Molecular Mechanism of CRISPR Interference in Type I-C CRISPR-Cas system

A thesis submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy

by

Siddharth Nimkar

Registration no: 136106023

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

June 2021

I dedicate this thesis to

my parents, relatives, friends and teachers.

Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

Statement

I do hereby declare that the matter embodied in this thesis entitled "**Molecular Mechanism of CRISPR Interference in Type I-C CRISPR-Cas system**" is the result of work carried out in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, under the supervision of **Dr B. Anand**.

In keeping with the general practice of reporting scientific observations, due acknowledgement has been made wherever the work described is based on the findings of other investigators.

Siddharth Nimkar

136106023

Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

Certificate

It is certified that the work described in this thesis entitled "**Molecular Mechanism of CRISPR Interference in Type I-C CRISPR-Cas system**" by **Mr Siddharth Nimkar** for the award of the degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision in the Department of Biosciences and Bioengineering, IITG. The work embodied in this thesis has not been submitted elsewhere for a degree.

Dr B. Anand

Thesis Supervisor

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Assam 781039, India

Acknowledgement

The work presented in this thesis is made possible by the efforts and support of several people to whom I am extremely grateful. Hence, I would like to offer my gratitude to all those who helped me achieve this incredible goal.

I would like to express my gratitude towards the head of the MAB Lab and my thesis supervisor, Dr B. Anand, who gave me the chance to pursue my dream. I am fortunate to experience his wealth of knowledge and receive his guidance that I could not have gotten under the direction of anyone else. I am privileged to receive the valuable scientific insights and life lessons that will prove beneficial for my future endeavours.

I would also extend my gratitude to the members of my doctoral committee, Prof. Kannan Pakshirajan (Chairperson), Dr Nitin Chaudhary and Dr Lal Mohan Kundu, for their unbiased opinions and valuable suggestions that have been the driving force required for the completion of my thesis. In addition, I thank them for their invaluable advice during annual progress seminars. I would also like to thank Dr Vishal Trivedi for his guidance and suggestions towards building my career.

I would also like to extend my acknowledgement towards Prof. Arun Goyal, Prof. V Venkata Dasu, Prof. Kannan Pakshirajan and Prof. Latha Rangan, Department of Biosciences and Bioengineering, IIT Guwahati, for lending their support as respectable Heads of the Department during my PhD tenure.

I owe my thanks to the Department of Biosciences and Bioengineering and Central Instrumentation Facility, IIT Guwahati, for providing the necessary research facilities to accomplish my PhD thesis objectives. I want to extend my thanks to all the Department of Biosciences and Bioengineering staff members for providing me with the logistical support essential to perform my research. I thank Mr Nurul Islam, Mrs Prarthana Swargari for helping in procuring consumables required to carry out my research. I thank Mr Niranjan Barah, Mr Chandan Kumar Nath and Mr Dipankar Barman for providing technical assistance. Further, I would like to acknowledge IIT Guwahati, Ministry of Human Resources Development, India and Department of Biotechnology, Government of India, for financial assistance and funding towards my PhD project. I also acknowledge the geniality of all the investigators who shared their plasmids and bacterial strains either directly or through the Addgene repository.

Even when life hits rock bottom, every situation should be considered an opportunity to be your best. I firmly believe all my lab members made me feel the same way. I would like to thank Dr Himanshu Sharma, Dr Yoganand KNR, Dr Ankita Punetha, Sunanda Chhetry and

Manasasri Murlidharan for the beautiful memories we made together. We have seen ups and downs during our tenure and tackled most of the problems together. I wish a bright future to all of them. I would also like to acknowledge my lab mates Rohan and Neha for their support and encouragement. I will always cherish the memories I have with all my past and present lab members.

Friendship is one of the beautiful emotions, and I am glad to have found true friendship at IIT Guwahati. I feel fortunate to have found Dimple as she was constantly on my side during joyful and sad phases. With her love and care, she was a beacon of light in my journey towards achieving this goal. I was also privileged to receive a lot of care, love, support, criticism and knowledge from Arun. Lastly, I feel lucky that somewhere along the way, I found my best friend in Payel. I will never forget the happy time, laughter, love and friendship I shared over the years with her. I would also like to thank Vartika, Debika, Ganesh, Shrikanth, Dhana, Ruchira, Ruchika, Mohan and Balwant for their help as a friend.

Lastly, I would like to thank my Mother, Father, Sister, Uncle, Aunt, and family for being the pillars of strength during my PhD. I am fortunate to have such a brave mother. She, despite ailments, supported and encouraged me to complete my PhD. The numerous personal sacrifices made by my family have enabled me to reach this juncture in life.
I also express my gratitude to all others whom I may have missed.

Thank you,

Siddharth Nimkar

Contents

Table of Figures

Abbreviations

Cas – CRISPR associated protein

MGEs – Mobile genetic elements

dsDNA – Double-stranded DNA

ssDNA – Single-stranded DNA

ATP – Adenosine triphosphate

ADP – Adenosine diphosphate

GTP – Guanosine triphosphate

NTP – Nucleoside triphosphate

AMP-PNP – adenylyl imidodiphosphate

6-FAM – 6-Carboxyfluorescein

CTD – C-Terminal Domain

WT – Wild type

EtBr – Ethidium bromide

PAGE – Polyacrylamide gel electrophoresis

SEC – Size exclusion chromatography

Kan^R – Kanamycin Resistance

$Spec^R$ – Spectinomycin Resistance

Amp^R –Ampicillin Resistance

Chapter 1: Introduction

1.1 Introduction

Life flourishes in a microbial-driven world, where the earth's initial hostile environment was made hospitable by archaea, bacteria and viruses by crafting the conditions suitable for eukaryotes to evolve. These microorganisms code for proteins involved in various biochemical pathways, creating a biochemical balance in the ecosystem (Field et al., 1998).

Viruses are the most abundant biological entities on the earth, and they impart a critical ecological and evolutionary role. All cellular organisms harbour a unique repertoire of viruses. Historically, three distinct hypotheses for viral evolution are considered (Forterre, 2006). According to the 'primordial virus world' hypothesis, viruses are the direct descendants of the first replicating molecules. In contrast, 'reductive virus origin' explains that the viruses result from the degradation of the ancestral cellular material rendering the viral particle an obligate parasite. Finally, the origin of viruses is defined in terms of the 'escaped gene' hypothesis, according to which the genes from bacteria or archaea escaped and gave rise to viruses.

Bacteria and Archaea have co-existed with bacteriophages (viruses) for billions of years, and thus they interact with each other routinely. Throughout evolution, bacteria are under the attack of parasitic bacteriophages. Bacteriophages lack cellular machinery to replicate and synthesise their nucleic acid or protein and thus hijack the host bacterial system. During this process, often bacteriophage ends up killing its bacterial host.

Due to the immense evolutionary pressure imposed by bacteriophages on bacteria, the complexity and diversity of anti-phage mechanisms developed by bacteria are astonishing. We have only recently begun to understand the complexities of the interaction between bacteria and bacteriophages. The anti-phage mechanism study has been used as a molecular biology tool, such as restriction enzymes for cutting DNA and other gene-editing techniques.

According to the red queen hypothesis, the living organism must continuously evolve to maintain its relative population. Taking a constant tug-of-war between bacteria and viruses into consideration, bacteria must always acquire/evolve anti-phage mechanism. On the other hand, phages are in a constant struggle to overcome such bacterial anti-phage mechanisms. This evolutionary arms race between bacteria and phages forced bacteria to produce a vast assembly of anti-phage systems referred to as the 'prokaryotic immune system'. Two of the highly studied bacterial immune system are restriction-modification system (R-M system) and abortive infection (Abi) systems. More recently, the research focus has been shifted to the

CRISPR-Cas system. This research on the bacteria has concluded that the prokaryotic immune system is much more complicated than previously believed. Apart from these, numerous new defence systems are being discovered regularly, unravelling the previously unknown prokaryotic defence systems.

An individual species of bacteria can encode several defence systems, and it has also been shown that such systems can be transferred horizontally from one species to another in a relatively shorter time scale. One of the most powerful drivers for horizontal gene transfer is transposable elements and viruses. Their capability to replicate, edit, change and transfer genetic information is an instrument for generating genetic diversity.

1.1.1 Defence Islands

The arms race between bacteria and phages has led to the evolution of several defence systems in bacteria and phages. The analysis of genes responsible for the anti-phage mechanism in bacteria has suggested that the defence genes and mobilomes (transposons) are frequently found to be clustered in genomic islands. These defence islands contain several operons encoding different proteins, and a few of them have already been established to be involved in anti-phage defence. The presence of an unknown gene in the defence island suggests that it may also participate in the defence against the phages. For example, genes for restriction-modification (R-M) system, toxin-antitoxin system, abortive infection are usually found on defence islands. The clustering of defence genes suggests functional cooperation between defence systems. However, it remains unclear to what extent these defence systems are associated with each other. Moreover, the correlation between the defence genes and the mobilomes strongly suggests the critical role performed by horizontal gene transfer (HGT) in the evolution of these islands. The diversity present in prokaryotic defence systems is explained in detail in the following section.

1.1.2 Innate Immune Systems

Immune systems are broadly classified into two groups: Innate immune systems and Adaptive immune systems. The critical difference between innate and adaptive immune system is that innate immune systems are present from birth, and they do not keep records of previous infections, which makes innate immunity non-specific. On the other hand, adaptive immune

systems remember previous infections and use this memory to gain leverage against invading phages. Few of the innate immune systems are described below.

1.1.2.1 Blocking phage adsorption

Phage invasion begins with recognising a specific surface receptor molecule, typically a protein, polysaccharide, or lipopolysaccharide (LPS) and attachment of phage particle to the cell wall of bacteria. Subsequently, the phage genome is injected inside the bacteria. Consequently, Bacteria attempt to use a phage specific mechanism to avoid viral DNA entry by modifying surface receptors. Receptor molecules must not only be present on the surface, but they should also be accessible to the phage particle. Thus, bacteria try to change the molecular structure of the receptor through mutations or by covering the receptor with additional barriers (Dy et al., 2014; Labrie et al., 2010). For example, *Bordetella bronchiseptica* has two distinct phases, Bvg^+ and Bvg^-. Bvg^+ phase is required for pulmonary colonisation, and during this phase, bacteria express several surface proteins, including pertactin. Phage particles attaching to pertactin have been identified. The absence of pertactin in Bvg^- strain hinders phage attachment on the bacterial surface (Liu et al., 2002). In *Escherichia coli*, the K1 capsule has been shown to obstruct the attachment of the T7 phage particle to one of its LPS receptors (Scholl et al., 2005).

1.1.2.2 Preventing the injection of DNA

After successful attachment of phage onto the bacterial surface, superinfection exclusion (Sie) systems can block the entry of phage DNA into the bacterial cytoplasm. Sie systems are phage derived and work to avoid the entry of other phages, thereby conferring protection to already lysogenised host. For example, an *E. coli* phage HK97 manufactures gp15 transmembrane protein, which stops the entry of other HK97 and one other closely related phage, HK75 (Cumby et al., 2012).

1.1.2.3 Restriction-Modification systems

Even after the successful insertion of viral DNA inside the cytoplasm, there are several mechanisms deployed by the bacteria in order to prevent further damage. One such mechanism

is the restriction-modification (R-M) system which can destroy the invading DNA. A typical R-M system includes a restriction endonuclease (REase) and a related methyltransferase (MTase) (Tock and Dryden, 2005)

An R-M system recognises short DNA sequences in the range from 4 to 8 bases pairs. Since these sequences may be present in both the host and the attacking viruses, the host uses MTase to methylate and modify its DNA. Host Restriction endonucleases do not recognise such self-methylated DNA. Since the attacking phage DNA is not methylated, it is cleaved upon entry into the cytoplasm. R-M systems are diverse, and they are categorised into four groups based on sequence recognition and cleaving mechanism (Roberts et al., 2003).

Based on subunit composition, ATP/GTP requirement and mechanism of cleavage, R-M systems are classified into four types (I-IV) (Pingoud et al., 2016; Roberts et al., 2003; Roberts et al., 2007). Type I R-M system is complex and contains three subunits; R (restriction), M (modification) and S (specificity). The R subunit contains an RNase domain and a superfamily II helicase domain (ATP-dependent helicase). The M subunit is a methylase that methylates the host nucleic acid. (Roberts et al., 2003). Type II system is the simplest of all the R-M systems and has been widely used in applications. Type II system consists of a REase, and MTase typically encoded within the same operon (Ershova et al., 2012). Type III system is similar to type II system and contains only two subunits, R and M. Like type I system, type III system also contains an ATP dependent helicase domain fused to R subunit (Butterer et al., 2014; Rao et al., 2014). Type IV R-M is different from others, as it contains only an endonuclease (R). Modification enzymes are not part of the R-M system. The R subunit in the type IV R-M system is a GTPase and endonuclease (Bourniquel and Bickle, 2002; Loenen and Raleigh, 2014).

1.1.2.4 Abortive infection

The majority of the protective mechanisms against phages aim at increasing the survival of the infected bacteria. In contrast, abortive infection (Abi) is a process by which the bacterial cells avoid the spread of phage particles by sacrificial death. Phage infected bacterium undergoes altruistic death to avoid the spread of phages to the surrounding bacterial cells. Cell death is either achieved through changes in the cell membrane or changes in internal cellular processes such as transcription and translation. However, the exact mechanism of the Abi

process is poorly understood. Abi systems are associated with mobile genetic elements that include prophages and plasmids (Samson et al., 2013).

In *Lactobacillus lactis,* several Abi systems from AbiA to AbiZ have been characterised (Chopin et al., 2005). AbiZ stimulates premature lysis of cells, leading to the incomplete assembly of viral particles and aborting further viral infection (Durmaz and Klaenhammer, 2007). Another example of an Abi system is found in *Staphylococcus epidermidis*, where a serine/threonine kinase Stk2 phosphorylates several molecular targets in cellular pathways such as transcription, translation and replication. Such inactivation of several metabolic pathways by phosphorylation leads to cell death (Depardieu et al., 2016).

1.1.2.5 Toxin-Antitoxin systems

Bacterial Toxin-Antitoxin (TA) systems are abundantly found genetic elements that protect against invading phages and plasmids. The TA module encodes a toxin that inhibits cell proliferation by impeding vital cellular processes and an anti-toxin that neutralises the toxin (Page and Peti, 2016; Unterholzner et al., 2013). The mechanism of action of toxin is not the same for all bacteria. A toxin can be a DNase or an RNase, and it can inhibit the replication of DNA or halt the synthesis of proteins. Six different types of TA systems have been discovered, and some inhibit the phage life cycle. For example, the TA system in *Pectobacterium atrosepticum* encodes an endoribonuclease ToxN (toxin). A non-coding RNA, ToxI (anti-toxin), blocks the activity of ToxN by binding to it. If there is a phage attack, the phage and the host RNA is destroyed by ToxN RNase activity. Subsequently, the host and phage are unable to survive, which arrests the infection (Blower et al., 2011; Fineran et al., 2009). Similarly, *E. coli* encodes MazF ribonuclease (toxin), which can suppress phage infection by cleaving phage nucleic acid. MazE (anti-toxin) suppresses the activity of MazF in the absence of a phage attack (Engelberg-Kulka et al., 2005).

1.1.2.6 Argonaute

The argonautes (Ago) were initially described in RNA interference (RNAi) in eukaryotes (eAgo) (Bohmert et al., 1998; Tabara et al., 1999). Later homologos of eAgo proteins were discovered in bacteria and archaea (pAgo) (Aravind et al., 2000). The structure of pAgo suggests that it contains PIWI, PAZ and MID domains responsible for nucleic acid

binding and cleavage (Wang et al., 2008; Yuan et al., 2005). Initially, the role of pAgo was poorly understood, but their presence in the defence island suggested their role in the defence (Makarova et al., 2009). Later it was experimentally established that pAgo is involved in small single-stranded RNA/DNA-guided recognition and nuclease-mediated cleavage of foreign genetic material (Makarova et al., 2009; Sheng et al., 2014; Swarts et al., 2015; Willkomm et al., 2016; Yuan et al., 2005). Some species harbours a nuclease-inactive form of pAgo, which, along with microRNAs, suppresses the translation of target mRNA without cleaving it (Hutvagner and Simard, 2008). For example, *Thermus thermophilus* (TtAgo) protects against foreign nucleic acids using DNA-guided DNA interference (Swarts et al., 2014).

1.1.2.7 Recently discovered bacterial defence systems

Recently a comprehensive study of defence islands in several bacteria and archaea has led to the discovery of many previously unknown defence systems. A few of them are described below.

1.1.2.7.1 The Gabija

This system consists of a DNA helicase (GajB) and an ATP dependent nuclease (GajA) which are arranged in an operon *gajAB* (Doron et al., 2018). Since this operon is found in defence island, it is thought to be involved in anti-phage defence. In most bacteria, *gajA* and *gajB* genes are found together with few exceptions. In few strains of *Ralstonia solanacearum* (HA4-1, IBSBF1503, CMR15, CFBP3059, and UW386), GajA and GajB genes are dispersed across the genome, which makes its role in anti-phage defence unlikely, at least in certain strains of *R. solanacearum* (Castillo et al., 2020; Doron et al., 2018). However, a thorough study is required to establish the role of the *gajAB* operon in anti-phage defence.

1.1.2.7.2 The Hachiman

The Hachiman operon *hamAB* translates to a helicase (HamB) and a protein of unknown function (HamA, Pfam ID 08878).

1.1.2.7.3 The Zorya

Based on the composition of the genes, there are two types of Zorya systems. Type I Zorya comprises of *zorABCD* operon. Proteins ZorA and ZorB are homologs of inner membrane protein MotA and MotB, which are part of the flagellar motor of bacteria. Protein *zorC* contains a domain of unknown function (DUF), whereas zorD is a helicase. The Zorya

type II lacks *zorC* and *zorD* and contains an additional HNH nuclease *zorE*. The presence of membrane proteins ZorA and ZorB suggest membrane depolarisation and cell suicide during phage invasion (Doron et al., 2018).

1.1.2.7.4 The Thoeris

The first gene in the Thoeris system is *thsA*. This gene is identified as a nicotinamide adenine dinucleotide (NAD) binding domain in some organisms, whereas sometimes it is annotated as sirtuin (SIR-like domain or Macro domain). The second gene is *thsB* which has a Toll-interleukin receptor (TIR) domain. TIR domain is an essential part of some innate immune systems in plants and animals where it works by transferring molecular signals after the cell has detected a pathogen. The presence of the TIR domain implies that the Thoeris system should also be involved in delivering immune signals upon phage attack (Doron et al., 2018).

1.1.2.7.5 The Druantia

The Druantia is characterised by a large *DurA* gene having a DUF domain and several other genes of unknown function (*DurB, DurC*, *DurD*). A second gene present in this system is a helicase with Walker A/B motif, which suggest its dependence on ATP. The total size of this system is typically 12kb, which exceeds most other defence systems. Therefore, the mechanism of action of this system is thought to be different from the existing defence systems in prokaryotes (Doron et al., 2018).

Apart from the above mentioned recently discovered defence systems, several operons are found on the defence islands. For example, the Shedu system consists of an endonuclease (*SduA*) (Doron et al., 2018). The Septu system has an ATP dependent helicase (*PtuA*) and an HNH nuclease (*PtuB*) (Doron et al., 2018). The Lamassu system and the Wadjet system show several genes with the DUF domain and/or a helicase (Doron et al., 2018). The recently discovered genes on the defence islands expand our understanding of the bacterial arsenal of defence. However, a systematic study is required to unravel their mechanisms.

1.1.3 The Adaptive Immune Systems

The hallmark of the adaptive immune system is the presence of immunological memory after the first pathogen encounter, which leads to an enhanced immune response on the subsequent invasion by the same pathogen. In bacteria, the CRISPR-Cas system employs a unique defence mechanism where a portion of foreign DNA is incorporated as immunological

memory. In the next step, the same portion of integrated foreign DNA helps the host to evade the subsequent attack. Hitherto, the CRISPR-Cas system is the only known adaptive immune system in prokaryotes.

