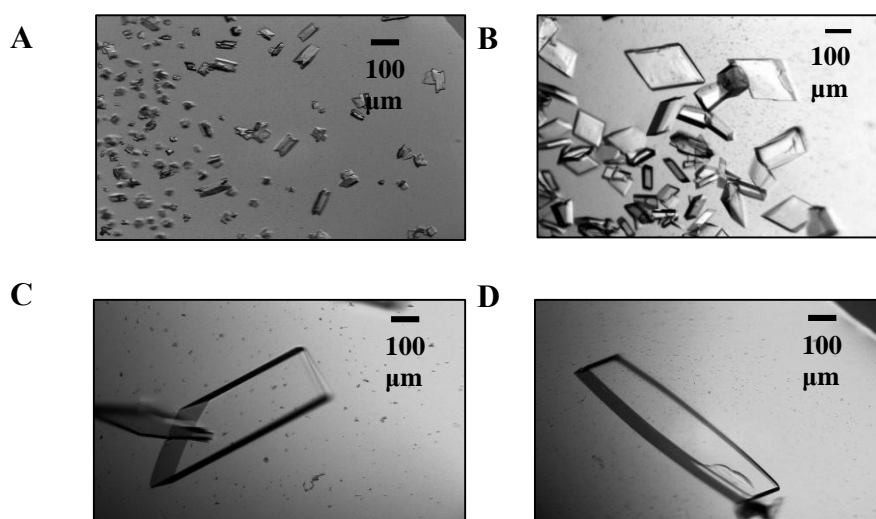


Figure 4.3. **Images of *McrB* $\Delta$ N (histidine-tagged) crystals.** A) Crystals of *McrB* $\Delta$ N obtained in absence of nucleotide, B) Crystals of *McrB* $\Delta$ N obtained in complex with GDPNP C) crystal of *McrB* $\Delta$ NC obtained with GDPNP and D) crystal of *McrB* $\Delta$ NC obtained with GDP. Protein concentration was 10 mg/ml and nucleotide concentration was 1 mM.

positions per hexamer. Due to the poor diffraction quality of the data efforts to locate the selenium position using single anomalous diffraction method failed.

Search for a solution by molecular replacement using other known AAA+ structures also failed. This could be due to the large inter-domain and inter-subunit movements and conformational heterogeneity, which has been observed amongst other AAA+ proteins.

*Table 4.1 Diffraction data-collection statistics of McrB $\Delta$ N+GDPNP crystals*

	Native	Native (synchrotron)	Selenomethionine		
	(in-house)		Peak	Inflection	Remote
<b>Space group</b>	<b>P2<sub>1</sub></b>	<b>P2<sub>1</sub></b>	<b>P2<sub>1</sub></b>	<b>P2<sub>1</sub></b>	<b>P2<sub>1</sub></b>
<b>a (Å)</b>	<b>99.8</b>	<b>89.3</b>	<b>97.5</b>	<b>97.7</b>	<b>97.7</b>
<b>b (Å)</b>	<b>129.2</b>	<b>118.6</b>	<b>129.0</b>	<b>128.9</b>	<b>129.2</b>
<b>c (Å)</b>	<b>155.8</b>	<b>151.7</b>	<b>158.1</b>	<b>157.9</b>	<b>158.2</b>
<b><math>\beta</math> (degrees)</b>	<b>93.6</b>	<b>97.1</b>	<b>95.8</b>	<b>95.7</b>	<b>95.8</b>
<b>Resolution (Å)</b>	<b>20.0 - 8.0</b>	<b>50.0 - 4.5</b>	<b>50.0 - 5.4</b>	<b>50.0 - 5.4</b>	<b>50.0 - 5.4</b>
<b>R<sub>merge</sub></b>	<b>5.6 (30.3)</b>	<b>9.8 (50.8)</b>	<b>7.7 (53.8)</b>	<b>7.0 (89.8)</b>	<b>4.5 (67.7)</b>
<b>Completion</b>	<b>90.7 (80.2)</b>	<b>94.6 (95.7)</b>	<b>99.5 (99.5)</b>	<b>99.6 (99.8)</b>	<b>99.5 (99.8)</b>
<b>I/s (I)</b>	<b>17.7 (4.3)</b>	<b>7.5 (2.2)</b>	<b>8.3 (2.6)</b>	<b>9.6 (2.0)</b>	<b>13.7 (2.4)</b>
<b>No. of molecules in ASU</b>	<b>6 to 12</b>	<b>6 to 12</b>	<b>6 to 12</b>	<b>6 to 12</b>	<b>6 to 12</b>

McrB $\Delta$ NC complexed with both GDPNP and GDP was also crystallized but these crystals diffracted poorly (Figure 4.3 C, D). Significant time and efforts were spent to improve the diffraction quality of these crystals. While the McrB $\Delta$ N+GDPNP crystal diffraction could be improved to 6.5 Å, the crystals obtained with GDP diffracted only to 9 Å.

#### 4.3.4 Self-rotation map

Diffraction data analysis for crystals of McrB $\Delta$ N+GDPNP lead to Matthew's coefficient (Matthews, 1968) of 4.8 Å<sup>3</sup>/Da for 6 molecules in ASU (74.23% solvent content) and 2.39 Å<sup>3</sup>/Da for 12 molecules in ASU with 48.5% solvent content. Since SEC-MALS data (Chapter 2 section 2.3.2.2) clearly showed the hexameric assembly of McrB $\Delta$ N oligomer, a self-rotation map was calculated in P2<sub>1</sub> space group to locate any six-fold in ASU. Self-rotation Patterson map calculated using the MOLREP (Vagin & Teplyakov, 2010) revealed peaks corresponding to non-crystallographic symmetries, in addition to the crystallographic 2-fold arising from monoclinic symmetry (Figure 4.4 A). Twelve peaks were located on the chi=180° self-rotation map section. Two significant peaks were observed in chi=60° section and three in chi=120° section. This observation suggested the occurrence of at least two 6-fold symmetry axes in the ASU and that the two 6-folds are related by 2-fold symmetry.

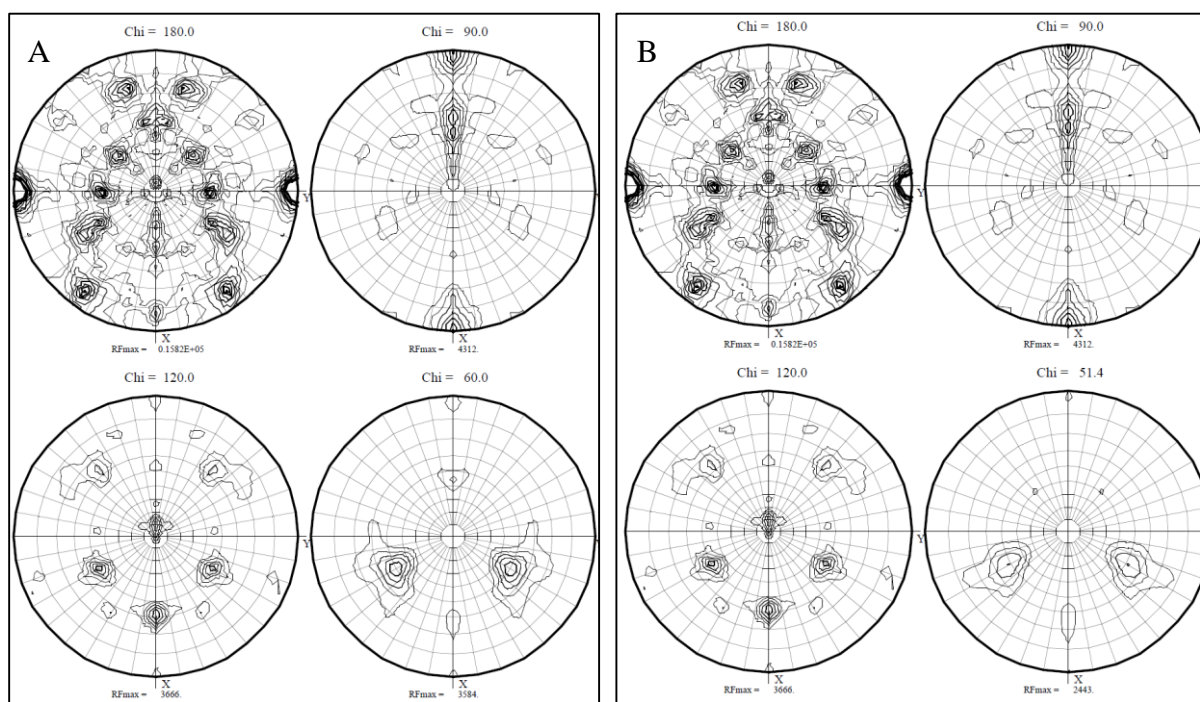


Figure 4.4. **Self-rotation function map showing stereographic projection of McrB $\Delta$ N+GDPNP in space group P2<sub>1</sub>.** Calculation with A) chi=60° show two peaks while in B) calculation with chi=51.4° do not show any significant peak, suggesting a hexameric symmetry in the ASU.

Since McrB was proposed to be a heptamer in earlier studies, a self-rotation map at chi=51.4° was also calculated. No significant peak was noted in chi=51.4° section (Figure 4.4 B). Based on these observations it was concluded that the ASU contains two hexameric rings. The absence of 7-fold symmetry suggested that



McrB $\Delta$ N does not form a heptameric ring, which was consistent with the conclusions of the SEC-MALS analysis and cryo-EM 2D class averages discussed in chapter 2, section 2.3.2.2 and 2.3.6.

#### 4.3.5 Crystallization trials with truncated protein McrB $\Delta$ N<sup>WT</sup> and McrB $\Delta$ N<sup>WT</sup>C complex

Another construct of McrB $\Delta$ N lacking the C-terminal histidine tag (McrB $\Delta$ N<sup>WT</sup>), was generated as an alternative for crystallization efforts. This construct showed biochemical properties of oligomerization in presence of GTP, complexation with McrC in presence of GTP and GTP hydrolysis (stimulated by McrC) similar to the tagged construct (data not shown). The crystals obtained with McrB $\Delta$ N<sup>WT</sup> complexed with GDPNP did not show any improvement in resolution but the McrB $\Delta$ N<sup>WT</sup>C complex bound to GDPNP diffracted better than its histidine-tagged counterpart. McrB $\Delta$ N<sup>WT</sup>C crystals were successfully optimized in three conditions – i) 0.2 M sodium malonate pH 6 and 16% PEG 3350, ii) 0.1 M succinic acid pH 7 and 15% PEG 3350 and iii) 0.2 M sodium citrate tribasic dehydrate and 16% PEG 3350. The crystals obtained with sodium malonate condition diffracted consistently better than the other two. Single-wavelength anomalous dispersion (SAD) data were collected for these crystals grown from selenomethionine derivative of McrB $\Delta$ N<sup>WT</sup>C (only McrB $\Delta$ N<sup>WT</sup> labeled with selenomethionine) (Figure 4.5 A). These crystals will be referred to as Form I. The summary of diffraction data is presented in Table 4.2.

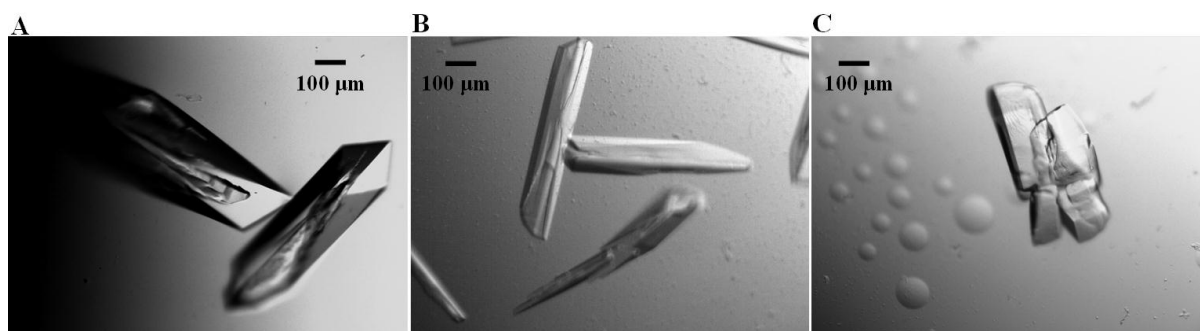


Figure 4.5. **Images of McrB $\Delta$ N<sup>WT</sup>C (McrB $\Delta$ N without histidine-tag) crystals.** A) Crystals of McrB $\Delta$ N<sup>WT</sup>C obtained with Na malonate condition, B) Crystals of McrB $\Delta$ N<sup>WT</sup>C obtained with tacsimate condition in presence of GDPNP C) Crystals of McrB $\Delta$ N<sup>WT</sup>C obtained with butanetriol condition. The protein concentration used for crystallization was 7 mg/ml in presence of 2.5 mM GDPNP

During my efforts to obtain better quality diffraction data, I went back to crystallization screen and a new crystallization hit was obtained with 20% tacsimate pH 7 and 15% PEG 3350. I optimised crystallization in this condition and collected diffraction data for these crystals (Figure 4.5 B). This crystal had unit cell parameter similar to Form I crystals and there was no improvement in their diffraction quality.

*Table 4.2 Diffraction data-collection statistics of McrB $\Delta$ NC+GDPNP crystals*

	<b>Crystal form I</b>	<b>Crystal form II</b>
<b>Space group</b>	<b>P2<sub>1</sub></b>	<b>P2<sub>1</sub></b>
<b>a (Å)</b>	<b>185.7</b>	<b>104.1</b>
<b>b (Å)</b>	<b>145.9</b>	<b>144.8</b>
<b>c (Å)</b>	<b>213.6</b>	<b>186.4</b>
<b><math>\beta</math> (degree)</b>	<b>93.9</b>	<b>91.7</b>
<b>Resolution (Å)</b>	<b>50 - 4.5</b>	<b>50-4.4</b>
<b>R<sub>merge</sub></b>	<b>11.3 (90.0)</b>	<b>9.9 (81.1)</b>
<b>Completion</b>	<b>98.4 (82.5)</b>	<b>99.9 (100 0)</b>
<b>I/<math>\sigma</math> (I)</b>	<b>7.6 (1.4)</b>	<b>16.6 (3.2)</b>
<b>No. of McrBC in ASU</b>	<b>2</b>	<b>1</b>
<b>Mathew's Coefficient</b>	<b>2.89 Å<sup>3</sup>/Da</b>	<b>2.81 Å<sup>3</sup>/Da</b>

Meanwhile, crystallization screen with a new molecular dimension screen MORPHEUS II yielded a third form of McrB $\Delta$ N<sup>WT</sup>C crystal (Form II). The condition consisted of 15% PEG3000, 20% 1,2,4-butanetriol, 1% NDSB 256 (Dimethyl-2-hydroxyethylammoniumpropane sulfonate), 0.1 M MOPSO-Bis-Tris pH 6.5, 0.03 M each of sodium chloride, lithium sulphate and potassium sulphate. With these crystals (Figure 4.5 C), the resolution was improved to 4.4 Å along with a change in crystal form. The SAD data for these crystals is presented in Table 4.2.

Mathew's coefficient for Form I was 2.89 Å<sup>3</sup>/Da (57.39% solvent content) with two McrB $\Delta$ N<sup>WT</sup>C tetradecamers in the ASU. Mathew's coefficient for Form II crystal was 2.81 Å<sup>3</sup>/Da (56.2% solvent content) with a tetradecamer in the ASU. A self-rotation

map calculated for the diffraction data from Form I or Form II crystals did not yield any information about the symmetry of the molecules in the ASU. Incorporation of selenium was judged from X-ray fluorescence scan of the crystal at the synchrotron. In Form I crystals only McrB had selenomethionine and there were 196 of them in the ASU. Form II had selenomethionine in both McrB and McrC and there were 112 of them in the ASU. A complete anomalous diffraction data was collected at the energy edge of selenium for both the crystals. Attempts were made to locate the position of selenium. Different programs including Phaser EP (McCoy et al, 2007), AUTOSOL (Terwilliger et al, 2009) and CRANK2 (Skubak & Pannu, 2013) were used to locate the positions. Selenium positions thus obtained were used to calculate phases. Density modified maps were calculated using the phases. None of the maps gave interpretable electron density. This could be because of failure to obtain the correct solution for selenium positions.

#### **4.3.6 The dumbbell-shaped density of McrB $\Delta$ NC**

While crystallographic studies were being carried out, we initiated cryo-EM studies on McrBC. This study resulted in three different maps. All the three maps at resolutions 7.4 Å, 8.5 Å and 9.8 Å provided a clear picture of the architecture of McrBC complex. The EM map revealed a double-tiered arrangement of two rings bridged by a neck resembling a dumbbell of length ~150 Å (Figure 4.6 A). The ring-like density had a diameter of ~105 Å and appeared to be made up of six segments (Figure 4.6 A, B). The two rings were interpreted to correspond to the oligomers of McrB $\Delta$ N, while the density bridging them to that of McrC. The double-tiered structure of stacked hexameric rings is not uncommon and has been observed in type I AAA+ proteins, such as MCM and SV40 Large T antigen (SV40 LTAG), in the form of a dodecamer. The similar stacked arrangement of rings is seen in type II AAA+ proteins that have two AAA+ domains in a single polypeptide. However, unlike the previous examples, the rings of McrB do not stack against each other but instead sandwich the nuclease McrC. Interestingly, this arrangement of the AAA+ rings sandwiching the nuclease is reminiscent of chaperone-protease assemblies, such as the eukaryotic 19S-20S proteasome complex, archaeal PAN-20S proteasome complex, and the bacterial ClpA-ClpP complex (Bedford et al, 2010; Carroni et al, 2014; Medalia et al, 2009). In all these complexes, the proteases are sandwiched between two parallel

rings of the AAA+ chaperones. In McrBC, the two rings are not parallel to one another but are tilted with respect to each other by about  $26^\circ$  (Figure 4.6 C, D).

