Structural and biochemical studies of the Type ISP restriction-modification enzyme LlaBIII

A Thesis

Submitted in partial fulfilment of the requirements

for the degree of
Doctor of Philosophy

By

Mahesh Kumar Chand
20123180

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE

2019

I would like to dedicate my thesis to my beloved grandparents

Indian Institute of Science Education and Research, Pune

Dr. Homi Bhabha Road, Pashan, Pune, 411008, India

Saikrishnan Kayarat, PhD
Associate Professor
Division of Biology

CERTIFICATE

Certified that the work incorporated in the thesis entitled **"Structural and biochemical studies of the Type ISP restriction-modification enzyme LlaBIII"** submitted by **Mr. Mahesh Kumar Chand** 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.

Dr. Saikrishnan Kayarat

Supervisor

Date: 19/03/2019

Tel.: +91 020 2590 8047   Fax: +91 020 2590 8186   email: saikrishnan@iiserpune.ac.in

DECLARATION

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

Mahesh Kumar Chand

Reg. No. 20123180

Date: 19/03/2019

Acknowledgements

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abbreviations

°C	Degree Celsius
µl	Microliters
µm	Micrometers
µM	Micromolar
Å	Angstrom
AdoMet	S-adenosyl-L-methionine
ADP	Adenosine diphosphate
AHJR	Archeal Holliday Junction Resolvase
AMPPNP	Adenylyl-imidodiphosphate
ATP	Adenosine triphosphate
bp	Base pair
CD	Circular dichroism
CTD	C-terminal domain
dsDNA	Double stranded deoxyribonucleic acid
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
GTP	Guanosine triphosphate
HsdM	Host specificity for DNA modification
HsdR	Host specificity for DNA restriction
HsdS	Host specificity for DNA specificity
ICR	Immigration control region
IDT	Integrated DNA technologies
Pi	Inorganic phosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Kelvin
Kb	Kilobase
kDa	Kilodalton

mg	Milligram
ml	Millilitre
Mod	Modification
Mrr	Methylated adenine recognition and restriction
MTase	Methyltransferase
nm	Nanometre
nM	Nanomolar
NTD	N-terminal domain
NTP	Nucleotide triphosphate
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
P-loop	Phosphate binding loop
REBASE	Restriction enzyme database
RecA	Recombinase A
RecB	Recombinase B
Res	Restriction
RM	Restriction-modification
SDS	Sodium dodecyl sulphate
SF2	Superfamily 2
ssDNA	Single stranded deoxyribonucleic acid
TFO	Triplex forming oligonucleotide
TRD	Target recognition domain

Synopsis

Structural and biochemical studies of the Type ISP restriction-modification enzyme LlaBIII

Name: Mahesh Kumar Chand

Roll number: 20123180

Name of supervisor: Dr. Saikrishnan Kayarat

Department: Biology, IISER Pune

Date of registration: 1st August 2012

Chapter 1: Type I and Type ISP restriction-modification enzymes

Bacteria are under constant threat from bacteriophage attacks. In order to survive, they have evolved various defence mechanisms to combat their natural parasites (Dryden et al., 2001). One such mechanism is to nucleolytically cleave the viral DNA entering the bacterial cells. These activities are carried out by a class of enzymes called restriction-modification (RM) systems. The modification component of the RM enzymes methylates a particular DNA target sequence, while the restriction component nucleolytically cleaves DNA having an unmodified target (Dryden et al., 2001). The restriction activity of the RM enzymes can be broadly classified into two categories- ATP-dependent and ATP-independent. The heteropentameric Type I and the heterotrimeric Type III RM enzymes have both the restriction and modification component together and hydrolyse ATP to cleave DNA (Butterer et al., 2014; Davies et al., 1999; Gupta et al., 2015). In contrast, the restriction and the modification components of Type II systems are, in general, separate enzymes and do not utilise ATP (Pingoud et al., 2014). Furthermore, Type I RM enzymes are unique in cutting DNA hundred to a few thousand base pair (bp) away from their target sites (Studier and Bandyopadhyay, 1988). A Type I RM enzyme bound to its target actively moves (translocate) along DNA powered by ATP hydrolysis and two converging and colliding enzymes of the same kind lead to DNA cleavage. In addition to their physiological importance, these enzymes serve as paradigms for understanding modular, multifunctional protein machines (Murray, 2000), particularly in formulating concepts of protein-DNA recognition, DNA methylation, nuclease activity (Rao et al., 2014; Smith et al., 2009a), double-stranded (ds) DNA

translocation by superfamily 2 (SF2) helicases (Durr et al., 2006; Stanley et al., 2006) and long-range communication by enzymes (Halford et al., 2004; Schwarz et al., 2013).

Recently characterised Type ISP RM enzymes are single polypeptide multidomain enzymes, comprising of the functional domains of Type I RM enzymes. These enzymes contain N-terminal Mrr nuclease domain followed by the SF2 helicase-like ATPase domain, N6 adenine methyltransferase and a C-terminal target recognition domain (TRD). Earlier biochemical studies demonstrated that Type ISP enzymes require at least two targets oriented head-to-head for nucleolytic cleavage (Sisakova et al., 2013; Smith et al., 2009a; Smith et al., 2009b, c; van Aelst et al., 2013). These features of Type ISP enzymes are, in general, similar to that of Type I enzymes. Based on the mechanism of DNA restriction by Type I enzymes deciphered using biophysical studies (Seidel et al., 2008), it was proposed that the Type ISP RM enzymes would also hydrolyse ATP and translocate DNA forming a loop (Smith et al., 2009c). Two such enzymes bound to targets oriented head-to-head would eventually converge and collide bringing together and activating the respective nucleases. The activated nucleases of the collision complex would nick one of the two strands, respectively, thus causing double-strand (ds) DNA break formation.

Unlike the well-studied Type II RM enzymes, the enzymatic mechanism of DNA restriction and modification by ATP-dependent RM enzymes are unclear owing to their molecular complexity. Multiple efforts for obtaining the atomic details of these enzymes provided us with the crystal structures of individual subunits (Kennaway et al., 2009; Kim et al., 2005; Lapkouski et al., 2009; Liu et al., 2017). The first insights into the molecular organisation of ATP-dependent enzymes came from structural analysis of the Type I RM enzymes EcoKI and EcoR124I using a combination of negative-stain electron microscopy, neutron scattering and structural modelling (Kennaway et al., 2012). Lack of structural information of these enzyme systems is a big lacuna, obtaining which has been hindered by their large size. The proposed work aims to gain insights into the mechanism of an ATP-dependent RM enzyme by structural studies using X-ray crystallography tools of the Type ISP RM enzyme. Structural studies on these ATP dependent RM enzymes will also help in understanding other biological phenomenon carried out by the enzymes containing homologous domains. For example, a) the homologues of the ATPase domain essential for the nucleolytic activity of the RM enzyme also perform chromatin remodelling and transcription modulation; b) coupling of homologous nuclease-ATPase domains are essential for DNA repair.

To understand the molecular architecture and mechanism of action of ATP dependent RM enzymes, the following objectives were proposed-

1. Purification, biochemical characterisation and crystallisation of Type ISP RM enzyme LlaBIII

2. Structural characterisation of LlaBIII and dsDNA translocation

3. DNA-mediated coupling of the ATPase, translocase and nuclease activities of a Type ISP restriction-modification enzyme

Chapter 2: Purification, biochemical characterisation and crystallisation of Type ISP RM enzyme LlaBIII

In our efforts to crystallise Type ISP RM enzyme LlaBIII, the first step was the purification of the active and highly pure protein. These multidomain enzymes are highly unstable and start degrading during purification. As part of a bigger goal to gain mechanistic insights into how multi-domain protein machines function, in this chapter we discuss our efforts towards purification of homogeneous monomeric LlaBIII, its biochemical characterisation and crystallisation strategy. The purified protein was active in its DNA cleavage activity. We further checked the ATPase and DNA translocation activity of the purified protein. It showed stimulation in the ATPase activity in the presence of dsDNA containing a target site and was efficient in displacing the triplex-forming oligonucleotide from the triplex DNA.

DNA binding experiments were instrumental in deciding the dsDNA for co-crystallisation with LlaBIII. The DNA binding experiment provided us with the stoichiometry of the protein to DNA as 1:1.3 for forming a homogeneous protein-DNA complex. It also helped in deciding the length of DNA for crystallisation. This paved the path towards obtaining a crystal of LlaBIII with 28 bp DNA that diffracted to 2.7 Å.

Chapter 3: Structural characterisation of LlaBIII and dsDNA translocation

The structure of LlaBIII bound to 28 bp DNA revealed six structural domains- the N-terminal Mrr family nuclease followed by the N-core and C-core RecA folds of the ATPase located upstream of the target site and an all α-helical domain that links the ATPase to the MTase and the C-terminal TRD. The target site on the DNA showed extensive interaction with the TRD and MTase domains and only a few interactions with a unique β-hairpin loop present in the N-core domain of the SF2 helicase-like ATPase domain of the enzyme. The TRD can be divided into 3 subdomains – the core domain (involved in major interaction with the DNA), the jaw (fastens the clamp on the DNA) and the guide (steers the DNA towards the ATPase domain). The structure also showed a flipped adenine base in the MTase catalytic pocket regardless of the absence of AdoMet in the crystallisation condition. This shows that the enzyme is competent to carry out DNA methylation

or translocation in the presence of appropriate cofactor. Bestowed with the slow rate of methylation, these enzymes will prefer to translocate on the DNA rather methylating the target site (Chand et al., 2015).

The organisation of LlaBIII on DNA suggests that translocation initiation along the DNA will require remodelling of the extensive interactions made by the TRD-MTase clamp leading to target release, contrary to the proposed translocation mechanism by DNA looping. Also, the upstream location of the nuclease domain seems too distant to either interact directly or produce a dsDNA break. In collaboration with Prof. Mark Szczulkun, Bristol University, UK, using single molecule magnetic tweezers assay it was shown that Type ISP RM enzymes translocate on DNA without looping (Chand et al., 2015). Based on these observations and single cleavage event mapping (Chand et al., 2015) we propose a DNA shredder model for Type ISP RM enzymes. According to this model, upon recognition of target sequence, the ATPase domain starts ATP hydrolysis and DNA translocation leading to the conformational change in the TRD-MTase clamp and target release. When two translocating enzymes converge, they form the collision complex of two active translocases. This complex nicks the DNA on both the strands and further translocation leads to multiple nicks on the DNA strand, resulting in shredding of the DNA (Chand et al., 2015).

Chapter 4: DNA-mediated coupling of the ATPase, translocase and nuclease activities of a Type ISP restriction-modification enzyme

After characterisation of the dsDNA translocation by Type ISP RM enzymes, we further embarked on understanding the mechanism of ATPase domain activation. Firstly, we tried to understand the role of DNA in activating the ATPase domain, as earlier reports suggested that presence of the target site on the DNA stimulates the ATPase activity of the RM enzymes (Smith et al., 2009c). Here we show that ATPase activation is a two-step process. The recognition of target site leads to partial activation of ATPase activity whereas the complete activation requires a sufficient length of DNA to interact with the ATPase domain. Based on ATPase assay in the presence of different DNA length, we show 21 bp upstream DNA length is required for the complete ATPase activation in case of Type ISP RM enzyme LlaBIII. Further by mutagenesis, we showed the involvement of the β-hairpin loop of the N-core and motif V of C-core ATPase domain in coupling the ATPase to DNA translocase and nuclease activity.

The β-hairpin loop deletion mutant showed a reduced ATPase activity, and it was unable to translocate on the DNA leading to no nucleolytic activity. While the motif III mutant which is known to couple ATPase activity to translocation activity in SF2 helicase was able to cleave the substrate DNA with less efficiency. Inspite of less ATPase activity of motif III mutant as compared to β-hairpin loop deletion mutant it was able to translocate on the DNA. Similarly, Arg564 to

alanine mutation, a motif V mutant showed ATPase activity same as that of LlaBIII wild type and a slow translocation activity but no DNA cleavage activity. We also show that the nuclease domain activation of these enzymes depends on the speed of DNA translocation, by a heterologous cooperation assay. The slow translocating LlaBIII[R564A] mutant showed DNA cleavage upon cooperation with another active Type ISP RM enzyme LlaGI whereas the β-hairpin loop deletion mutant still showed no DNA cleavage activity. This suggested that the translocation activity not only brings two nuclease domains together cleave dsDNA but the translocation initiation and rate of translocation regulates the activation of the nucleases.

In summary, these studies provided new mechanistic insights into the functioning of the macromolecular machines. The X-ray crystal structure of an ATP-dependent Type ISP RM enzyme LlaBIII was determined, providing detailed information on the molecular architecture of the protein. This structure formed the basis for the postulation of a new mechanism of double-strand DNA break formation by the concerted action of two distal nucleases that converge to make multiple nicks on the DNA strands. The studies on the ATPase domain led us to identify a new structural motif β-hairpin loop, involved in the coupling of the ATP hydrolysis to DNA translocation. The studies on β-hairpin loop and motif V showed dsDNA break by Type ISP requires both ATP hydrolysis and initiation of DNA translocation. We show that the nuclease activation depends on the efficiency of translocation and force produced on the collision of two enzymes.

References:

Butterer, A., Pernstich, C., Smith, R.M., Sobott, F., Szczelkun, M.D., and Toth, J. (2014). Type III restriction endonucleases are heterotrimeric: comprising one helicase-nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. Nucleic acids research *42*, 5139-5150.

Chand, M.K., Nirwan, N., Diffin, F.M., van Aelst, K., Kulkarni, M., Pernstich, C., Szczelkun, M.D., and Saikrishnan, K. (2015). Translocation-coupled DNA cleavage by the Type ISP restriction-modification enzymes. Nature chemical biology *11*, 870-877.

Davies, G.P., Martin, I., Sturrock, S.S., Cronshaw, A., Murray, N.E., and Dryden, D.T. (1999). On the structure and operation of type I DNA restriction enzymes. Journal of molecular biology *290*, 565-579.

Dryden, D.T., Murray, N.E., and Rao, D.N. (2001). Nucleoside triphosphate-dependent restriction enzymes. Nucleic acids research *29*, 3728-3741.

Durr, H., Flaus, A., Owen-Hughes, T., and Hopfner, K.P. (2006). Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures. Nucleic acids research *34*, 4160-4167.

Gupta, Y.K., Chan, S.H., Xu, S.Y., and Aggarwal, A.K. (2015). Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. Nature communications *6*, 7363.

Halford, S.E., Welsh, A.J., and Szczelkun, M.D. (2004). Enzyme-mediated DNA looping. Annual review of biophysics and biomolecular structure *33*, 1-24.

Kennaway, C.K., Obarska-Kosinska, A., White, J.H., Tuszynska, I., Cooper, L.P., Bujnicki, J.M., Trinick, J., and Dryden, D.T. (2009). The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. Nucleic acids research *37*, 762-770.

Kennaway, C.K., Taylor, J.E., Song, C.F., Potrzebowski, W., Nicholson, W., White, J.H., Swiderska, A., Obarska-Kosinska, A., Callow, P., Cooper, L.P., *et al.* (2012). Structure and operation of the DNA-translocating type I DNA restriction enzymes. Genes & development *26*, 92-104.

Kim, J.S., DeGiovanni, A., Jancarik, J., Adams, P.D., Yokota, H., Kim, R., and Kim, S.H. (2005). Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. Proceedings of the National Academy of Sciences of the United States of America *102*, 3248-3253.

Lapkouski, M., Panjikar, S., Janscak, P., Smatanova, I.K., Carey, J., Ettrich, R., and Csefalvay, E. (2009). Structure of the motor subunit of type I restriction-modification complex EcoR124I. Nature structural & molecular biology *16*, 94-95.

Liu, Y.P., Tang, Q., Zhang, J.Z., Tian, L.F., Gao, P., and Yan, X.X. (2017). Structural basis underlying complex assembly and conformational transition of the type I R-M system. Proceedings of the National Academy of Sciences of the United States of America *114*, 11151-11156.

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

Pingoud, A., Wilson, G.G., and Wende, W. (2014). Type II restriction endonucleases--a historical perspective and more. Nucleic acids research *42*, 7489-7527.

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

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

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

Sisakova, E., van Aelst, K., Diffin, F.M., and Szczelkun, M.D. (2013). The Type ISP Restriction-Modification enzymes LlaBIII and LlaGI use a translocation-collision mechanism to cleave non-specific DNA distant from their recognition sites. Nucleic acids research *41*, 1071-1080.

Smith, R.M., Diffin, F.M., Savery, N.J., Josephsen, J., and Szczelkun, M.D. (2009a). DNA cleavage and methylation specificity of the single polypeptide restriction-modification enzyme LlaGI. Nucleic acids research *37*, 7206-7218.

Smith, R.M., Josephsen, J., and Szczelkun, M.D. (2009b). An Mrr-family nuclease motif in the single polypeptide restriction-modification enzyme LlaGI. Nucleic acids research *37*, 7231-7238.

Smith, R.M., Josephsen, J., and Szczelkun, M.D. (2009c). The single polypeptide restriction-modification enzyme LlaGI is a self-contained molecular motor that translocates DNA loops. Nucleic acids research *37*, 7219-7230.

Stanley, L.K., Seidel, R., van der Scheer, C., Dekker, N.H., Szczelkun, M.D., and Dekker, C. (2006). When a helicase is not a helicase: dsDNA tracking by the motor protein EcoR124I. The EMBO journal *25*, 2230-2239.

Studier, F.W., and Bandyopadhyay, P.K. (1988). Model for how type I restriction enzymes select cleavage sites in DNA. Proceedings of the National Academy of Sciences of the United States of America *85*, 4677-4681.

van Aelst, K., Sisakova, E., and Szczelkun, M.D. (2013). DNA cleavage by Type ISP Restriction-Modification enzymes is initially targeted to the 3'-5' strand. Nucleic acids research *41*, 1081-1090.

Chapter: 1   Type I and Type ISP restriction-modification enzymes

1.1 Restriction-Modification enzymes

Bacteria are under constant threat from viral attacks. In order to survive, they have evolved various defence mechanisms to combat their natural parasite. One such system is the Restriction-Modification (RM) system. These enzymes cleave viral DNA/foreign DNA entering into the bacterial system with its restriction subunit after finding its target sequence, while at the same time protecting the host genome by its modification activity (Figure 1-1).
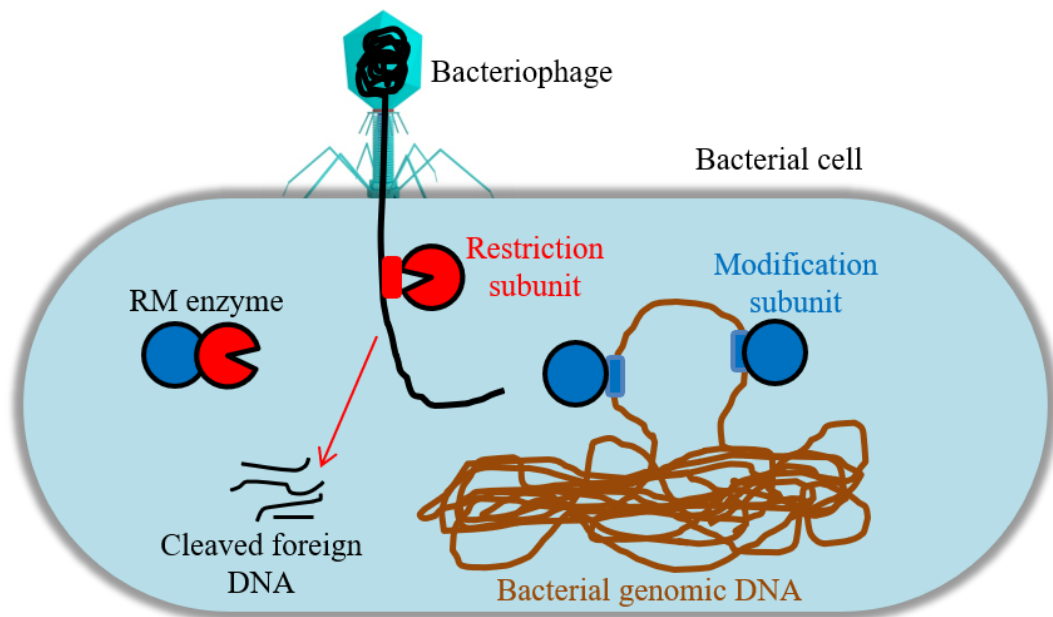
Figure 1-1: Restriction-modification system. The restriction subunit (shown in red) recognises the target sequence (red) on the foreign DNA (in this case its bacteriophage DNA) and cleaves it. The modification subunit methylates the target sequence present in the bacterial genome. Black colour shows foreign DNA whereas a bacterial genome is shown in brown colour.

The modification domain methylates the target sequence present in its genome, hence preventing it from the adverse effect of the restriction domain. They were first reported by Luria and Human in 1952 (Luria and Human, 1952). In the 1960s' Werner Arber demonstrated that these RM enzymes bind to specific DNA sequence (Linn and Arber, 1968), Hamilton Smith in 1970s' verified that these enzymes cut in the middle of the binding sequence (Smith and Wilcox, 1970) and Daniel Nathans pioneered the use of RM enzymes in genetic engineering. In 1978 they shared Nobel Prize in Medicines for their valuable discovery. Till date, 11,340 restriction modification enzymes (including restriction enzymes and methyltransferases) are biochemically characterised, and more than 2.35 lakh putative restriction-modification enzymes are reported (REBASE database as on 26/12/2018). Based on the recognition site, molecular architecture, site of cleavage and cofactor requirements, RM enzymes are classified into four different groups - Type I, Type II, Type III and Type IV. Type II RM enzymes do not require any ATP for their activity, and they cleave at the or close to the binding site. These are the common restriction enzymes which we use

daily for cloning purposes. Type IV enzymes cleaves methylated DNA, whereas Type I and III require ATP and cleave non-methylated DNA. In general Type I and Type III are hetero-oligomeric in nature. Type III contains two subunits (Res and Mod) whereas Type I is pentameric (2 HsdR, 2 HsdM and 1 HsdS subunit). There is also a difference in ATP usage by Type I and Type III to perform dsDNA break. Type III utilises <<1 ATP to translocate 1 bp (84 ATP to translocate 1000bp) while Type I requires 1 to 1.5 ATP to translocate 1 bp (Reiser and Yuan, 1977; Szczelkun et al., 1996). The differences between Type I, Type II, Type III and Type IV RM enzymes are listed in Table 1.1.

Table 1-1: Difference between Type I, Type II, Type III and Type IV RM enzymes.

Features	Type I	Type II	Type III	Type IV
Genes	*hsdS, hsdM, hsdR*	*ecoIR, ecoIM*	*res, mod*	*mcrB, mcrC*
Subunits	HsdS, HsdM, HsdR	Res, Mod	Res, Mod	McrBl, McrBs, McrC
Enzyme activity	REase, MTase, ATPase, DNA translocation	REase, MTase	REase, MTase, ATPase, DNA translocation	REase, GTPase, DNA translocation
Protein composition	3 subunits	monomeric	2 subunits	3 subunits
Cofactors				
For methylation	AdoMet	AdoMet	AdoMet (Mg^{2+})	-
For restriction	AdoMet, ATP, Mg^{2+}	Mg^{2+}	AdoMet, ATP, Mg^{2+}	GTP, Mg^{2+}
Recognition site	Asymmetric, bipartite	Palindromic	Asymmetric	Bipartite
Nature of substrate DNA	Non-methylated DNA	Non-methylated DNA	Non-methylated DNA	Methylated DNA
Restriction Site	1-7kb from recognition site	Inside recognition site	24-26 bp from recognition site	30 bp from recognition site

1.2 Type I RM enzymes

Type I RM systems were the first to be discovered in spite of that they did not find a place in the tool box of molecular biologists. Before genome sequencing, only a few Type I RM systems were

known. By 1980s only two enzymes were characterised namely- EcoBI and EcoKI (Kan et al., 1979; Lautenberger et al., 1979; Ravetch et al., 1978; Sommer and Schaller, 1979). Even today the characterised Type I RM systems are very less as compared to Type II enzymes. This difference in the number of characterised enzymes shows the difficulty in identifying and characterising these enzymes. Characterisation of these enzymes has been difficult because of the ability to cleave the DNA far away from the target sites. But, recent development in DNA sequencing provides a platform for characterising these enzymes. The Pac Bio Single-molecule real-time DNA (SMRT) sequencing can help in characterising these enzymes based on the methylation pattern in the host genome (Kasarjian et al., 2005; Murray et al., 2012) as these enzymes methylate N6 adenine of the target sequence. These are very powerful sequencing methods which can distinguish one nucleotide from other as well as the modification state of the specific nucleotide from no modification (Clark et al., 2012).

1.2.1 Gene and its organisation in genome

Type I RM enzymes are encoded by three genes namely- *hsdS*, *hsdM* and *hsd*R. hsd stands for host specificity determinant. *hsdS* encodes for specificity subunit, *hsdM* encodes for modification subunit, and *hsdR* encodes for restriction subunit (nuclease domain and ATPase domain). These enzymes can be encoded by the bacterial genome (EcoKI), or it can be plasmid encoded (EcoR124). EcoKI and related Type I RM enzymes found in enteric bacteria are often located with *mcrBC* and *mrr* genes in highly divergent locus called immigration control region (Figure 1-2). The alternate DNA segment of this cluster contains Type I RM enzymes (Arber and Wauters-Willems, 1970; Barcus et al., 1995; Bullas et al., 1980; Ryu et al., 1988; Sibley and Raleigh, 2004).

Based on the genetic complementation tests, DNA hybridization, antibody cross-reactivity and amino acid (aa) sequence these Type I RM enzymes are divided into 5 different groups- Type IA (EcoKI), Type IB (EcoAI), Type IC (EcoR124I), Type ID (StySBLI) and Type IE (KpnBI) (Chin et al., 2004; Dryden et al., 2001; Titheradge et al., 2001). High throughput genome sequencing has shown that almost half the bacterial and archaeal population contains Type I RM systems. Most genomes have mainly one kind of Type I RM enzymes, but this number can change from two to eight (e.g. *Desulfococcus oleovorans*) (Loenen et al., 2014a). Some species of *Mycoplasma* contains a higher number of genes coding for HsdS subunit whereas they code only 1 or 2 genes for HsdM and HsdR subunits (e.g. *Mycoplasma haemofelis* contains genes coding for 22 HsdS subunits) (Loenen et al., 2014a). The higher number of HdsS is probably to cope up with the changing defence against the intruder (Dybvig et al., 1998; Sitaraman and Dybvig, 1997).
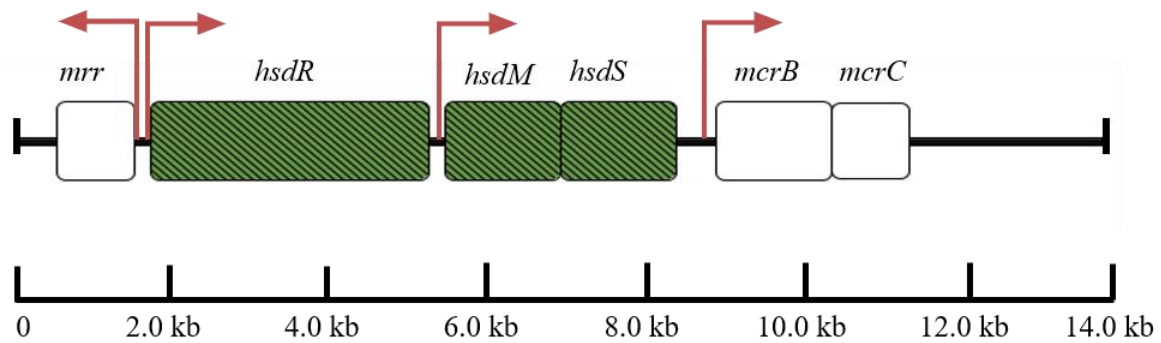
Figure 1-2: The immigration control region of *E. coli* K-12. It contains three restriction enzymes, the hsd cluster coding for Type IA (EcoKI) RM enzyme- hsdR (restriction subunit), hsdM (modification subunit) and hsdS (specificity subunit). Type IV restriction enzymes are coded by the *mcrB* and *mcrC* genes found in this cluster. The mrr enzymes are coded by the *mrr* gene. (Bourniquel and Bickle, 2002)

1.2.2 Enzyme Assembly

Type I RM enzymes perform DNA methylation as well as DNA cleavage activity. These enzymes require $Mg^{2+,}$ S-adenosylmethionine (SAM) and ATP for their activity. SAM is required for methylation activity whereas in the presence of ATP these enzymes start DNA translocation and cleave dsDNA. The active methylase complex is formed from one HsdS and two HsdM subunits whereas the heteropentameric assembly formed by one HsdS, two HsdM and two HsdR functions as both methylation and restriction enzymes (Dryden et al., 2001; Kennaway et al., 2009; Kennaway et al., 2012). These enzymes recognise a bipartite target sequence where one or both adenines are methylated (one from each strand) in the host genome. The holoenzyme prefers to methylate a hemimethylated DNA substrate whereas cleave a non-methylated foreign DNA.

1.2.3 Specificity subunit (HsdS)

Specificity subunit is involved in the target sequence recognition by Type I RM enzymes. It is also the core subunit, which allows the formation of active MTase or restriction complex. It contains two independent target recognition domains, separated by a spacer to recognise two halves of the bipartite sequence (Figure 1-3A). 5' target sequence contains 3-4 bp followed by a non-specific 6-8 bp sequence and 4-5 bp specific 3' sequence (Table 1.2). The N-terminal TRD recognises 5' specific sequence while the C-terminal recognise 3'specific sequence (Cowan et al., 1989; Fuller-Pace and Murray, 1986; Gann et al., 1987; Gubler et al., 1992; Thorpe et al., 1997). HsdS consists of globular target recognition domain and a helical dimerisation domain. This dimerisation domain is highly conserved within a Type I family and less conserved between families. The α-helical dimerisation domain forms a coiled-coil leucine-zipper (Figure 1-3B).
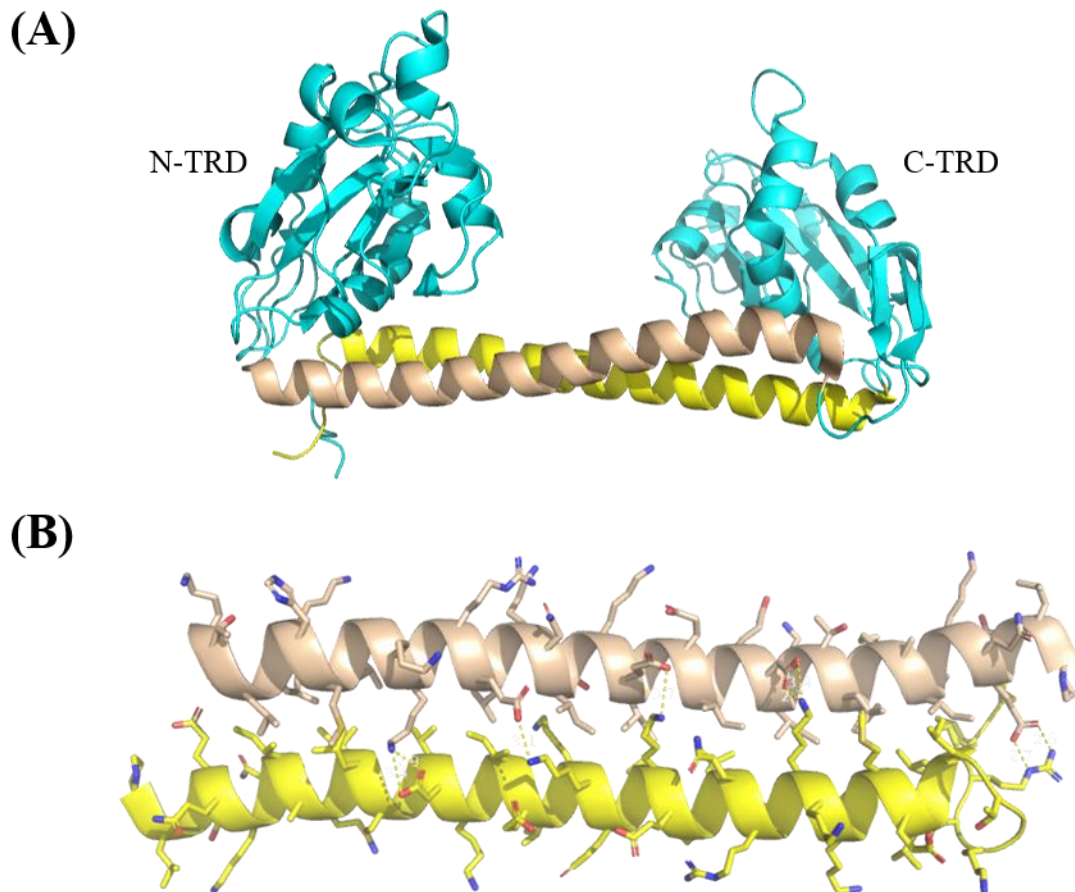
Figure 1-3: The HsdS subunit of Type I RM enzymes. **(A)** The globular domain (shown in cyan) N-TRD and C-TRD are involved in recognition of two halves of the bipartite target sequence. The two TRD (specificity domains) are separated by the dimerisation domain D1 (helix in sandy tan) and D2 (helix in yellow). **(B)** Zoomed in view of dimerisation domains showing an array of H-bonds. (PDB: 1YF2) (Kim et al., 2014).

1.2.3.1 Specificity Changes-

The exchange of TRD from different Type I RM enzymes leads to recognition of hybrid sequences, providing the host bacterium to deal with the modified target sequence of bacteriophages with new specificities (Loenen et al., 2014). Specificity subunit being central to DNA methylation of the host genome and restriction of foreign DNA, changing the recognition sequence of this domain will change the specificities for both MTase and restriction endonucleases (Loenen et al., 2014a). The multiple ways by which these enzymes can change their specificity are discussed below-

1.2.3.1.1 TRD exchange

Specificity of Type I RM enzymes can be changed in multiple ways; one such example is TRD exchange. This was first noticed in 1976 during chromosomal DNA transfer from a strain of *Salmonella* (expressing StySP1) to its derivative (expressing StyLTIII). The transductant contained a new enzyme known as StySQ as a result of genetic cross over between the HsdS gene

of both the enzymes (Bullas et al., 1976). StySQ contains an N-terminal TRD of the StySPI gene and C terminal TRD of the StyLTIII gene (Fuller-Pace et al., 1984; Fuller-Pace and Murray, 1986). Accordingly, it recognises a 5' half site of StySPI and 3' half site of StyLTIII (Nagaraja et al., 1985a; Nagaraja et al., 1985b) (Table 1.2). The domain architecture of StySQ was same as that of the other members of Type IA RM enzymes with N- terminal variable region, a central conserved region, another variable region and a C- terminal conserved region. This domain swapping by genetic cross-over is not only limited to Type IA family, it is also observed in members of other Type I RM families as well (like N-terminal TRD of StySK1 and C terminal TRD of EcoA1 of Type 1B and EcoR124I and EcoDXXI of Type 1C) (Gubler et al., 1992; Thorpe et al., 1997).

1.2.3.1.2 Gap length Change

Apart from TRD exchange, there are other ways by which these enzymes can change their specificity. These enzymes can change the gap length between two specific sequences of the bipartite target site by changing the length of the dimerisation domain. This dimerisation determines the separation between two variable regions involved in the recognition of the target sequence. Larger the separation between the two variable regions more will be the gap between the two specific sequences. For example, EcoR124I and EcoR124II both recognise the same specific sequence ((5' end) GAA and (3' end) RTCG) but EcoR124 I recognises these sequences separated by 6 bp non-specific sequence whereas EcoR124 II requires 7 bp separation between the two (Firman et al., 1985; Glover et al., 1983; Price et al., 1989; Price et al., 1987a). Later it was found to be dependent on the repeat sequence of 4 amino acids Thr-Ala-Glu-Leu (T-A-E-L). The increase in the length of dimerisation helix by one helical turn leads to an increase in the separation between the bipartite sequences by 1 bp (Price et al., 1989; Price et al., 1987a; Price et al., 1987b). This is observed in Type IC RM family as a result of unequal crossing over of short sequence repeat in D1.

1.2.3.1.3 Homodimeric S subunits

Earlier reports suggest that change in the order of specificity domains do not implicate biological significance as they are rotationally symmetric. The HsdS subunits consist of two identical regions arranged antiparallel to each other. Each repetitive region consists of a globular part containing TRD interacting with one M subunit and alpha-helical region interacting with a helical region of other identical unit. The helical structures help in maintaining the distance between TRDs and also invert their orientation. The DNA is arranged such that bipartite recognition sequence at 5' end of DNA is recognised by N-terminal region and sequence at 3' end of DNA by C-terminal region of S-subunit. The respective methylation domains present in opposite orientation methylate the

adenine in recognition sequence. Study by Abadjieva et al., showed that an EcoR124I mutant, HsdS with 160 aa deletion in the C-terminus was active and its 3' recognition palindromic to the 5' recognition sequence (GAA N7 TTC) (Abadjieva et al., 1993). This suggested the substitution of C-TRD function by N-TRD, which recognises the palindromic sequence. Even the helices forming coiled coil can function interchangeably. The sequence similarity between dimerisation domains of Type IC HsdS subunits explains the former.

Another mutant of EcoDXXI S subunit with Tn5 insertion resulted in truncated C-TRD. However, it was still able to function due to its homodimer nature and substitution of C-TRD function by N-TRD (Meister et al., 1993). A deliberately generated mutant with a deletion in N-TRD still showed activity by homodimers interchanging their function and C-TRD recognising the palindromic sequence of its specificity (MacWilliams and Bickle, 1996). Further experiments, with co-expression of N- and C-terminal HsdS subunit genes with truncation resulted in wild-type EcoDXXI activity (MacWilliams and Bickle, 1996). This suggested that it is not essential for TRD genes to be aligned and give rise to a single protein chain to function together (MacWilliams and Bickle, 1996). However, they are found to be aligned in characterised and putative type I systems. The fusion has an advantage in systems with multiple TRDs as it maintains stoichiometry to 1:1 and governs their pairing probabilities. Apart from Type IC enzymes EcoR124I, EcoDXXI and NgoAV, other Type I RM enzymes may not possess homodimeric properties due to lack of sequence similarity in dimerization domain. E.g. Type IA (Loenen et al., 2014a).

1.2.3.1.4 Circular permutation of HsdS subunits

Even though the composition of HsdS subunits is similar in all members of Type I families, a considerable number of permutations is seen in its internal units (S1, S2, D1 and D2). The HsdS subunits of Type IA, IB and IC enzymes possess different circular permutations of dimerisation and specificity domains D1, D2 and S1, S2 respectively. The various combinations differ in the position of N- and C-terminal regions. Experimental evidence suggested that the position of S1 and S2 domain does not affect the function of RM enzymes whereas changing the position of the N-terminal or C-terminal domain (S-D) affected the enzyme activity (Janscak and Bickle, 1998). The position of N and C-terminal domains (S1-D1 and S2-D2) determines the proper enzyme folding or function or resistance to protein-degrading enzymes (Janscak and Bickle, 1998).

1.2.3.1.5 Recognition sequence orientations

In Type I RM enzymes, the two TRDs of HsdS subunit are asymmetric, and recognise two different sequences in the bipartite recognition sequence. The N-TRD recognises 5' half sequence and the C-TRD recognises 3' half sequence. This also suggests that the recognition sequence

depends on which strand is considered. E.g. the recognition sequence can be 'AAC N6 GTGC' or its complement 'GCAC N8 GTT for EcoKI (Loenen et al., 2014a). Ideally both the sequences will be recognised by the same enzyme but this is particularly valuable for understanding the specificities of TRDs present at different positions in various Type I subtypes (Thorpe et al., 1997).

Table 1-2: Target sequence for Type I RM enzymes.

RM Enzyme	Type of RM enzyme	Target Sequence
EcoBI	IA	TGA (N8) TGCT
EcoDI	IA	TTA (N7) GTCY
EcoKI	IA	AAG (N6) GTGC
StyLTIII	IA	GAG (N6) RTAYG
StySPI	IA	AAC (N6) GTRC
StySQ	IA	AAC (N6) RTAYG
EcoAI	IB	GAG (N7) GTCA
EcoEI	IB	GAG (N7) ATGC
CfrAI	IB	GCA (N8) GTGG
StySK1	IB	CGAT (N7) GTTA
EcoDXXI	IC	TCA (N7) RTTC
EcoprrI	IC	CCA (N7) RTGC
EcoR124I	IC	GAA (N6) RTCG
EcoR124II	IC	GAA (N7) RTCG
EcoR124IΔ50	IC	GAA (N7) TTC
NgoAV	IC	GCA (N8) TGC
StySBLI	ID	CGA (N6) TACC
KpnBI	IE	CAAA (N6) RTCA

1.2.4 Modification subunit (HsdM)

HsdM subunit is involved in protecting the host DNA from the adverse effect of the restriction subunit. This domain recognises the methylation status of the host DNA and methylates the adenine bases (N-6 position) of the hemimethylated target sequence in the presence of S-adenosylmethionine (AdoMet) as a co-factor (Dryden et al., 1993). The adenine from the top strand in the 5' half and adenine from the bottom strand in 3' half target sequence is methylated by the MTases. Generally, the target site for Type I RM enzymes are hemimethylated in the host genome which acts as the preferred target site for methylation by maintenance methyl transferases EcoKI (Dryden et al., 1993) and EcoR124I (Taylor et al., 1993). It takes hours for these enzymes

to methylate an unmethylated DNA sequence. On the other hand, Type IB EcoAI methylates unmethylated and hemimethylated target sites with the same affinity (Suri and Bickle, 1985). This methylated adenine act as a signature for the host DNA.

The HsdM subunit is highly conserved within a family of Type I RM enzyme whereas it shares 20-30% identity within the members of different families. The conservation is mainly in the motif sequences. It resembles γ-class of Type II N-6 adenine methyltransferases containing six motifs common to the prototype M. TaqI (Labahn et al., 1994). Mutagenesis studies have shown the involvement of motif I (D/E/SXFXGXG) in AdoMet binding and motif IV (N/DPPF/Y/W) in catalysis activity (Dryden et al., 1995). These methyltransferases flips out the adenine bases in the catalytic pocket (as seen in M.TaqI; PDB: 1G38) unlike α- and β- class of methyltransferases which flips the base from the complementary strand (α-MTase T4Dam; PDB: 1YFL) (Goedecke et al., 2001; Horton et al., 2005). The flipped base is stabilised by H-bonding (Blumenthal and Cheng, 2001; Goedecke et al., 2001) and stacking interaction by the aromatic amino acid (Pues et al., 1999) to facilitate the methyl transfer reaction by the catalytic residues.

Apart from the catalytic activity, HsdM is also involved in target recognition along with HsdS subunit. A structural study on M. EcoKI showed that a four-helix bundle formed the constant region of the HsdS subunit and C-terminal α-helices of the two HsdM (Figure1-4) (Liu et al., 2017b). This study further showed that it is not the TRD of HsdS rather it's the four helix bundle that determines the specificity of interaction between the two subunits (Liu et al., 2017b).
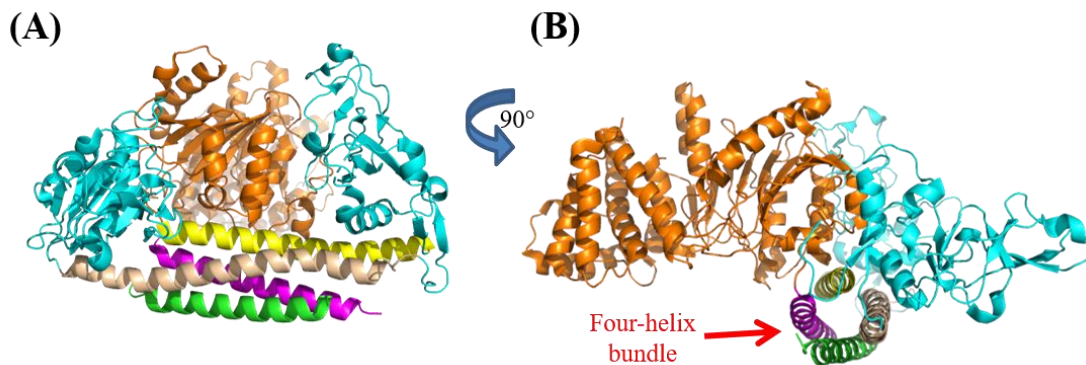
Figure 1-4: Interaction of HsdM with HsdS subunit of Type I RM enzymes. **(A)** The dimerisation domain of HsdS (shown in yellow and tan) interacts with the C-terminal helices of two HsdM subunits (magenta coloured α-helix of one HsdM and green coloured α-helix from the other HsdM subunit). The HsdS is shown in cyan and one HsdM subunit is shown in orange colour. Only α-helix (green) is shown of the second HsdM subunit. **(B)** The four-helix bundle (shown by red arrow) formed from the dimerisation region of the HsdS (yellow and tan) and two α-helices from the HsdM (magenta and green) (PDB: 5YBB) (Liu et al., 2017b).

1.2.5 Restriction endonuclease and translocation subunit (HsdR)

HsdR subunit is responsible for the nucleolytic activity of the Type I RM enzyme complex. These enzymes cleave the dsDNA in between two target sequences upon translocation of few to

thousands of base pairs on DNA. The translocation is performed by the superfamily 2 (SF2) helicase-like ATPase domain of HsdR subunit, which converts the chemical energy obtained from ATP hydrolysis to mechanical event of walking on DNA (Seidel et al., 2008; Seidel et al., 2004; Stanley et al., 2006; Szczelkun et al., 1996). When two translocating enzymes collide, they cleave the dsDNA by the nucleolytic activity of the nuclease domain which belongs to RecB nuclease-family a member of nuclease superfamily I PD-(ExK) (Davies et al., 1999b; Sisakova et al., 2008a).

The structural study on HsdR domain of Type IC RM enzyme EcoR124I has shown that it contains 4 globular domains in square-planar geometry – endonuclease domain (aa residues 13-260), RecA-like helicase domain 1 and 2 (aa residues 261-461 and 470-731) and helical domain (aa residues 732-892) (PDB: 2W00) (Lapkouski et al., 2009) (Figure 1-5 A, B). The amino acid residues 893-1038 was disordered in the structure. The endonuclease domain contains an αβα core of the Type II restriction enzymes (Niv et al., 2007) with the catalytic residues (Asp151, Glu165 and Lys167) clustered opposite to the helical domain (Lapkouski et al., 2009). The sequence comparison of endonuclease region of EcoR124I with other member of Type I RM enzymes has shown a conserved motif QxxxY along with motif I (Glu37), motif II (Asp151) and motif III (Glu165-x-Lys167) of RecB-like nuclease family (Figure 1-5A) (Sisakova et al., 2008b). Mutational studies of these motifs have shown reduced or loss of DNA cleavage upon mutation of motif II and motif III which also affects the ATPase and DNA translocation activity (Sisakova et al., 2008b). The mutation of motif QxxxY does not affect the DNA cleavage activity but reduces the efficiency of DNA cleavage, revealing its involvement in stabilising the nuclease domain (Sisakova et al., 2008a; Sisakova et al., 2008b).

The ATPase domain contains RecA-like fold (Singleton et al., 2007) for ATP binding/hydrolysis and is involved in the movement on DNA. Two RecA-like folds (Domain 1 and 2) forms the active core for ATP binding/hydrolysis and DNA binding (Lapkouski et al., 2009; Singleton et al., 2007). Based on sequence alignment motifs involved in ATP binding and hydrolysis (motif Q, I, II and VI), DNA binding (motif Ia, Ib, Ic, IV, IVa, V and Vb) and coupling of ATP hydrolysis to DNA translocation (motif III and Va) of the ATPase domain has been identified as belonging to SF2 helicase family (Fairman-Williams et al., 2010), although, all the motifs are not present in all members of this family. The levels of sequence conservation in motifs are high within the family and decrease across different family (Fairman-Williams et al., 2010; Singleton et al., 2007). Among these motifs, the ATP binding/hydrolysis motifs (motif I and motif II) and energy coupling arginine finger (motif VI) are highly conserved. Motif I (GxxxxGK(T/S)) is also called as Walker A/phosphate binding loop or P-loop, which is involved in binding with the phosphate group of the

ATP through the amino group of lysine and hydroxyl group of threonine/serine. Motif II (DExx), aspartate is involved in catalysing the ATP hydrolysis via coordination of Mg^{2+} ion. It is also known as D-loop/Walker B. All other less conserved helicase motifs along with these highly conserved motifs are arranged within the RecA domain to form the active core.

(A)

(B)

(C)
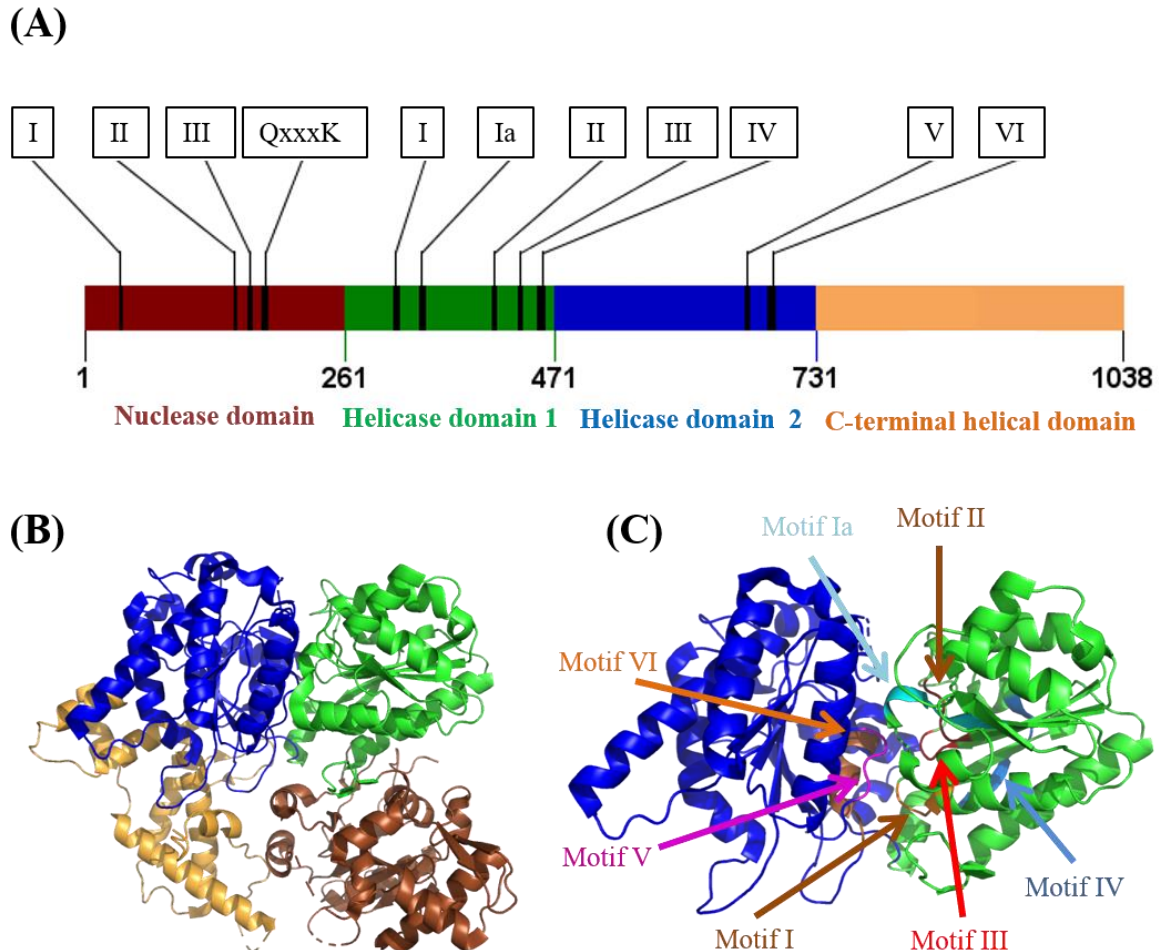
Figure 1-5: HsdR subunit of Type I RM enzyme. (A) The schematic representation of the domain of the HsdR subunit with the conserved nuclease and helicase motifs. Numbers show the boundaries of each domain (Bourniquel and Bickle, 2002). **(B)** Ribbon diagram of HsdR subunit of EcoR124I showing a planar organisation of the endonuclease domain, two RecA-like helicase domains and the C-terminal helical domain. The colour codes are the same as shown in the schematic representation in Figure 1-6A **(C)** The RecA-like helicase domain showing conserved helicase motifs in the active ATPase core (PDB: 2W00) (Lapkouski et al., 2009).