1.1.3.1 The discovery of the CRISPR-Cas system

The first report of unknown repeat sequences adjacent to the alkaline phosphatase (*iap*) gene in *E. coli* was published in a CRISPR unrelated work in 1984 (Ishino et al., 1987). Fifteen years later, in the year 2002, *in silico* analysis of the genomes of prokaryotes (both bacteria and archaea) showed repeating DNA sequences, which were not present in eukaryotes or viruses. These repeats vary in size from 21 to 37 nucleotides and are interspaced by similarly sized non-repetitive DNA sequence. To highlight this distinguishing feature, this portion of the genome was referred to as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) (Jansen et al., 2002). In the same study, four CRISPR-associated (*cas*) genes were identified in the CRISPR loci of bacteria. These genes are absent in the bacteria lacking CRISPR loci. The presence of *cas* genes (*cas1*, *cas2*, *cas3* and *cas4*) in the vicinity of the CRISPR locus indicated the functional relationship between the CRISPR locus and *cas* genes (Jansen et al., 2002). Further analysis showed the presence of a superfamily II helicase (Cas3'), HD-type phosphohydrolase (Cas3") and RecB family exonuclease (Cas4), among others in the *cas* operon, suggesting that these might be involved in DNA repair (Makarova et al., 2002), chromosome partitioning (Mojica et al., 1995) or regulation of gene expression (Jansen et al., 2002). With the revolution in genome sequencing methods, numerous bacterial and viral genomes were sequenced, and their sequences were freely available for analysis. Comparative genome analysis of CRISPR loci in the bacterial and the phage genomes by three independent studies showed that the unique non-repetitive DNA sequences are homologous to DNA sequences in foreign plasmids and phages or mobile genetic elements (MGEs) (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005).

1.1.3.2 Comparison between the CRISPR-Cas system and RNA interference

Further analysis suggested several similarities between CRISPR loci and the eukaryotic RNA interference (RNAi) system (Makarova et al., 2006). The critical evidence showing the similarity between the system emanates from the functional role of proteins in CRISPR loci and RNAi protein components, particularly RISC (Filipowicz, 2005; Sontheimer, 2005; Tang, 2005). The functional part of RISC entails a helicase-nuclease protein, dicer, exonuclease, slicer, argonaute and sometimes an RNA binding protein (Hammond, 2005; Miyoshi et al., 2005; Scadden, 2005; Sontheimer, 2005). Functional analogs of these proteins can be recognised in CRISPR loci as well. It was initially thought that dicer is an analogue of a superfamily 2 helicase fused to HD family nuclease (Cas3). Similarly, the RecB family nuclease (Cas4) present in the CRISPR loci was believed to play the functional role of the slicer in the CRISPR system. The RNAi pathway in eukaryotes uses short interfering RNA (siRNA), microRNA (miRNA) or piwi-interfering RNA (piRNA) as a guide to locate the target nucleic acid (Cagan and Ready, 1989; Carthew and Sontheimer, 2009; Talansky and Gordon, 1988). Similarly, in the CRISPR system, short RNA called CRISPR RNA (crRNA) helps Cas proteins to locate the target DNA/RNA. Both siRNA and crRNA are derived from invading nucleic acids and are processed from a longer parent RNA (Brouns et al., 2008; Carte et al., 2008; Carthew and Sontheimer, 2009). Both RNAi and CRISPR pathways are adaptive and target nucleic acids in a sequence-dependent manner (Andersson and Banfield, 2008; Barrangou et al., 2007). Even though both the systems function alike, there are a few dissimilarities between the two systems. Unlike siRNA and miRNA, crRNA is matured from a single-stranded precursor, and the amount of crRNA is not amplified after transcription (Brouns et al., 2008; Carte et al., 2008; Hale et al., 2008; Marraffini and Sontheimer, 2008). RNA from RNAi is directly borrowed from invading RNA, whereas crRNA is derived from a previously integrated portion of the invading nucleic acid (Andersson and Banfield, 2008; Barrangou et al., 2007).

1.1.4 CRISPR-Cas overview

The CRISPR-Cas system is exclusively present in prokaryotes. Nearly 50% of sequenced bacteria and >80% of sequenced archaea harbour the CRISPR-Cas system (Grissa et al., 2007). However, recent metagenomic data from a groundwater sample suggests that CRISPR-Cas may be present in only 10% of the bacterial and archaeal species, which is lower than previously thought (Burstein et al., 2016). A typical CRISPR consists of a repeating unit of

short palindromic repeats (25-40 nt) and interspaced by similarly sized spacers (Figure 1.1). Usually, spacers are mapped to foreign DNA (homologue to plasmids and viruses), but a few spacers matching the host genome have also been reported in some bacteria (Stern et al., 2010). The number of spacers in a CRISPR array varies from as low as two to several hundred. CRISPR array is preceded by a leader sequence (Figure 1.1), which holds a promoter for the transcription of the CRISPR array to form a precursor CRISPR RNA (pre-crRNA). The leader region is an A-T rich region and harbours binding sites for regulatory elements (Hale et al., 2012; Lillestol et al., 2009; Pougach et al., 2010; Pul et al., 2010).

A CRISPR-Cas system functions in three stages (Figure 1.1): (1) The first stage is the Adaptation stage, where a portion of the foreign genetic element is identified and incorporated inside the host bacterial genome. This acquired DNA portion acts as memory during subsequent infection; (2) The second stage is the expression/maturation stage, during which the CRISPR array is transcribed into pre-crRNA, which is ultimately processed by Cas/host proteins to form a mature crRNA. Mature crRNA remains associated with the Cas protein(s) to form an effector ribonucleoprotein complex (RNP); (3) the last stage is the Interference stage, where the RNP complex locates a specific DNA sequence in the invading DNA through complementary crRNA base pairing. Depending on the type of CRISPR-Cas system, the invading DNA is cleaved by either the RNP complex or a separate effector nuclease. Each stage in the CRISPR-Cas system will be discussed in detail in the later sections.

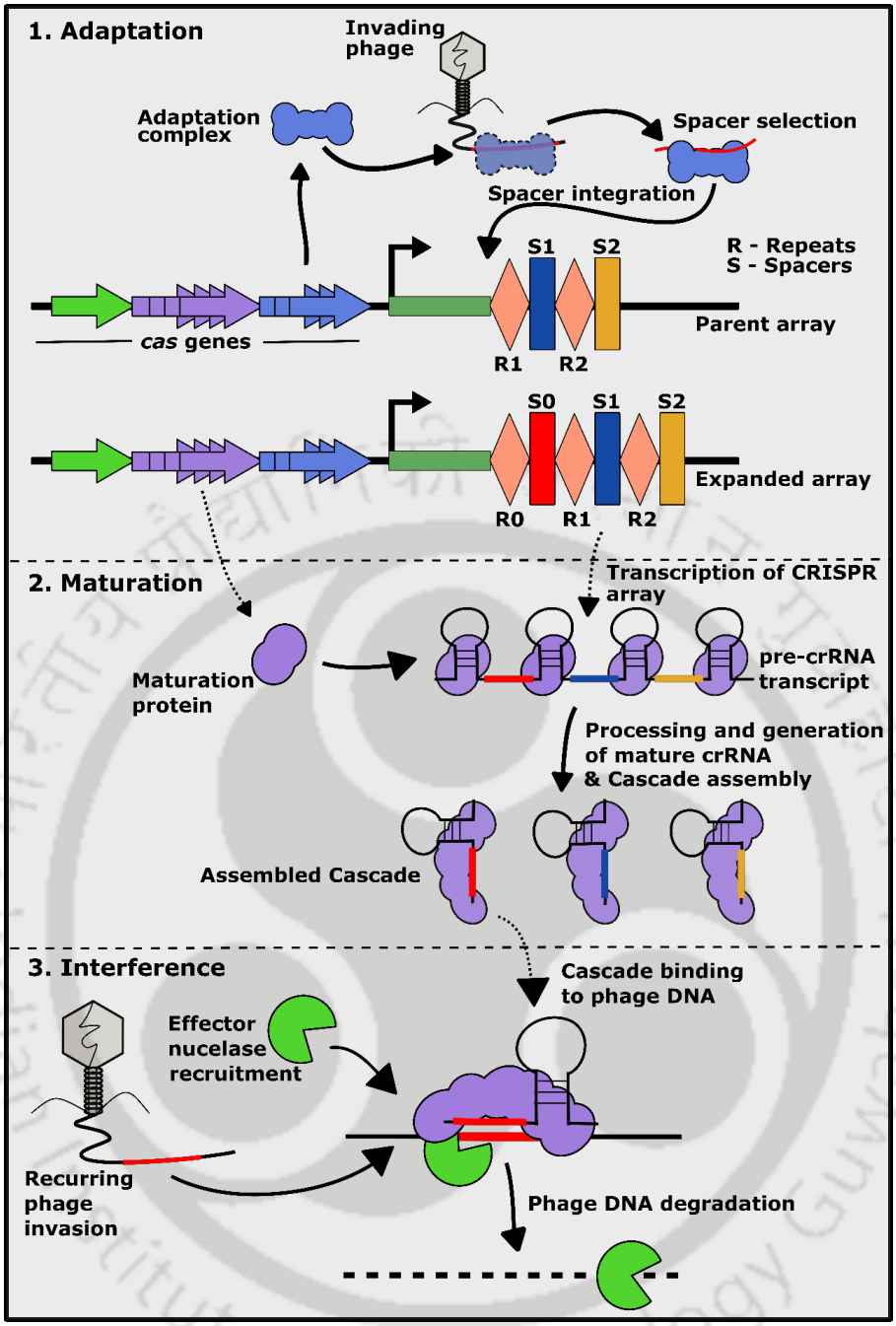

Figure 1. 1: An Overview of the CRISPR-Cas system.

A typical CRISPR-Cas locus consists of an array of conserved repeats which are interspaced by variable spacers. The 'leader' region (dark green bar) is present upstream to the CRISPR array and holds a promotor for CRISPR-RNA expression. CRISPR-associated genes (*cas* genes) are found in the vicinity of the array. **(1) Adaptation:** Adaptation complex (Cas1-Cas2-Cas4) locates PAM in the phage DNA / prespacer. Prespacer is trimmed and integrated into the leader proximal region of the CRISPR array (S0). Depending on CRISPR types, nucleases involved in prespacer trimming may vary. Here, S1 denotes the first spacer, S2 denotes the second, and so on. Similarly, R1 denotes the first repeat and so on. A new spacer (S0) is acquired, followed by repeat duplication (R0) and expansion of the CRISPR array. **(2)**

Maturation: Expression of CRISPR array starts from the 'leader' region forming a pre-crRNA transcript having a stem-loop structure. Cas proteins (Cas6/Cas5) cleave pre-crRNA at a specific location to form a final/mature crRNA, which subsequently assembles into a Cascade complex. **(3) Interference:** In the final stage, invading phage DNA is identified, and an effector nuclease is recruited to cleave the phage genome.

1.1.5 Classification of the CRISPR-Cas system

Despite the functional similarity, the CRISPR-Cas system exhibits remarkable diversity in Cas proteins, gene compositions, the architecture of the genomic loci, and the stoichiometry of the RNP complex (Barrangou and Horvath, 2017; Hille and Charpentier, 2016; Ishino et al., 2018; Klompe and Sternberg, 2018; Mohanraju et al., 2016; Wright et al., 2016). Owing to such vast diversity in the CRISPR-Cas system, the categorisation and classification of the CRISPR-Cas systems is a significant challenge. With the ever increase of genomic and metagenomic data, our knowledge about the diversity in the CRISPR-Cas systems is incessantly expanding. In order to keep pace with such expansion in the CRISPR-Cas, several classification schemes were put forward (Koonin et al., 2017a; Makarova and Koonin, 2015). However, initial classifications became obsolete, and the latest classification was adopted in 2020 (Makarova et al., 2020). Since none of the genes is shared across CRISPR-Cas systems, a computational approach is used. Identification of signature genes, comparison of genomic loci and composition, sequence similarity-based clustering and phylogenetic analysis, neighbourhood analysis and comparison, and experimental evidence were some of the parameters used to categorize the CRISPR-Cas system (Makarova et al., 2020). Based on the roles played by Cas proteins, all the *cas* genes are divided into four distinct functional modules (Figure 1.2). The first module is the Adaptation module, which includes the critical *cas* genes involved in the spacer integration process (Figure 1.2). This module includes a CRISPR array and three genes; *cas1, cas2 and cas4*. In the Expression module, genes for the processing of crRNA are grouped. The Interference module contains genes that encode subunits of the effector complex (Figure 1.2). Finally, in the fourth module, all other ancillary proteins are grouped (Figure 1.2). Based on the architecture of the effector complex, the CRISPR-Cas system is divided into two broad classes; Class 1, where the effector complex consists of multiple Cas proteins such as Cas3, Cas5, Cas6, Cas7, Cas8, Cas10 and Cas11, in different combinations depending on the type and subtypes. Class 2 CRISPR-Cas consists of a single effector protein such as Cas9, Cas12 or Cas13.

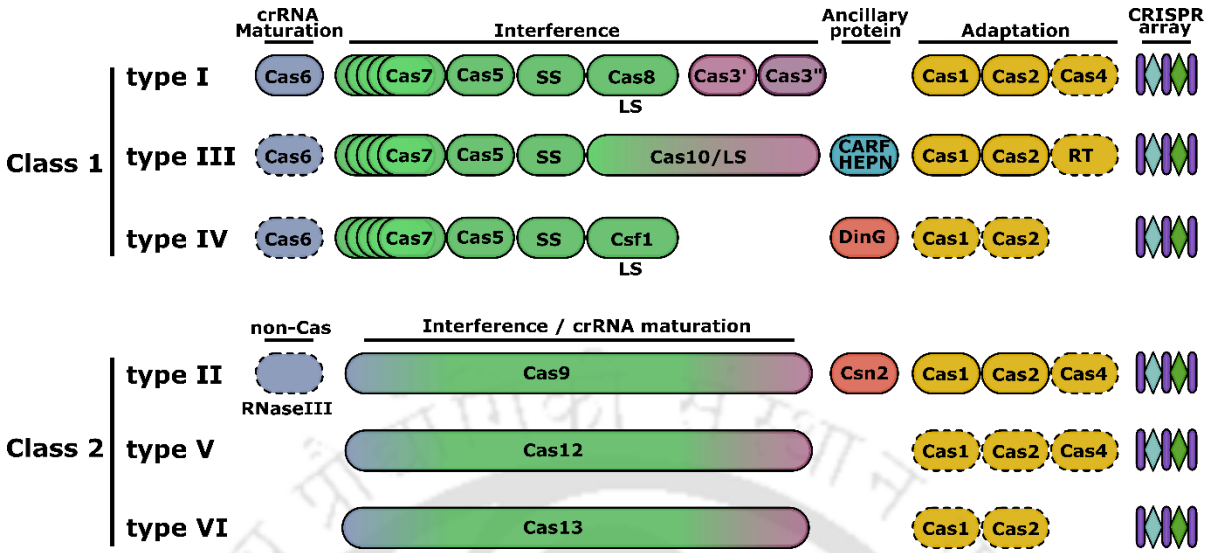

Figure 1. 2: CRISPR-Cas classification:

The figure depicts a generalized representation of the organization of genes in various CRISPR Cas systems. *Cas* genes are grouped according to their functional role, so the actual arrangement of genes in a particular organism can vary. Genes in the dotted line indicate that it is crucial in some subtypes and not present in others. Dual-colour in Cas10 and Class 2 (Cas9, Cas12, and Cas13) means functional convergence compared to gene positions in the type I system. Multiple CRISPR types in a single bacterial species are possible. CRISPR array is shown for all the systems; however, in several bacteria and archaea, the CRISPR array is either absent or shared with other existing CRISPR types. SS – Small subunit, LS – Large subunit. (Makarova et al., 2020)

1.1.5.1 Class 1 CRISPR-Cas system

The Class 1 CRISPR-Cas system is further divided into three types depending on CRISPR loci architecture: type I, type III, and type IV. Type I and type III are the most common CRISPR-Cas types which are present in archaea and bacteria. Type I is marked by Cas3 effector nuclease, whereas the function of Cas3 in type III is carried out by Cas10 protein. The effector protein in type IV is uncharacterized, and Csf2, a subunit of the surveillance complex, is chosen as the signature protein (Figure 1.2) (Koonin et al., 2017a; Makarova and Koonin, 2015; Makarova et al., 2020; Pinilla-Redondo et al., 2020). Type IV system either lack adaptation protein or harbours a rudimentary form of adaptation module (Makarova et al., 2020). The effector complex in type I and III is generally multi-subunit and contains paralogous Cas

proteins of recently discovered Repeat Associated Mysterious Proteins (RAMPs) family, such as Cas5 and Cas7. The RAMP family comprises an RNA Recognition Motif (RRM) containing proteins found in bacteria and archaea, and it is primarily associated with CRISPR loci (Wang and Li, 2012). Along with the RRM fold containing Cas5, Cas6 and Cas7, the effector complex is also comprised of a 'large' and a 'small' subunit (Hochstrasser et al., 2014; Hochstrasser et al., 2016; Jackson et al., 2014a; Jackson and Wiedenheft, 2015; Staals et al., 2013; van der Oost et al., 2014; Zhao et al., 2014). Typically, Cas5 and Cas6 contain duplicated RRM fold in tandem, whereas Cas7 contains only one RRM fold. Effector complex usually contains one unit of Cas5 and several (5-7) units of Cas7 along with crRNA containing a spacer and some portion of a repeat (Hochstrasser et al., 2016; Jackson et al., 2014a; van der Oost et al., 2014; Zhao et al., 2014). The number of Cas7 may change according to the length of the spacer in crRNA (Hatoum-Aslan et al., 2013; Kuznedelov et al., 2016; Luo et al., 2016). The binding of Cas7 to crRNA forms the backbone of the effector complex, which may also contain several copies of the 'small' subunit. Cas5 remains bound to the 5' end of the crRNA, and it directly interacts with the 'large' subunit. Protein that forms the 'large' subunit varies depending on the CRISPR type: Cas8 in type I and Cas10 in type III. Although the RAMP proteins show poor sequence similarity, a comparison of the available structure of the effector complex suggests a common evolutionary origin of these proteins (Jackson and Wiedenheft, 2015; Makarova et al., 2011a; Makarova et al., 2015).

1.1.5.1.1 Type I CRISPR-Cas system

Type I system is characterised by the presence of Cas3 effector nuclease and is further divided into nine subtypes (type I-A to I-G) (Figure 1.3). The adaptation proteins (Cas1, Cas2 and Cas4) are present across all the major subtypes. Cas4, which is required for the processing of prespacers, is absent in type I-E and I-F, and it is fused to Cas1 in type I-G (Almendros et al., 2019; Makarova et al., 2020). Cas6, which helps in the maturation of crRNA, is absent in several subtypes. In type I-C, the function of Cas6 is replaced by Cas5 protein (Nam et al., 2012; Punetha et al., 2014). Interestingly, Cas3 shows variability in the domain architecture across subtypes. A typical Cas3 protein contains an N-terminal HD-type nuclease domain fused to DExD/H box superfamily 2 helicase and a C-terminal domain (CTD). In type I-A, the two functional domains of Cas3 (nuclease-helicase) are split and co-operate during interference (Majumdar and Terns, 2019; Makarova and Koonin, 2015). The domains in type I-G Cas3 are interchanged, and it contains N-terminal helicase followed by C-terminal nuclease (Makarova

et al., 2020). Cas3 is fused to Cas2 in type I-F, and its nuclease domain is fused to Cas10 in type I-D (Figure 1.3) (Makarova et al., 2020).