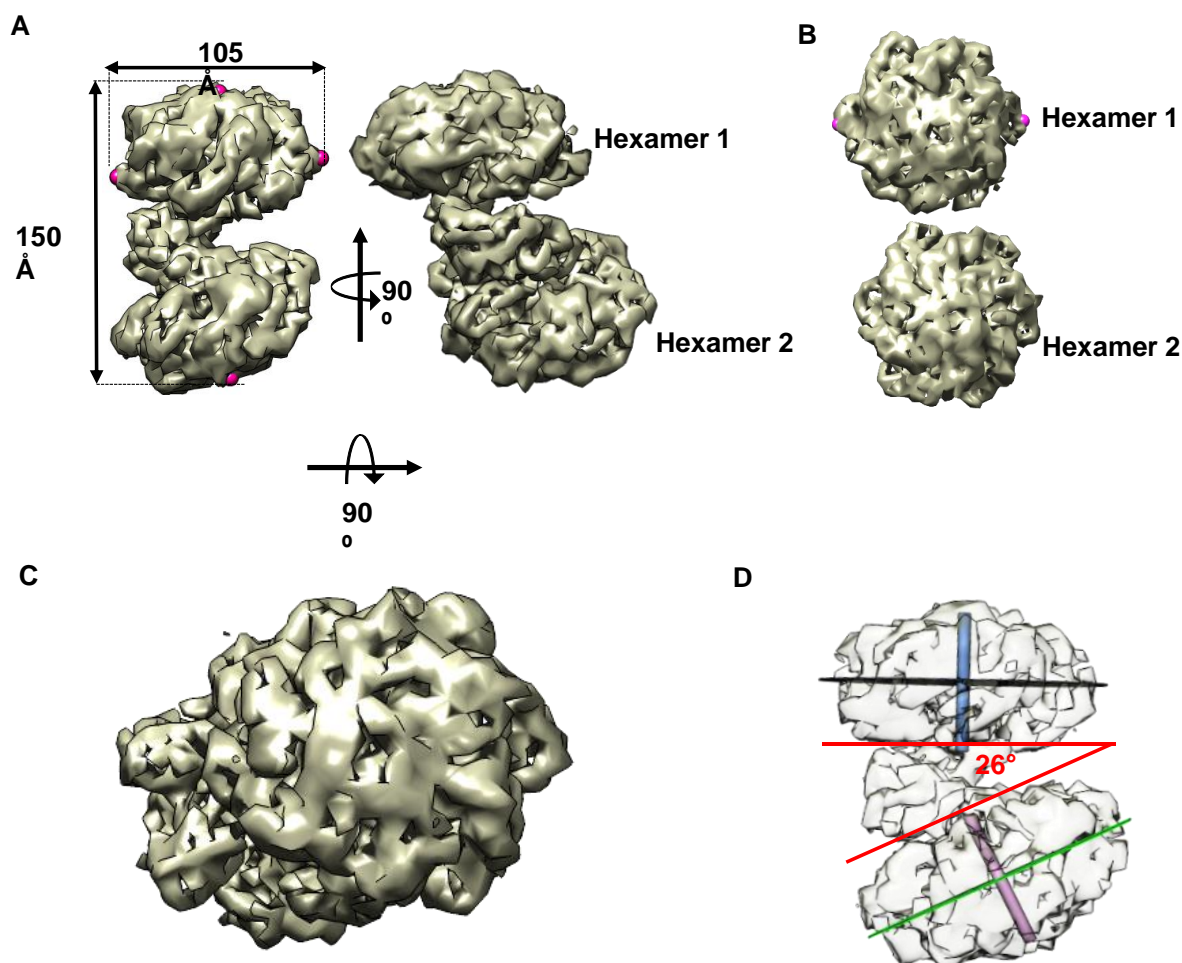


Figure 4.6. **Architecture of McrB $\Delta$ NC complex.** A) Front and side view of dumbbell shaped 7.4 Å electron density map of McrB $\Delta$ NC complex. B) Top view of both hexamers separated from the complex, showing density for six subunits. C) Top view of McrB $\Delta$ NC complex showing both hexamer 1 along with hexamer 2 indicating a staggered arrangement of the two hexamers. D) An image of McrB $\Delta$ NC complex showing the angle between the two hexamers measured as the angle between the planes passing through the hexameric rings. The planes were generated and angles calculated using Chimera

Comparison of the 8.5 and 9.8 Å maps revealed differences in the orientation of the two hexamers with respect to each other. Superposition of the densities of hexamer 2 from the three map clearly showed the difference in orientation of hexamer 1 (Figure 4.7 A, B). This inherent conformational plasticity of McrBC hindered our efforts to obtain a higher-resolution EM map, despite the addition of more particles for averaging.

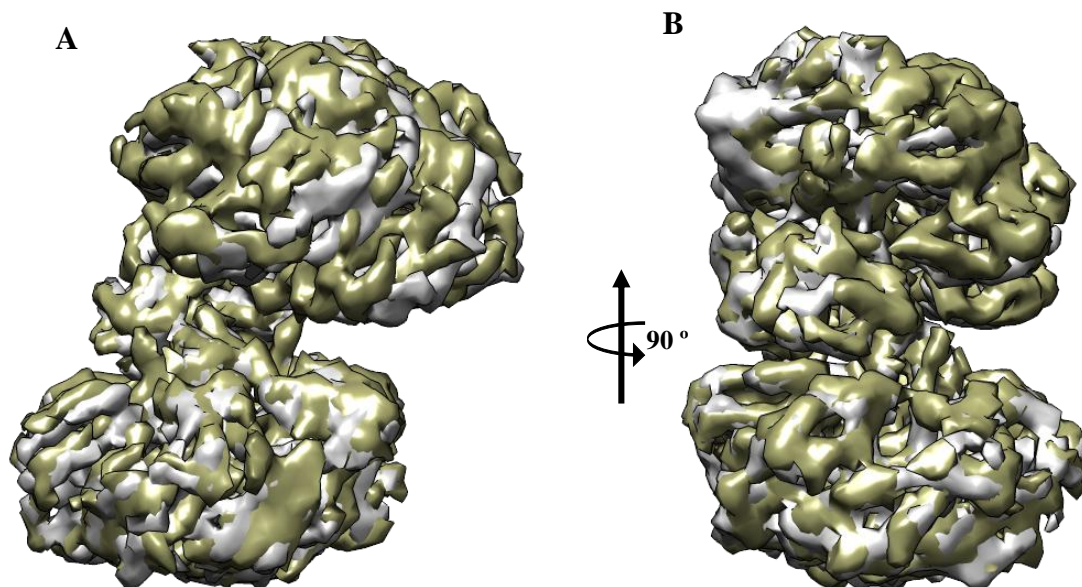


Figure 4.7. **Superimposition of two McrB $\Delta$ NC maps.** A) Front and B) side view of superimposed 9.8 Å (Grey) and 8.5 Å (Olive) maps showing the conformational differences between hexamer 1 of the two maps. The 9.8 Å map was superimposed on hexamer 2 of 8.5 Å map with a correlation of 0.62 and images were prepared using Chimera.

#### 4.3.7 Modeling the structure of McrB $\Delta$ N hexamers

The quality of the 7.4 Å map of McrB $\Delta$ NC was sufficient to visualize many of the secondary structures. This gave us the confidence to build a model of the assembly. The modeling involved two stages – (i) building a model of the two hexameric McrB $\Delta$ N rings, and (ii) building a model of McrC. This section discusses the results of my effort to model the McrB $\Delta$ N hexamer. The boundaries of the six subunits that constitute a McrB $\Delta$ N oligomer could be visually identified in the map. Also, the densities corresponding to the N-terminal and the C-terminal domains of each of the AAA+ folds could be delineated. The characteristic structure of the N-terminal domain (NTD) of a AAA+ fold having a  $\beta$ -sheet sandwiched between  $\alpha$ -helices was clearly visible, and so was the largely  $\alpha$ -helical C-terminal domain (CTD).

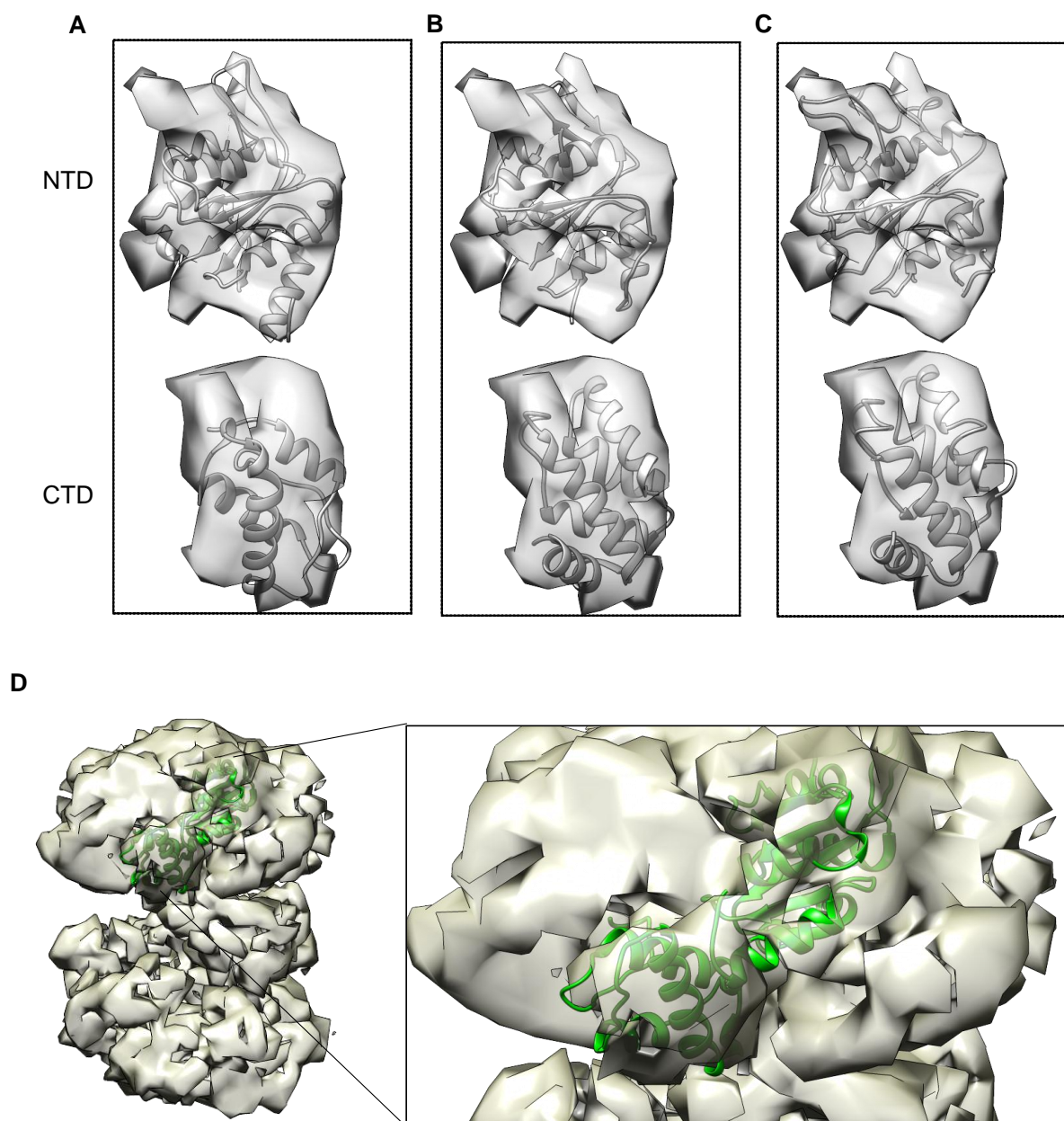


Figure 4.8. **Building the model for McrB $\Delta$ N subunit.** Images showing the fitting of different molecular models - A) PspF (PDB ID 2C96), B) CsoCbbQ (PDB ID 5C3C) and C) molecular model generated *in silico* using homology modeling (see Results). D) Image of a complete McrB $\Delta$ N<sup>WT</sup> model fit into electron density of one subunit of McrB $\Delta$ N<sup>WT</sup>C complex. Inset shows the magnified view highlighting the goodness of fit of the model into the electron density. The images were prepared in Chimera.

The features of the density facilitated modeling of McrB $\Delta$ N through docking. Based on amino acid sequence analysis, McrB $\Delta$ N has been grouped along with the bacterial enhancer binding protein (bEBP) into clade VI of AAA+ superfamily

(Erzberger & Berger, 2006). Therefore, for docking, I initially started with the structure of a subunit of the bEBP PspF (PDB ID: 2C96). The AAA+ fold of PspF was individually docked onto the density of each subunit of McrB $\Delta$ N that was delineated as described above, and the NTD and the CTD rigid-body fitted using COOT (ref). An inspection of the fitted coordinates showed that the CTD did not fit as well as the NTD (Figure 4.8 A). One of the striking differences was that the density for the CTD of the subunits of McrB $\Delta$ N indicated the presence of five helices, while the CTD of PspF, as in the case of most other AAA+ proteins, has only four.

As an alternative, homology models of McrB $\Delta$ N were generated using the I-TASSER server (Roy et al, 2010; Yang & Zhang, 2015; Zhang, 2008). One of the templates identified by the program for the model building was the AAA+ protein Cso-CbbQ (PDB ID: 5C3C). It was noticed that the C-terminal domain of Cso-CbbQ was made of five helices and that this domain fit well into the density. The N-terminal domain of Cso-CbbQ also fit reasonably well into the density (Figure 4.8 B). Despite McrB $\Delta$ N and Cso-CbbQ share a relatively low sequence identity of about 29%, and the two proteins are grouped into different clades of AAA+ family (Cso-CbbQ is included in Clade VII), Cso-CbbQ was found to be a good structural template for McrB $\Delta$ N. I used I-TASSER to homology model the N-terminal domain of McrB $\Delta$ N and Swiss-Model (Arnold et al, 2006; Biasini et al, 2014; Bordoli et al, 2009) to model the C-terminal domain using Cso-CbbQ as the template. Using these homology models, the structure of McrB $\Delta$ N was built by docking and rigid-body fitting the two domains in each of the protomer density (Figure 4.8 C & 4.9). In addition, a rigid body fit for the Loop 1 (residues 245 to 255) of the N-terminal domain was also done as it was significantly outside the density.

Despite the quality of map being good, the lower resolution of the data prevented us from locating the nucleotides (GDPNP) bound to the McrB $\Delta$ N hexamers with confidence.

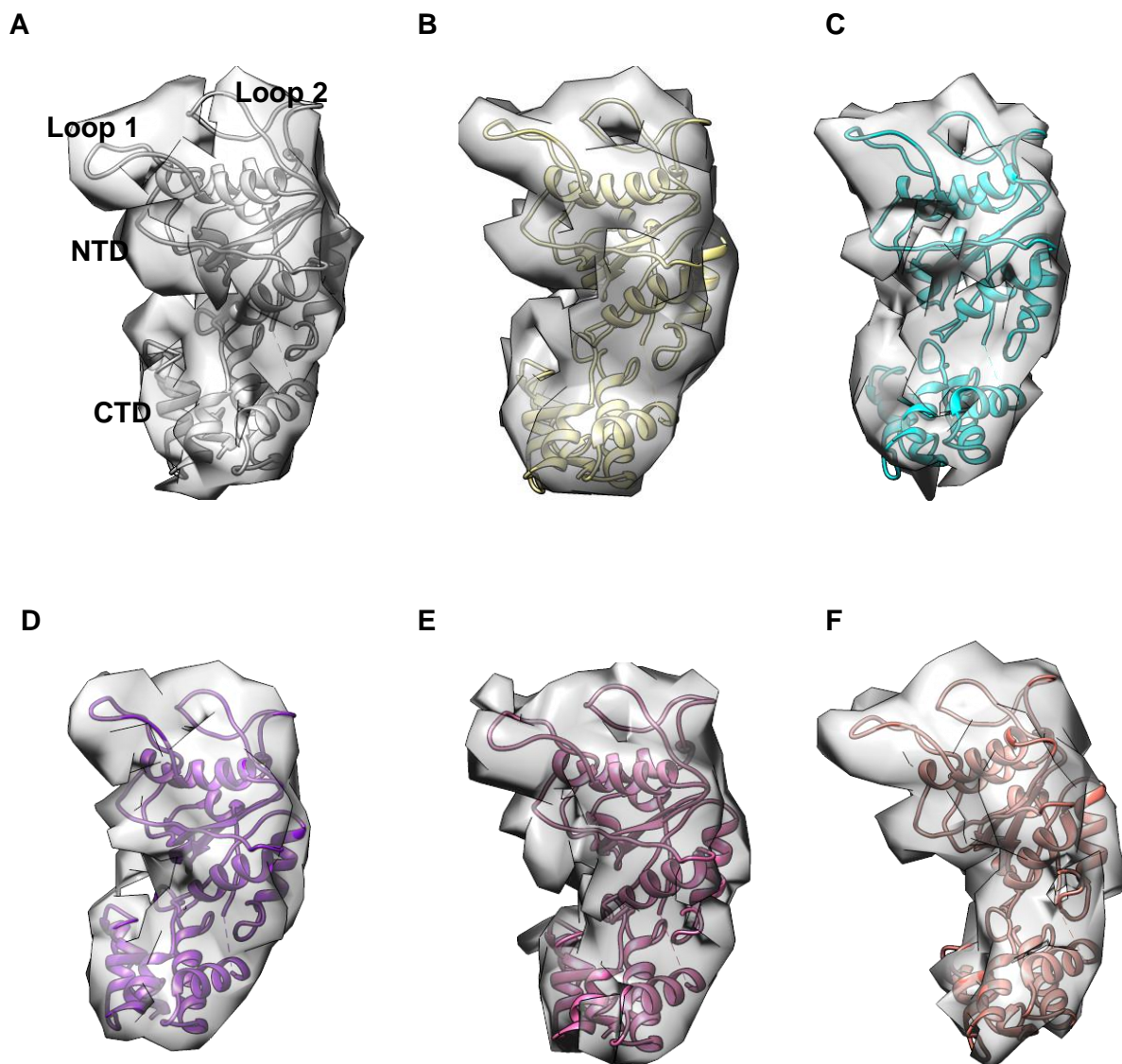


Figure 4.9. **Goodness of fit between model and electron density map.** Images showing the overall fitting of the final model- N-terminal domain (NTD), C-terminal domain (CTD) and loops, Loop 1 and Loop 2,- into segmented electron density corresponding to the six protomers of hexamer 1.

#### 4.3.8 Architecture of the *McrB $\Delta$ N* hexamer

As discussed above, the model of the two domains of *McrB $\Delta$ N* could be fit well into the density corresponding to each subunit. The density of Loop 1 and Loop 2, which are characteristic of Clade VI of AAA+ family, was clearly visible. Both the hexamers form a ring-like structure, in which the subunits are related not by a perfect but rather by a pseudo 6-fold symmetry. The angle of rotations relating the neighboring segments



of density corresponding to the individual subunits varies from  $56^\circ$  to  $68^\circ$ . These values were obtained by superimposing the neighboring segments using Chimera. The quality of superimposition was judged by cross-correlation values, which was better than 0.93.

The deviation from 6-fold symmetry appears to be the result of differences in the interface formed between neighboring subunits. Visually clear gaps could be seen at interfaces between some of the neighboring subunits. The modeled structures of the two McrB $\Delta$ N hexamers were used to calculate the area buried between the six different interfaces that exist per ring using the software QtPISA (Krissinel, 2010; Krissinel & Henrick, 2007) (Figure 4.10 A). Three of the subunit interfaces in each ring had a buried surface area approximately equal to or greater than  $1500 \text{ \AA}^2$ . The other three interfaces had buried surface area less than  $1200 \text{ \AA}^2$ . The calculation included side chain positions as obtained from homology modeling. Consequently, the buried surface area calculated is an approximate indicator of the interface contacts between the subunits.

For convenience, the subunits of the hexamer were named A to F in the clockwise direction when viewed from the top (Figure 4.10 A). The subunit that transitioned from being part of an interface with the lower buried surface area to an interface with the higher buried surface area was called A. I found that the centroid of each of the subunit did not deviate much from the plane of the ring (Figure 4.10 B). This suggested that the subunits did not adopt a spiral arrangement, which has been noted in case of many other AAA+ proteins (Huang et al, 2016; Yokom et al, 2016). Apart from differences in the interface area, the subunits also showed variation in inter-domain orientation. Superimposition of NTD of all the subunits of McrB $\Delta$ N hexamer showed the inter-domain movement of the CTD (Figure 4.10 C). However, no direct correlation could be deduced between the change in inter-domain orientation with that of either the interface buried surface area or the location of McrC (see below).