Type I HsdR contains seven conserved motifs of SF2 helicases- motif I (GTGKT), motif Ia (FLVDRR), motif II (DEAH), motif III (TAT), motif IV (FGEPVYRY), motif V (LTTGVD) and motif VI (QMKGRAT) (Figure 1-5 A and 1-5 C) (Bourniquel and Bickle, 2002). An extensive mutagenesis study of these motifs has shown their role in ATP hydrolysis and DNA translocation, which affects DNA cleavage (Davies et al., 1999a; Davies et al., 1999b; Davies et al., 1998). In spite of containing the conserved helicase motifs, these enzymes do not unwind dsDNA; rather they translocate on the dsDNA (Seidel et al., 2005). These translocases utilise entire dsDNA

structure for translocation by forming primary contacts with the sugar-phosphate backbone and bases of the 3'-5' strand (Stanley et al., 2006). The contacts with 5'-3' strand are not important for translocation on dsDNA, but it stabilises the enzyme on the DNA (Stanley et al., 2006). These enzymes along with chromatin remodelers (Durr et al., 2005; Liu et al., 2017a; Thoma et al., 2005; Zhang et al., 2006) represent a new class of dsDNA translocases that does not unwind DNA but contains the common SF2 helicase motifs.

Based on the recent study on Type IC RM enzyme (EcoR124I) HsdR domain, the C-terminal helical region after the ATPase domain can be further divided into two parts – the helical region and the C-terminal domain (Grinkevich et al., 2018). The helical domain is shown to be communicating with the ATPase domain and modulate its activity. Residues Lys527 (ATPase) is H-bonded to helical domain residues Asp796 and Asn794 whereas the Glu730 (ATPase) is H-bonded to Tyr736 (helical domain) through the side chains of the amino acid residues. Mutational studies to abolish these contacts has affected the ATP hydrolysis, DNA translocation and DNA cleavage (Bialevich et al., 2017). The crystal structure of the C-terminal region, when combined with the 4 domain planar structure of HsdR shows a unique α-helical structure protruding outward of the plane (Grinkevich et al., 2018). The surface charge calculation of HsdR with modelled C-terminal region shows that this domain extends DNA binding cleft between the two RecA domains leading to the proposal that it plays a role in correct positioning of the DNA on the active site of the endonuclease (Grinkevich et al., 2018). In agreement with the earlier studies (Davies et al., 1999b; Obarska-Kosinska et al., 2008), Grinkevich et al., propose participation of conserved motif [1016]KKxxxxxK[1023] in binding to the methyltransferase (Grinkevich et al., 2018). The deletion studies of this region have shown its involvement in complex assembly and DNA binding (Grinkevich et al., 2018).

1.2.6 Translocation and DNA cleavage

DNA cleavage by Type I RM enzyme is a multistep process. It involves target sequence recognition on the DNA molecule followed by ATP hydrolysis and finally, the nucleolytic activity of the two converging HsdR subunits. An active restriction complex consists of 1 HsdS, 2 HsdM and 2 HsdR subunits (SM_2R_2) (Figure 1-6A). This complex is unstable and dissociates to form the MTase core with one HsdS and 2 HsdM subunits (Janscak and Bickle, 1998). Some reports have also shown partial dissociation of the complex leading to the formation of MTase core with one HsdR subunit (Janscak and Bickle, 1998; Suri et al., 1984). Both the active restriction complex (SM_2R_2) and partially dissociated complexes (SM_2R) are active DNA translocases (Firman and Szczelkun, 2000). Both the complexes translocate at a similar speed (100-500 bp/s) with a

difference in their processivity. The active HsdR complex translocates 4300 ± 900 bp, suggesting three times more processivity than the partially dissociated complex with one HsdR subunit (which translocates 1320 ± 150 bp) (McClelland et al., 2005; Seidel et al., 2004).
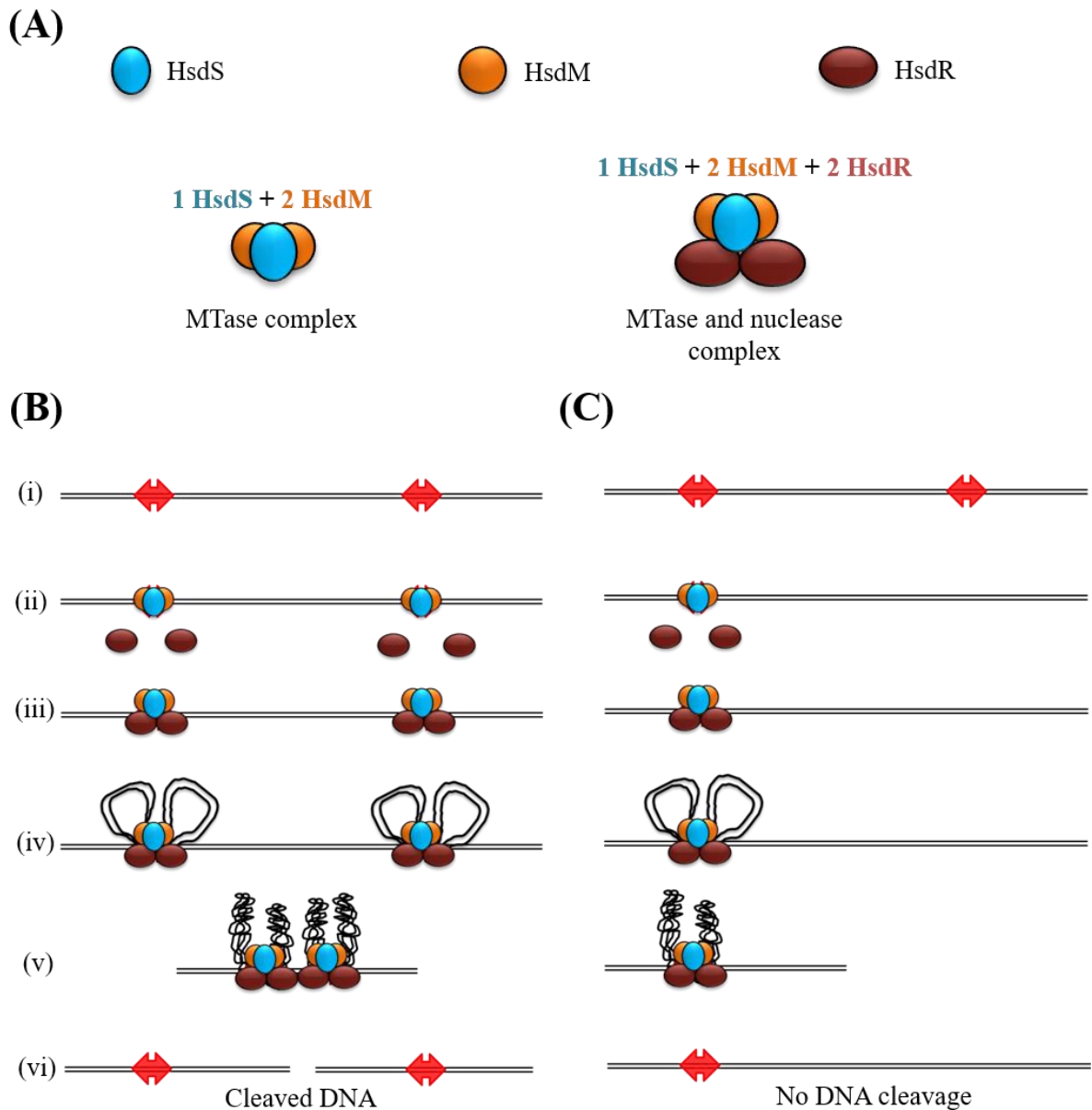
Figure 1-6: Collision model for DNA cleavage by Type I RM enzymes. (A) Subunits forming active MTase and DNA cleavage complex. **(B)** Linear DNA cleavage by Type I RM enzyme containing two target sites. (i) The dsDNA containing two target site (shown in red) separated by thousands of base pairs. (ii) The target site is recognised by the HsdS and HsdM. (iii) After checking the methylation state of the target site, the MTase complex loads the HsdR subunit. (iv) This active complex starts bidirectional DNA translocation by DNA looping upon ATP hydrolysis. (v) The collision of two translocating enzyme brings two nuclease domains together. (vi) The nick by the two nuclease domains on each strand leads to dsDNA break. **(C)** The single site containing DNA is refrained from nucleolytic activity as the two translocating HsdR are unable to converge on the DNA for nuclease domain activation (Studier and Bandyopadhyay 1988).

The translocation speed of EcoR124I SM_2R_2 is 1000 bp/s with two HsdR subunits translocating bi-directionally whereas SM_2R translocates uni-directionally with a speed of 550 bp/s (Seidel et

al., 2004). The SM$_2$R complex can change the direction of translocation after its dissociation from one HsdM subunit of the MTase core and re-binding on the second HsdM subunit (Firman and Szczelkun, 2000; Seidel et al., 2004). It implies that HsdM subunit continuously loads the HsdR subunit from the solution onto the DNA, indicating that the concentration of HsdR in the solution decides the re-initiation of the translocation event.

A collision model for DNA restriction by Type I RM enzymes was proposed by Studier and Bandyopadhyay in 1988 (Studier and Bandyopadhyay, 1988) (Figure 1-6 B and 1-6 C). This model explains DNA cleavage by Type I enzymes at non-specific sites in between two target sequences. The first step for restriction of DNA is the recognition of target sequence (specific site) by the MTase core. If the DNA is hemimethylated (if host DNA), the MTase core will methylate the DNA and disassemble from the target site. If the target site is unmethylated, the MTase core binds to the target sequence (Figure 1-6 B (ii)) and loads the HsdR onto the adjacent DNA (non-specific site) on either side of the MTase core (Seidel et al., 2005) (Figure 1-6 B(iii)). The HsdR domain interacts with the DNA with its ATPase domain while maintaining its contacts with the MTase core through the C-terminal region. The HsdR domain starts translocation on the dsDNA by hydrolysing ATP whereas the MTase core remains bound to the target site (Firman and Szczelkun, 2000; Seidel et al., 2005; Seidel et al., 2004). This translocation leads to the formation of two large DNA loops by two HsdR translocating in opposite direction (Rosamond et al., 1979; van Noort et al., 2004; Yuan et al., 1980) (Figure 1-6 B (iv)). This translocation brings two endonucleases together acting as a roadblock (Janscak et al., 1999), which activates the nucleolytic activity in the presence of Mg^{2+} ions (Figure 1-6 B (v and vi)). dsDNA break happens as a result of two successive DNA nicking activity by each endonuclease (Janscak et al., 1999; Studier and Bandyopadhyay, 1988). Accordingly, a linear substrate with two target sites will be preferred substrate for nucleolytic reactions (Figure 1-6 B) whereas a linear DNA with one site cannot be cleaved by these enzymes (Figure 1-6 C). On a single site substrate, the MTase core will be assembled and will load the HsdR subunits. The HsdR subunit will translocate in opposite direction and ultimately fall off from the DNA ends resulting in no nucleolytic activity (Figure 1-6 C). Contrary to this, a single target site on a circular plasmid DNA will lead to efficient DNA cleavage event (Janscak et al., 1999). In the case of circular DNA, the two opposite translocating HsdR subunit of the enzyme bound to the target site will collide, leading to DNA cleavage (Figure 1-7).
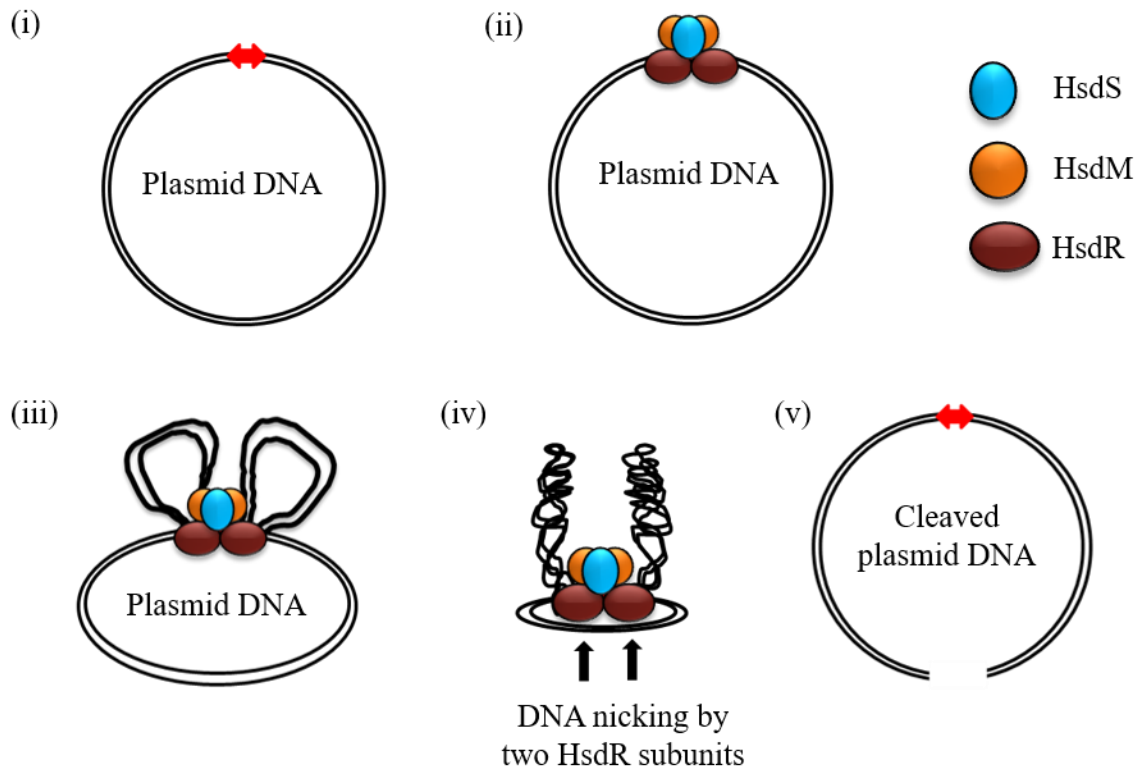
Figure 1-7: Collision model for circular DNA cleavage by Type I RM enzymes. (i) The dsDNA circular containing a single target site (shown in red). (ii) The non-methylated target site (shown in red) is recognised by the HsdS and HsdM which loads the HsdR subunit. (iii) This active complex starts bidirectional DNA translocation by DNA looping upon ATP hydrolysis. (iv) The collision of two translocating HsdR subunits brings two nuclease domains together (similar to the convergence of HsdR subunits upon collision of two translocating enzyme). (v) The nick by the two nuclease domains on each strand leads to dsDNA break.

1.3 Type ISP RM enzymes

Type ISP (single polypeptide) RM enzymes have been discovered recently. These enzymes appear to be similar to Type I RM enzymes in their DNA cleavage activity and ATP usage. However, unlike the hetero-pentameric nature of Type I RM enzymes, Type ISP RM enzymes are single polypeptide protein containing all the functional domains required for their activity (Smith et al., 2009c). These enzymes contain all the four domains (nuclease, ATPase, methylase and target recognition domain) required for restriction of viral DNA and modification of host DNA in a single polypeptide chain (Figure 1-8). LlaGI and LlaBIII belong to Type ISP RM enzymes group. The two enzymes share an overall sequence identity of 80% whereas it shows 98% sequence identity in first 1200 amino acid residues comprising the nuclease, ATPase and MTase domains. LlaGI recognises a 7 bp sequence CTnGAyG (Smith et al., 2009a) and LlaBIII recognises 6 bp sequence TnAGCC (Sisakova et al., 2013) (n is A, T, G or C whereas y is T or C). These enzymes require two target sites in head-to-head orientation to cleave dsDNA. Although the DNA cleavage activity of Type ISP RM enzyme is similar to Type I RM enzyme, still there is a variation between

the two. Type I RM enzymes recognise an asymmetric bipartite sequence whereas these enzymes recognise a short asymmetric and degenerate target sequence. Also the length of Specificity domain of Type I RM enzymes is double the length of TRD of Type ISP RM enzymes. Similarly, the active enzyme contains single MTase domain, ATPase domain and nuclease domain.
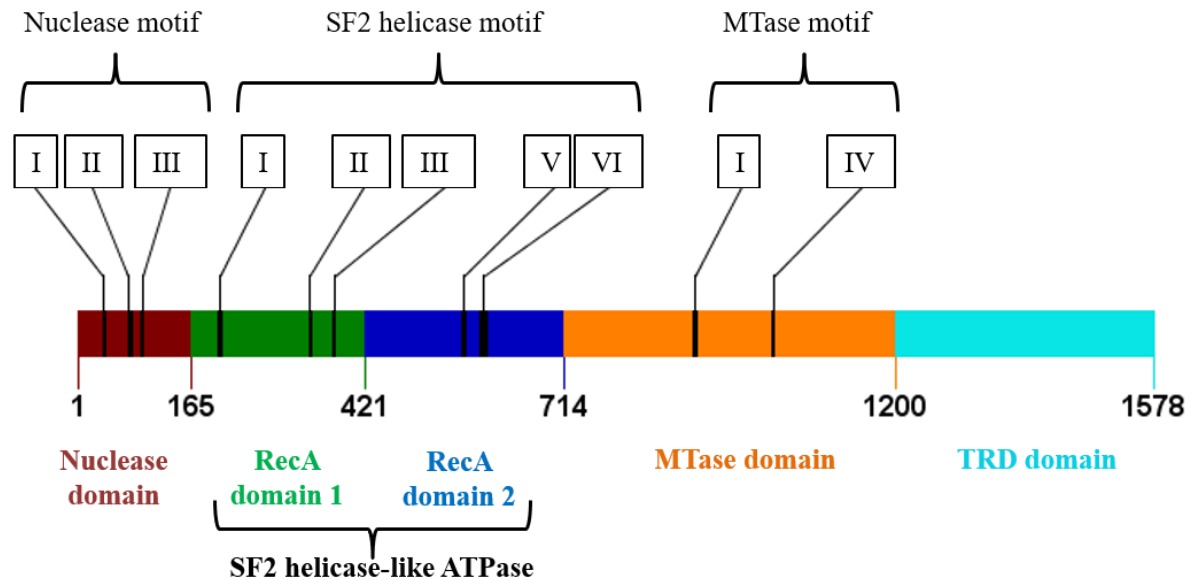
Figure 1-8: Organisation of domains and motifs of Type ISP RM enzymes (LlaBIII and LlaGI). The schematic showing N-terminal nuclease domain followed by ATPase, methyltransferase and C-terminal target recognition domain. Numbers show the boundaries of each domain.

1.3.1 DNA methylation by Type ISP RM enzymes

Type ISP RM enzymes methylate only one adenine in the target sequence leading to hemimethylated DNA (Smith et al., 2009a). Hemimethylated during its replication cycle will provide another hemimethylated and non-methylated DNA, where a non-methylated DNA can be cleaved by these enzymes, making organism incapable of division. But this problem is solved by the requirement of the head-to-head oriented target site for DNA cleavage (Sisakova et al., 2013; Smith et al., 2009a; Smith et al., 2009b, c). This orientation ensures the methylation of both the DNA strands of host genome. DNA cleavage occurs only when both the target sites are non-methylated (head-to-head oriented) if one of the two sites or both the sites are methylated the DNA substrate refrains from cleavage. These DNA substrates with one methylated site lead to methylation of other site by these enzymes in the presence of AdoMet, Mg^{2+} and ATP (Smith et al., 2009a). Like Type I RM enzymes, these enzymes also contain a γ-class adenine methyltransferase. It transfers the methyl group to the N6 position of the adenine base in the target sequence. LlaGI and LlaBIII both methylate bottom strand adenine complementary to first thymine in the target sequence (LlaGI- CTnGAyG and LlaBIII- TnAGCC) (Sisakova et al., 2013; Smith et al., 2009a).

1.3.2 Nuclease domain of Type ISP RM enzymes

Nucleolytic cleavage is involved in many biological processes like holiday junction resolvase, restriction of foreign DNA by the RM enzymes, etc. A number of novel motifs have been identified which are involved in these activities. Among the RM enzymes, Type II restriction enzymes show the maximum diversity. Nuclease domain of Type II restriction is distributed in five superfamily- (i) PD-(D/E)XK superfamily, (ii) HNH superfamily, (iii) Phospholipase D (PLD) superfamily, (iv) GIY-YIG superfamily and (v) Half-pipe (HP) fold. The nuclease domain of majority of RM enzymes belongs to PD-(D/E)XK superfamily. The nuclease domain of Type I and III RM enzymes belong to RecB and AHJR (Archaeal Holiday Junction Resolvase) sub-families of PD-(D/E)XK superfamily (Janscak et al., 2001). Although the cleavage activity of Type ISP is similar to Type I, still there is variation in the nuclease domain. The nuclease of Type ISP contains motif I (glutamate 38), motif II (aspartate 74 and aspartate 78) and motif III (glutamine 92 and lysine 94) which is similar to the motifs of other methylation specific Mrr endonucleases (Figure 1-8), a sub-group of AHJR family (Sisakova et al., 2008a). The mutagenesis study of these residues has shown their importance in DNA cleavage by Type ISP RM enzymes. Among these residues, aspartate 74 (which is 4 amino acid before the catalytic aspartate 78) of motif II is highly conserved throughout the family and absent in other sub-group of superfamily I endonucleases. It also plays a crucial role in nuclease activity and, suggesting that it can form the catalytic triad with D78 and K94 and play an important role in dsDNA break (Smith et al., 2009b). These enzymes form nicks on the 3'-5' strand of the dsDNA with respect to the target site (van Aelst et al., 2013). Translocation is mainly required to bring two nucleases together so that they nick on opposite strands leading to dsDNA break.

1.3.3 ATPase domain of Type ISP RM enzyme

Like Type I RM enzymes, they also contain a SF2 helicase-like ATPase domain with two RecA domains (Sisakova et al., 2013; Smith et al., 2009c) (Figure 1-8). The ATPase domain is involved in converting the chemical energy from ATP hydrolysis to dsDNA translocation event establishing a long-range communication between two target sequences. These enzymes require 1 to 2 ATP molecules to translocate 1 bp on the helical pitch of the DNA molecule (Smith et al., 2009c). Unlike Type I RM enzymes, they translocate unidirectional and require two target sites for DNA cleavage in head-to-head orientation only, in other orientation i.e. head-to-tail, it can only nick dsDNA whereas a tail-to-tail oriented target site on DNA cannot activate nucleolytic activity.

1.3.4 Mechanism of DNA cleavage by Type ISP RM enzymes

For DNA cleavage by Type ISP RM enzymes, a looping mechanism has been hypothesised similar to Type I RM enzymes (Smith et al., 2009c). It states that upon binding to target sequence if these enzymes encounter methylation it will fall off else it will start ATP hydrolysis by ATPase domain.
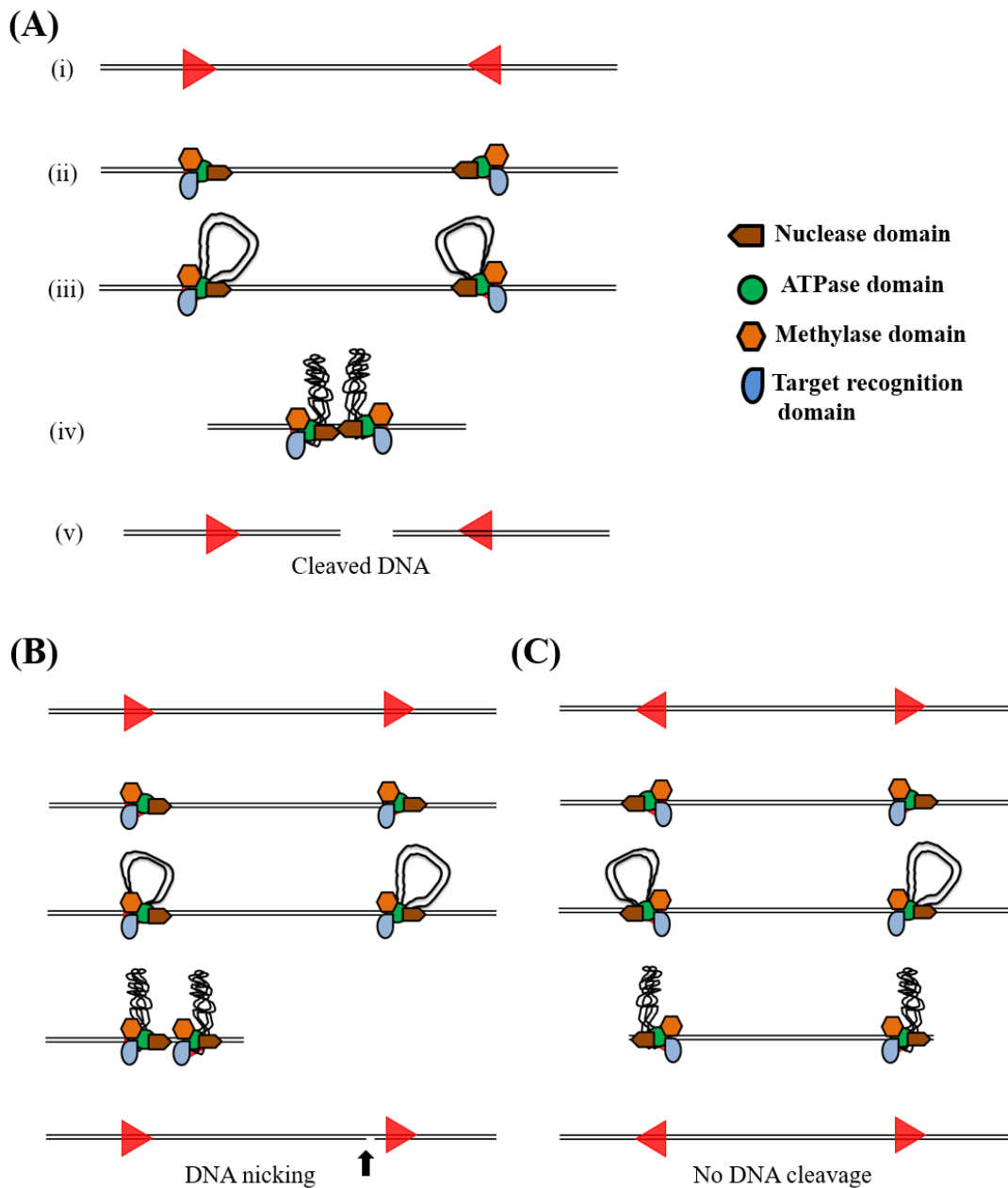
Figure 1-9: Collision model for DNA cleavage by Type ISP RM enzymes. **(A)** Linear DNA cleavage by Type I RM enzyme containing head-to-head oriented target sites. (i) The dsDNA containing two target sites (shown in red) separated by thousands of base pairs. (ii) The non-methylated target site (shown in red) is recognised by the TRD and MTase domain of the enzyme. (iii) This enzyme starts unidirectional DNA translocation by DNA looping upon ATP hydrolysis. (iv) The collision of two translocating domains brings two nuclease domains together. (v) The nick by the two nuclease domains on each strand leads to dsDNA break. **(B)** Type ISP RM enzymes cannot cleave the head-to-tail oriented target site on linear DNA. As a result of the collision between two enzymes, the dsDNA is nicked. **(C)** The tail-to-tail oriented target site containing DNA refrains from nucleolytic activity as the translocating enzyme will fall off from the DNA ends (Smith et al., 2009c).

This ATP hydrolysis will trigger conformational changes which will allow the enzyme to perform restriction activity. Binding of these enzymes to the target site allows the ATPase domain to interact with the non-specific DNA downstream to the target site. Since these enzymes contain only one ATPase domain, they translocate unidirectionally upon ATP hydrolysis, leading to the formation of a single loop by each enzyme (Smith et al., 2009c). When two enzymes translocating from head-to-head oriented target sites collides at a position in between two target sequences, leads to a dsDNA break (Figure 1-9 A). These enzymes cannot cleave a tail-to-tail oriented or a head-to-tail oriented target sequence as they are unidirectional translocases. Translocation on a head-to-tail site substrate will only be nicked in a strand-specific manner upon collision (Figure 1-9 B) whereas the tail-to-tail target site will lead to movement of these enzymes in opposite direction (Figure 1-9 C). However, a single target site on the circular DNA leads to nicking of the DNA (Figure 1-10). This shows that a similar loop translocating mechanism can be used by single and multi-subunit proteins where the ATPase domain can be a part of nucleoprotein complex assembly. The translocating module can be linked by a protein-protein interaction or by a covalent bond but the question remains is how these ATPase modules are coordinated and controlled in such a complex system.
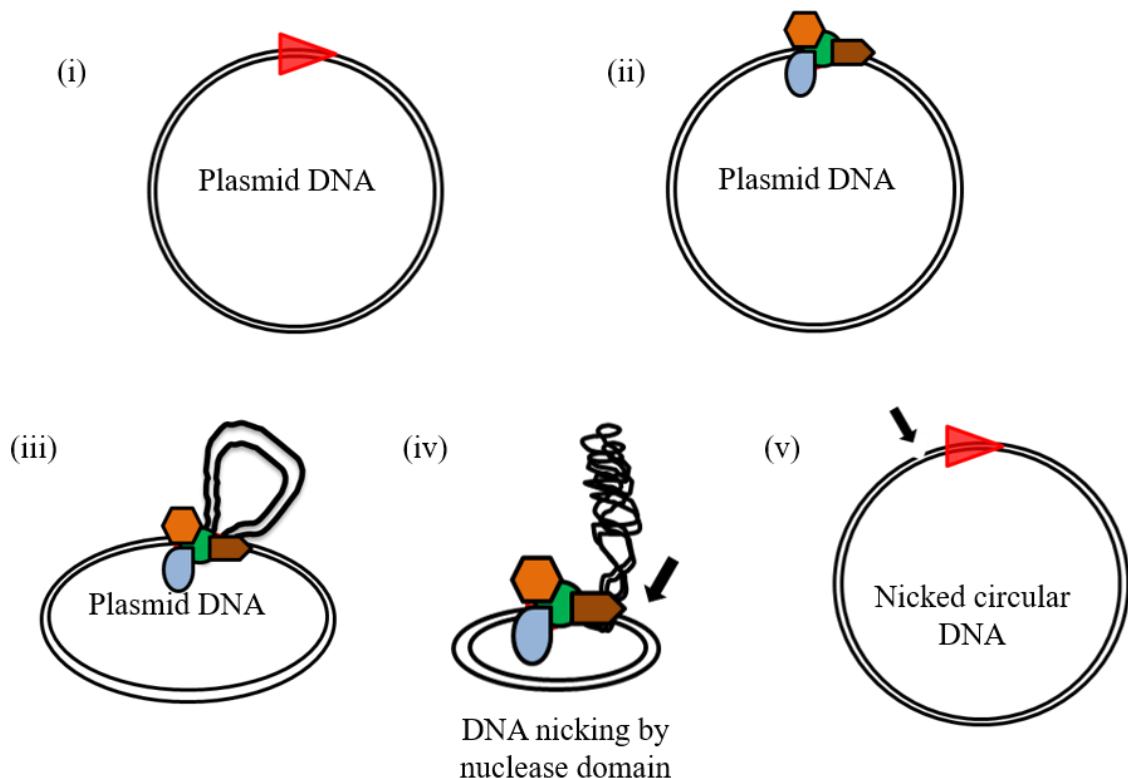
Figure 1-10: Collision model for Circular DNA cleavage by Type ISP RM enzymes. The circular DNA with single target sequence will be nicked by the Type ISP RM enzymes as the translocation of the enzyme will be stopped by the same enzyme sitting on the target site on the circular DNA, leading to activation of the nucleolytic activity by the one enzyme. This will lead to the formation of nicked circular DNA.

1.4 Objectives

Type ISP RM enzymes contains the functional domains in a single polypeptide chain as compared to the heteropentameric Type I RM enzyme, which make them a better system to address the modularity of the complex Type I RM enzymes. The main focus of the work will be to characterise molecular architecture of Type ISP on dsDNA and gain insights into its mechanism of action. It will provide information about spatial organisation of domains and motifs. The binary complex will explain the crucial interactions required for stable and active protein DNA complex formation upon binding to a recognition sequence. It will also provide structural information of the MTase domain, the ATPase domain and nuclease domain. Complementary biochemical experiments will be carried out to complement the information obtained from structural studies. Towards this, I proceeded to address the following objectives-

1. Purification, biochemical characterisation and crystallisation of Type ISP RM enzyme LlaBIII

2. Structural characterisation of LlaBIII and dsDNA translocation

3. DNA-mediated coupling of the ATPase, translocase and nuclease activities of a Type ISP restriction-modification enzyme

Chapter: 2   Purification and biochemical characterisation of Type ISP RM enzyme LlaBIII

2.1 Introduction

Nucleotide triphosphate (NTP)-dependent restriction-modification (RM) enzymes are potent weapons of the bacterial defence system (Dryden et al., 2001). These enzymes prevent the entry of foreign DNA in the host bacterium by introducing double-strand DNA (dsDNA) breaks in the DNA. The same enzyme complex modifies the host DNA in a site-specific manner, which protects the genome from the nucleolytic activity of the nuclease domain (Dryden et al., 2001). This task of opposing activities is carried out by the coordinated action of target recognition, methyltransferase, ATPase and nuclease domains that constitute the enzymes. Classical Type I RM enzymes are hetero-pentamers, comprising two HsdR subunits (nuclease and translocase domains), two HsdM subunits (methyltransferase domains), and one HsdS subunit (target recognition domain or TRD). The target recognition domain recognises a specific target DNA sequence, and the methyltransferase modifies the target adenine base (Dryden et al., 1993). Nucleolytic cleavage of unmodified DNA is triggered by the recognition of two DNA binding sites on the same DNA, ATP-dependent long-range communication and collision of two RM enzymes (Dryden et al., 2001).

The first restriction-modification enzymes to be discovered and purified were Type I RM enzymes more than 45 years ago (Linn and Arber, 1968; Meselson and Yuan, 1968). Since then, these enzymes have served as a paradigm for understanding how complex protein assemblies work (Murray, 2000). Our current understanding of how Type I RM enzymes function is based on extensive genetic, biochemical, biophysical and structural analysis. Several efforts towards structure determination of the enzymes in the last ten years have resulted in the high-resolution crystal structures of individual domains (Calisto et al., 2005; Kim et al., 2005; Lapkouski et al., 2009) and a low-resolution 3D electron microscopy map of the entire EcoKI and EcoR124I enzymes (Kennaway et al., 2012). However, due to their complex nature, the structure of a Type I holo-enzyme at, or near, the atomic resolution has been elusive.

The recently characterised LlaGI and LlaBIII enzymes represent a new class of Type I RM enzyme, called Type ISP RM enzyme. These enzymes contain all the four functional domains- i) N-terminal nuclease domain, (ii) SF2 helicase-like ATPase domain, (iii) Methylase domain and (iv) C-terminal target recognition domain (Figure 2-1) required for its activity in a single polypeptide chain (Sisakova et al., 2013; Smith et al., 2009a). LlaGI and LlaBIII have nearly identical amino acid sequences except for the regions involved in DNA recognition (the TRD) and catalysis of DNA methylation. Accordingly, they bind different sequences: LlaGI recognises 5′-CTnGAyG-3′ while LlaBIII recognises 5′-TnAGCC-3′, with methylation of only one strand taking place (Sisakova et al., 2013; Smith et al., 2009a). DNA cleavage requires two sites in a head-to-

head orientation with cleavage occurring at distant sites between the two sites somewhere close to the midpoint position. Type ISP RM enzymes are monomeric in functionally active form (Smith et al., 2009a), which makes them a good candidate for crystallographic studies of ATP-dependent RM enzymes. As a part of the bigger goal of gaining mechanistic insights into how this multi-domain enzymes function, we report in this chapter our efforts towards biochemical characterisation and crystallisation of LlaBIII bound to its DNA substrate. Structural studies on Type ISP RM enzyme LlaGI were previously initiated in the laboratory. LlaGIΔN (nuclease domain deletion mutant of LlaGI) was purified and co-crystallised with 22 bp DNA harbouring a single nucleotide 5'-overhang (Kulkarni et al., 2016). However, crystallisation attempts for full-length LlaGI yielded poor diffracting crystals. Efforts to improve diffraction quality involved variations in the initial crystallisation condition of this complex such as temperature, the concentration of precipitants, use of additives etc. Despite this, these crystals diffracted to a best of 7.4 Å (Kulkarni et al., 2016). Hence, I started my work on LlaBIII as an alternate strategy towards obtaining high-resolution diffraction data. LlaBIII was purified to its homogeneity and crystallised with 28 bp DNA. Also, the DNA binding, ATPase and DNA translocase activities were characterised.
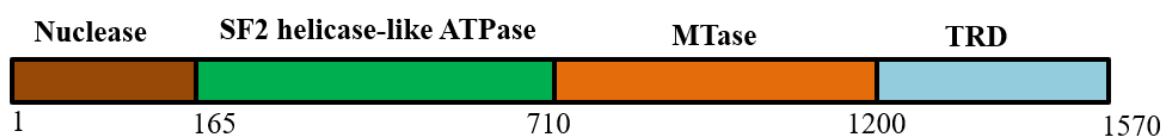
Figure 2-1: Schematic diagram depicting the domain arrangement of LlaBIII. The boundaries of each domain are labelled with the amino acid residue numbers.

2.2 Materials and methods

2.2.1 Cloning of LlaBIIIΔN

LlaBIIIΔN (1-165 amino acid residues deletion) was ligated into the pHIS17 vector under NdeI and BamHI sites. The nuclease domain of the *llabiii* gene was deleted by polymerase chain reaction (PCR) using forward primer LBdN-pHIS-F and reverse primer LB-pHIS-R and *llabiii* gene as the template. The amplified gene of ~ 4.3 kb containing ATPase, MTase and target recognition domain was double digested with NdeI and BamHI restriction enzymes and PCR purified to remove the restriction enzymes and short cleaved product.

Similarly, the pHIS17 vector was also digested with the respective enzymes. The digested gene product was ligated with linear pHIS vector harbouring complementary overhangs using NEB T4 DNA ligase. The linear vectors were treated with thermosensitive alkaline phosphatase (TSAP) to remove 5' phosphate before ligation reaction. The ligated product was transformed in *E. coli* NEB

turbo electrocompetent cells by electroporation. After transformation, the positive clone was screened on LB agar plates containing ampicillin as the selection marker. Colonies were inoculated in LB broth with (100 µg/ml) ampicillin for 10-12 hours. Cells were pelleted down, and plasmid DNA was isolated using the QIAGEN miniprep kit. Positive clones were checked by double digestion, and the nucleotide sequence of the mutant was confirmed by DNA sequencing.

2.2.2 Purification of LlaBIII and LlaBIIIΔN

LlaBIII was overexpressed in 6 L of *Escherichia coli* BL21 (DE3) from the recombinant clone in pET28a vector. A higher amount of soluble protein in the crude lysate was observed upon inducing the culture at 25°C with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at $OD_{600} = 0.6$. The incubation temperature was lowered from 37°C to 25°C before the addition of IPTG. Induced cells were harvested after 5 hours of incubation. To minimise proteolytic degradation of the protein, protease inhibitors (Roche) were added to the lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA and 1 mM DTT), all the steps of purification were carried out at 4°C, and the protein was purified to homogeneity within 24 hours of lysis. The cells were lysed by sonication (1 second on, 3 seconds off for 3 minutes at 60% amplitude) at 4°C. Two rounds of sonication were done to completely lyse the cells with 10 minutes gap between the two rounds. The lysate was centrifuged in the Beckman Coulter floor model centrifuge at 37000 RPM for 45 minutes at 4°C in the Ti45 rotor. The supernatant obtained after ultracentrifugation was fractionated by ammonium sulphate. 40-70% w/v ammonium sulphate pellet was re-suspended in buffer B0 (50 mM Tris-HCl pH 8.0, 1 mM EDTA and 1 mM DTT). The re-suspended protein was loaded on HiPrep Heparin FF 16/10 column (GE Healthcare) and eluted with 0-100% linear gradient of buffer B1000 (50 mM Tris-HCl pH 8.0, 1000 mM NaCl, 1mM EDTA and 1 mM DTT). The eluted fractions were analysed on 10% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS PAGE) (acrylamide:bisacrylamide=29:1). Fractions containing LlaBIII were dialysed against buffer B50 (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 1mM EDTA and 1 mM DTT) for 2 hours and further loaded on MonoQ 10/100 GL column (GE Healthcare) pre-equilibrated with B50. Elution was done with 0-50% linear gradient of B1000. Finally, size exclusion chromatography using Superdex 200 10/300 (GE Healthcare) ensured homogenous monomeric protein. Purified LlaBIII was stored in 50 mM Tris-HCl pH 8.0, 100 mM KCl and 1 mM DTT. Purified protein stored at –80°C remained intact and was used for experiments later. LlaBIIIΔN was expressed in 6 L of *Escherichia coli* BL21 (AI) with 0.2 % arabinose induction from the recombinant clone in pHIS17 vector and purified in the same way as LlaBIII wild type.

2.2.3 Purification of DNA substrate for crystallisation and DNA cleavage assay

The dsDNA substrates used for co-crystallisation were obtained by annealing two chemically synthesised, PAGE-purified, complementary single-strand DNA purchased from Integrated DNA Technologies, USA. Both the strands were annealed using master cycler Eppendorf vapo.protect. Initially, it was heated to 99°C for 1 minute and then cooled to 25°C with a rate of 1°C/min and stored at 4°C. The duplex DNA was purified to remove single-strand DNA using an anion exchange column MonoQ 10/100 GL (GE Healthcare). The annealed DNA was loaded on MonoQ 10/100 GL (GE Healthcare) pre-equilibrated with buffer B50. Elution was done with 0-100% linear gradient of buffer B1000. The MonoQ 10/100 GL (GE Healthcare) purified dsDNA was concentrated using 3 kDa molecular weight cutoff Vivaspin 2 (GE Healthcare) centricons. The purified fractions were added to the Vivaspin 2 centricons and the solution was centrifuged at 3000 RCF (relative centrifugal force) to concentrate the DNA. The dsDNA was washed with MilliQ to remove the buffer containing EDTA in it. The concentrated duplex DNA was stored at -20°C.

1085 bp long DNA substrate was produced by polymerase chain reaction using the LlaBIII gene as a template and forward primer LB-1085-F and reverse primer LB-1085-R. List of oligomers used as primer and substrates are listed in Table 2.1.

2.2.4 DNA cleavage assay

A 1085 bp long DNA with two head-to-head oriented target sites separated by 155 bp was used as the substrate for nucleolytic cleavage assay. Cleavage assay was done in TMDK buffer (50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 1 mM DTT and 150 mM KCl) (van Aelst et al., 2013). Varying concentrations of LlaBIII was incubated with 10 nM DNA in TMDK buffer and the reaction was started by addition of 4 mM ATP. The reaction mix was incubated for 30 minutes at 25°C. Reactions were stopped by addition of half the volume of STEB (100 mM Tris-HCl pH 7.5, 200 mM EDTA, 40% (w/v) sucrose, 0.4 mg/ml bromophenol blue) containing 1% SDS and analysed on 1% agarose gel.

Table 2-1: DNA oligomers used for the study.

Primer	Sequence
LBdN-pHIS-F	CTTTAAGAAGGAGATATACATATGCGTCCAGAAAATGTTGTTG
LB-pHIS-R	CGCAGCAGCGGTTTCTTTACCAGACTCGAGTTATAGTCCCTGTACTA CTCTTG
LB-1085-F	GCAACGGACGCTCGCTGATCCAG
LB-1085-R	CCATCGCTTGGGAGACGGGGTTTTG
Oligo 23	TTAGCTAATAGACTGAGCCGAGG
Oligo 23_C	TCCTCGGCTCAGTCTATTAGCTA
Oligo 28	GCTCTAGCTAATAGACTGAGCCGAGGTG
Oligo 28_C	CACCTCGGCTCAGTCTATTAGCTAGAGC
Oligo 46	GTCTTATGCAGGTCACCTGCTCTAGCTAATAGACTGAGCCGAGGTG
Oligo 46_C	CACCTCGGCTCAGTCTATTAGCTAGAGCAGGTGACCTGCATAAGAC
Oligo 40_NSP	GTACTCAGCAGTATCCTGTATGCTACGTATTGCTATCGTG
Oligo 40_C_NSP	CACGATAGCAATACGTAGCATACAGGATACTGCTGAGTAC
Cy5_Oligo 50	Cy5(G)GAGTCAGCTCTAGCTAATAGACTGAGCCGAGGCGTACGACG TGCTGCGG
Cy5_Oligo 50_C	CCGCAGCACGTCGTACGCCTCGGCTCAGTCTATTAGCTAGAGCTGA CTCC
TFO	TTCTTTTCTTTCTTCTTTCTTT

2.2.5 DNA binding assay

Binding assays were performed with 50 bp long oligomer Cy5 labelled dsDNA in binding buffer (50mM Tris-HCl pH 7.4, 10 mM MgCl$_2$, 1 mM DTT, 50 mM NaCl, 0.1 mg/ml BSA and 10% glycerol) to characterise the binding affinity of LlaBIII to DNA. 10 nM Cy5 labelled DNA was allowed to bind with varying concentration of LlaBIII (0, 2.5, 5, 10, 25, 50, 100, 250, 500 nM) on ice for 10 minutes. The reaction was stopped with half the volume of stop buffer STB (40% w/v sucrose and 0.1 M Tris-HCl pH 8.0) and loaded on 6% native PAGE. The gel was imaged on Typhoon TRIO+ variable mode imager (GE Healthcare).

Binding assays for optimisation of DNA length for crystallisation were performed with 23 bp, 28 bp and 46 bp dsDNA in binding buffer TMDK. The 23 bp and 28 bp dsDNA was selected based on the crystallization trials on LlaGI. LlaGIΔN was crystallized with 23 bp DNA and LlaGI full length was crystallized with 28 bp DNA. Although the crystal of LlaGI with 28 bp DNA diffracted poorly still we considered it in our crystallization trails with LlaBIII. 46 bp DNA harbor 34 bp upstream DNA length to engage the ATPase and nuclease domains. 250 nM DNA was allowed to

bind with varying concentrations of LlaBIII (0, 10, 50, 100, 250, 500, 1000, 2500, 5000 nM). Protein and DNA were mixed and incubated on ice for 10 minutes. Reaction was stopped with half the volume of STB (0.1 M Tris-HCl pH 7.5, 40% w/v sucrose, 0.4 mg/ml bromophenol blue) and loaded on 6% native PAGE. The gel was stained in ethidium bromide (EtBr) solution for 5 minutes and imaged on Invitrogen lifetech imager.

2.2.6 ATPase assay

NADH coupled ATPase assay was used to estimate the ATP hydrolysis rate of LlaBIII in the presence of specific and non-specific DNA. A 46 bp long DNA harbouring one LlaBIII target site was used as a specific substrate and a 40 bp long DNA with no target sequence was used as non-specific DNA. 50 nM LlaBIII was incubated with 500 nM of specific or non-specific DNA for 5 minutes. This enzyme-DNA complex was added to ATPase reaction master mix (600 μM NADH, 1 mM phosphoenolpyruvate, 10 mM MgCl$_2$, pyruvate kinase/lactate dehydrogenase mix). The reaction volume was made up to 180 μl with TDK (Tris-HCl pH 8.0, 1 mM DTT and KCl). 20 μl 10 mM ATP was added, and the reaction was shaken for 30 seconds to remove air bubbles before the measurements were taken. Absorbance at 340 nM was recorded every 10 seconds for 800 seconds at 25°C. Fluorescence intensities were measured using Varioskan Flash (Thermo Scientific). Varying concentrations of ADP (0, 100, 200, 300, 400, 500 and 600 μM) were used to obtain a standard plot with each set of experiment. The concentration of ATP hydrolysed was calculated at each time interval using a straight line equation $Y = mX + C$ (where Y = absorbance, m = slope, X = concentration of NADH or ADP produced and C = intercept on Y-axis) obtained from the standard plot with different ADP concentrations. The graph was plotted using GraphPad PRISM 5.

Michaelis-Menten plot of rate of ATP hydrolysis versus concentration of ATP was used for calculation of K_m, k_{cat} and V_{max}. Time dependent ATPase assay was performed with 50 nM LlaBIII and 500 nM 46 bp specific DNA at different ATP concentrations (0, 0.25, 1, 4, 16, 64, 128, 256, 512, 1024 and 2048 μM) to calculate the rate of ATP hydrolysis. Line equation fit of the data at different ATP concentration provided the rate for different ATP concentrations.