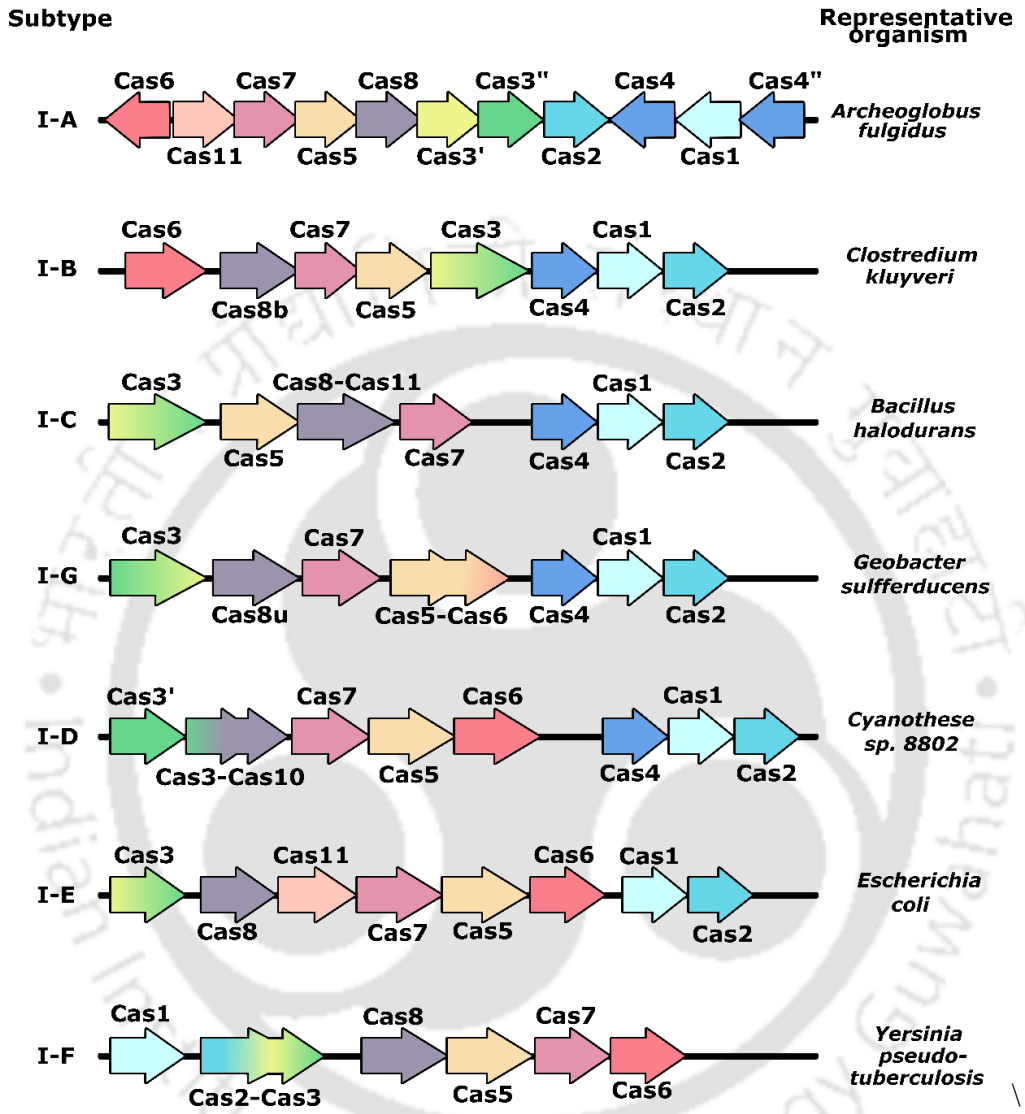

Figure 1. 3: Type I CRISPR-Cas classification.

The figure shows the schematic representation of the genetic composition and architecture of type I systems. A list of representative organisms is mentioned on the right, and on the left, CRISPR subtypes are mentioned. Homologous genes are depicted in the same colours. In type I-A, Cas3' (yellow) and Cas3" (green) represent DExD/H box helicase and HD-nuclease, respectively, whereas, in another subtype, the fusion of helicase-nuclease is represented using gradient colouring (yellow-green). As an exception in type I-G, Cas3 domains are reversed and depicted accordingly using gradient colour.

1.1.5.1.2 Type III CRISPR-Cas system

The presence of an HD-type nuclease Cas10 protein characterises the type-III CRISPR-Cas system. Cas10 is the largest subunit in the type-III system and contains a palm domain (cyclase and polymerase). (Osawa et al., 2013; Osawa et al., 2015; Staals et al., 2014; Staals et al., 2013). All type-III systems (except type III-A and III-E) lack a CRISPR array and possibly share the CRISPR array from other subtypes present in the same organism. Like type-I subtypes, type-III also exhibits several variations in CRISPR loci architecture, and therefore it is further divided into six subtypes (type III-A to III-F) (Makarova et al., 2020). The adaptation module is not present in most of the type-III system, and it is believed that it borrows adaptation proteins from other CRISPR-Cas subtypes present at a different genomic location. This system is also marked by genes encoding the small subunits, Csm (in type III-A, III-E and III-D) and Cmr (in type III-B and III-C) (Tamulaitis et al., 2017). Apart from the *cas* genes, the type-III system is associated with accessory proteins that are shown to be involved in membrane transport or signal transduction. These proteins contain a CRISPR Associated Rossmann Fold (CARF) domain and a Higher Eukaryotes Prokaryote Nucleotide-binding (HEPN) domain. This system has been shown to synthesise a signalling molecule, cyclic oligoA, which binds to the CARF domain, which then activates the HEPN domain, cleaving RNA indiscriminately (Jia et al., 2019; Molina et al., 2019).

1.1.5.1.3 Type IV CRISPR-Cas system

Several bacterial and archaeal species contain uncharacterised type-IV CRISPR-Cas system. This system is further divided into four subtypes; type-IV-A to type IV-D (Pinilla-Redondo et al., 2020). This subtype is considered highly mobile because it is usually found in plasmids. Like type-III, this system lacks the adaptation proteins, Cas1, Cas2 or Cas4 and an effector nuclease (Makarova et al., 2015; Makarova et al., 2020). Csf2 is the signature protein for this system (Makarova et al., 2015; Makarova et al., 2020; Pinilla-Redondo et al., 2020). This system includes a minimal effector complex (Csf1, Csf3 (Cas5), Csf2 (Cas7) and crRNA) which consists of a reduced large subunit. Type IV-A has a DinG helicase - Csf4 and a Cas6 like protein - Csf5 (Makarova et al., 2020; Pinilla-Redondo et al., 2020). A putative signal transduction protein CysH is present in type IV-C (Makarova et al., 2020; Pinilla-Redondo et al., 2020). Type IV system is found mostly associated with type I system. Due to this reason and the absence of adaptation proteins, type IV is believed to share Cas proteins with type I subtypes (Pinilla-Redondo et al., 2020).

1.1.5.2 Class 2 CRISPR-Cas system

Class 2 system shows a minimal architecture with a single multidomain protein (Cas9 or Cpf1), which performs all the interference related functions and crRNA processing in some variants. The presence of a single effector protein in Class 2 makes it suitable for several applications such as genetic manipulation and gene silencing. Since its discovery, our understanding of the Class 2 CRISPR-Cas system has been rapidly expanding. Like Class 1, Class 2 is also sub-divided into three types (Type II, V and VI), and these types are further divided into >25 subtypes and still expanding (Makarova et al., 2020).

1.1.5.2.1 Type II CRISPR-Cas system

The gene encoding a multidomain protein Cas9 is the signature gene for the type II CRISPR-Cas system. Type II is further divided into three subtypes: Type II-A, II-B and II-C (Makarova et al., 2020). Type II contains the usual adaptation gene *cas1* and *cas2*. Gene *cas4* endonuclease is present in some variants, and it may be required for the adaptation process (Makarova et al., 2020). In type II-A, Cas4 is absent and replaced by Csn2 protein, which, along with Cas9, is found to be essential for prespacer acquisition (Heler et al., 2015; Nussenzweig et al., 2019; Wei et al., 2015). Most type II systems also harbour a tracrRNA with a complementary sequence to the repeats in the CRISPR array. The tracrRNA binds to the repeat region on the pre-crRNA transcript, followed by an RNaseIII mediated cleavage leading to the maturation of crRNA (Deltcheva et al., 2011). Both Cas9 and tracrRNA are required for crRNA maturation. Cas9 shows a bilobed structure having a Recognition lobe (REC) and a Nuclease lobe (NUC). NUC lobe contains juxtaposed RuvC endonuclease and HNH nuclease domain (Jinek et al., 2014a). In the interference stage, the effector complex (Cas9, tracrRNA and crRNA) locates the target DNA, and the RuvC and HNH domains of Cas9 makes a double-stranded cut on both the strands of DNA (Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012a).

1.1.5.2.2 Type V CRISPR-Cas system

Type V is the most diverse of all the CRISPR-Cas types and is further divided into ten subtypes (Makarova et al., 2020; Yan et al., 2019). The information about several subtypes in type V systems are derived from metagenomic studies, and hence only computational evidence is available for some subtypes (Yan et al., 2019). However, many subtypes are highly studied,

which includes Cas12a (Cpf1) from type V-A (Moreno-Mateos et al., 2017; Safari et al., 2019; Zetsche et al., 2015; Zetsche et al., 2017). Cas12 is the signature protein in this type, and similar to Cas9, Cas12 is also bilobed (both REC and NUC present), and it also contains two RuvC endonuclease domains which cleave both the strands on the target DNA. (Zetsche et al., 2015). Unlike Cas9, Cas12 lacks the HNH nuclease domain. All the adaptation genes, *cas1*, *cas2* and *cas4,* are present in type V-A, V-B (*cas4* fused to *cas1*), V-F and V-E (Makarova et al., 2020; Yan et al., 2019). Type V-C and V-D contain only the *cas1* gene, whereas the adaptation genes are entirely absent in type V-G, V-I, V-H, and V-U, suggesting a shared adaptation module with other existing CRISPR loci (Makarova et al., 2020; Yan et al., 2019). Apart from this, certain types (V-F, V-G and V-U) contains a comparatively smaller version of Cas12 protein. Cas12-crRNA effector complex typically targets double-stranded DNA (dsDNA). Exceptionally, Cas12 from type V-G (Cas12g) has been shown to possess RNA-dependent RNA targeting (Karvelis et al., 2020). Upon binding to target RNA, Cas12g also shows a non-specific collateral cleavage of RNA and single-stranded DNA (ssDNA) (Chen et al., 2018; Harrington et al., 2018; Yan et al., 2019).

1.1.5.2.3 Type VI CRISPR-Cas system

Type VI is the minimalistic CRISPR-Cas type and contains Cas13 as a signature protein (Makarova et al., 2020). Cas13 shows bilobed organisation with REC and NUC lobes (Zhang et al., 2018). It is further divided into four subtypes (type VI-A, VI-B, VI-C and VI-D) (Makarova et al., 2020). The presence of two HEPN domain is the peculiar feature of the type VI system (O'Connell, 2019; Yan et al., 2018; Zhang et al., 2019; Zhang et al., 2018). HEPN domain is also a common feature of proteins in several toxin-antitoxin systems. Thus it is believed that Cas13 has originated by a duplication of the HEPN domain from a common ancestor possibly belonging to the toxin-antitoxin system (Makarova et al., 2020). The adaptation module is almost absent or not well distinguished, suggesting shared adaptation modules with other existing CRISPR loci (Makarova et al., 2020). Like type V systems, type VI also exhibits non-specific collateral cleavage of RNA (Abudayyeh et al., 2016; Knott et al., 2017).

1.1.6 CRISPR adaptation

The variable DNA sequence "spacer" within the CRISPR array is the fundamental part of immunological memory. Spacer sequences are acquired from mobile genetic elements (MGEs, like plasmids and viruses) to provide immunity against them. Studies have shown that some spacers are also acquired from the host genome (Levy et al., 2015; Yosef et al., 2012); however, this phenomenon is poorly understood. Research based on data mining suggests that around 90% of the acquired spacer sequences are derived from MGE (Shmakov et al., 2017). The number of spacers in a CRISPR array can vary from 2 to more than 200, and they usually do not belong to the same MGE. Thus, the presence of variable spacer sequences can provide immunity against a wide range of MGEs during the interference stage (Shmakov et al., 2017). The acquisition of new spacers leads to the expansion of the CRISPR array, which is also responsible for acquiring immunological memory through Lamarckian type evolution (Barrangou et al., 2007; Garneau et al., 2010; Yosef et al., 2012).

The adaptation process is a complex and multistep process where a portion of MGE is chosen and subsequently stored in the CRISPR array. In the first step, Cas proteins identify a short sequence motif (2-7 nt) called protospacer-adjacent motif (PAM) on the invading MGE. In the next step, this portion of DNA is processed into 20-40 bp sequence with the help of Cas proteins and in some prokaryotes by host exonucleases (Barrangou et al., 2007; Garneau et al., 2010; Marraffini and Sontheimer, 2010; Yoganand et al., 2019; Yosef et al., 2012). This small portion of MGE is termed as "prespacers" (spacer precursors). In the last step, the processed portion of DNA is integrated into the CRISPR array by Cas integrases (Barrangou et al., 2007; Garneau et al., 2010). The integrated portion of DNA is termed spacers. Usually, the adaptation process is polarised where new spacers are integrated towards the 'leader' proximal region of CRISPR loci (Barrangou et al., 2007; Díez-Villaseñor et al., 2013; Erdmann and Garrett, 2012; Yosef et al., 2012). In general, the leader region harbours the promoter for the transcription of the CRISPR array during the later stages. Targeting the phages/plasmids, which undergo rapid mutations (escape mutants), can be challenging for the CRISPR-Cas system (Amitai and Sorek, 2016; Sternberg et al., 2016). Occasionally, multiple spacers are also acquired from a single phage genome, thereby increasing the chances of interference when some portion of phage DNA is mutated (Paez-Espino et al., 2013; Staals et al., 2016b; van Houte et al., 2016). The frequent and multiple acquisitions of spacers from MGEs keep the CRISPR system in pace with the rapidly evolving phages/plasmids.

Acquisition of spacers from entirely unfamiliar MGEs is called naïve adaptation. The process of naïve adaptation is slow, and the spacers are acquired randomly. MGEs undergo rapid mutation, and the slow process of naïve adaptation is not sufficient to act against mutated MGEs. In order to overcome this problem, spacers are rapidly acquired from mutated MGE through a process of 'priming' called primed adaptation or priming aquisition. Naïve and primed adaption processes are discussed in details in later sections.

1.1.6.1 Self and non-self-discrimination during adaptation

Self-targeting is a potential threat because of two reasons. First, the adaptation proteins can acquire the host genome sequence that leads to the targeting of the 'self' genome (Levy et al., 2015; Wei et al., 2015; Yosef et al., 2012). Second, the crRNA, which is transcribed from the CRISPR array, can pair with the corresponding spacer in the host genome due to sequence complementarity leading to cell death (Bikard et al., 2012; Edgar and Qimron, 2010; Jiang et al., 2013). Thus, the CRISPR system must discriminate between self and non-self-DNA during the interference and adaptation stages. To avoid such autoimmune response CRISPR system uses several mechanisms to bias specificity towards foreign MGEs.

1.1.6.1.1 RecBCD complex in adaptation

For the generation of spacers during naïve adaptation, the CRISPR system uses the RecBCD pathway in gram-negative bacteria (Ivancic-Bace et al., 2015; Levy et al., 2015) and AddAB in gram-positive bacteria (Modell et al., 2017). The primary function of RecBCD is to assist the repair of double-stranded breaks in DNA via homologous recombination (Figure 1.4) (Dillingham and Kowalczykowski, 2008). It also degrades single-stranded DNA protecting bacteria from foreign MGEs (Modell et al., 2017). Activated RecBCD complex unwinds dsDNA through ATP dependent helicase subunit, and the nuclease subunit degrades DNA until it encounters a crossover hotspot instigator (Chi) site (Dillingham and Kowalczykowski, 2008). Chi, which is an eight-nucleotide sequence motif, decelerates the RecBCD mediated cleavage. Since Chi sites are present abundantly in the host genome (every 4.6 kb), it limits the prespacer generation by RecBCD complex from 'self' DNA. On the contrary, MGEs, which generally lack Chi-site, becomes an easy target for the RecBCD complex (Figure 1.4).

Furthermore, the linear phage DNA injected during infection is preferentially utilised by CRISPR machinery to generate spacers, as the host genomic DNA is circular and lacks double-stranded breaks. These properties of MGEs allows immediate recognition and acquisition of spacers, making the CRISPR system more effective against invading phages (Modell et al., 2017). Alternatively, it is also possible that Cas proteins may physically associate with RecBCD (Babu et al., 2011). Moreover, in several CRISPR types, spacer acquisition happens in the absence of RecBCD, which suggest alternate self/non-self-recognition mechanisms.

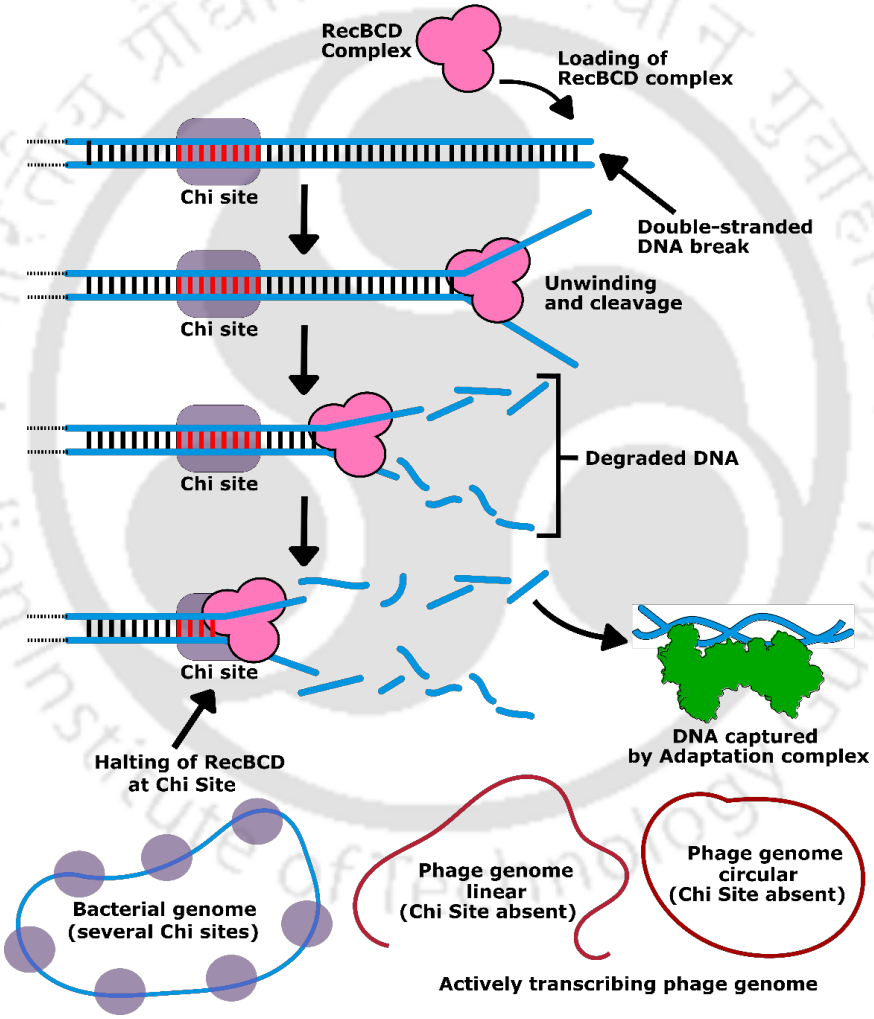

Figure 1. 4: Self vs non-self-recognition.

Schematic diagram representing naïve adaptation through the RecBCD pathway. The DNA degraded by the RecBCD DNA repair complex with the appropriate size and containing the PAM region is identified and captured by the Adaptation complex. Since the phage genome lack Chi sites, the RecBCD complex can actively degrade it. Additionally, the frequency of spacer uptake is higher for actively transcribing phage genome.

1.1.6.1.2 Regulation of spacer acquisition

The CRISPR-Cas autoimmunity is also controlled by limiting the rate of spacer acquisition. In the laboratory condition, the rate of DNA uptake from MGEs has been reported to be very slow (occurring in 1 in 10^7 cells) (Heler et al., 2015; Hynes et al., 2014; McGinn and Marraffini, 2016). This rate is slower in natural conditions, and the rate of uptake of 'self' DNA is equally rare. On the other hand, overexpression of Cas integrases to increase spacer acquisition rate showed a higher level of toxicity (Heler et al., 2017). The slow rate of spacer acquisition is evolutionarily significant that nullifies the deleterious effect of autoimmunity. In order to save the bacterial population from the effects of autoimmunity, the rate of spacer acquisition may be regulated by an unknown mechanism. Indeed, the recent discovery of quorum sensing in the CRISPR-Cas system suggests molecular regulation in the fight against phages (Hoyland-Kroghsbo et al., 2017; Patterson et al., 2016).

1.1.6.1.3 Selection of prespacer

CRISPR adaptation cannot acquire a spacer from a random location on MGEs. Only a subsection of DNA bordering the PAM sequence on MGEs is qualified to become a spacer in type I and II CRISPR systems (Deveau et al., 2008; Mojica et al., 2009; Wang et al., 2015). However, during adaptation, the PAM sequence is omitted, and it is not integrated into the CRISPR array. Since the CRISPR array lacks PAM, it cannot become a target for the CRISPR interference machinery protecting the 'self' genome. However, the MGE with the cognate PAM sequence can be targeted by CRISPR interference. Additionally, it has been shown that the frequency of spacer uptake is higher from actively transcribing and replicating part of the DNA (Deng et al., 2013; Goldberg et al., 2014; Peng et al., 2015; Samai et al., 2015). The high transcription and replication rate are the hallmark of phage and plasmid DNA, whereas host genomic DNA is comparatively less active (Figure 1.4). Therefore, spacers are preferentially acquired from MGEs sparing the host genome.