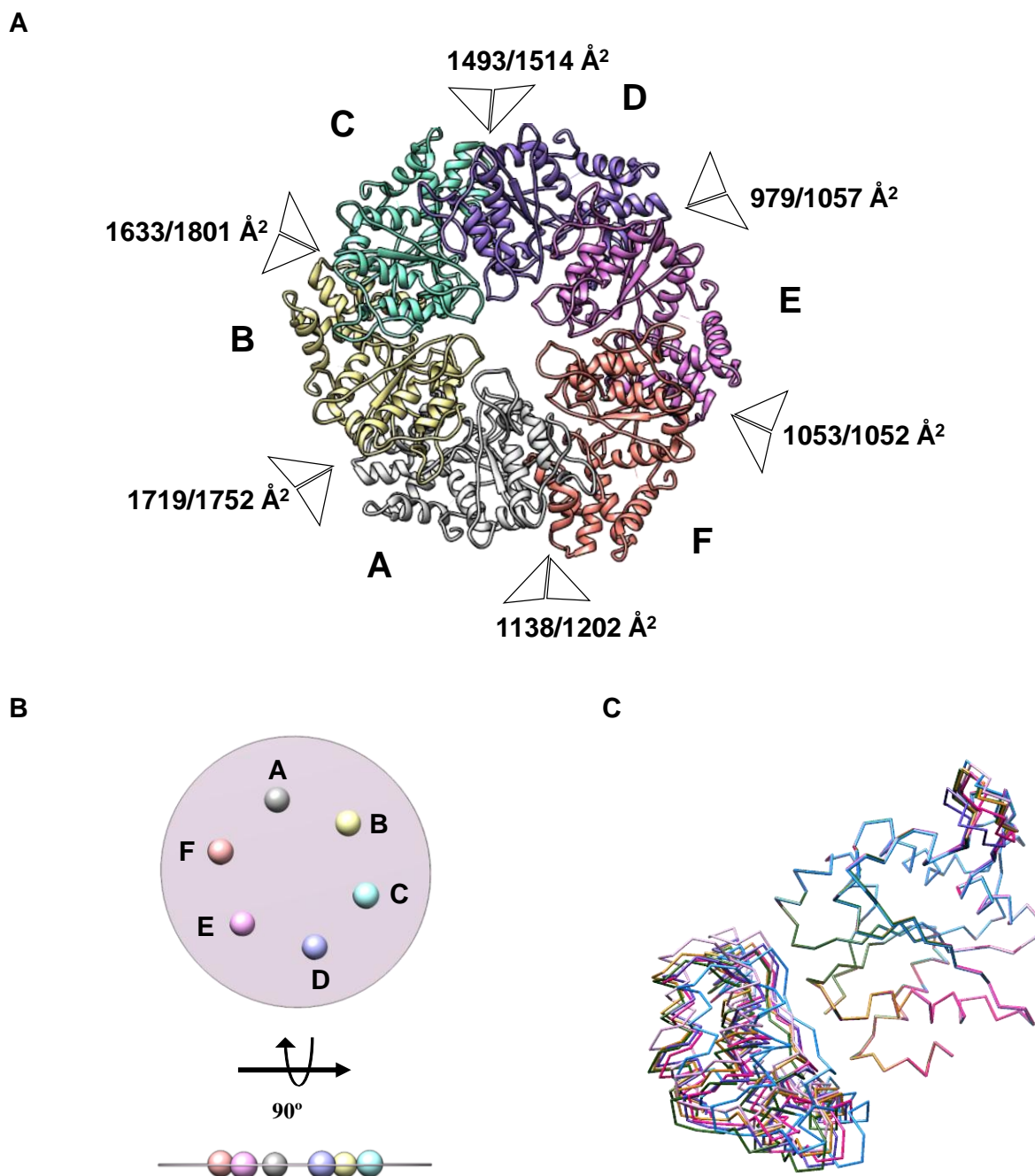


Figure 4.10. **The asymmetric interface of McrB $\Delta$ N hexamer.** A) Segmented electron density of corresponding to the subunits of McrB $\Delta$ N highlighting the asymmetry in the interface of different subunits due to variable buried surface area. The buried surface area at the subunit interfaces for both hexamer 1 and hexamer 2 are given. B) Top view and side view of a plane passing through NTD of different subunits and their centroids indicating lack of a spiral opening in the ring. C) Alignment of NTD of different subunits showing conformational heterogeneity of CTD. The figure also highlights the conformational flexibility of Loop 1, which was rigid body fitted into the density individually for each subunit.

#### 4.3.9 Model of McrC

Modeling the two hexameric rings of McrB $\Delta$ N allowed to clearly delineate the density corresponding to McrC (Figure 4.11 B). The density appeared to have 2-fold symmetry (Figure 4.11 C), possibly corresponding to a dimer of McrC. The density could be divided into three segments, a segment that appeared to be made of  $\alpha$ -helices sandwiching a  $\beta$ -sheet, a second segment corresponding to all  $\alpha$ -helical densities and from which jutted out a shaft-like long and thin density. The  $\alpha$ - $\beta$ - $\alpha$  segment formed the dimeric interface (Figure 4.11 D).

As there is no structure of a close homolog of McrC available, as a first step to model the structure of McrC I carried out a sequence based secondary structure prediction of McrC (Figure 4.11 A). The prediction indicated that the secondary structure till first 180 amino acids was predominantly  $\alpha$ -helical, while the remaining C-terminal residues were formed of the  $\alpha$ - $\beta$ - $\alpha$  structure. Amidst the  $\alpha$ -helices at the N-terminus, there was a long stretch of residues (from residue number 60 to 100) that appeared to take an extended conformation to possibly form a long loop corresponding to the shaft-like density. As the conformation of the loop could not be ascertained from the density, this region was left unstructured. Based on the density, the helices could be modeled in the all  $\alpha$ -helical segment, which will be referred to as the helical domain. However, the polarity of the helices and the connectivity between them could not be assigned unambiguously.

*In silico* homology modeling of the N-terminal sequence did not yield a model resembling the arrangement of the helices that was obtained through interpretation of the density. Modelling of the C-terminal  $\alpha/\beta$  region of McrC using I-TASSER yielded a model with a hypothetical protein from *Vibrio cholerae* (PDB id: 1XMX) as the template (C-score -3.5). This model could be docked on to the  $\alpha$ - $\beta$ - $\alpha$  segment and rigid body fit using COOT.

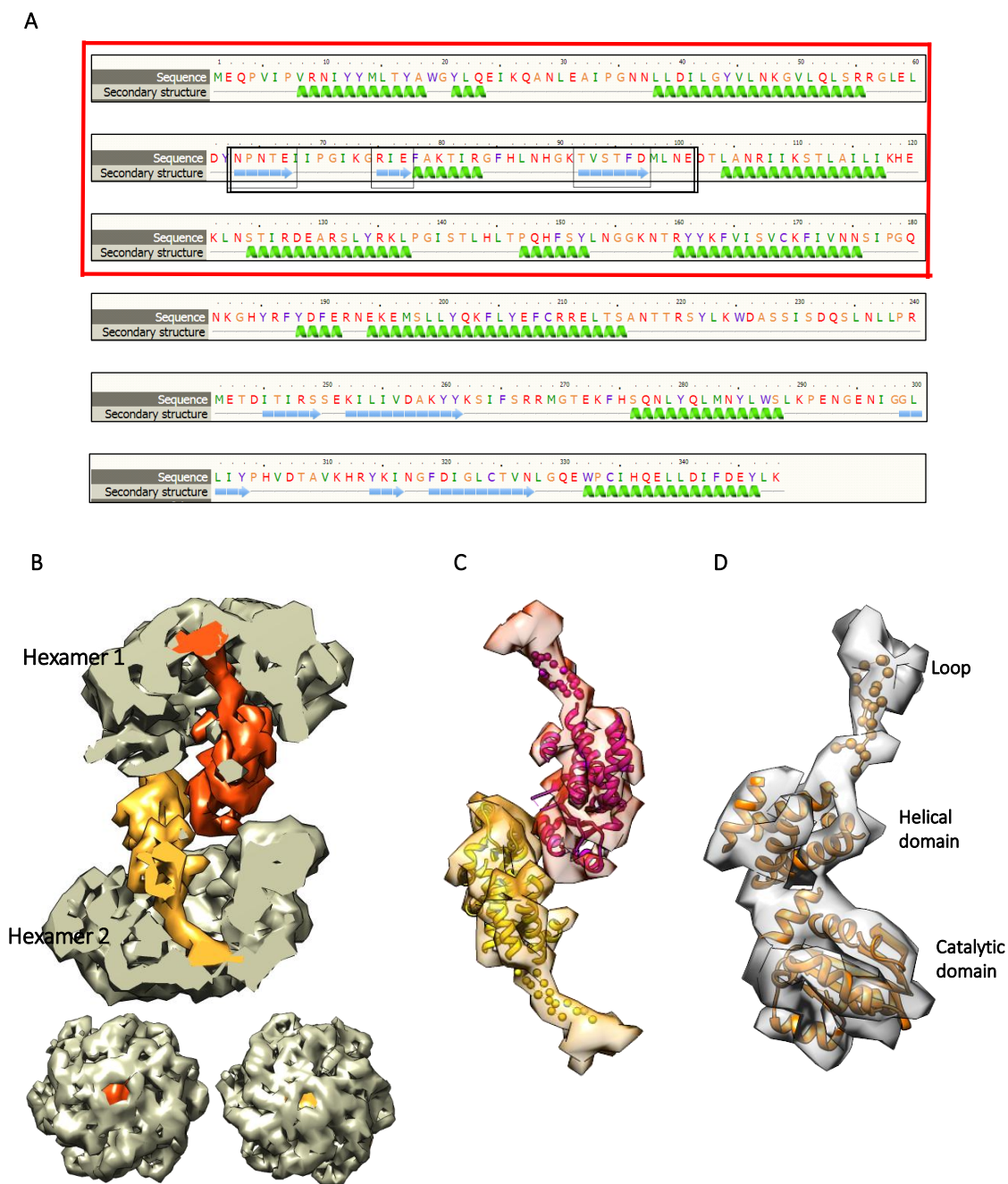


Figure 4.11. **Building a model for McrC.** A) Secondary structure prediction of McrC using Phyre server. Red box indicates the N-terminal helical domain of McrC, the region in black box highlight the possible extended loop residues. B) Cut-through side view of McrB $\Delta$ N<sup>WT</sup>C showing the density for McrC. The hexamer 1 and hexamer 2 top views indicate that the central pore of McrB $\Delta$ N<sup>WT</sup> ring is blocked by McrC B) Segmented electron density of McrC dimer showing a 2-fold symmetry. C) Fitting of McrC model in catalytic domain and helices in helical domain. Spheres in the loop region were placed due to lack of information about loop conformation at this resolution. All the images were prepared in Chimera

#### **4.3.10 Architecture of McrC**

The model of McrC allowed us to study its architecture. Based on the density, two subunits of McrC related by 2-fold symmetry could be modeled. McrC belongs to the PD...D/EXK family of nucleases (Pieper & Pingoud, 2002), which are generally found to be dimers (Kosinski et al, 2005). Each subunit of McrC comprises of three domains – the N-terminal helical domain, the shaft, and the C-terminal  $\alpha$ - $\beta$ - $\alpha$  domain. The  $\alpha$ - $\beta$ - $\alpha$  structure is consistent with other members of the PD...D/EXK family of nucleases, such as FokI, which have their catalytic domain made of  $\alpha$ -helices sandwiching a  $\beta$ -sheet. Henceforth, the  $\alpha$ - $\beta$ - $\alpha$  domain will be referred to as the catalytic domain. The catalytic domain forms the dimeric interface (Figure 4.11 D). As mentioned above, the fold of the helical domain and the structure of loop could not be ascertained.

#### **4.3.11 Interaction of McrC with McrB $\Delta$ N**

The architecture of the complex revealed that the helical domain and the shaft from each subunit of McrC interacted with the subunits of the hexameric McrB $\Delta$ N to form the dumbbell structure. The shaft inserts into the pore of the hexameric ring from the bottom (Figure 4.11 B). The C-terminal catalytic domain of McrC did not interact with McrB $\Delta$ N. The interaction between McrC and the subunits of the McrB $\Delta$ N hexamer was found to be asymmetric (Figure 4.12 and 4.13). Amongst the six subunits, the interaction of McrC with the C subunit was most extensive (Figure 4.12 and 4.13). The helical domain of McrC interacted with both NTD (approximately around residue 335 to 338) and CTD (approximately around residue 437 to 439). The shaft interacted with Loop 1 and the region preceding it. The helical domain also interacted with the region around residues 335-338 and 437-439 of subunit D, while the tip of the shaft interacted with the tip of Loop 1. The subunits A, B, E and F, had little interaction with McrC except for the interaction of the tip of Loop 1 with the tip of the shaft (Figure 4.13). The interaction of the tip of the shaft with the tip of Loop 1 of all the six subunits of McrB $\Delta$ N blocks the mouth of the pore of the ring (Figure 4.11 B).

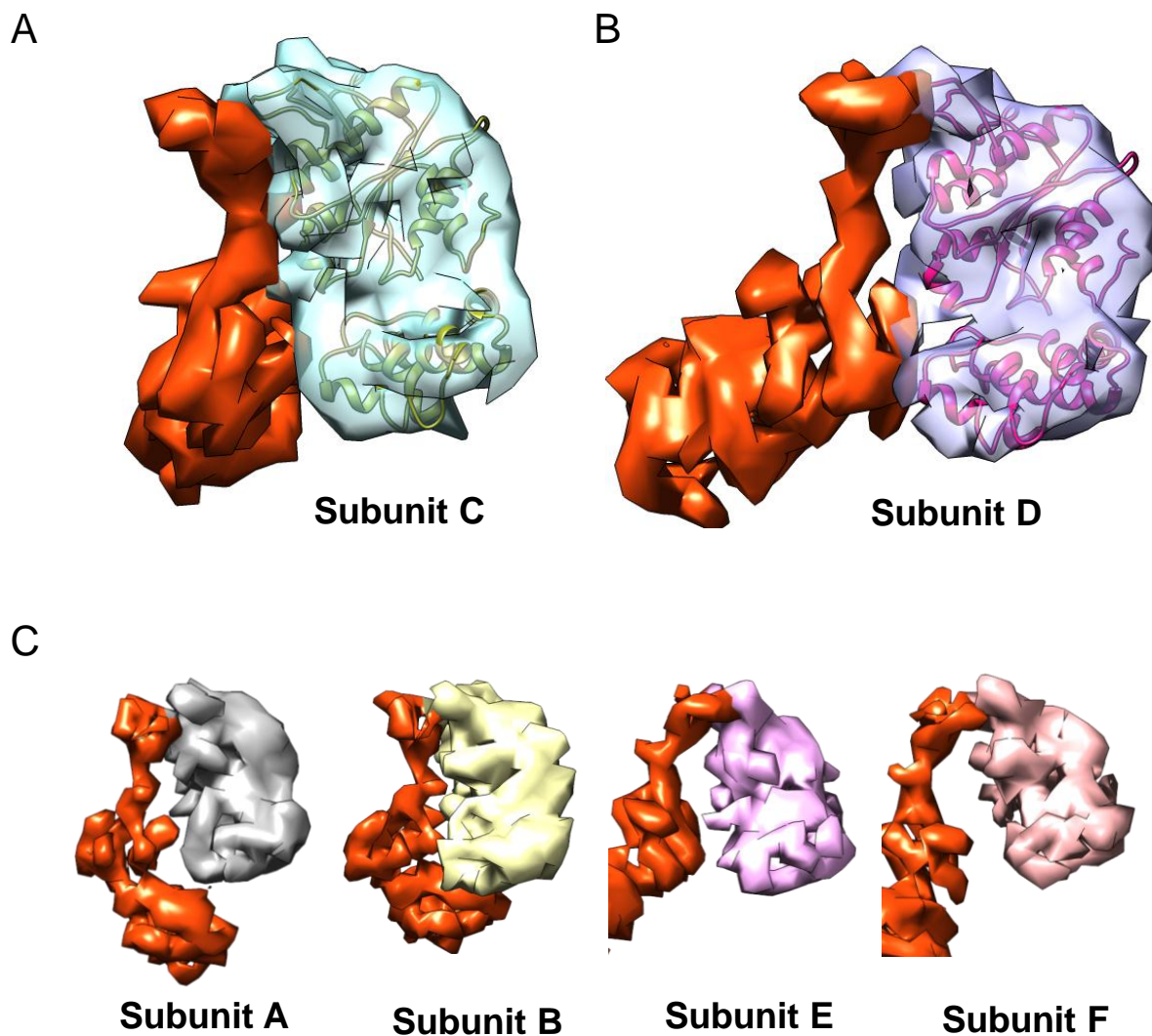


Figure 4.12. **Interaction of McrC with different subunits of McrB $\Delta$ N<sup>WT</sup>.** Segmented electron density of McrC (Orange) showing A) interaction with subunit C via its loop and helical domain, B) interaction with subunit D via the tip of the loop and helical domain and C) interaction with other subunits via tip of the loop.

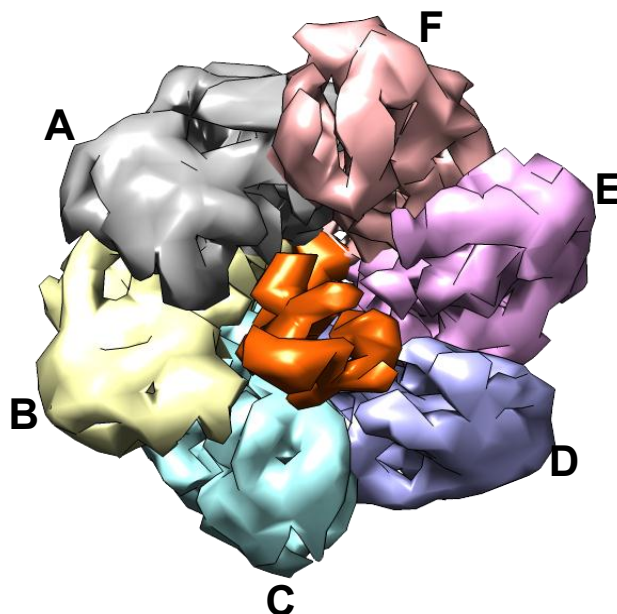


Figure 4.13. **Interaction of McrC with different subunits of McrB $\Delta$ N.** Bottom view of the hexamer showing interaction of helical domain of McrC (orange) with different subunits. The helical domain interacts only with subunit C and D.

#### 4.4 Discussion

In this chapter, I have presented the architecture of the McrB $\Delta$ NC complex. This study provides unambiguous proof that McrB $\Delta$ NC is made of two hexameric rings of McrB $\Delta$ N bridged by a dimer of McrC. This assembly is consistent with the results from SEC-MALS analysis, which revealed that both McrBC and McrB $\Delta$ NC are made of twelve subunits of McrB/McrB $\Delta$ N and two subunits McrC. The density map and the structure modeled based on the map revealed a pseudo 6-fold symmetrical arrangement of the six subunits in each hexamer and highlighted its asymmetric interaction with McrC. This architecture is reminiscent of the assembly of the  $\gamma$  subunit inside the F1 ATPase catalytic subunit.

##### 4.4.1 *A model for McrC-stimulated GTPase activity of McrB*

The structure of McrB $\Delta$ NC assembly provides insights into a possible mechanism of McrB GTPase activity and its stimulation by McrC. In the complex, interfaces AB, BC, and CD have higher buried surface area ( $>1500 \text{ \AA}^2$ ) than DE, EF and FA ( $<1200 \text{ \AA}^2$ ). Studies on AAA+ proteins, E1 protein from bovine papillomavirus, Hsp104, HslVU and

PspF, have shown nucleotide-dependent inter-subunit movement resulting in higher buried surface area at interfaces where the nucleotide is bound (Enemark & Joshua-Tor, 2006; Taylor et al, 2012; Wang et al, 2001; Wendler et al, 2009). Accordingly, in McrB $\Delta$ N hexamer, although the density did not reveal the presence of nucleotide, it was assumed that the interfaces AB, BC and CD having higher buried surface area to be GDPNP bound, as the experiment was performed in presence of GDPNP.

Previous studies found that McrB has a higher affinity for GTP than GDP (Pieper et al, 1997), leading to the conclusion that the stimulation is a result of McrC promoting catalysis rather than modulation of GDP-GTP exchange. The structure reveals that McrC makes extensive contact only with the C and D subunits of McrB. As discussed above, the CD interface is expected to be GTP bound. Consequently, the interaction of McrC with subunits C and D, I propose, would promote hydrolysis. The structure suggests that the helical domain of McrC interacts with a region corresponding to residues 335-338 of McrB. It has previously been shown that mutation of the R337 of McrB increases rate of GTP hydrolysis (Pieper et al, 1999), mimicking the situation when McrB is complexed with McrC. It is possible that McrC remodels this region to stimulate GTP hydrolysis. The nature of the remodeling is not clear from the density.