2.2.7 Triplex DNA and Triplex Displacement

22 nucleotides long pyrimidine-rich single-stranded triplex-forming oligonucleotide (TFO) was radiolabelled with polynucelotidyl kinase (PNK) and ATPγ^{32}P at 37°C for 30 minutes. Labelled TFO was purified using mini quick spin oligo (Roche) column pre-equilibrated with MM Buffer (10 mM MES pH 5.5 and 25 mM MgCl$_2$) and kept at -20 °C till further use. 50 nM pONE plasmid

containing a single site for LlaBIII (Sisakova et al., 2013) was linearised with HindIII. 25 nM TFO was mixed with the linearised pONE plasmid in MM Buffer and incubated at 57°C for 15 minutes. The reaction mix was further incubated at 20°C overnight to allow triplex DNA formation (Firman and Szczelkun, 2000). Triplex DNA was stored at -20°C until further use.

50 nM enzyme and was incubated with 0.25 nM triplex DNA in TMDK buffer for 5 minutes to allow binding to the target site. The reaction was initiated by addition of ATP and incubated for 10 minutes at 20°C. The reaction mixture also contains unlabelled TFO to prevent non-specific sticking of displaced TFO to LlaBIII. In the case of time-dependent triplex displacement assay, the incubation time of reaction mixes were varied. Half the volume of GSMB (15% w/v glucose, 3% w/v SDS, 250 mM MOPS pH 5.5 and 0.4 mg/ml bromophenol blue) was added to stop the reaction and analysed on 6% Native PAGE prepared in 40 mM Tris-acetate, 5 mM sodium acetate, 5 mM $MgCl_2$, 0.1 mM EDTA pH 5.5 at 4°C. The gel was dried in Gel dryer model 583 attached with HydroTech^{TM} Vaccum Pump supplied by BioRad at 80°C for 1 ½ hour. The dried gels were exposed to phosphorimager plate (GE Healthcare) overnight. Gels were imaged on Typhoon TRIO+ variable mode imager (GE Healthcare), and the intensities of the bands were quantified using ImageJ software. All the graphs were plotted using GraphPad PRISM 5. Band intensities were considered 100% for reactions wherein the enzyme was incubated with triplex DNA and the displacement was calculated for lanes where ATP was added. In the time-dependent experiment, the intensities for zero time point was considered as 100% and the intensities for all time points were calculated.

2.2.8 Crystallisation and data collection

A complex of the purified LlaBIII and DNA substrate was formed by mixing the two in 1:1.3 molar ratios at 4°C. A protein concentration of 5 mg/ml and a crystallisation drop size of 200 nl were used for all the initial crystallisation trials by sitting drop vapour diffusion method. The nanodrops were set using a robotic liquid handler. More than 750 different conditions were tested for crystallising LlaBIII with 28 bp DNA. One of the nanodrop conditions yielded a single crystal by vapour diffusion at 291 K using the sitting-drop method. Improvement in crystal size was achieved upon equilibrating 4 µl drops by the hanging drop method. The reservoir buffer contained 200 mM KCl and 20% w/v PEG 3350. This crystal diffracted to 2.7 Å at 100 K with ethylene glycol as a cryo-protectant. Data were collected at the I02 beamline, Diamond Light Source, at a wavelength of 0.9793 Å.

2.3 Results

2.3.1 LlaBIII and LlaBIIIΔN purification

LlaBIII expression trials were done at different temperatures (37°C, 25°C and 18°C) for 3, 5 and 16 hours respectively using *E. coli* BL21 (DE3) (Figure 2-2A and 2-2B). The concentration of IPTG used for induction as well as the cell density of induction was optimised for better expression. Best expression was achieved upon induction at OD600 nm = 0.6 with 0.5 mM IPTG for 5 hours at 25°C (Figure 2-2B).
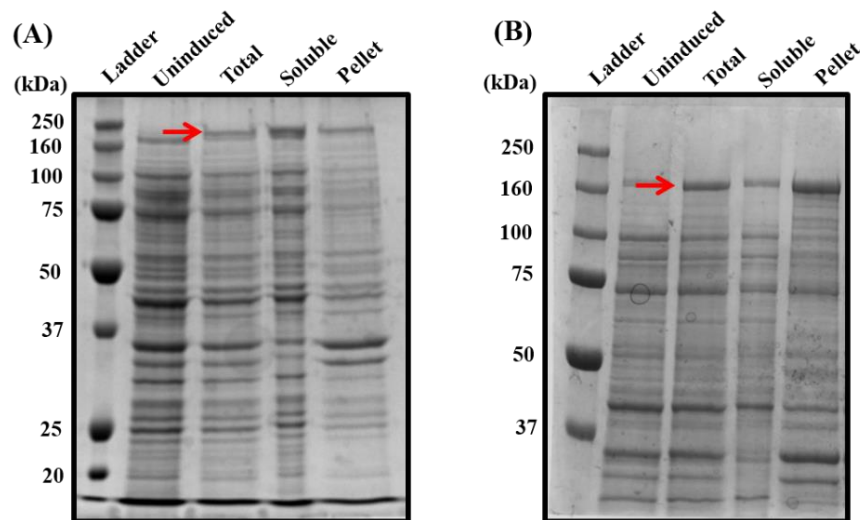
Figure 2-2: LlaBIII expression check in *E. coli* BL21 (DE3) cells at (A) 18°C and (B) 25°C respectively. Uninduced lane has all the proteins expressed in the uninduced cells whereas Total, Soluble and Pellet stands for protein in the induced cells. The Total lane has all the proteins expressed in the cell, Soluble has all proteins in the soluble fraction (supernatant), and Pellet shows all the insoluble proteins. The red arrow in the gel marks LlaBIII band on the gel.

To study the mechanism of action by this multi-domain enzyme, we purified LlaBIII in its native form. For this, the designed strategies included ammonium sulfate fractionation and a combination of 3 different columns, i.e. affinity, ion exchange and size exclusion columns (Figure 2-3A). The first column used was HiPrep Heparin FF 16/10 (GE Healthcare) which has an affinity for DNA binding proteins. Elution was carried out using a linear gradient of 50% to 100% of buffer B1000 (Figure 2-3B). Relatively pure fractions were pooled and loaded on anion exchange column MonoQ 10/100 GL (GE Healthcare) to further remove impurities (Figure 2-3C). The strongly ionic MonoQ column removed not only protein impurities but also any cellular DNA fragments co-purified with the protein. Finally, size exclusion chromatography (SEC) using Superdex 200 yielded homogeneous and monomeric LlaBIII (Figure 2-3D and 2-3E). Highly pure fractions were pooled, concentrated and stored at -80°C in 5 µl aliquots. LlaBIIIΔN was purified using the same strategy.

The initially designed strategy for purification was able to remove contaminating protein at each step of purification, but we used to see degraded protein bands on SDS PAGE. This degradation was time dependent. Hence, we further optimised the purification protocol to reduce the total time required for purification, which resulted in pure and intact protein.
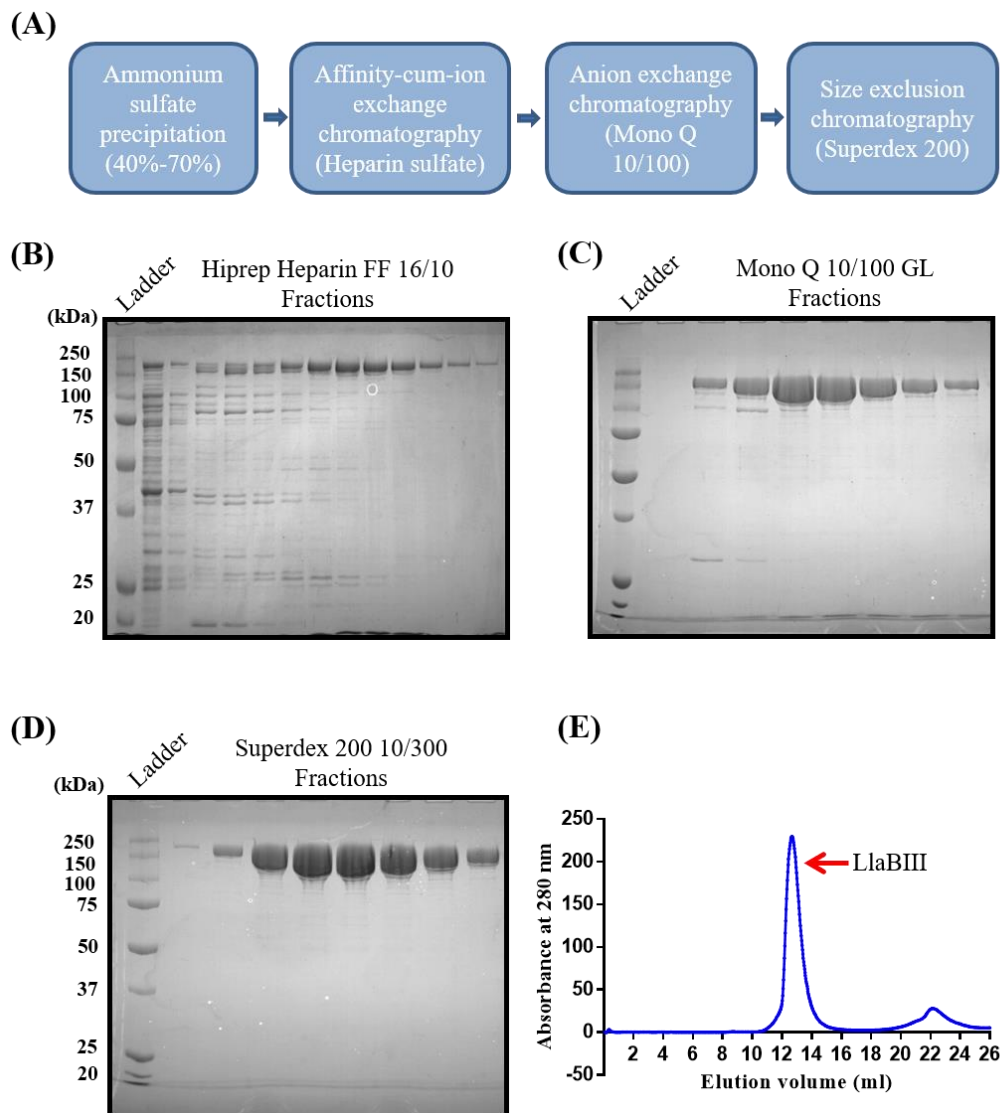
Figure 2-3: LlaBIII purification. **(A)** Flow chart depicting the purification strategy used for LlaBIII purification. **(B-D)** SDS PAGE gels showing the purity of the protein after each column. **(E)** The Abs_{280nm} chromatogram showing a monomeric peak after elution from Superdex200 10/300 column.

2.3.2 DNA cleavage assay

DNA cleavage assay was performed to check the activity of LlaBIII. LlaBIII concentration-dependent DNA cleavage was observed on the agarose gel with 25 nM enzyme concentrations, however efficient DNA cleavage was observed after 200 nM of enzyme (Figure 2-4). The DNA cleavage product was a smear with centring around 400 bp and 700 bp, unlike regular Type II restriction enzyme cleavage where sharp bands are observed on the gel. For these enzymes to cleave the DNA, there should be an effective collision event, which brings two nucleases together

resulting in nicks on both the DNA strands at the collision site. This collision is a stochastic event with the maximum probability for cleavage at the midpoint between the two head-to-head oriented sites. This leads to a smear of cleaved DNA centred around 400 bp and 700 bp. Also, this smearing was not because of any contaminating nuclease as in the absence of ATP no smears were seen at the highest enzyme concentration used for DNA cleavage assay. This enzyme did not cleave a single-site DNA or DNA with no site in the presence of ATP.
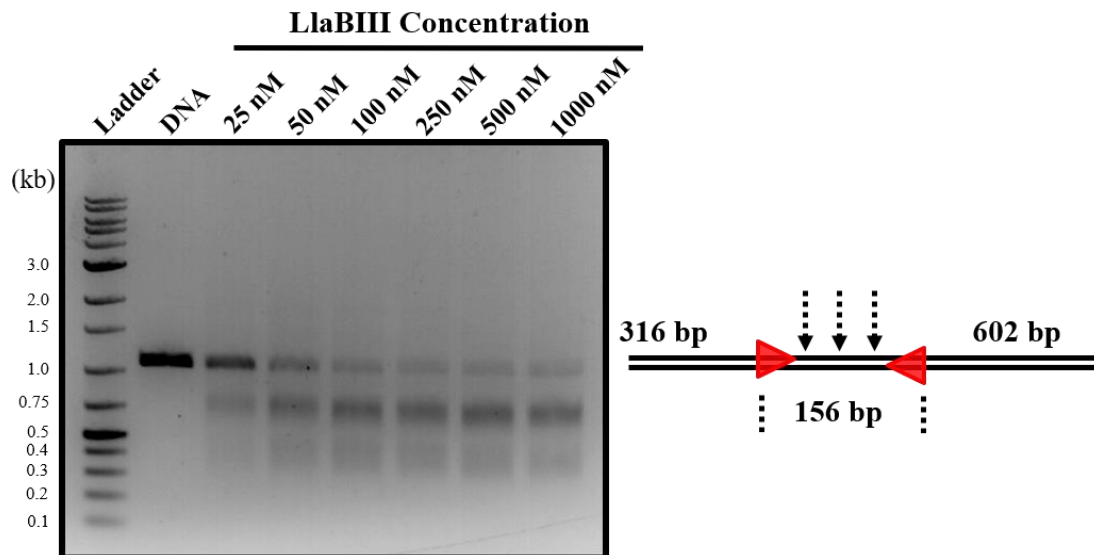
Figure 2-4: DNA cleavage assay. 10 nM DNA was incubated with a varying concentration of LlaBIII in TMDK buffer with 4 mM ATP at 25°C for 30 minutes. The right panel depicts the substrate used for the DNA cleavage assay. The red arrow shows the head-to-head oriented target site on the substrate DNA. Black arrows indicate that the cleavage occurs between the two target sites.

2.3.3 ATPase assay

The ATP hydrolysis rate of LlaBIII in the presence of specific DNA and non-specific DNA was calculated using NADH coupled ATPase assay. The ATP concentration is maintained constant in the coupled assay throughout the reaction time. Each ATP molecule hydrolysed by LlaBIII produces ADP and inorganic phosphate. Pyruvate kinase regenerates ATP from ADP during conversion of phosphoenolpyruvate to pyruvate. This pyruvate is further converted into lactate by lactate dehydrogenase. During the conversion of pyruvate to lactate, lactate dehydrogenase utilises one NADH molecule. For each ATP hydrolysed by the LlaBIII in the reaction, one NADH molecule is consumed, and ATP is regenerated by the pyruvate kinase and lactate dehydrogenase hence maintaining the initial ATP concentration in the reaction (Figure 2-5A). The important part of a coupled enzyme system is that the coupling enzymes should be faster than the enzyme for which the activity is being measured. To check this, we doubled the concentration of the coupling enzymes; still, we did not see any change in the apparent rate of ATP hydrolysis by LlaBIII. The

(A)

LlaBIII
ATPase

ATP $\longrightarrow$ ADP + P$_i$ (Inorganic Phosphate)

ADP + P$_i$

Phosphoenol
pyruvate

Pyruvate
kinase

ATP

Pyruvate

Lactate

Lactate
dehydrogenase

NADH$^+$ NAD$^+$ + H$^+$

Absorbance at 340 nm

(B)

5' TGAGCC 3'
3' ACTCGG 5'

34 bp   6 bp   6 bp

40 bp

(C)

(D)

Enzymes	LlaBIII
E$_t$	0.05 μM
k$_{cat}$	12.2 ± 0.5 s^{-1}
K$_m$	23.4 ± 5.6 μM
V$_{max}$	0.6 μM/s

(E)

— LlaBIII SP_DNA
— LlaBIII NSP_DNA
— LlaBIII No DNA
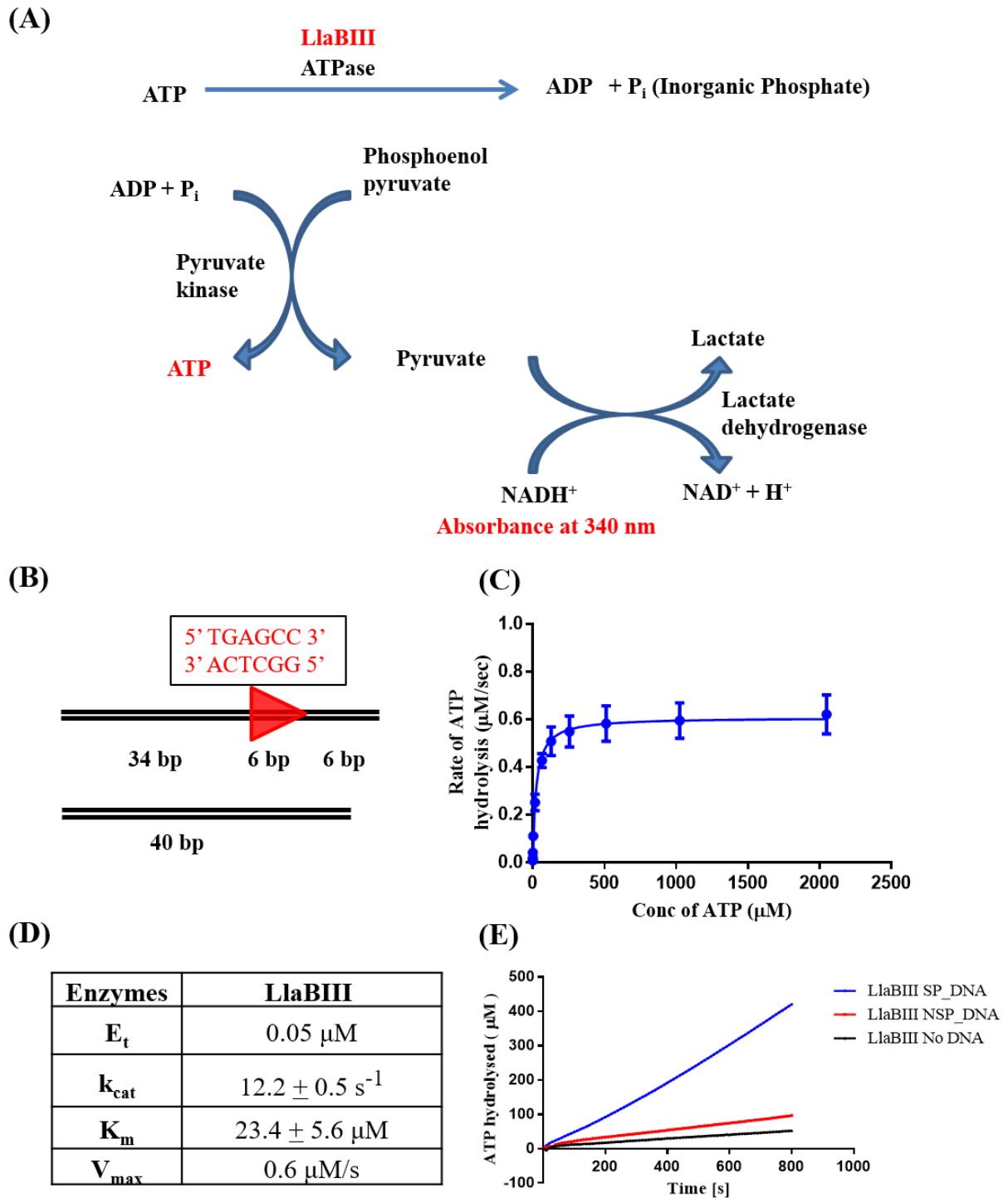
Figure 2-5: NADH coupled ATPase assay of LlaBIII. **(A)** Schematic of NADH-coupled ATPase assay depicting regeneration of ATP by using NADH. **(B)** Substrates used for ATPase assay. The dsDNA with red arrow is 46 bp specific DNA for LlaBIII whereas the lower DNA is 40 bp with no target site **(C)** ATP concentration (0.05, 0.25, 1, 4, 16, 64, 128, 256, 512, 1024 and 2048 μM) dependent ATPase assay was performed to determine the Michaelis-Menten plot for LlaBIII (blue). 50 nM LlaBIII with 500 nM DNA concentration was used in NADH coupled ATPase assay. **(D)** The table showing kinetic parameters obtained from the graph in figure 2-5C. **(E)** ATPase assay in the presence of specific, non-specific and no DNA. 1 mM ATP was used in the reaction. Reactions were carried out at 25°C. The experiments were performed in triplicates.

observed rate was proportional to the concentration of LlaBIII. Based on LlaBIII concentration-dependent ATPase assay we used 50 nM LlaBIII and 500 nM DNA concentrations. 10-fold excess DNA was used to reduce the possibility of free enzyme in the reaction.

We performed an ATP concentration-dependent ATPase assay to obtain the enzyme kinetics parameter for ATP hydrolysis by LlaBIII in the presence of 46 bp specific DNA substrate (Figure 2-5B). Initially, the rate of ATP hydrolysis increased with the increase in ATP concentration and reached the maximum rate at 512 μM (Figure 2-5C). We obtained Michaelis-Menten plot with K_m = 23.4 $\pm$ 5.6 μM, V_{max} = 0.6 μM/s and k_{cat} = 12.2 $\pm$ 0.5 s^{-1}(Figure 2-5C and 2-5D). We also performed an ATPase assay in the absence of DNA and the presence of specific and non-specific DNA (Figure 2-5B). LlaBIII showed stimulation of ATPase activity in the presence of specific DNA (Figure 2-5E). We observed a 5-fold increase in the ATPase activity of LlaBIII in the presence of specific DNA as compared to non-specific DNA. The ATPase rate for non-specific and no DNA was similar suggesting the ATPase activity of LlaBIII in the presence of non-specific DNA is the basal activity.

2.3.4 DNA translocation assay

To investigate the translocation of LlaBIII on DNA, we used triplex displacement assay. A triplex DNA is formed when a single-stranded DNA anneals to a dsDNA. The triplex DNA is mainly formed between a purine-rich dsDNA and a pyrimidine-rich third stand through Hoogsteen base pairing. When a translocating enzyme encounters the third strand in its path, it displaces the Hoogsteen base paired third strand (TFO) and this displacement signals that enzyme is translocating on the substrate DNA (Figure 2-6A and 2-6B). A linear DNA substrate shown in Figure 2-6C was used to form the triplex DNA. In our experiment, we annealed a radiolabeled TFO to its binding site 1581 bp downstream of the LlaBIII target site on the triplex DNA (Figure 2-6). Triplex DNA was formed at lower pH (pH 5.5) as the protonation of the N3 position of cytosine takes place at pH less than 6.0, which is crucial for triplex DNA formation. The reaction for triplex DNA formation was kept at 20°C for overnight as the on-rate for triplex formation was very slow. The triplex DNA was also stable during freeze and thaw cycle on ice. Once the triplex DNA was formed, it was stable at pH 8.0. We also checked the optimal temperature for the triplex displacement reaction and performed all our reaction at 20°C.

After forming the triplex DNA, the stability of the triplex DNA was checked on a 6% native PAGE in TMDK buffer. Initially, we tried to separate the triplex DNA from the displaced TFO on 1% agarose gel. The band intensity for the triplex DNA was visible but the TFO was diffused hence we used native PAGE for proper visualization of TFO and triplex DNA. The triplex DNA

incubated at 20°C was stable and migrated as a slow-moving band on the gel. TFO did not dissociate from triplex DNA even after 10 minutes incubation at 20°C in TMDK buffer. As a control to check the formation and displacement of TFO, the triplex DNA was heated at 80°C for 10 minutes in TMDK buffer and cooled to room temperature, which dislodged the TFO from the triplex DNA. Analysis of this sample on a PAGE gel showed the migration of the free radiolabeled TFO (Figure 2-7A). This also implied that under the experimental conditions the displaced TFO did not anneal back. In the presence of LlaBIII and ATP, TFO from ~80% of the triplex DNA was displaced in 10 minutes (Figure 2-7B and 2-7C). Time-dependent triplex displacement assay showed that LlaBIII was able to displace ~70% TFO from the triplex DNA in 2 minutes and reached a maximum of 80% in 4 minutes of incubation at 20°C in the presence of ATP. No triplex displacement was seen in the absence of ATP or LlaBIII (Figure 2-7B and 2-7D). We do see non-specific sticking of displaced single-stranded labelled TFO to LlaBIII which was ruled out by the addition of cold TFO to the reaction mixture whereas this cold TFO does not affect the stability of the triplex DNA.

DNA cleavage assay and ATPase assay was carried out at 25°C whereas the triplex displacement assay was performed at 20°C because triplex DNA is stable at 20°C in TMDK buffer and stability decreases as the temperature increases. We also observed TFO dissociation with the removal of Mg^{2+} ions from the reaction. The reaction stop buffer (STEB) contains EDTA which chelates out Mg^{2+} ion as a result of which we see complete dissociation of TFO from the triplex DNA. It seems that the divalent metal ion stabilised the local pKa so that is why we could shift from the MES (pH 5.5) to the reaction buffer (pH 8.0). The gel running buffer was kept acidic and contained $MgCl_2$ to stabilise the triplex DNA during electrophoresis. In spite of all precautions and performing the reaction at 20°C we always see some free TFO in the gel containing triplex DNA. This can be attributed to the salt conditions we are using in the experiments. We are using 150 mM KCl in the reaction, and triplex DNA is least stable in the presence of K^+ as compared to other monovalent cations (Beck et al., 2013). KCl cannot be removed from the reaction as it provides specificity to LlaBIII for the target sequence.
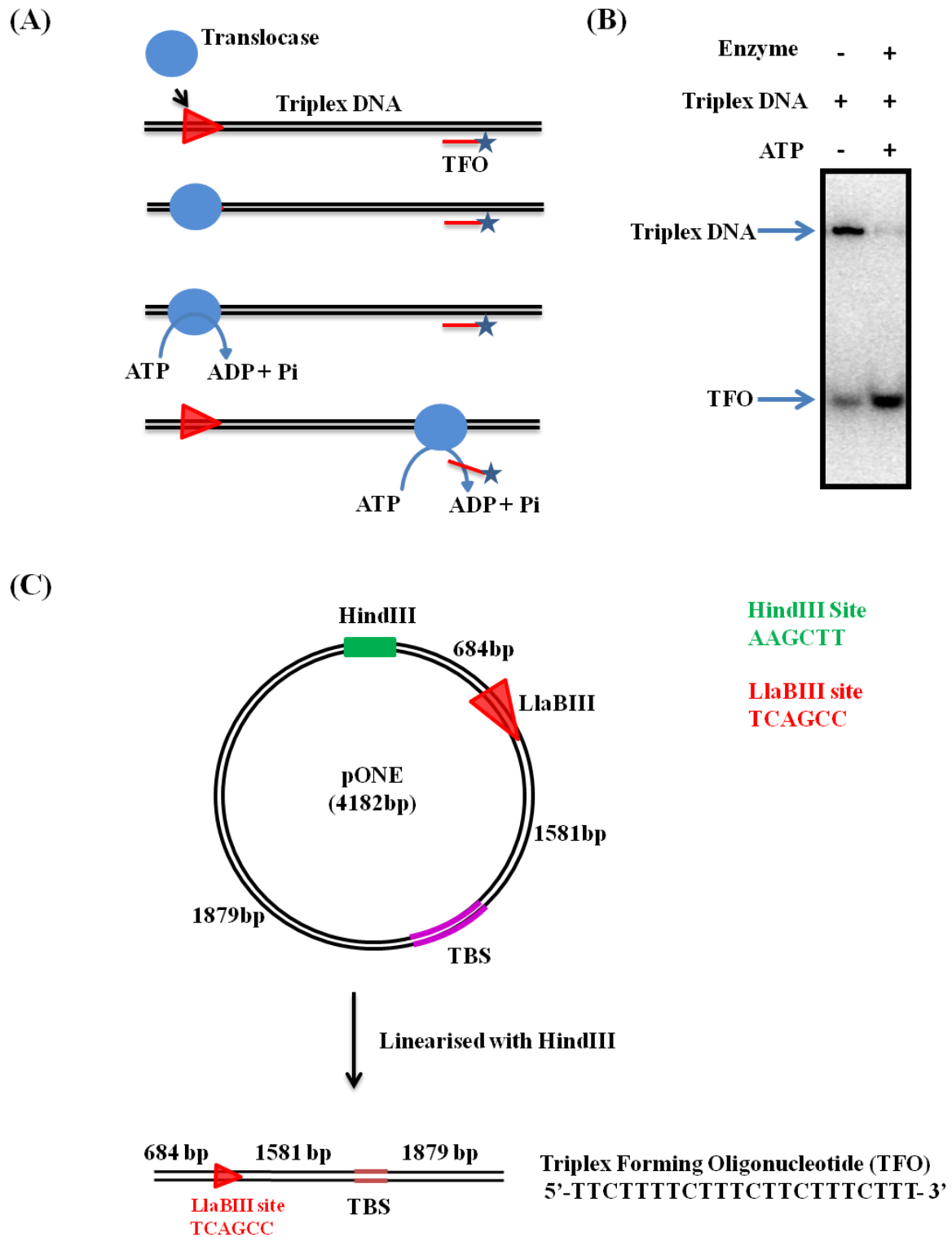
Figure 2-6: Triplex displacement and substrate generation. **(A)** Phenomenon of TFO displacement by an active translocase. TFO (shown in red with an asterik mark) is displaced by an actively translocating enzyme. **(B)** Gel showing triplex displacement. When a translocating enzyme encounters a third strand (TFO), it displaces it leading to an increase in the intensity for free and decrease in intensity for Triplex DNA. The positions of the triplex and TFO are marked with a blue arrow. **(C)** Generation of substrate DNA for the translocation assay from pONE plasmid. TBS (shown in magenta) is the TFO binding site.
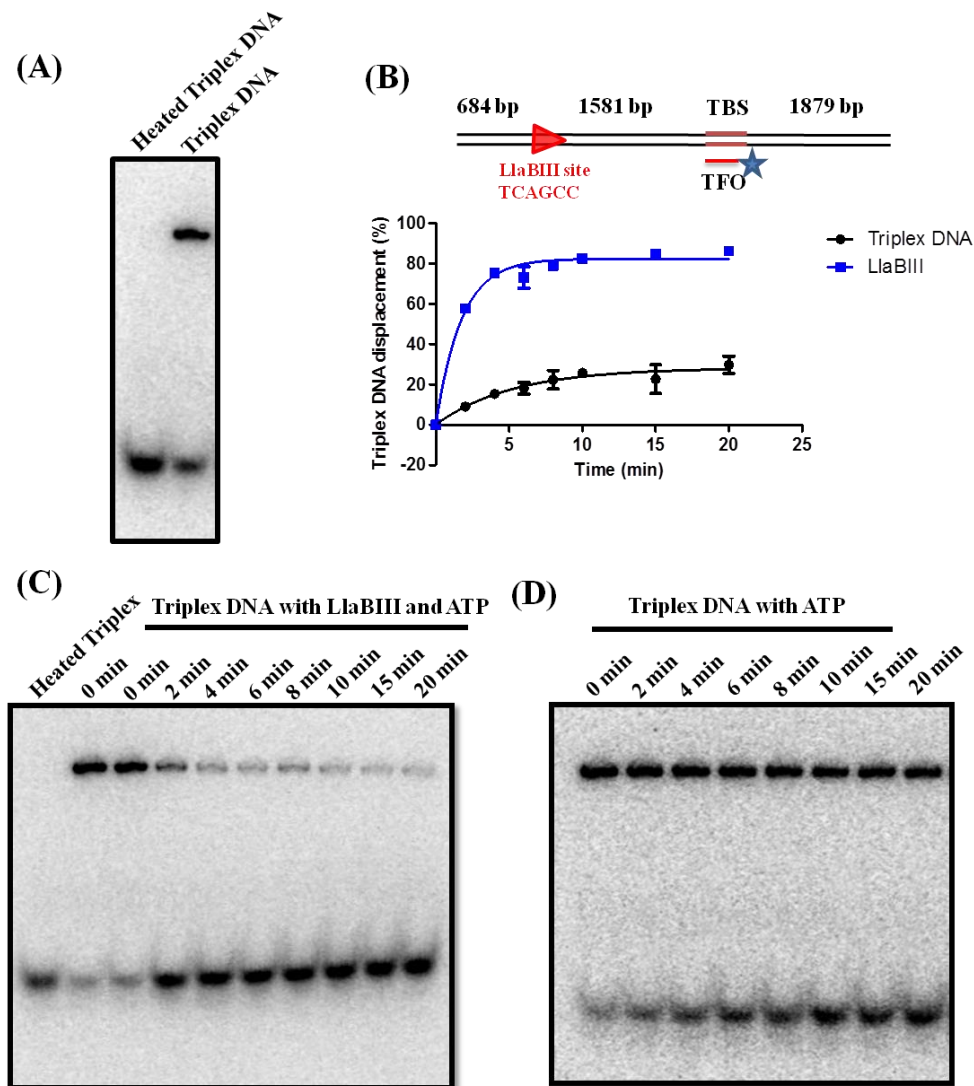
Figure 2-7: Triplex displacement assay of LlaBIII. **(A)** Gel showing the feasibility of triplex displacement assay in TMDK buffer. Lane 1 shows that the displaced TFO does not reanneal in the reaction conditions whereas lane 2 shows that triplex DNA is stable in TMDK buffer for the duration of the experiment. **(B)** Quantification of the triplex displacement by LlaBIII at different time points in the presence of 4 mM ATP. The experiments were performed in triplicates. **(C and D)** Representative gels showing triplex DNA displaced in the presence and absence of LlaBIII respectively.

2.3.5 DNA Binding Assay

As shown above, binding of these enzymes to the target sequence is crucial for DNA cleavage, ATP hydrolysis and DNA translocation activities. Consequently, we focused our efforts on the crystallisation of the target sequence-bound enzyme complex. The crystallisation of a protein-DNA complex, in general, is critically dependent on the length of the DNA substrate hence, it was important to characterise the binding of LlaBIII to DNA before starting crystallisation trials. Short DNA substrates that would be suitable for crystallographic studies were chosen for the binding studies by electrophoretic mobility shift assay (EMSA). Towards determining the binding affinity

of LlaBIII, we performed binding studies on cy5 labelled 50 bp DNA. Target site (TGAGCC) for LlaBIII was in the centre of the DNA with 23 bp upstream and 21 bp downstream of the site.
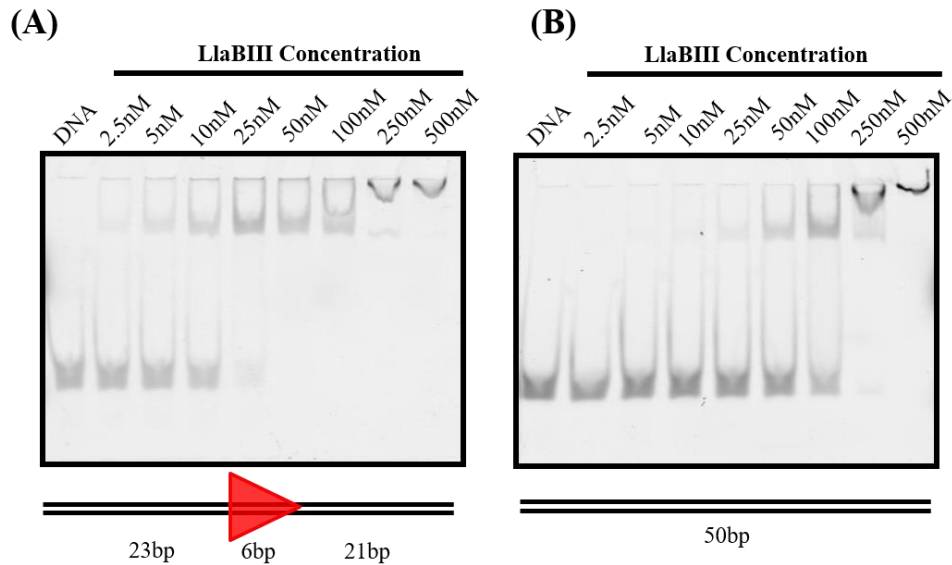
Figure 2-8: DNA binding assay with cy5 labelled DNA. **(A)** Binding with DNA containing target sequence. **(B)** Binding with DNA containing no target sequence. 10 nm Cy5 labelled DNA was used for the experiment.

EMSA showed the appearance of either a single or a double shift on the 6% native PAGE for both specific and non-specific DNA. We interpreted the first (lower) shift as LlaBIII binding to the target sequence on the specific DNA, which started at 2.5 nM of LlaBIII. The second shift (supershift) was observed when the enzyme concentration was 250 nM or 500 nM. In the case of non-specific DNA, the first shift occurred at a higher LlaBIII concentration of 25 nM, while the supershift occurred at 250 nM LlaBIII (Figure 2-8A). The supershift of both specific and non-specific DNA possibly corresponded to another LlaBIII molecule binding nonspecifically to the excess unbound DNA in an already formed complex of LlaBIII-DNA. The binding studies showed an apparent dissociation constant (K_d) of 1 nM, whereas the non-specific DNA bound with an apparent K_d of 10 nM (Figure 2-8B).

Using the assay, we tried to identify shorter specific DNA for crystallographic studies. The substrate DNA used for the binding studies were asymmetric with DNA length on upstream and downstream to the target sequence. The upstream length was varied and downstream length was kept constant with 6 bp. DNA binding assays with 22 bp, 28 bp and 46 bp DNA (Figure 2-9A) were performed with EtBr staining. The specific DNA 22 bp (Figure 2-9B), 28 bp (Figure 2-9C) and 46 bp (Figure 2-9D) showed an apparent K_d of 2 nM, while a non-specific 28 bp DNA had an affinity of 10 nM (Figure 2-9E). EMSA with the 22 bp and 28 bp DNA did not show a second shift suggesting that the DNA length was not enough for another LlaBIII molecule to bind.

(A)

(B) LlaBIII Concentration

(C) LlaBIII Concentration

(D) LlaBIII Concentration

(E) LlaBIII Concentration
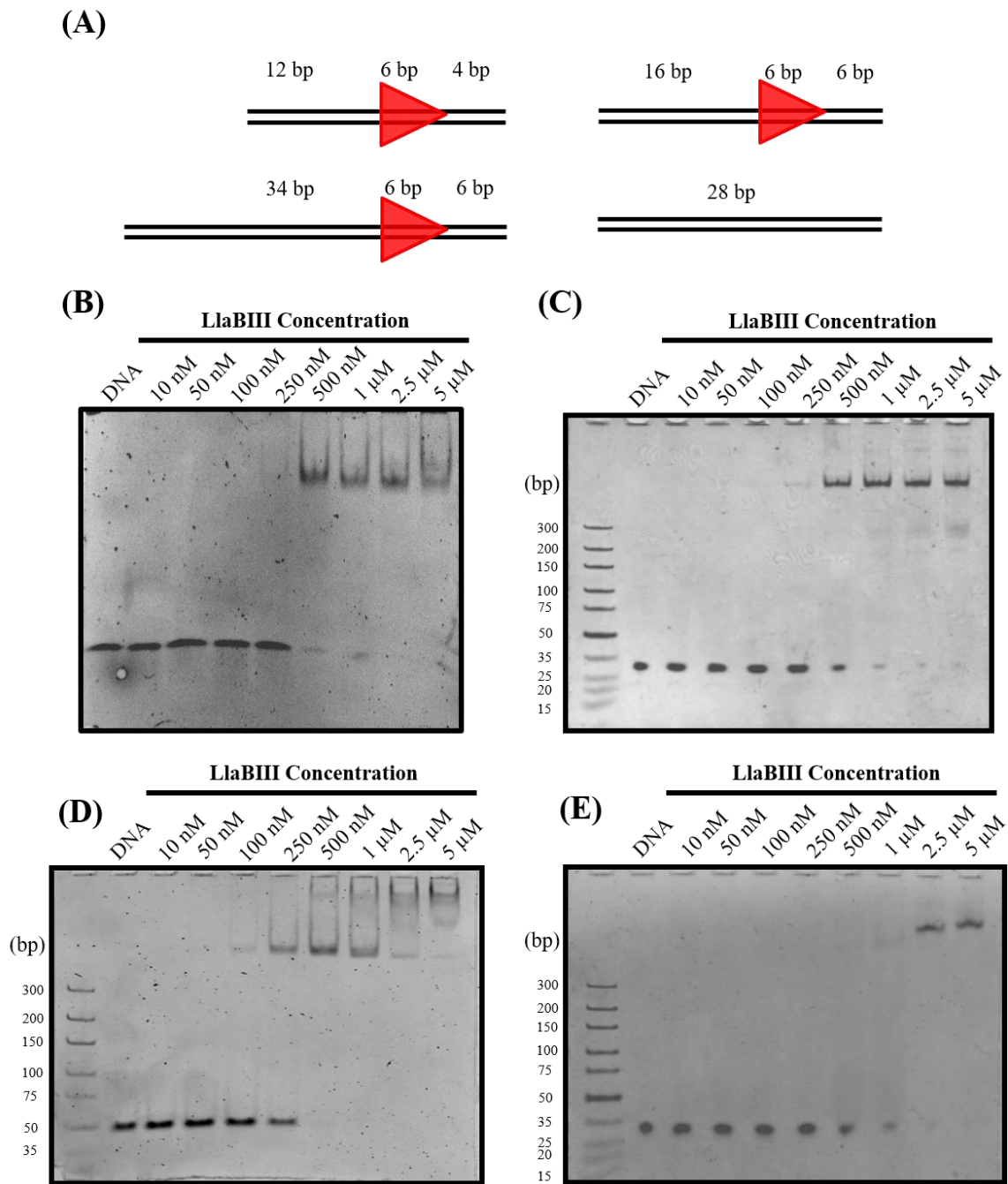
Figure 2-9: LlaBIII binding with variable dsDNA lengths. **(A)** Schematic representation of DNA substrates of variable lengths used for the EMSA studies. 6% native PAGE gels showing binding of LlaBIII with **(B)** 22 bp DNA **(C)** 28 bp DNA **(D)** 46 bp DNA **(E)** 28 bp non-specific DNA. All the gels were stained using ethidium bromide for 5 minutes.

2.3.6 Crystallisation of LlaBIII with DNA

The DNA binding studies provided a framework for crystallisation of LlaBIII bound to its target sequence. The 28 bp substrate described in the DNA binding studies was used in the initial set of crystallisation, as it also corresponded to the length of the DNA used for crystallisation of full-length LlaGI-DNA complex (Kulkarni et al., 2016) (Figure 2-10A). We could crystallise LlaBIII in complex with 28 bp DNA after screening 768 conditions (Figure 2-10B). Data collection

statistics are listed in Table 2-2. The space group was identified as P1 and confirmed using POINTLESS (Evans, 2006), and Matthew's coefficient (Matthews, 1968) indicated two molecules of LlaBIII in the asymmetric unit.
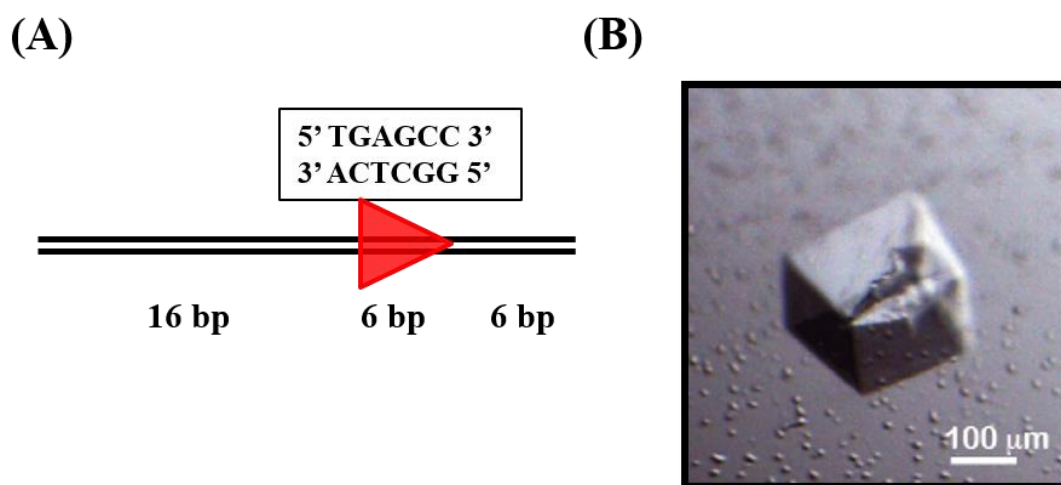
Figure 2-10: Crystal of LlaBIII bound to 28 bp DNA. **(A)** The schematic of the DNA used for the crystallisation experiment. The red arrow shows the LlaBIII target sequence and the direction of translocation. **(B)** Image of the crystal of LlaBIII bound to 28 bp DNA.

Table 2-2: Data collection statistics for LlaBIII bound to 28 bp DNA.

Data collection	LlaBIII-DNA
Space group	P1
a,b,c (Å)	62.9, 112.7, 146.2
α,β,γ (^{0})	103.3, 90.9, 105.8
Resolution (Å)	50-2.70 (2.77-2.70)
R_{sym} (%)	5.6 (66.8)
I/σI	11.2 (1.2)
Completeness (%)	96.4 (96.3)
Redundancy	1.8

2.4 Discussion

Large-scale purification of Type ISP RM enzymes, when carried out carefully to avoid proteolytic degradation, yielded pure and homogenous samples for crystallisation. Purification was a crucial step as this protein degrades during the purification steps, making it inappropriate for the

crystallisation studies (as the homogeneity of the sample is the prime requirement for crystallisation). Further optimization of the purification protocol with a reduced purification time of 24 hours yielded homogeneous and active LlaBIII. We did not use LlaBIIIΔN for the crystallisation trials as our priority was to obtain the crystal structure of wild-type LlaBIII with DNA. The biochemical characterisation showed that the ATPase which is stimulated upon recognition of specific DNA as reported earlier (Smith et al., 2009a). The decrease in ATPase rate for non-specific DNA was not because of the DNA length (40 bp), but it was the absence of target sequence as 22 bp DNA with target site also showed stimulation in the ATPase activity (Chapter 4; DNA length dependent ATPase assay). This showed that target sequence recognition brings conformational changes leading to activation of the enzyme. These enzymes contain ATPase domain which is proficient in ATP hydrolysis and DNA translocation (Sisakova et al., 2013; Smith et al., 2009b). As expected, it showed an active DNA translocation activity, displacing ~70% TFO from the triplex DNA in 2 minutes. In agreement with the ATPase and DNA translocase activity, LlaBIII actively cleaved the substrate DNA with the head-to-head oriented target site.

DNA binding experiment was pivotal in deciding the DNA for co-crystallization with LlaBIII. It leads us to decide the optimal length of DNA required to form the specific interaction with the substrate DNA and also provided us with the stoichiometry of protein to DNA for forming a homogeneous protein-DNA complex. DNA binding experiments also suggested that LlaBIII effectively interacts with the target sequence and upstream DNA. LlaBIII showed two kinds of shift upon interaction with DNA. The first shift was LlaBIII binding specifically to the target sequence on specific DNA at lower enzyme concentration where the enzyme preferred the target sequence. The supershift was the result of non-specific interaction of the excess LlaBIII (at 25 fold higher than DNA) with the flanking region of the substrate DNA. We also saw the same supershift with non-specific DNA binding at higher enzyme concentration. Supershift disappeared when the DNA length was shortened suggesting the binding of the second LlaBIII molecule lead to the second shift.

In conclusion, the designed strategy for Type ISP RM enzymes purification and biochemical characterisation paved the path towards obtaining crystal and diffraction data for Type ISP enzyme LlaBIII bound to 28 bp DNA.

Chapter: 3   Structural characterisation of LlaBIII and dsDNA translocation

3.1 Introduction

The prokaryotic ATP-dependent restriction-modification (RM) enzymes provide a potent defence against infection by foreign and bacteriophage DNA and accordingly have a widespread distribution (Dryden et al., 2001; Labrie et al., 2010). Although recognition of specific sequences (targets) in the foreign DNA leads to nucleolytic cleavage (restriction), cleavage of self DNA is prevented by the methylation (modification) of the target by the same enzyme or enzyme complex. Since the isolation of the first such enzymes in 1968 (Linn and Arber, 1968; Meselson and Yuan, 1968), which helped launch the molecular biology revolution, many ATP-dependent RM enzymes have been characterised, including the classical hetero-pentameric Type I systems (Loenen et al., 2014a) and the closely related but monomeric single polypeptide Type ISP systems (Loenen et al., 2014a; Smith et al., 2009a). These enzymes are a model for understanding modular, multifunctional protein machines (Murray, 2000), particularly in formulating concepts of protein-DNA recognition, DNA methylation and base flipping, nuclease activity (Rao et al., 2014; Smith et al., 2009a), double-stranded (ds)DNA translocation by superfamily 2 (SF2) helicases (Durr et al., 2006; Stanley et al., 2006) and long-range communication by enzymes (Halford et al., 2004; Schwarz et al., 2013). The first insights into the molecular organization of ATP-dependent enzymes came from structural analysis of the Type I RM enzymes EcoKI and EcoR124I using a combination of negative-stain electron microscopy, neutron scattering and structural modelling (Kennaway et al., 2012). However, despite over 40 years of research, the molecular details of these enzymes' actions are unclear because of the lack of high-resolution structures.

To address the lack of high-resolution structural data, we undertook structure-function studies of Type ISP enzymes LlaGI and LlaBIII from *Lactococcus lactis*. The Type ISP enzymes perform the disparate functions of restriction and modification by coordinating four functional domains within a single polypeptide: a target recognition domain (TRD) that recognizes a 6 or 7 base pair (bp) asymmetric sequence, an N6-adenine methyltransferase (MTase) that modifies one strand of the sequence using *S*-adenosyl methionine (AdoMet) stimulated by ATP, an SF2 helicase-like ATPase motor and a nuclease6. Unlike the well-studied SF1 and SF2 helicases that translocate single-stranded (ss) DNA and ssRNA, much less is known about the mechanism of dsDNA translocation by SF2 ATPases, which is important in many biological phenomena such as nucleosome remodelling, removal of stalled RNA polymerases or modulation of gene expression (Hopfner et al., 2012; Narlikar et al., 2013; Thoma et al., 2005; Wollmann et al., 2011). Structure-Function analysis of Type ISP enzymes would further our understanding of dsDNA translocation, particularly in the context of a fully functioning molecular machine.

We also sought to understand how the Type ISP nuclease domains interact to produce a dsDNA break. Previous biochemical analysis pointed to a loop-translocation mechanism leading to DNA cleavage, most similar to that of classical Type I RM enzymes (Sisakova et al., 2013; Smith et al., 2009a; Smith et al., 2009c; van Aelst et al., 2013). After Type ISP target recognition, extensive ATP hydrolysis (~1−2 ATP per base pair) produces unidirectional DNA translocation downstream of the target (Sisakova et al., 2013; Smith et al., 2009c; van Aelst et al., 2013). Because of the directional motion, DNA cleavage only occurs when there is at least one pair of head-to-head targets on the same DNA. DNA with isolated sites, or pairs of sites in other orientations are, at best, only nicked (Smith et al., 2009a). dsDNA breaks are produced at nonspecific loci between the head-to-head targets consistent with convergent enzyme translocation and collision (Sisakova et al., 2013; Smith et al., 2009c; van Aelst et al., 2013).