1.1.6.2 Selection and processing of prespacers

In type I, II and V CRISPR-Cas system, the selection of prespacers is guided by a DNA sequence in the PAM region on the MGEs. The PAM sequence is generally compatible with the interference machinery (Datsenko et al., 2012; Deveau et al., 2008; Mojica et al., 2009; Swarts et al., 2012). Even though CRISPR-Cas systems are highly diverse, Cas1 and Cas2

proteins, crucial for adaptation machinery, are universally conserved across most CRISPR-Cas types. Most of our understanding of CRISPR adaptation in type I system has been gained from *E. coli* which belongs to the type I-E system. In *E. coli,* Cas1-Cas2 assemble to form a heterohexameric integrase complex – $(Cas1)_4(Cas2)_2$ – that aids in the integration of new spacers into the CRISPR array (Figure 1.5A) (Nunez et al., 2015a; Nunez et al., 2014; Wang et al., 2015; Wright et al., 2017). There are four Cas1 active sites on the integration complex, out of which only two Cas1 sites are utilised during spacer processing and integration (Wang et al., 2015). Cas1-Cas2 complex identifies PAM region and Cas1 subunit preferably bind the PAM complementary sequence (Figure 1.5B and 5C) (Datsenko et al., 2012; Swarts et al., 2012; Wang et al., 2015; Yoganand et al., 2019). The dual-forked substrates are generated as a result of helicase-nuclease activity of the RecBCD complex (Levy et al., 2015) or by interference driven primed adaptation (Kunne et al., 2016; Staals et al., 2016a). Cas1-Cas2 complex identifies such dual-forked DNA substrates, which contain a 23 bp duplex region and 3' extended overhangs. In E. *coli,* the architecture and dimension of the Cas1-Cas2 complex are tuned to accommodate approximately 33 nt long DNA substrates (Wang et al., 2015). The processing of DNA substrates usually begins with the docking of the Cas1-Cas2 complex onto forked DNA substrate, which is subsequently trimmed by host exonucleases such as ExoIII, ExoT and DnaQ (Figure 1.5B) (Ramachandran et al., 2020; Yoganand et al., 2019). As a minor variation in type I-E, the CRISPR system in *Streptococcus thermophilus* encodes Cas2 intrinsically fused to DnaQ nuclease, which engenders prespacers compatible for integration (Drabavicius et al., 2018). Apart from Cas1-Cas2, the several CRISPR types encode an accessory nuclease Cas4 as a part of the integration complex for prespacer processing and integration. The role of Cas4 is discussed in detail in the later section.

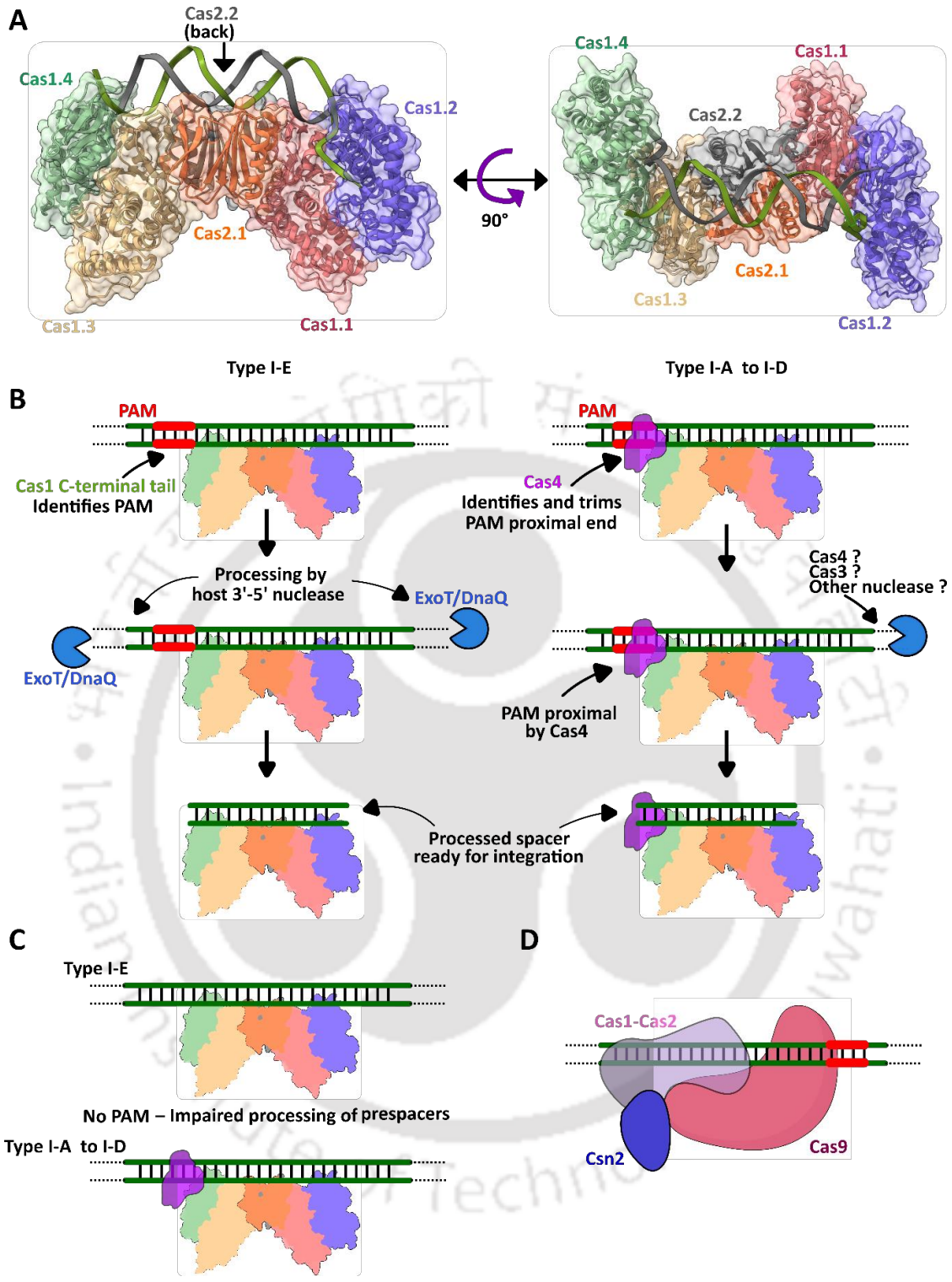

Figure 1. 5: Selection and processing of prespacers.

A. Molecular structure of type I-E adaptation complex containing four subunits of Cas1 and two subunits of Cas2 (PDB ID: 5DS6) **B**. In CRISPR types I-A to I-D, Cas4 endonuclease recognises and trims PAM proximal end of the prespacer, and an unknown nuclease trims PAM distal end. In type I-E, Cas4 is absent, and PAM is recognised by the C-terminal tail of the Cas1 subunit. In the subsequent step, host 3'-5' nucleases (ExoIII, ExoT or DnaQ) trims unprocessed prespacers. The processed prespacer are asymmetric and contains a 3'-OH group.

The binding of the adaptation complex to the *bonafide* prespacer acts as a molecular ruler conserving the length of spacers.

C. Absence of the cognate PAM sequence does not support prespacer processing and integration.

D. In type III-A, an auxiliary protein Csn2 is required along with Cas1-Cas2 and Cas9. Cas9 is only required for PAM identification, whereas Csn2 nuclease trims prespacer.

In type II-A CRISPR-Cas system, all the Cas proteins *viz.*, Cas9 (effector nuclease), Cas1, Cas2 (core integrase complex), Csn2 (an auxiliary protein) and tracrRNA are required for spacer selection and integration (Heler et al., 2015; Wei et al., 2015). Here, Cas9 is responsible for PAM recognition and mutations in the PAM recognition domain leads to random spacer acquisition (Figure 1.5D). Nuclease null mutants of Cas9 support spacer integration, suggesting Cas9 nuclease is not involved in prespacer processing but required only for PAM identification (Heler et al., 2015; Wei et al., 2015). However, it is yet to be established how exactly Cas9 assembles and interacts with the Cas1-Cas2-Csn2 complex during spacer acquisition. In a proposed model, Cas9 may bind to a degraded PAM-containing portion of MGEs. Csn2 (ring-shaped tetramer) may slide on DNA till it encounters Cas9 and subsequently trim and load this processed DNA onto the adaptation complex (Figure 1.5D) (Wilkinson et al., 2019). Since Csn2 is absent in type II-B and II-C, the exact mechanism of spacer acquisition is poorly understood in these systems (Makarova et al., 2020).

1.1.6.3 Role of Cas4 in CRISPR adaptation

Type I (I-A, I-B, I-C and I-G), type II-B and type V (V-A, V-B, V-E, V-F) encode for an accessory protein Cas4, which is involved in the adaptation process (Almendros et al., 2019; Kieper et al., 2018; Lee et al., 2019; Lee et al., 2018; Makarova et al., 2020; Rollie et al., 2018; Swarts et al., 2012). Cas4 is homologous to RecB and additionally contains four conserved cysteine residues that are involved in the coordination of an iron-sulphur cluster (Zhang et al., 2012). Its nuclease activity ensures the selection of prespacers with correct PAM and trimming of prespacers at PAM proximal end before integration (Figure 1.5B). A recent study in *Bacillus halodurans* (type I-C) revealed that Cas4 directly interacts with Cas1-Cas2 (Lee et al., 2019). Recent Cryo-EM structure of Cas4/I-C showed Cas4-Cas1-Cas2 are present as a complex in 2:4:2 stoichiometry ($(Cas4)_2$-$(Cas1)_4$-$(Cas2)_2$) (Lee et al., 2019). However, a majority of the

particles observed under Cryo-EM contained a single copy of Cas4 in the complex, i.e. $(Cas4)_1$-$(Cas1)_4$-$(Cas2)_2$. In this complex, Cas4 is present near the PAM region in unprocessed prespacers, which suggest the role of Cas4 in the orientation of prespacers at the leader-repeat-spacer junction in the CRISPR array (Lee et al., 2019).

1.1.6.4 Integration of prespacers into CRISPR array

Each of the four Cas1 subunits in the adaptation complex contains a PAM sensing region; however, only one of the Cas1 subunit holds PAM in its active site (Nunez et al., 2015a; Wang et al., 2015). This safeguards the correct orientation of spacers during the integration into the CRISPR array (Shipman et al., 2016; Shmakov et al., 2014; Swarts et al., 2012). The unidirectional integration is essential to ensure the alignment of the PAM region with the crRNA sequence during CRISPR interference (Figure 1.6). Another crucial requirement during integration is the presence of the leader and repeats (Figure 1.6). Prespacer integration is polarised, and newly acquired spacers are incorporated at the leader proximal end of the CRISPR array. This polarised integration maintains the chronology of MGE encounters and thus favours efficient interference against the most recent infections (Barrangou et al., 2007; McGinn and Marraffini, 2016; Pourcel et al., 2005). The mechanism of prespacer integration has been mainly studied in type I-E system, and it has been shown that the integration mechanism is analogous to retroviral integrases and transposases (Nunez et al., 2015b). The integration process initiates with a nucleophilic attack of 3' OH of the prespacer followed by ligation to 5' phosphate of the leader-repeat junction and forms a half-site intermediate (Figure 1.6). The recognition of leader-repeat boundary is guided by binding of leader sequence with Integration Host Factor (IHF) (Figure 1.6) (Nunez et al., 2016; Nunez et al., 2015b; Yoganand et al., 2017). After binding to IHF, the leader region sharply bends, guiding the Cas1-Cas2 complex towards the leader-repeat boundary (Figure 1.6) (Yoganand et al., 2017). Subsequently, a second nucleophilic attack at the existing first repeat-spacer junction leads to full-site integration of prespacers into the CRISPR array (Figure 1.6) (Arslan et al., 2014; Nunez et al., 2015b; Rollie et al., 2015). The correct orientation of integration is guided by the presence of partial/full PAM in the prespacer. However, after integration, only one nucleotide from PAM is retained (Figure 1.6) (Shipman et al., 2016; Shmakov et al., 2014; Wang et al., 2015).

On the contrary, the adaptation process in type II systems do not require IHF. Instead, adaptation is dependent on a short (~5 nt) motif on the leader region termed as Leader Anchoring Site (LAS) (McGinn and Marraffini, 2016; Wright and Doudna, 2016; Xiao et al., 2017b). LAS is directly recognised by the Cas1-Cas2 complex and is sufficient to produce correctly oriented spacers. Mutations in LAS lead to abnormal integration of prespacers (McGinn and Marraffini, 2016). Like type I-E system, spacer integration starts with a 3' OH nucleophilic attack at the first LAS-repeat junction leading to half-site integration products (Xiao et al., 2017b). Subsequently, depending on the spacer size and binding of the Cas1-Cas2 complex, a second nucleophilic attack on the first repeat-spacer junction leads to full-site integration (Wright and Doudna, 2016; Xiao et al., 2017b). Failure to produce full-site integration leads to improper integration, which is likely to be disintegrated by host DNA repair enzymes (Wright and Doudna, 2016).

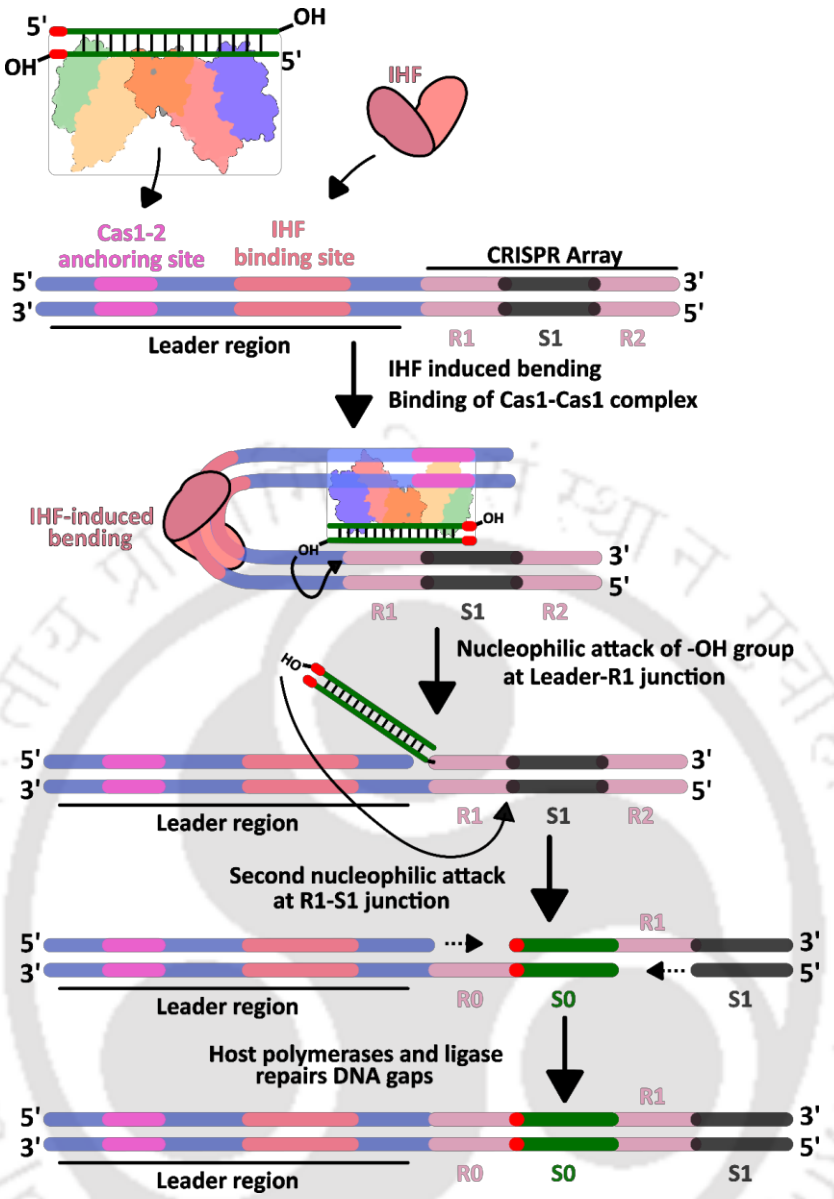

Figure 1. 6: Integration of prespacer into the CRISPR array.

The above figure depicts the process of spacer integration that occurs in the type I-E CRISPR-Cas system. Cas1-Cas2 complex (Cas1-Cas2-Cas4 complex in type I-A to I-D and I-G) captures PAM containing DNA fragment generated through RecBCD mediated degradation (Naïve adaptation) or generated as a result of CRISPR interference (Primed adaptation). Only a single base from PAM is retained during integration (The base 'G' from 'AAG' PAM is retained in type I-E system). Cas1-Cas2 complex harbouring PAM containing prespacer binds to Cas1-Cas2 anchoring site (coloured magenta). Subsequently, IHF bends the leader region by 120° and localizes the Cas1-Cas2 complex at the leader-R1 junction. Nucleophilic attack by 3' -OH group at leader-R1 junction leads to half-site integration intermediate product. Subsequently, a second nucleophilic attack by free 3'-OH at the R1-S1 junction produces a full-site integration product. Finally, the breaks and gaps in the array are repaired by host polymerases and ligases. R1 – first repeat in parent array, S1 – First spacer in parent array, R0

– newly generated repeat after integration, S0 – newly acquired spacer after integration, IHF – Integration Host Factor.

1.1.6.5 Primed Adaptation

The mutation rate in MGEs is very high, and often, mutations in the PAM and 'seed' sequence (7-13 nt long PAM proximal region on MGE is crucial for interference) of the prespacer can lead to impaired CRISPR interference (Deveau et al., 2008; Fineran et al., 2014; Semenova et al., 2011). Thus, to avoid such 'escape' MGEs, the CRISPR system has evolved with a mechanism to increase spacer acquisition frequency using the existing spacers as bait (Andersson and Banfield, 2008; Swarts et al., 2012). During this process, multiple spacers are acquired from the priming region of the interference complex during the interference stage leading to a strong immune response from newly acquired spacers (Figure 1.7) (Fineran et al., 2014; Savitskaya et al., 2013; Swarts et al., 2012). Additionally, a naturally occurring fusion of Cas2 and Cas3 (Cas2-3) in type I-F explains the evolutionary advantage of having an overlap between adaptation and interference stages (Fagerlund et al., 2017; Makarova et al., 2020; Staals et al., 2016a).

Primed adaptation begins with the identification of target DNA/MGEs by interference complex. After binding to target DNA, the interference complex recruits Cas3 – an active nuclease-helicase – that unwinds and degrades target DNA into smaller fragments (Figure 1.7B). In *E. coli,* the *in vitro* activity of Cas3 produces ~30 to 100 nt fragments which may anneal to form suitable prespacers for the Cas1-Cas2 complex (Kunne et al., 2016). In a separate study, the conformation of the Cascade complex after binding to cognate or mutated target DNA decides the prevalence of interference or adaptation, respectively. With mutation in PAM or 'seed' region, Cascade complex attains a confirmation that is not favourable for interference; instead, Cas1-Cas2 and Cas3 are recruited and forms a Primed Adaptation Complex (PAC), which leads to the acquisition of new spacers (Figure 1.7A) (Blosser et al., 2015; Hayes et al., 2016; Redding et al., 2015; Xue et al., 2016).