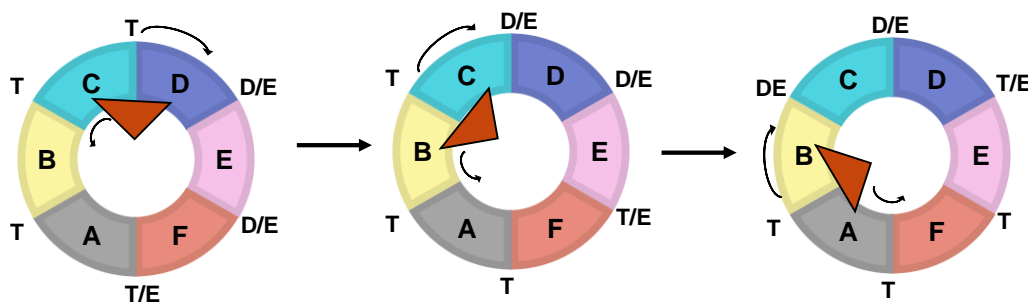


Figure 4.14. **Model for sequential GTP hydrolysis by McrB $\Delta$ NC** Each subunit is depicted by different blocks color coded as in figure 5.9. The subunits are labelled at the center of colored blocks while the three possible states- T- GTP bound, D/E- GDP bound or Empty and T/E- GTP bound or empty- are mentioned at the interfaces. McrC (orange triangle at centre) interacts asymmetrically with hexamer making additional contacts at CD interface for GTP hydrolysis. After hydrolysis, McrC moves to next nucleotide bound site i.e the BC interface while CD interface becomes empty or ADP bound.



Based on the asymmetric interaction of McrC and variable interface indicating different nucleotide occupancy, I propose a sequential GTP hydrolysis model for the McrBC motor (Figure 4.14). According to this model, the CD interface, interacting with McrC, is the hydrolysis activated state (T). Upon GTP hydrolysis, the CD interface would convert to an interface with ADP bound (D) or empty (E) state. With the current data, it cannot be ascertained whether the DE and EF interfaces are empty or ADP bound. I propose that following these states the AF interface would be ready to bind to incoming GTP thus being in a T/E state. This would lead to the formation of the GTP bound AB interface. For sequential hydrolysis of GTP, McrC will have to rotate in a direction opposite to the direction of GTP hydrolysis (Figure 4.14).

#### 4.4.2 The architecture of McrB $\Delta$ NC shows lack of putative Sensor 2

In chapter 3, I tried to locate the Sensor 2 motif in McrB. In the model for McrB $\Delta$ NC, it was observed that the proposed Sensor 2 arginine in the CTD is actually located far away from the P-loop (Figure 4.15). Instead, two polar residue serine and histidine are positioned close to the P-loop. In the sequence alignment of McrB with its homologs (Chapter 3 section 3.3.1), these residues were conserved instead of the arginine. Thus, McrB might have a non-canonical Sensor 2 having residues serine or histidine instead of arginine.

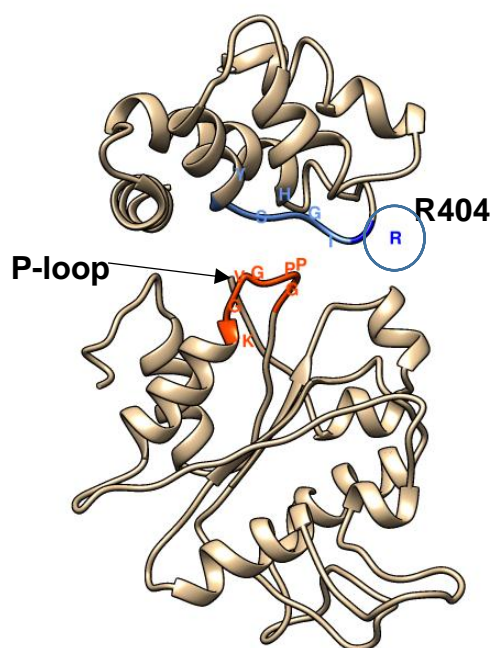


Figure 4.15. **Position of proposed Sensor 2 residue R404.** The model for subunit A showing the position of Arg404 which was proposed to be the Sensor 2 of McrB in Chapter 3. The model shows the position of the residue away from the P-loop thus explaining no effect of mutagenesis.

#### **4.4.3 Implication of McrB $\Delta$ NC architecture on probable modes of DNA loading**

Hexameric ring helicases have been shown to interact and translocate their DNA substrates via loops projecting at the center of the pore (Abid Ali et al, 2016; Gai et al, 2016; Georgescu et al, 2017). Different modes of assembly have been observed for different helicases. For example, the hexameric ring of SV40 LTag and Papillomavirus E1 protein have been shown to assemble around the origin DNA, while eukaryotic MCM which exist as a split ring loads with the help of the helicase loader ORC/Cdc6. Whether McrBC disassembles and then assemble around DNA or the rings of McrB open in presence of DNA need to be deciphered with further studies.

The current structure of McrB $\Delta$ NC, observed in this work, however, adds a complication for possible McrBC DNA interaction. The complex has a closed architecture with the entry of the pore blocked by McrC. For the DNA to thread through the pore, either by assembling around DNA or by ring opening, will require dislocation of the McrC shaft. Since McrC dislocation does not seem favorable for maintaining McrB and McrC interactions that appear important for coupled GTPase and nuclease activity, the only possibility in such a scenario will be opening of the ring wide enough to accommodate both DNA and McrC. Such a process would require a large rearrangement of the complex. The other possibility for McrBC and DNA interaction can be via wrapping of DNA around the molecule without being threaded into the central pore. With the current information, these models cannot be substantiated and further studies are required to dissect out the mode of DNA loading by this translocase.

To summarise, in this chapter, I showed the architecture of McrB $\Delta$ NC complex showing an asymmetric hexameric ring along with the asymmetric interaction of McrC with different subunits. Based on these observations, I proposed a sequential rotary GTPase mechanism for McrB $\Delta$ NC. I also showed that McrC plugs the pore of the McrB ring. With the knowledge of its molecular architecture, I now initiated stopped-flow kinetic studies of the complex to understand the kinetic mechanism of GTPase activity of McrBC and its DNA loading. The final chapter of this thesis describes my efforts to observe and decipher the kinetics of McrBC motor in real time.

**4.4.4 Structure determination using crystal diffraction data:** In this chapter I also described crystallographic studies on McrB $\Delta$ N and McrB $\Delta$ NC, which allowed me to obtain higher resolution diffraction data sets for both the systems. I proceeded with

using the models of McrB $\Delta$ N hexameric ring and McrB $\Delta$ NC obtained from the cryo-EM map for molecular replacement (MR) by PHASER (McCoy, 2007). Unambiguous MR solutions were obtained for the different data set. Using the calculated phases obtained using the MR solutions, we plan to identify the positions of seleniums in the SAD data sets using anomalous difference Fourier map. We plan to combine the experimental phases obtained from the selenium positions with the partial calculated phases obtained from the MR solution to determine the crystal structures of McrB $\Delta$ N and McrB $\Delta$ NC. A detailed comparison of the two structures will provide us with the mechanism of McrC-stimulated GTP hydrolysis by McrB.

## References

Abid Ali F, Renault L, Gannon J, Gahlon HL, Kotecha A, Zhou JC, Rueda D, Costa A (2016) Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. *Nature communications* **7**: 10708

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**: 195-201

Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica Section D, Biological crystallography* **67**: 271-281

Bedford L, Paine S, Sheppard PW, Mayer RJ, Roelofs J (2010) Assembly, structure, and function of the 26S proteasome. *Trends in cell biology* **20**: 391-401

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research* **42**: W252-258

Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T (2009) Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* **4**: 1-13

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72**: 248-254

Carroni M, Kummer E, Oguchi Y, Wendler P, Clare DK, Sinning I, Kopp J, Mogk A, Bukau B, Saibil HR (2014) Head-to-tail interactions of the coiled-coil domains regulate ClpB activity and cooperation with Hsp70 in protein disaggregation. *eLife* **3**: e02481

Enemark EJ, Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**: 270-275

Erzberger JP, Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annual review of biophysics and biomolecular structure* **35**: 93-114

Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? *Acta crystallographica Section D, Biological crystallography* **69**: 1204-1214

Gai D, Wang D, Li SX, Chen XS (2016) The structure of SV40 large T hexameric helicase in complex with AT-rich origin DNA. *eLife* **5**

Georgescu R, Yuan Z, Bai L, de Luna Almeida Santos R, Sun J, Zhang D, Yurieva O, Li H, O'Donnell ME (2017) Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. *Proceedings of the National Academy of Sciences of the United States of America* **114**: E697-E706

Grant T, Grigorieff N (2015) Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**: e06980

Huang R, Ripstein ZA, Augustyniak R, Lazniewski M, Ginalski K, Kay LE, Rubinstein JL (2016) Unfolding the mechanism of the AAA+ unfoldase VAT by a combined cryo-EM, solution NMR study. *Proceedings of the National Academy of Sciences of the United States of America* **113**: E4190-4199

Kabsch W (2010) Xds. *Acta crystallographica Section D, Biological crystallography* **66**: 125-132

Kosinski J, Feder M, Bujnicki JM (2005) The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC bioinformatics* **6**: 172

Krissinel E (2010) Crystal Contacts as Nature's Docking Solutions. *Journal of computational chemistry* **31**: 133-143

Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**: 774-797

LeMaster DM, Richards FM (1982) Preparative-scale isolation of isotopically labeled amino acids. *Analytical biochemistry* **122**: 238-247

Matthews BW (1968) Solvent content of protein crystals. *Journal of molecular biology* **33**: 491-497

McCoy AJ (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta crystallographica Section D, Biological crystallography* **63**: 32-41

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *Journal of applied crystallography* **40**: 658-674

Medalia N, Beer A, Zwickl P, Mihalache O, Beck M, Medalia O, Navon A (2009) Architecture and molecular mechanism of PAN, the archaeal proteasome regulatory ATPase. *The Journal of biological chemistry* **284**: 22952-22960

Panne D, Raleigh EA, Bickle TA (1999) The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. *Journal of molecular biology* **290**: 49-60

Pieper U, Brinkmann T, Kruger T, Noyer-Weidner M, Pingoud A (1997) Characterization of the interaction between the restriction endonuclease McrBC from *E. coli* and its cofactor GTP. *Journal of molecular biology* **272**: 190-199

Pieper U, Pingoud A (2002) A mutational analysis of the PD...D/EXK motif suggests that McrC harbors the catalytic center for DNA cleavage by the GTP-dependent restriction enzyme McrBC from *Escherichia coli*. *Biochemistry* **41**: 5236-5244

Pieper U, Schweitzer T, Groll DH, Gast FU, Pingoud A (1999) The GTP-binding domain of McrB: more than just a variation on a common theme? *Journal of molecular biology* **292**: 547-556

Rhou A, Grigorieff N (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of structural biology* **192**: 216-221

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**: 725-738

Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology* **180**: 519-530

Skubak P, Pannu NS (2013) Automatic protein structure solution from weak X-ray data. *Nature communications* **4**: 2777

Sutherland E, Coe L, Raleigh EA (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *Journal of molecular biology* **225**: 327-348

Tan YZ, Cheng A, Potter CS, Carragher B (2016) Automated data collection in single particle electron microscopy. *Microscopy* **65**: 43-56

Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ (2007) EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology* **157**: 38-46

Taylor JL, White SR, Lauring B, Kull FJ (2012) Crystal structure of the human spastin AAA domain. *Journal of structural biology* **179**: 133-137

Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, Zwart PH, Hung LW (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta crystallographica Section D, Biological crystallography* **65**: 582-601

Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. *Acta crystallographica Section D, Biological crystallography* **66**: 22-25

Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *Journal of molecular biology* **229**: 105-124

Wang J, Song JJ, Franklin MC, Kamtekar S, Im YJ, Rho SH, Seong IS, Lee CS, Chung CH, Eom SH (2001) Crystal structures of the HslIVU peptidase-ATPase complex reveal an ATP-dependent proteolysis mechanism. *Structure* **9**: 177-184

Wendler P, Shorter J, Snead D, Plisson C, Clare DK, Lindquist S, Saibil HR (2009) Motor mechanism for protein threading through Hsp104. *Molecular cell* **34**: 81-92

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta crystallographica Section D, Biological crystallography* **67**: 235-242

Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic acids research* **43**: W174-181

Yokom AL, Gates SN, Jackrel ME, Mack KL, Su M, Shorter J, Southworth DR (2016) Spiral architecture of the Hsp104 disaggregase reveals the basis for polypeptide translocation. *Nature structural & molecular biology* **23**: 830-837

Zhang K (2016) Gctf: Real-time CTF determination and correction. *Journal of structural biology* **193**: 1-12

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* **9**: 40

## Chapter 5

### Pre-steady-state kinetic studies on McrB and McrBC

#### 5.1 Introduction

As discussed earlier, the McrB GTPase motor belonging to AAA+ family assembles into higher oligomers both in presence and absence of nucleotides (Chapter 2). GTP hydrolysis has been shown to be essential for the endonucleolytic activity of McrBC (Gast et al, 1997; Panne et al, 1999; Pieper et al, 1997; Sutherland et al, 1992). In Chapter 2, I showed that GTP drives the formation of a stable hexameric assembly of McrB and tetradecameric assembly of McrBC independent of protein concentration. This was followed by observations on the structural features of McrB $\Delta$ N complex based on which I had proposed that McrBC complex might hydrolyze GTP in a sequential manner. To further understand the mechanism of the assembly of McrBC and its GTPase activity, it is important to understand the rate-limiting steps that govern these reactions. In order to understand whether the two complexes have a similar mode of assembly, I probed the oligomerization kinetics of McrB both in presence and absence of McrC. Towards this, I carried out stopped-flow studies using intrinsic tryptophan fluorescence of McrB and McrC to monitor oligomerization in conjunction with pre-steady state GTP hydrolysis by McrBC. This study was important to establish if oligomerization was a prerequisite for GTP hydrolysis. Previous studies have shown that DNA has no or little effect on steady-state GTPase activity of McrBC (Panne et al, 1999; Pieper et al, 1997). To understand whether this observation holds true even for pre-steady-state, I carried out GTPase study of McrB in presence of DNA.

As GTP is ubiquitous in the cellular milieu, it is expected that McrBC will exist as a tetradecameric complex before encountering substrate DNA, which could be linear, closed circular or super-coiled. Many of the DNA metabolizing enzymes like helicases and translocases have been shown to have a multi-subunit hexameric ring architecture. These enzymes mostly belong to either AAA+ family or RecA family of ATPase and function by translocating DNA through the central pore of the ring. While the toroidal shape of these motors provides a higher processivity it also poses an inherent challenge to load and encircle DNA. The process of DNA loading by these protein complexes requires a high regulation necessary to control the DNA replication. Unfoldases, associated with proteasome machinery also translocate peptides through



the center of the pore. In these motors, the substrate, upon recognition, is threaded from the top of the ring. However, in case of hexameric helicases and translocases like FtsK, DnaB, T7gp4, MCM, SV40 LTag, rho terminase (both AAA+ and RecA family) (O'Shea & Berger, 2014), the threading model does not seem to hold true as these proteins work on large DNA substrates which are sometimes circular or do not have free ends. Even though the shape and architecture of these helicases and translocases are known for quite some time, the knowledge of the mode of their loading onto DNA has just started to emerge. With the availability of high-resolution structures of these helicases coupled with *in vitro* studies, we now have a better mechanistic description of how a helicase interacts with the separated single strands. But such information is lacking in case of double-strand DNA translocases which are important for cellular processes like DNA recombination and repair.

There are two possible modes for assembly of these complexes- either they assemble around the substrate or the ring has to open in a clamp-like fashion to accommodate sliding of DNA into the central pore. The former model is shown for proteins which assemble into oligomeric structures in presence of substrate like-

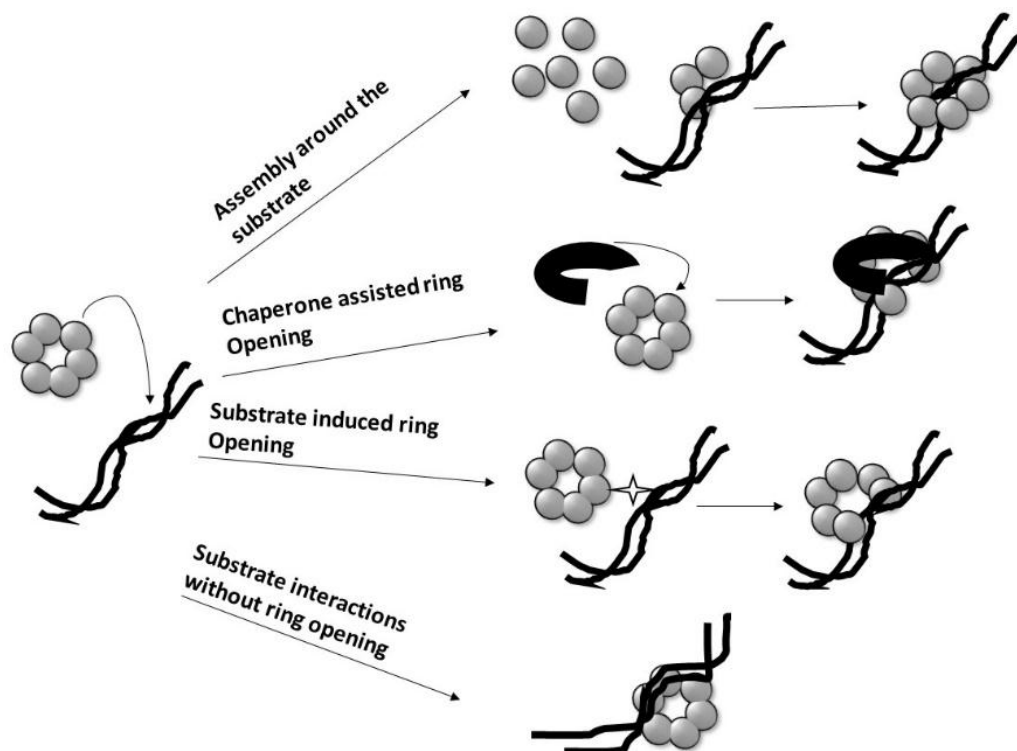


Figure 5.1. **Possible modes for a close ring molecule to load onto DNA.** Schematic diagram showing binding of a closed hexameric ring to DNA by A) disassembling upon DNA interaction and then assembling around substrate, B) opening with the help of a chaperone, C) opening without chaperone or D) instead of ring opening, wrapping DNA around the ring during translocation.

papillomavirus protein E1 (Schuck & Stenlund, 2005; Schuck & Stenlund, 2011), SV40 LTag (Chang et al, 2013; Dean et al, 1992; Gomez-Lorenzo et al, 2003; Valle et al, 2000) and FtsK (Ben-Yehuda et al, 2003; Crozat & Grainge, 2010; Massey et al, 2006). In case of helicases like T7 gp4 and MCM, it is proposed that these proteins exist as heptamers (Toth et al, 2003; Yu et al, 2002), and upon DNA interaction lose one subunit, converting into a hexamer along with creating a gap through which substrate can slip into the ring (Costa et al, 2006; Crampton et al, 2006). Some hexameric helicases like DnaB, employ a chaperone or loader like DnaC which help open the preformed ring to load onto substrate DNA (Arias-Palomo et al, 2013). Then there is an example of rho transcription terminator whose crystal structure showed an open split “Lock washer” like the shape of the ring in presence its substrate (Skordalakes & Berger, 2003). Based on these studies and knowledge about the architecture of McrBC, one can imagine it to bind to its substrates in one of the four following ways. A) Assembly around the substrate; B) use McrC as a chaperone to open the ring and load onto the substrate; C) undergo ring opening in response to the substrate; D) interact only through its N-terminal domains by DNA wrapping (Figure 5.1 ).