The prevailing Type ISP model predicts that the ATPase domain will be situated downstream of the MTase-TRD and that, despite the complexity of communication, cleavage will result from direct dimerization of nuclease domains in a collision complex. However, the structure we present shows that the nuclease-ATPase domains of Type ISP enzymes are located upstream of the MTase-TRD domains relative to the bound directional target sequence. This unexpected arrangement has important implications for both DNA translocation and cleavage. Guided by the structure and based on single-molecule biophysical and biochemical studies, we propose a model for dsDNA-break production resulting from multiple, random strand nicks.

3.2 Materials and methods

3.2.1 Structure determination of LlaBIII-DNA

The data were processed using XDS (Kabsch, 2010), and the intensities were scaled and merged using XSCALE (Kabsch, 2010), and converted to structure factors using TRUNCATE (French and Wilson, 1978). LlaBIII-DNA crystallized in the space group P1. The structure solution for LlaBIII-DNA was obtained by molecular replacement with MOLREP (Vagin and Teplyakov, 2010) using the coordinates of an N-terminal deletion mutant of LlaGI (PDB ID: 5FFJ) (Kulkarni et al., 2016). Default parameters values as in CCP4I (Potterton et al., 2003) were used for the search resulting in a unique solution. Alternate cycles of model building using Coot (Emsley et al., 2010), and refinement of coordinates and individual B-factors using phenix.refine (Afonine et al., 2010) was carried out to build the structure of the full-length LlaBIII-DNA complex. Hydrogens were automatically added to the model during refinement. The refinement statistics for the structures are provided in Table 3.1. The final model had 96% of the residues in the favoured,

4% in the allowed and 0% in the disallowed region of the Ramachandran plot. PyMOL (http://www.pymol.org) was used to generate figures.

3.2.2 Modeling studies

The models of LlaBIII bound to longer DNA and the collision complex were generated using Coot (Emsley et al., 2010). The former was generated by addition of nucleotides upstream to the co-crystal DNA. The conformation of the nucleotides was approximated to that of a B-form DNA. A model of the collision complex was generated by placing the structure of two molecules of LlaBIII-DNA side by side such that the two TRDs and the downstream ends of the DNA were contiguous. The pseudo-continuous DNA thus formed has two molecules of LlaBIII oriented head-to-head. The upstream ends were extended by the addition of nucleotides having the conformation of B-form DNA.

3.3 Results

3.3.1 Architecture of Type ISP RM enzymes

 X-ray crystallographic studies initiated using full-length LlaGI yielded crystals that diffracted X-rays poorly. As an alternate strategy, we carried out studies on a close homolog, LlaBIII, which has a very high sequence identity of ~80% (the first 1,020 amino acids have ~98% sequence identity, and the remaining sequence has 48% identity). The two proteins can functionally cooperate to cleave DNA (van Aelst et al., 2013). We determined the structure of full-length LlaBIII bound to a 28 bp DNA substrate mimic (Figure 3-1) to a resolution of 2.7 Å (Table 3-1). The asymmetric unit contained two LlaBIII−DNA molecules related by two-fold rotational symmetry, resulting in a pseudo-continuous DNA of 56 bp (Figure 3-2). This assembly appeared to be a result of crystal packing and may not be functionally relevant as the targets are oriented tail-to-tail (which does not support cleavage). The remaining crystal contacts arose from protein-protein and protein-DNA interactions between symmetry-related molecules. Characterisation by size-exclusion chromatography (SEC) coupled to multi-angle light scattering (MALS) confirmed that the protein is monomeric (Chand et al., 2015).

The structure revealed six structural domains with an unexpected arrangement: the N-terminal Mrr family nuclease followed by the N-core and C-core RecA folds of the ATPase located upstream of the target, and an all α-helical domain (referred hereafter as the coupler) that links the ATPase to the MTase and the C-terminal TRD (Figure 3-1 and 3-3). The DNA interacted extensively with the TRD and MTase but made fewer contacts with the ATPase as the DNA length limited the interaction to only a part of the N-core (Figure 3-4). Among the six domains, quality of electron

density for the nuclease was comparatively poor, possibly owing to intradomain and interdomain conformational mobility.

Table 3-1: Refinement statistics.

Refinement statistics	LlaBIII-DNA
Resolution (Å)	48.2-2.7
No. reflections (total/test)	99188/4969
Rwork / Rfree	21.8/26.0
No. of atoms	
Protein	23312
DNA	2283
Ion	3
Water	148
B-factors	
Protein (Å^2)	75.2
Ion (Å^2)	83.6
Water (Å^2)	57.7
R.m.s.d.	
Bond lengths(Å)	0.003
Bond angles (^{0})	0.686
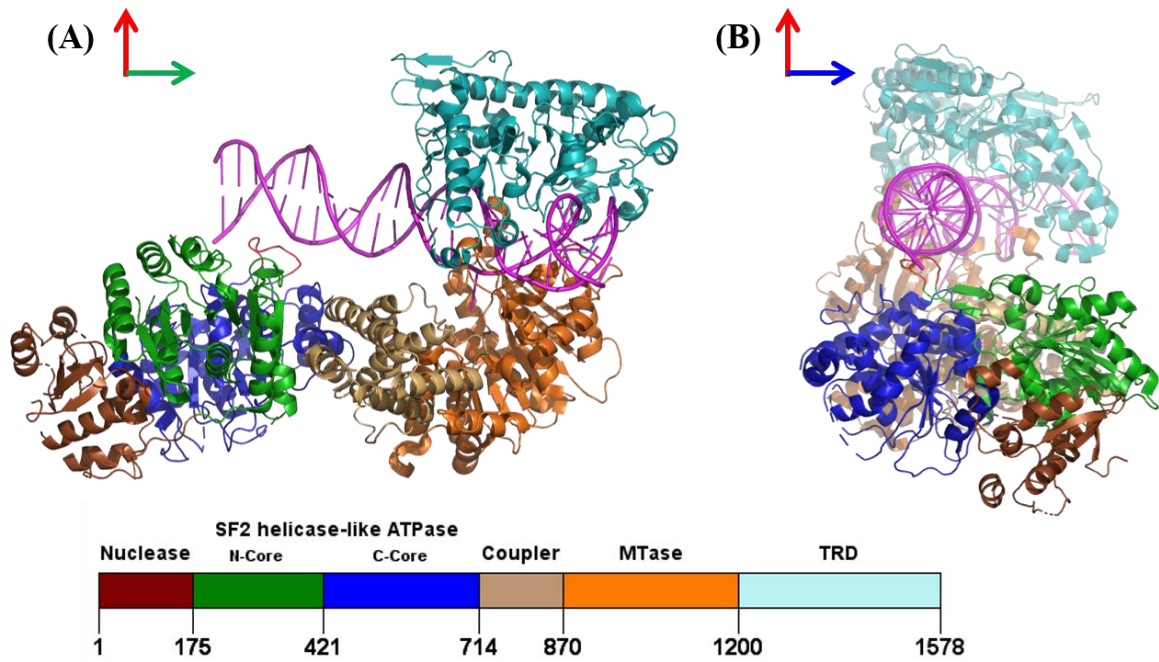
Figure 3-1: Ribbon diagram showing architecture of the modular Type ISP RM enzyme LlaBIII bound to 28 bp DNA substrate. TRD and MTase are recognizing the target sequence whereas the ATPase domain is interacting with the upstream DNA. DNA is shown in magenta and domains of LlaBIII are colored as mentioned in the lower panel with respective domain boundaries (PDB: 4XQK).
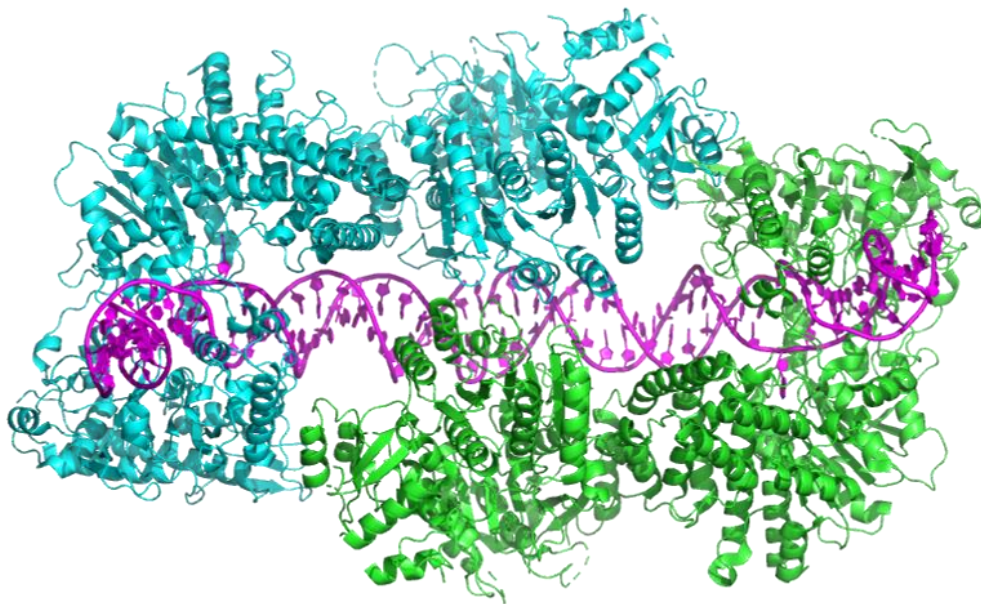
Figure 3-2: Ribbon diagram of two molecules of LlaBIII-DNA in asymmetric unit of the crystal. The magenta color shows DNA, green and cyan are two molecules of LlaBIII. The molecules are related by 2-fold rotational symmetry.
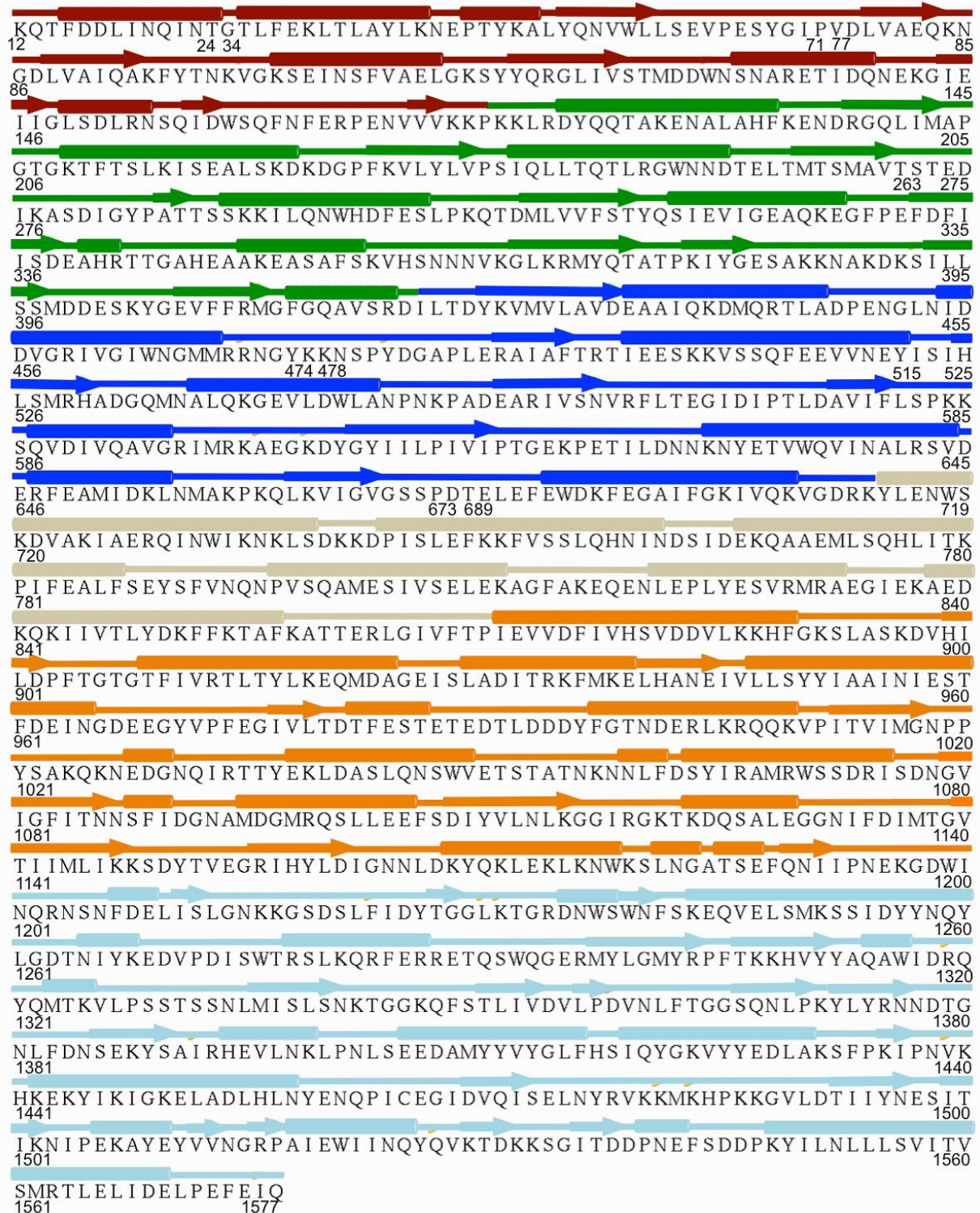
Figure 3-3: Secondary structure representation of LlaBIII. The secondary structure information was obtained using the program STRIDE (Frishman and Argos, 1995). The cylinder represents the helices and the strands are shown with arrows.
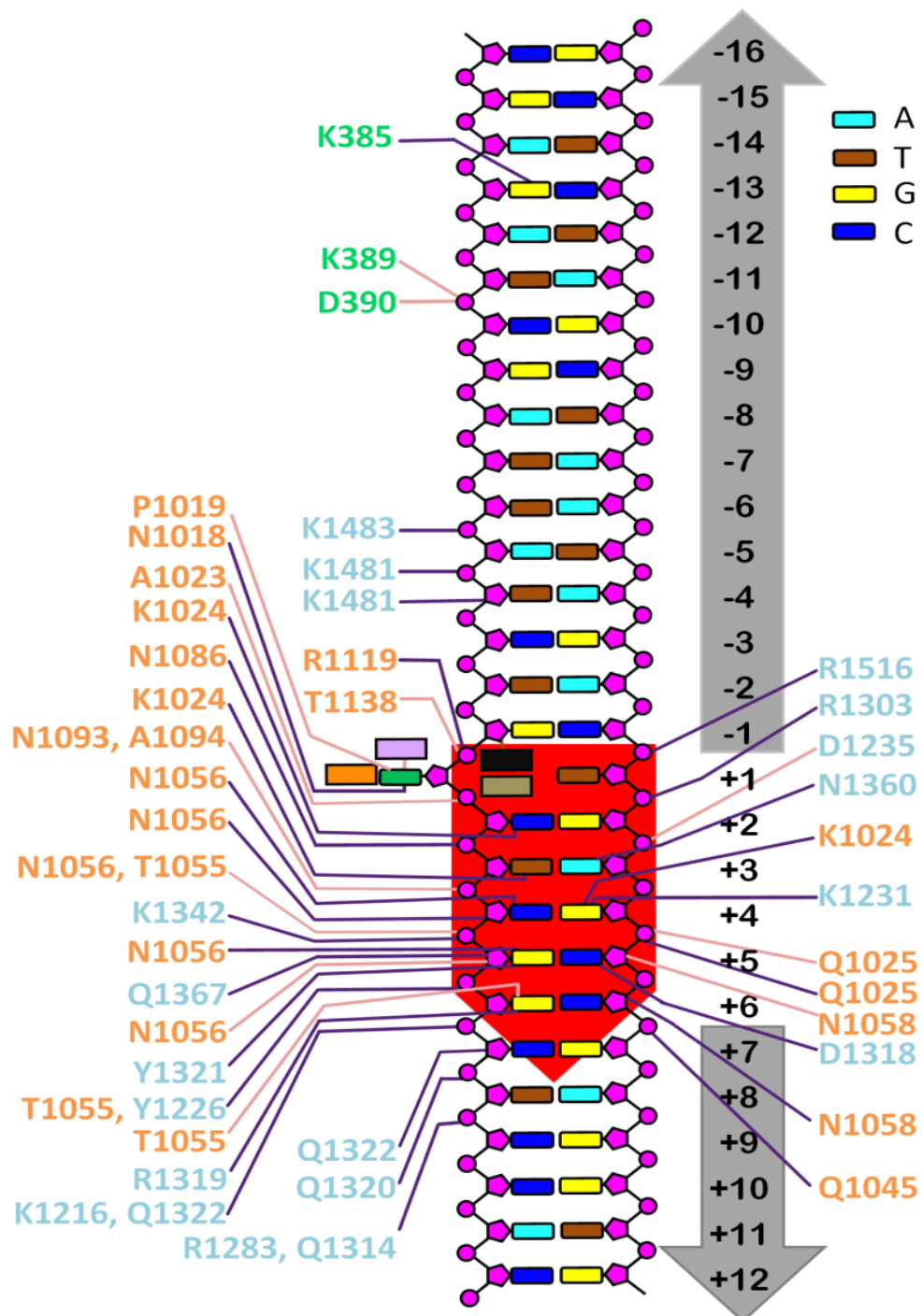
Figure 3-4: LlaBIII-DNA interaction. Amino acids are colored according to the domain color in figure 3-1 and 3-3. Cyan numbers represent TRD, orange for MTase and green for N-core ATPase domain. Main chain interactions are shown in pink lines whereas side chain interactions are represented with purple lines. The flipped adenine base is shown in green colored box. R1119 is shown with black box, M1137 with tan, F1134 orange and Y1021 with lavender box. The red box on the DNA represents the target sequence for LlaBIII. The flipped adenine is labeled as +1, upstream sequences are labeled with negative sign (-) whereas downstream sequences are labeled with positive sign (+).

Table 3-2: Inter-domain movement in two chains of LlaBIII. The rotation angles were obtained by superimposing a domain of chain B to the corresponding domain of chain A. Prior to this chain B was superimposed onto chain A with respect to the coupler domain (residues 715 to 870) to ensure a common frame of reference.

Domain	Rotation angle between the respective domains of chain A and B of LlaBIII	Polar angles of rotation (omega, phi, kappa)
Coupler	0°	0.0°, 0.0°, 0.0°
Nuclease	7.2°	68.6°, 24.6°, 7.2°
N-core domain of ATPase	4.6°	81.3°, 26.7°, 4.6°
C-core domain of ATPase	3.3°	82.8°, 26.2°, 3.3°
Nuclease-ATPase	4.5°	77.5°, 24.3°, 4.5°
Methylase	1.1°	85.8°, 140.4°, 1.1°
TRD	1.4°	107.8°, 131.4°, 1.4°

Comparison of the LlaBIII−DNA molecules in the asymmetric unit highlighted interdomain conformational plasticity (Table 3-2). In particular, the nuclease-ATPase domains moved as a unit with respect to the coupler about a hinge-like interface formed by two laterally packed α-helices, one from the C-core of the ATPase (E649−N656) and one from the coupler (P823−I835) (Figure 3-5). The HsdM subunit of classical Type I enzymes (PDB identifiers 2AR0 and 3UFB) (Park et al., 2012) and the multidomain Type IIG enzyme BpuSI (Shen et al., 2011) contain a similar all−α-helical domain of unknown function located N-terminally to the MTase domain. The domain has been postulated to be involved in switching between different conformational and functional states (Shen et al., 2011).
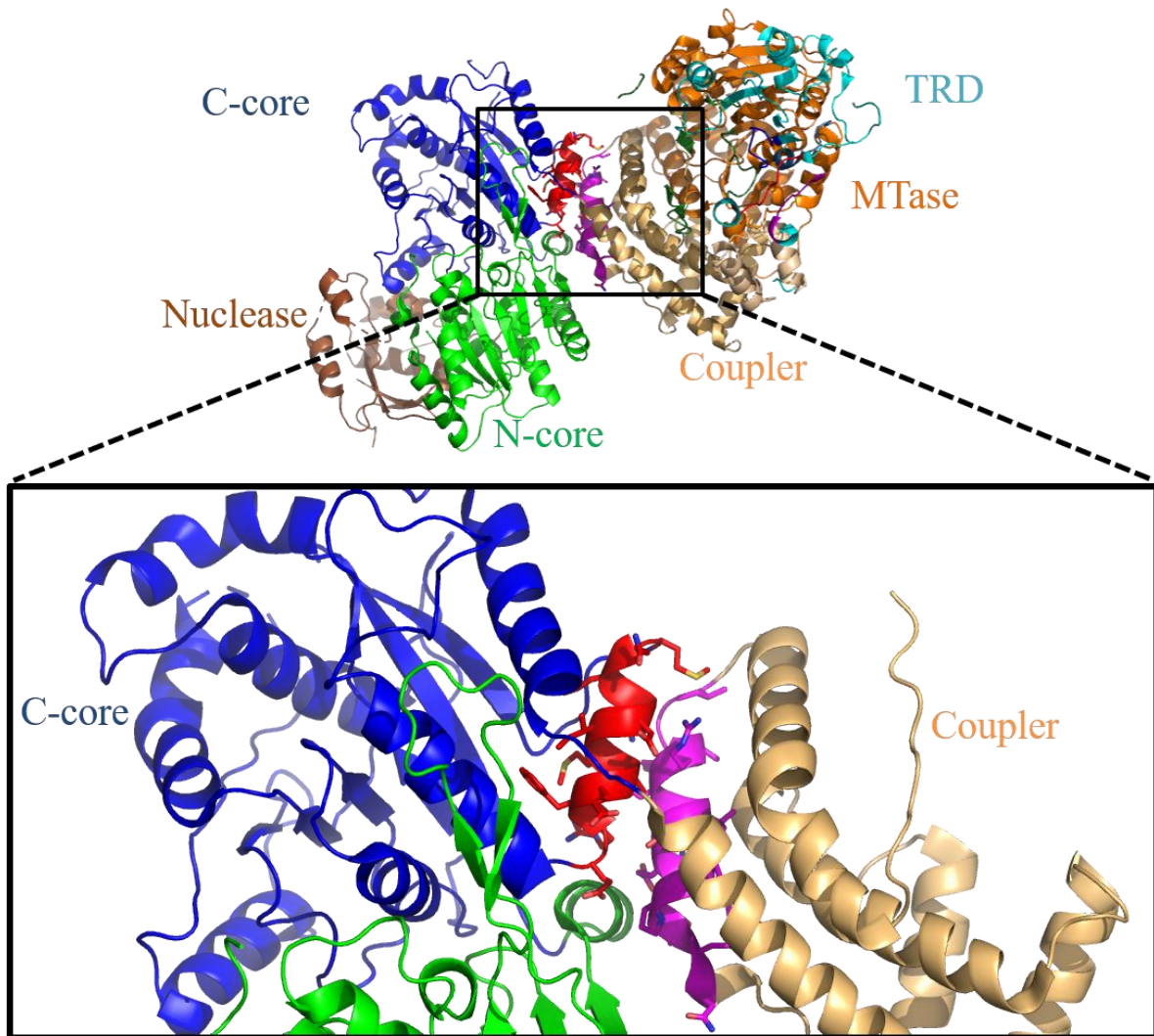
Figure 3-5: The hinge about which the nuclease-ATPase move about the coupler. A zoomed in view of two laterally packed α-helices (E649-N656 and P823-I835). The red colored E649-N656 is from C-core ATPase domain and the magenta colored P823-I835 is from coupler. The side chains for amino acid resides of these helices are shown in sticks.

3.3.2 The target sequence recognition by LlaBIII (TRD-MTase)

In the structure, the target was clamped by the MTase and the TRD (Figure 3-6A). The MTase of LlaBIII had a structural fold similar to that of the prototype γ-class N6-adenine MTase, M•TaqI (Goedecke et al., 2001), and also of the MTase domains of the classical Type I enzymes (PDB identifiers 2AR0, 2Y7C (Kennaway et al., 2012) and 3UFB (Park et al., 2012)) and the Type IIG enzyme BpuSI (Shen et al., 2011). However, the TRD of Type ISP (LlaBIII) shows much more interaction as compared to M.TaqI (Figure 3-4). The TRD of LlaBIII is also ~200 amino acid larger than a prototype N6-adenine methyltransferase M. *Taq*I (Goedecke et al., 2001), and it can be divided into three subdomains- the core (1205-1241 and 1291-1448), the jaw (1242-1288) and the guide (1441-1578) (Figure 3-6B). The core subdomain of the TRD forms the major interaction

(A)

(B)

Target Recognition Domain

Target Recognition Domain

Methyltransferase
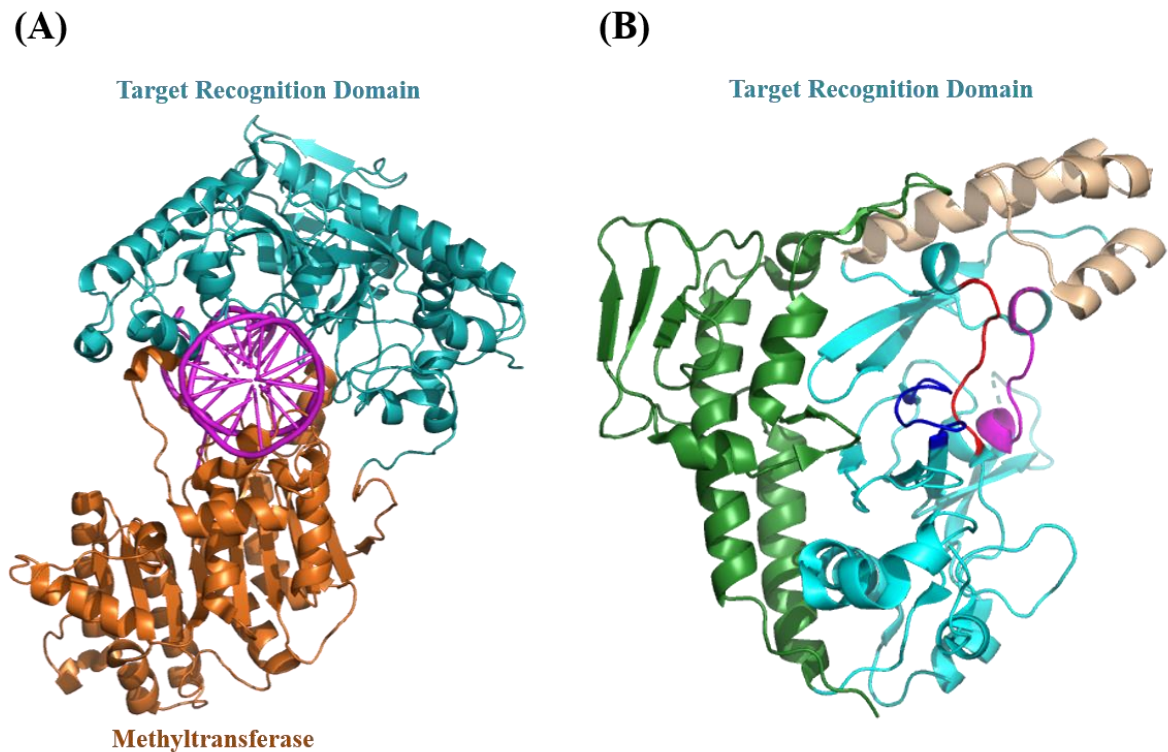
Figure 3-6: Target sequence recognition by TRD. **(A)** Ribbon diagram of TRD-MTase clamp encircling the target sequence. **(B)** Target recognition domain of LlaBIII with three sub-domains. The guide is shown in green, jaw in wheat color and the core is colored cyan. The structural elements of the core subdomain involved in recognition of target sequence are loop IV (red), loop V (magenta) and loop VI (blue).

with the target and contains the same fold as that of the TRD of M. *Taq*I. The jaw is made up of the three-helix bundle. These three helices along with MTase fastens the clamp around the DNA, aiding into the processivity of the enzyme. The C-terminal of the TRD is the guide. The guide has two layer open- faced β-sandwich structure. It contains four stranded N-type Greek key motifs. This Greek key motif is stacked against a layer of three helices with a β-hairpin loop insertion between them (Figure 3-6B). Kulkarni et al., 2016 has shown that among all these three subdomains of TRD, the amino acid sequence for the guide is highly conserved followed by the core and the jaw subdomain among Type ISP RM enzymes. The structural fold of the core is same as TRD1, and TRD2 of Type I (EcoR124I) and TRD of Type IIG (BpuSI) RM enzymes whereas similar structural folds for jaw and guide was not found in other RM enzymes (Kulkarni et al., 2016). The core of the TRD and MTase domain primarily recognizes the target sequence. The N-terminus and C-terminus helices of the guide subdomain form a helical bundle which is evocative of Type I RM enzyme HsdS subunit helical bundle formed from the central and distal conserved region (Kim et al., 2005; Loenen et al., 2014b). This helical bundle in Type I RM enzymes act as a molecular ruler, i.e. separates the two TRD to recognize the bipartite target sequence (Kulkarni et al., 2016). In LlaBIII, the guide (helical bundle with the β-hairpin loop) forms the concave surface for the core packing (Figure 3-6B). The guide does not take part in the target recognition,

but it steers the DNA towards the ATPase domain with the help of N-type Greek key motif situated at the end of the helical bundle.

(A)

(B)
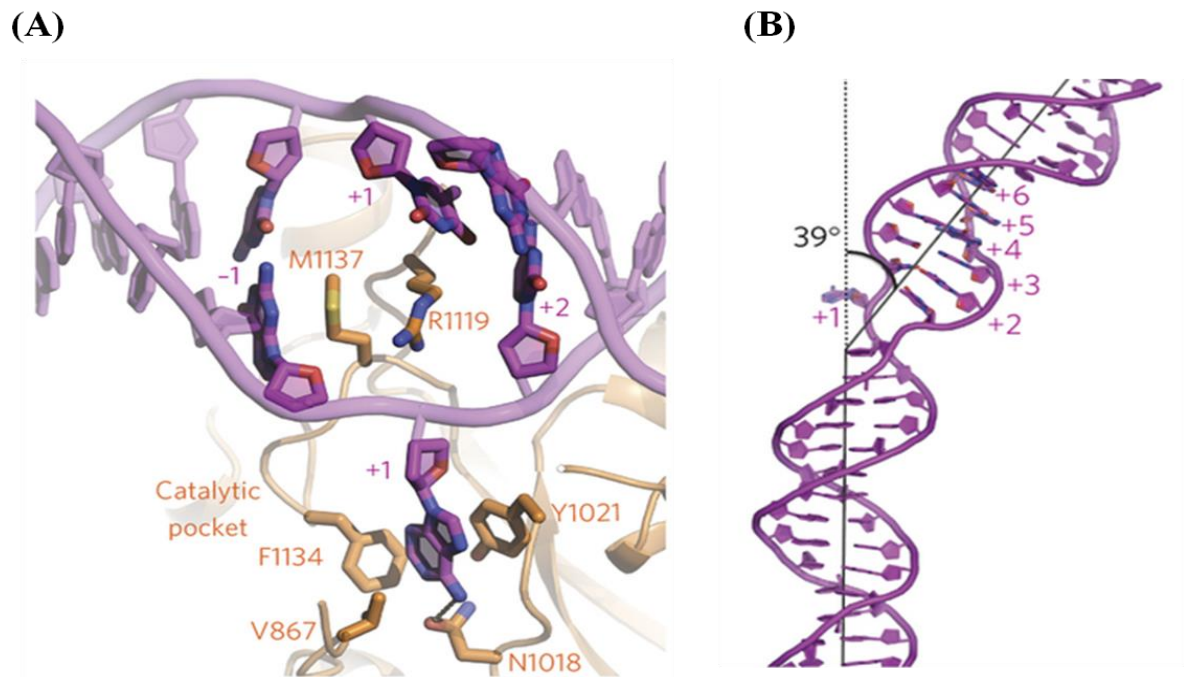
Figure 3-7: Adenine base flipping and DNA bending. **(A)** MTase-DNA contacts stabilising the flipped adenine base in the MTase catalytic pocket. Amino acid residues M1137 and R1119 plugging the cavity created due to adenine flipping. N6-position of adenine is in H-bond distance from catalytic asparagine residue. **(B)** DNA bending upon recognition of target sequence by TRD and MTase.

MTase domain of LlaBIII shows structural similarity to other γ-class adenine-N6 methyl transferases including M.*Taq*I and HsdM subunit of Type I RM enzymes. In the structure of LlaBIII bound to 28 bp DNA, the target sequence adenine (the adenine to be methylated by the MTase domain) is flipped out into the active site of MTase domain regardless of the absence of AdoMet in the crystallisation condition (Figure 3-7A). The flipped adenine is stacked by tyrosine (Y1021), and the N6 position is hydrogen bonded to the catalytic asparagine (N1018). This adenine also makes an energetically favourable T-shaped aromatic-aromatic interaction with the phenylalanine (F1134) which would sterically block guanine from entering the pocket. The cavity generated in the DNA because of adenine flipping out is plugged by the two amino acid residues R1119 and M1137. These two amino acid residues form a continuous stack with +2 and -1 position of the DNA. The cavity created by DNA distortion in case of LlaBIII is larger whereas in case of M. *Taq*I, domain movement compresses the DNA locally and a proline residue intercalates with the DNA bases at the flipped position (Goedecke et al., 2001). In most other DNA modification enzymes which flip the nitrogenous bases, the cavity is frequently filled a single hydrophobic residue (Cheng and Roberts, 2001). LlaBIII bends the DNA by 39^0 at the target site (+1 position) leading to drastically increased width of the major and minor grooves, unlike the prototype M.

*Taq*I which only increases the width of the minor groove by 3Å (Goedecke et al., 2001)(Figure 3-7B).
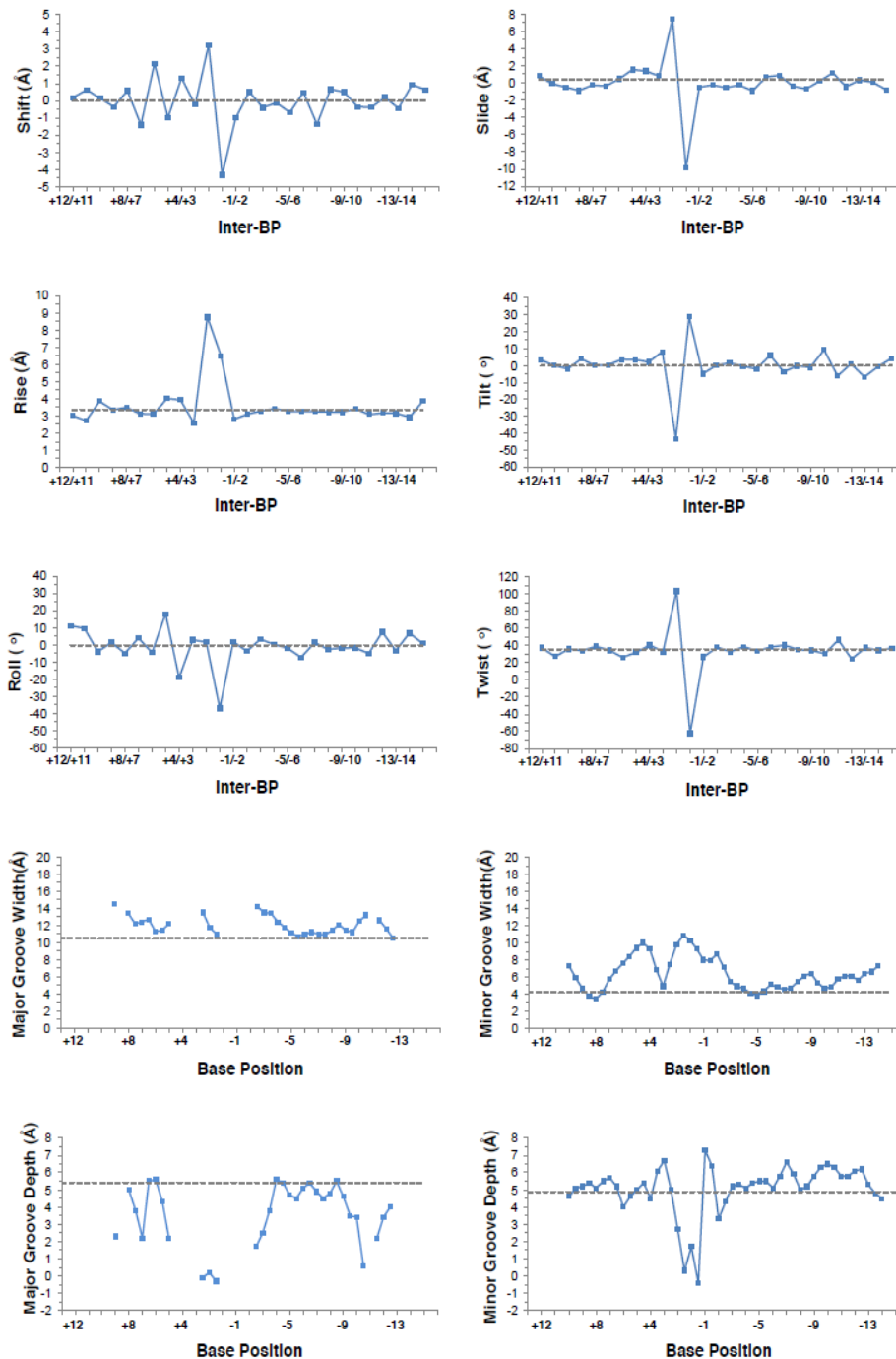
Figure 3-8: Geometrical parameters of the DNA bound to LlaBIII. The geometrical parameters were calculated for the DNA chains E and F using the program Curves+ (Lavery et al., 2009). The major groove width and depth at some of the regions could not be calculated due to large deformation. The broken grey lines indicate the average parameter values for B-DNA obtained using the coordinates 1BNA.

The MTase-TRD clamp deformed the DNA, leading to altered geometrical parameters at the target (Figure 3-8). The changes in the groove geometry facilitated sequence-specific protein-DNA interactions. The MTase read the sequence via the minor groove (Figure 3-9A), whereas the TRD read it via the major groove (Figure 3-9B). LlaBIII recognizes a 6 bp target sequence TnAGCC.
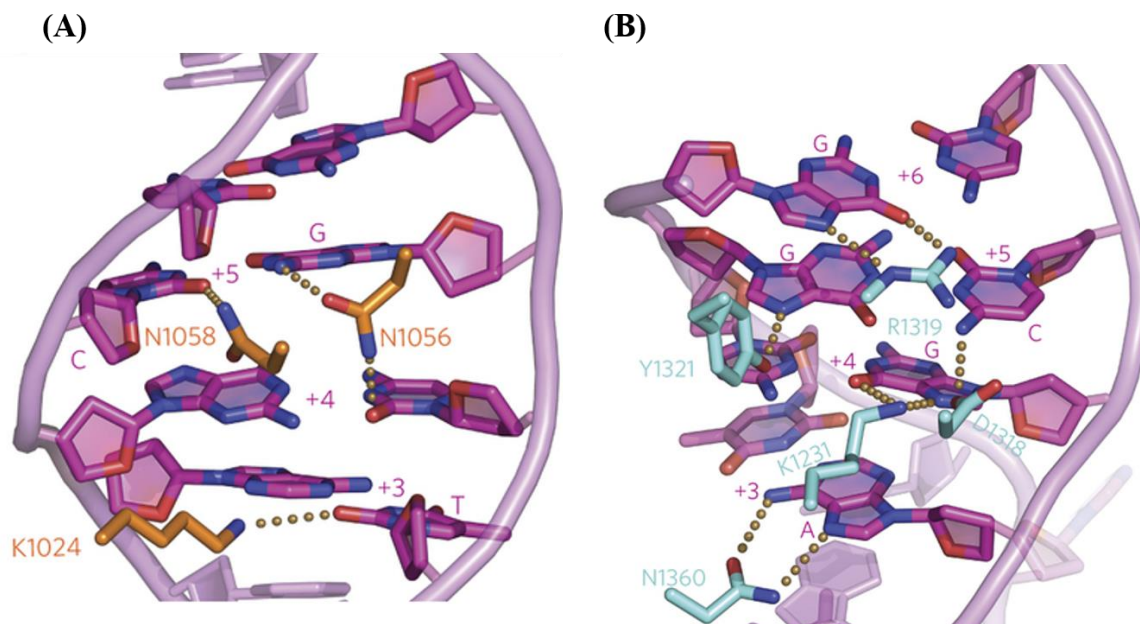
Figure 3-9: Target recognition by TRD and MTase. (A) Minor groove readout by MTase (orange). **(B)** Major groove readout by TRD (cyan). Dotted lines show H-bond interactions.

The structural elements involved in recognition of the target sequence are: Loop I (G1017 - N1027), Loop II (S1052 - F1061) and Loop III (K1123 – M1144) in the MTase domain; and Loop IV (G1228 – R1234), Loop V (W1316 – T1324) and Loop VI (P1357 –S1366) of the core subdomain in the TRD (Figure 3-10). Loop I contains the catalytic asparagine (N1018), which catalyses the transfer of the methyl group. It interacts with the bottom strand flipped adenine at +1 position and thymine at position +3 with N1018 and K1024 respectively. Y1021 also stacks the flipped adenine. Loop VI (N1360) recognizes adenine of top strand at position +3 whereas thymine at position +1 is not being recognized specifically by any amino acid residue. Loop II interacts with bases at position +4, +5 and +6. N1058 and N1056 recognize top strand guanine and complementary cytosine at position +4 respectively. These two amino acid residues also interact with the cytosine-guanine base pair at position +5 whereas the guanine of the bottom strand at position +6 is recognized by T1055. Loop IV (K1231) recognises top strand guanine at position +4. Specificity at position +5 is brought by the interaction of D1318 of Loop V, where it interacts with top strand cytosine. Y1321 interacts with bottom strand guanine. R1319 interacts with bottom strand guanine at position +6. Loop III is involved in stabilising adenine in MTase catalytic pocket. F1133 forms a T-shaped aromatic interaction with flipped adenine and plugs the cavity created

due to base flipping. The guanine-cytosine base pair at position +2 lies at the mouth of TRD-MTase clamp, but no specific interaction was observed.
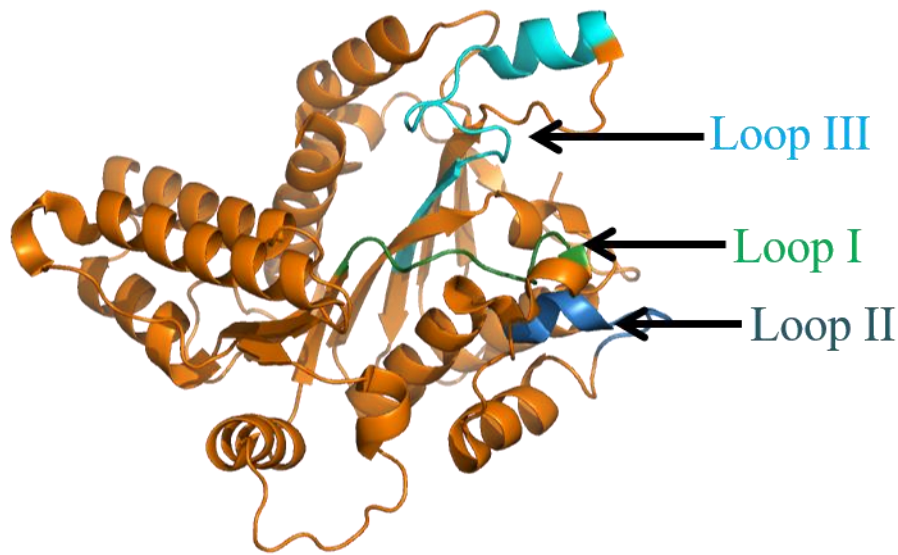
Figure 3-10: MTase domain of LlaBIII. Three structural elements involved in target sequence recognition and methylation of adenine- loop I (green), loop II (blue) and loop III (cyan).

3.3.3 Mechanism of dsDNA translocation

Sequence-specific target binding by the MTase-TRD is essential for initiation of DNA translocation (Sisakova et al., 2013; Smith et al., 2009c). The structure revealed that upon target recognition, the MTase-TRD steered the upstream DNA toward the ATPase, thereby engaging the N-core and C-core (Figure 3-1 and 3-11), illustrating a mechanism that couples target recognition and DNA translocation, mediated by DNA conformational change. Structural comparison with other SF1 and SF2 ATPases revealed that the disposition of the RecA cores bound to dsDNA (Figure 3-11) was similar to that seen among ssDNA-bound helicases oriented to bind ATP (Buttner et al., 2007; Gu and Rice, 2010; Lee and Yang, 2006; Saikrishnan et al., 2009; Velankar et al., 1999) and to the apo structure of the Swi2/Snf2 ATPase from zebrafish Rad54 (Thoma et al., 2005) (Figure 3-12). In contrast, the orientation of the cores in the only other structure of a helicase-like ATPase bound to dsDNA, i.e., the Swi2/Snf2 ATPase from *Sulfolobus solfataricus*, differs by almost 180° (Durr et al., 2005). The relevance of this conformation is unclear (Durr et al., 2006; Durr et al., 2005).

The Type ISP N-core−DNA interaction was mediated by a β-hairpin loop (S383−D390) that interacted with the major groove and gripped the 3′-terminated strand. The main-chain amides of K389 and D390 interacted with phosphate at position −11, whereas in one of the monomers of LlaBIII, K385 inserted into the major groove and contacted the base at −13 (Figure 3-4 and 3-11). In the other monomer, the lysine side chain was disordered, suggesting that the interactions involving the main-chain atoms were the primary contributors to the grip between the DNA and

the hairpin loop. It is, however, likely that new interactions are formed between the hairpin loop and the DNA in other conformational states that may occur during a complete ATPase cycle.
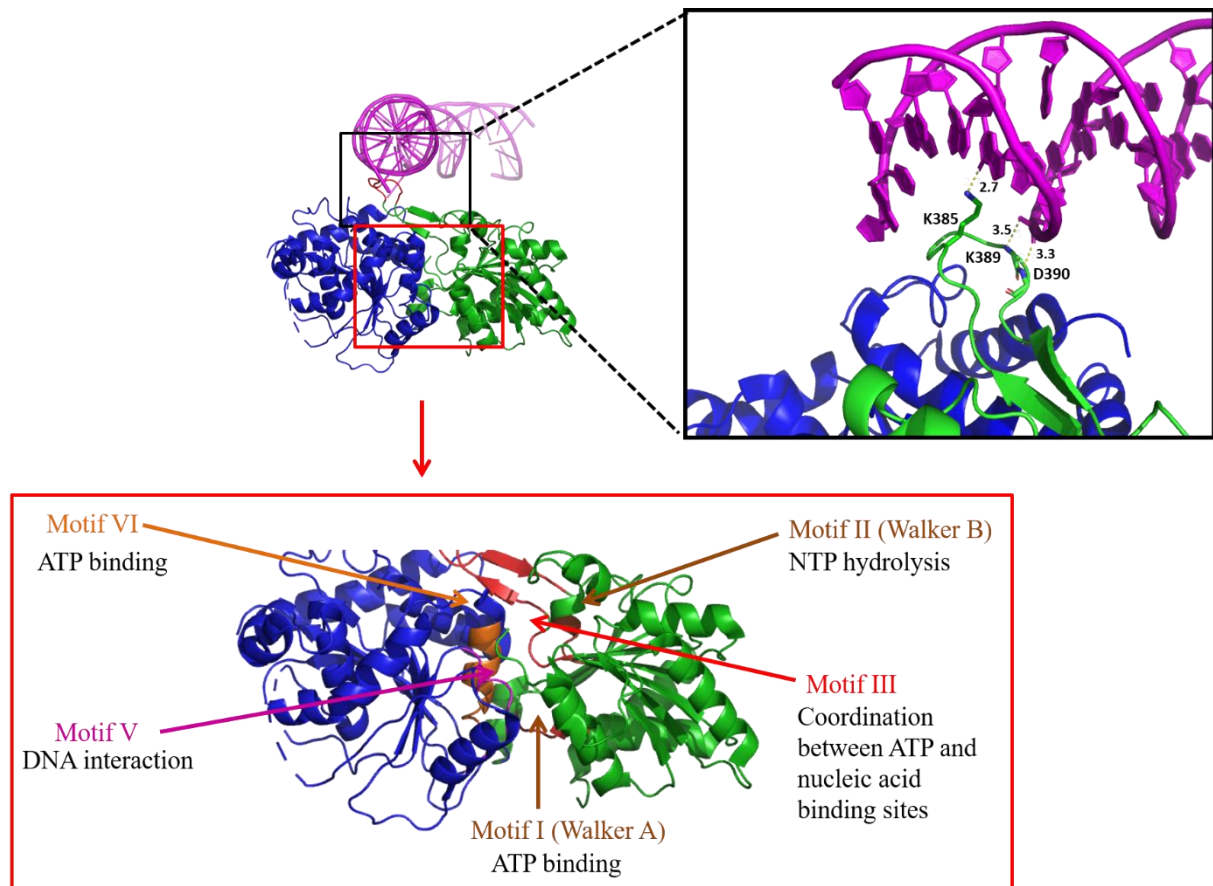
Figure 3-11: ATPase domain of LlaBIII. The β-hairpin loop residues (E382 to D390) interacting with the DNA (shown in the top right panel) and conserved SF2 helicase-like motifs (shown in the bottom panel).