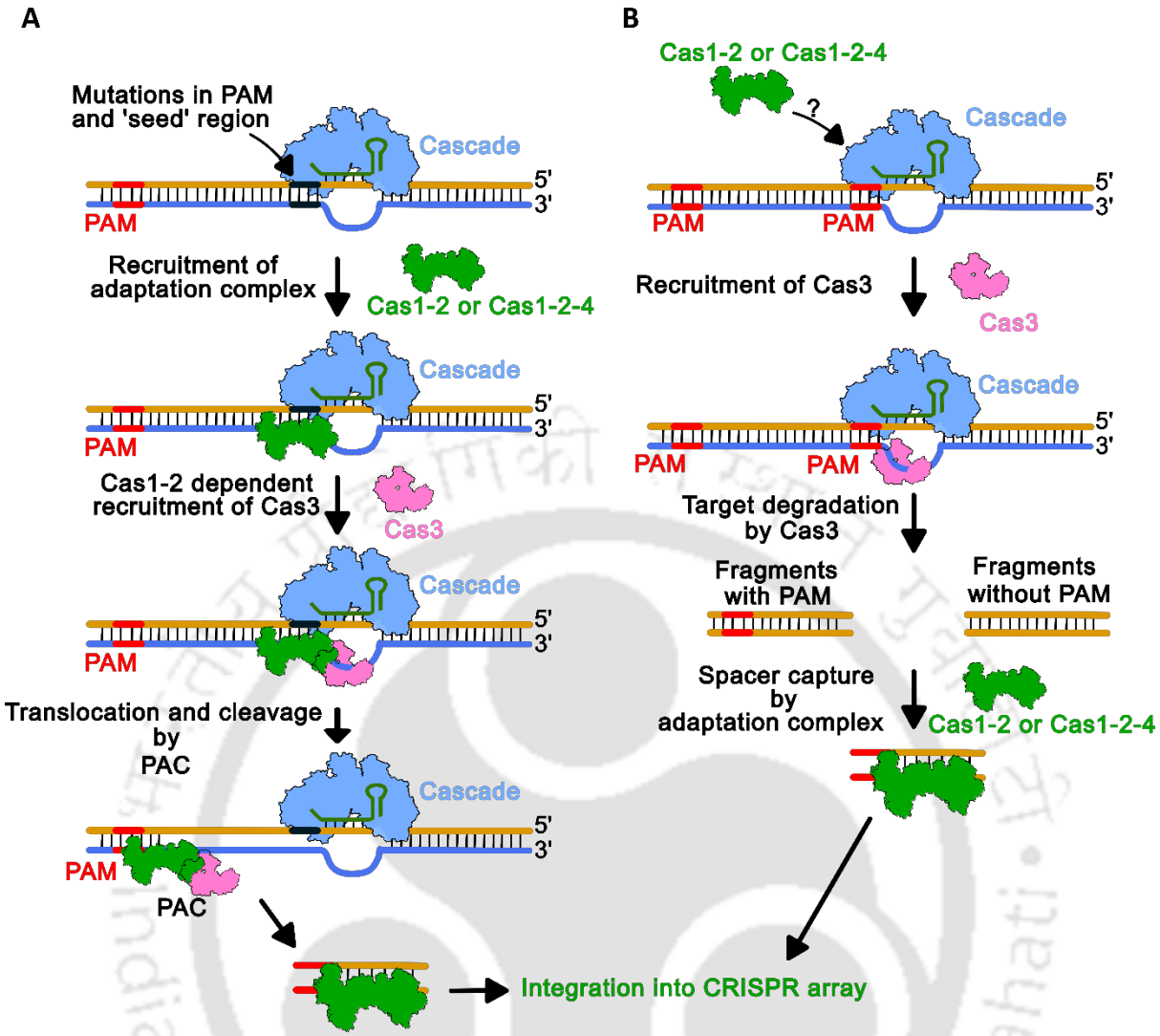

Figure 1. 7: Interference dependent spacer acquisition.

A. Mutations (coloured black in DNA strands) in the PAM and 'seed' region can lead to impaired interference. In some cases, the Cascade guided by crRNA can bind to the imperfect target sequence; however, the recruitment of Cas3 is conformationally blocked. In such a scenario, the adaptation complex is recruited to the Cascade priming region, which subsequently recruits Cas3 nuclease. The adaptation complex and Cas3 associate to form a Primed Adaptation Complex (PAC) in type I-E. The PAC utilizing helicase activity of Cas3 translocates unidirectionally (bidirectionally in some subtypes) on the target DNA, producing DNA fragments. On encountering PAM (coloured red), the adaptation complex halts and integrates the corresponding portion of DNA into the CRISPR array.

B. According to an alternate pathway in type I-E, in the presence of canonical PAM, the Cascade complex recruits Cas3 nuclease, which subsequently degrades target DNA. The degraded fragment of target DNA containing cognate PAM is selected and captured by adaptation machinery and later integrated into the CRISPR array. The association of adaptation complex with interference complex, in this case, is undetermined. The rate of target

degradation is higher when crRNA and target are perfectly matched, and thus the Adaptation-Interference association, in this case, is poorly understood.

Primed adaptation is at least 50 times more frequent than naïve acquisition, and it can also tolerate mutations up to 13 nucleotides (including PAM and 'seed' region) in the MGE (Datsenko et al., 2012; Fineran et al., 2014; Xue et al., 2015). Primed adaptation is strand biased in type I-E, and most of the spacers are acquired from the strand that triggered priming (Datsenko et al., 2012; Shmakov et al., 2014; Swarts et al., 2012). However, during naïve adaptation, spacers are acquired equally from both strands. Whereas, in type I-B and I-F, such strand bias is not observed (Li et al., 2014; Richter et al., 2014). Overall, the present understanding of primed adaptation is minimal, and several questions from strand selection to molecular events that occur during priming is yet to be answered. Since primed adaptation is not studied in most CRISPR-Cas subtypes, understanding its mechanism in different systems will significantly elevate our knowledge.

1.1.7 Biogenesis of crRNA and effector complex

The characteristic of the CRISPR-Cas system is the degradation of MGEs by forming a ribonucleoprotein complex that includes an invader derived small RNA and Cas proteins. Herein the small RNA acts as a guide to recognise the MGE based on sequence complementarity. The process starts with the transcription of the CRISPR locus into a long precursor CRISPR RNA (pre-crRNA), which is later processed to form a mature CRISPR RNA (crRNA) (Carte et al., 2008). Each crRNA contains a partial repeat sequence and a complete spacer sequence which can target a unique MGE. The transcription is facilitated by the AT-rich leader region upstream to the CRISPR array (Brouns et al., 2008; Pougach et al., 2010; Pul et al., 2010). In *E. coli* and *Salmonella enterica*, the transcription of the CRISPR array is regulated by transcription repressor H-NS and activator LeuO (Medina-Aparicio et al., 2011; Pougach et al., 2010; Pul et al., 2010; Westra et al., 2010). As an exception, each repeat in the CRISPR array of *Neisseria meningitides* contains a 9 nt promoter sequence resulting in the individual transcription of the downstream spacer (Zhang et al., 2013).

1.1.7.1 Maturation of crRNA in Class 1 system

CRISPR-Cas systems are diverse, and therefore the process of crRNA maturation is not similar between Class 1 and Class 2 (Punetha et al., 2018). In Class 1, the process of crRNA maturation seems to be similar in type I and III (Figure 1.8). The critical protein involved in the processing of crRNA belongs to the Cas6 family (a metal-independent enzyme containing two RRM-type RNA-binding domains) endoribonuclease (Carte et al., 2008; Gesner et al., 2011; Haurwitz et al., 2010; Richter et al., 2012; Sashital et al., 2011). Due to the presence of a palindromic sequence in repeats, most type I pre-crRNAs contain stem-loop structures. Cas6 recognises these stem-loop structures and cleaves RNA at the bottom of the 3' end of the stem (Figure 1.8) (Gesner et al., 2011; Haurwitz et al., 2010; Sashital et al., 2011). The cleavage generates mature crRNA containing a full spacer sequence and short repeat derived 5' handle (8 nt handle in *E. coli*) and 3' stem-loop (Figure 1.8). Exceptionally, in the type I-C system, Cas6 family protein is absent, and Cas5 plays the functional role of crRNA maturation. Here, the recognition and cleavage mechanism exhibited by Cas5 is similar to Cas6 (Figure 1.8) (Garside et al., 2012; Nam et al., 2012; Punetha et al., 2014). After cleavage, both Cas5 and Cas6 remain bound to the structured repeat of crRNA (Figure 1.8) and provide a framework for the assembly of CRISPR-associated ribonucleoprotein (RNP) complex for antiviral defence (Cascade complex) (Jore et al., 2011; Sashital et al., 2011). In contrast, type I-A and I-B repeats are non-palindromic in nature (Figure 1.8) (Koonin et al., 2017b; Kunin et al., 2007). Here, Cas6 dimers remodel repeats to form a transient stem-loop structure favourable for its maturation (Reeks et al., 2013; Richter et al., 2013; Sefcikova et al., 2017; Shao and Li, 2013; Shao et al., 2016). Cas6 in type I-A and I-B is a multi-turnover enzyme that leaves crRNA after cleavage and does not associate with the Cascade complex (Charpentier et al., 2015). In bacteria like *Thermus thermophilus* and *Pyrococcus furiosus*, multiple type I systems are found. Here Cas6 is not present in all the subtypes, and evidently, CRISPR array/crRNA is shared during maturation/interference (Majumdar et al., 2015; Staals et al., 2014).

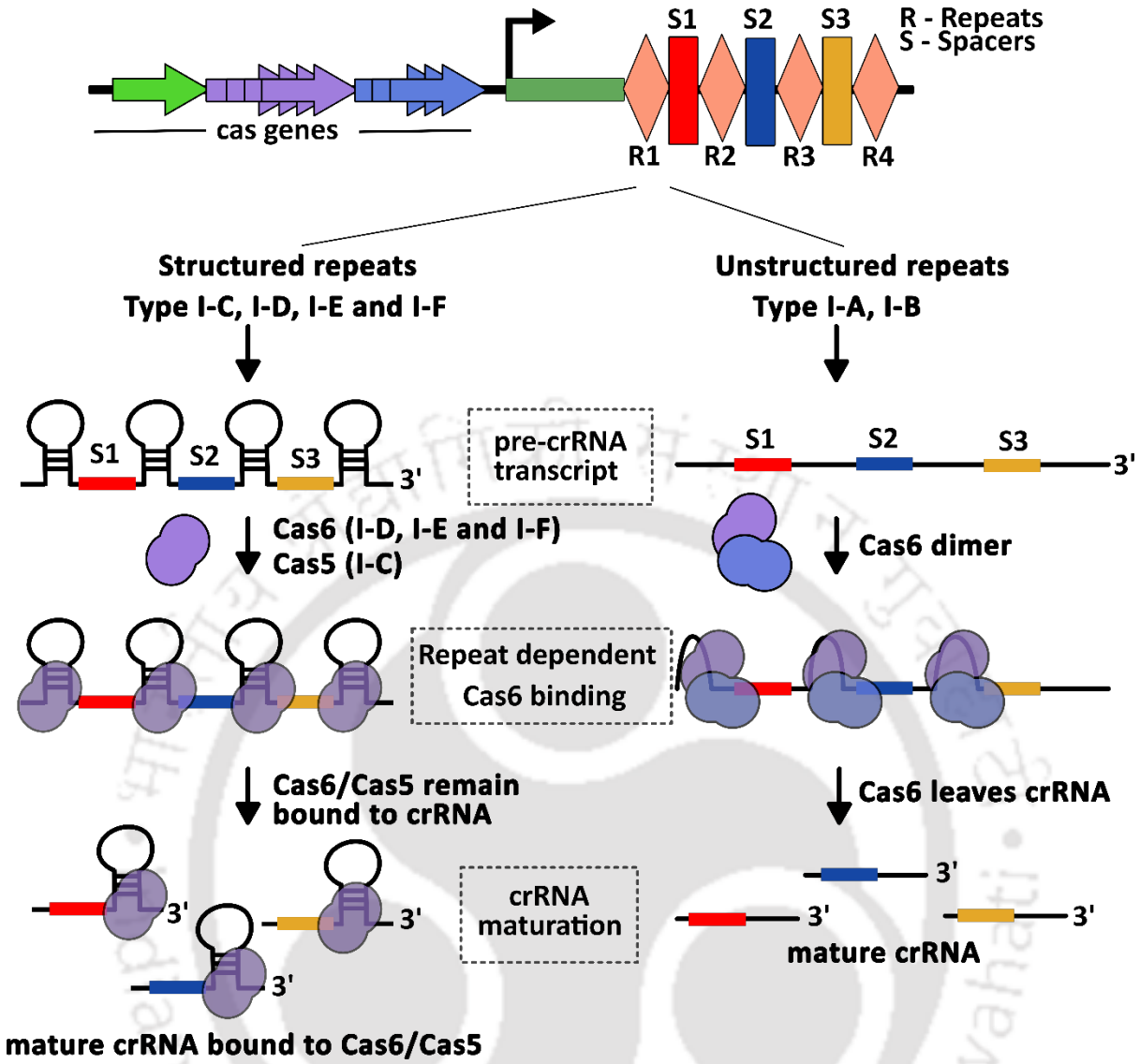

Figure 1. 8: Processing of pre-crRNA in Class 1 CRISPR-Cas system.

Type I systems contain both structured and unstructured repeat sequences. **Left**: The CRISPR array is transcribed into a long pre-crRNA transcript. Cas6-family protein / Cas5 (I-C) recognises the stem-loop structure and processes the crRNA by cleaving at the base of the step-loop. The resulting mature crRNA contains a complete spacer sequence and partial repeat region. The crRNA bound to Cas6/Cas5(I-C) allows the assembly of other Cascade proteins. **Right**: The repeats in I-A and I-B are non-palindromic. Cas6 induces a conformational change in repeat sequence to form a hair-pin like structure. Cas6 recognises and cleaves within the repeats. Here, Cas6 is a multi-turnover enzyme and does not form a part of the Cascade complex.

In the type III system, Cas6 exhibits high sequence similarity to corresponding type I-A and I-B homologs (Figure 1.8). Moreover, type III repeats are not structured or partially structured (Kunin et al., 2007). Therefore, the crRNA maturation process in type III systems follow a similar remodelling of repeats in pre-crRNA and subsequent generation of the final crRNA (Carte et al., 2008; Hatoum-Aslan et al., 2011). Like type I-A and I-B, Cas6 in type III releases crRNA after cleavage and does not form a part of the Cascade complex (Plagens et al., 2014; Richter et al., 2012). Interestingly, like type I-C, Cas6 homologs are not present in type III-C and III-D, and Cas5 may complement the role of Cas6 (Makarova et al., 2020). In type IV, Csf5 (Cas6 homolog) facilitates the maturation of crRNA (Ozcan et al., 2019).

1.1.7.2 Maturation of crRNA in Class 2 systems

Unlike Class 1, Class 2 systems use a single-subunit multidomain effector nuclease and, in some cases, non-CRISPR host protein for crRNA maturation (Deltcheva et al., 2011; Makarova et al., 2020; Makarova et al., 2018). Additionally, in type II and V-B, a trans-encoded small RNA called trans-activating crRNA (tracrRNA) is also crucial during the maturation process (Deltcheva et al., 2011; Shmakov et al., 2015; Zhang et al., 2013). The initiation of the maturation process in the type II-A system starts with the binding of tracrRNA to the complementary sequence on the repeat region on pre-crRNA leading to the formation of tracrRNA:crRNA duplex (Figure 1.9). The effector protein Cas9 (II-A) stabilises the duplex and further recruits the host RNaseIII protein for the initial processing of crRNA (Figure 1.9) (Deltcheva et al., 2011). Subsequently, another unknown host nuclease trims 5' repeat derived overhang (Figure 1.9). The tracrRNA bound mature crRNA with ~20 nt 5' spacer and ~19-22 nt 3' repeat-derived region along with Cas9 completes the interference complex required for target cleavage (Deltcheva et al., 2011; Gasiunas et al., 2012; Jinek et al., 2012a). Conversely, each repeat contains promoter elements in the type II-C system that lead to transcription of individual crRNA (Dugar et al., 2013; Zhang et al., 2013). Like type II-A, the processing of crRNA in type II-C also depends on tracrRNA:crRNA duplex; however, functional interference complex can also form in the absence of matured crRNA (Zhang et al., 2013).

Like type II, type V and VI also requires an effector nuclease for both maturation and interference. Cas12a is the effector nuclease in type V-A systems and helps in tracrRNA independent maturation of crRNA (Figure 1.9). Cas12a recognises structured repeats in pre-crRNA and cleaves within the repeats (Dong et al., 2016; Fonfara et al., 2016). Later, an

unknown host nuclease trims both 5' and 3' ends, leading to the maturation of crRNA with a 5' repeat derived hairpin loop. In type V-C and V-D, a short complimentary untranslated RNA (scoutRNA) is required for maturation (Figure 1.9) (Harrington et al., 2020). Compared to tracrRNA, scoutRNA is short and attains a different secondary structure. However, scoutRNA remains bound to Cas12 after crRNA maturation and helps in interference (Harrington et al., 2020). Here, host RNaseIII is not required for pre-crRNA trimming. On the other hand, type V-B, V-E, V-F and V-G contains tracrRNA, and the crRNA maturation mechanism is expected to be similar to type II-A (Cas9) (Makarova et al., 2020; Shmakov et al., 2015).

In type VI, tracrRNA is absent, and Cas13 (effector nuclease) is required for crRNA maturation (Figure 1.9) (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Makarova et al., 2020; Shmakov et al., 2015). Cas13 recognises structural and sequence motifs on repeat region and cleaves upstream of the stem-loop on pre-crRNA (East-Seletsky et al., 2016). Intriguingly, in type VI-A, crRNA maturation is not a compulsory requirement for functional interference; however, a mature crRNA increases the overall efficacy of target cleavage during interference (East-Seletsky et al., 2016). In type VI-B, the repeat lengths are not constant within the CRISPR array; however, the length of the spacer sequence is uniform (~30 nt). A mature crRNA may contain either 36 nt or 88 nt long repeats, and both the crRNAs are equally efficient in target cleavage (Smargon et al., 2017).

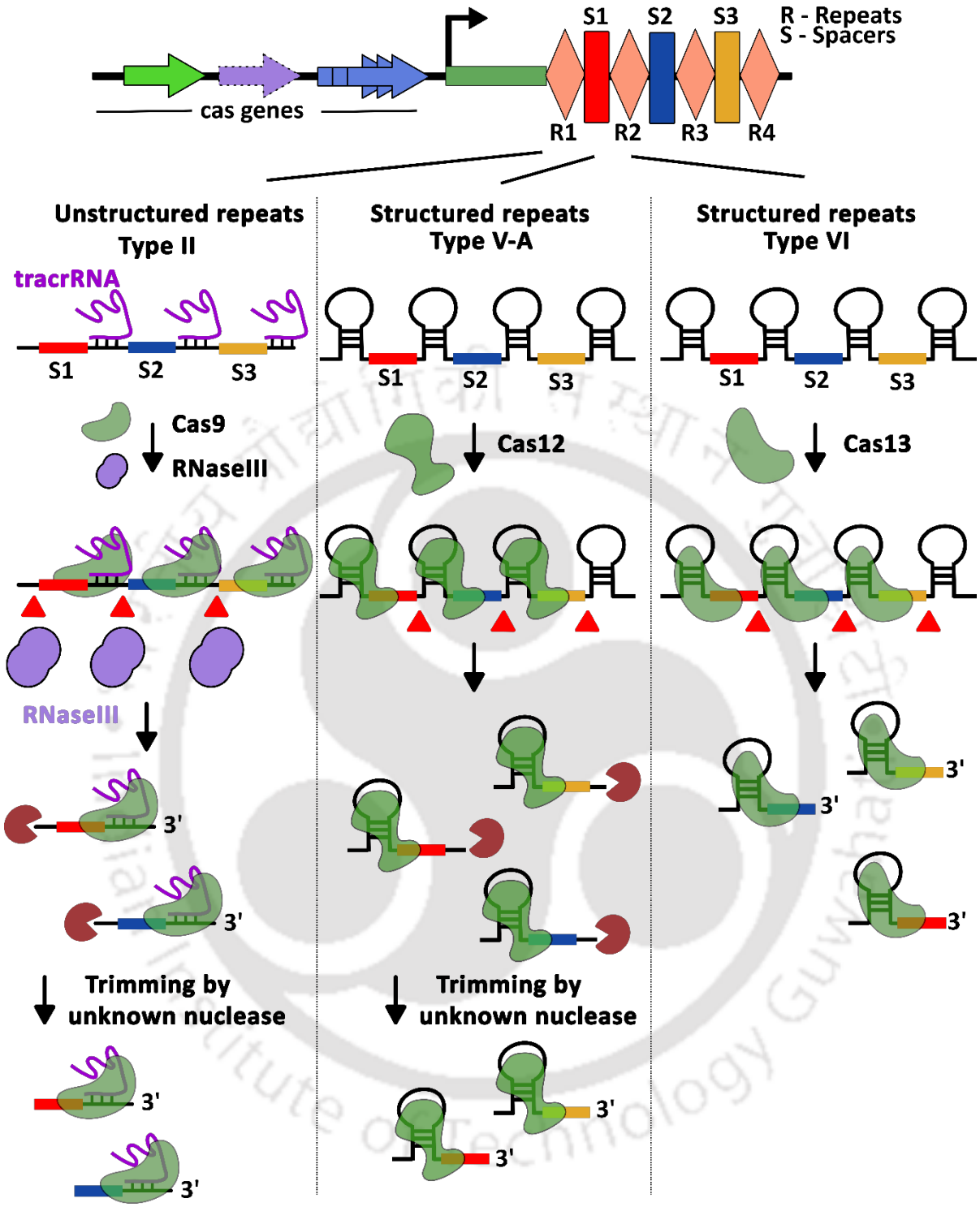

Figure 1. 9: Processing of pre-crRNA in Class 1 CRISPR-Cas system.