In this Chapter, I used the intrinsic tryptophan fluorescence of the protein to observe conformational changes under different protein DNA mixing regimes in a millisecond time resolution stopped-flow spectroscopy set up. This study was complemented with studies on DNA binding and GTPase activity of the protein under similar mixing regimes. In the light of results obtained, this chapter discusses the plausibility of different models for McrBC loading on double-stranded DNA (dsDNA).

## **5.2 Materials and Methods**

### ***5.2.1 Protein purification***

McrB and McrC were purified as described previously. Before starting the experiments, protein concentration was re-estimated from the absorbance at OD<sub>280</sub> using extinction coefficient 74720 M<sup>-1</sup> cm<sup>-1</sup> for McrB, 54780 M<sup>-1</sup> cm<sup>-1</sup> for McrC and 37360 M<sup>-1</sup> cm<sup>-1</sup> for McrBΔN. In all the experiments discussed below, McrB and McrC were added in 4:1 molar ratios. In the following sections, a mix of McrB and McrC without nucleotide will be referred to as McrB+McrC and the mix incubated with GTP will be referred as McrBC.

### 5.2.2 Oligonucleotide

The specific methylated DNA, mDNA, was designed similar to the 60 bp binding oligonucleotide used in DNA binding studies (EMSA) in Chapter 2, except that mDNA contained methylation only in top strand. For mDNA, following oligonucleotides were annealed in 10 mM Tris pH 8.

MB60MSPI-1F	GCCGGGTAACCCGGGTAAGTCCGGGTAAGA/i5HydMe- dC/CGGTAGTTCGGTATCGAGGGGTAGGCCGC
MB60SPI-1R	GCGGCCTACCCCTCGATACCGAACTACCGGTCTTACCCGGAC TTACCCGGGTTACCCGGC

The non-specific DNA substrate (sequence-wise similar to mDNA), nmDNA (non-methylated), was also generated by annealing following oligonucleotides in 10 mM Tris pH 8.

MB60SPI-1F	GCCGGGTAACCCGGGTAAGTCCGGGTAAGACCGGTAGTTCG GTATCGAGGGGTAGGCCGC
MB60SPI-1R	GCGGCCTACCCCTCGATACCGAACTACCGGTCTTACCCGGAC TTACCCGGGTTACCCGGC

Annealing was done by mixing the two oligonucleotides in equimolar amounts followed by heating at 99°C and the cooling gradually till 25°C at a rate of 1°C per second using a thermocycler.

For anisotropy experiments, 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (Hex) was added to 5'-terminus of MB60MSPI-1F or MB60SPI-1F. The labeled MB60MSPI-1F and MB60SPI-1F were annealed with unlabeled MB60SPI-1R to generate methylated and non-methylated substrate DNA respectively. The oligonucleotides were obtained from Integrated DNA Technologies, Inc. (IDT).

### 5.2.3 Steady-state tryptophan fluorescence measurements

Steady-state tryptophan fluorescence was measured at 25°C using Horiba FluoroMax 4 spectrometer with  $\lambda_{\text{ex}} = 297$  nm (slit width 5 nm). Emission spectra was collected from 307 nm to 407 nm (slit width = 5 nm). Final concentration of reactants (as mentioned in different experiments) were 500 nM McrB, 125 nM McrC, 1 mM

nucleotide (GTP, GDP, GDPNP or GTP $\gamma$ S) and 500 nM mDNA in 50 mM KCl, 10 mM Tris pH 8, 5 mM MgCl<sub>2</sub> and 1 mM DTT.

#### **Correcting inner filter effect:**

The inner filter effect of different ligands (nucleotides or DNA) was corrected by using 500 nM tryptophan solution with 1 mM nucleotide (GTP, GDP, GDPNP or GTP $\gamma$ S) or 500 nM DNA (specific mDNA1) in 50 mM KCl, 10 mM Tris pH 8, 5 mM MgCl<sub>2</sub> and 1 mM DTT.

#### **5.2.4 Stopped-flow intrinsic tryptophan fluorescence kinetic study**

Tryptophan fluorescence was measured at 25°C using the SF61-DX2 stopped-flow fluorimeter (TgK Scientific, UK) with excitation at 297 nm (4 nm bandwidth) and a 320 nm band pass filter (Schott WG320). Reactants in all experiments were mixed in 1:1 ratio from syringe C and syringe D and the reaction was monitored over a time period of 0 s-180 s in logarithmic acquisition mode. Final concentrations of reactants in all experiments (as mentioned in different experiments) were 500 nM McrB/McrB $\Delta$ N, 125 nM McrC, 1 mM GTP, 1 mM GTP $\gamma$ S, 500 nM mDNA and 500 nM nmDNA in 50 mM KCl, 10 mM Tris pH 8, 5 mM MgCl<sub>2</sub> and 1 mM DTT.

#### **5.2.5 Stopped-flow pre-steady-state GTPase rate measurements**

GTP hydrolysis by McrB and McrB+McrC were measured by using the phosphate binding protein (PBP) labeled with N-[2-(1-maleimidyl)ethyl]-7-(diethylamino)coumarin-3-carboxamide (MDCC) (Webb, 2007). The labeled PBP shows an increase in fluorescence upon binding to phosphate. A calibration curve was prepared with an inorganic phosphate standard and experimental data were analyzed within the linear range of the standard. Inorganic phosphate contamination in the nucleotide was removed by treatment with a phosphate “mop” (Brune et al, 1994) – bacterial purine nucleoside phosphorylase (PNPase) and 7-methylguanosine (7-MG) (Sigma) for 1 hour followed by filtering the mopped solution using a protein concentrator (Vivaspin20, 10 kDa Sartorius). Fluorescence intensity was measured using SF61-DX2 stopped-flow fluorimeter (TgK Scientific, UK) with excitation at 436 nm (2 nm bandwidth) and a 455 nm band pass filter (Schott GG455). All measurements were performed at 25°C. In each reaction, PBP-MDCC was added at

12  $\mu\text{M}$  concentration in syringe C with protein making the final concentration as 6  $\mu\text{M}$ . All GTPase experiments were carried out in buffer containing 50 mM KCl, 10 mM Tris pH 8, 5 mM  $\text{MgCl}_2$  and 1 mM DTT. The concentration of nucleotide and protein will be mentioned with each data in the text.

### Data Fitting:

The GTPase rates were derived by fitting linear regression line to the steady-state. The data was then fit to following models as mentioned in main text or figure legends:

- 1) Michaelis-Menten model for enzyme kinetics

$$Y = V_{max} * \frac{X}{K_m + X}$$

Where, Y = Steady-state GTPase rate at a given nucleotide concentration, X = concentration of nucleotide,  $V_{max}$  = maximum steady-state GTPase rate,  $K_m$  = concentration of GTP at which half maximal GTPase rate is attained.

- 2) Allosteric Sigmoidal model for enzyme kinetics

$$Y = V_{max} * X^h / (K_{prime} + X^h)$$

Where, Y = Steady-state GTPase rate at a given nucleotide concentration, X = concentration of nucleotide,  $V_{max}$  = maximum steady-state GTPase rate,  $K_{prime} = K_m^h$  and h = hill slope.

- 3) Specific Binding Model with Hill Slope for ligand binding

$$Y = B_{max} * \frac{X^h}{K_d^h + X^h}$$

Where, Y = Initiation rate constant at a given nucleotide concentration, X = concentration of McrB,  $B_{max}$  = maximum initiation rate constant,  $K_d$  = dissociation rate constant and h = hill slope.

- 4) Substrate Inhibition Model for enzyme kinetics

$$Y = Y_{max} * X / (K_d + X * (1 + X/K_i))$$

Where, Y = Initiation rate constant at a given nucleotide concentration, X = concentration of nucleotide,  $Y_{max}$  = maximum initiation rate constant,  $K_d$  = concentration of GTP at which half maximal initiation is attained and  $K_i$  = dissociation constant for substrate binding such that two substrates can bind to an enzyme.

### **5.2.6 Stopped-flow DNA anisotropy measurements**

DNA anisotropy was measured at 25°C using the SF61-DX2 stopped-flow fluorimeter (TgK Scientific, UK) with excitation wavelength set at 534 nm (6 nm bandwidth) and a 550 nm long pass filter (Schott OG550). G-factor was set to unity as described earlier (Schwarz et al, 2013). Reactants from syringe C and D were mixed in a 1:1 ratio. Data were acquired in linear mode for 20 s. About 8 traces were averaged for each different experiment. The final reaction contained 10 nM labeled DNA, 500 nM McrB, 1250 nM McrC and 1 mM GTP in 50 mM KCl, 10 mM Tris pH 8, 5 mM MgCl<sub>2</sub> and 1 mM DTT.

### **5.2.7 Steady-state DNA anisotropy measurements**

To measure DNA binding by McrB and McrB+McrC, steady-state anisotropy studies were carried out using Fluorolog-3 spectrofluorometer. Measurements were taken at 25°C with  $\lambda_{\text{excitation}} = 534$  nm;  $\lambda_{\text{emission}} = 551$  nm, and slits, 5/5 nm. Protein (McrB, McrC or a mix of McrB and McrC in a McrB:McrC::4:1 ratio) was added sequentially to 5.2 nM DNA (methylated or non-methylated) in 50 mM KCl, 10 mM Tris pH 8, 5 mM MgCl<sub>2</sub> and 1 mM DTT. The concentrations were re-calculated at every step to take care of dilution effect. Anisotropy measurements were made after 5 min incubation at each concentration. Data were plotted in GraphPad Prism 5 and fit by nonlinear regression to a single site binding equation.

## **5.3 Results**

### **5.3.1 McrB shows change in Trp fluorescence in presence of nucleotide**

Size exclusion studies showed that both McrB and McrBC can oligomerize into higher oligomers in presence of either GTP or GDP. In my structural studies I observed that while McrB $\Delta$ N formed hexameric rings with 6-fold symmetry (observed in 2-D class averages; Chapter 2 and self-rotation function; Chapter 4), the McrB $\Delta$ NC complex

lacked a 6-fold symmetry despite having two hexamers (Chapter 4; self-rotation function for McrB $\Delta$ NC diffraction data).

The tryptophan (Trp) fluorescence has been used as a probe in many studies to indicate conformational changes in a protein (Lakowicz, 2006; Lakowicz & Gryczynski, 1992; Yengo et al, 1999). Changes in Trp emission spectra can be due to ligand binding, conformational changes, protein oligomerization or protein degradation. To further understand the assembly of McrB hexamers, I investigated whether Trp fluorescence of the protein can be used to monitor conformational changes associated with the assembly pathway. Similar experiments were carried out to study the pathway of oligomeric assembly by the AAA+ protein ClpA (Kress et al, 2007). McrB has six tryptophans; three in the N-terminal DNA binding domain and three in the C-terminal AAA+ domain. McrC has four tryptophans. I first used steady-state fluorescence to measure the changes in the emission profiles upon addition of different nucleotides. To see whether the two states have similar conformations, I observed steady-state

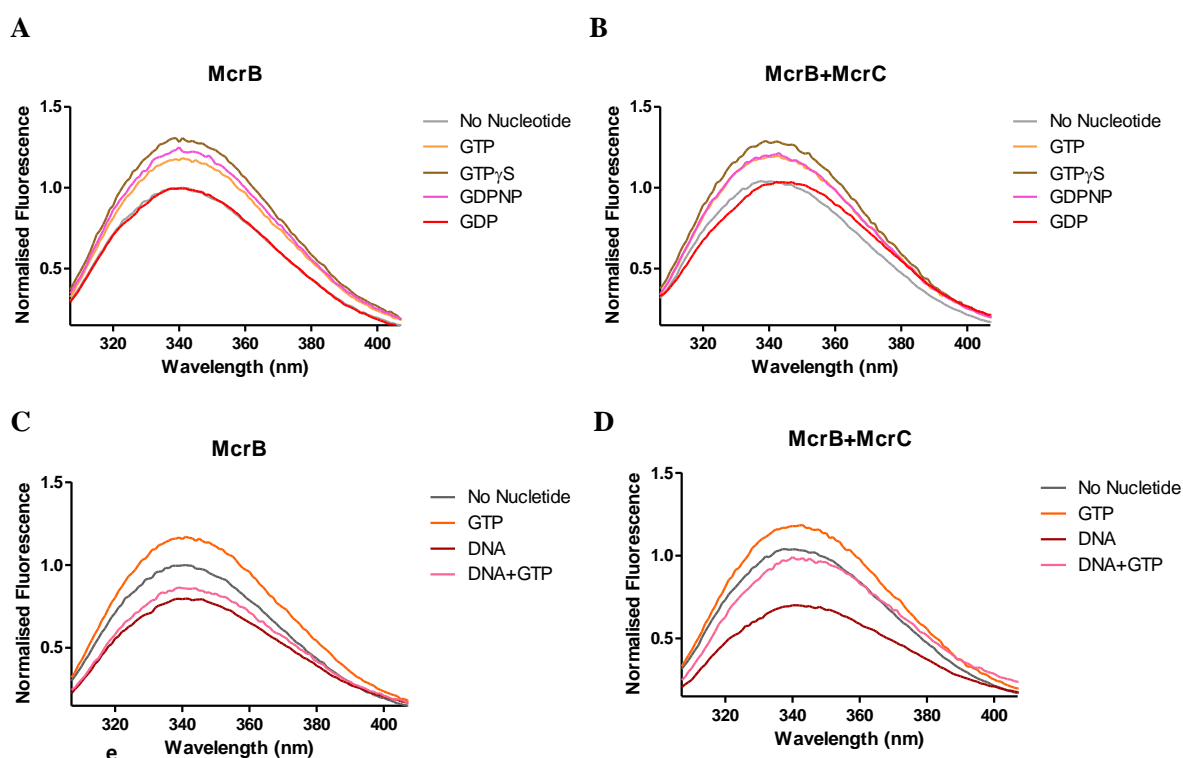


Figure 5.2. **Steady-state tryptophan fluorescence emission scans** of A) 500 nM McrB in presence of 1 mM of different nucleotides, B) 500 nM McrB and 125 nM McrC in presence of 1 mM of different nucleotides, C) 500 nM McrB in presence of 500 nM mDNA and 1 mM GTP, D) 500 nM McrB and 125 nM McrC in presence of 500 nM mDNA and 1 mM GTP. The fluorescence intensity in each experiment was normalized with respect to the fluorescence intensity of protein (McrB or McrB+McrC) without nucleotide or DNA at 350 nm. The data were plotted in GraphPad Prism 5.

tryptophan fluorescence of McrB and McrBC in presence of different nucleotides- GTP, GDP, GDPNP and GTP $\gamma$ S (Figure 4.2 A, B). McrB showed about 20% increased tryptophan fluorescence in presence of GTP and GDPNP and about 30% increase in case of GTP $\gamma$ S without any spectral shift (Figure 4.2 A). In contrast, the presence of GDP did not increase tryptophan fluorescence. McrB+McrC also showed enhanced, but lower than that of McrB, tryptophan fluorescence - 15% in case of GTP and GDPNP, 23% in case of GTP $\gamma$ S, but no fluorescence change was observed in presence of GDP (Figure 4.2 B). Identical experiments carried out at different times had a wavelength of emission maxima varying from 337 to 341 nm in case of both McrB and McrB+McrC (data not shown). However, in presence of nucleotide, the wavelength of emission maxima was always 341 or 342 nm. The fluorescence emission increased upon binding of a nucleotide having a  $\gamma$  phosphate. This suggested that there is a difference in the structural states of the oligomers formed in complex with GTP and GDP and/or that the  $\gamma$  phosphate modulated the environment of one or more tryptophans.

Steady-state emission spectra of tryptophan fluorescence of McrB showed a reduction in fluorescence intensity by 20% when DNA was added (Figure 5.2 C). A similar experiment carried out in presence of both McrB+McrC, showed a higher quenching (33%) than McrB (Figure 5.2 D). In presence of GTP, the fluorescence quenching by DNA for both McrB and McrB+McrC was lowered to 15% and 5% respectively. Thus, while GTP affected the amount of fluorescence quenching by DNA for McrB+McrC, there was little effect in case of McrB alone. Even though a decrease in intensity upon DNA addition was similar in repeat experiments, the emission maxima were variable for both McrB and McrB+McrC. The observation that McrB+McrC, when mixed with DNA+GTP, showed a much larger difference in fluorescence intensity change than when mixed only with DNA suggested a different mode of binding of the complex rather than the monomeric proteins. To observe these fluorescence changes in real time and understand the process better, I carried out stopped-flow studies with both McrB and McrB+McrC.



### 5.3.2 Complexation of McrB with McrC does not affect Trp fluorescence

Trp fluorescence intensity changes in McrB or a mix of McrB and McrC (McrB+McrC) were monitored over time after mixing with nucleotide to investigate the assembly process (Figure 5.3). Protein and GTP were rapidly mixed from separate syringes (Figure 5.3 A) and change in Trp fluorescence was followed in real time using a millisecond time resolution stopped-flow fluorimeter. Figure 5.3 shows the uncorrected fluorescence intensity as a function of time for different combinations of protein and

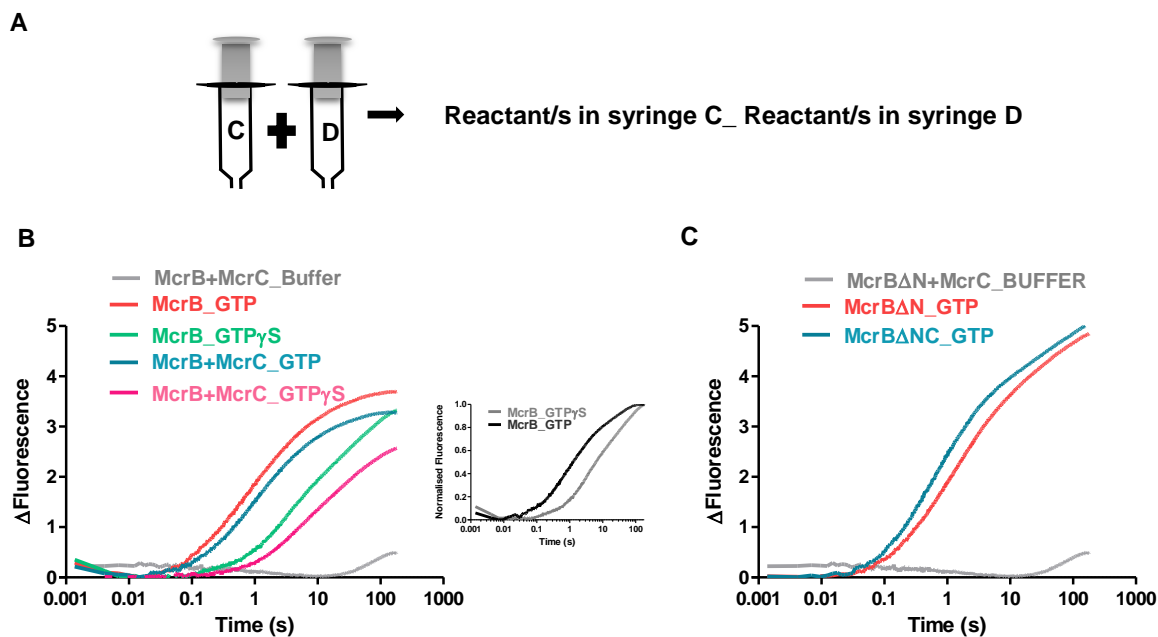


Figure 5.3. **Real time observation of McrB oligomerization.** A) Schematic to depict reaction mixing regime and labeling of different experiment. B) Comparison of stopped-flow tryptophan fluorescence change of 0.5  $\mu\text{M}$  McrB or a mix of 0.5  $\mu\text{M}$  McrB and 0.125  $\mu\text{M}$  McrC (McrB+McrC) upon mixing with 1 mM GTP or 1 mM GTPyS. Inset shows the comparison of normalised fluorescence traces of McrB mixed with GTP and GTPyS. C) Comparison of stopped-flow tryptophan fluorescence change of 0.5  $\mu\text{M}$  McrB $\Delta$ N or a mix of 0.5  $\mu\text{M}$  McrB and 0.125  $\mu\text{M}$  McrC (McrB $\Delta$ N +McrC) upon mixing with 1 mM GTP. The traces are shown in logarithmic time scale. The data were plotted and normalized in GraphPad Prism 5.

nucleotide. Change in the fluorescence intensity signal was observed when nucleotide was added to the protein indicating nucleotide-dependent conformations attained by protein. Since McrB undergoes oligomerization upon binding to the nucleotide, the signal could be because of binding of nucleotide or oligomerization of McrB or both.