The structure revealed an enzyme that was competent to carry out either methylation or translocation/cleavage in the presence of the appropriate cofactor: AdoMet or ATP. It appears that the tussle between the two activities is avoided by the apparent rate of methylation being slower than ATP hydrolysis (Chand et al., 2015). As a consequence, ATPase activity would more often initiate translocation on unmodified foreign DNA, leading to cleavage rather than methylation. On the host DNA, the absolute requirement for two unmodified head-to-head targets for cleavage (Sisakova et al., 2013; Smith et al., 2009a; van Aelst et al., 2013) ensures that the newly replicated host DNA, with one strand methylated and the other unmethylated, would not be cleaved. After multiple redundant translocation events, maintenance methylation would eventually occur.
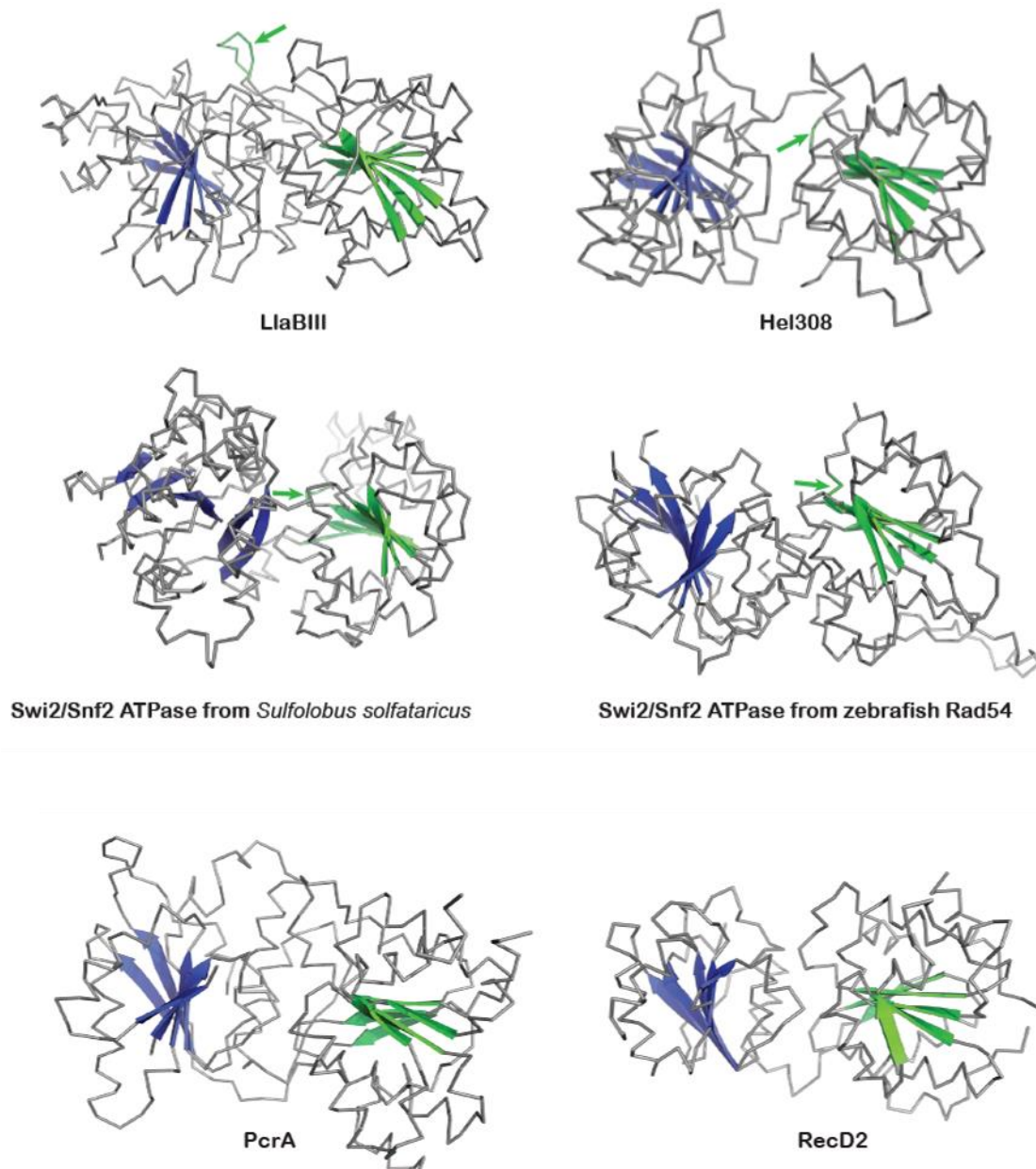
Figure 3-12: Relative orientation of the N-core and the C-core domains in SF2 and SF1 helicase-like ATPases. Ribbon diagram showing relative orientation of N-core (green) and C-core (blue)of the ATPase domain. β-hairpin loop in LlaBIII is shown as ribbon for clarity. Green arrow shows the equivalent position in other members of SF2 helicase. LlaBIII (PDB: 4XQK) (Chand et al., 2015), Hel308 PDB: 2P6R (Buttner et al., 2007), Swi2/Snf2 from Sulfolobus solfataricus (PDB: 1Z63) (Durr et al., 2005), Swi2/Snf2 ATPase from zebra fish Rad54 (PDB: 1Z3I) (Thoma et al., 2005), PcrA (PDB: 2PJR) (Velankar et al., 1999), RecD2 (PDB: 3GP8) (Saikrishnan et al., 2009).

3.3.4 Mechanism of long-range communication along DNA

In contrast to expectations of the prevailing model (Figure 3-13A), the location of the Type ISP ATPase upstream of the MTase-TRD−target complex appeared completely inconsistent with Type I−like loop translocation, as downstream ATPase motion would push the MTase-TRD off the target (Figure 3-13B and 3-13C). In addition, collision of two converging Type ISP enzymes at

the MTase-TRD interface will leave the upstream nucleases at locations apparently too distant to either interact directly or produce a dsDNA break.

(A)

(B)

(C)

LlaBIII bound to DNA

LlaBIII starts ATP hydrolysis and translocation
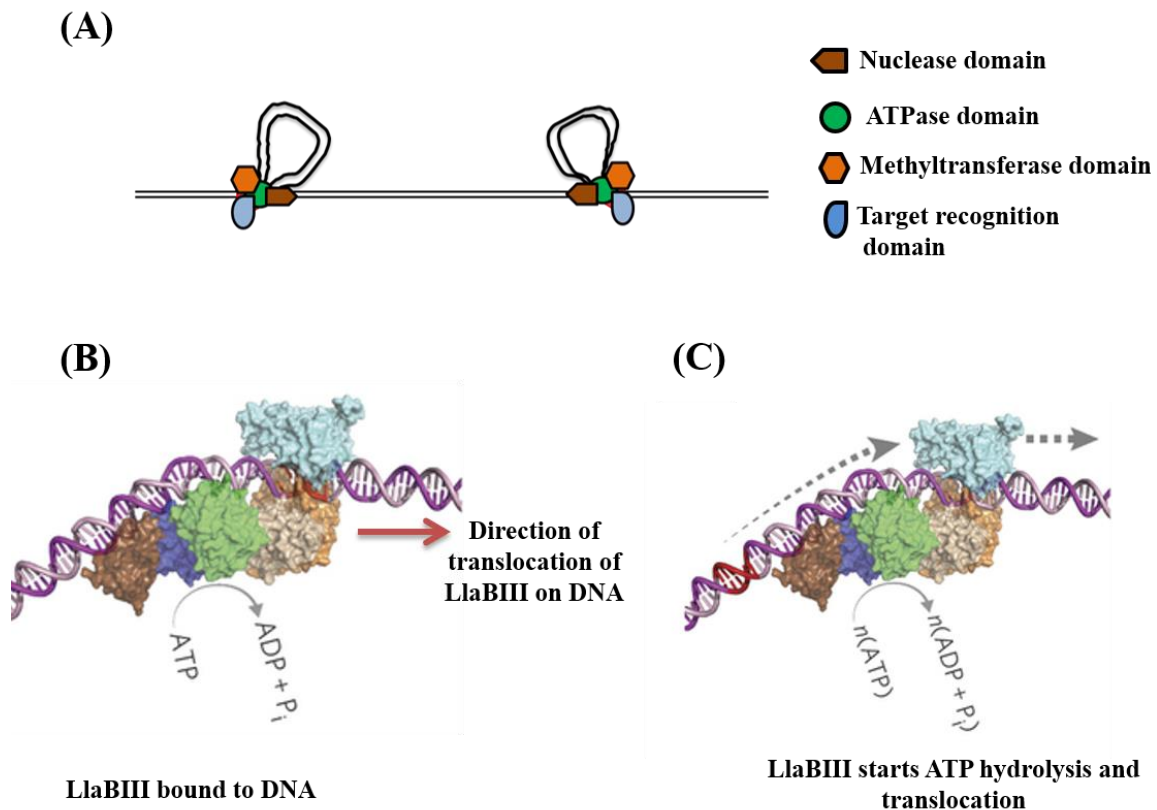
Figure 3-13: Mechanism of long range communication by Type ISP RM enzymes. **(A)** The prevailing DNA looping model for communication between target sites. This model shows that for looping ATPase domain should be downstream to the TRD and MTase domain. **(B, C)** Contrary to this LlaBIII contains ATPase domain (green and blue) upstream to the TRD (cyan) and MTase (orange), suggesting that it will remodel the TRD-MTase clamp upon ATP hydrolysis during translocation leading to the release of target site.

Magnetic tweezers microscope (MTM) assays have been used to distinguish between loop-dependent and loop-independent translocation models for ATP-dependent RM enzymes (Ramanathan et al., 2009; Seidel et al., 2005). We (in collaboration with Prof. Mark D. Szczulkun, Bristol University, UK) used a similar approach here, in which cleavage of DNA with two head-to-head Type ISP targets would be observed as a loss of magnetic bead tracking, whereas loop translocation would be observed as a uniform downward bead movement (Figure 3-14). This experiment was performed with LlaGI. Cleavage occurred rapidly after addition of LlaGI and ATP, independent of the DNA stretching force and without any evidence for DNA shortening and hence loop formation (Figure 3-14) (Chand et al., 2015). The cleavage rate was nonetheless dependent on ATP concentration, consistent with translocation-driven communication and cleavage (Chand et al., 2015).
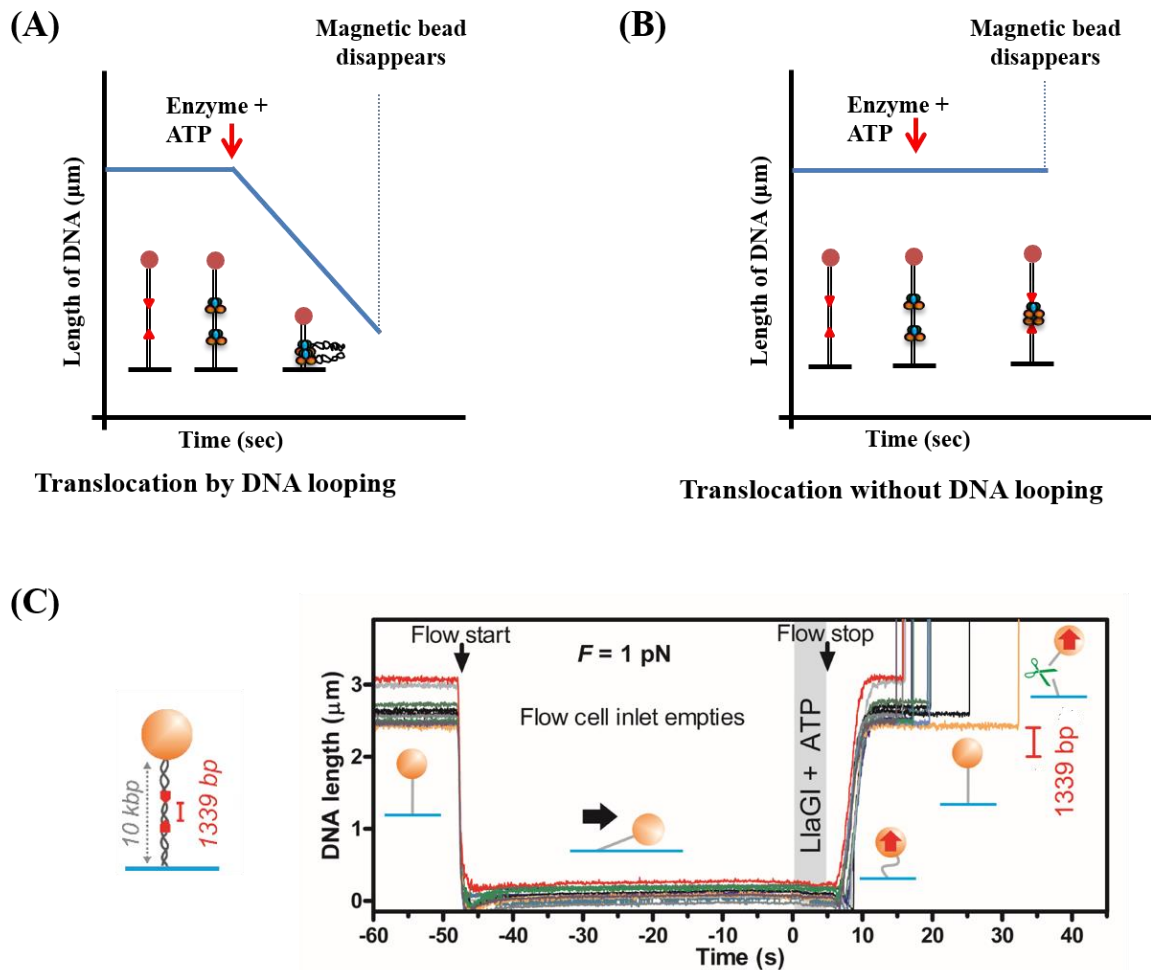
Figure 3-14: Loop independent translocation by Type ISP RM enzymes. Schematic showing translocation by DNA looping and loop independent translocation respectively. **(A)** Upon ATP hydrolysis enzyme starts translocation by pulling the DNA leading to loop formation, in turn the length of DNA decreases (shown in image A). **(B)** In case of loop independent translocation enzyme releases the target sequence and translocates along the DNA, hence no decrease in DNA length is observed. **(C)** In case of Type ISP RM enzyme (LlaGI) when ATP is supplied (at 0 time point shown on X-axis) DNA reaches its maximum length and no decrease in length is observed till the magnetic bead is lost upon cleavage. Each chromatogram represents single experiment (This experiment was performed in Prof. Mark Szczelkun's Laboratory, University of Bristol, UK) (Chand et al., 2015).

3.3.5 Regulation of the nuclease activity

DNA cleavage does not occur during translocation but is activated upon convergent protein collision (Sisakova et al., 2013; Smith et al., 2009a; Smith et al., 2009c; van Aelst et al., 2013). The key catalytic residues of the Mrr family nuclease (Smith et al., 2009b) were clustered in a putative active site cleft (Figure 3-15), which pointed away from the DNA path (Figure 3-1). The loop containing the essential D74 (Smith et al., 2009b) was unstructured (Figure 3-15), suggesting a catalytically impaired conformation that becomes activated on engaging with the DNA upon collision. The loop could also be unstructured owing to the absence of divalent cation in the buffer, which is thought to interact with the catalytic triad of D74, D78 and K94 (Smith et al., 2009b). We note that the side chains of D78 and K94 were also disordered in the structure.
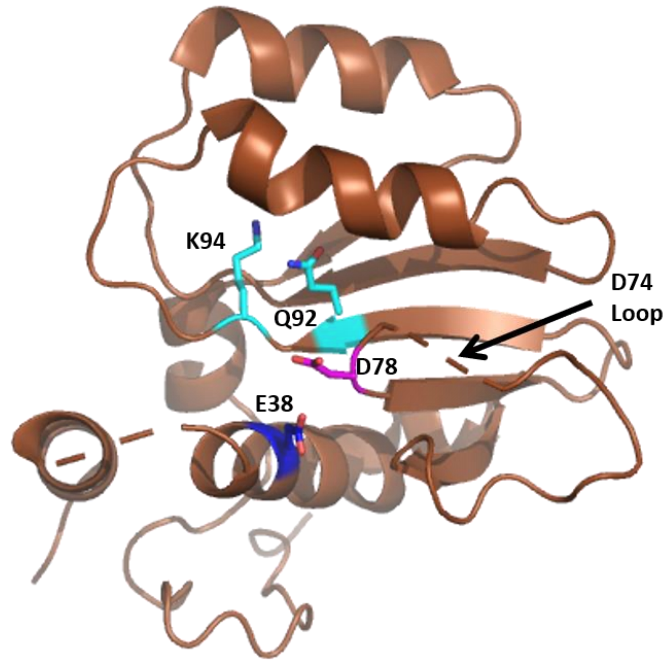
Figure 3-15: Structure of nuclease domain of LlaBIII. Nuclease domain with catalytic motifs- motif I (E38), motif II (D78 and unstructured D74 loop) and motif III (Q92 and K94).

Nuclease activation, along with structuring of the D74 loop, may additionally arise because of ATP-dependent or ATP-independent nucleoprotein rearrangements that engage the active site with the DNA. Simple modelling based on the structures of the SF2 helicase NS3 from hepatitis virus C bound to ssDNA and an ATP analogue (Gu and Rice, 2010) suggested that ATP-induced domain closure would move the nuclease closer to the DNA path. However, in this orientation, the nuclease active site still pointed away from the DNA, and the active orientation is possibly only attained upon collision. In addition, pivoting of the nuclease-ATPase unit about the hinge between the ATPase and the coupler (Figure 3-5) could help engage the nuclease with the DNA.

Activation of the upstream nuclease effected by interdomain plasticity was supported by the mapping of a slow, ATP-independent nicking activity, which exclusively targeted the bottom 3′-5′ strand at position −30 (Chand et al., 2015).Time dependent DNA cleavage mapping on head-to-head substrate DNA showed that the minimum separation of two DNA nicks by 30 bp and maximum 120 bp with a median of 57 bp which further increased to ~68 bp (Chand et al., 2015). The minimum separation of 30 bp was achieved as a result of conformational change in the collision complex which was mapped as the separation between two nuclease-ATPase complex (Chand et al., 2015; van Aelst et al., 2015). This increase in the separation between two nicks with time was achieved because of the movement of collision complex on the DNA leading to multiple nicks (Chand et al., 2015).

3.4 Discussion

The data presented here indicate that the upstream ATPase of a Type ISP enzyme remodels its downstream MTase-TRD−target complex to allow target release (Figure 3-16). Upon binding of ATP, Type ISP domain movement and ensuing rearrangements in the ATPase-DNA interactions, including those involving the β-hairpin loop, would pull the DNA in an upstream direction, leading to movement downstream of the complete enzyme without a necessity for DNA loop formation. This model is quite distinct from that of the classical Type I RM enzymes. We suggest that the coupler will play a key role in transferring conformational strain during initiation. The flipped base seen in the MTase active site would also act as an 'anchor' to prevent target release; this must be returned to its base-paired location in the DNA before initiation can proceed, and this may be accelerated by ATP-induced strain relayed by the DNA and coupler.

Remodeling of the MTase-TRD must involve loss of target interactions. However, the remodeled MTase-TRD may remain associated with the DNA to act as a sliding clamp, enhancing processivity. Alternatively, the clamp may open fully, allowing the coupler-MTase-TRD units to swing completely off the DNA. This motion may help explain the ~30 bp closest approach of the nuclease domains.

The Type ISP remodeling activity is somewhat similar to the role of the ATPase subunit of the Type III RM enzymes (Schwarz et al., 2013). However, although those enzymes use a burst of ATPase activity to break the MTase-target interactions, subsequent bidirectional movement along DNA is thermally driven without the need for ATP hydrolysis. In contrast, the Type ISP enzymes continue to consume ATP during unidirectional translocation (Smith et al., 2009c). It remains to be seen whether the Type III enzymes also move toward their cognate MTase subunits, 'pushing' them off the target as in the Type ISP scheme (Figure 3-16) or move in the opposite direction to 'pull' the MTase subunits off the target.

The Type ISP structure also provides a simple, single-polypeptide framework to understand other remodeling processes that may involve active disruption of protein-nucleic acid complexes by an ATPase motor, such as the DExH/D RNA helicase (Jankowsky et al., 2001), SF1 DNA helicase PcrA (Fagerburg et al., 2012; Park et al., 2010), the Swi2/Snf2 family ATPase Mot1 (Hopfner et al., 2012; Wollmann et al., 2011) or nucleosome remodelers (Hopfner et al., 2012; Narlikar et al., 2013). The above structural comparison suggests that the ATPase cycle proposed for bonafide helicases, involving domain closure and opening of N- and C-cores upon ATP binding and hydrolysis (Gu and Rice, 2010; Lee and Yang, 2006; Saikrishnan et al., 2009; Velankar et al., 1999) is conserved in dsDNA-translocating ATPases. In turn, the Type ISP ATPase would, therefore, use an equivalent inchworm mechanism for both remodeling and translocation. The β-hairpin loop is unique to Type ISP ATPases (Figure 3-12), indicating variation in the translocation

mechanisms among SF2 ATPases despite conservation of canonical motifs and overall structures. The wide spacing of the DNA strand breaks produced is consistent with the upstream locations of the nuclease domains as predicted in a head-on collision complex (Figure 3-16D). As the nucleases cannot interact directly, activation upon collision is most likely due to conformational strain from the ATPases transferred via the couplers. Nonetheless, additional remodeling and movement of the collision complex must be required to further process the DNA to produce a dsDNA break. We propose a 'DNA shredder' model in which cumulative nicking events eventually result in a dsDNA break (Figure 3-16). This contrasts with the precise DNA cleavage produced by two closely spaced nucleases, as suggested for Type I and III enzymes and observed in many other, simpler nucleases.
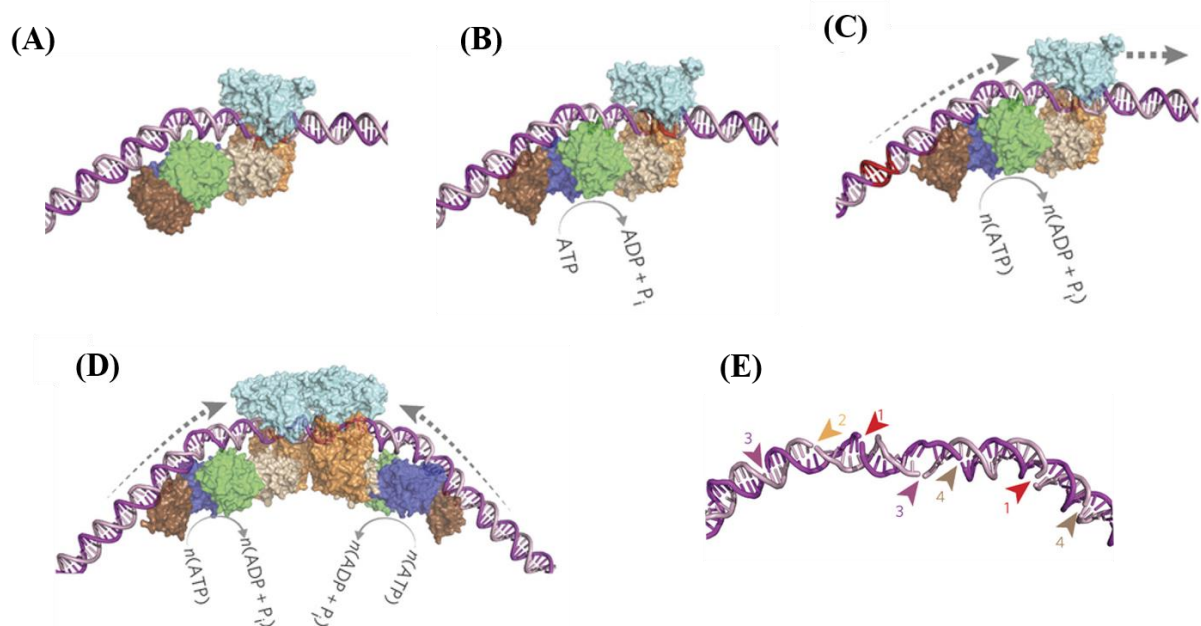
Figure 3-16: Model for loop-independent DNA translocation and extensive nucleolytic DNA processing. **(A)** The pre-initiation complex. The nuclease is in an inactive conformation. **(B)** The ATPase cycle loosens the MTase-TRD grip on the DNA. **(C)** dsDNA translocation downstream of the target (red). **(D)** Convergence of two enzymes initially brings the nucleases ~75 bp apart. **(E)** Example of stochastic 'DNA shredding' by a collision complex (Chand et al., 2015). Numbers are the order of the nicking events.

The collision complex would initially result in the formation of one or two strand breaks. The location of this initial cleavage will be random, dictated by the preceding stochastic translocation events. Rearrangements of the complex due to interdomain plasticity together with the nuclease's sequence preference, results in varied loci of cleavage and hence a range of initial spacings between the nicks by the two enzyme (Chand et al., 2015). The observed 30 bp minimum reflects a collision complex where both coupler-MTase-TRD units have swung off the DNA, allowing the closest approach of two helicase-nuclease units that still allows for stress-mediated activation (Chand et al., 2015; van Aelst et al., 2015). Continued ATPase activity can then remodel the

collision complex, leading to further upstream and downstream movement and nicking events, causing compound damage culminating in a dsDNA break (Figure 3-16).

We speculate that, depending on the cellular host, broken DNA with variable 3′ overhangs produced by Type ISP enzymes (Chand et al., 2015) would not be a substrate for the corresponding dsDNA break repair enzymes (RecBCD or AddAB), which have a preference for blunt ends (Dillingham and Kowalczykowski, 2008; Taylor and Smith, 1985; Yeeles et al., 2009). As a consequence, even if the foreign DNA were to have the regulatory sequences required for homologous recombination, they would not enter that repair pathway. dsDNA break formation by multiple nicking events would additionally prevent simple religation of the DNA ends. Where the cell also encodes a CRISPR-Cas system, the small DNA fragments generated during cleavage by Type ISP enzymes (or during post-cleavage processing by classical Type I enzymes (Endlich and Linn, 1985)) may feed into the spacer-acquisition pathway (adaptation), as suggested for the DNA fragments generated by RecBCD (Levy et al., 2015).

In conclusion, the distinct nucleolytic activity of the Type ISP enzymes contrasts with mechanisms used by dimeric nucleases and illustrates yet another strategy evolved toward resistance in the perpetual bacteriophage-host arms race (Labrie et al., 2010).

Results presented in this chapter have been published in the following research article-

Chand, M.K., Nirwan, N., Diffin, F.M., van Aelst, K., Kulkarni, M., Pernstich, C., Szczelkun, M.D., and Saikrishnan, K. (2015). Translocation-coupled DNA cleavage by the Type ISP restriction-modification enzymes. Nature chemical biology *11*, 870-877.

Chapter: 4 DNA-mediated coupling of the ATPase, translocase and nuclease activities of a Type ISP restriction-modification enzyme

4.1 Introduction

Helicases and translocases are the primary motors facilitating nucleic acids transactions in a cell (Singleton et al., 2007). They hydrolyse nucleoside triphosphate and use the chemical energy to perform mechanical work. Helicases unwind double-stranded (ds) DNA and directionally move (translocate) along the unwound single-stranded DNA, while translocases translocate along dsDNA (Singleton et al., 2007). The motor is often part of a multidomain and multifunctional protein and orchestrates its activity with those of the other functional domains to accomplish a specific task. The detailed mechanism of how the different domains are functionally coupled to the motor in many of these molecular machines is not well understood (Fairman-Williams et al., 2010; Singleton et al., 2007). Here, we have studied the ATPase-driven activities of the Type ISP restriction-modification (RM) enzyme to gain insights into the coupling of the different functional domains.

Like many helicases whose ATPase activities are stimulated by DNA having specific structures, ATP hydrolysis by a Type ISP RM enzyme is activated by dsDNA containing its target sequence (Chand et al., 2015; Smith et al., 2009c). The rate of ATP hydrolysis by the Type ISP RM enzyme is significantly higher in the presence of specific DNA than in the presence of nonspecific DNA (Smith et al., 2009c). The regulation of the ATPase activity by the target sequence is observed in the case of the other ATP-dependent RM enzymes, the Type I and Type III RM enzymes, too (Davies et al., 1999; Meisel et al., 1995). The stimulation of the ATPase motors by specific DNA structures or sequences could be a means of avoiding wasteful hydrolysis of ATP in the cell. The structure of the Type ISP RM enzymes LlaBIII and LlaGI bound to a DNA substrate mimic revealed that the target sequence is recognised by the TRD and MTase (Chand et al., 2015; Kulkarni et al., 2016). The MTase is coupled to the ATPase by a coupler domain, and N-terminal to the ATPase is the nuclease domain. Furthermore, the structure revealed that at the target site the DNA is bent by ~39° causing it to steer towards the ATPase.

The ATPase is physically more than 15 bp upstream to the target sequence anchored to the TRD and MTase (Chand et al., 2015). Previously, it has been shown that at least 23 bp upstream to the target sequence is required for LlaBIII to initiate translocation along the DNA (Chand et al., 2015). This prompted us to ask if the binding of the target sequence to the TRD and MTase is sufficient to stimulate the ATPase or, like in the case of translocation initiation, if a DNA long enough to interact with the ATPase is also required? The ATPase and the translocase activity of a helicase-like ATPases are in general tightly coupled. It has been demonstrated in the case of some SF2 helicases that motifs III and V transduce the conformational changes arising from ATP hydrolysis to DNA translocation (Banroques et al., 2010; Christiansen et al., 2003; Fernandez et al., 1997;

Fitzgerald et al., 2017; Gu and Rice, 2010; Papanikou et al., 2004; Pause and Sonenberg, 1992; Smith and Peterson, 2005). Using mutagenesis, we also sought to find if motifs III and V play a role in coupling the two activities in Type ISP RM enzymes.

Translocation of the Type ISP RM enzyme along the DNA is essential for nucleolytic cleavage (van Aelst et al., 2013). The convergence of two such translocating enzymes in a head-to-head direction brings the respective nucleases in proximity to catalyse closely spaced nicks on the two strands of the duplex leading to double-strand DNA break (Chand et al., 2015; Sisakova et al., 2013; Smith et al., 2009b, c). Convergence of the nucleases from two enzymes is a requirement for DNA cleavage by the ATP-dependent Type I (Seidel et al., 2008; Studier and Bandyopadhyay, 1988) and Type III RM enzymes (Meisel et al., 1995), too. It is, however, not clear if the sole purpose of DNA translocation is to bring the nucleases close in space, or if the rate of translocation and, consequently, the force with which the enzymes collide also contributes to the activation of the nucleases. It has been hypothesised that the collision of the converging enzymes alters the conformation of the Type ISP RM enzymes and activates their nuclease (van Aelst et al., 2015). Through the biochemical characterisation of the wild-type (WT) and mutant enzymes designed to study the coupling of the ATPase to the other functional activities, we demonstrate that nucleolytic cleavage not only requires an active translocase but is also dependent on the rate of translocation. The study reported here provides new insights into the role of substrate DNA in coupling the different functional activities and the importance of the ATPase activity for the functioning of a Type ISP RM enzyme.

4.2 Materials and methods

4.2.1 Site directed mutagenesis and DNA substrates

Site-directed mutagenesis was carried out, and the ATPase domain mutants (LlaBIII^{K385A}, LlaBIII$^{\Delta Loop}$, LlaBIII^{PolyALA}, LlaBIII^{T376A} and LlaBIII^{R564A}) were cloned into a pRSF vector under NcoI and XhoI sites using forward LB-pRSF-F and reverse primer LB-pRSF-R alongwith mutation site-specific primers K385A-R, ΔLoop-2G-R, ΔLoop-2G-F, PolyALA-R, T376A-R and R564A-R, respectively. The sequences of all the primers used are listed in Table 4.1. The gene sequences of the mutants were confirmed by DNA sequencing. A 1085 bp long DNA substrate having two target sites LlaBIII in head-to-head orientation was generated by PCR using LlaBIII gene as a template with forward primer LB-1085-F and reverse primer LB-1085-R. 1758 bp DNA substrate for cooperation assay was amplified in a similar manner with forward primer LJ1HISF

and reverse primer LJ1HISR1. All the DNA oligomers used as primers and as substrates for the ATPase assays were purchased from Sigma/Integrated DNA Technologies.

Table 4-1: DNA oligos used for the study.

Primer	Sequence
LB-pRSF-F	ACTTTAATAAGGAGATATACCATGGTGGCATTTTGGAAGGAATG
LB-pRSF-R	CGCAGCAGCGGTTTCTTTACCAGACTCGAGTTATAGTCCCTGTACTACTCTTG
K385A-R	GATACTCTTGTCTTTGGCATTTTTTGCCGCACTTTCCCCATATATCTTTGG
ΔLoop-2G-F	CACCAAAGATATATGGGGAAGGTGGTAAGAGTATCTTAC
ΔLoop-2G-R	GATGAAAGTAAGATACTCTTACCACCTTCCCCATATATC
PolyALA-R	GATACTCTTGTCTGCGGCAGCTGCTGCCGCAGCTTCCCCATATATC
T376A-R	CTTTCCCCATATATCTTTGGTGCAGCCGTTTGGTACATTC
R564A-R	CGATTCCTTCCGTTAAGAAGGCGACATTAGAAACAATTCG
LB-1074-F	GCAACGGACGCTCGCTGATCCAG
LB-1074-R	CCATCGCTTGGGAGACGGGGTTTG
LJ1HISF	GAAGGAGATATACATATGGGTAAAATCGTCCTGCC
LJ1HISR1	GATGATGATGATGGGATCCTTATTCTTCCGTGGAC
Oligo23-F	TTAGCTAATAGACTGAGCCGAGG
Oligo23-R	TCCTCGGCTCAGTCTATTAGCTA
Oligo28-F	GCTCTAGCTAATAGACTGAGCCGAGGTG
Oligo28-R	CACCTCGGCTCAGTCTATTAGCTAGAGC
Oligo32-F	GCCTGCTCTAGCTAATAGACTGAGCCGAGGTG
Oligo32-R	CACCTCGGCTCAGTCTATTAGCTAGAGCAGGC
Oligo33-F	GACCTGCTCTAGCTAATAGACTGAGCCGAGGTG
Oligo33-R	CACCTCGGCTCAGTCTATTAGCTAGAGCAGGTC
Oligo34-F	GTACCTGCTCTAGCTAATAGACTGAGCCGAGGTG
Oligo34-R	CACCTCGGCTCAGTCTATTAGCTAGAGCAGGTAC
Oligo35-F	GTCACCTGCTCTAGCTAATAGACTGAGCCGAGGTG
Oligo35-R	CACCTCGGCTCAGTCTATTAGCTAGAGCAGGTGAC
Oligo46-F	GTCTTATGCAGGTCACCTGCTCTAGCTAATAGACTGAGCCGAGGTG
Oligo46-R	CACCTCGGCTCAGTCTATTAGCTAGAGCAGGTGACCTGCATAAGAC
TFO	TTCTTTTCTTTCTTCTTTCTTT
NSP40-F	GTACTCAGCAGTATCCTGTATGCTACGTATTGCTATCGTG
NSP40-R	CACGATAGCAATACGTAGCATACAGGATACTGCTGAGTAC

4.2.2 Purification of LlaBIII

LlaBIII and its mutants were purified as described earlier in chapter 2 (Chand et al., 2015).

4.2.3 DNA cleavage assay

A 1085 bp long DNA with two LlaBIII target sites in head-to-head orientation separated by 155 bp was used as the substrate for nucleolytic DNA cleavage assay. The enzyme (WT LlaBIII or mutants) and DNA were mixed in TMDK buffer (50 mM Tris-Cl pH 8.0, 10 mM $MgCl_2$, 1 mM DTT, 150 mM KCl) on the ice and the reaction was started by adding 4 mM ATP (van Aelst et al., 2013). The reaction mix was incubated for 30 minutes at 25°C. Reactions were stopped by addition of half the volume of STEB (0.1 M Tris-Cl pH 7.5, 0.2 M EDTA, 40% w/v sucrose, 0.4 mg/ml bromophenol blue) and analysed on 1% agarose gel.

LlaBIII and LlaGI can cooperate and cleave the substrate DNA containing one LlaBIII and one LlaGI site (van Aelst et al., 2013). For cooperation assay, a 1758 bp long DNA containing a LlaBIII site in a head-to-head orientation with a LlaGI site was used. 500 nM LlaBIII or its mutant and LlaGI were incubated with 10 nM DNA for 5 minutes in TMDK buffer at 4°C, and the reaction was started by addition of 4 mM ATP. The reaction mixes were incubated for 30 minutes at 25°C and were stopped by adding 1% SDS and 70 mM EDTA. The samples were treated with proteinase K for 1 hour at 37°C and analysed on 1% agarose gel.

4.2.4 Circular dichroism spectroscopy and nanoDSF

Circular dichroism (CD) spectra of LlaBIII and its mutants were recorded at 25°C using Jasco 815 spectrometer. 200 µl of 0.06 mg/ml protein samples in 20 mM phosphate buffer pH 8.0 with 100 mM KCl was added to 2 mm quartz cuvette and absorption spectra of the circularly polarised light were measured from 195 nm to 250 nm wavelength. Data was acquired at bandwidth 1 nm and scanning speed 100 nm/minute. Sensitivity was kept high with 5 accumulations and data integration time (DIT) as 1 second. For accuracy, protein concentration was estimated by both Bradford assay and absorption measurement at 280 nm.

Temperature-dependent protein unfolding assay was carried out using NanoDSF (Prometheus NT.48, Nanotemper Technologies). 5 µl protein solution (1 mg/ml protein in buffer containing 50 mM Tris-Cl pH 8.0, 1 mM DTT and 100 mM KCl) was loaded in a capillary and the sample was heated from 20°C to 95°C with a rate of 1°C/minute. Fluorescence intensity of protein was measured at 330 nm and 350 nm. The ratio of the fluorescence intensity at 330 nm and 350 nm, i.e. F350/F330, was plotted against the temperature gradient. The first derivative of F350/F330 was used to deduce the melting temperature.

4.2.5 DNA binding assay

Binding assays were performed with 28 bp specific/non-specific DNA in binding buffer (50 mM Tris pH 8.0, 10 mM MgCl$_2$, 1 mM DTT, 150 mM KCl). 250 nM DNA was allowed to bind with varying concentrations of LlaBIII or LlaBIII$^{\Delta Loop}$ (0, 10, 50, 100, 250, 500, 1000, 2500, 5000 nM). Protein and DNA were mixed and incubated on ice for 10 minutes. Reactions were stopped by adding half the volume of STB (0.1 M Tris-Cl pH 7.5, 0.2 M EDTA, 40% w/v sucrose, 0.4 mg/ml bromophenol blue) and loaded on 6% native PAGE run at 4°C. The gel was stained with ethidium bromide and imaged on Typhoon TRIO+ variable mode imager (GE Healthcare).

4.2.6 ATPase assay

NADH coupled ATPase assay was performed to compare the activity of LlaBIII and its mutants. Fluorescence intensities were measured using Varioskan Flash (Thermo Scientific). A 46 bp long DNA having one LlaBIII target site was used as a specific substrate and a 40 bp long DNA with no target sequence was used as nonspecific DNA. 50 nM of WT or mutant enzyme was incubated with 500 nM of specific or non-specific DNA for 5 minutes. ATP was added, and the reaction was shaken for 30 seconds before the measurements were taken. Absorbance at 340 nM was recorded every 10 seconds for 800 seconds at 25°C. ADP standard was performed for each set of experiment. The concentration of ATP hydrolyzed was calculated at each time interval using a line equation Y=mX + C (where Y=absorbance, m=slope, X=concentration of ADP produced or ATP hydrolyzed and C=intercept on Y-axis) obtained from the standard plot with different ADP concentrations.

Kinetics parameters- V_{max}, K_m and k_{cat} were obtained by ATP concentration-dependent assay with ATP concentrations of 0.05 μM, 0.25 μM, 1 μM, 4 μM, 16 μM, 64 μM, 128 μM, 256 μM, 512 μM, 1024 μM and 2048 μM. The data with ATP hydrolysis rates plotted against different ATP concentration were fit to Michaelis-Menten equation using GraphPad PRISM 5.

4.2.7 Triplex DNA and Triplex Displacement

Triplex forming oligonucleotide (TFO) was radiolabelled with T4 Polynucleotide Kinase (New England Biolabs *Inc.*) and ATPγ^{32}P at 37°C for 30 minutes. Labelled TFO was purified by mini-quick spin oligo (Roche, Germany) column and kept at -30°C till further use. 50 nM pONE plasmid with a single site for LlaBIII (Sisakova et al., 2013) was linearised with HindIII. 25 nM of TFO mixed with the linearised plasmid in MM Buffer (10 mM MES pH 5.5 and 25 mM MgCl$_2$) and incubated at 57°C for 15 minutes. The reaction mix was further incubated at 20°C overnight to

allow the triplex DNA formation (Firman and Szczelkun, 2000). Triplex DNA was stored at -20°C until further use.

50 nM enzyme and 0.25 nM triplex DNA was used in TMDK buffer (50 mM Tris-Cl pH 8.0, 10 mM MgCl$_2$, 1 mM DTT and 150 mM KCl). The enzyme was incubated with triplex DNA in TMDK buffer for 5 minutes to allow binding to the target site. The reaction was initiated by addition of 4 mM ATP and incubated for 10 minutes at 20°C. The reaction mixture also contained unlabelled TFO to prevent non-specific binding of displaced TFO to LlaBIII. In the case of time-dependent ATPase assay, the incubation time of the reaction mixture was varied. Half the volume of GSMB (15% w/v glucose, 3% w/v SDS, 250 mM MOPS pH 5.5 and 0.4 mg/ml bromophenol blue) was added to stop the reaction and analyzed on 6% Native PAGE (acrylamide:bisacrylamide 29:1) prepared in 40 mM Tris-acetate, 5 mM sodium-acetate, 5 mM MgCl$_2$, 0.1 mM EDTA pH 5.5 at 4°C. The gel was dried in Gel dryer model 583 attached with HydroTech^{TM} Vaccum Pump supplied by BioRad at 80°C for 1 ½ hour. The dried gels were exposed to phosphorimager plate (GE Healthcare) overnight. Gels were imaged on Typhoon TRIO+ variable mode imager (GE Healthcare), and the intensities of the bands were quantified using ImageJ software. All the graphs were plotted in GraphPad PRISM 5. The intensities for the enzyme and triplex DNA complex were considered as 100%, and the displacement was calculated for the lanes where ATP was added. In the time-dependent experiment, the intensities for zero time point was considered as 100% and the intensities for all time points were normalised.

4.3 Results

4.3.1 ATPase activation is a two-stage process dependent on DNA length

To find if DNA target recognition is sufficient to stimulate the LlaBIII ATPase, we measured the ATPase activity of the enzyme in the presence of a 22 bp DNA having the target sequence. As defined previously (Chand et al., 2015), each base pair (bp) in the substrate DNA is numbered such that the target adenine for methylation and its complementary base is +1, the base pairs upstream (towards the ATPase domain) of the target sequence have a negative numbering, and those downstream have a positive numbering (Figure 4-1A, B). The 22 bp DNA had 4 bp (+10) downstream of the target sequence and 12 bp (-12) upstream. This DNA was not long enough to interact with the ATPase. In comparison to a non-specific DNA, the 22 bp DNA stimulated the ATPase activity by ~2 folds (Figure 4-1C).

Increasing the length upstream of the target site in the direction of the ATPase (Figure 4-1B) from -12 to -20 did not affect the ATPase activity (Figure 4-1C). However, an upstream length longer

than 20 bp significantly increased the ATPase activity (Figure 4-1C). The clustering of the ATPase activity into two distinct sets with one having a higher ATP hydrolysis rate than the other beyond a threshold length suggested that there are two stages in achieving complete activation of the ATPase. The first stage involves the binding of the target sequence to the MTase and TRD. This stage does not require the DNA to be long enough to engage with the ATPase motor (Figure 4-1B). The second stage requires a DNA long enough (> -20) to interact with the ATPase (Figure 4-1B).

This is consistent with our previous observation that DNA translocation by LlaBIII, which is driven by ATP hydrolysis, is dependent on the upstream length of the specific DNA (Chand et al., 2015). Interestingly, though, activation of the ATPase requires at least 21 bp upstream of the target sequence (Figure 4-1C), translocation initiation requires at least 23 bp (Chand et al., 2015). This implies that full activation of the ATPase is not sufficient for DNA translocation to initiate, and the latter requires an additional 2 bp upstream.
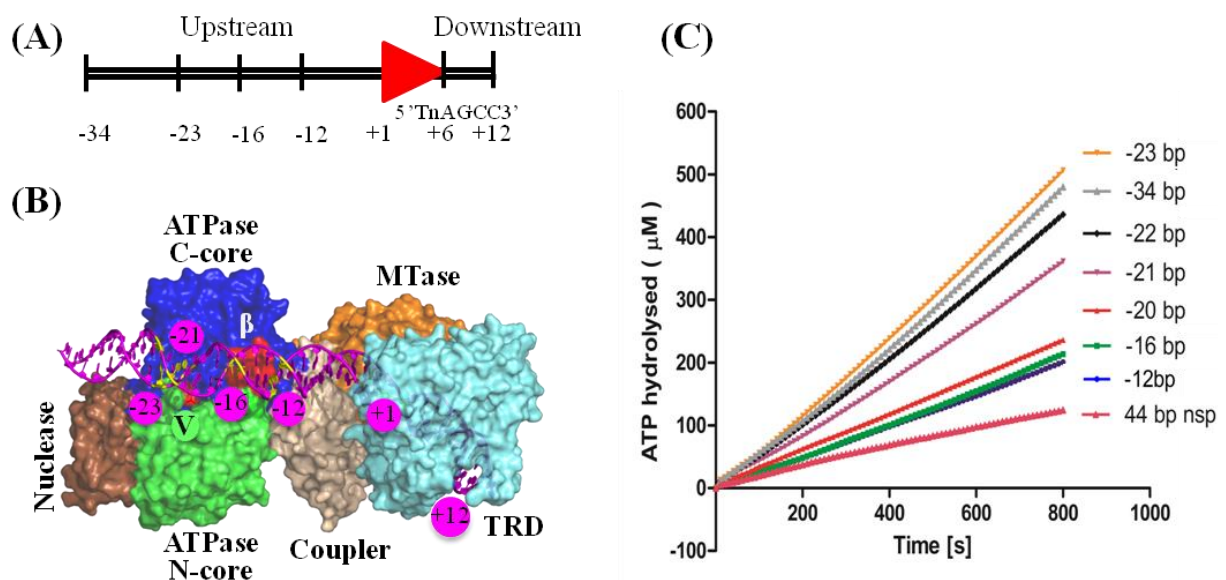
Figure 4-1: DNA length-dependent stimulation of ATPase activity. **(A)** The scheme used for numbering the base pairs in a DNA substrate. The red arrowhead represents the target sequence. The target sequence of LlaBIII is written below the arrowhead. **(B)** Model of LlaBIII bound to DNA (magenta). The yellow base pairs show the position of the upstream end of some of the substrates used for the study. The β-hairpin loop and Arg564 of motif V are coloured in red (see text for details). **(C)** The rate of ATP hydrolysis by WT LlaBIII in the presence of specific DNA with varying length upstream of the target sequence. The plot also shows the amount of ATP hydrolyzed in the presence of non-specific (nsp) DNA. The assay was performed with 50 nM LlaBIII, 500 nM DNA and 1 mM ATP. Shown are the averages of three replicates.

4.3.2 β-hairpin loop of the ATPase N-core is essential for nucleolytic activity

The analysis in the previous section reveals that full activation of the ATPase requires not only DNA target recognition but also engagement of the DNA with the ATPase domain. The first
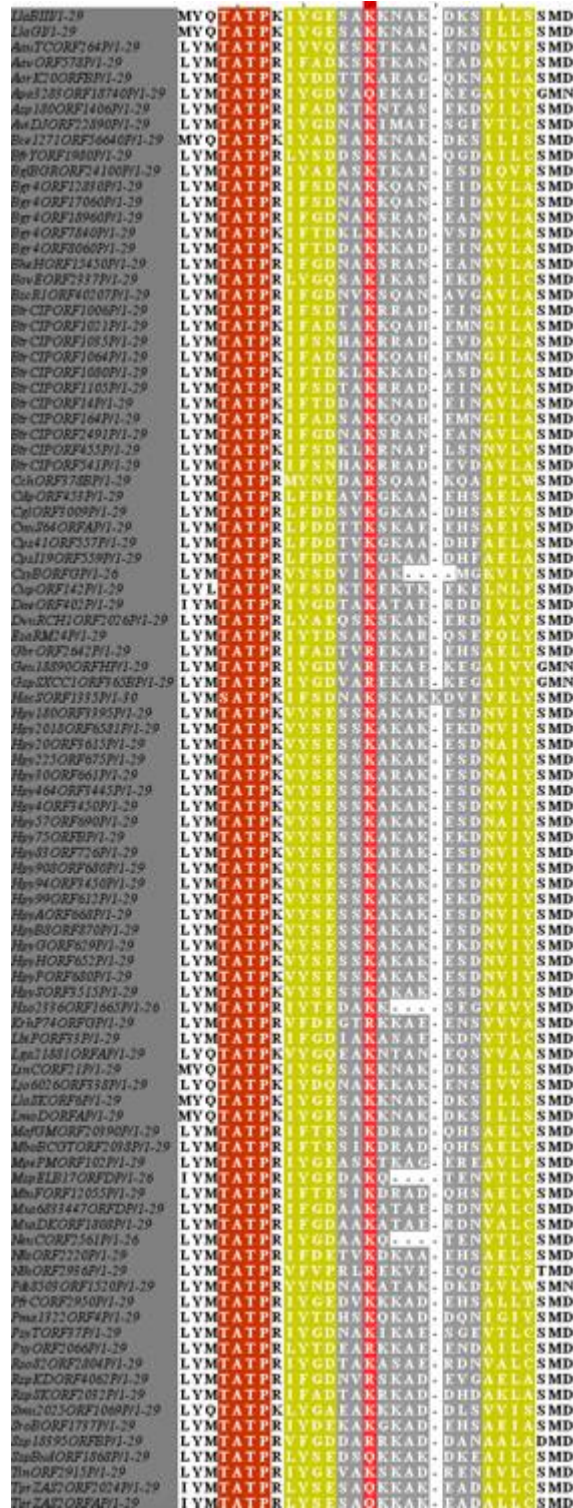
Figure 4-2: Amino acid sequence alignment of the β-hairpin loop region in Type ISP RM enzymes. The conserved lysine highlighted in red corresponds to LlaBIII-Lys385.

structural element of the ATPase that the traversing DNA encounters is a β-hairpin loop in its N-core (Figure 4-1B) (Chand et al., 2015; Kulkarni et al., 2016). This hairpin loop interacts with the phosphodiester backbone of the DNA through bonds formed by the main chain amino group of K389 and D390. We had earlier proposed that this region might be involved in paddling the DNA

during translocation (Chand et al., 2015). To find if the loop was important for LlaBIII activities, we mutated the loop. We made three sets of mutants – LlaBIII[K385A] in which a highly conserved lysine at position 385 (Figure 4-2) was mutated to alanine; LlaBIII[PolyALA] in which all the loop residues from 383 to 389 were mutated to alanine; and LlaBIII[ΔLoop] in which residues 383-390 were replaced with 2 glycines. Nucleolytic assay revealed that LlaBIII[ΔLoop] failed to cleave DNA (Figure 4-3A). In contrast, LlaBIII[K385A] and LlaBIII[PolyALA] had nucleolytic activities comparable to that of the WT enzyme (Figure 4-3A). Comparison of the secondary and tertiary structural profile of the WT and the mutant enzymes using circular dichroism and nanoDSF indicated that the mutations did not affect the structural integrity of LlaBIII (Figure 4-3B, 3C).
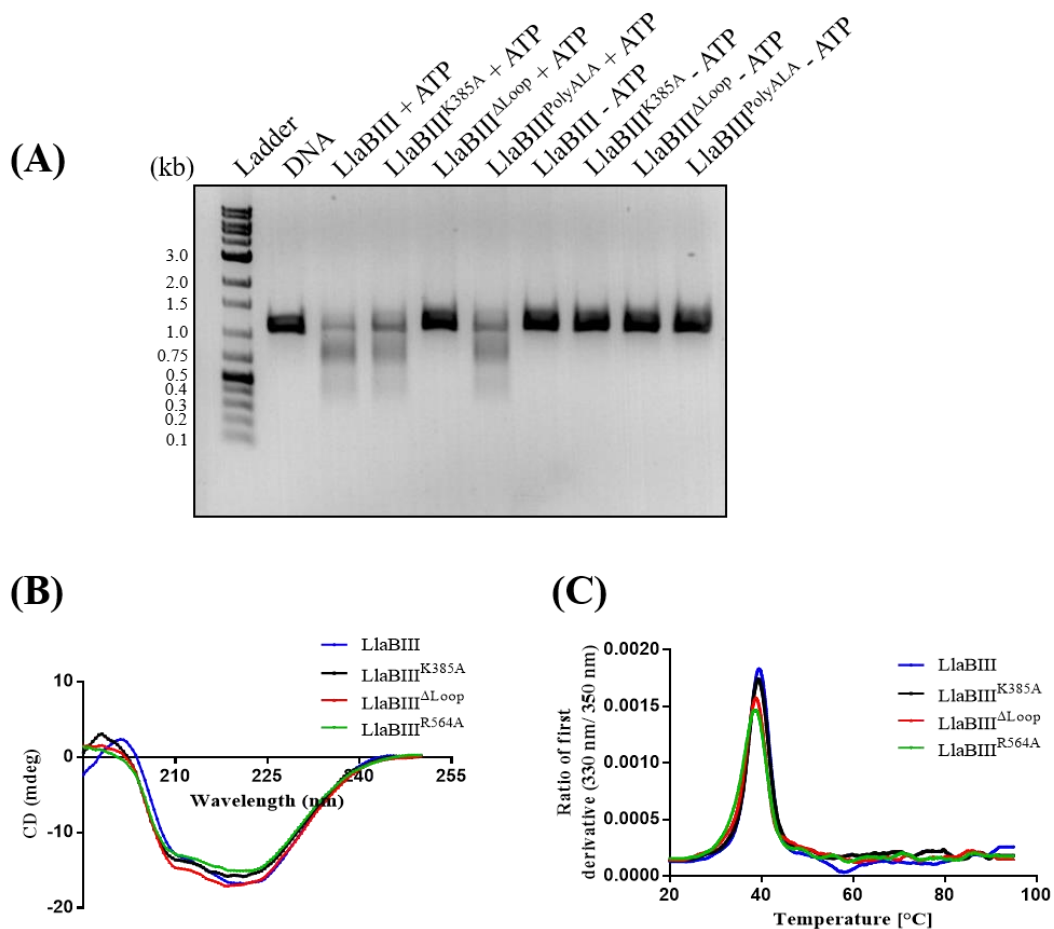
Figure 4-3: DNA cleavage and structural conservation of β-hairpin loop mutants of LlaBIII. (A) Results of the DNA cleavage assay showing the nucleolytic activity of LlaBIII and its β-hairpin loop mutants (LlaBIII[K385A], LlaBIII[ΔLoop] and LlaBIII[PolyALA]) on 1085 bp DNA with head-to-head oriented sites. (B) Circular dichroism and (C) NanoDSF showing the conservation of secondary and tertiary structures of LlaBIII and its mutants.