In class 2 systems, there are no reserved Cas proteins for crRNA processing. For crRNA maturation, these systems employ the effector nuclease and host RNaseIII (only in the type II system).

Left: Type II systems require tracrRNA, Cas9 and RNaseIII for crRNA maturation. The tracrRNA binds to the complementary repeat region acting on the long pre-crRNA transcript. The structure of tracrRNA guides the binding of Cas9, which later outsources the processing to host RNaseIII. RNaseIII nicks the 5' side of the Cas9 and separates individual crRNA units, which are then processed by an unknown host nuclease.

Centre: The structured repeats in type V-A is recognised by Cas12, which makes a nick at the 3' end of the spacer. The individual crRNAs are then further processed by an unknown nuclease. Type V shows variations in crRNA processing; in type V-C and V-D, scoutRNA is also required, whereas, in type V-B, V-E, V-F and V-G, tracrRNA is present. In this figure, only type V-A processing is depicted.

Right: In type VI, only Cas13 is responsible for crRNA maturation. Here, complete maturation of crRNA is not a compulsory requirement for active CRISPR interference, and immature crRNA can also help in the degradation of target DNA. Red triangles denote nicking positions. Unknown nucleases are depicted using brown colour half-circles.

1.1.8 CRISPR interference

The final step or the hallmark of the CRISPR-Cas system is the sequence-specific destruction of invaders/MGEs. The sequence specificity is attained through the target DNA complementary spacer region on crRNA within the interference complex. Since the crRNA is derived from the CRISPR array, the possibility of interference complex binding to 'self' CRISPR DNA and leading to cleavage of the 'self' genome cannot be avoided. However, such an autoimmune response cannot be favoured during evolution. Hence, there is a mechanism to differentiate between self and non-self during CRISPR interference. In most CRISPR-Cas systems, the PAM sequence is not acquired or partially integrated into the CRISPR array during adaptation (Goren et al., 2012; Heler et al., 2015). Hence, PAM is only present in invading MGEs and absent in the CRISPR locus. Since PAM is crucial for functional CRISPR interference, cleavage of 'self' CRISPR DNA is ruled out (Abudayyeh et al., 2016; Anders et al., 2014; Deveau et al., 2008; Hayes et al., 2016; Mojica et al., 2009; Westra et al., 2013; Yosef et al., 2012; Zetsche et al., 2015). On the contrary, type III systems do not employ the PAM sequence for such discrimination. Here, 5' and 3' portions of repeats on the matured crRNA act as a marker to identify invading MGEs. The complementary base pairing between crRNA and 'self' CRISPR DNA is not favourable for interference machinery to degrade DNA, thereby protecting the 'self' genome (Marraffini and Sontheimer, 2010).

1.1.8.1 Class 1 CRISPR-Cas Interference

Class 1 CRISPR-Cas systems are subdivided into type I, III and IV, where a multi-subunit interference complex is responsible for interference. Most of our understanding of interference in Class1 is derived from type I and III. Knowledge about the recently discovered type IV systems' functioning is in a nascent stage and requires a thorough study. In the type IV system, Cas5 and Cas7 are present in all the subtypes (IV-A to IV-D), with Cas8 and Cas11 appearing in a few subtypes (Makarova et al., 2020; Ozcan et al., 2019). Overall, full interference machinery in type IV has not been identified yet. The co-existence of type IV with other CRISPR-Cas systems suggests shared interference machinery during target cleavage (Pinilla-Redondo et al., 2020).

1.1.8.1.1 Type I CRISPR-Cas Interference

Type I is the most abundant CRISPR-Cas system across bacteria and archaea (Koonin et al., 2017a; Makarova and Koonin, 2015; Makarova et al., 2015). Invading MGE is cleaved with the help of a multi-subunit ribonucleoprotein surveillance complex called Cascade. Cascade complex harbours crRNA, which imparts sequence specificity to the system. Surveillance complex (Cascade) locates the invading MGE, and an effector nuclease, Cas3 (nuclease-helicase), cleave the invading MGE through its nuclease activity (Brouns et al., 2008). The overall architecture of Cascade is conserved; however, its composition can vary depending on the CRISPR subtypes. Among all the seven type I subtypes (Makarova et al., 2015), type I-E is the most characterised subtype; thus, it is a suitable model for understanding type I interference.

1.1.8.1.2 Type I Cascade complex composition and architecture

Cascade complex in type I-E system consists of all the identified proteins in type I *viz.* Cas5, Cas6, Cas7, Cas8 and Cas11 along with crRNA (Jackson et al., 2014a; Makarova et al., 2020). Cas5, Cas6 and Cas7 contain RNA recognition motif (RRM) and directly interact with crRNA in the complex (Brouns et al., 2008; Jore et al., 2011). After the maturation of crRNA, Cas6 remains bound to 3’ repeat derived stem-loop of crRNA, which acts as a scaffold for other subunits to assemble (Gesner et al., 2011; Sashital et al., 2011). One molecule of Cas6 forms the head of the structure (Figure 1.10A-B). The helical backbone of the Cascade complex is composed of seven subunits of Cas7 molecules that make direct contact with the 32 nt spacer

region of crRNA (Figure 1.10A). Here, Cas7 molecules attain a closed palm-like shape with interconnected thumb domains, which stabilises the structure. Cas5 is present at 5' repeat derived end of crRNA (Jackson et al., 2014a; Jore et al., 2011; Mulepati et al., 2014; Wiedenheft et al., 2011; Zhao et al., 2014). The interwoven arrangement of Cas7 and Cas5 molecules segregates the spacer region of crRNA into six sections (Figure 1.10A-B and 1.11A). The thumb domain in Cas7 flips one base in the reverse direction after every five bases on crRNA (Figure 1.11A). Although Cas5 bears no noticeable sequence similarity with Cas7 thumb, it performs a similar base flipping function at the last nucleotide of the 5' repeat handle. The resulting flipped base is not available for base pairing with target DNA (Figure 1.11A). As a result, six units of 5 nucleotide stretch on crRNA base pairs with the target DNA during interference (Jackson et al., 2014a; Mulepati et al., 2014; Zhao et al., 2014). Cas8 and Cas11 are large subunit (LS) and small subunit (SS) of the Cascade complex, respectively. Two subunits of Cas11 interacts with Cas7 in the 'belly region of the Cascade. Subunit Cas8 interacts with Cas7 and Cas5 and forms a part of the tail region (Jackson et al., 2014a; Mulepati et al., 2014; Zhao et al., 2014). After the assembly of all the Cascade subunits, the final structure resembles a seahorse shape (Figure 1.10A) (Jackson et al., 2014a; Jore et al., 2011; Wiedenheft et al., 2011).

The interference module in type I-C CRISPR operon contains four genes viz. *cas5, cas7*, *cas8 and cas11* (Figure 1.10C). Unlike type I-E, Cas6 is absent in the type I-C system, and as a result, the 3' repeat derived stem-loop end of the crRNA is not capped (Makarova et al., 2015; Nam et al., 2012; Punetha et al., 2014). After maturation of crRNA, Cas5 subunit caps 5' repeat derived handle and acts as a scaffold for other subunits to assemble (Garside et al., 2012; Koo et al., 2013; Nam et al., 2012). Like type I-E, in type I-C, seven subunits of Cas7 forms the backbone, interacting directly with ~32 nt spacer region of crRNA (Hochstrasser et al., 2016; O'Brien et al., 2020). The cryo-EM structure of type I-C Cascade in *Desulfovibrio vulgaris* shows that Cas8 is a fusion of large and small subunit. Cas8 subunit assembles near the 5' end of crRNA interacting with Cas5 and the seventh subunit of Cas7 (Figure 1.10C) (Hochstrasser et al., 2016). Recently it has been identified that the standalone Cas11 (SS), which is internally translated from the *cas8* gene, also forms a part of the Cascade complex (Figure 1.10C) (McBride et al., 2020; O'Brien et al., 2020). The presence of the ribosome-binding site (followed by start codon) within the *cas8* gene leads to an independent translation of Cas11. Two subunits of Cas11 make the 'belly' of the seahorse-shaped Cascade complex (Figure 1.10C) (O'Brien et al., 2020). Overall, the Cascade complex in type I-C displays the

following composition: $(Cas5)_1$-$(Cas7)_7$-$(Cas8)_1$-$(Cas11)_2$ (Figure 1.10C) (McBride et al., 2020; O'Brien et al., 2020).

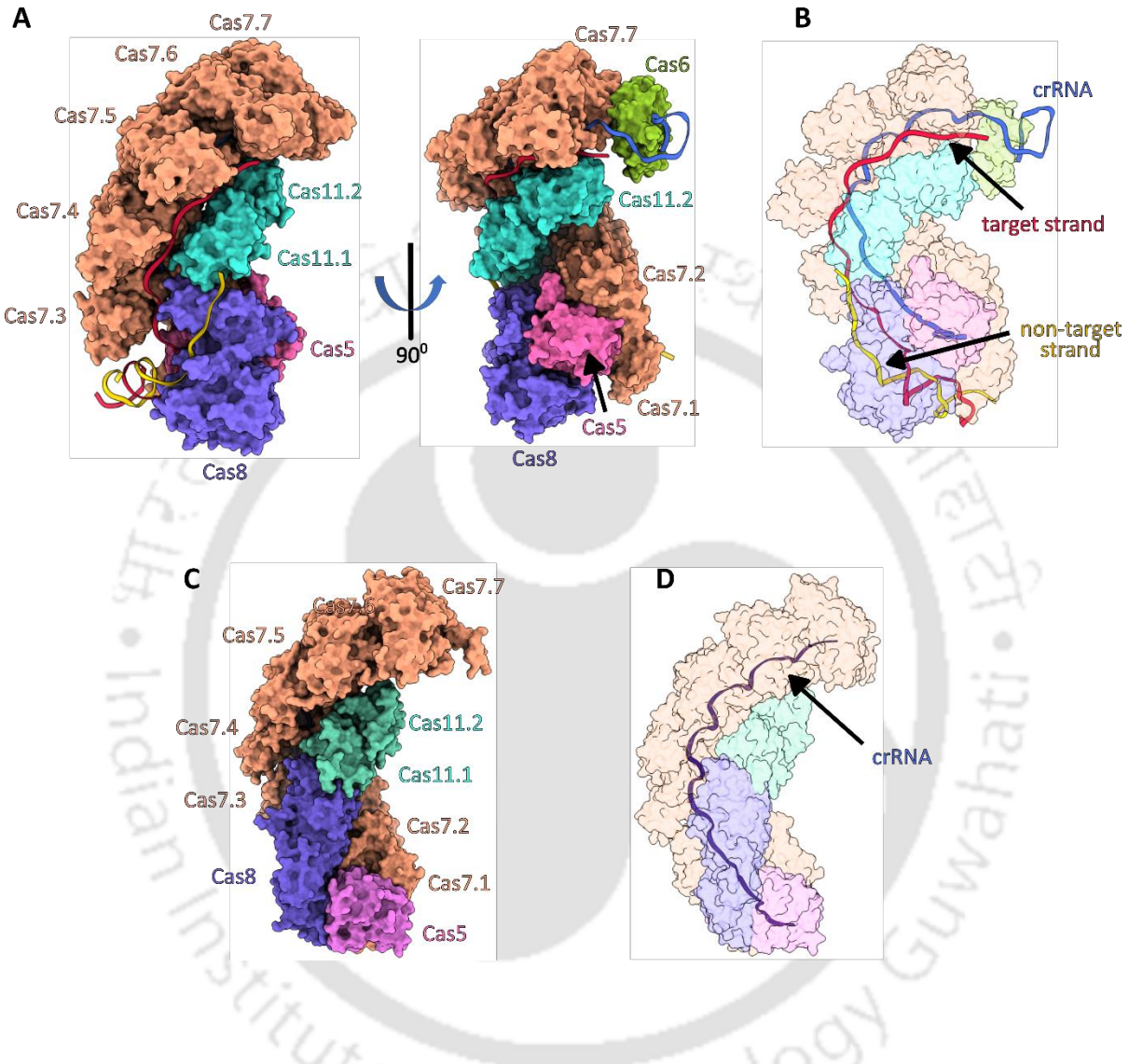

Figure 1. 10: Cascade composition in type I-E and type I-C CRISPR-Cas system.

A. Seahorse shaped structure of type I-E Cascade complex bound to target-DNA from *E. coli* is shown above. Cas6, after crRNA processing, remain bound to the 3' end, allowing the assembly of the whole Cascade complex. Seven units of Cas7 forms the helical backbone. Cas11 (SS) from the belly region and Cas8 (LS) form the tail region of the structure. PDB ID: 5H9F.

B. The structure shows crRNA and target DNA present within the complex. The target strand makes a complementary base pairing with the spacer region of the crRNA displacing the non-target strand and forming an R-loop. A single subunit of Cas6 caps the stem-loop region of the crRNA. PDB ID: 5H9F.

C. Atomic model of *D. vulgaris* type I-C Cascade (PDB ID: 7KHA) derived from the associated Cryo-EM structure (EMD-22876) is shown in the figure. Cas6 is absent in type I-C and 3'end of crRNA is capped by Cas7 subunit. Two subunits of Cas11 and the C-terminal region of Cas8 form the belly region of the Cascade. The N-terminal of Cas8 forms the large-subunit and interacts with Cas5 and Cas7.

D. The figure shows type I-C crRNA present within the Cascade complex. The 3' stem-loop region is poorly resolved and not present in the above model.

Cascade complex in type I-A, I-B I-D, I-F and I-G shows a few variations in the architecture (Makarova et al., 2020). Unfortunately, information about type I-A, I-B, and I-G Cascade complex is not available. In type I-F, large and small subunits are absent, and Cas5 and Cas7 replace their functional role (Gleditzsch et al., 2016; Pausch et al., 2017). In type I-D, the functional role of the large subunit is replaced by Cas10, which also makes this subtype a connecting link between type I and III (Makarova et al., 2020; McBride et al., 2020).

1.1.8.1.3 Mechanism of target identification in type I system

CRISPR interference starts with the surveillance and identification of invading MGE through sequence specificity derived from crRNA within the Cascade complex. A critical step during target recognition is identifying the PAM sequence by Cas8 (LS) (Figure 1.11B) (Hayes et al., 2016). The L1 loop, Gly-rich loop and Gln-wedge present in the Cas8 or the large subunit directly interacts with the PAM sequence (Figure 1.11B) (Hayes et al., 2016). Once the PAM is identified, conformational changes in the glycine-rich regions on the Cas8 subunit of the Cascade complex leads to the unwinding of target DNA from the PAM proximal end (Figure 1.11) (van Erp et al., 2017). As the unwinding proceeds further towards the PAM distal end, a strong target binding is enabled by complementary base pairing between crRNA and the target stand (Figure 1.11A) (Xiao et al., 2018; Xiao et al., 2017a). The three-stranded DNA:RNA hybrid with an unpaired DNA strand is generated after successful target binding, and this structure is called R-loop (Figure 1.11A). The unpaired DNA strand in the R-loop is the non-target strand and performs a crucial role in effector nuclease loading (Hayes et al., 2016; Jackson et al., 2014a; Mulepati et al., 2014; Xiao et al., 2017a). Mutation in the PAM and 'seed' (PAM proximal crRNA paired 8 nt on target strand) region hampers target DNA binding and interference (Datsenko et al., 2012; Semenova et al., 2011; Xiao et al., 2017a). Disruption of the L1 loop weakens target DNA binding drastically (Hayes et al., 2016). The curvature of

the Cascade complex bends the target strand to ~120°, and the backbone subunit (Cas7) maintain the stability of the structure (Hayes et al., 2016; Xiao et al., 2018; Xiao et al., 2017a). The non-target strand is bound to Cas11 subunits, which stabilize R-loop (Figure 1.11A) (Hayes et al., 2016; Xiao et al., 2017a). Since the type I system is diverse, some variations can be seen across different subtypes. However, the overall mechanism of target identification and binding seems to be conserved.

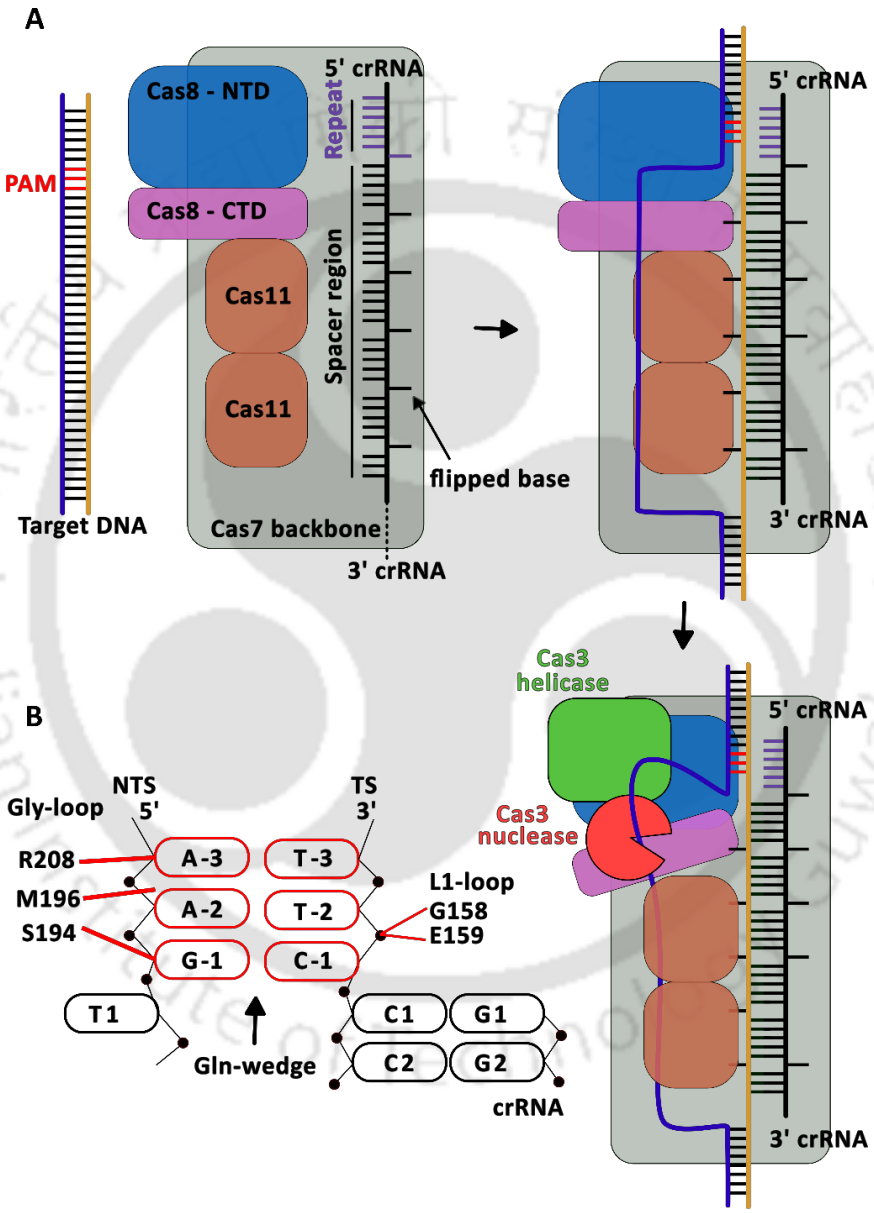

Figure 1. 11: Target identification and R-loop formation.

A. The above figure shows the schematic representation of target DNA identification and R-loop formation in the *E. coli* type I-E system. The double-stranded target DNA with target strand (yellow) and non-target strand (blue) containing cognate PAM 'AAG' (for *E. coli*) is shown in the figure. The binding of Cas7 to crRNA induces base flipping after every 5th base

starting from the last nucleotide of the 5' repeat handle. The dotted line at the 3' end of the crRNA represents the Cas6 bound stem-loop structure (Figure 1.8, left). The search for the target starts with a sampling of the minor grooves of the invading phages/MGEs. On locating the correct PAM sequence, the Gly-loop, Gln-wedge and L1-loop of the Cas8 (LS) initiate the R-loop formation from PAM proximal end. A further unwinding of DNA leads to complete displacement of the non-target strand and complementary base pairing between target-strand and spacer region of crRNA. Due to Cas7 induced base flipping, every 6th base on the target strand does not take part in binding. Local conformational changes in the Cas8 CTD lead to sequestration of non-target strand by Cas11 and forming a flexible bulge at PAM proximal end. Subsequently, the recruitment of Cas3 results in target degradation.

B. Glycine loop, Glutamine wedge and L1 loop present in the Cas8 subunit are essential for target identification and initiation of R-loop formation. The figure represents the amino acid residues that directly interact with the bases at the PAM region in *E. coli*.