Fluorescence intensity change was similar both in presence or absence of McrC suggesting that McrC does not affect the reaction leading to the change in fluorescence intensity (Figure 5.3 B). Similar observations were made with the N-

terminal deletion mutant McrB $\Delta$ N, where the presence of McrC did not lead to a dramatic change in the rate of intensity change (Figure 5.3 C). The rate of intensity change for McrB and McrB $\Delta$ N also appeared similar, suggesting that the signal is mostly contributed by the C-terminal GTPase domain. If the Trp fluorescence signal was due to the formation of hexamers/tetradecamers, then one would expect to observe a lag in the signal due to the multiple steps needed to form the oligomer. However, if the first step in oligomerization or the association of the nucleotide were to trigger the Trp signal, one would expect an exponential dependence without a lag.

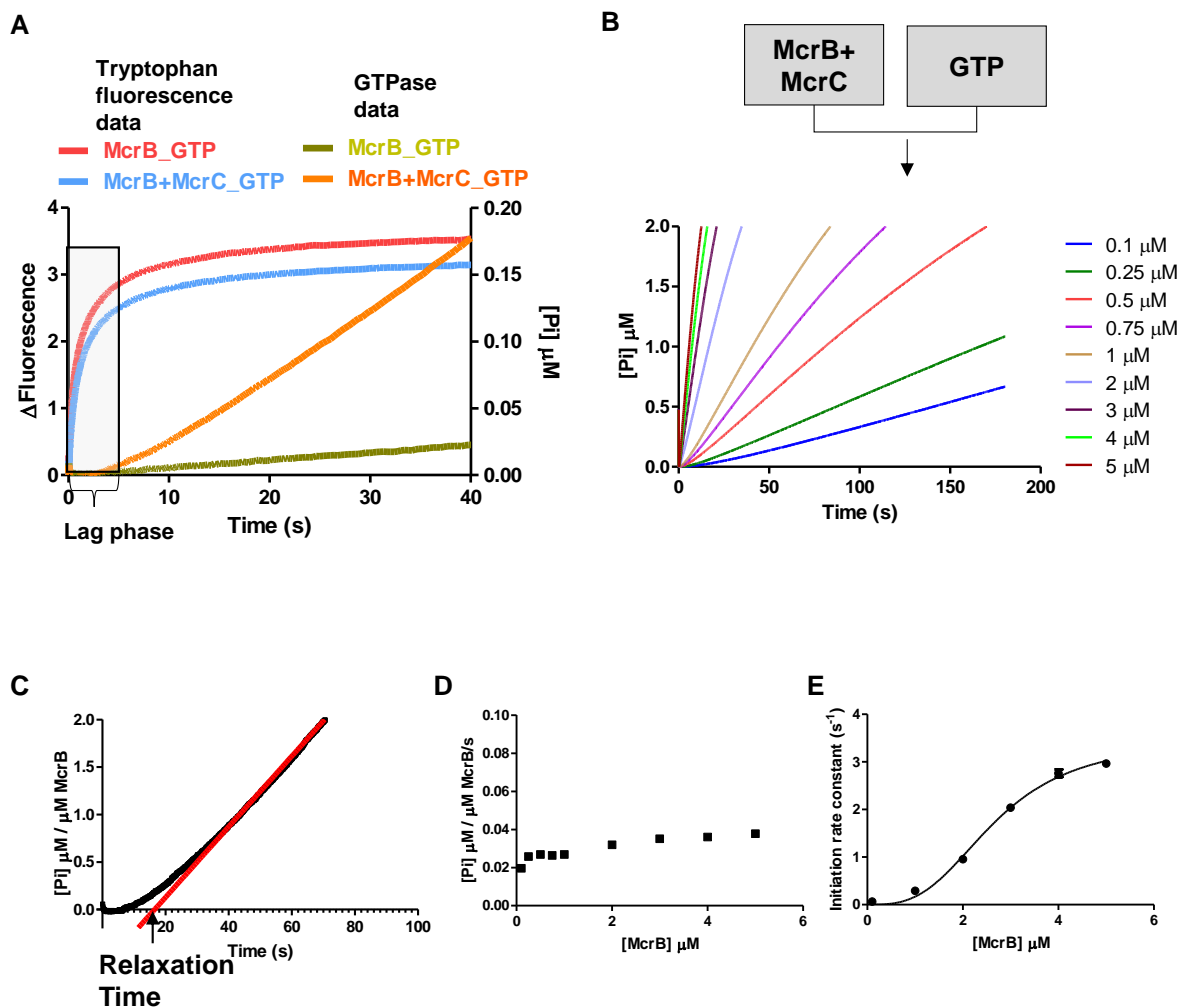
The data for McrB or McrB+McrC, when mixed with GTP, were fit to both in-built “exponential” and a “lag preceding exponential” models in GraphPad Prism 5. The data did not show a lag and most closely fit to the sum of multiple exponential processes. Therefore the Trp fluorescence change that was observed was a signal on the pathway to oligomer formation, but it could not be necessarily determined if it was the end point of the reaction. Because of the complexity of the signals, I did not further investigate its dependence on the concentration of nucleotides.

Slower reactions were observed when GTP $\gamma$ S was mixed with McrB or McrB+McrC (Figure 5.3 A; inset), where the reaction profile became spread over more time decades. This suggested the possibility that the assembly rates were slower in presence of this analog. Alternatively, one can speculate that the profile in presence of the non-hydrolysable GTP $\gamma$ S is different from that of GTP because in case of the latter oligomerization is accompanied by hydrolysis simultaneously.

### **5.3.3 GTP hydrolysis follows oligomerization of McrB**

To examine if oligomerization and hydrolysis are simultaneous events, I sought to compare the rate of GTP hydrolysis using the same stopped-flow set-up and mixing regime. For this, the phosphate binding protein (PBP) was used, which produces an increase in fluorescence upon rapid and tight binding to free phosphate (Brune et al, 1994; Webb, 2007). The use of rapid mixing to measure the GTPase activity has a distinct advantage over previous steady-state studies which used techniques that required long measurement times (in order to stabilize the signal).

I first compared the rate of phosphate release of McrB and McrB+McrC at saturating GTP concentration (1 mM). The profile showed two clear features (Figure 5.4 A). Firstly, as reported previously (Panne et al, 1999; Pieper et al, 1997), the presence of McrC accelerated the steady-state GTPase rate and secondly, both



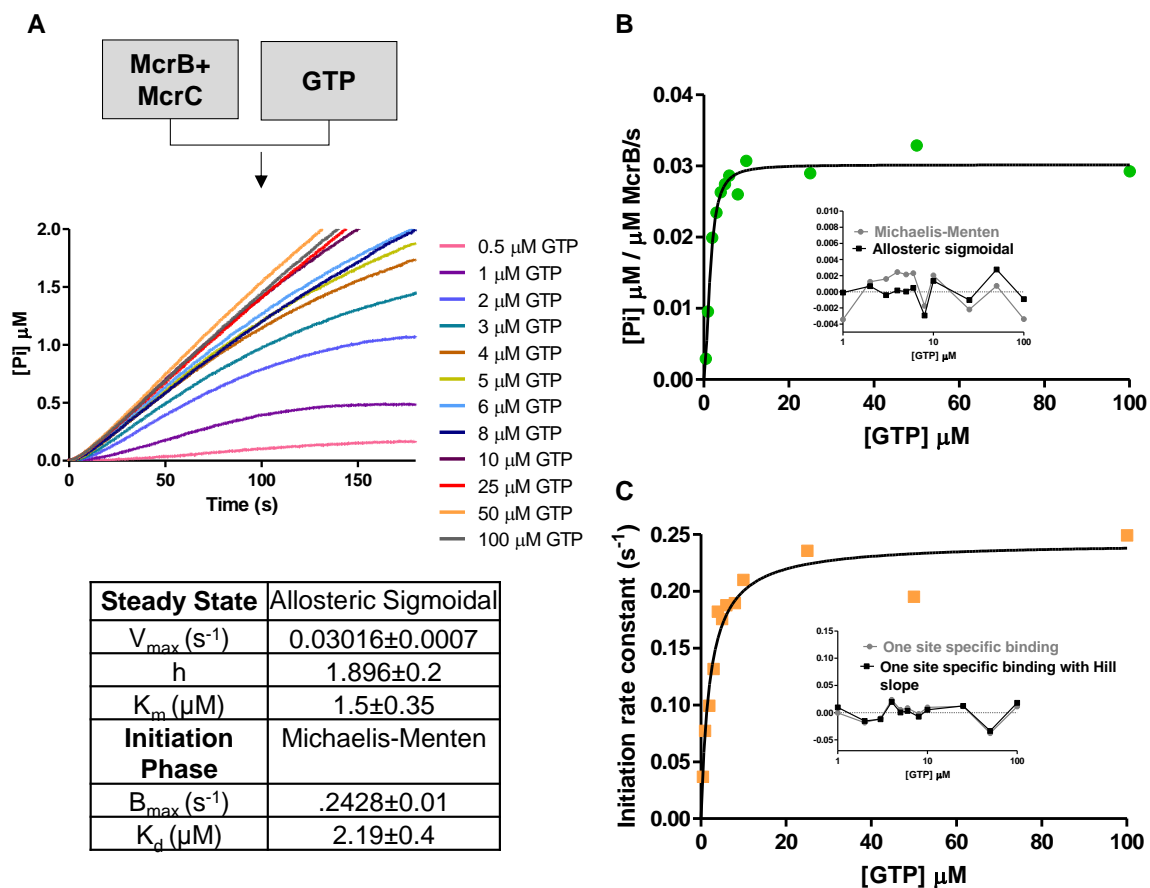
**Figure 5.4. Relation between GTP hydrolysis and oligomerization.** A) Time trace of GTP hydrolysis with 0.5 μM McrB and 0.5 μM McrB+0.125 μM McrC in comparison with corresponding time trace of tryptophan fluorescence. GTP concentration in both types of experiment was 1 mM. B) GTPase data for McrB+McrC collected at different concentrations of protein. McrB concentration is mentioned in the figure while corresponding McrC concentration was 1/4<sup>th</sup> of McrB concentration. Measurements were carried out with 1 mM GTP. C) An example of data analysis showing the linear straight line fit for steady-state rate and intercept at X-axis for initiation time constant (reciprocal of relaxation time). D) Graph showing effect of protein concentration on steady-state GTP hydrolysis rates and E) graph showing effect of protein concentration on rate at which steady-state is attained (Initiation) fit to a specific binding model with Hill slope (Hill coefficient = 3) The data were plotted and fit in GraphPad Prism 5. All the experiments are average of three traces.

profiles showed a clear lag phase. By plotting the GTPase profile with the tryptophan fluorescence data collected with similar mixing regimes, it was observed that the steady-state hydrolysis was reached only when maximum Trp fluorescence intensity changes were attained. This indicated that the lag observed in GTPase data represented the assembly steps in the formation of higher order species and that the lower order species might have had no or very low GTPase activity.

To further investigate the assembly process, I measured the GTPase rate upon mixing different concentrations of McrB+McrC with 1 mM GTP (Figure 5.4 B). In each case the McrB:McrC ratio was kept as 4:1. One would expect that for each concentration of McrBC, the steady-state phase would return the same microscopic rate (i.e. phosphates released per McrB) while the lag phase, from a simple consideration of the law of mass action, would decrease with an increase in concentration. The steady-state rate and an apparent lag time were estimated by fitting the steady-state phases by straight line using non-linear-regression, where the gradient is the steady-state rate and the X-axis intercept is the lag time (relaxation time). The reciprocal of relaxation time gave an apparent initiation rate constant. This rate constant represents one or sum of two or more steps involved in oligomerization (Figure 5.4 C). The relationship between steady-state rates ( $\mu\text{M Pi}/\mu\text{M McrB/s}$ ) and concentration of McrB was approximately constant (Figure 5.4 D). The apparent initiation rate showed a sigmoidal relationship with protein concentration indicating the effect of concentration on the rate of oligomerization (Figure 5.4 E). This observation agrees with the previous study discussed in Chapter 2, where I showed that increase in the concentration of protein shifts the equilibrium towards higher oligomers.

#### **5.3.4 GTP hydrolysis by McrBC is a cooperative process**

Another feature of a hexameric ring-shaped GTPase is that the GTPase kinetics may have a cooperative dependence on the GTP concentration if more than one site in the ring is occupied during hydrolysis. Such cooperativity has been observed in other AAA+ motors but no such observation has been reported for McrB or McrBC. As was proposed in Chapter 4, McrBC might hydrolyze GTP in a sequential manner. To probe



**Figure 5.5. GTP hydrolysis in McrBC show cooperativity.** A) Stopped-flow GTP hydrolysis traces (average of three traces) of 500 nM McrB in presence of 125 nM McrC at varying GTP concentrations. B) Graph showing effect of GTP on steady-state GTPase reaction. Data was fit to an allosteric sigmoidal model. Inset shows the comparison of residuals for a fit obtained with Michaelis-Menten model and Allosteric sigmoidal model. C) Graph showing initiation phase of GTPase reaction fit to one site specific binding model. Inset shows the comparison of residuals for a fit obtained with one site specific binding with and without Hill slope. The table shows the kinetic parameters obtained from data fitting. The steady state rates and initiation rate constants were derived from average of two separate experiments (three GTPase traces were averaged in a single experiment). The data were plotted and fit in GraphPad Prism 5

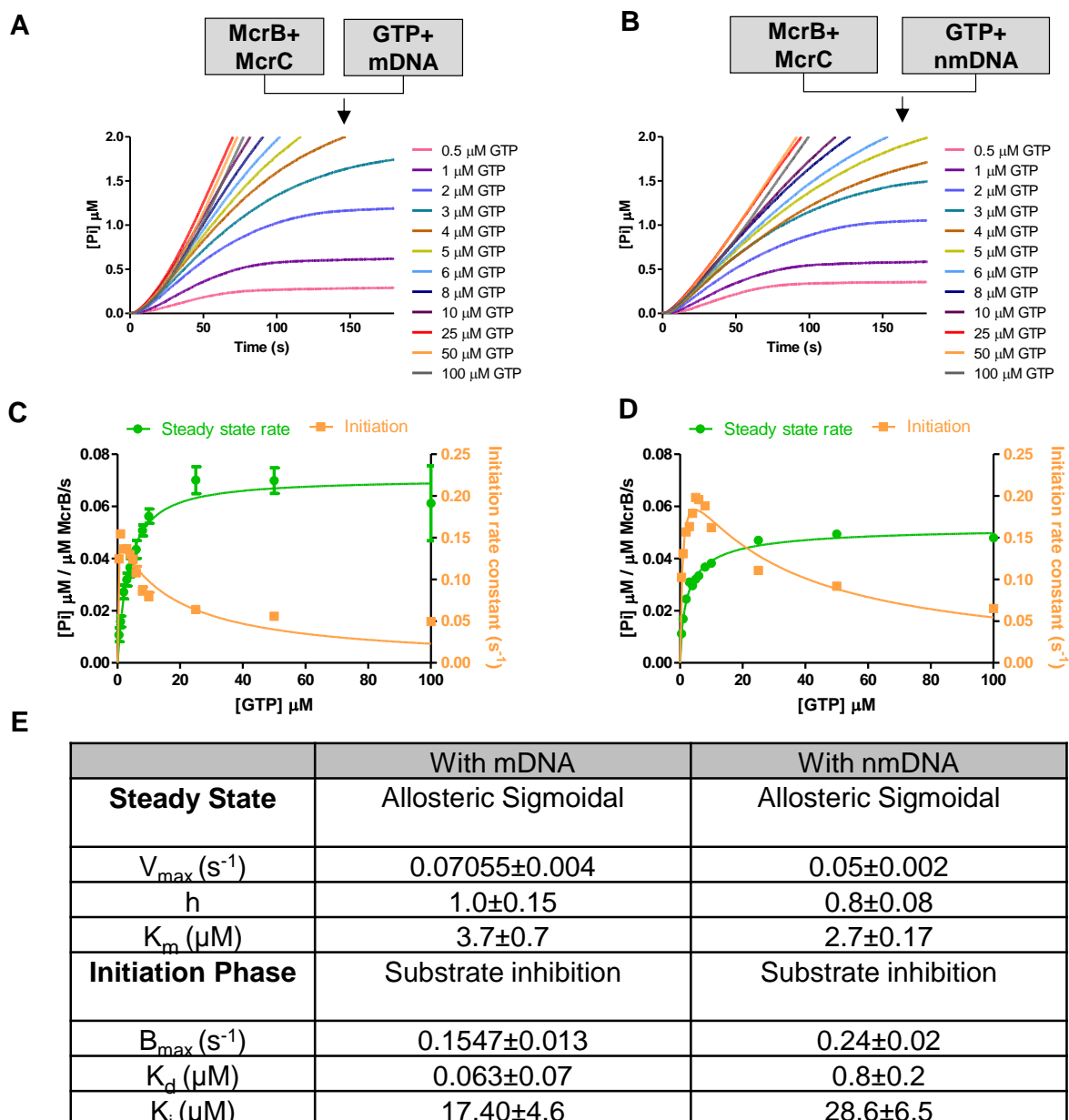
the sequential mode of GTP hydrolysis, GTPase assay of McrBC was carried out with increasing GTP concentration. Ideally, this reaction should be done by mixing preformed rings with GTP in the stopped-flow. However, this could not be done in this case as GTP is necessary for McrBC oligomerization. Therefore I measured the GTPase rate by rapidly mixing McrB+McrC (4:1 molar ratio) with different concentrations of GTP (Figure 5.5 A). By fitting the steady-state phase of the reaction by a straight line, I determined the dependence of both the steady-state rate and the initiation rate constant, which was a complex mix of oligomeric assembly stages and

any initial hydrolysis rates. The data (Figure 5.5 B and C) showed that the steady-state rates had a sigmoidal dependence on the GTP concentration, with Hill coefficient of 1.9 while the apparent initiation rate showed a Michaelis-Menten dependence on GTP. In agreement with GTP binding and hydrolysis studies carried out on McrBC previously (Pieper et al, 1997), both steady-state rate and initiation rate constant gave similar  $K_m$  and  $K_d$  values of 1.5  $\mu\text{M}$  and 2  $\mu\text{M}$ , respectively. The observed cooperativity in GTPase data suggested a requirement for more than one GTP during hydrolysis.