To find why LlaBIII[ΔLoop] failed to cleave DNA, we tested the affinity of LlaBIII[ΔLoop] for specific DNA. The mutant and the WT enzyme bound to DNA equally well and could discriminate between specific and non-specific DNA (Figure 4-4A-D). Increasing the concentration of LlaBIII[ΔLoop] also failed to elicit DNA cleavage (Figure 4-5). Next, we compared the ATPase
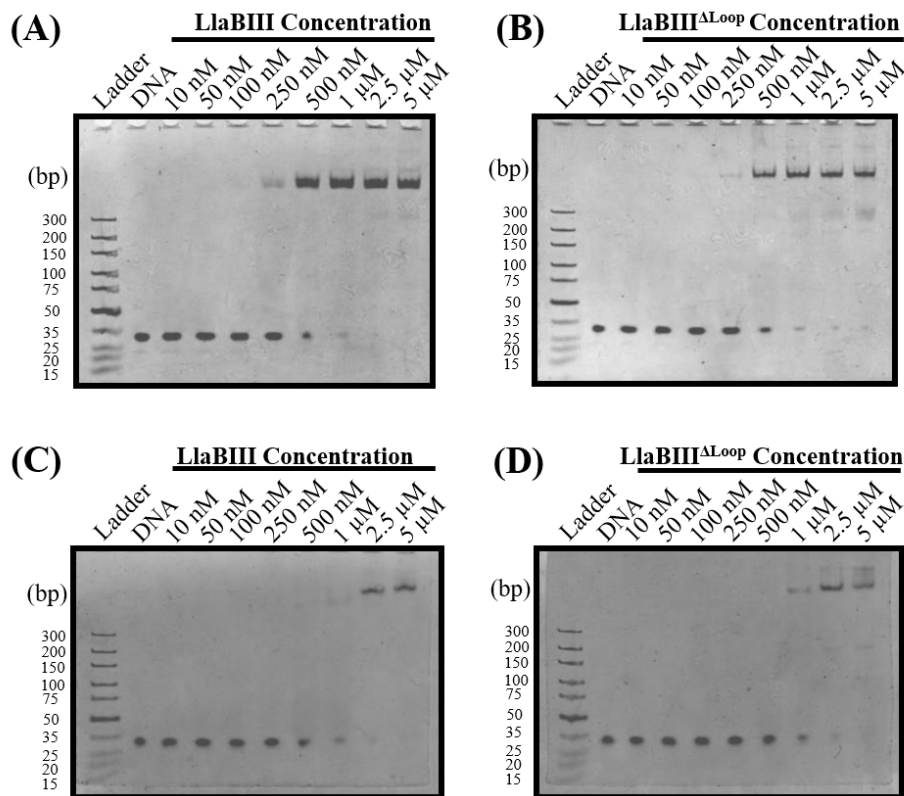
Figure 4-4: DNA binding studies of LlaBIII$^{\Delta Loop}$ mutant. Electrophoretic mobility shift assays comparing the DNA binding affinities of WT LlaBIII and LlaBIII$^{\Delta Loop}$ for **(A, B)** a specific 28 bp DNA. **(C, D)** a non-specific 28 bp DNA. 250 nM DNA was used for the binding studies at 4°C.

activities of WT LlaBIII and LlaBIII$^{\Delta Loop}$. The ATPase activities of LlaBIII^{K385A} and LlaBIII^{PolyALA}, which have nucleolytic activities comparable to the WT enzyme, were also measured. The rate of ATP hydrolysis was measured in the presence of either a specific or a non-specific DNA. The ATPase activities of WT and the mutant enzymes were significantly higher in the presence of specific DNA than non-specific DNA (Figure 4-6A-D), implying that, despite the mutations of the β-hairpin loop, the ATPase activity was stimulated by specific DNA. Though specific DNA stimulated the ATPase activities of the four enzymes, they showed different rates of ATP hydrolysis (Figure 4-6A-D). WT and LlaBIII^{PolyALA} had similar ATPase activities. However, LlaBIII^{K385A} and LlaBIII$^{\Delta Loop}$ had relatively lower ATPase activities.

LlaBIII-Lys385 is close enough to interact with the DNA *via* its side chain (Chand et al., 2015), and, as a consequence, its mutation, modulates the stimulation of the ATPase by the substrate DNA. Interestingly, replacement of all the loop residues with alanine (LlaBIII^{PolyALA}) overcame the decrease in the ATPase due to K385A mutation (Figure 4-6B, 6C). As the nucleolytic activity of LlaBIII^{K385A} was found to be comparable to the WT enzyme, we think that the lower ATPase activity of LlaBIII^{K385A} is sufficient for efficient DNA cleavage to occur. Also, both LlaBIII^{K385A} and LlaBIII$^{\Delta Loop}$ displayed the DNA length-dependent two-stage ATPase activation implying that the mutations did not affect the mode of activation (Figure 4-6B, 6D). As LlaBIII^{K385A} and
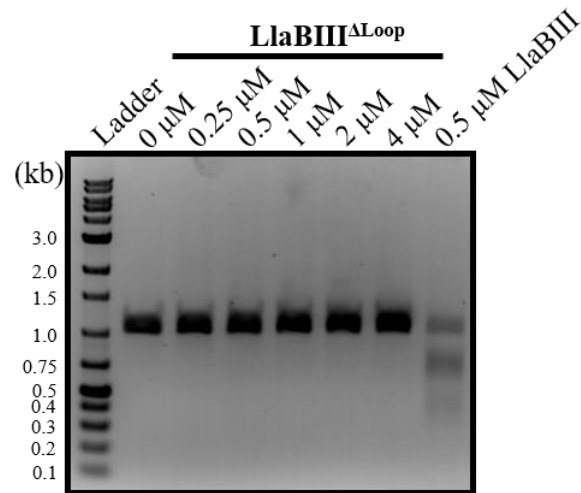
Figure 4-5: Concentration-dependent DNA cleavage assay of LlaBIII$^{\Delta Loop}$ mutant. LlaBIII$^{\Delta Loop}$ was unable to cleave substrate DNA even at a concentration of 4 µM. The last Lane shows DNA cleavage by LlaBIII wild type at 500 nM concentration. 10 nM DNA was mixed with varying concentrations of the enzyme and reaction was started with the addition of 4 mM ATP at 25ºC.

LlaBIII^{PolyALA} were nucleolytically active while LlaBIII$^{\Delta Loop}$ was inactive, we concluded that the side chains of the residues constituting the β-hairpin loop were not important for the activity, while the interactions made by the main chains atoms of residues 389 and 390 with the DNA were important.
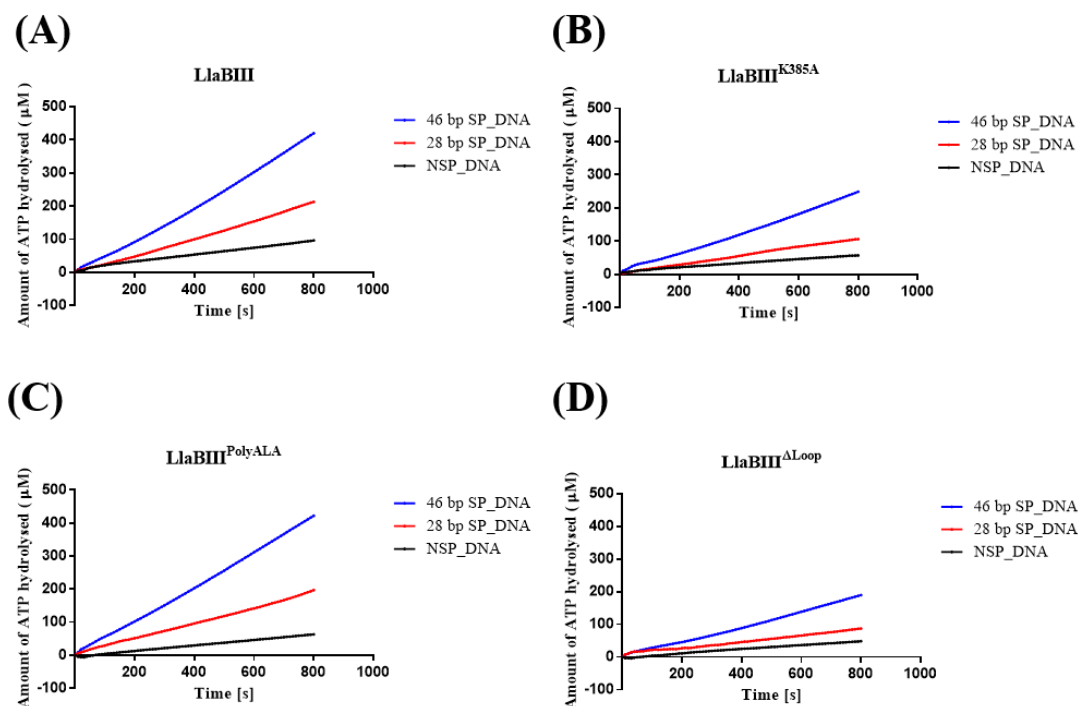
Figure 4-6: NADH coupled ATPase assay of β-hairpin loop mutants of LlaBIII. Comparison of the ATPase activities of **(A)** LlaBIII, **(B)** LlaBIII^{K385A}, **(C)** LlaBIII^{PolyALA}, and **(D)** LlaBIII$^{\Delta Loop}$ in the presence of 46 bp and 28 bp specific DNA (SP_DNA) and a 44 bp non-specific DNA (NSP_DNA). 50 nM LlaBIII or its mutants, 500 nM DNA and 1 mM ATP were used for the NADH coupled ATPase assay carried out at 25ºC. Shown are the averages of three replicates.

4.3.3 Deletion of the β-hairpin loop affects DNA translocation

The biochemical analysis of LlaBIII$^{\Delta Loop}$ demonstrated that the mutant enzyme is proficient in DNA binding and ATP hydrolysis, but lacks nucleolytic activity. Nucleolytic cleavage by LlaBIII
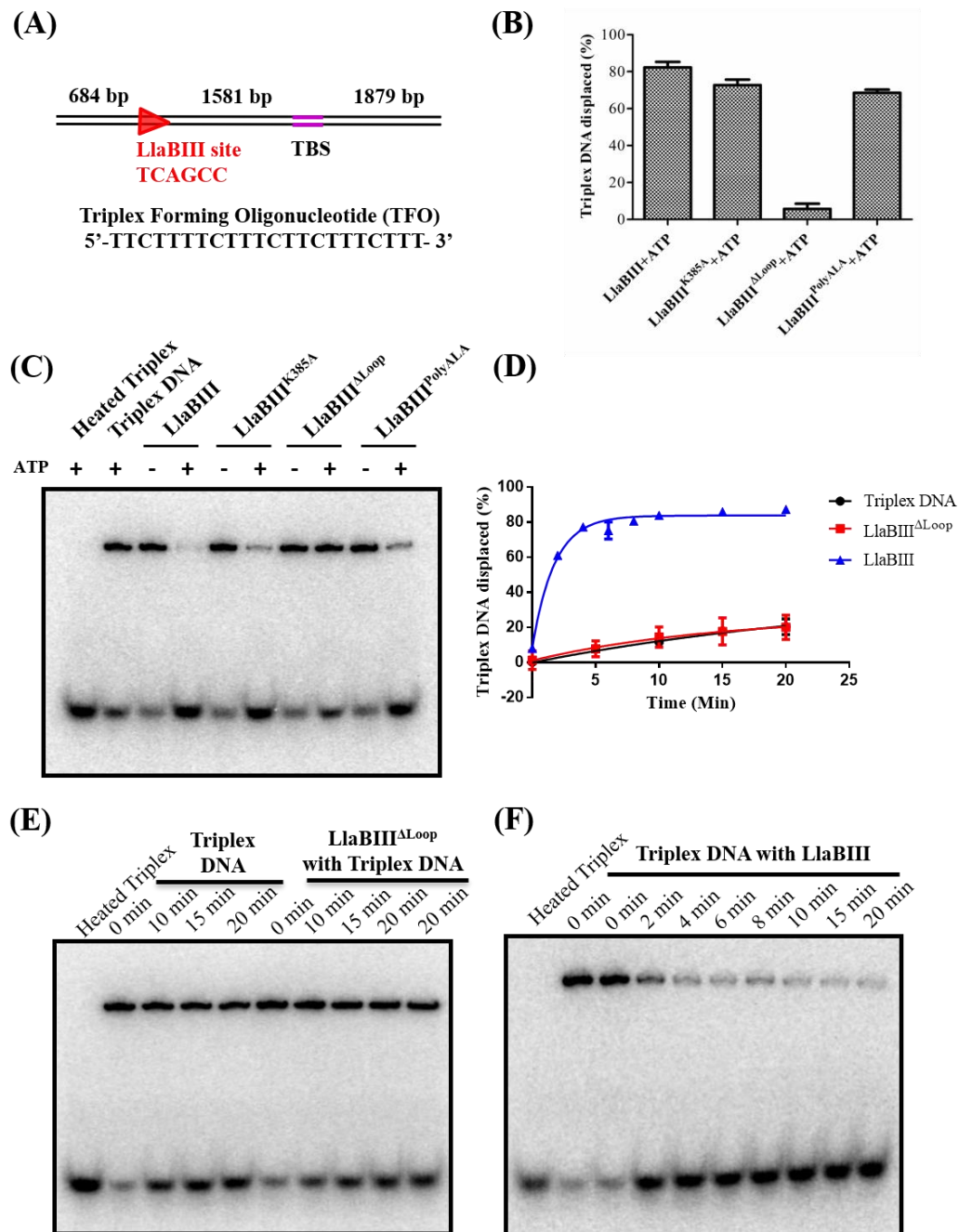
Figure 4-7: DNA translocation assay of β-hairpin loop mutants of LlaBIII. **(A)** The triplex DNA substrate used for DNA translocation assay. Red arrow shows the LlaBIII target site and direction of translocation. TBS is the TFO binding site. **(B)** Bar graph showing ATP-dependent triplex displacement by LlaBIII, LlaBIII^{K385A}, LlaBIII^{PolyALA} and LlaBIII$^{\Delta Loop}$ in 10 minutes at 20ºC. 4 mM ATP was used in the reaction. Shown are the averages of three replicates. **(C)** A representative gel of the triplex displacement assay showing the displacement of TFO by β-hairpin loop mutants in 10 minutes. **(D)** Time dependent triplex displacement assay showing that LlaBIII$^{\Delta Loop}$ is an inactive translocase. Shown are the averages of three replicates. **(E, F)** A representative gel of the triplex displacement assay showing time dependent displacement of TFO by **(E)** Triplex DNA control and LlaBIII$^{\Delta Loop}$ **(F)** LlaBIII.

requires two translocating enzymes to converge on the DNA. Hence, we sought to find if LlaBIII$^{\Delta Loop}$ is active as a translocase. Towards this, we used a triplex displacement assay that has previously been used to study the translocation activity of LlaBIII (Sisakova et al., 2013). The assay used a ~4.2 kb long linear DNA having a single LlaBIII target site located 1581 bp upstream of a triplex binding site, which was bound to a 22 nt long oligo to form the triplex (Figure 4-7A). As expected, the WT enzyme was an active translocase, which displaced almost ~83% of the TFO from the substrate DNA in 10 minutes.

LlaBIII^{K385A} and LlaBIII^{PolyALA}, which are active nucleases, displaced ~73% and ~69% of the triplex oligo. Interestingly, the decreased ATPase activity of LlaBIII^{K385A} did not affect its translocase activity. In contrast, LlaBIII$^{\Delta Loop}$ displaced a mere ~6% of the triplex oligo bound to the substrate DNA, which was comparable to the displacement noted in control experiments without protein (Figure 4-7B, 7C). Time-dependent triplex displacement assay showed that LlaBIII displaces ~61% and 83% of triplex DNA in 2 and 4 minutes respectively whereas LlaBIII$^{\Delta Loop}$ did not displace the triplex even after 20 minutes of incubation (Figure 4-7D-F). This study clearly demonstrated that the β-hairpin loop is important for DNA translocation by LlaBIII, and deletion of the loop affected the nuclease activity because of deficiency in translocation. The role of the β-hairpin loop in translocation is consistent with the earlier proposal that this structural element might be involved in paddling the DNA (Chand et al., 2015). As LlaBIII$^{\Delta Loop}$ is an active ATPase but an inactive translocase, we concluded that the β-hairpin loop contributes to coupling the two activities.

4.3.4 Effect of mutation of SF2 helicase motifs III and V on LlaBIII activities

In SF2 helicase-like ATPases, motifs III and V are attributed to couple ATPase and translocase activities (Banroques et al., 2010; Christiansen et al., 2003; Fernandez et al., 1997; Fitzgerald et al., 2017; Gu and Rice, 2010; Papanikou et al., 2004; Pause and Sonenberg, 1992; Smith and Peterson, 2005). In the previous section, we demonstrated that the β-hairpin loop in LlaBIII couples the two activities. A structural model of LlaBIII bound to longer DNA suggested that the presence of the β-hairpin loop hides the proximal motif III from interacting with the DNA, while motif V was positioned to interact with the DNA. In SF2 ATPases where the role of motif III in substrate binding and/or coupling of the ATPase to the translocase activity has been demonstrated, the substrate is found to interact with motif III (Fitzgerald et al., 2017). This led us to compare the significance of motifs III and V for the LlaBIII activities.

We mutated Thr376 and Arg564, the most conserved residues of motifs III and V, respectively, to alanine. LlaBIII^{T376A} cleaved the substrate DNA, but in comparison to the WT enzyme, the activity was noticeably weak (Figure 4-8A). In contrast, the motif V mutant LlaBIII^{R564A} did not show any
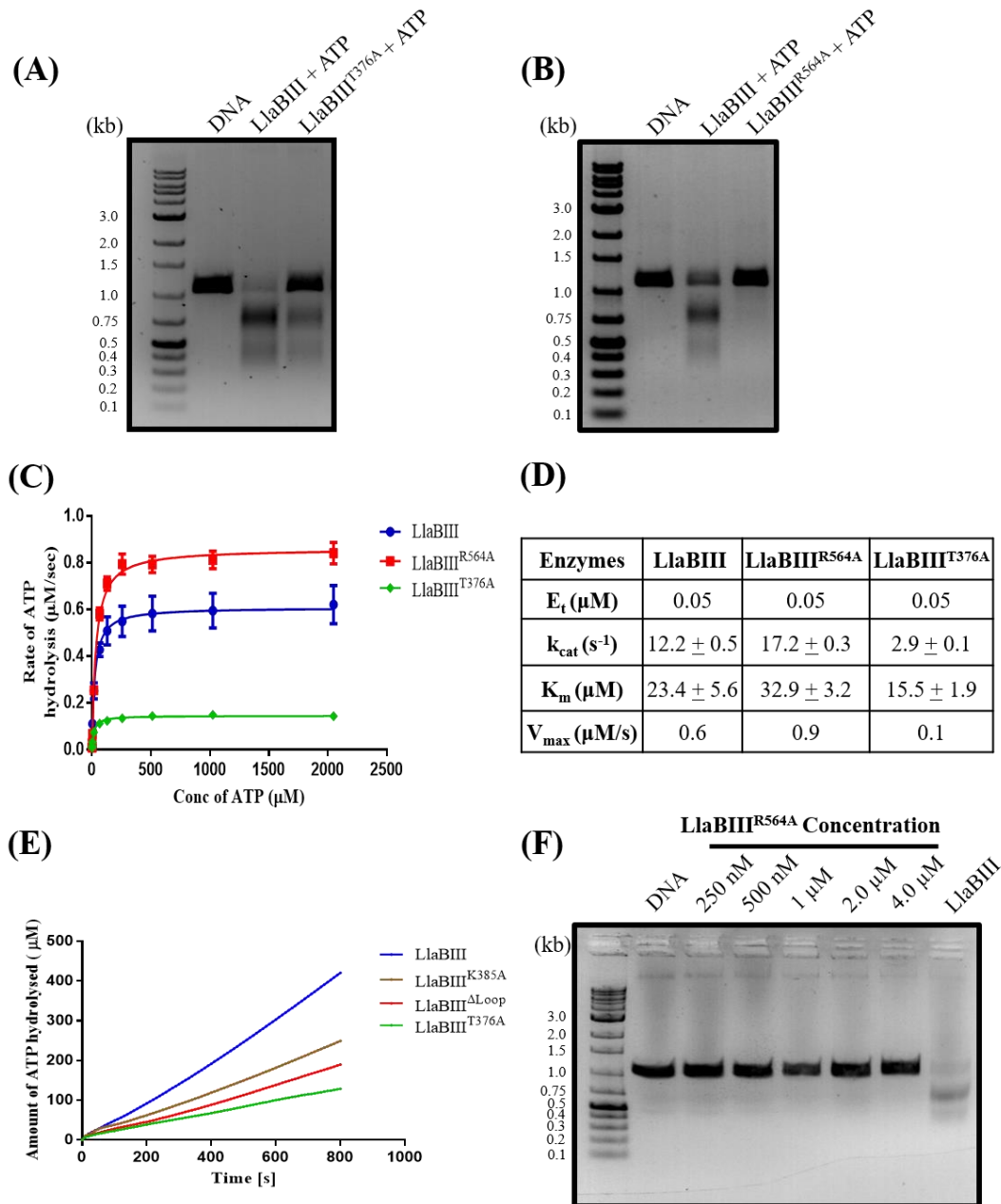
Figure 4-8: Biochemical characterisation of motifs III and V of LlaBIII ATPase domain. (A, B) DNA cleavage by LlaBIII^{T376A} and LlaBIII^{R564A}. In comparison to WT LlaBIII, LlaBIII^{T376A} shows less DNA cleavage activity while the motif V mutant LlaBIII^{R564A} is nucleolytically inactive. **(C)** The kinetics of ATP hydrolysis by WT LlaBIII (blue), LlaBIII^{T376A} (green) and LlaBIII^{R564A} (red). 50 nM LlaBIII or the mutants with 500 nM DNA concentration were used in NADH coupled ATPase assay. The ATP concentrations used were 0.05, 0.25, 1, 4, 16, 64, 128, 256, 512, 1024 and 2048 µM. Reactions were carried out at 25°C. Shown are the averages of three replicates. **(D)** Kinetic parameters for ATP hydrolysis obtained from Figure 4-8C. **(E)** Comparison of the ATPase activity of LlaBIII^{T376A} with LlaBIII^{K385A} and LlaBIII$^{\Delta Loop}$ in the presence of 46 bp specific DNA and 1 mM ATP carried out at 25°C. **(F)** Concentration-dependent DNA cleavage assay for LlaBIII^{R564A} shows that this mutant is unable to cleave substrate DNA even at a concentration of 4 µM. The last lane shows DNA cleavage by LlaBIII wild type at 500 nM concentration. 4 mM ATP was used for DNA cleavage reaction.

nucleolytic activity (Figure 4-8B). We proceeded to measure the ATPase activity of LlaBIII^{T376A}, which was found to be significantly lower than WT LlaBIII, and slightly lower than LlaBIII^{K385A} and LlaBIII$^{\Delta Loop}$ (Figure 4-8C-E). LlaBIII^{R564A}, on the other hand, displayed ATPase activity comparable to the WT enzyme (Figure 4-8C, 8D). Despite this, LlaBIII^{R564A} could not cleave the substrate DNA even at higher enzyme concentrations (Figure 4-8F).

(A)

(B)

(C)
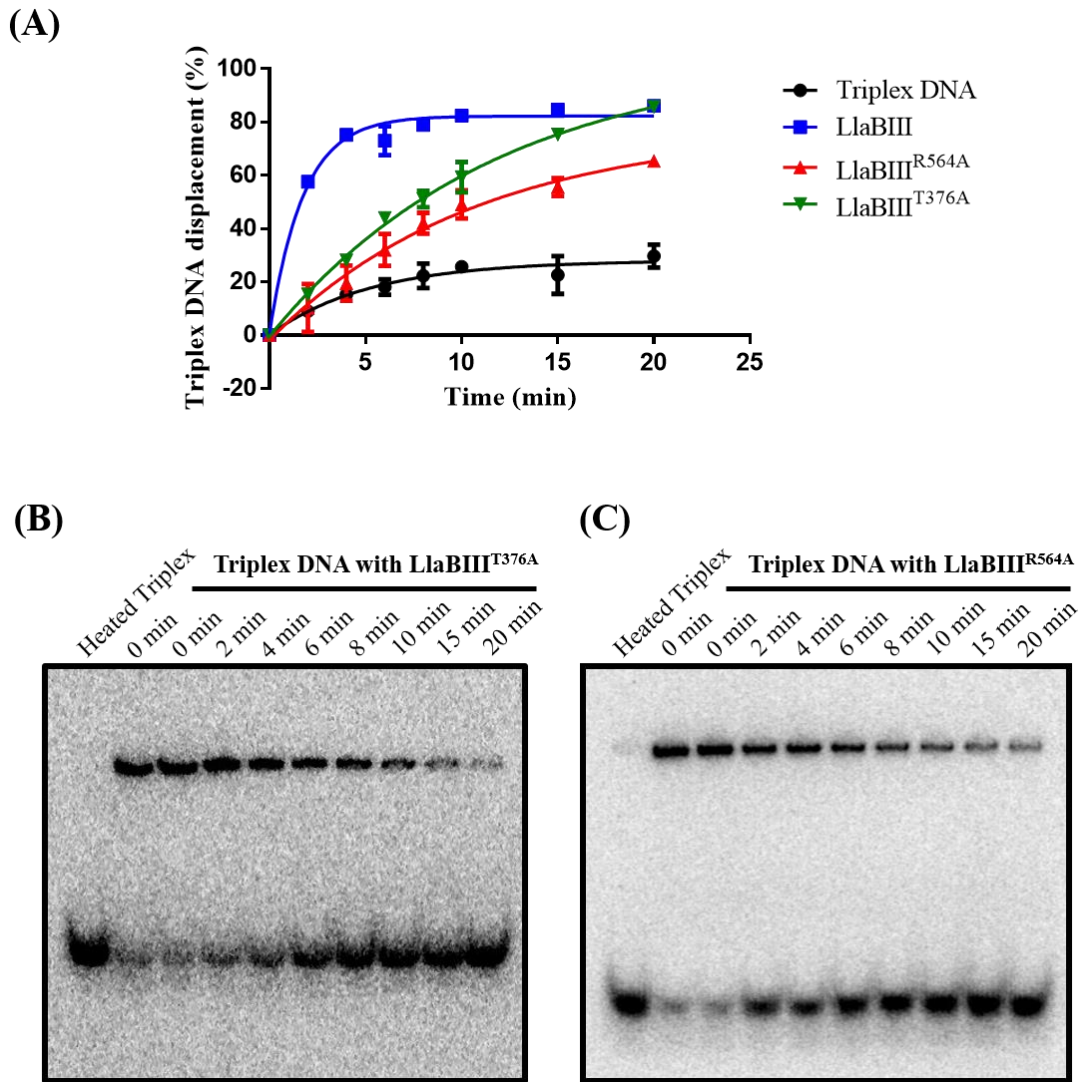
Figure 4-9: DNA translocation assay of motifs III (LlaBIII^{T376A}) and V (LlaBIII^{R564A}) mutants of LlaBIII ATPase domain. **(A)** Triplex displacement assay showing the percentage of triplex DNA displaced by LlaBIII (blue), LlaBIII^{T376A} (green) and LlaBIII^{R564A} (red) over different time points (0, 2, 4, 6, 8, 10, 15 and 20 minutes) at 20°C. The black line is triplex DNA control showing the dissociation of TFO from the DNA in the absence of protein over the same time points. The reaction was performed with 4 mM ATP concentration. Shown are the averages of three replicates. **(B, C)** A representative gel of the triplex displacement assay showing time-dependent displacement of TFO by **(B)** LlaBIII^{T376A} and **(C)** LlaBIII^{R564A}.

We proceeded to compare the translocase activities of LlaBIII^{T376A} and LlaBIII^{R564A} with WT LlaBIII using triplex displacement assay as described above. In comparison to WT LlaBIII, LlaBIII^{R564A} took longer time to displace the triplex than LlaBIII^{T376A}. WT LlaBIII was able to

displace ~77% of the triplex DNA in 4 minutes, while LlaBIII^{T376A} displaced a similar amount (~70%) in 10 minutes and LlaBIII^{R564A} took 20 minutes to displace ~67% of the triplex DNA (Figure 4-9A-C). This was indicative of reduction in the rate of DNA translocation by LlaBIII^{T376A} and LlaBIII^{R564A} (McClelland et al., 2005). Interestingly, while LlaBIII^{T376A} displayed a visible though a less efficient nucleolytic activity, R564A mutation had negligible nuclease activity (Figure 4-8B). This suggested that the weak translocase activity of LlaBIII^{R564A} was not sufficient to activate the nuclease, while that of LlaBIII^{T376A} was sufficient to display visible DNA cleavage.

4.3.5 DNA cleavage by LlaBIII^{R564A} in cooperation with a wild-type enzyme

Since LlaBIII^{R564A} was ATPase-competent but less efficient translocase, we hypothesised that the two converging LlaBIII^{R564A} did not collide with sufficient force to activate the nuclease.
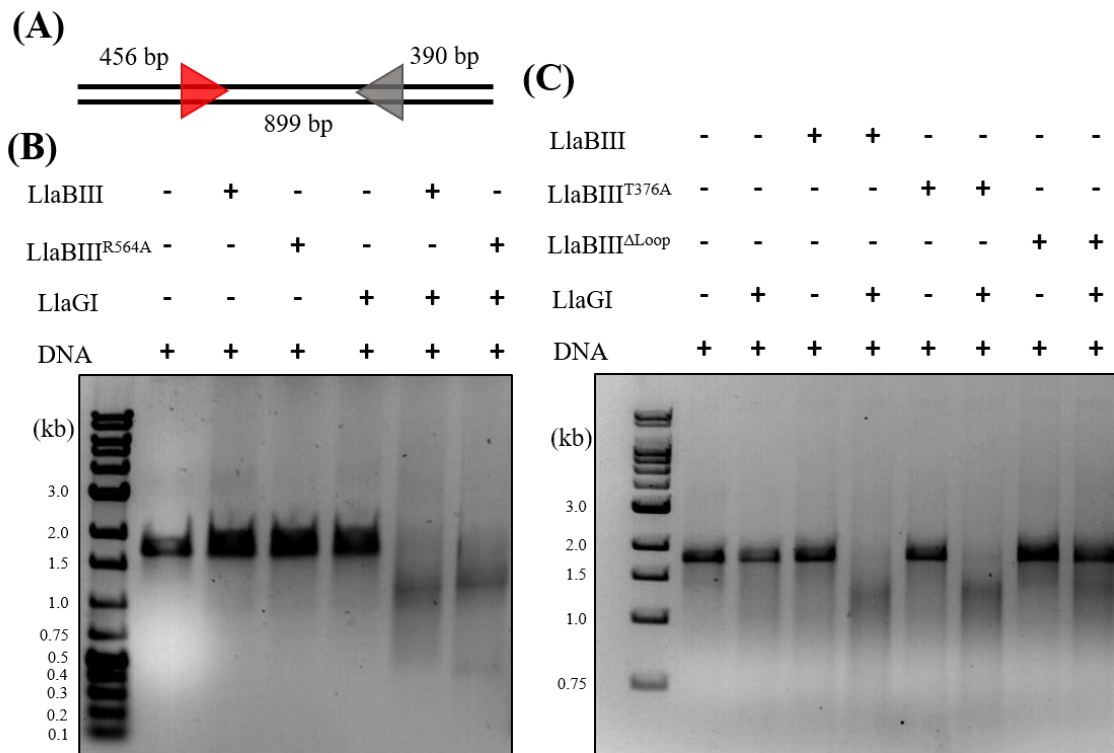
Figure 4-10: Heterologous cooperation assay. **(A)** Schematic of the DNA substrate used with LlaBIII site in red and the LlaGI site in grey. **(B, C)** DNA cleavage by LlaBIII or its mutants in cooperation with LlaGI. 500 nM of each enzyme was incubated with 10 nM of DNA and 4 mM of ATP for 30 minutes at 25°C.

Consequently, we predicted that the collision of a slower LlaBIII^{R564A} with a faster WT enzyme might result in sufficient force to activate the nuclease to cleave DNA. To address this prediction, we used the homologue LlaGI, which recognises a different recognition sequence (CTnGAYG). A 1758 bp long DNA containing a LlaBIII site and a LlaGI site in head-to-head orientation was used as the substrate (Figure 4-10A). As the DNA substrate had a single target site for the respective enzymes, WT LlaBIII, LlaBIII^{R564A} or LlaGI were unable to cleave the DNA on their

own (Figure 4-10B). However, in cooperation with LlaGI, WT LlaBIII and LlaBIII^{R564A} cleaved the DNA efficiently (Figure 4-10B).

This assay established that the nucleolytic activity of LlaBIII^{R564A}, a mutant with an inefficient translocase, could be rescued in cooperation with a WT enzyme. Similarly, LlaBIII^{T376A}, which had a lower ATPase and nuclease activities, displayed better nucleolytic activity in cooperation with LlaGI (Figure 4-10C). However, LlaBIII$^{\Delta Loop}$, which had an ATPase activity comparable to LlaBIII^{T376A} but little translocation activity, displayed negligible nuclease activity in cooperation with LlaGI (Figure. 4-10C). The results of these experiments were consistent with the above prediction that the nucleolytic cleavage is dependent on the translocation efficiencies of the colliding enzymes and also highlighted that for DNA to be cleaved the Type ISP RM enzyme should have the ability to translocate DNA, even if inefficiently.

4.4 Discussion

The study described here aimed at understanding the mechanism of stimulation of the ATPase activity and its coupling to the other functional activities of the Type ISP RM enzyme. Like many other helicases that require substrates for stimulation of the ATPase, in the case of a Type ISP RM enzyme stimulation requires DNA containing a target sequence. The study revealed that just the recognition of the target sequence by the TRD and MTase stimulates the ATPase partially, while the subsequent engagement of the DNA with the ATPase stimulates it fully (Figure 4-11). Consequently, the ATPase is not fully active unless the DNA is long enough to be gripped by the domain to initiate translocation. This ensures that ATP hydrolysis is kept negligibly low unless the DNA has a target sequence and sufficient length upstream to initiate DNA translocation.

We speculate that interdomain movement between and intradomain conformational changes within the N-core and the C-core upon DNA binding would make the ATPase competent to hydrolyse ATP. This speculation is supported by structural changes seen in SF2 ATPases upon binding to their substrates (Buttner et al., 2007; Fitzgerald et al., 2017; Gu and Rice, 2010; Kulkarni et al., 2016; Liu et al., 2017a; Yan et al., 2016). The observation that mere binding of the target sequence to the MTase and TRD stimulates the ATPase, though partially, suggests a direct allosteric communication between the target recognition unit (MTase-TRD) and the ATPase, possibly *via* the coupler domain. Complete stimulation is achieved *via* a substrate DNA long enough to interact with the ATPase. Determination and analysis of structures of Type ISP RM enzymes bound to DNA and ATP will be required to obtain the molecular basis of these conformational changes and allosteric communication.

Our study found the β-hairpin loop of the N-core of the ATPase, a structural element that the DNA interacts with, to be essential for nucleolytic cleavage as it coupled the ATPase to the translocase activity (Figure 4-11). The β-hairpin loop interacts with the DNA and facilitates translocation possibly by paddling the DNA resulting from the conformational changes in the ATPase domain upon nucleotide hydrolysis. The β-hairpin loop is not common to all SF2 helicase-like ATPases and was first noticed in Type ISP RM enzymes (Chand et al., 2015). A structure and sequence analyses of Type I RM enzymes revealed that some of these enzymes have a loop at an equivalent position and that they may perform a function similar to that of the β-hairpin loop of Type ISP RM enzymes (Figure 4-12).
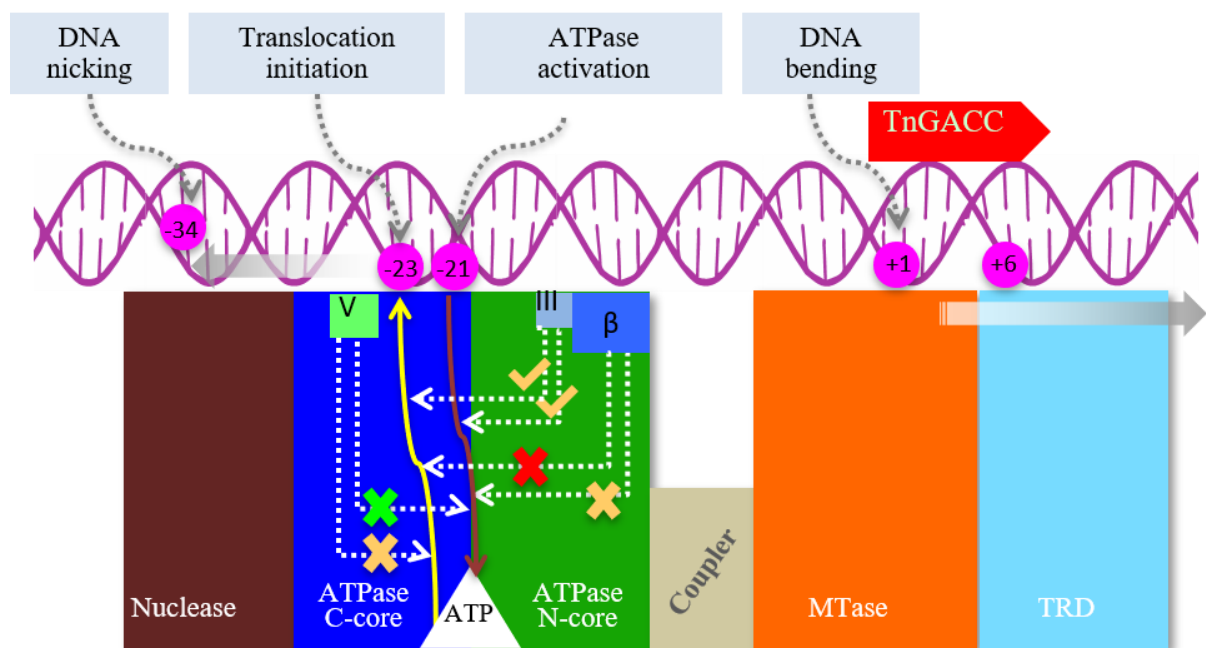
Figure 4-11: Schematic representation of the DNA-mediated coupling of the ATPase to translocase and nuclease activities. The target sequence (+1 to +6) and the region of DNA bending upon target binding, which steers the DNA towards the ATPase domain, are shown. The continuous brown arrow from the DNA at -21 to the ATP binding pocket (white) indicates the activation of the ATPase by the substrate, while the yellow line to the DNA at -23 indicates ATPase-coupled DNA translocation. The effect of the deletion of the β-hairpin loop, and mutation of motif III LlaBIII-Thr376 and motif V LlaBIII-Arg564 on the ATPase and translocase activities are illustrated using dotted arrow lines connecting the respective structural elements to the brown and yellow arrows coupling ATPase and translocase. A tick mark indicates detectable nucleolytic cleavage, while a cross mark indicates no cleavage. An amber tick mark indicates a diminished ATPase or translocase, but detectable nuclease activity, the green and the amber cross in the case of motif V indicate an active ATPase and a diminished translocase resulting in an inactive nuclease. The red and amber cross in the case of β-hairpin deletion mutant indicates reduced ATPase activity but inactive translocase and nuclease activities. The solid grey arrows indicate the direction of translocation of the DNA and the Type ISP RM enzyme, with respect to each other. -34 is the position of nicking by a static enzyme.

The conserved Arg564 of motif V in Type ISP RM enzyme LlaBIII also contributes to coupling the ATP hydrolysis to the translocase activity (Figure 4-11). LlaBIII[R564A] was an active ATPase but was slower at DNA translocation, suggesting a partial decoupling of the translocase and

ATPase. This conclusion is consistent with the role of motif V proposed in case of SF2 ATPases such as chromatin remodeler SWI/SNF (Smith and Peterson, 2005). The arginine, which in LlaBIII is located on the path traversed by the DNA (Figure 4-1B), is a conserved residue of motif
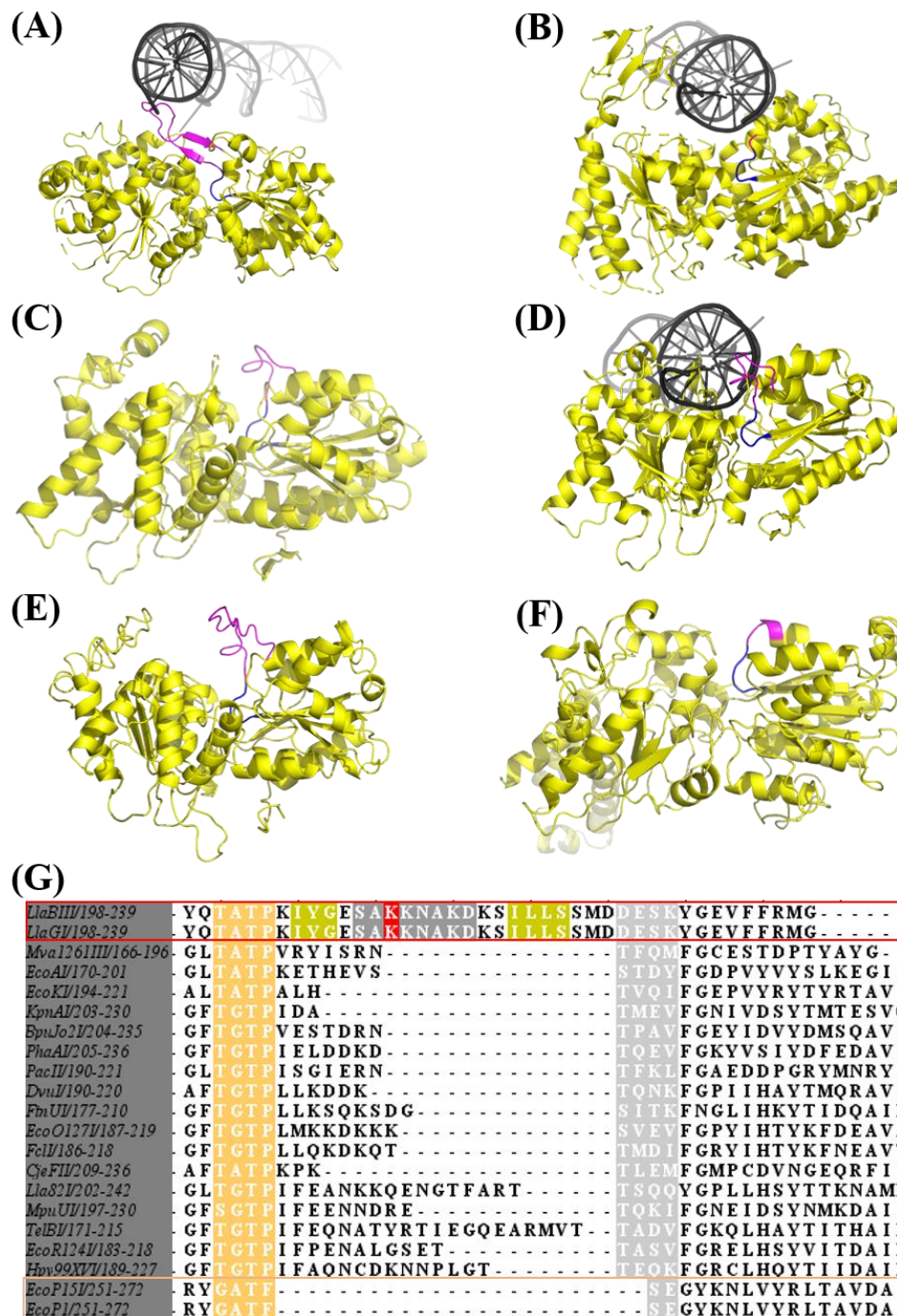
Figure 4-12: Structural comparison of the ATPase domains of RM enzymes. (A) DNA-bound LlaBIII, (B) DNA-bound Type III RM enzyme EcoP15I (Gupta et al., 2015), (C) Type I RM enzyme EcoR124I (Lapkouski et al., 2009), (D) EcoR124I with a modeled DNA indicating the possibility of the interaction of hairpin loop (magenta) with the DNA, (E) Type I RM enzyme TelBI modelled using I-TASSER (Yang et al., 2015) and (F) EcoKI modelled using ITASSER (Yang et al., 2015). The position of the β-hairpin loop is coloured magenta. Blue colour represents motif III. The β-hairpin loop appears to be present in EcoR124I and TelBI but is absent in EcoKI and EcoP15I. (G) Sequence alignment of a part of the helicase domain of Type ISP (LlaBIII and LlaGI), Type I and Type III (EcoP15I and EcoPI) indicating the possible presence of an equivalent loop in Type I RM enzymes.

V of SF2 ATPases and is found to interact with the phosphate backbone of bound DNA in many of them (Liu et al., 2017a; Yan et al., 2016). The interaction is expected to act as a grip on the DNA during translocation. Consequently, we think that the loss of the interaction with the DNA affected traction, and hence reduced the rate of translocation by LlaBIII^{R564A}.
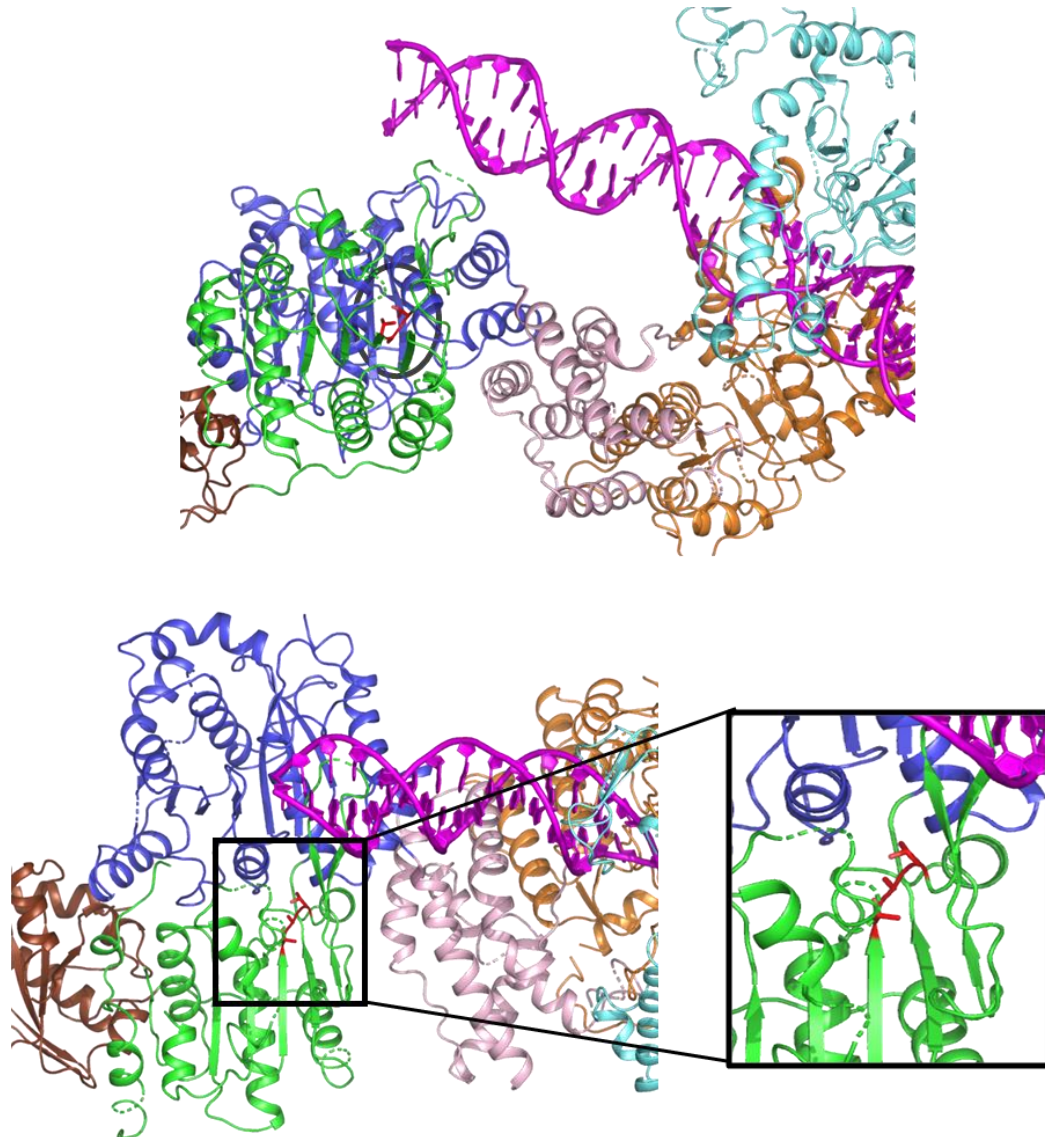
Figure 4-13: Structure of LlaBIII-DNA complex highlighting the position of motif III (red) with respect to the DNA and β-hairpin loop. Inset showing the motif III.