1.1.8.1.4 The effector nuclease/Cas3 and target degradation

After binding to target DNA, the LS and SS of the Cascade complex undergo a few structural and spatial changes that lead to the recruitment of effector nuclease Cas3 for target cleavage (Hayes et al., 2016; Xiao et al., 2017a). Typically, Cas3 is a fusion of N-terminal HD-nuclease domain and C-terminal superfamily 2 DExD/H Box helicase (Figure 1.12A) (Gong et al., 2014; Huo et al., 2014a; Jackson et al., 2014b; Sinkunas et al., 2011; Sinkunas et al., 2015); however, in some variants, either nuclease and helicase domains are expressed separately (type I-A) or fused to other Cas proteins (e.g. Cas2-Cas3 fusion in type I-F) (Huo et al., 2014a; Makarova et al., 2015; Makarova et al., 2020). In type I-G, Cas3 domains are interchanged, containing N-terminal helicase and C-terminal nuclease (Makarova et al., 2020). The function of the nuclease domain is supported by metal ions such as magnesium and cobalt (Beloglazova et al., 2011; Mulepati and Bailey, 2011; Sinkunas et al., 2011). After loading onto the single-stranded DNA (ssDNA), the helicase domain of Cas3 unwinds DNA in a 3' to 5' direction in the presence of metal ion and using energy from ATP-hydrolysis (Sinkunas et al., 2011). The highly conserved arginine-rich channel on the helicase domain is believed to aid stable capture and translocation on ssDNA (Xiao et al., 2017a). Apart from two major domains, Cas3 also harbours a C-terminal domain (CTD), although its precise function is not adequately understood (Figure 1.12A).

In *Pseudomonas aeruginosa* (type I-F), Cas2 and Cas3 are fused, and the nuclease activity of Cas2-Cas3 is inhibited when Cas1 is present in the complex. However, the Cascade

complex bound *bonafide* target reactivates Cas2-Cas3 nuclease activity, leading to target degradation (Rollins et al., 2015). The fusion of Cas2-Cas3 in type I-F hints towards cooperation between adaptation and interference machinery. In *Thermobifida fusca* (type I-E), standalone Cas3 or Cas3-Cascade complex is incompetent to dislocate a DNA binding protein (e.g. RNA polymerase) barrier during translocation (Dillard et al., 2018). Interestingly, the addition of the Cas1-Cas2 complex to the target bound Cas3-Cascade complex ensues the formation of a larger Primed Adaptation Complex (PAC), which can displace protein barriers efficiently (Dillard et al., 2018). However, the mechanism of the displacement of 'roadblock' is not understood.

After the validation of *bonafide* target DNA by the cascade complex, the change in conformation of the Cas8 subunit divulges the Cas3 interaction site (Figure 1.11A). It seems Cas3 and Cascade complex co-operate via a combination of steric hindrance, hydrophobic and electrostatic interaction (Xiao et al., 2018; Xiao et al., 2017a). An initial nick by nuclease domain facilitates the loading of Cas3 on the displaced non-target strand in the R-loop created by the Cascade complex (Redding et al., 2015). Subsequently, Cas3 unwinds double-stranded DNA (dsDNA) in 3' to 5' direction and regularly feeds ssDNA to HD-nuclease domain, creating a single-stranded gap of approximately 200-300 nt on the non-target strand (Figure 1.12B) (Huo et al., 2014a; Westra et al., 2012a; Zhao et al., 2014). However, this is not sufficient to completely degrade invading MGE. It is believed that unknown host nucleases might help destroy target DNA or the Cascade independent non-specific nuclease activity of Cas3, which has been shown *in vitro*, might be sufficient for large scale destruction (Mulepati et al., 2014; Redding et al., 2015; Sinkunas et al., 2013).

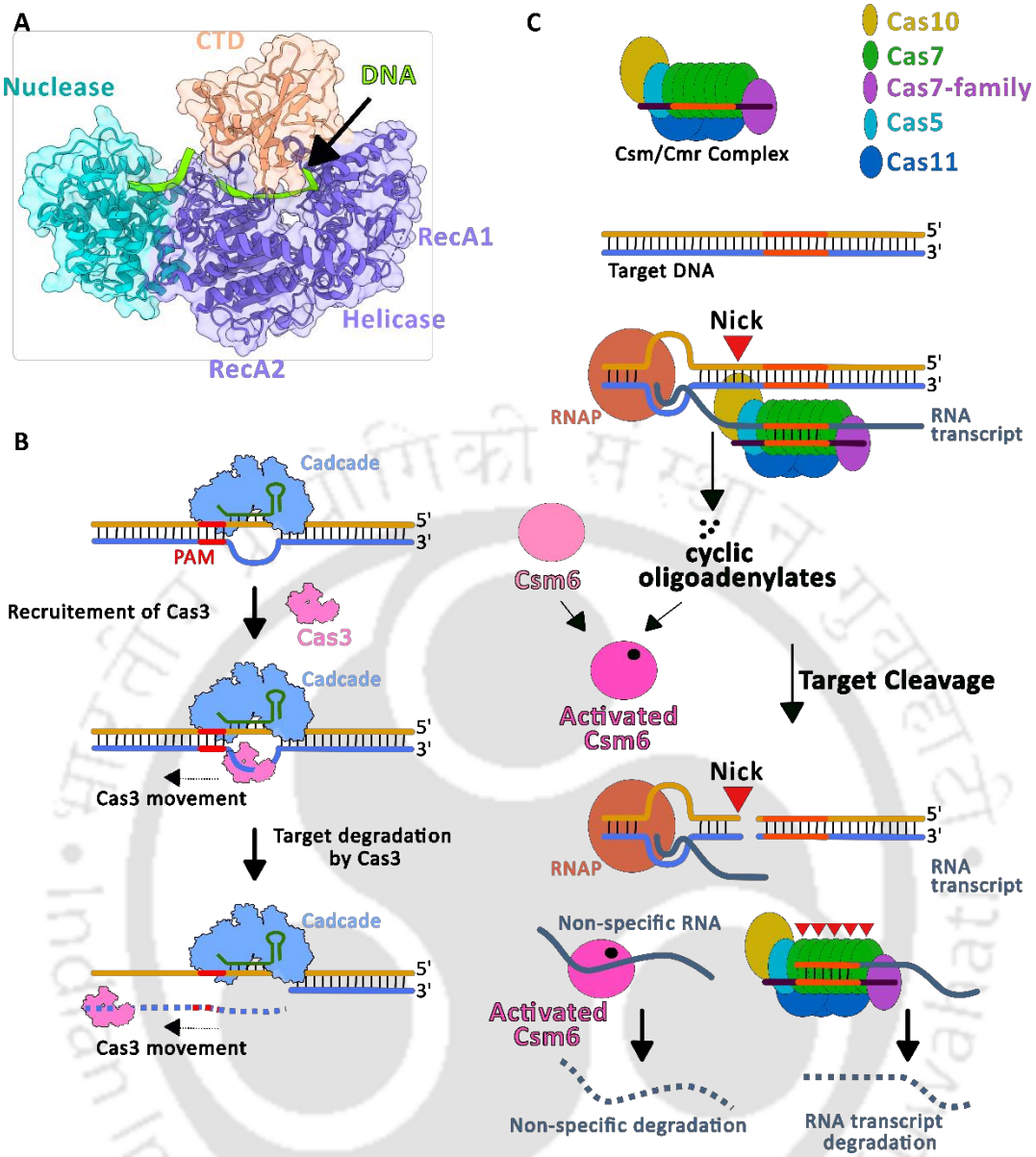

Figure 1. 12: Target degradation in Class 1 CRISPR-Cas system.

A. Molecular structure of type I-E Cas3 from *T. fusca* is shown in the figure. Cas3 consists of three domains; Nuclease, Helicase and C-terminal domain (CTD).

B. The figure represents the interference mechanism in the type I system. After binding to target DNA, Cascade undergoes a conformational change and recruits Cas3. Cas3 loads onto the non-target strand and subsequently unwinds and degrades DNA in a 3' to 5' direction.

C. Type III interference machinery has two crucial components; the Cas10-Csm complex harbouring crRNA and Cas6 (RNase with HEPN domain). The Cas10 shows dual activities – nicking the target DNA and conversion of ATP to cyclic oligoadenylates (cOA). The identification of the target begins with searching the crRNA complementary sequence on the transcribing nascent RNA. The binding of the Csm complex to nascent RNA tethers it to the

transcribing DNA, thereby activating the double-stranded DNA cleavage by Cas10. On the other hand, cOA molecules activate Csm6, which subsequently degrade non-specific RNA present in the vicinity.

1.1.8.1.5 Type III CRISPR-Cas Interference

The type III Interference complex contains a Cascade like structure (Csm for III-A and Cmr for III-B) and shows high similarity to type I interference complex in architecture and composition (Abbondanzieri et al., 2005; Hochstrasser et al., 2014; Jackson et al., 2014a; Makarova et al., 2011b; Mulepati et al., 2014; Osawa et al., 2015; Staals et al., 2014; Taylor et al., 2015; Zhao et al., 2014). However, unlike type I, type III can target both invading RNA and DNA of MGE (Deng et al., 2013; Elmore et al., 2016; Goldberg et al., 2014; Samai et al., 2015). Most of our understanding about type III systems is from III-A and III-B, and most recently discovered III-C, III-D and III-F are not yet studied. After crRNA maturation, Cas5 binds to a 5' repeat derived handle to which other Cas proteins assemble. Cas7 family proteins (Csm3 and Csm5 in III-A and Cmr1, Cmr4 and Cmr6 in III-B) form the backbone of the complex by holding on to the spacer region of crRNA. Cas10 (Csm1 in III-A and Cmr2 in III-B) and Cas11 (Csm2 in III-A and Cmr5 in III-B) form large and small subunits, respectively (Makarova et al., 2020; Osawa et al., 2015; Staals et al., 2014; Taylor et al., 2015). In type III, the degradation of the target DNA can happen only when it is transcriptionally active (Deng et al., 2013; Elmore et al., 2016; Estrella et al., 2016; Goldberg et al., 2014; Kazlauskiene et al., 2016; Samai et al., 2015). After binding to nascent target RNA using crRNA mediated complementary base pairing, the Cas7 family subunits of interference complex cleaves target RNA after every sixth base. The transcriptionally active DNA is cleaved by the HD-nuclease domain of the Cas10 subunit (Figure 1.12C) (Osawa et al., 2015; Samai et al., 2015; Staals et al., 2014; Tamulaitis et al., 2014; Taylor et al., 2015). Csm6 and Cxs1 family RNase frequently associate with type III complex and have an auxiliary role in interference (Figure 1.12C) (Koonin et al., 2017a; Makarova et al., 2020). Apart from target cleavage, Cas10 (target bound) also converts ATP to cyclic adenylate, which activates RNase Csm6 leading to degradation of non-specific proximate RNA transcripts (host and invader RNA) (Figure 1.12C) (Kazlauskiene et al., 2017; Kazlauskiene et al., 2016; Niewoehner et al., 2017).

1.1.8.2 Class 2 CRISPR-Cas Interference

In Class 2, a multi-subunit cascade-like complex is absent, and a single multi-domain Cas protein carries out target degradation. The effector nuclease notation varies based on the size and domain architecture (Cas9 in type II, Cas12 in type V and Cas13 in type VI). The application of Cas9 in genome editing makes it one of the highly characterized CRISPR-Cas systems.

1.1.8.2.1 Type II CRISPR-Cas interference

The interference complex in type II consists of Cas9, crRNA and a tracrRNA (Figure 1.13A) (Barrangou et al., 2007; Deltcheva et al., 2011; Gasiunas et al., 2012; Jinek et al., 2012a; Jinek et al., 2014a; Nishimasu et al., 2014). Cas9 has a bilobed structure containing an alpha-helical recognition lobe (REC), and a nuclease lobe (NUC) fused through a linker with an arginine-rich helix. The positive charge on the bridge helix interacts directly with the crRNA. The NUC lobe contains HNH and RuvC domains that cleave target DNA during interference (Anders et al., 2014; Hirano et al., 2016; Jiang and Doudna, 2015; Jinek et al., 2014a; Nishimasu et al., 2015; Nishimasu et al., 2014). Cas9, guided by crRNA, locates and binds to the target DNA. Subsequently, binding induced conformational change promotes PAM identification by a variable C-terminal domain and unwinding of target DNA from PAM proximal end (Anders et al., 2014; Jiang et al., 2016; Jinek et al., 2012a; Sternberg et al., 2015; Sternberg et al., 2014). Like, type I, incorrect PAM, or ‘seed’ sequence does not favour target degradation. Perfectly matched target DNA leads to conformational activation of HNH and RuvC nuclease domains and stabilization of R-loop. HNH domain cleaves target strand whereas RuvC domain nicks non-target strand resulting in a blunt double-stranded DNA break (Figure 1.13A) (Jinek et al., 2012a; Sternberg et al., 2015).

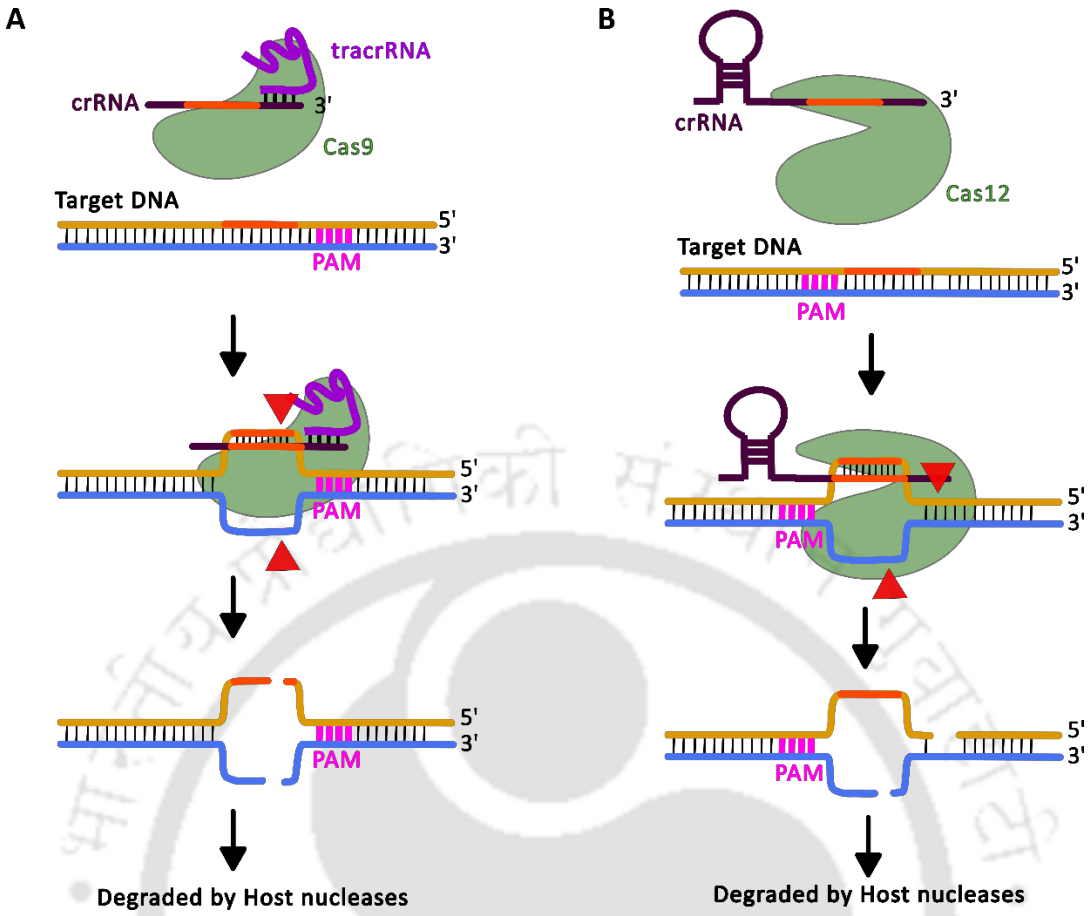

Figure 1. 13: Target Degradation in type II and V CRISPR-Cas interference.

A. The type II CRISPR interference is aided by Cas9, crRNA and tracrRNA. Cas9 identifies the crRNA complementary sequence of the target DNA. Later, the HNH and RuvC domains of Cas9 realigns, leading to double-stranded breaks on the target DNA, followed by complete degradation by host nucleases.

B. Type V-A CRISPR interference does not require tracrRNA and relies solely on structured crRNA and Cas12. After the binding to bonafide target DNA, the RuvC domain of Cas12 makes a staggered cut on the target DNA.

1.1.8.2.2 Type V CRISPR-Cas interference

The functional requirement of the effector nuclease Cas12 varies in type V systems (Fonfara et al., 2016; Makarova et al., 2020; Shmakov et al., 2015; Zetsche et al., 2015; Zetsche et al., 2017). Cas12a (type V-A) does not require tracrRNA, whereas, for an active Cas12b (type V-B), tracrRNA is mandatory. The characterization of Cas12c (type V-C) is yet to be carried out. Like Cas9, Cas12 has a bilobed structure with both REC and NUC lobes. However,

it harbours only a single catalytic RuvC domain (Dong et al., 2016; Gao et al., 2016; Liu et al., 2017a; Yamano et al., 2017). Perfect target binding leads to proper positioning and activation of the RuvC domain, which subsequently cleaves both the target and the non-target stands. The crystal structure of Cas12 with target DNA reveals that both the strands are positioned on the RuvC active site (Yang et al., 2016). However, it is not clearly understood whether the strands are cleaved successively or simultaneously. The cleavage product usually contains 5-7 nt sticky ends (Figure 1.13B) (Fonfara et al., 2016; Shmakov et al., 2015; Yang et al., 2016; Zetsche et al., 2015; Zetsche et al., 2017).

1.1.8.2.3 Type VI CRISPR-Cas interference

The effector nuclease Cas13 is a HEPN containing protein that is specialised to cleave ssRNA (Figure 1.14). Like Cas9 and Cas12, the overall structure is bilobed, and it contains two HENP domains in the NUC lobe (Anantharaman et al., 2013; Makarova et al., 2020; Shmakov et al., 2015). Upon binding to target ssRNA, the HEPN domains are activated, which cleave target ssRNA and collateral RNA (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Shmakov et al., 2015; Smargon et al., 2017). Activation of Cas13 is tolerated by mismatches at the edges of crRNA:target-ssRNA base pairing, but the central region is crucial (Abudayyeh et al., 2016; Liu et al., 2017b). Interaction of Cas13 with protospacer flanking sequence (PFS) is also essential during interference (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Smargon et al., 2017). Once the target-ssRNA meets all the required conditions, two HEPN domains cleave both target and proximally present ssRNAs. During interference, host RNAs are also cleaved, leading to cell dormancy or suicide (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Smargon et al., 2017).

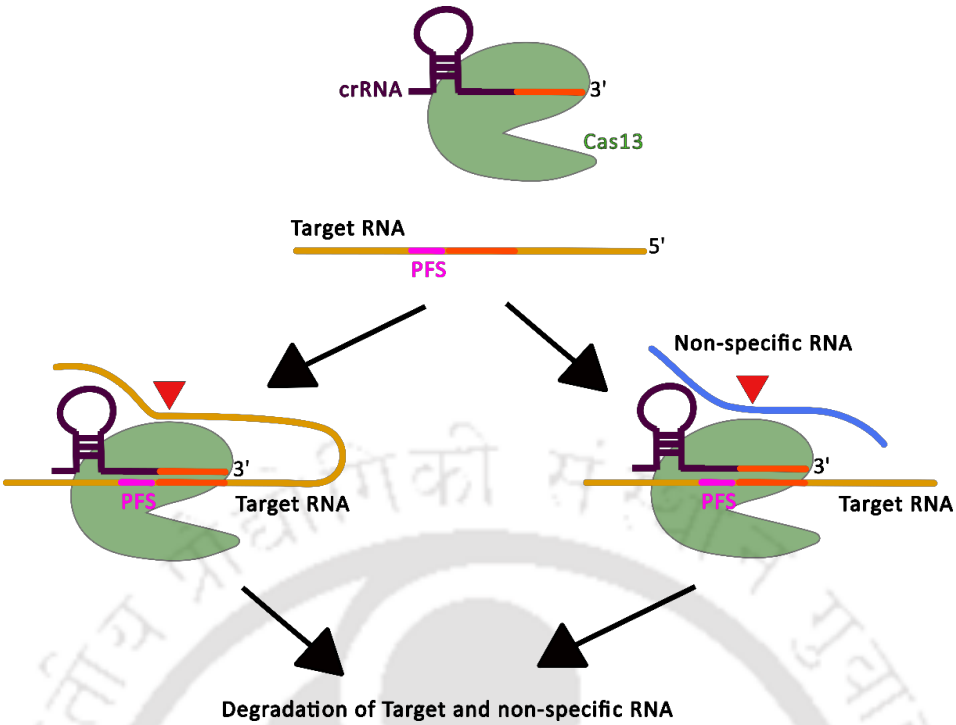

Figure 1. 14: Target Degradation in type VI CRISPR-Cas interference.