### **5.3.5 DNA binding affects the kinetics of GTPase activity of McrBC**

Based on the suggestion that the GTPase activity of McrBC is required for translocation along the 1D contour of the DNA (Panne et al, 1999), one might expect that the binding of DNA will activate the GTPase rate as the motor becomes coupled to the track. To investigate this, I carried out GTPase assays as a function of GTP concentration in the presence of a saturating concentration of methylated or non-methylated oligonucleotide. The data obtained from measuring GTPase activity with increasing GTP concentration were fit by a straight line to extract the steady-state rate and the apparent initiation rate. The rates thus obtained were fit as a function of GTP concentration to a Hill equation and substrate inhibition respectively (Figure 5.6). The methylated DNA (mDNA) showed a very moderate effect on the  $K_m$  of the steady-state rate but produced a more noticeable 2.5-fold increase in the  $V_{\text{max}}$  (Figure 5.6 A). The fitted Hill coefficient decreased and was close to unity. An interpretation of this is that binding of the mDNA by the McrB N-terminal DNA binding domain allowed loading of DNA to the AAA+ motor which then hydrolyzed GTP at just one site. This increase in the GTPase activity was relatively modest, and far below the activation of many DNA helicases by binding to DNA. For e.g. SV40 shows 7-fold (Giacherio & Hager, 1979), mtMCM shows 13-fold (Kelman et al, 1999), and SNF2/SWI2 shows 5-fold stimulation (Laurent et al, 1993). The DNA-activated GTPase rate is still rather modest for a translocating motor. Interestingly, the non-methylated DNA (nmDNA) also reduced the Hill coefficient, suggesting a change in coupling, although the  $K_m$  increased slightly and the steady-state rate elevated by  $\sim 1.5$ -fold (Figure 5.6 E).

I also observed that data was that both types of DNA (mDNA and nmDNA) had effects on the apparent initiation rate as a function of GTP concentration. Both the DNAs reduced the  $K_d$  and  $V_{\text{max}}$ . One explanation for these observations can be that at



**Figure 5.6. DNA affects the kinetics of GTP hydrolysis of McrBC.** A) Stopped-flow GTP hydrolysis traces (average of three traces) of 500 nM McrB in presence of 125 nM McrC and 500 nM mDNA at varying GTP concentrations along with graph showing effect of GTP on steady-state and initiation phase of GTPase reaction in presence of mDNA B) Stopped-flow GTP hydrolysis traces of 500 nM McrB in presence of 125 nM McrC and 500 nM nmDNA at varying GTP concentrations along with graph showing effect of GTP on steady-state and initiation phase of GTPase reaction in presence of nmDNA. C) Graph showing kinetics of GTPase in presence of mDNA. The steady state rates were fit to allosteric sigmoidal model and initiation phase fit to one site specific binding model. The steady state rates and initiation rate constants were derived from average of two separate experiments (three GTPase traces were averaged in a single experiment). D) Graph showing kinetics of GTPase in presence of nmDNA. The steady state rates were fit to allosteric sigmoidal model and initiation phase fit to one site specific binding model. E) The table shows the kinetic parameters obtained from data fitting. The data were plotted and fit in GraphPad Prism 5

lower GTP concentration, the protein might bind to DNA weakly. As GTP increases the affinity of McrB/McrBC for DNA (observed from EMSA studies in Chapter 2 and previous reports (Kruger et al, 1995; Panne et al, 1999)), the inhibitory effect of DNA on McrBC oligomerization might get pronounced thus resulting in a longer lag at higher GTP concentrations.

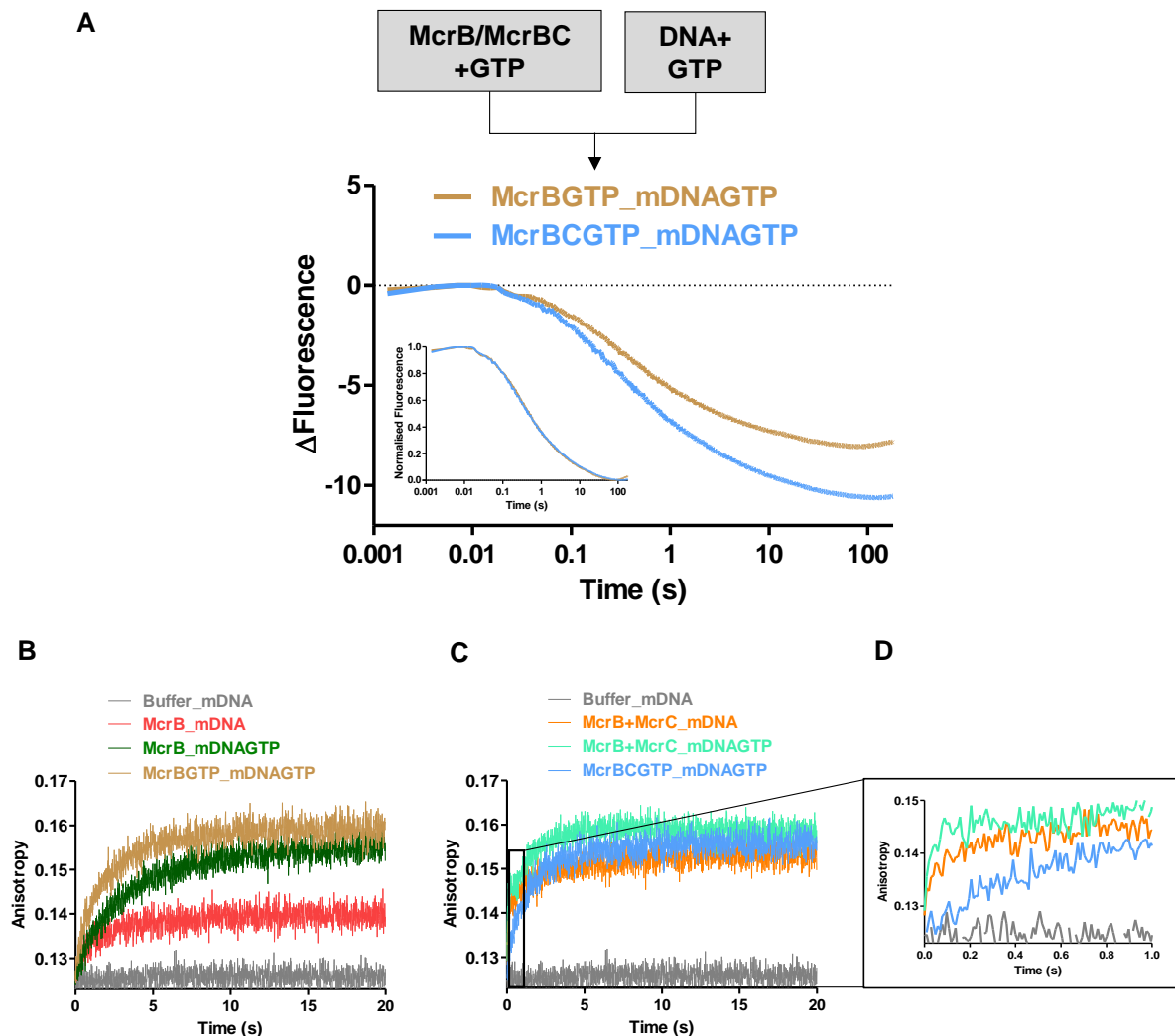
### **5.3.6 The effect of DNA binding on the McrB oligomers**

The GTPase data above suggested that a combination of GTP and DNA binding by lower order oligomeric species inhibits the formation of the hexamer that is necessary for full GTPase activity. I showed in section 5.3.1 that DNA affected the Trp fluorescence of McrB and McrC. Thus to investigate the effect of DNA on protein, I used the Trp fluorescence measurements again to follow loading of McrB complex on DNA. To avoid the signal associated with oligomer assembly, I measured the Trp intensity by mixing proteins pre-mixed with GTP (thus formed into oligomers), with DNA pre-mixed with GTP. For both McrBC and McrB, I observed a decrease in Trp intensity. The decrease in Trp fluorescence intensity (Figure 5.7 A) was greater for McrBC than McrB but the kinetics of change in fluorescence of both McrB and McrBC, observed by comparing their respective normalized traces, overlaid perfectly (Figure 5.7 A; inset). The reaction appeared complex indicating multiple transient states spread over several time decades. A simple explanation of this signal is that DNA binding causes quenching of tryptophan residues, possibly one or more of those in the McrB N-terminal DNA binding domain. The increased fluorescence intensity could be due to additional McrC tryptophans.

Additionally, I measured DNA anisotropy changes upon mixing McrB under different mixing regimes to confirm that the signal in Trp fluorescence study is due to DNA association (Figure 5.7 B). The experiment was carried out both in presence and absence of McrC (Figure 5.7 B, C). It was observed that the apparent kinetics of the DNA association profiles measured using anisotropy were faster for preformed oligomers of both McrB and McrBC (Figure 5.7 B, C). This suggested that the oligomers bind to DNA more efficiently than the lower order species. However, I also observed that whenever McrC was mixed with DNA as McrB+McrC mix, there was a sudden jump in the anisotropy signal (Figure 5.7 C, D). When McrC was present as a complex with McrB (McrBCGTP), this jump was not observed. A control experiment



with McrC showed that this was due to the non-specific interaction of McrC with labeled nucleotide (data not shown). I think this is a non-specific interaction because, neither I nor previous studies (Gast et al, 1997; Kruger et al, 1995) observed DNA binding by McrC.



**Figure 5.7. Presence of McrC does not affect DNA induced Trp signal.** A) Stopped-flow traces of preformed oligomers of McrB and McrBC (McrB+McrC pre-mixed with GTP) mixed with GTP and mDNA. Inset shows the normalized Trp signal change for both traces highlighting the similarity of kinetics between the two. The traces are shown in logarithmic time scale. B) DNA anisotropy measurement as a function of time under different mixing regimes with McrB and C) in presence of McrC. In all experiments, concentration of McrB was 500 nM, McrC 125 nM, mDNA in GTPase study was 500 nM, mDNA in anisotropy study was 10 nM and GTP was 1 mM. D) A) Time traces of DNA anisotropy change upon McrB binding in presence of McrC showing the sudden jump in anisotropy when McrB+McrC is not a preformed complex (zoomed from Figure 5.7 C). The data were plotted and normalized in GraphPad Prism 5.

### **5.3.7 DNA binding exerts an inhibitory effect on the assembly of McrB oligomers**

To further explore whether DNA binding inhibits the assembly of McrBC oligomers as suggested by the GTPase data (Figure 5.6), I wanted to repeat the tryptophan fluorescence stopped-flow assay by simultaneously mixing McrB+McrC monomers with GTP and DNA. The idea was that the signal associated with the oligomerization would be delayed. I wasn't able to continue these experiments in the presence of McrC as I observed that when McrC bound to DNA before oligomerization there were additional confounding effects. Therefore these assays were pursued with McrB alone.

I compared 3 McrB mixing regimes with GTP $\gamma$ S (Figure 5.8 A): i) McrB mixed with DNA (McrB\_mDNA), to follow the DNA binding signal without oligomerization; ii) McrB mixed with GTP $\gamma$ S (McrB\_GTP $\gamma$ S), to follow the oligomerization signal; and, iii) McrB simultaneously mixed with GTP $\gamma$ S and DNA (McrB\_mDNAGTP $\gamma$ S). It was observed that the signal from mixing McrB with GTP $\gamma$ S and DNA at the same time (McrB\_mDNAGTP $\gamma$ S) closely matched the combined signal from the addition of the oligomerization (McrB\_GTP $\gamma$ S) and DNA-binding (McrB\_mDNA) profiles (Figure 5.8 A). In other words, the oligomerization and DNA-binding signals in the case of GTP $\gamma$ S were completely independent. However, when I repeated the same experiment with GTP (Figure 5.8 B), I found that the signal from mixing McrB with GTP and DNA simultaneously (McrB\_mDNAGTP) did not overlay with the combined signal (McrB\_GTP+ McrB\_mDNA). Instead, the Trp fluorescence signal reduced faster in the former profile. This can be because of DNA binding before oligomerization and causing off-pathway states that only re-enter the assembly process slowly. This effect was not observed with GTP $\gamma$ S as the assembly pathway was already slow (observed by comparison of normalized McrB\_GTP $\gamma$ S and McrB\_GTP traces (Figure 5.3 A)) and was not rate limited by the equilibrium binding to specific DNA.

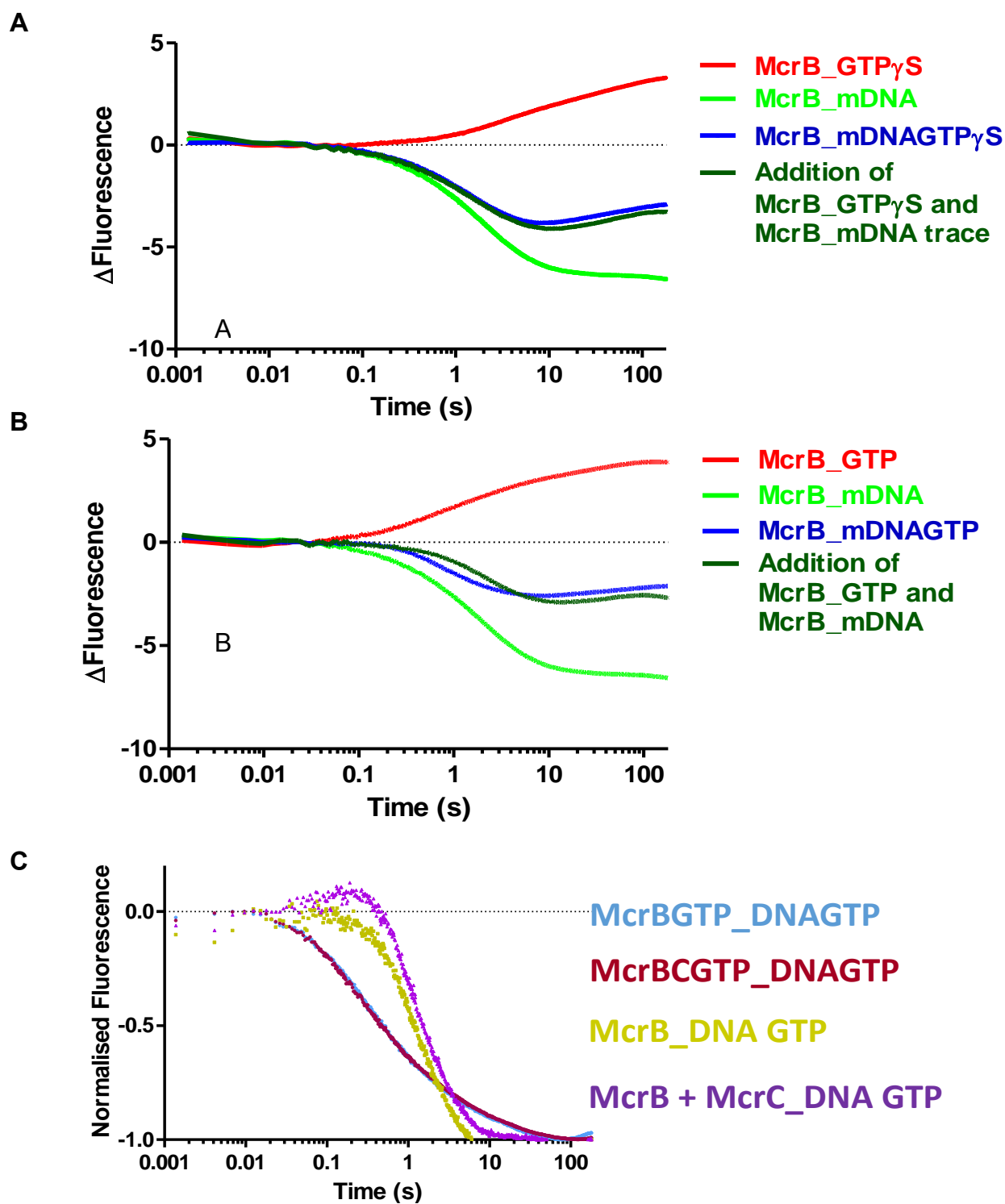


Figure 5.8. **Effect of DNA on oligomerization of McrB.** A) Comparison of Trp fluorescence change ( $\Delta$ Fluorescence) of McrB\_GTP $\gamma$ S, McrB\_DNA, McrB\_DNAGTP $\gamma$ S B) Comparison of Trp fluorescence change of McrB\_GTP, McrB\_DNA, McrB\_DNAGTP. C) Comparison of normalized Trp fluorescence changes of McrB and McrBC when mixed with DNA and GTP as preformed complex (McrBGTP\_DNAGTP, MCrBCGTP\_DNA GTP) or as monomers (McrB\_DNAGTP, MCrB+McrC\_DNA GTP). The lag observed in simultaneous mixing of monomeric protein to DNA and GTP is not observed in the preformed

complex mixed with DNA and GTP. All traces are shown in logarithmic time scale. The data were plotted and fit in GraphPad Prism 5. Data normalization was done in GraphPad Prism 5.

Next, I compared the Trp fluorescence change of preformed complex upon DNA binding and assembly in presence of DNA (Figure 5.8 C). Data in Figure 5.7 A, McrBGTP\_mDNAGTP and McrBCGTP\_mDNAGTP, mimic *in vivo* conditions while data in Figure 5.8 B, McrB\_DNA+GTP and McrB+McrC\_DNA GTP, mimic assembly of an already disassembled complex in presence of DNA. Comparison of normalized traces of these two experiments showed that the kinetics of Trp fluorescence change with preformed complex was very different from the simultaneous mixing of DNA and GTP (Figure 5.8 C). In the simultaneous mixing experiment, a lag was observed which is a result of DNA-induced delay in oligomerization (Section 5.3.6; Figure 5.8 B). If a preformed oligomer was to disassemble and then assemble, a lag in the profile should have been observed. No lag was observed in the DNA anisotropy study with preformed oligomers either (Figure 5.7 B, C). In fact, anisotropy changes measured with preformed complex showed faster kinetics of binding. Based on these observations a disassembly followed by an assembly model for McrBC can be refuted.

### **5.3.8 McrC affects DNA binding by McrB**

As discussed in the previous section, I observed that when McrC is not a part of McrBC complex i.e in experiments with McrB+McrC mix, pre-steady-state kinetics of DNA association showed complex behavior. The effect was observed both in Trp fluorescence studies and DNA anisotropy studies (Figure 5.7). To see whether the same effect is manifested at steady-state, I carried out steady-state DNA anisotropy studies of McrB both in presence and absence of McrC (Figure 5.9). As observed in EMSA studies, DNA anisotropy studies also indicated that presence of GTP increased the affinity of the protein for DNA (data not shown). In presence of nucleotide, McrBC ( $K_d = 0.65 \pm 0.2 \mu\text{M}$ ) also showed higher affinity for DNA, than McrB, ( $K_d = 1.4 \pm 0.4 \mu\text{M}$ ). This data suggested that McrC might be able to modulate DNA binding of McrB along with stimulating its GTP hydrolysis. Whether this effect is a manifestation of GTPase

activation of McrB due to complexation with McrC or it is due to another physical modulation caused by McrC could not be ascertained in this study.