In Type ISP RM enzymes, unlike in many other SF2 ATPases (Banroques et al., 2010; Fitzgerald et al., 2017; Papanikou et al., 2004), motif III is masked from interacting with the substrate by the β-hairpin loop that is positioned adjacent to the motif (Figure 4-13). As a consequence, motif III is not expected to be involved in DNA translocation *via* direct interaction with the substrate. But being adjacent to the β-hairpin loop, residues of motif III may be affected by the changes in the loop on DNA binding. We find that the mutation of Thr376 of motif III decreases ATPase activity, which decreases the rate of translocation by LlaBIII^{T376A}. However, unlike LlaBIII^{R564A},

LlaBIII^{T376A} displayed a visible nucleolytic activity. This suggested the requirement of a threshold rate of translocation for the nuclease to be activated.

The slowing down of DNA translocation LlaBIII^{R564A} hindered its nucleolytic activity, which, however, was restored in cooperation with a WT enzyme, revealing the significance of the rate of translocation on the enzymatic activity. We interpret that the diminished rate of translocation affected the impact with which two LlaBIII^{R564A} collided and, thus, failed to activate the nuclease. Here, we envisage activation of the nuclease by conformational changes resulting from the impact of the collision, as proposed previously (van Aelst et al., 2015). Unlike LlaBIII^{T376A} and LlaBIII^{R564A}, we found that LlaBIII$^{\Delta Loop}$, which is an active ATPase but a completely inactive translocase, failed to cleave DNA on collision with a translocating WT enzyme. This implied that translocation initiation was required for DNA cleavage.

It has previously been shown that an ATPase inactive mutant of LlaBIII, which remains statically bound to its target site, on colliding with a translocating WT enzyme nicks the DNA but fails to form a double-strand DNA break (van Aelst et al., 2015). Combined with the data from this study, we conclude that double-strand DNA break by Type ISP RM enzymes not only requires ATP hydrolysis but also translocation initiation. We think that the activation of the nuclease of the ATP-dependent Type I RM enzymes, which actively translocate DNA, would also be dependent on the efficiency of translocation and force produced on enzyme collision. In the case of Type III RM enzymes, the other class of ATP-dependent RM enzymes, ATP hydrolysis and interaction between partner proteins are essential for activation of the nuclease for DNA nicking and double-strand DNA break (Ahmad et al., 2018). However, as the convergence of two Type III RM enzymes is driven by diffusion rather than active translocation (Ahmad et al., 2018; Crampton et al., 2007; Ramanathan et al., 2009), their nucleolytic activity will be independent of translocation-mediated activation and collision.

Conclusion and future directions

This thesis presents the first complete structure of an ATP dependent Type ISP RM enzyme (LlaBIII). The structure of LlaBIII bound to DNA formed the basis for the postulation of a new mechanism of dsDNA translocation and cleavage by the concerted action of two distal nucleases that converge to make multiple nicks on the DNA strands. The studies on the ATPase domain led us to identify a new structural motif, a β-hairpin loop, involved in the coupling of the ATP hydrolysis to DNA translocation. The studies on the β-hairpin loop and motif V showed that dsDNA break by Type ISP RM enzyme requires both ATP hydrolysis and initiation of DNA translocation. Motif V mutant (LlaBIII[R564A]), a slow translocating Type ISP RM enzyme was unable to cleave the DNA, but in partnership with a wild type enzyme, it efficiently cleaved the DNA suggesting a translocation dependent DNA cleavage activity. This study led us to conclude that the nuclease activation depends on the efficiency of translocation and force produced upon the collision of two converging enzymes.

Structural studies on Type ISP RM enzymes provided detailed information about the interaction of the TRD, MTase and ATPase domains of LlaBIII with dsDNA. This study along with the structure of LlaGIΔN (a nuclease domain deletion of LlaGI) bound to 22 bp DNA (Kulkarni et al., 2016) provided us with valuable insights into target sequence recognition by two closely related Type ISP RM enzymes. Structural comparison of residues involved in recognition of two different target sites by LlaGI and LlaBIII showed the difference in DNA specificity is achieved using the same set of structural elements but through variation in amino acids residues that constitute them (Kulkarni et al., 2016). A bioinformatics approach by Kulkarni et al., 2016, showed that a target change does not always involve an appropriate substitution of amino acids at the respective positions, but it also requires a compensatory mutation at distant locations to maintain the integrity of the structural elements (Kulkarni et al., 2016). By combining the consensus amino acid code for target site recognition (Kulkarni et al., 2016) and modular nature of Type ISP RM enzymes, these enzymes can be used as excellent model system for engineering new recognition sequences. Because of the modular nature, these engineered enzymes can be useful tools for biological processes such as site-specific DNA methylation and target site dependent chromatin remodelling.

In chapter 4, we have discussed the role of upstream dsDNA length in activation of the ATPase domain. The observation that recognition of the target sequence stimulates the ATPase partially, suggests a direct allosteric communication between the target recognition unit (MTase-TRD) and the ATPase, possibly through the coupler domain. Complete stimulation is achieved upon a substrate DNA long enough to interact with the ATPase. For a better understanding of the molecular basis of this allostery and conformational changes, the structure determination and

analysis of ternary complex (LlaBIII with DNA and ATP) is required. Towards this, we were able to crystallise the LlaBIII wild type enzyme with a longer DNA in the presence of non-hydrolysable ATP analogue. This crystal diffracted to 5.5 Å which was not helpful in providing the information about the conformational changes. Future crystallisation attempts by changing DNA length and use of ATP analogues may help in improving the resolution of the complex.

References

Abadjieva, A., Patel, J., Webb, M., Zinkevich, V., and Firman, K. (1993). A deletion mutant of the type IC restriction endonuclease EcoR1241 expressing a novel DNA specificity. Nucleic acids research *21*, 4435-4443.

Afonine, P.V., Mustyakimov, M., Grosse-Kunstleve, R.W., Moriarty, N.W., Langan, P., and Adams, P.D. (2010). Joint X-ray and neutron refinement with phenix.refine. Acta crystallographica Section D, Biological crystallography *66*, 1153-1163.

Ahmad, I., Kulkarni, M., Gopinath, A., and Saikrishnan, K. (2018). Single-site DNA cleavage by Type III restriction endonuclease requires a site-bound enzyme and a trans-acting enzyme that are ATPase-activated. Nucleic acids research *46*, 6229-6237.

Arber, W., and Wauters-Willems, D. (1970). Host specificity of DNA produced by Escherichia coli. XII. The two restriction and modification systems of strain 15T. Molecular & general genetics : MGG *108*, 203-217.

Banroques, J., Doere, M., Dreyfus, M., Linder, P., and Tanner, N.K. (2010). Motif III in superfamily 2 "helicases" helps convert the binding energy of ATP into a high-affinity RNA binding site in the yeast DEAD-box protein Ded1. Journal of molecular biology *396*, 949-966.

Barcus, V.A., Titheradge, A.J., and Murray, N.E. (1995). The diversity of alleles at the hsd locus in natural populations of Escherichia coli. Genetics *140*, 1187-1197.

Beck, A., Vijayanathan, V., Thomas, T., and Thomas, T.J. (2013). Ionic microenvironmental effects on triplex DNA stabilization: cationic counterion effects on poly(dT).poly(dA).poly(dT). Biochimie *95*, 1310-1318.

Bialevich, V., Sinha, D., Shamayeva, K., Guzanova, A., Reha, D., Csefalvay, E., Carey, J., Weiserova, M., and Ettrich, R.H. (2017). The helical domain of the EcoR124I motor subunit participates in ATPase activity and dsDNA translocation. PeerJ *5*, e2887.

Blumenthal, R.M., and Cheng, X. (2001). A Taq attack displaces bases. Nature structural biology *8*, 101-103.

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

Bullas, L.R., Colson, C., and Neufeld, B. (1980). Deoxyribonucleic acid restriction and modification systems in Salmonella: chromosomally located systems of different serotypes. Journal of bacteriology *141*, 275-292.

Bullas, L.R., Colson, C., and Van Pel, A. (1976). DNA restriction and modification systems in Salmonella. SQ, a new system derived by recombination between the SB system of Salmonella typhimurium and the SP system of Salmonella potsdam. Journal of general microbiology *95*, 166-172.

Buttner, K., Nehring, S., and Hopfner, K.P. (2007). Structural basis for DNA duplex separation by a superfamily-2 helicase. Nature structural & molecular biology *14*, 647-652.

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

Chand, M.K., Nirwan, N., Diffin, F.M., van Aelst, K., Kulkarni, M., Pernstich, C., Szczelkun, M.D., and Saikrishnan, K. (2015). Translocation-coupled DNA cleavage by the Type ISP restriction-modification enzymes. Nature chemical biology *11*, 870-877.

Cheng, X., and Roberts, R.J. (2001). AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic acids research *29*, 3784-3795.

Chin, V., Valinluck, V., Magaki, S., and Ryu, J. (2004). KpnBI is the prototype of a new family (IE) of bacterial type I restriction-modification system. Nucleic acids research *32*, e138.

Christiansen, M., Stevnsner, T., Modin, C., Martensen, P.M., Brosh, R.M., Jr., and Bohr, V.A. (2003). Functional consequences of mutations in the conserved SF2 motifs and post-translational phosphorylation of the CSB protein. Nucleic acids research *31*, 963-973.

Clark, T.A., Murray, I.A., Morgan, R.D., Kislyuk, A.O., Spittle, K.E., Boitano, M., Fomenkov, A., Roberts, R.J., and Korlach, J. (2012). Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic acids research *40*, e29.

Cowan, G.M., Gann, A.A., and Murray, N.E. (1989). Conservation of complex DNA recognition domains between families of restriction enzymes. Cell *56*, 103-109.

Crampton, N., Roes, S., Dryden, D.T., Rao, D.N., Edwardson, J.M., and Henderson, R.M. (2007). DNA looping and translocation provide an optimal cleavage mechanism for the type III restriction enzymes. The EMBO journal *26*, 3815-3825.

Davies, G.P., Kemp, P., Molineux, I.J., and Murray, N.E. (1999a). The DNA translocation and ATPase activities of restriction-deficient mutants of Eco KI. Journal of molecular biology *292*, 787-796.

Davies, G.P., Martin, I., Sturrock, S.S., Cronshaw, A., Murray, N.E., and Dryden, D.T. (1999b). On the structure and operation of type I DNA restriction enzymes. Journal of molecular biology *290*, 565-579.

Davies, G.P., Powell, L.M., Webb, J.L., Cooper, L.P., and Murray, N.E. (1998). EcoKI with an amino acid substitution in any one of seven DEAD-box motifs has impaired ATPase and endonuclease activities. Nucleic acids research *26*, 4828-4836.

Dillingham, M.S., and Kowalczykowski, S.C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiology and molecular biology reviews : MMBR *72*, 642-671, Table of Contents.

Dryden, D.T., Cooper, L.P., and Murray, N.E. (1993). Purification and characterization of the methyltransferase from the type 1 restriction and modification system of Escherichia coli K12. The Journal of biological chemistry *268*, 13228-13236.

Dryden, D.T., Murray, N.E., and Rao, D.N. (2001). Nucleoside triphosphate-dependent restriction enzymes. Nucleic acids research *29*, 3728-3741.

Dryden, D.T., Willcock, D.F., and Murray, N.E. (1995). Mutational analysis of conserved amino-acid motifs in EcoKI adenine methyltransferase. Gene *157*, 123-124.

Durr, H., Flaus, A., Owen-Hughes, T., and Hopfner, K.P. (2006). Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures. Nucleic acids research *34*, 4160-4167.

Durr, H., Korner, C., Muller, M., Hickmann, V., and Hopfner, K.P. (2005). X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. Cell *121*, 363-373.

Dybvig, K., Sitaraman, R., and French, C.T. (1998). A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. Proceedings of the National Academy of Sciences of the United States of America *95*, 13923-13928.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta crystallographica Section D, Biological crystallography *66*, 486-501.

Endlich, B., and Linn, S. (1985). The DNA restriction endonuclease of Escherichia coli B. II. Further studies of the structure of DNA intermediates and products. The Journal of biological chemistry *260*, 5729-5738.

Evans, P. (2006). Scaling and assessment of data quality. Acta crystallographica Section D, Biological crystallography *62*, 72-82.

Fagerburg, M.V., Schauer, G.D., Thickman, K.R., Bianco, P.R., Khan, S.A., Leuba, S.H., and Anand, S.P. (2012). PcrA-mediated disruption of RecA nucleoprotein filaments--essential role of the ATPase activity of RecA. Nucleic acids research *40*, 8416-8424.

Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. Current opinion in structural biology *20*, 313-324.

Fernandez, A., Guo, H.S., Saenz, P., Simon-Buela, L., Gomez de Cedron, M., and Garcia, J.A. (1997). The motif V of plum pox potyvirus CI RNA helicase is involved in NTP hydrolysis and is essential for virus RNA replication. Nucleic acids research *25*, 4474-4480.

Firman, K., Price, C., and Glover, S.W. (1985). The EcoR124 and EcoR124/3 restriction and modification systems: cloning the genes. Plasmid *14*, 224-234.

Firman, K., and Szczelkun, M.D. (2000). Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. The EMBO journal *19*, 2094-2102.

Fitzgerald, M.E., Rawling, D.C., Potapova, O., Ren, X., Kohlway, A., and Pyle, A.M. (2017). Selective RNA targeting and regulated signaling by RIG-I is controlled by coordination of RNA and ATP binding. Nucleic acids research *45*, 1442-1454.

French, S., and Wilson, K. (1978). On the treatment of negative intensity observations. Acta crystallographica Section A, Foundations and advances *34*, 517–525.

Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment. Proteins *23*, 566-579.

Fuller-Pace, F.V., Bullas, L.R., Delius, H., and Murray, N.E. (1984). Genetic recombination can generate altered restriction specificity. Proceedings of the National Academy of Sciences of the United States of America *81*, 6095-6099.

Fuller-Pace, F.V., and Murray, N.E. (1986). Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes. Proceedings of the National Academy of Sciences of the United States of America *83*, 9368-9372.

Gann, A.A., Campbell, A.J., Collins, J.F., Coulson, A.F., and Murray, N.E. (1987). Reassortment of DNA recognition domains and the evolution of new specificities. Molecular microbiology *1*, 13-22.

Gupta, Y.K., Chan, S.H., Xu, S.Y., and Aggarwal, A.K. (2015). Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. Nature communications *6*, 7363.

Glover, S.W., Firman, K., Watson, G., Price, C., and Donaldson, S. (1983). The alternate expression of two restriction and modification systems. Molecular & general genetics : MGG *190*, 65-69.

Goedecke, K., Pignot, M., Goody, R.S., Scheidig, A.J., and Weinhold, E. (2001). Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. Nature structural biology *8*, 121-125.

Grinkevich, P., Sinha, D., Iermak, I., Guzanova, A., Weiserova, M., Ludwig, J., Mesters, J.R., and Ettrich, R.H. (2018). Crystal structure of a novel domain of the motor subunit of the Type I restriction enzyme EcoR124 involved in complex assembly and DNA binding. The Journal of biological chemistry *293*, 15043-15054.

Gu, M., and Rice, C.M. (2010). Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. Proceedings of the National Academy of Sciences of the United States of America *107*, 521-528.

Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A., and Bickle, T.A. (1992). Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. The EMBO journal *11*, 233-240.

Halford, S.E., Welsh, A.J., and Szczelkun, M.D. (2004). Enzyme-mediated DNA looping. Annual review of biophysics and biomolecular structure *33*, 1-24.

Hopfner, K.P., Gerhold, C.B., Lakomek, K., and Wollmann, P. (2012). Swi2/Snf2 remodelers: hybrid views on hybrid molecular machines. Current opinion in structural biology *22*, 225-233.

Horton, J.R., Liebert, K., Hattman, S., Jeltsch, A., and Cheng, X. (2005). Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. Cell *121*, 349-361.

Jankowsky, E., Gross, C.H., Shuman, S., and Pyle, A.M. (2001). Active disruption of an RNA-protein interaction by a DExH/D RNA helicase. Science *291*, 121-125.

Janscak, P., and Bickle, T.A. (1998). The DNA recognition subunit of the type IB restriction-modification enzyme EcoAI tolerates circular permutions of its polypeptide chain. Journal of molecular biology *284*, 937-948.

Janscak, P., MacWilliams, M.P., Sandmeier, U., Nagaraja, V., and Bickle, T.A. (1999). DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. The EMBO journal *18*, 2638-2647.

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

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

Kan, N.C., Lautenberger, J.A., Edgell, M.H., and Hutchison, C.A., 3rd (1979). The nucleotide sequence recognized by the Escherichia coli K12 restriction and modification enzymes. Journal of molecular biology *130*, 191-209.

Kasarjian, J.K., Kodama, Y., Iida, M., Matsuda, K., and Ryu, J. (2005). Four new type I restriction enzymes identified in Escherichia coli clinical isolates. Nucleic acids research *33*, e114.

Kennaway, C.K., Obarska-Kosinska, A., White, J.H., Tuszynska, I., Cooper, L.P., Bujnicki, J.M., Trinick, J., and Dryden, D.T. (2009). The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. Nucleic acids research *37*, 762-770.

Kennaway, C.K., Taylor, J.E., Song, C.F., Potrzebowski, W., Nicholson, W., White, J.H., Swiderska, A., Obarska-Kosinska, A., Callow, P., Cooper, L.P., *et al.* (2012). Structure and operation of the DNA-translocating type I DNA restriction enzymes. Genes & development *26*, 92-104.

Kim, J.S., DeGiovanni, A., Jancarik, J., Adams, P.D., Yokota, H., Kim, R., and Kim, S.H. (2005). Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. Proceedings of the National Academy of Sciences of the United States of America *102*, 3248-3253.

Kulkarni, M., Nirwan, N., van Aelst, K., Szczelkun, M.D., and Saikrishnan, K. (2016). Structural insights into DNA sequence recognition by Type ISP restriction-modification enzymes. Nucleic acids research *44*, 4396-4408.

Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I., and Saenger, W. (1994). Three-dimensional structure of the adenine-specific DNA methyltransferase M.Taq I in complex with the cofactor S-adenosylmethionine. Proceedings of the National Academy of Sciences of the United States of America *91*, 10957-10961.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nature reviews Microbiology *8*, 317-327.

Lapkouski, M., Panjikar, S., Janscak, P., Smatanova, I.K., Carey, J., Ettrich, R., and Csefalvay, E. (2009). Structure of the motor subunit of type I restriction-modification complex EcoR124I. Nature structural & molecular biology *16*, 94-95.

Lautenberger, J.A., Edgell, M.H., Hutchison, C.A., 3rd, and Godson, G.N. (1979). The DNA sequence on bacteriophage G4 recognized by the Escherichia coli B restriction enzyme. Journal of molecular biology *131*, 871-875.

Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D., and Zakrzewska, K. (2009). Conformational analysis of nucleic acids revisited: Curves+. Nucleic acids research *37*, 5917-5929.

Lee, J.Y., and Yang, W. (2006). UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. Cell *127*, 1349-1360.

Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., and Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature *520*, 505-510.

Linn, S., and Arber, W. (1968). Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. Proceedings of the National Academy of Sciences of the United States of America *59*, 1300-1306.

Liu, X., Li, M., Xia, X., Li, X., and Chen, Z. (2017a). Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. Nature *544*, 440-445.

Liu, Y.P., Tang, Q., Zhang, J.Z., Tian, L.F., Gao, P., and Yan, X.X. (2017b). Structural basis underlying complex assembly and conformational transition of the type I R-M system. Proceedings of the National Academy of Sciences of the United States of America *114*, 11151-11156.

Loenen, W.A., Dryden, D.T., Raleigh, E.A., and Wilson, G.G. (2014a). Type I restriction enzymes and their relatives. Nucleic acids research *42*, 20-44.

Loenen, W.A., Dryden, D.T., Raleigh, E.A., Wilson, G.G., and Murray, N.E. (2014b). Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic acids research *42*, 3-19.

Luria, S.E., and Human, M.L. (1952). A nonhereditary, host-induced variation of bacterial viruses. Journal of bacteriology *64*, 557-569.

MacWilliams, M.P., and Bickle, T.A. (1996). Generation of new DNA binding specificity by truncation of the type IC EcoDXXI hsdS gene. The EMBO journal *15*, 4775-4783.

Matthews, B.W. (1968). Solvent content of protein crystals. Journal of molecular biology *33*, 491-497.

McClelland, S.E., Dryden, D.T., and Szczelkun, M.D. (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., Mackeldanz, P., Bickle, T.A., Kruger, D.H., and Schroeder, C. (1995). Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. The EMBO journal *14*, 2958-2966.

Meister, J., MacWilliams, M., Hubner, P., Jutte, H., Skrzypek, E., Piekarowicz, A., and Bickle, T.A. (1993). Macroevolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn5 transposition. The EMBO journal *12*, 4585-4591.

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

Murray, I.A., Clark, T.A., Morgan, R.D., Boitano, M., Anton, B.P., Luong, K., Fomenkov, A., Turner, S.W., Korlach, J., and Roberts, R.J. (2012). The methylomes of six bacteria. Nucleic acids research *40*, 11450-11462.

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

Nagaraja, V., Shepherd, J.C., and Bickle, T.A. (1985a). A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. Nature *316*, 371-372.

Nagaraja, V., Shepherd, J.C., Pripfl, T., and Bickle, T.A. (1985b). Two type I restriction enzymes from Salmonella species. Purification and DNA recognition sequences. Journal of molecular biology *182*, 579-587.

Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. Cell *154*, 490-503.

Niv, M.Y., Ripoll, D.R., Vila, J.A., Liwo, A., Vanamee, E.S., Aggarwal, A.K., Weinstein, H., and Scheraga, H.A. (2007). Topology of Type II REases revisited; structural classes and the common conserved core. Nucleic acids research *35*, 2227-2237.

Obarska-Kosinska, A., Taylor, J.E., Callow, P., Orlowski, J., Bujnicki, J.M., and Kneale, G.G. (2008). HsdR subunit of the type I restriction-modification enzyme EcoR124I: biophysical characterisation and structural modelling. Journal of molecular biology *376*, 438-452.

Papanikou, E., Karamanou, S., Baud, C., Sianidis, G., Frank, M., and Economou, A. (2004). Helicase Motif III in SecA is essential for coupling preprotein binding to translocation ATPase. EMBO reports *5*, 807-811.

Park, J., Myong, S., Niedziela-Majka, A., Lee, K.S., Yu, J., Lohman, T.M., and Ha, T. (2010). PcrA helicase dismantles RecA filaments by reeling in DNA in uniform steps. Cell *142*, 544-555.

Park, S.Y., Lee, H.J., Song, J.M., Sun, J., Hwang, H.J., Nishi, K., and Kim, J.S. (2012). Structural characterization of a modification subunit of a putative type I restriction enzyme from Vibrio vulnificus YJ016. Acta crystallographica Section D, Biological crystallography *68*, 1570-1577.

Pause, A., and Sonenberg, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. The EMBO journal *11*, 2643-2654.

Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E., (2003). A graphical user interface to the CCP4 program suite. Acta crystallographica Section D, Biological crystallography *59,* 1131-1137.

Price, C., Lingner, J., Bickle, T.A., Firman, K., and Glover, S.W. (1989). Basis for changes in DNA recognition by the EcoR124 and EcoR124/3 type I DNA restriction and modification enzymes. Journal of molecular biology *205*, 115-125.

Price, C., Pripfl, T., and Bickle, T.A. (1987a). EcoR124 and EcoR124/3: the first members of a new family of type I restriction and modification systems. European journal of biochemistry / FEBS *167*, 111-115.

Price, C., Shepherd, J.C., and Bickle, T.A. (1987b). DNA recognition by a new family of type I restriction enzymes: a unique relationship between two different DNA specificities. The EMBO journal *6*, 1493-1497.

Pues, H., Bleimling, N., Holz, B., Wolcke, J., and Weinhold, E. (1999). Functional roles of the conserved aromatic amino acid residues at position 108 (motif IV) and position 196 (motif VIII)

in base flipping and catalysis by the N6-adenine DNA methyltransferase from Thermus aquaticus. Biochemistry *38*, 1426-1434.

Ramanathan, S.P., van Aelst, K., Sears, A., Peakman, L.J., Diffin, F.M., Szczelkun, M.D., and Seidel, R. (2009). Type III restriction enzymes communicate in 1D without looping between their target sites. Proceedings of the National Academy of Sciences of the United States of America *106*, 1748-1753.

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

Ravetch, J.V., Horiuchi, K., and Zinder, N.D. (1978). Nucleotide sequence of the recognition site for the restriction-modification enzyme of Escherichia coli B. Proceedings of the National Academy of Sciences of the United States of America *75*, 2266-2270.

Reiser, J., and Yuan, R. (1977). Purification and properties of the P15 specific restriction endonuclease from Escherichia coli. The Journal of biological chemistry *252*, 451-456.

Rosamond, J., Endlich, B., and Linn, S. (1979). Electron microscopic studies of the mechanism of action of the restriction endonuclease of Escherichia coli B. Journal of molecular biology *129*, 619-635.

Ryu, J., Rajadas, P.T., and Bullas, L.R. (1988). Complementation and hybridization evidence for additional families of type I DNA restriction and modification genes in Salmonella serotypes. Journal of bacteriology *170*, 5785-5788.

Saikrishnan, K., Powell, B., Cook, N.J., Webb, M.R., and Wigley, D.B. (2009). Mechanistic basis of 5'-3' translocation in SF1B helicases. Cell *137*, 849-859.

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

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

Seidel, R., Bloom, J.G., van Noort, J., Dutta, C.F., Dekker, N.H., Firman, K., Szczelkun, M.D., and Dekker, C. (2005). Dynamics of initiation, termination and reinitiation of DNA translocation by the motor protein EcoR124I. The EMBO journal *24*, 4188-4197.

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

Shen, B.W., Xu, D., Chan, S.H., Zheng, Y., Zhu, Z., Xu, S.Y., and Stoddard, B.L. (2011). Characterization and crystal structure of the type IIG restriction endonuclease RM.BpuSI. Nucleic acids research *39*, 8223-8236.

Sibley, M.H., and Raleigh, E.A. (2004). Cassette-like variation of restriction enzyme genes in Escherichia coli C and relatives. Nucleic acids research *32*, 522-534.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annual review of biochemistry *76*, 23-50.

Sisakova, E., Stanley, L.K., Weiserova, M., and Szczelkun, M.D. (2008a). A RecB-family nuclease motif in the Type I restriction endonuclease EcoR124I. Nucleic acids research *36*, 3939-3949.

Sisakova, E., van Aelst, K., Diffin, F.M., and Szczelkun, M.D. (2013). The Type ISP Restriction-Modification enzymes LlaBIII and LlaGI use a translocation-collision mechanism to cleave non-specific DNA distant from their recognition sites. Nucleic acids research *41*, 1071-1080.

Sisakova, E., Weiserova, M., Dekker, C., Seidel, R., and Szczelkun, M.D. (2008b). The interrelationship of helicase and nuclease domains during DNA translocation by the molecular motor EcoR124I. Journal of molecular biology *384*, 1273-1286.

Sitaraman, R., and Dybvig, K. (1997). The hsd loci of Mycoplasma pulmonis: organization, rearrangements and expression of genes. Molecular microbiology *26*, 109-120.

Smith, C.L., and Peterson, C.L. (2005). A conserved Swi2/Snf2 ATPase motif couples ATP hydrolysis to chromatin remodeling. Molecular and cellular biology *25*, 5880-5892.

Smith, H.O., and Wilcox, K.W. (1970). A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. Journal of molecular biology *51*, 379-391.

Smith, R.M., Diffin, F.M., Savery, N.J., Josephsen, J., and Szczelkun, M.D. (2009a). DNA cleavage and methylation specificity of the single polypeptide restriction-modification enzyme LlaGI. Nucleic acids research *37*, 7206-7218.

Smith, R.M., Josephsen, J., and Szczelkun, M.D. (2009b). An Mrr-family nuclease motif in the single polypeptide restriction-modification enzyme LlaGI. Nucleic acids research *37*, 7231-7238.

Smith, R.M., Josephsen, J., and Szczelkun, M.D. (2009c). The single polypeptide restriction-modification enzyme LlaGI is a self-contained molecular motor that translocates DNA loops. Nucleic acids research *37*, 7219-7230.

Sommer, R., and Schaller, H. (1979). Nucleotide sequence of the recognition site of the B-specific restriction modification system in E. coli. Molecular & general genetics : MGG *168*, 331-335.

Stanley, L.K., Seidel, R., van der Scheer, C., Dekker, N.H., Szczelkun, M.D., and Dekker, C. (2006). When a helicase is not a helicase: dsDNA tracking by the motor protein EcoR124I. The EMBO journal *25*, 2230-2239.

Studier, F.W., and Bandyopadhyay, P.K. (1988). Model for how type I restriction enzymes select cleavage sites in DNA. Proceedings of the National Academy of Sciences of the United States of America *85*, 4677-4681.

Suri, B., and Bickle, T.A. (1985). EcoA: the first member of a new family of type I restriction modification systems. Gene organization and enzymatic activities. Journal of molecular biology *186*, 77-85.

Suri, B., Shepherd, J.C., and Bickle, T.A. (1984). The EcoA restriction and modification system of Escherichia coli 15T-: enzyme structure and DNA recognition sequence. The EMBO journal *3*, 575-579.

Szczelkun, M.D., Dillingham, M.S., Janscak, P., Firman, K., and Halford, S.E. (1996). Repercussions of DNA tracking by the type IC restriction endonuclease EcoR124I on linear, circular and catenated substrates. The EMBO journal *15*, 6335-6347.

Taylor, A.F., and Smith, G.R. (1985). Substrate specificity of the DNA unwinding activity of the RecBC enzyme of Escherichia coli. Journal of molecular biology *185*, 431-443.

Taylor, I., Watts, D., and Kneale, G. (1993). Substrate recognition and selectivity in the type IC DNA modification methylase M.EcoR124I. Nucleic acids research *21*, 4929-4935.

Thoma, N.H., Czyzewski, B.K., Alexeev, A.A., Mazin, A.V., Kowalczykowski, S.C., and Pavletich, N.P. (2005). Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. Nature structural & molecular biology *12*, 350-356.

Thorpe, P.H., Ternent, D., and Murray, N.E. (1997). The specificity of sty SKI, a type I restriction enzyme, implies a structure with rotational symmetry. Nucleic acids research *25*, 1694-1700.

Titheradge, A.J., King, J., Ryu, J., and Murray, N.E. (2001). Families of restriction enzymes: an analysis prompted by molecular and genetic data for type ID restriction and modification systems. Nucleic acids research *29*, 4195-4205.

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

van Aelst, K., Saikrishnan, K., and Szczelkun, M.D. (2015). Mapping DNA cleavage by the Type ISP restriction-modification enzymes following long-range communication between DNA sites in different orientations. Nucleic acids research *43*, 10430-10443.

van Aelst, K., Sisakova, E., and Szczelkun, M.D. (2013). DNA cleavage by Type ISP Restriction-Modification enzymes is initially targeted to the 3'-5' strand. Nucleic acids research *41*, 1081-1090.

van Noort, J., van der Heijden, T., Dutta, C.F., Firman, K., and Dekker, C. (2004). Initiation of translocation by Type I restriction-modification enzymes is associated with a short DNA extrusion. Nucleic acids research *32*, 6540-6547.

Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S., and Wigley, D.B. (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell *97*, 75-84.

Wollmann, P., Cui, S., Viswanathan, R., Berninghausen, O., Wells, M.N., Moldt, M., Witte, G., Butryn, A., Wendler, P., Beckmann, R.*, et al.* (2011). Structure and mechanism of the Swi2/Snf2 remodeller Mot1 in complex with its substrate TBP. Nature *475*, 403-407.

Yan, L., Wang, L., Tian, Y., Xia, X., and Chen, Z. (2016). Structure and regulation of the chromatin remodeller ISWI. Nature *540*, 466-469.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction. Nature methods, 12, 7-8.

Yeeles, J.T., Cammack, R., and Dillingham, M.S. (2009). An iron-sulfur cluster is essential for the binding of broken DNA by AddAB-type helicase-nucleases. The Journal of biological chemistry *284*, 7746-7755.

Yuan, R., Hamilton, D.L., and Burckhardt, J. (1980). DNA translocation by the restriction enzyme from E. coli K. Cell *20*, 237-244.

Zhang, Y., Smith, C.L., Saha, A., Grill, S.W., Mihardja, S., Smith, S.B., Cairns, B.R., Peterson, C.L., and Bustamante, C. (2006). DNA translocation and loop formation mechanism of chromatin remodeling by SWI/SNF and RSC. Molecular cell *24*, 559-568.

Author's Publication

Translocation-coupled DNA cleavage by the Type ISP restriction-modification enzymes

Mahesh K Chand[1], Neha Nirwan[1], Fiona M Diffin[2], Kara van Aelst[2], Manasi Kulkarni[1], Christian Pernstich[2], Mark D Szczelkun[2]* & Kayarat Saikrishnan[1]*

Production of endonucleolytic double-strand DNA breaks requires separate strand cleavage events. Although catalytic mechanisms for simple, dimeric endonucleases are known, there are many complex nuclease machines that are poorly understood. Here we studied the single polypeptide Type ISP restriction-modification (RM) enzymes, which cleave random DNA between distant target sites when two enzymes collide after convergent ATP-driven translocation. We report the 2.7-Å resolution X-ray crystal structure of a Type ISP enzyme–DNA complex, revealing that both the helicase-like ATPase and nuclease are located upstream of the direction of translocation, an observation inconsistent with simple nuclease-domain dimerization. Using single-molecule and biochemical techniques, we demonstrate that each ATPase remodels its DNA-protein complex and translocates along DNA without looping it, leading to a collision complex in which the nuclease domains are distal. Sequencing of the products of single cleavage events suggests a previously undescribed endonuclease model, where multiple, stochastic strand-nicking events combine to produce DNA scission.

The prokaryotic ATP-dependent restriction-modification (RM) enzymes provide a potent defense against infection by foreign and bacteriophage DNA, and accordingly have a widespread distribution[1,2]. Although recognition of specific sequences (targets) in the foreign DNA leads to nucleolytic cleavage (restriction), cleavage of self DNA is prevented by the methylation (modification) of the target by the same enzyme or enzyme complex. Since the isolation of the first such enzymes in 1968 (refs. 3,4), which helped launch the molecular biology revolution, many ATP-dependent RM enzymes have been characterized, including the classical heteropentameric Type I systems[5] and the closely related but monomeric single polypeptide Type ISP systems[5,6]. These enzymes are a model for understanding modular, multifunctional protein machines[7], particularly in formulating concepts of protein-DNA recognition, DNA methylation and base flipping, nuclease activity[6,8], double-stranded (ds)DNA translocation by superfamily 2 (SF2) helicases[9,10] and long-range communication by enzymes[11,12]. The first insights into the molecular organization of ATP-dependent enzymes came from structural analysis of the Type I RM enzymes EcoKI and EcoR124I using a combination of negative-stain electron microscopy, neutron scattering and structural modeling[13]. However, despite over 40 years of research, the molecular details of these enzymes' actions are unclear because of the lack of high-resolution structures.

To address the lack of high-resolution structural data, we undertook structure-function studies of Type ISP enzymes LlaGI and LlaBIII from *Lactococcus lactis*. The Type ISP enzymes perform the disparate functions of restriction and modification by coordinating four functional domains within a single polypeptide: a target recognition domain (TRD) that recognizes a 6 or 7 base pair (bp) asymmetric sequence (**Supplementary Results**, **Supplementary Fig. 1a**), an N6-adenine methyltransferase (MTase) that modifies one strand of the sequence using *S*-adenosyl methionine (AdoMet) stimulated by ATP, an SF2 helicase-like ATPase motor and a nuclease[6]. Unlike the well-studied SF1 and SF2 helicases that translocate single-stranded (ss)DNA and ssRNA, much less is known about the mechanism of dsDNA translocation by SF2 ATPases, which is important in many

biological phenomena such as nucleosome remodeling, removal of stalled RNA polymerases or modulation of gene expression[14–17]. A structure-function analysis of Type ISP enzymes would further our understanding of dsDNA translocation, particularly in the context of a fully functioning molecular machine.

We also sought to understand how the Type ISP nuclease domains interact to produce a dsDNA break. Previous biochemical analysis pointed to a loop-translocation mechanism leading to DNA cleavage, most similar to that of classical Type I RM enzymes (**Supplementary Fig. 1b**)[6,18–20]. After Type ISP target recognition, extensive ATP hydrolysis (~1–2 ATP per base pair) produces unidirectional DNA translocation downstream of the target[18–20]. Because of the directional motion, DNA cleavage only occurs when there is at least one pair of head-to-head targets on the same DNA–DNA with isolated sites or pairs of sites in other orientations are, at best, only nicked[6]. dsDNA breaks are produced at nonspecific loci between the head-to-head targets consistent with convergent enzyme translocation and collision[18–20].

The prevailing Type ISP model predicts that the ATPase domain will be situated downstream of the MTase-TRD and that, despite the complexity of communication, cleavage will result from direct dimerization of nuclease domains in a collision complex. However, the structure we present shows that the nuclease-ATPase domains of Type ISP enzymes are located upstream of the MTase-TRD domains relative to the bound directional target sequence. This unexpected arrangement has important implications for both DNA translocation and cleavage. Guided by the structure and based on single-molecule biophysical and biochemical studies, we propose a model for dsDNA-break production resulting from multiple, random strand nicks.

RESULTS
Architecture of Type ISP RM enzymes

X-ray crystallographic studies initiated using full-length LlaGI yielded crystals that diffracted X-rays poorly. As an alternate strategy, we carried out studies on a close homolog, LlaBIII, which has
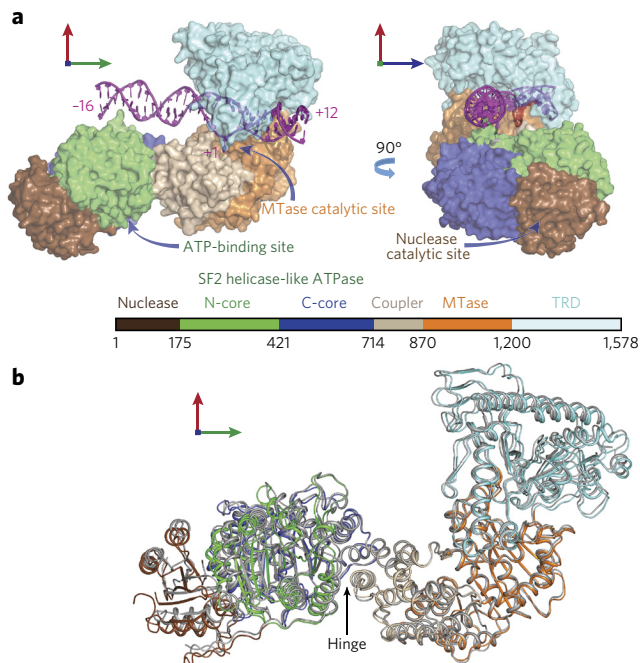
Figure 1 | Architecture of the modular Type ISP RM enzymes. (a) Surface representation of LlaBIII-DNA. **(b)** Superposition of the two chains of LlaBIII revealed conformational plasticity. The coupler domain (residues 715–870) of chain B (gray) was superimposed onto the coupler domain of chain A using the secondary structure–matching method in Coot[46]. The structural domains of chain A are colored as in **a**. The hinge about which the nuclease-ATPase domains move with respect to the coupler is indicated (**Supplementary Table 1**).

very high sequence identity of ~80% (the first 1,020 amino acids have ~98% sequence identity, and the remaining sequence has 48% identity). The two proteins can functionally cooperate to cleave DNA[20]. We determined the structure of full-length LlaBIII bound to a 28-bp DNA substrate mimic (**Supplementary Fig. 2**) to a resolution of 2.7 Å (**Supplementary Table 1**).

The asymmetric unit contained two LlaBIII–DNA molecules related by a twofold rotational symmetry, resulting in a pseudo-continuous DNA of 56 bp (**Supplementary Fig. 3a**). This assembly appeared to be a result of crystal packing and may not be functionally relevant as the targets are oriented tail to tail (which does not support cleavage). The remaining crystal contacts arose from protein-protein and protein-DNA interactions between symmetry-related molecules (**Supplementary Fig. 3b**). Characterization by size-exclusion chromatography (SEC) coupled to multi-angle light scattering (MALS) confirmed that the protein is monomeric (**Supplementary Fig. 3c**).

The structure revealed six structural domains with an unexpected arrangement: the N-terminal Mrr family nuclease followed by the N-core and C-core RecA folds of the ATPase located upstream of the target, and an all α-helical domain (referred hereafter as the coupler) that links the ATPase to the MTase and the C-terminal TRD (**Fig. 1a** and **Supplementary Fig. 4**). The DNA interacted extensively with the TRD and MTase, but made fewer contacts with the ATPase as the DNA length limited the interaction to only a part of the N-core (**Fig. 1a** and **Supplementary Fig. 2**). Among the six domains, quality of electron density for the nuclease was comparatively poor, possibly owing to intradomain and interdomain conformational mobility.

Comparison of the LlaBIII–DNA molecules in the asymmetric unit highlighted interdomain conformational plasticity (**Fig. 1b** and **Supplementary Table 2**). In particular, the nuclease-ATPase domains moved as a unit with respect to the coupler about a hinge-like

interface formed by two laterally packed α-helices, one from the C-core of the ATPase (E649–N656) and one from the coupler (P823–I835) (**Fig. 1b** and **Supplementary Fig. 5**). The HsdM subunit of classical Type I enzymes (PDB identifiers 2AR0 and 3UFB[21]) and the multidomain Type IIG enzyme BpuSI[22] contain a similar all–α-helical domain of unknown function located N-terminally to the MTase domain. The domain has been postulated to be involved in switching between different conformational and functional states[22].

Target recognition and adenine base flipping
In the structure, the target was clamped by the MTase and the TRD (**Fig. 2a**). The MTase of LlaBIII had a structural fold similar to that of the prototype γ-class N6-adenine MTase, M•TaqI[23], and also of the MTase domains of the classical Type I enzymes (PDB identifiers 2AR0, 2Y7C[13] and 3UFB[21]) and the Type IIG enzyme BpuSI[22]. The core region of the C-terminal TRD of LlaBIII was structurally homologous to that of the TRD of M•TaqI[23], TRD1 and TRD2 of a Type I HsdS subunit[24,25], and the TRD of BpuSI[22].

The MTase-TRD clamp deformed the DNA, leading to altered geometrical parameters at the target (**Supplementary Fig. 6**). The grip bent the DNA (**Fig. 2b**), increasing the width and decreasing the depth of the major and minor grooves (**Fig. 2b–d** and **Supplementary Fig. 6**). The changes in the groove geometry facilitated sequence-specific protein-DNA interactions. The MTase read the sequence via the minor groove (**Fig. 2c**), whereas the TRD read it via the major groove (**Fig. 2d**). Despite the absence of AdoMet in the crystals, the target adenine for methylation (position +1; **Supplementary Fig. 2**) was flipped out into the MTase catalytic pocket (**Fig. 2e**). The adenine stacked against Y1021 and hydrogen bonded with the catalytic N1018. F1134 made an energetically favorable T-shaped aromatic-aromatic interaction with the adenine and would sterically block a guanine from entering the pocket. The intercalating R1119 and M1137 plugged the resulting DNA cavity (**Fig. 2e**).

Mechanism of dsDNA translocation
Sequence-specific target binding by the MTase-TRD is essential for initiation of DNA translocation[18,19]. The structure revealed that upon target recognition, the MTase-TRD steered the upstream DNA toward the ATPase, thereby engaging the N-core and C-core (**Figs. 1a** and **3a,b**), illustrating a mechanism that couples target recognition and DNA translocation, mediated by DNA conformational change. Structural comparison with other SF1 and SF2 ATPases revealed that the disposition of the RecA cores bound to dsDNA (**Fig. 3a**) was similar to that seen among ssDNA-bound helicases oriented to bind ATP[26–30] and to the apo structure of the Swi2/Snf2 ATPase from zebrafish Rad54 (ref. 14) (**Supplementary Fig. 7**). In contrast, the orientation of the cores in the only other structure of a helicase-like ATPase bound to dsDNA, i.e., the Swi2/Snf2 ATPase from *Sulfolobus solfataricus*, differs by almost 180° (ref. 31). The relevance of this conformation is unclear[9,31].

The Type ISP N-core–DNA interaction was mediated by a β-hairpin loop (S383–D390) that interacted with the major groove and gripped the 3′-terminated strand. The main-chain amides of K389 and D390 interacted with phosphate at position −11, whereas in one of the monomers of LlaBIII, K385 inserted into the major groove and contacted the base at −13 (**Fig. 3a**). In the other monomer, the lysine side chain was disordered, suggesting that the interactions involving the main-chain atoms were the primary contributors to the grip between the DNA and the hairpin loop. It is, however, likely that new interactions are formed between the hairpin loop and the DNA in other conformational states that may occur during a complete ATPase cycle. Assuming 3′-5′ motion, the LlaBIII ATPase will thus move downstream, consistent with the requirement for head-to-head targets for cleavage (**Supplementary Fig. 1**). A model of
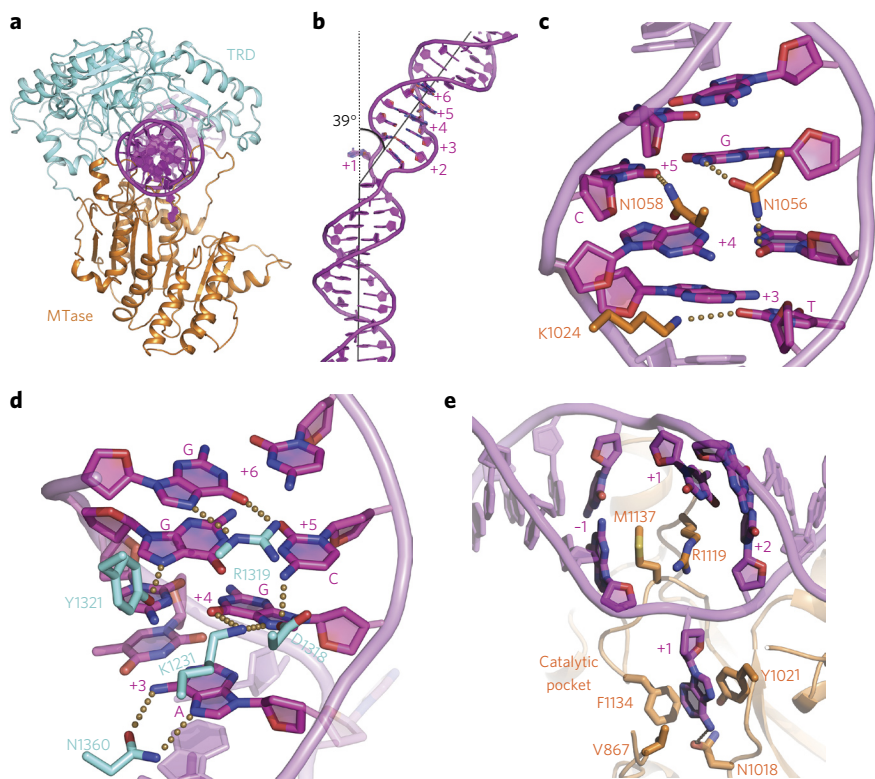
Figure 2 | DNA target recognition by LlaBIII. (**a**) Ribbon diagram of the MTase-TRD clamp encircling the target. (**b**) Extent of (magenta) DNA bending. (**c**) Minor groove readout by MTase (orange). (**d**) Major groove readout by TRD (cyan). (**e**) MTase-DNA contacts stabilizing the flipped adenine. Dotted lines represent interactions within hydrogen bonding distance (≤3.5 Å).

a longer DNA bound to the enzyme, suggested that the ATPase becomes fully engaged when the DNA extends to −22 (**Fig. 3b**). Accordingly, effective downstream DNA translocation (**Fig. 3c**) and cleavage (**Supplementary Fig. 8**) required at least 23 bp upstream of the target.

The structure revealed an enzyme that was competent to carry out either methylation or translocation/cleavage in the presence of the appropriate cofactor: AdoMet or ATP. It appears that the tussle between the two activities is avoided by the apparent rate of methylation being slower than ATP hydrolysis (**Supplementary Fig. 9**). As a consequence, ATPase activity would more often initiate translocation on unmodified foreign DNA, leading to cleavage rather than methylation. On the host DNA, the absolute requirement for two unmodified head-to-head targets for cleavage[6,19,20] ensures that the newly replicated host DNA, with one strand methylated and the other unmethylated, would not be cleaved. After multiple redundant translocation events, maintenance methylation would eventually occur.