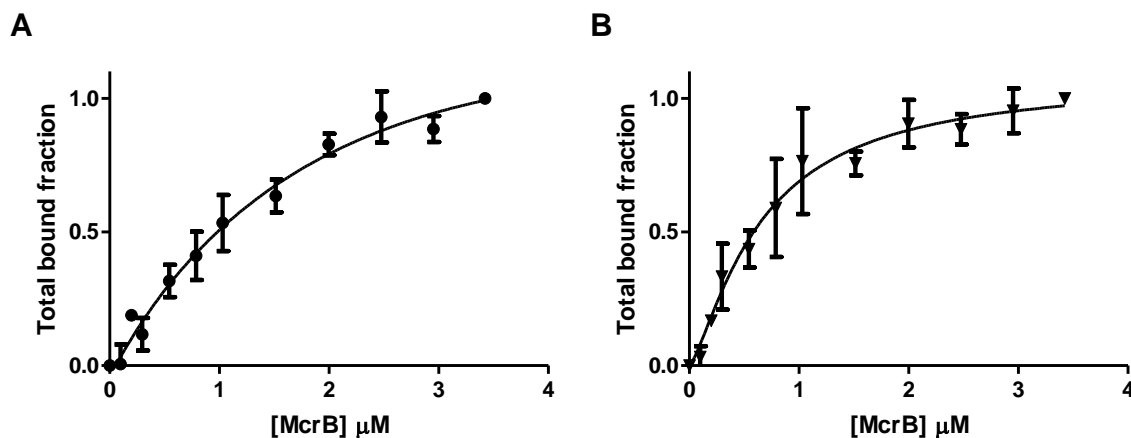


Figure 5.9 **Effect of McrC on DNA binding by McrB.** A) Steady-state anisotropy change of 500 nM McrB and B) steady-state anisotropy change of 500 nM McrB and 125 nM McrC. GTP concentration was 1 mM and labelled DNA (mDNA) concentration was 5 nM. The data is obtained by averaging four experiments and fit to a single site specific binding model using GraphPad Prism 5

## 5.4 Discussion

### 5.4.1 *DNA binding affects the coupling of GTPase reaction rather than its kinetics*

The data in this chapter demonstrates that oligomerization is driven by nucleotide binding and that full oligomerization is required before GTP hydrolysis starts. In the absence of DNA, the GTPase reaction shows a sigmoidal dependence on GTP but this effect is lost upon DNA binding, which also activates the rate moderately indicating a different kinetic mode of the motor in presence of DNA.

The rate of GTP hydrolysis of McrBC was relatively low compared to the rates expected of a stepping motor. In the case of the Type III RM enzymes, a similarly low ATPase rate was shown by stopped-flow to comprise two phases- a rapid burst phase followed by a slower steady-state. The rapid burst phase was due to specific DNA association and a key to the communication mechanism of the Type III enzymes (Schwarz et al, 2013). Thus, the GTPase activation elicited by binding of a methylated

target DNA in this study is rather small and points towards the possibility of an alternative role for GTP than translocation.

#### **5.4.2 *McrBC does not disassemble and reassemble around the substrate DNA***

Ring-like hexameric helicases and translocases engage and translocate their substrate DNA through loops or  $\beta$ -hairpin strands extending from central channel. The ring-like shape requires an additional mechanism for these enzymes to load onto their substrates (Figure 4.1). Based on studies on several helicases and translocases, two models for DNA loading are very popular: 1) either these complexes assemble as rings around the substrate or, 2) if they exist as a complex, the ring either disassembles to assemble around the substrate completely or open partially to accommodate the slipping of DNA into the central channel.

Under physiological conditions, McrBC exists as a preformed complex. *In vitro* experiments (Chapter 2) have conclusively shown that DNA is not required for the assembly of this complex. Thus assembly in presence of substrate is not a mode of loading for McrBC. Although a small substrate was used for these studies, previous studies have shown that this enzyme is capable of binding and cleaving long linear DNA substrates and circular plasmid substrates (Raleigh & Wilson, 1986; Stewart & Raleigh, 1998; Sutherland et al, 1992). Thus threading of substrate from the top of the ring is ruled out as a mode of loading of this complex. The other two possible modes are either disassembly of the complex upon DNA interaction followed by assembly or ring opening to facilitate entry of DNA into the central channel.

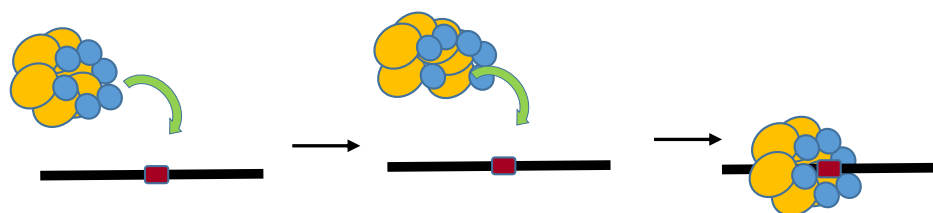
Comparison of normalized traces of these two experiments showed that the kinetics of Trp fluorescence change with preformed complex was very different from the simultaneous mixing of DNA and GTP (Figure 5.8 C). The lag observed in simultaneous mixing experiment suggested an inhibition of oligomerization by DNA. If a preformed complex undergoes disassembly, such inhibition of oligomerization would be reflected in both Trp fluorescence change and DNA anisotropy change. Figure 5.7 shows that no lag is observed in Trp fluorescence change when preformed complex bound DNA. The anisotropy changes (Figure 5.7 B and C) were also faster indicating faster binding by preformed oligomers.

### 5.4.3 Possible mode of assembly of McrBC complex

In the studies presented in this Chapter, it was observed that presence of McrC did not affect kinetics of change in fluorescence during oligomerization (Figure 5.3 A) or during DNA association by preformed McrBC complex (Figure 5.7 A). But knowing the architecture of the complex, a DNA threading through the central pore of McrB ring would call for a large structural rearrangement. Whether the fast Trp fluorescence decay in a preformed oligomer (Figure 5.7 A) is due to conformational changes caused by ring opening or due to DNA wrapping around the ring (Figure 5.10), as suggested in earlier studies (Sukackaite et al, 2012), could not be distinguished from these experiments. The hexameric helicase Rho transcription terminator has been shown to bind to substrate RNA at the specific site first via a primary N-terminal oligonucleotide binding site followed by RNA wrapping and then binding to a secondary binding site that resides inside the central channel (Skordalakes & Berger, 2003; Thomsen & Berger, 2009). No such secondary binding site was identified in case of McrB and thus further makes a threading model less anticipated.

The work carried out in this study clearly show that nucleotide and DNA interactions are not completely independent events though the extent of DNA stimulated GTPase activity is only 2 fold. Since DNA is capable of affecting the oligomerization kinetics of McrB/McrBC, it is important to understand whether there exists a regulatory mechanism for McrB exercised by DNA. Further studies on this complex can help understand the basis of its regulation via assembly of McrBC. Since

#### Substrate induced ring opening



#### Substrate interaction without ring opening

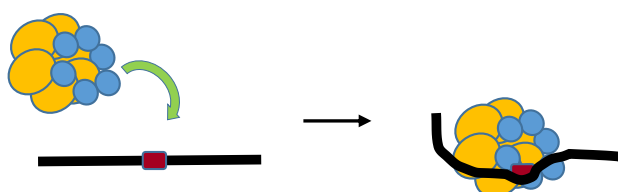


Figure 5.10. Plausible models for DNA loading by McrB/McrBC.

the GTPase activity of McrBC is very low and fold stimulation by DNA is also quite small, it is worth exploring whether GTP hydrolysis by McrBC acts as an activation switch like Type III enzymes. Thus further structure-function studies are required to add more dimensions to the understanding of this complex.

## **5.5 Conclusion and Future Directions**

The thesis presented studies aimed at understanding the mechanism of the modification dependent restriction enzyme McrBC. This complex specifically cleaves DNA containing two or more R<sup>m</sup>C sites in a GTP hydrolysis dependent manner (Panne et al, 2001; Panne et al, 1999; Pieper et al, 1997; Pieper & Pingoud, 2002). The oligomeric structure of McrBC is essential for its functional activities. SEC-MALS and cryo-EM studies conclusively established that McrB exists as a hexamer in presence of GTP and addition of McrC leads to formation of a teradecamer consisting of two McrB hexamers and two McrC monomers (Chapter 2 and 4). Mutation of a cluster of positively charged residues, RRK, in the pre-Sensor loop of McrB affected its oligomerization (Chapter 3). The study also established that mutation of the Walker B aspartate prevented nucleotide-driven oligomerization, while mutation of the Walker B glutamate inhibited nucleotide hydrolysis (Chapter 3). Both mutational and structural analysis of McrB suggested lack of a Sensor 2 arginine (Chapter 3 and 4).

Architecture of McrB $\Delta$ NC determined at 7.4 Å by cryo-EM revealed a unique arrangement of the AAA+ ring of McrB $\Delta$ N and its GTPase-activator McrC (Chapter 4). Analysis of structure showed that the buried surface areas at the interface of the subunits of McrB were variable, and that McrC interacted with the ring-like McrB asymmetrically. Based on the interface area between the subunits of the McrB hexamer, we proposed that GTP binds McrB $\Delta$ NC at three of the six binding pockets per McrB hexamer sequentially. These observations led us to propose a sequential mode of nucleotide hydrolysis stimulated by McrC. We do note that the interface area calculated from the structural model derived from the 7.4 Å map may not be a conclusive evidence for the presence of non-hydrolyzed nucleotide. Based on the structural information we cannot rule out the possibility of nucleotide hydrolysis being stochastic, though literature suggests that a larger number of AAA+ proteins employ a sequential mode over a stochastic mode. The asymmetric nature of the structure, however, rules out the possibility of a concerted mode of nucleotide hydrolysis.



The structural model of McrB $\Delta$ NC at 7.4 Å does not provide in atomic detail the molecular basis of how McrC activates McrB GTPase. Towards this, efforts have to be made in future to improve the resolution by carrying out further cryo-EM studies to obtain more micrographs and thus increase the number of particles used in 3D reconstruction. As described in Chapter 4, the available crystallographic data is of a much higher resolution and phase information has to be improved experimentally and computationally to build the crystal structure. In addition, structure determination of McrB $\Delta$ N hexamer and its comparison with the structure of McrB $\Delta$ NC would provide additional molecular details of activation of the GTPase. The high-resolution structural information would also provide the precise location and stoichiometry of the nucleotides and mechanistic insights into the coupling of the GTPase with the nuclease.

Chapter 5 of the thesis described results of the study carried out to observe real time protein dynamics during oligomerization, GTP hydrolysis and DNA binding by McrB/McrBC. For this, change in intrinsic tryptophan fluorescence was monitored using stopped-flow spectroscopy. These studies indicated that oligomerization of McrBC preceded the steady state of GTPase activity. This result is consistent with the assumption made in the preceding chapters that oligomerization is essential for the functional activities of McrBC. The fluorescence change upon GTP binding and oligomerization, observed in real time, showed multiple exponential steps. But this data was not analyzed further due to lack of complementary experiments such as observation of tryptophan fluorescence change as a function of nucleotide concentration, type of nucleotide etc. These experiments need to be performed in future to better understand the oligomerization pathway

A puzzling feature revealed by the structure was that McrC interacts with the asymmetric ring of McrB $\Delta$ N such that it blocks the central pore. This is perplexing because in helicases and translocases that have been studied in detail, the substrate DNA threads and gets translocated through the open central pore (Ahnert et al, 2000; Carney & Trakselis, 2016; O'Shea & Berger, 2014; Patel & Picha, 2000; Sanders et al, 2007). Consequently, an important question that is raised by this study is the mode of DNA binding by McrBC. Since a loop of McrC is blocking the pore, it is not clear if the substrate DNA would thread through the pore displacing the McrC loop or adopt an alternate mode of binding, such as wrapping around the oligomer. One of the

results described in Section 5.4.2 rules out the possibility of McrBC disassembling on encountering substrate DNA and reassembling around it at the recognition sequence. The other alternate models include either a) a large conformational change facilitating ring opening to accommodate both the McrC loop and the DNA or just the DNA, or b) the DNA wraps around the complex without entering the central pore. Future biophysical and structural studies will aim at deciphering which of the two models are right.

In conclusion, the work described in this thesis provides important insights and lays foundation for future work towards understanding the working of McrBC and the fundamental functional principles of the ubiquitous ring shaped AAA+ motor.

## References:

Abrahams JP, Leslie AG, Lutter R, Walker JE (1994) Structure at 2.8 Å resolution of F<sub>1</sub>-ATPase from bovine heart mitochondria. *Nature* **370**: 621-628

Ahnert P, Picha KM, Patel SS (2000) A ring-opening mechanism for DNA binding in the central channel of the T7 helicase-primase protein. *The EMBO journal* **19**: 3418-3427

Arias-Palomo E, O'Shea VL, Hood IV, Berger JM (2013) The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell* **153**: 438-448

Bedford L, Paine S, Sheppard PW, Mayer RJ, Roelofs J (2010) Assembly, structure, and function of the 26S proteasome. *Trends Cell Biol* **20**: 391-401

Ben-Yehuda S, Rudner DZ, Losick R (2003) Assembly of the SpoIIIE DNA translocase depends on chromosome trapping in *Bacillus subtilis*. *Current biology : CB* **13**: 2196-2200

Brune M, Hunter JL, Corrie JE, Webb MR (1994) Direct, real-time measurement of rapid inorganic phosphate release using a novel fluorescent probe and its application to actomyosin subfragment 1 ATPase. *Biochemistry* **33**: 8262-8271

Carney SM, Trakselis MA (2016) The excluded DNA strand is SEW important for hexameric helicase unwinding. *Methods* **108**: 79-91

Chang YP, Xu M, Machado AC, Yu XJ, Rohs R, Chen XS (2013) Mechanism of origin DNA recognition and assembly of an initiator-helicase complex by SV40 large tumor antigen. *Cell reports* **3**: 1117-1127

Costa A, Pape T, van Heel M, Brick P, Patwardhan A, Onesti S (2006) Structural studies of the archaeal MCM complex in different functional states. *Journal of structural biology* **156**: 210-219

Crampton DJ, Ohi M, Qimron U, Walz T, Richardson CC (2006) Oligomeric states of bacteriophage T7 gene 4 primase/helicase. *Journal of molecular biology* **360**: 667-677

Crozat E, Grainge I (2010) FtsK DNA translocase: the fast motor that knows where it's going. *Chembiochem : a European journal of chemical biology* **11**: 2232-2243

Dean FB, Borowiec JA, Eki T, Hurwitz J (1992) The simian virus 40 T antigen double hexamer assembles around the DNA at the replication origin. *The Journal of biological chemistry* **267**: 14129-14137

Gast FU, Brinkmann T, Pieper U, Kruger T, Noyer-Weidner M, Pingoud A (1997) The recognition of methylated DNA by the GTP-dependent restriction endonuclease McrBC resides in the N-terminal domain of McrB. *Biological chemistry* **378**: 975-982

Giacherio D, Hager LP (1979) A poly(dT)-stimulated ATPase activity associated with simian virus 40 large T antigen. *The Journal of biological chemistry* **254**: 8113-8116

Gomez-Lorenzo MG, Valle M, Frank J, Gruss C, Sorzano CO, Chen XS, Donate LE, Carazo JM (2003) Large T antigen on the simian virus 40 origin of replication: a 3D snapshot prior to DNA replication. *The EMBO journal* **22**: 6205-6213

Kelman Z, Lee JK, Hurwitz J (1999) The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* DeltaH contains DNA helicase activity. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 14783-14788

Kress W, Mutschler H, Weber-Ban E (2007) Assembly pathway of an AAA+ protein: tracking ClpA and ClpAP complex formation in real time. *Biochemistry* **46**: 6183-6193

Kruger T, Wild C, Noyer-Weidner M (1995) McrB: a prokaryotic protein specifically recognizing DNA containing modified cytosine residues. *The EMBO journal* **14**: 2661-2669

Lakowicz JR (2006) *Principles of fluorescence spectroscopy*, 3rd edn. New York: Springer.

Lakowicz JR, Gryczynski I (1992) Tryptophan fluorescence intensity and anisotropy decays of human serum albumin resulting from one-photon and two-photon excitation. *Biophysical chemistry* **45**: 1-6

Laurent BC, Treich I, Carlson M (1993) The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. *Genes & development* **7**: 583-591

Le DT, Eckert T, Woehlke G (2013) Computer simulation of assembly and cooperativity of hexameric AAA ATPases. *PLoS one* **8**: e67815

Martin A, Baker TA, Sauer RT (2008) Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Mol Cell* **29**: 441-450

Massey TH, Mercogliano CP, Yates J, Sherratt DJ, Lowe J (2006) Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. *Molecular cell* **23**: 457-469

Medalia N, Beer A, Zwickl P, Mihalache O, Beck M, Medalia O, Navon A (2009) Architecture and molecular mechanism of PAN, the archaeal proteasome regulatory ATPase. *The Journal of biological chemistry* **284**: 22952-22960

O'Shea VL, Berger JM (2014) Loading strategies of ring-shaped nucleic acid translocases and helicases. *Curr Opin Struct Biol* **25**: 16-24

Panne D, Muller SA, Wirtz S, Engel A, Bickle TA (2001) The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *The EMBO journal* **20**: 3210-3217

Panne D, Raleigh EA, Bickle TA (1999) The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. *J Mol Biol* **290**: 49-60

Patel SS, Picha KM (2000) Structure and function of hexameric helicases. *Annu Rev Biochem* **69**: 651-697

Pieper U, Brinkmann T, Kruger T, Noyer-Weidner M, Pingoud A (1997) Characterization of the interaction between the restriction endonuclease McrBC from *E. coli* and its cofactor GTP. *J Mol Biol* **272**: 190-199

Pieper U, Pingoud A (2002) A mutational analysis of the PD...D/EXK motif suggests that McrC harbors the catalytic center for DNA cleavage by the GTP-dependent restriction enzyme McrBC from *Escherichia coli*. *Biochemistry* **41**: 5236-5244

Raleigh EA, Wilson G (1986) Escherichia coli K-12 restricts DNA containing 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America* **83**: 9070-9074

Rodriguez-Aliaga P, Ramirez L, Kim F, Bustamante C, Martin A (2016) Substrate-translocating loops regulate mechanochemical coupling and power production in AAA+ protease ClpXP. *Nature structural & molecular biology* **23**: 974-981

Sanders CM, Kovalevskiy OV, Sizov D, Lebedev AA, Isupov MN, Antson AA (2007) Papillomavirus E1 helicase assembly maintains an asymmetric state in the absence of DNA and nucleotide cofactors. *Nucleic Acids Res* **35**: 6451-6457

Sauer RT, Baker TA (2011) AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem* **80**: 587-612

Schuck S, Stenlund A (2005) Assembly of a double hexameric helicase. *Molecular cell* **20**: 377-389

Schuck S, Stenlund A (2011) Mechanistic analysis of local ori melting and helicase assembly by the papillomavirus E1 protein. *Molecular cell* **43**: 776-787

Schwarz FW, Toth J, van Aelst K, Cui G, Clausing S, Szczelkun MD, Seidel R (2013) The helicase-like domains of type III restriction enzymes trigger long-range diffusion along DNA. *Science* **340**: 353-356

Siddiqui SM, Sauer RT, Baker TA (2004) Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes Dev* **18**: 369-374

Skordalakes E, Berger JM (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* **114**: 135-146

Stewart FJ, Raleigh EA (1998) Dependence of McrBC cleavage on distance between recognition elements. *Biological chemistry* **379**: 611-616

Sukackaite R, Grazulis S, Tamulaitis G, Siksnyys V (2012) The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. *Nucleic acids research* **40**: 7552-7562

Sutherland E, Coe L, Raleigh EA (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *Journal of molecular biology* **225**: 327-348

Thomsen ND, Berger JM (2009) Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* **139**: 523-534

Toth EA, Li Y, Sawaya MR, Cheng Y, Ellenberger T (2003) The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Molecular cell* **12**: 1113-1123

Valle M, Gruss C, Halmer L, Carazo JM, Donate LE (2000) Large T-antigen double hexamers imaged at the simian virus 40 origin of replication. *Molecular and cellular biology* **20**: 34-41

Webb MR (2007) Development of fluorescent biosensors for probing the function of motor proteins. *Molecular bioSystems* **3**: 249-256

Yengo CM, Chrin L, Rovner AS, Berger CL (1999) Intrinsic tryptophan fluorescence identifies specific conformational changes at the actomyosin interface upon actin binding and ADP release. *Biochemistry* **38**: 14515-14523

Yu X, VanLoock MS, Poplawski A, Kelman Z, Xiang T, Tye BK, Egelman EH (2002) The Methanobacterium thermoautotrophicum MCM protein can form heptameric rings. *EMBO reports* **3**: 792-797