Mechanism of long-range communication along DNA

In contrast to expectations of the prevailing model, the location of the Type ISP ATPase upstream of the MTase-TRD–target complex appeared completely inconsistent with Type I–like loop translocation, as downstream ATPase motion would push the MTase-TRD off the target (**Supplementary Fig. 10a**). In addition, collision of two converging Type ISP enzymes at the MTase-TRD interface will leave the upstream nucleases at locations apparently too distant to either interact directly or produce a dsDNA break (**Supplementary Fig. 10a**).

Magnetic tweezers microscope (MTM) assays have been used to distinguish between loop-dependent and loop-independent translocation models for ATP-dependent RM enzymes[32,33]. We used a

similar approach here, in which cleavage of DNA with two head-to-head Type ISP targets would be observed as loss of magnetic bead tracking, whereas loop translocation would be observed as a uniform downward bead movement (**Supplementary Fig. 10b**). Cleavage occurred rapidly after addition of LlaGI and ATP, independent of the DNA stretching force and without any evidence for DNA shortening and hence loop formation (**Fig. 4a–c**). The cleavage rate was nonetheless dependent on ATP concentration (**Fig. 4d,e**), consistent with translocation-driven communication and cleavage. Above the previously measured apparent Michaelis constant ($K_{M,app}$) values for ATP hydrolysis and translocation[18], the cleavage rate was invariant, most likely because events after collision (for example, nuclease activity) are rate-limiting. Below the $K_{M,app}$ values, the cleavage rate decreased rapidly, most likely due to a combination of slower translocation, and a concurrent decrease in processivity (as the translocation rate approaches the DNA dissociation rate), which would necessitate multiple translocation events before successful collision.

Changes in DNA topology

The previous suggestion that these enzymes translocate DNA loops was based on an indirect topoisomerase assay that measured changes in DNA topology during translocation of a nuclease mutant[18]. Wild-type enzyme was not used as it would have rapidly cleaved the DNA substrate. The principle of the assay is that a motor complex that forms a DNA loop and translocates around the DNA helical pitch will trap changes in DNA supercoiling: the twin-domain supercoiling model[11,34]. The upstream DNA loop expands in size and has reduced twist (leading to negative supercoiling) whereas the downstream DNA has increased twist (leading to positive supercoiling). Using a one-site DNA, a LlaGI nuclease mutant and *Escherichia coli* topoisomerase I, we observed conversion of a relaxed DNA to a positively supercoiled one, consistent with twin-domain supercoiling[18].

We re-examined the effect of translocation by a LlaGI nuclease mutant on DNA topology using the MTM assay (**Supplementary Fig. 11a**). On negatively supercoiled one-site DNA at low force, we observed transient increases in DNA height consistent with loop translocation (the plectonemes being released by the increased twist downstream of the motor). In theory, upon reaching relaxed DNA the motor could continue to translocate, leading to formation of positive supercoils and a reduction in bead height. However, we never observed reductions in bead position below the resting height of the negatively supercoiled DNA. This suggests that upon reaching relaxed DNA (or soon before or after), the loop is released, allowing re-equilibration to negatively supercoiled DNA. On positively supercoiled DNA at low force, the bead height transiently reduced, also consistent with loop translocation (the increased twist producing additional plectonemes). However, the events were longer-lived and more frequent.

If loop translocation on a negatively supercoiled DNA produced relaxed DNA, the difference in bead height compared to relaxed DNA in the absence of looping would indicate the length of DNA trapped in the loop[35]. The differences we observed were larger than expected given a model where the motor initiates looping immediately next to the target (**Supplementary Fig. 1**). In **Supplementary Figure 11b**, relaxed DNA should have formed after translocation

of ~168 bp (+16 turns assuming 10.5 bp per turn), corresponding to a height difference of ~57 nm. The larger difference observed (for example, 136 nm) could be explained if LlaGI trapped a larger loop initially. However, we cannot explicitly state that the relaxed DNA state had been reached (i.e., looping may be short-lived and always collapse before reaching 168 bp).

On topologically unconstrained one-site DNA, neither the nuclease mutant nor wild-type LlaGI produced loop-translocation events (**Supplementary Figs. 11c,d** and **12a**). Wild-type LlaGI produced a much lower frequency of loop translocation events on positively supercoiled DNA (**Supplementary Fig. 11d**). Using the wild-type enzyme, we also tested DNA cleavage on two-site DNA (**Supplementary Fig. 11e**). We observed individual strand-cleavage events: the first event relaxed the supercoils, lengthening the DNA, before a subsequent event generated a dsDNA break. However, we did not observe reductions in bead height consistent with twin-domain supercoiling. Therefore it seems that although LlaGI can occasionally trap loops during translocation, these are not prerequisites for cleavage of either supercoiled DNA (**Supplementary Fig. 11e**) or relaxed DNA (**Fig. 4a,b**).

We suggest an alternative model to account for supercoiling changes observed here and previously[18] (**Supplementary Fig. 11f**). An individual enzyme cannot trap a DNA loop. However, where a translocating motor interacts with a static motor, continued translocation leads to twin-domain supercoiling. These interactions are promoted and maintained by DNA supercoils, particularly positive DNA nodes. However, they are facile and are not necessary for DNA translocation or cleavage. The higher event frequency observed with nuclease mutants may reflect changes in motor properties; for the classical Type I RM enzymes, nuclease domain mutations caused decreased translocation and ATPase rates, enzyme population heterogeneity, failure to initiate and altered turnover dynamics[36].

Evidence for release of the target during translocation

Molecular motors have a finite chance of dissociating during each ATP-dependent step along DNA. The forward stepping rate divided by the sum of the stepping rate and the dissociation rate defines the processivity. During loop-independent translocation, the enzyme must leave its target site and subsequent dissociation will release the complete enzyme. This contrasts with translocation termination by the classical Type I RM enzymes, where the motor dissociates from the core MTase, which remains at the site[33]. Translocation termination can be quantified using a triplex-displacement assay. We initiated translocation by mixing ATP with a preformed LlaGI–DNA triplex complex in a stopped flow fluorimeter. We probed the termination pathway by including 'traps' along with the ATP (**Supplementary Fig. 12b**). We trapped dissociated enzymes using an excess of DNA carrying the target sequence or, as a control, a random, nonspecific DNA (**Supplementary Fig. 12c**). Alternatively, we blocked the exposed DNA target by adding a helicase mutant (LlaGI(K210A)), which can bind the site but not translocate[20].

Both traps caused a reduction in the amplitude of triplex displacement, which scaled with the translocation distance (**Supplementary Fig. 12d**). The burst amplitude represented the proportion of prebound LlaGI proteins that initiated translocation and reached the triplex without being trapped. The specific DNA
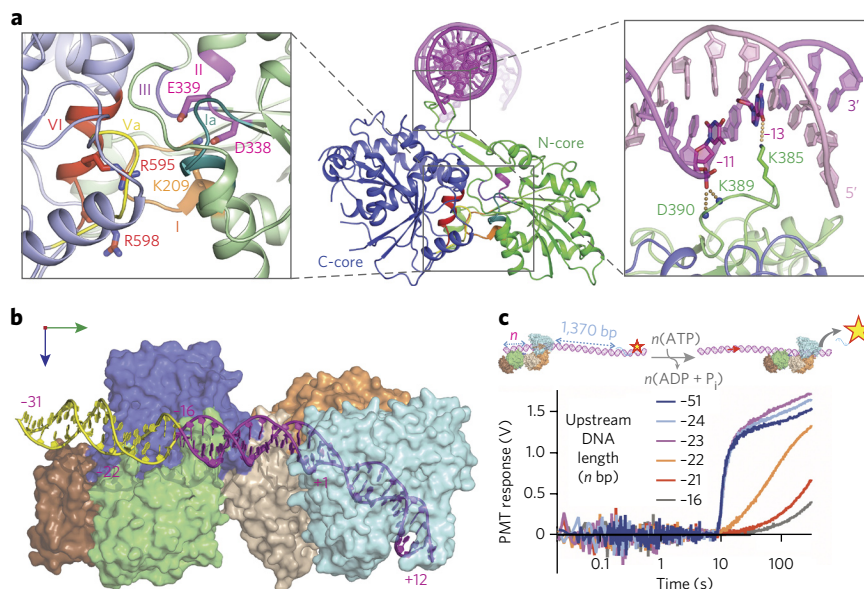

Figure 3 | Architecture and upstream positioning of the ATPase domains. (**a**) ATPase architecture highlighting: conserved SF2 helicase-like motifs, with ATP-interacting residues as sticks (left inset), and DNA interaction of loop S383–D390, with side chain of K385 represented as sticks and main-chain nitrogens of K389 and D390 as spheres (right inset). (**b**) Model of nuclease-ATPase interactions with an extended DNA in yellow. (**c**) Effect of upstream DNA length on triplex displacement. LlaGI motor activity initiated with ATP was monitored on DNA with varying upstream DNA (schematic at top).

was an efficient trap over short time scales, but protein dissociation from the trap allowed reassociation with the triplex substrate, leading to a background triplex-displacement rate. The nonspecific oligoduplex had a minor trapping effect, but turnover was more rapid. The helicase mutant was more efficient in sequestering the site and preventing reassociation by dissociated enzymes (i.e., the background turnover rate was slower). But for technical reasons we could only add an equimolar concentration of mutant to wild-type protein. Therefore, some wild-type protein could rebind the site, and the apparent amplitude was higher than that seen with the DNA trap. Nonetheless, the distance-dependent changes in displacement amplitude seen by trapping either the site or the dissociated enzyme are consistent with release of the target during loop-independent translocation.

Regulation of the nuclease activity

DNA cleavage does not occur during translocation, but is activated upon convergent protein collision[6,18–20]. The key catalytic residues of the Mrr family nuclease[37] were clustered in a putative active site cleft (**Fig. 5a**), which pointed away from the DNA path (**Fig. 1a**). The loop containing the essential D74 (ref. 37) was unstructured (**Fig. 5a**), suggesting a catalytically impaired conformation that becomes activated on engaging with the DNA upon collision. The loop could also be unstructured owing to the absence of divalent cation in the buffer, which is thought to interact with the catalytic triad of D74, D78 and K94 (ref. 37). We note that the side chains of D78 and K94 were also disordered in the structure.

Nuclease activation, along with structuring of the D74 loop, may additionally arise because of ATP-dependent or ATP-independent nucleoprotein rearrangements that engage the active site with the DNA. Simple modeling based on the structures of the SF2 helicase NS3 from hepatitis virus C bound to ssDNA and an ATP analog[30] suggested that ATP-induced domain closure would move the nuclease closer to the DNA path. However, in this orientation, the nuclease active site still pointed away from the DNA, and the active orientation

is possibly only attained upon collision. In addition, pivoting of the nuclease-ATPase unit about the hinge between the ATPase and the coupler (**Fig. 1b** and **Supplementary Fig. 5**) could help engage the nuclease with the DNA. Activation of the upstream nuclease effected by interdomain plasticity was supported by the mapping of a slow, ATP-independent nicking activity, which exclusively targeted the bottom 3′-5′ strand at position −30 (**Supplementary Fig. 13a,b**).

Mechanism of dsDNA break formation

Although activation must occur upon collision, the enzyme architecture suggested that the nuclease domains in a collision complex will be located too far apart to interact directly and, additionally, that the DNA nicks introduced would be too far apart to form a dsDNA break (**Supplementary Fig. 10a**). To investigate the locations of the strand breaks resulting from stochastic translocation and randomized collision, we mapped single DNA cleavage events at different time points during the reaction using a DNA with two head-to-head sites spaced 97 bp apart (**Fig. 5b**). The assay (**Supplementary Fig. 14**) reports either on the outermost cleavage loci for 3′ overhangs or the innermost cleavage loci for 5′ overhangs. The majority (>98%) of collision events produced 3′ overhangs, with an average centroid of collision midway between the sites, as predicted[18] (**Fig. 5b**). The Mrr family nuclease had an unexpected dinucleotide preference (**Supplementary Fig. 13c**), although this did not influence the distributions observed on the DNA used here.

Early (10 s) cleavage events clustered between the sites and in the immediate upstream regions (as defined by **Fig. 5b**). The 3′-3′ distances between the outermost nicks on the two strands revealed a distinct minimum of 30 bp but also a wide range of longer values up to 120 bp, with a median value of 57 bp (**Fig. 5c,d**). The minimum represents the closest approach of two nuclease-ATPase units during collision. The wide range of values observed supports a distant arrangement of nuclease domains. As the reaction progressed, the 3′-3′ distance increased, consistent with DNA end processing as observed in gel assays[6] (**Supplementary Fig. 8**), raising the median value to ~68 bp. In addition, we began to observe cleavage events further upstream than position −30, consistent with an ability of the collision complex to move from its primary location.

DISCUSSION

The data presented here indicate that the upstream ATPase of a Type ISP enzyme remodels its downstream MTase-TRD–target complex to allow target release (**Fig. 6**). Upon binding of ATP, Type ISP domain movement and ensuing rearrangements in the ATPase-DNA interactions, including those involving the β-hairpin loop, would pull the DNA in an upstream direction, leading to movement downstream of the complete enzyme without a necessity for DNA loop formation. This model is quite distinct from that of the classical Type I RM enzymes. We suggest that the coupler will play a key role in transferring conformational strain during initiation. The flipped base seen in the MTase active site would also act as an 'anchor' to prevent target release; this must be returned to its base-paired location in the DNA before initiation can proceed, and this may be accelerated by ATP-induced strain relayed by the DNA and/or coupler.

Remodeling of the MTase-TRD must involve loss of target interactions. However, the remodeled MTase-TRD may remain associated with the DNA to act as a sliding clamp, enhancing processivity. Alternatively, the clamp may open fully, allowing the coupler-MTase-TRD units to swing completely off the DNA. This motion may help explain the ~30 bp closest approach of the nuclease domains (**Fig. 5c,d**).

The Type ISP remodeling activity is somewhat similar to the role of the ATPase subunit of the Type III RM enzymes[12]. However, although those enzymes use a burst of ATPase activity to break the MTase-target interactions, subsequent bidirectional movement
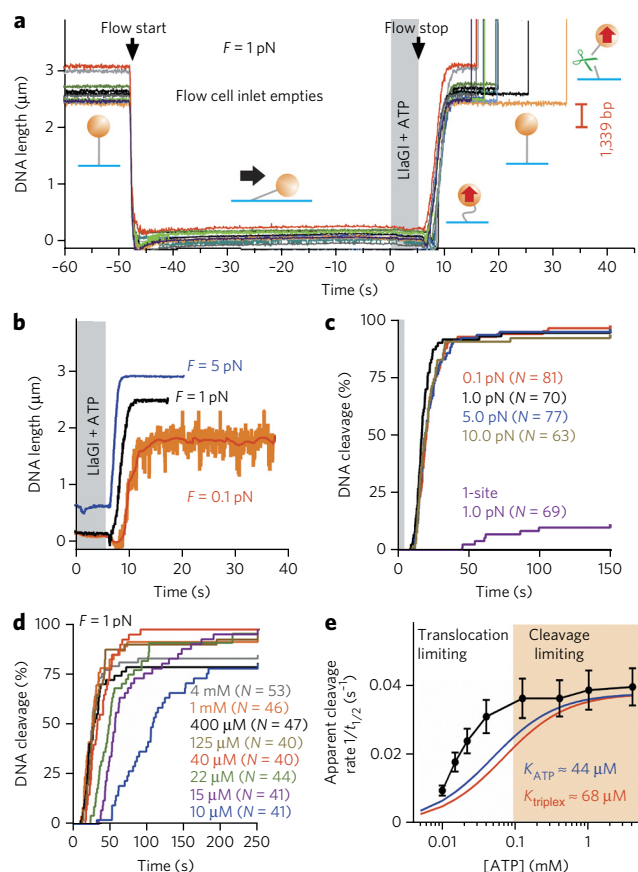
Figure 4 | Loop-independent translocation. (a) DNA cleavage events (upward spikes) after transient introduction of LlaGI and ATP into the flow cell. DNA length change expected of looping is indicated. Lines are raw data (31 Hz). Flow was started as indicated to empty the inlet. Bead positions changed because of flow and cell distortion owing to hydrodynamic pressure. The LlaGI + ATP solution was added and the flow stopped after ~5 s. The DNA relaxed back to full extension slowly because of hydrostatic re-equilibration of the cell. **(b)** Examples of loop-independent cleavage at a range of stretching forces. Lines are raw data (31 Hz) except the red line (1 s filter). **(c)** Cumulative dsDNA cleavage for *N* events (as indicated on the graph) on head-to-head DNA at 0.1–10 pN and on a one-site DNA (**Supplementary Fig. 12**) at 1.0 pN. **(d)** Cumulative dsDNA cleavage for *N* events (as indicated on the graph) on head-to-head DNA at 1 pN at a range of ATP concentrations. **(e)** ATP dependence of the apparent cleavage rate (taken from $t_{1/2}$, the time to reach 50% cleavage in **d**). Blue and red lines are simulations of a hyperbolic relationship for the previously determined apparent $K_{M,app}$ values for ATP hydrolysis (K_{ATP}) and DNA translocation ($K_{triplex}$)[18]. Error bars were calculated from the average values divided by the square root of number of points (*N* values in **d**) in the bin.

along DNA is thermally driven without the need for ATP hydrolysis. In contrast, the Type ISP enzymes continue to consume ATP during unidirectional translocation[18]. It remains to be seen whether the Type III enzymes also move toward their cognate MTase subunits, 'pushing' them off the target as in the Type ISP scheme (**Fig. 6**) or move in the opposite direction to 'pull' the MTase subunits off the target.

The Type ISP structure also provides a simple, single-polypeptide framework to understand other remodeling processes that may involve active disruption of protein-nucleic acid complexes by an ATPase motor, such as the DExH/D RNA helicase[38], SF1 DNA helicase PcrA[39,40], the Swi2/Snf2 family ATPase Mot1 (refs. 15,17) or nucleosome remodelers[15,16]. The above structural comparison
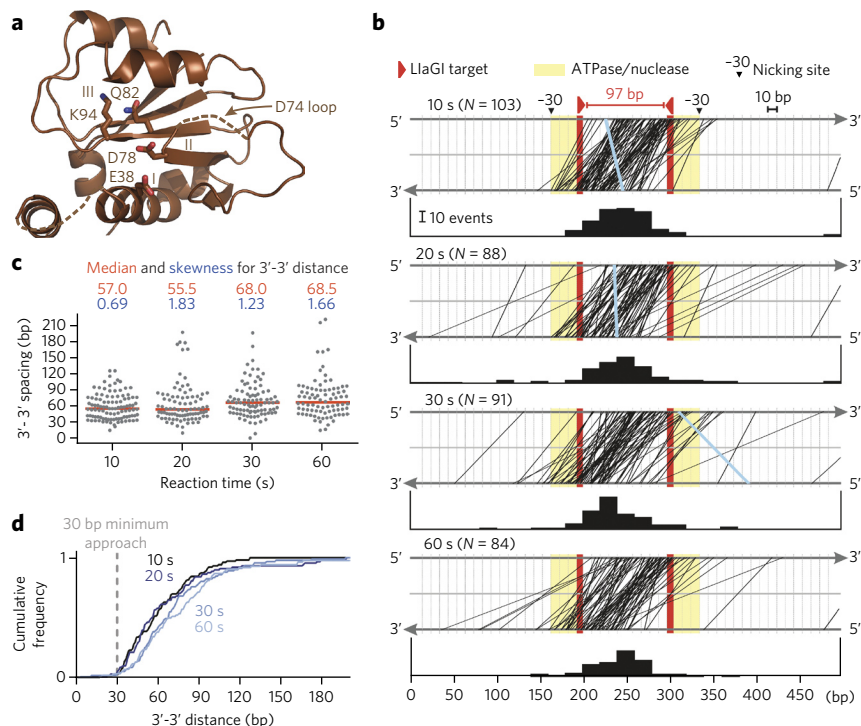
Figure 5 | Single-cleavage-event mapping. (**a**) Mrr family nuclease, highlighting catalytic motifs and residues. (**b**) DNA cleavage by LlaGI at each time point is plotted as top/bottom strand locations joined by a line (rare 5′-5′ overhangs in blue) for *N* events (as indicated on each plot). Bar graphs are the collision centroids for the *N* events in 20 bp bins. (**c**) Scatter plot of the outermost 3′-3′ cleavage distances, with median (red line) and skewness values (the latter showing a time-dependent increase in data asymmetry). *N* values in **b**. (**d**) Cumulative frequency (normalized for *N* events in **b**) of outermost 3′-3′ cleavage distances.

The wide spacing of the DNA strand breaks produced (**Fig. 5**) is consistent with the upstream locations of the nuclease domains as predicted in a head-on collision complex (**Supplementary Fig. 10a**). As the nucleases cannot interact directly, activation upon collision is most likely due to conformational strain from the ATPases transferred via the couplers. Nonetheless, additional remodeling and movement of the collision complex must be required to further process the DNA to produce a dsDNA break. We propose a 'DNA shredder' model in which cumulative nicking events eventually result in a dsDNA break (**Fig. 6**). This contrasts with the precise DNA cleavage produced by two closely spaced nucleases, as suggested for Type I and III enzymes (**Supplementary Fig. 1**) and observed in many other, simpler nucleases.

The collision complex would initially result in the formation of one or two strand breaks. The location of this initial cleavage will be random, dictated by the preceding stochastic translocation events. Rearrangements of the complex due to interdomain plasticity together with the nuclease's sequence preference (**Supplementary Fig. 13c**), can result in varied loci of cleavage and hence a range of initial 3′-3′ spacings. The observed 30 bp minimum could reflect a collision complex where both coupler-MTase-TRD units have swung off the DNA, allowing the closest approach of two helicase-nuclease units that still allows for stress-mediated activation. Continued ATPase activity can then remodel the collision complex, leading to further upstream and downstream movement and nicking events, causing compound damage culminating in a dsDNA break (**Fig. 6**). The increases in the median and skewness values (**Fig. 5**) are thus due to the longer incubation times during which collision complex motion leads to further cleavage events at other sites. Post-collision mobility can also account for cleavage further upstream than position −30 and for rare 5′ overhangs (**Fig. 5b**), where presumably the collision complex nicks one strand and immediately moves in a 3′ direction before nicking the opposite strand.

suggests that the ATPase cycle proposed for bona fide helicases, involving domain closure and opening of N- and C-cores upon ATP binding and hydrolysis[26,27,29,30] is conserved in dsDNA-translocating ATPases. In turn, the Type ISP ATPase would therefore use an equivalent inchworm mechanism for both remodeling and translocation. The β-hairpin loop is unique to Type ISP ATPases (**Supplementary Fig. 7**), indicating variation in the translocation mechanisms among SF2 ATPases despite conservation of canonical motifs and overall structures.

Figure 6 | Model for loop-independent DNA translocation and extensive nucleolytic DNA processing. (**a**) The preinitiation complex. The nuclease is in an inactive conformation. (**b**) The ATPase cycle loosens the MTase-TRD grip on the DNA. (**c**) dsDNA translocation downstream of the target (red). (**d**) Convergence of two enzymes initially brings the nucleases ~75 bp apart. (**e**) Example of stochastic 'DNA shredding' by a collision complex. Numbers are the order of the nicking events.

The loading of multiple translocating enzymes on the DNA that is a potential consequence of target release could result in a pileup of enzymes on both sides of the primary collision complex. If these enzymes were activated, they could cause a time-dependent increase in 3′-3′ cleavage distance. However, post-collision mobility can more readily account for cleavage events upstream of position −30 and the 5′ overhangs. Nonetheless, rear-end collisions may have a supplementary role in DNA processing. We discounted an alternative model where strand nicking by multiple, separate collision events at random sites generates the widely spaced overhangs because this would generate 5′ overhangs with equal frequency and would not show a time-dependent increase in distance.

We speculate that, depending on the cellular host, broken DNA with variable 3′ overhangs produced by Type ISP enzymes would not be a substrate for the corresponding dsDNA break repair enzymes (RecBCD or AddAB), which have a preference for blunt ends[41–43]. As a consequence, even if the foreign DNA were to have the regulatory sequences required for homologous recombination, they would not enter that repair pathway. dsDNA break formation by multiple nicking events would additionally prevent simple religation of the DNA ends. Where the cell also encodes a CRISPR-Cas system, the small DNA fragments generated during cleavage by Type ISP enzymes (or during post-cleavage processing by classical Type I enzymes[44]) may feed into the spacer-acquisition pathway (adaptation), as suggested for the DNA fragments generated by RecBCD[45].

In conclusion, the distinct nucleolytic activity of the Type ISP enzymes contrasts with mechanisms used by dimeric nucleases and illustrates yet another strategy evolved toward resistance in the perpetual bacteriophage-host arms race[2].

METHODS

Methods and any associated references are available in the online version of the paper.

Accession code. Protein Data Bank: 4XQK.

References

1. Dryden, D.T.F., Murray, N. & Rao, D.N. Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res.* **29**, 3728–3741 (2001).
2. Labrie, S.J., Samson, J.E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327 (2010).
3. Linn, S. & Arber, W. Host specificity of DNA produced by *Escherichia coli*, X. In vitro restriction of phage fd replicative form. *Proc. Natl. Acad. Sci. USA* **59**, 1300–1306 (1968).
4. Meselson, M. & Yuan, R. DNA restriction enzyme from *E. coli*. *Nature* **217**, 1110–1114 (1968).
5. Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A. & Wilson, G.G. Type I restriction enzymes and their relatives. *Nucleic Acids Res.* **42**, 20–44 (2014).
6. Smith, R.M., Diffin, F.M., Savery, N.J., Josephsen, J. & Szczelkun, M.D. DNA cleavage and methylation specificity of the single polypeptide restriction-modification enzyme LlaGI. *Nucleic Acids Res.* **37**, 7206–7218 (2009).
7. Murray, N.E. type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.* **64**, 412–434 (2000).
8. Rao, D.N., Dryden, D.T. & Bheemanaik, S. Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.* **42**, 45–55 (2014).
9. Dürr, H., Flaus, A., Owen-Hughes, T. & Hopfner, K.-P. Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures. *Nucleic Acids Res.* **34**, 4160–4167 (2006).
10. Stanley, L.K. *et al.* When a helicase is not a helicase: dsDNA tracking by the motor protein EcoR124I. *EMBO J.* **25**, 2230–2239 (2006).
11. Halford, S.E., Welsh, A.J. & Szczelkun, M.D. Enzyme-mediated DNA looping. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 1–24 (2004).
12. Schwarz, F.W. *et al.* The helicase-like domains of type III restriction enzymes trigger long-range diffusion along DNA. *Science* **340**, 353–356 (2013).
13. Kennaway, C.K. *et al.* Structure and operation of the DNA-translocating type I DNA restriction enzymes. *Genes Dev.* **26**, 92–104 (2012).
14. Thomä, N.H. *et al.* Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. *Nat. Struct. Mol. Biol.* **12**, 350–356 (2005).
15. Hopfner, K.-P., Gerhold, C.-B., Lakomek, K. & Wollmann, P. Swi2/Snf2 remodelers: hybrid views on hybrid molecular machines. *Curr. Opin. Struct. Biol.* **22**, 225–233 (2012).
16. Narlikar, G.J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**, 490–503 (2013).
17. Wollmann, P. *et al.* Structure and mechanism of the Swi2/Snf2 remodeller Mot1 in complex with its substrate TBP. *Nature* **475**, 403–407 (2011).
18. Smith, R.M., Josephsen, J. & Szczelkun, M.D. The single polypeptide restriction-modification enzyme LlaGI is a self-contained molecular motor that translocates DNA loops. *Nucleic Acids Res.* **37**, 7219–7230 (2009).
19. Šišáková, E., van Aelst, K., Diffin, F.M. & Szczelkun, M.D. The type ISP Restriction-Modification enzymes LlaBIII and LlaGI use a translocation-collision mechanism to cleave non-specific DNA distant from their recognition sites. *Nucleic Acids Res.* **41**, 1071–1080 (2013).
20. van Aelst, K., Šišáková, E. & Szczelkun, M.D. DNA cleavage by type ISP Restriction-Modification enzymes is initially targeted to the 3′-5′ strand. *Nucleic Acids Res.* **41**, 1081–1090 (2013).
21. Park, S.Y. *et al.* Structural characterization of a modification subunit of a putative type I restriction enzyme from *Vibrio vulnificus* YJ016. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1570–1577 (2012).
22. Shen, B.W. *et al.* Characterization and crystal structure of the type IIG restriction endonuclease RM.BpuSI. *Nucleic Acids Res.* **39**, 8223–8236 (2011).
23. Goedecke, K., Pignot, M., Goody, R.S., Scheidig, A.J. & Weinhold, E. Structure of the N6-adenine DNA methyltransferase M•TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **8**, 121–125 (2001).
24. Kim, J.S. *et al.* Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc. Natl. Acad. Sci. USA* **102**, 3248–3253 (2005).
25. Calisto, B.M. *et al.* Crystal structure of a putative type I restriction-modification S subunit from *Mycoplasma genitalium*. *J. Mol. Biol.* **351**, 749–762 (2005).
26. Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S. & Wigley, D.B. Crystal structures of complexes of PcrA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**, 75–84 (1999).
27. Lee, J.Y. & Yang, W. UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* **127**, 1349–1360 (2006).
28. Büttner, K., Nehring, S. & Hopfner, K.-P. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* **14**, 647–652 (2007).
29. Saikrishnan, K., Powell, B., Cook, N.J., Webb, M.R. & Wigley, D.B. Mechanistic basis of 5′-3′ translocation in SF1B helicases. *Cell* **137**, 849–859 (2009).
30. Gu, M. & Rice, C.M. Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc. Natl. Acad. Sci. USA* **107**, 521–528 (2010).
31. Dürr, H., Körner, C., Müller, M., Hickmann, V. & Hopfner, K.P. X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**, 363–373 (2005).
32. Ramanathan, S.P. *et al.* type III restriction enzymes communicate in 1D without looping between their target sites. *Proc. Natl. Acad. Sci. USA* **106**, 1748–1753 (2009).
33. Seidel, R. *et al.* Dynamics of initiation, termination and reinitiation of DNA translocation by the motor protein EcoR124I. *EMBO J.* **24**, 4188–4197 (2005).
34. Liu, L.F. & Wang, J.C. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* **84**, 7024–7027 (1987).
35. Seidel, R. *et al.* Real-time observation of DNA translocation by the type I RM enzyme EcoR124I. *Nat. Struct. Mol. Biol.* **11**, 838–843 (2004).
36. Šišáková, E., Weiserova, M., Dekker, C., Seidel, R. & Szczelkun, M.D. The interrelationship of helicase and nuclease somains during DNA translocation by the molecular motor EcoR124I. *J. Mol. Biol.* **384**, 1273–1286 (2008).
37. Smith, R.M., Josephsen, L. & Szczelkun, M.D. An Mrr-family nuclease motif in the single polypeptide restriction-modification enzyme LlaGI. *Nucleic Acids Res.* **37**, 7231–7238 (2009).
38. Jankowsky, E., Gross, C.H., Shuman, S. & Pyle, A.M. Active disruption of an RNA-protein interaction by a DExH/D RNA helicase. *Science* **291**, 121–125 (2001).
39. Park, J. *et al.* PcrA helicase dismantles RecA filaments by reeling in DNA in uniform steps. *Cell* **142**, 544–555 (2010).
40. Fagerburg, M.V. *et al.* PcrA-mediated disruption of RecA nucleoprotein filaments—essential role of the ATPase activity of RecA. *Nucleic Acids Res.* **40**, 8416–8424 (2012).
41. Taylor, A.F. & Smith, G.R. Substrate specificity of the DNA unwinding activity of the RecBC enzyme of *Escherichia coli*. *J. Mol. Biol.* **185**, 431–443 (1985).
42. Dillingham, M.S. & Kowalczykowski, S.C. RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.* **72**, 642–671 (2008).

43. Yeeles, J.T., Cammack, R. & Dillingham, M.S. An iron-sulfur cluster is essential for the binding of broken DNA by AddAB-type helicase-nucleases. *J. Biol. Chem.* **284**, 7746–7755 (2009).

44. Endlich, B. & Linn, S. The DNA restriction endonuclease of *Escherichia coli* B. II. Futher studies of the structure of DNA intermediates and products. *J. Biol. Chem.* **260**, 5729–5738 (1985).

45. Levy, A. *et al.* CRISPR adaptation biases explain preferences for acquisition of foreign DNA. *Nature* **520**, 505–510 (2015).

46. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to K.S. or M.D.S.

ONLINE METHODS

Generation of a LlaGIΔN mutant. pRSFLlaGI[6] was used as template in PCR protocol using Phusion polymerase (Finnzymes) as per manufacturer's guidelines. The phosphorylated primers 5′-CATGGTATATCTCCTTATTAAAGT-3′ and 5′-CGTCCAGAAAATGTAGTCGT-3′ were designed to delete the region of the gene corresponding to amino acids 2–165. Template DNA was removed by digestion with DpnI and clones selected and sequenced after transformation of *E. coli* Top10 cells (Invitrogen).

Purification of LlaGI, LlaGIΔN and LlaBIII for crystallization. LlaGI, LlaGIΔN, selenomethionine containing LlaGIΔN (LlaGIΔN^{Se}) and LlaBIII were purified using the same strategy. The protein was overexpressed in 10 L of *E. coli* BL21 (DE3) from recombinant clones of the genes in pRSF vector. LlaGIΔN^{Se} was overexpressed in *E. coli* B834 (DE3) grown in LeMaster medium[47]. A higher amount of soluble protein in the crude lysate was observed upon inducing the culture at 25 °C with 0.5 mM IPTG at OD$_{600}$ = 0.6. The incubation temperature was lowered from 37 °C to 25 °C before addition of IPTG. Induced cells were harvested after 5 h further incubation. To minimize proteolytic degradation of the protein, protease inhibitors (Roche) were added to the lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM MgCl$_2$, 1mM EDTA and 1 mM DTT), all the steps of purification were carried out at 4 °C, and the protein was purified to homogeneity within 24 h of lysis. The cells were lysed by sonication. The protein was salted-out using 70% w/v ammonium sulfate. The resuspended protein pellet was further purified using heparin and MonoQ columns (GE Healthcare), respectively. Finally, size-exclusion chromatography using Superdex 200 10/300 (GE Healthcare) ensured homogenous monomeric protein. Purified LlaGI and LlaGIΔN were stored in a buffer containing 10 mM Tris-HCl pH 7.4, 100 mM NaCl and 1 mM DTT, and LlaBIII was stored in 10 mM Tris-HCl pH 7.4, 100 mM KCl and 1 mM DTT. Purified samples stored at −80 °C remained intact, and could be thawed and used for crystallization later.

Purification of DNA substrates for crystallization. The dsDNA substrates used for co-crystallization were obtained by annealing two chemically synthesized, PAGE-purified, complementary ssDNA purchased from Integrated DNA Technologies. The dsDNA was purified to remove ssDNA using a MonoQ column. The pure duplex DNA was concentrated and stored in sterile water at −20 °C.

Crystallization of LlaBIII−DNA and data collection. LlaBIII−DNA complex was crystallized in 200-nl drops (1:1 of protein: reservoir buffer) by vapor diffusion at 291 K using sitting-drop method (**Supplementary Fig. 15**). The reservoir buffer contained 200 mM KCl and 20% w/v PEG 3350. This crystal diffracted to 2.7 Å at 100 K with ethylene glycol as a cryoprotectant. Data were collected at the I02 beamline, Diamond Light Source, at a wavelength of 0.9793 Å. The data collection statistics are provided in **Supplementary Table 1**.

Crystallization of LlaGIΔN^{Se}-DNA and data collection. A complex of the purified protein and DNA was formed by mixing the two in 1:1.3 molar ratio at 4 °C. A protein concentration of 5 mg/ml and a crystallization drop size of 200 nl were used for all the initial crystallization trials by sitting-drop vapor-diffusion method. Conditions of varying buffers, additives, precipitants, DNA substrates and temperatures were screened for crystallization of LlaGI-DNA and LlaGIΔN-DNA complex. Crystals of LlaGI-DNA diffracted poorly. However, a crystal form of LlaGIΔN bound to a DNA substrate mimic (**Supplementary Fig. 15**) grown in 4 µl (1:1 of protein: reservoir buffer) hanging drops at 291 K by the vapor-diffusion method diffracted better. The reservoir buffer contained 100 mM Tris-HCl pH 7.4, 20% w/v PEG 20,000, 4% w/v PEG 550 MME and 150 to 250 mM sodium acetate. This condition was used to grow crystals of LlaGIΔN^{Se}-DNA. A 3.2 Å multi-wavelength anomalous data set was collected from one of these crystals at 100 K. Data were collected at I03 beamline, Diamond Light Source.

Structure solution of LlaGIΔN^{Se}-DNA. The data set was processed using XDS[48], and the intensities were scaled and merged using XSCALE[48] and converted to structure factors using TRUNCATE[49] (**Supplementary Table 2**). LlaGIΔN^{Se}-DNA crystallized in the space group P2$_1$ and Matthews coefficient suggested two molecules in the asymmetric unit. Experimental phases to a resolution of 3.2 Å were determined by single-wavelength anomalous diffraction (SAD) method from the data collected at the peak wavelength of 0.9797 Å

(**Supplementary Table 2**). The quality of the anomalous signal was good enough to obtain the substructure of 46 of the 58 selenomethionines in the asymmetric unit using ShelxD as implemented in the program PHASER-EP[50]. Phases were calculated by PHASER-EP. Density modification with solvent flattening and histogram matching was carried out using DM[51]. The quality of the map, though of medium resolution, was sufficiently good to visually identify the correct hand (**Supplementary Fig. 15**). The maps were visualized and model building carried out using Coot[46]. Initial cycles of structure refinements were carried out by REFMAC5 (ref. 52) and subsequently by phenix.refine[53]. This map was used to manually build most of the structure of LlaGIΔN-DNA. This structure was used as a model to solve the structure of LlaBIII-DNA by molecular replacement.

Structure determination of LlaBIII-DNA. The data was processed using XDS[48] and the intensities were scaled and merged using XSCALE[48], and converted to structure factors using TRUNCATE[49]. LlaBIII-DNA crystallized in the space group P1. The structure solution for LlaBIII-DNA was obtained by molecular replacement with MOLREP[54] using the coordinates of an N-terminal deletion mutant of LlaGI. Alternate cycles of model building using Coot[46], and refinement of coordinates and individual B-factors using phenix.refine[53] were carried out to build the structure of the full-length LlaBIII-DNA complex. Hydrogens were automatically added to the model during refinement. The refinement statistics for the structures are provided in **Supplementary Table 1**. The final model had 96% of the residues in the favored, 4% in the allowed and 0% in the disallowed region of the Ramachandran plot. PyMOL (http://www.pymol.org) was used to generate figures.

Size-exclusion chromatography coupled to multi-angle light scattering. The Type ISP enzymes (0.5–1.0 mg/ml) were run in reaction buffer (for LlaGI, 50 mM Tris-Cl, pH 8.0, 10 mM MgCl$_2$, 1 mM DTT; for LlaBIII, 50 mM Tris-Cl, pH 8.0, 10 mM MgCl$_2$, 150 mM KCl, 1 mM DTT) at 1 ml/min on a Superdex 200 analytical size exclusion column (GE Healthcare), attached to a light scattering diode array and a differential refractive index detector (Wyatt Technology). Chromatograms were analyzed using the Astra software (v6.0.5.3, Wyatt Technology). Absolute molecular masses of elutes were calculated from the ratio of light scattering and the differential refractive index. Reproducibility of the column chromatography was checked using bovine serum albumin as in ref. 55.

Modeling studies. The models of LlaBIII bound to longer DNA shown in **Figure 3b** and the collision complex in **Supplementary Figure 10a** were generated using Coot[46]. The former was generated by addition of nucleotides upstream to the co-crystal DNA. The conformation of the nucleotides was approximated to that of a B-form DNA. A model of the collision complex was generated by placing the structure of two molecules of LlaBIII-DNA side by side such that the two TRDs and the downstream ends of the DNA were contiguous. The pseudocontinuous DNA thus formed has two molecules of LlaBIII oriented head to head. The upstream ends were extended by addition of nucleotides having the conformation of B-form DNA.

DNA cleavage assay. To produce the linear substrates used in **Supplementary Figure 8**, pHH-3 (ref. 18) was used as a substrate for PCR with the downstream primer FOli LlaGI-529 and the following upstream primers (**Supplementary Table 3**): ROli LlaGI-16 for upstream DNA length (n) = 16 bp; ROli LlaGI-21 (for n = 21); ROli LlaGI-22 (for n = 22); ROli LlaGI-23 (for n = 23); ROli LlaGI-24 (for n = 24); and, ROli LlaGI-51 (for n = 51). All substrates were purified by phenol-chloroform extraction and ethanol precipitation. Assays contained 2 nM DNA, 4 mM ATP and 200 nM LlaGI in TMD buffer. Reactions were started by adding ATP, and incubated at 25 °C for the durations indicated. Reactions were stopped by adding half volume of STEB (0.1 M Tris (pH 7.5), 0.2 M EDTA, 40% (w/v) sucrose, 0.4 mg/ml bromophenol blue), and samples were analyzed by agarose gel electrophoresis. Gels were stained for 30 min in 0.5 µg/ml ethidium bromide and destained for at least 2 h. The gels were quantified from 8-bit images using ImageQuant TL (GEHealthcare), correcting band intensity for DNA length. Average values and s.d. were calculated in GraphPad Prism.

Triplex-displacement assays. To produce the linear substrate used in **Figure 3c**, pRMA03F[20] was used as a substrate for PCR with the downstream primer FOli LlaGITr-529 and the following upstream primers (**Supplementary Table 3**): ROli LlaGI Tr-16 (for n = 16); ROli LlaGI Tr-21 (for n = 21); ROli LlaGI

Tr-22 (for n = 22); ROli LlaGI Tr-23 (for n = 23); ROli LlaGI Tr-24 (for n = 24); and, ROli LlaGI Tr-51 (for n = 51). To produce the linear substrates used in **Supplementary Figure 12**, either pRMA03R[20] was cut with BlpI (for the 4,663-bp spacing) or pRMA03F was cut with EcoRI (for the 168-bp spacing). All substrates were purified by phenol-chloroform extraction and ethanol precipitation. Triplexes were formed overnight as described previously using a tetramethylrhodamine-labeled triplex-forming oligonucleotide[18]. Reactions were carried out using 1 nM DNA (0.5 nM triplex), 100 nM LlaGI, 4 mM ATP in TMD (50 mM Tris-Cl, pH 8.0, 10 mM $MgCl_2$ and 1 mM DTT) at 25 °C. Fluorescence intensity measurements were performed using an SF61-DX2 stopped-flow fluorimeter (TgK Scientific)[18]. DNA and enzyme were premixed, and the reactions initiated by mixing with ATP. In **Supplementary Figure 12**, translocation traps were added to the ATP syringe, and gave final concentrations of 1 μM oligoduplex or 100 nM LlaGI(K210A)[20].

ATP-independent nick-mapping assay. To produce the linear substrate (**Supplementary Fig. 14a**), pHH-3 was used as a substrate for PCR with the primers FOli LlaGI +75 (^{32}P-labeled for top-strand substrates) and ROli LlaGI −75 (^{32}P-labeled for bottom-strand substrates) (**Supplementary Table 3**). Assays contained 2 nM DNA and 200 nM wild-type LlaGI, nuclease mutant LlaGI(D74A) or nuclease domain deletion mutant LlaGIΔN in TMD buffer. Reactions were started by adding enzyme (or blank dilution buffer), and incubated at 25 °C for 60 min. Samples were stopped using 10 mM EDTA, 1 mM NaOH, 80% (v/v) formamide, 0.1% (w/v) bromophenol blue) and analyzed by electrophoresis on a urea denaturing 8% (w/v) polyacrylamide gel. Sequencing lanes were produced using labeled primers following the USB T7 sequencing kit (Affymetrix) according to the manufacturer's instructions. The pixel positions of the 16-bit scanned wild-type lanes in **Supplementary Figure 13a** were corrected for DNA length using an exponential function calculated based on the positions of DNA bands in the sequencing lanes, according to ref. 19.

Magnetic tweezers assay. pKA20 was generated by QuikChange multi-site directed mutagenesis of pSFV-1 (Life Technologies) using the following upstream primers (**Supplementary Table 3**): KA40F; KA41F; KA42F; KA43F; KA44F; KA45F; KA46F; KA47F to remove LlaGI sites at 131, 1539, 2120, 3637, 3863, 5619, 7252 and 9499 bp, respectively, to leave two head-to-head sites 1339 bp apart. pKA21 was generated by QuikChange mutagenesis of pKA20 to remove the LlaGI site at 2.640 bp using primer pair KA69F and KA69R (**Supplementary Table 3**). The DNA were cleaved using BamHI and SpeI. Biotin- or digoxigenin-modified attachment handles were made using 1.0 kbp DNA fragments that where labeled with biotin- or digoxigenin-dUTP by PCR of pUC19 (ref. 56), using primers ms293SpeF and primers ms293NotR (**Supplementary Table 3**), and which were digested with either BamHI or SpeI. The pKA20 or pKA21 fragment was ligated with the biotin/dig-labeled handles using T4 DNA ligase[56].

Single-molecule experiments were carried out as previously described[32,56] using a commercial magnetic tweezers instrument (Picotwist, Saint Romain de Popey, France, equipped with a Jai CM-140GE camera, image acquisition at 31 Hz)[57]. Fluidic cells were constructed from an uncoated 24 mm ×60 mm coverslip (Menzel-Gläser No. 1), double-sided adhesive tape (3M 467MP, 50 μm depth) and polyester film (Melinex 401, DuPont, 50 μm depth). Anti-digoxigenin (Roche) and BSA were adsorbed directly to the glass by incubation for >3 h at room temperature. Each DNA construct was bound at its biotin-modified end to excess streptavidin-coated magnetic beads (1 μm diameter, MyOne T1, Invitrogen) and added into the fluidic cell to allow the DNA to bind the surface via its digoxigenin-modified end. Beads attached directly to the coverslip were monitored to correct for instrument drift. The three-dimensional position of the magnetic bead and thus the orientation and length of the attached DNA molecule was determined from video images at 31 Hz using real-time 3D particle tracking with sub-nm accuracy[57]. Topologically unconstrained (nicked) and constrained (supercoilable) DNA were identified from rotations curves. The reaction temperature within the cell was maintained by Peltier elements to 25 ± 1 °C.

Experiments with LlaGI were carried out in TMD buffer. In all time trajectories depicted, raw DNA length data taken at the camera acquisition rate is shown, except for 0.1 pN trace in **Figure 4b** where the data was also smoothed with a 1 Hz moving average. Flow was started by manually opening a valve and 10 nM LlaGI and ATP (4 mM except where stated) were added to the flow cell

in-port. After 5 ± 1 s, the flow was stopped by closing the valve and the beads thereafter relaxed back to their stretched lengths. Cleavage times were taken from the initiation of flow to the point where bead tracking was lost. Error were calculated from the average values from the square root of number of points (N) in the bin. For topologically-constrained DNA, supercoiling levels were set by rotating the magnets at low force where the rotation curves are symmetrical.

Single DNA cleavage event mapping. The method was adapted from that of ref. 58 (**Supplementary Fig. 14**). pKA20.100 was made from pKA20 by reverse PCR using primers ROli LlaGI int 0 and FOli LlaGI int +100 (**Supplementary Table 3**), followed by recircularization of the linear product using T4 DNA ligase. PCR was used to amplify the chloramphenicol gene from pACYC184 (ref. 59) using primers KA082F and KA082R (**Supplementary Table 3**). The amplified product was purified following agarose gel electrophoresis using a QIAquick Gel Extraction Kit (Qiagen). Cleavage reactions contained 2 nM pKA20.100, 200 nM LlaGI, 4 mM ATP in TMD. Reactions were initiated by adding ATP, and the reactions at 25 °C stopped at 10, 20, 30 or 60 s by adding one half-volume of 100 mM Tris-Cl, pH 8.0, 100 mM EDTA. Cleaved DNA was purified using a QIAquick PCR Purification Kit (Qiagen), and treated with 1.5 units T4 DNA polymerase (New England Biolabs) per nM DNA and 100 μM dNTPs in the supplied reaction buffer supplemented with 100 μg BSA at 12 °C for 30 min. The reaction was stopped by heating to 75 °C for 30 min. Blunt-ended DNAs were purified using a QIAquick PCR Purification Kit and ligated to the chloramphenicol gene insert using T4 DNA ligase (New England Biolabs) at 16 °C for 16 h. *E. coli* TOP10 cells (Life Technologies) were transformed with the ligations and single colonies selected by overnight growth on LB agar plates with 50 μg/ml ampicillin and 34 μg/ml chloramphenicol. Plasmid DNAs representing single cleavage events were purified from single colonies and sent for DNA sequencing (Dundee Sequencing Service, Dundee, Scotland) using the primer KA050F (**Supplementary Table 3**). Data in **Figure 5c,d** and **Supplementary Figure 14d** are from three independent repeat cleavage reactions (average values and s.d. were calculated in GraphPad Prism). 6–12% of clones had deletions of >225 bp that could have resulted from illegitimate recombination[58], and were thus excluded from the analysis. Clones without a sequence deletion or duplication, characteristic of a blunt end (**Supplementary Fig. 14c**), were not found. Note, however, that our model (**Fig. 6**) does not exclude the possibility of strand breaks occurring immediately opposite one another in rare instances. Median and skewness values in **Figure 5c** were calculated using Excel (Microsoft).

47. LeMaster, D.M. & Richards, F.M. NMR sequential assignment of *Escherichia coli* thioredoxin utilizing random fractional deuteriation. *Biochemistry* **27**, 142–150 (1988).

48. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).

49. French, S. & Wilson, K. On the treatment of negative intensity observations. *Acta Crystallogr. A* **34**, 517–525 (1978).

50. Read, R.J. & McCoy, A.J. Using SAD data in Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 338–344 (2011).

51. Cowtan, K.D. & Zhang, K.Y.J. Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* **72**, 245–270 (1999).

52. Murshudov, G.N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).

53. Afonine, P.V. *et al.* Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 1153–1163 (2010).

54. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25 (2010).

55. Butterer, A. *et al.* type III restriction endonucleases are heterotrimeric: comprising one helicase-nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Res.* **42**, 5139–5150 (2014).

56. Szczelkun, M.D. *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. USA* **111**, 9798–9803 (2014).

57. Lionnet, T. *et al.* Magnetic trap construction. *Cold Spring Harb. Protoc.* **2012**, 133–138 (2012).

58. Jindrova, E., Schmid-Nuoffer, S., Hamburger, F., Janscak, P. & Bickle, T.A. On the DNA cleavage mechanism of type I restriction enzymes. *Nucleic Acids Res.* **33**, 1760–1766 (2005).

59. Chang, A.C. & Cohen, S.N. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **134**, 1141–1156 (1978).