In type VI, Cas13, guided by crRNA, binds to target DNA flanked by protospacer flanking sequence (PFS). The conformational changes in Cas13 activate the HEPN domains on the protein. Once activated, Cas13 acts as an RNase and degrades target RNA as well as non-specific RNAs leading to global RNA degradation.

1.1.9 Evasion of bacterial anti-phage defence

Our knowledge about bacterial anti-phage mechanisms has grown several folds; however, we have only begun to understand the regulation, control, and implications of these defence systems on host-pathogen interaction. Even though both bacteria and phages compete for survival, they coexist and exhibit a population equilibrium. To maintain this balance, phages have evolved several measures to counter bacterial defence systems. In order to escape bacterial defence systems, phages may deactivate the host nucleases, lose the target sequence through random mutation or show a higher rate of replication (Hampton et al., 2020). For example, some phages encode genes that modify bases on phage DNA through hydroxymethylation (Warren, 1980), gycosylation (Kruger and Bickle, 1983) or acetamidation (Swinton et al., 1983). These modifications assist certain phages to evade R-M systems (Vasu and Nagaraja, 2013). Many other phages encode methyltransferases (MTases) that protect the 'self' genome from multiple REases (Bickle and Kruger, 1993; Tock and Dryden, 2005). Additionally, some phages have evolved proteins that block host nucleases through direct

interaction. For example, several anti-RM (Goryanin et al., 2018) and anti-CRISPR (Bondy-Denomy et al., 2013) have been discovered that neutralizes the activities of R-M and CRISPR system, respectively.

1.1.9.1 Anti-CRISPR system

Since the CRISPR-Cas interference cleaves target-DNA based on the sequence complementarity, mutations in the genome may help phage escape degradation by the effector nuclease. However, by taking advantage of faulty interference machinery, bacteria can acquire new spacers through the primed adaptation process and subsequently cleave the phage DNA. Therefore, some phages have developed more advanced strategies to evade the attack by directly inhibiting CRISPR interference. Several phages express anti-CRISPR proteins (Acr), which bind to the interference complex and stops target cleavage. The first Acr protein was discovered in phages that infect *Pseudomonas* spp. These phages could infect and propagate in *Pseudomonas aeruginosa* even after the presence of an active type I-F CRISPR-Cas system and the targeting spacers (Bondy-Denomy et al., 2013). Later, it was found that these phages encode Acr proteins (AcrF1 and AcrF2) that directly bind to subunits of the Cascade complex and prevent it from binding to target DNA (Bondy-Denomy et al., 2015; Chowdhury et al., 2017; Peng et al., 2017). In addition, AcrF3 obstructs the binding of Cas3 to the Cascade complex, preventing DNA cleavage (Bondy-Denomy et al., 2015). Several Acr proteins (AcrE1-AcrE4) against type I-E CRISPR-Cas system in *P. aeruginosa* have also been found. Interestingly, these Acr proteins neither block the type I-F system in *P. aeruginosa* nor a related type I-E system in *E. coli* (Bondy-Denomy et al., 2013; Pawluk et al., 2014). This suggests that some of the Acr proteins are specialized for a given subtype and species. On the contrary, AcrF6 protein inhibits both type I-F and type I-E systems in *P. aeruginosa* (Pawluk et al., 2016). Similarly, Acr proteins have been discovered for type I-E, II-A, II-B and II-C CRISPR-Cas system (Dong et al., 2017; Pawluk et al., 2014; Rauch et al., 2017; Shin et al., 2017).

As an exception, Vibrio *cholerae* phages harbour a fully functional type I-F CRISPR-Cas system. This system contains spacers that target a portion of the defence island in *V. cholarae* to silence a non-CRISPR anti-phage defence system (Seed et al., 2013). Furthermore, *Campylobacter jejuni* harbours a type II-C system having Cas1, Cas2, Cas9 and tracrRNA, whereas Cas4 is absent. However, here a Cas4-like protein is complemented through *Campylobacter* phages. Interestingly, all the acquired spacers are of host origin, suggesting

that the Cas4-like protein activates spacer acquisition, preferably from the host genome. Consequently, the host genome is targeted by the interference machinery leading to cell death (Hooton and Connerton, 2014).

1.2 Applications of the CRISPR-Cas system

DNA binding proteins that modify a DNA locus possess a tremendous advantage in genome engineering applications. Restriction enzymes (Danna and Nathans, 1971; Kim and Chandrasegaran, 1994), TALENS (Ramalingam et al., 2014), zinc fingers (Kim et al., 1996; Kim et al., 1997), recombinases are some of the prokaryotic proteins being harnessed for gene editing. However, the process of developing a customized DNA binding protein to bind to a specific target is challenging, slow and requires expertise in protein engineering. On the contrary, the CRISPR-Cas systems rely on the spacer sequence on the crRNA for guidance and sequence specificity. After binding to the target sequence positioned adjacent to the 2-5 nt PAM or PFS sequence, the Cas nuclease cleaves the target nucleic acid. Thus, a sequence-specific cleavage at any location adjacent to PAM or PFS can be achieved using CRISPR-Cas systems by fabricating a unique crRNA or engineered guide-RNA (gRNA) comprising a suitable spacer sequence.

A single effector complex in the Class 2 CRISPR-Cas system makes it suitable for genome engineering applications. Among all the CRISPR variants present in the Class 2 system, *Streptococcus pyogenes* Cas9 (SpCas9) is the first and most widely used genome editing tool (Jinek et al., 2012b; Jinek et al., 2013; Jinek et al., 2014b; Wiedenheft et al., 2012). Cas9 is a dual-RNA guided nuclease requiring crRNA and tracrRNA (figure 1.13A). The function of crRNA and tracrRNA can be replicated with an engineered single gRNA (fusion of crRNA and tracrRNA). Cas9-mediated genome-editing functions by generating a double-stranded break (DSB) and consequently initiating a DNA damage response and inducing an endogenous DNA repair mechanism. There are two repair mechanisms: non-homologous end-joining (NHEJ) and homology-directed repair (HRD). NHEJ is error-prone and often results in insertion or deletion (indel) and can thus create knockouts. On the other hand, HDR relies on homologous recombination and can be diverted to a precise location using a template DNA sequence. This leads to the exact alteration of the genome, as is specified by the homologous template. Additionally, a nuclease mutant of Cas9 (dCas9) can only bind to the target DNA without cleaving it. The stable binding of dCas9 can stop the translocation of RNA-polymerase,

leading to reduced or arrested expression of the downstream gene (Bikard et al., 2013). Separate studies showed that Cas9 fused to cytidine deaminase (APOBEC1) (Komor et al., 2016) or uracil DNA glycosylase inhibitor (UGI) (Komor et al., 2017) can be used to edit a single base by modifying one base to another. Similarly, Cas12 (type V) and Cas13 (type VI) have also been used in genome editing (Cox et al., 2017; Zetsche et al., 2017).

1.3 Definition of the problem

CRISPR-Cas system is highly diverse, and it is broadly divided into two classes (1 and 2), which are further divided into six types (I–VI) and several subtypes. Type I (Class 1) system is the most abundant CRISPR-Cas system present in the bacteria, and it is further divided into seven subtypes. Befitting the enormous diversity exhibited by the type I system, the composition and architecture of the Cascade effector complex have undergone commensurate variation. In tune with this, the domain architecture of Cas3 – the effector nuclease in type I system – in some subtypes has evolved in a very distinctive way such that Nuclease and Helicase domains, which are fused in some sub-types, have undergone domain fission in others. The current understanding of CRISPR interference in type I comes majorly from the type I-E system with sporadic investigations from other type I systems such as I-A, I-F and I-G. In type I-C, the Cascade complex is composed of four subunits, viz. Cas5, Cas7, Cas8 and Cas11, as against five (Cas5, Cas6, Cas7, Cas8 and Cas11) in case of type I-E. Cas6, which is essential for crRNA maturation, is absent in type I-C, and Cas5, which is inert in type I-E, has replaced the role of Cas6. Given the architectural differences among the Cascade complex of type I-C and type I-E, it is reasonable to assume that this will be translated into functional differences. However, when the author embarked on this study, apart from type I-E, studies that systematically investigated the interference stage in type I-C were virtually non-existent. Hence, the author has attempted to understand the functional mechanism of CRISPR interference in the *B. halodurans* type I-C system and its role during CRISPR adaptation.

1.3.1 Objectives of the study

- Characterization of type I-C effector nuclease.
- Development of an *in vivo* assay system to study the interference mechanism.
- Understand the molecular events that occur during target recognition and degradation.
- Unravel the involvement of interference module in spacer acquisition.

Chapter 2: Biochemical Characterization of Type I-C Effector Nuclease

The work presented in this chapter is published

Nimkar, S., and Anand, B. (2020). Cas3/I-C mediated target DNA recognition and cleavage during CRISPR interference are independent of the composition and architecture of Cascade surveillance complex. Nucleic Acids Res. 48, 2486-2501.

2.1 Introduction

The effector nuclease Cas3 is indispensable for target DNA cleavage during CRISPR interference in type I systems. Cascade complex locates and binds to the target DNA with crRNA induced sequence specificity, forming an R-loop (Hayes et al., 2016; Jackson et al., 2014a; Mulepati et al., 2014; Xiao et al., 2017a). Favourable conformational changes in the Cascade complex recruits Cas3, which further degrades target DNA (Brouns et al., 2008; Hayes et al., 2016; Jackson et al., 2014a; Mulepati et al., 2014; Xiao et al., 2017a). Cas3 is present only in type I systems making it a 'signature' protein for the type I CRISPR-Cas system (Koonin et al., 2017a; Makarova et al., 2020). Our understanding of Cas3 is primarily attributed to the studies on *E. coli* and *T. fusca,* which belong to the type I-E CRISPR-Cas system (Gong et al., 2014; Hayes et al., 2016; Huo et al., 2014a; Ivancic-Bace et al., 2013; Jackson et al., 2014a; Majsec et al., 2016; Makarova et al., 2020; Mulepati et al., 2014; Sinkunas et al., 2011; Xiao et al., 2018; Xiao et al., 2017a). However, the arrangement and composition of genes in the CRISPR-Cas systems are diverse, and Cas3 also exhibits multiplicity in its domain architecture. In a typical type I Cas3, the N-terminal HD-nuclease domain is fused to a C-terminal DExD/H box Superfamily 2 helicase domain (Jackson et al., 2014b; Makarova et al., 2020; Sinkunas et al., 2011; Xiao et al., 2018; Xiao et al., 2017a). However, in Cas3 in type I-A, the nuclease and helicase domains are split and expressed as two standalone proteins (Makarova et al., 2020). In type I-G, Cas3 exhibits domain swapping where the helicase domain forms the N-terminus, and the nuclease domain is present at C-terminal. Additionally, in type I-F, Cas2 is C-terminally fused to Cas3 (Makarova et al., 2020). In type I-C, the molecular size of Cas3 is small; however, the domain architecture is similar to type I-E. Overall, it is reasonable to assume that these architectural differences will be reflected in the function and molecular mechanism of Cas3. Hence, to understand the interference mechanism in the type I-C system, it was essential to investigate and understand type I-C Cas3 (Cas3/I-C) in detail.

Since the information about the type I-C Cas3 was practically non-existence, we set out to thoroughly study both nuclease and helicase domains of Cas3/I-C. We tested nuclease activity on various substrates such as dsDNA, ssDNA, ssRNA and other DNA modifications. Since Cas3 is a metal-dependent nuclease, we also checked the metal ions required to activate Cas3/I-C nuclease activity. Subsequently, we also carried out experiments to assess DNA unwinding by the Cas3/I-C helicase domains. Nuclease and helicase domains are fused in

Cas3/I-C; consequently, the nuclease domain has to cooperate with the helicase domain. We tried to understand this co-operation *in vitro*. Overall in chapter 2, we have tried to characterise Cas3/I-C as a standalone protein *in vitro*.

2.2 Materials and Methods

2.2.1 Molecular cloning

B. halodurans C-125 strain was procured from The Microbial Type Culture Collection and Gene Bank (MTCC), India. Gene encoding Cas3/I-C (BH0336) and its mutants (D48A and D395A) were cloned in 1R (Addgene #29664) at the SspI restriction site by employing the Gibson Assembly method. Point mutations were created in Cas3/I-C using PCR-based mutagenesis. The 1R vector encodes Strep II tagged protein.

2.2.2 Protein purification

E. coli BL21 (DE3) harbouring pWT-1 (Cas3/I-C mentioned-above cloned in 1R) was grown in LB broth supplemented with 50 µg/ml kanamycin at 37 °C till the OD_{600} reached 0.6. At this point, Cas3/I-C expression was induced with 0.2 mM IPTG and cells were allowed to grow for 12 h at 18 °C. Cells were harvested and re-suspended in binding buffer (20 mM Tris-Cl (pH 8.0), 300 mM NaCl, 10% glycerol, 1 mM phenylmethylsulphonyl fluoride (PMSF) and 6 mM β-mercaptoethanol). Cells were lysed using a cell disruptor (20 kpsi), and cell debris was removed by centrifugation at 4 °C and 16000 × g for 30 min. After cell lysis, DNase I (50 µg/ml) was added to reduce the viscosity due to the presence of intact genomic DNA. The clear supernatant was loaded onto a pre-equilibrated 5 ml StrepTrap HP (GE Healthcare) column. After loading, the column was washed with 10 CV of binding buffer to remove unbound protein, and Cas3/I-C was eluted in a binding buffer containing 2.5 mM desthiobiotin. Cas3/I-C was further loaded onto the HiTrap Heparin HP column (GE Healthcare) to remove impurities. Subsequently, it was eluted with a linear gradient of 0.15–2 M NaCl in the binding buffer. Later, the protein was further purified through HiLoad 16/600 Superdex 200 prep grade column (GE Healthcare) and eluted in buffer containing 20 mM Tris-Cl (pH 7.5), 150 mM NaCl, 10% glycerol and 6 mM β-mercaptoethanol. After concentration, samples were flash-frozen in liquid nitrogen and stored at −80 °C until use. Protein samples were run on SDS-PAGE to check the purity, and concentration was measured by absorbance at 280 nm. Cas3/I-

C mutants cloned in 1R (pD48A, pQ253A) were purified using a similar protocol mentioned above.

2.2.3 ATP hydrolysis assay

ATP hydrolysis reactions were conducted at 37 °C in the reaction buffer containing 20 mM Tris-Cl (pH 8.0), 50 mM NaCl, 10 mM Mg^{2+}, 1 mM ATP, 6 mM β-Mercaptoethanol and 1 μM Cas3/I-C. Hydrolysis reactions were performed in the presence of 10 ng of ssDNA, dsDNA and 16S rRNA. Malachite Green assay was used to determine the liberated inorganic phosphate. Reactions were initiated by the addition of ATP and stopped by adding a malachite green reaction mixture (ammonium molybdate: malachite green in the ratio 1:4). To this mixture, 34 % sodium citrate was added to make a final concentration of 3.7 %. The volume was made up to 1 ml using water, and absorbance was measured at 630 nm. A standard curve between absorbance and inorganic phosphate concentration was established by using a varying concentration of KH_2PO_4. All the reactions were performed in triplicates.

2.2.4 Assay for nuclease activity

All nuclease activity assays were performed in cleavage buffer containing 20 mM Tris-Cl (pH 8.0), 60 mM NaCl, 1 mM dithiothreitol (DTT) and 10 mM $MgCl_2$. About 500 nM Cas3/I-C was incubated with 100 nM DNA or 50 nM RNA unless specified otherwise. To check metal ion dependency, 2 mM of various metal salts and 500 nM Cas3/I-C were used. The reactions were carried out at 37 °C for 10–60 min. Cleavage products were visualized by either 0.8% agarose gel electrophoresis or 10–15% (w/v) 8 M urea denaturing PAGE.

2.2.5 Assay for helicase activity

Partial DNA duplexes were used to test the helicase activity of Cas3/I-C and its variants. 6-FAM labelled oligonucleotides were obtained from IDT. To generate the construct with 3′ overhang, 3′ 6-FAM labelled short oligonucleotide (36 nt), and unlabelled long oligonucleotide (70 nt) were annealed. Similarly, in order to generate the construct with 5′ overhang, 5′ 6-FAM labelled short oligonucleotide (34 nt) and unlabelled long oligonucleotide (70 nt) were annealed. In addition, an excess of an unlabelled short oligonucleotide having an identical sequence as that of 6-FAM labelled was used as trap DNA in order to avoid the re-annealing

of the separated strand. Helicase activity was assessed by mixing Cas3/I-C (0.5–1 μM), and 6-FAM labelled DNA substrates (0.5 μM) in the reaction buffer containing 20 mM Tris-Cl, 60 mM NaCl, 6 mM β-mercaptoethanol, 10 mM $MgCl_2$, 50 μM Trap-DNA and 1 mM ATP (or other triphosphate variants). The reaction was carried out at 37 °C for 2 h. DNA bands were visualized on a 12% (w/v) native PAGE.

2.3 Results

2.3.1 Characterization of the Nuclease activity of Cas3/I-C

To test the substrate preference for HD nuclease, different nucleic acid substrates were utilized in the presence of ATP and Mg^{2+} ions. This showed that Cas3/I-C cleaves both double-stranded (pQE2) and single-stranded (M13mp18) DNA proficiently (Figure 2.1A).

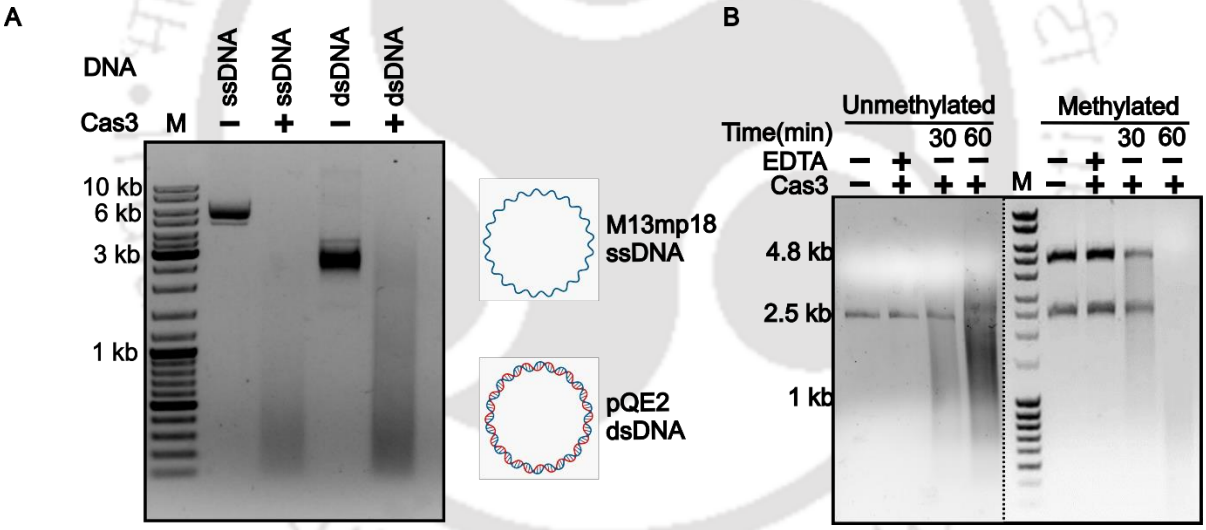

Figure 2. 1: Substrate preference for Cas3 nuclease.

(A) ssDNA and dsDNA represent single-stranded and double-stranded DNA substrates, respectively. Cas3/I-C cleaves both single (M13mp18) and double (pQE2) stranded DNA in the presence of divalent metal ion (Mg^{2+}) and ATP. DNA marker (M) positions are shown on the left.

(B) Methylated and unmethylated DNA fragments are of the same size (2.5 kb). Unmethylated DNA was generated using Pfu polymerase that generates blunt-end products. The methylated DNA fragment of 2.5 kb was released from the plasmid using restriction enzymes NdeI and KpnI. An additional band of 4.8 kb corresponds to the linearized vector pQE2. DNase activity

is not sensitive to methylation. The results were visualized using 0.8% agarose gel electrophoresis. The dotted line indicates a discontinuity in the gel for clarity, and M denotes the position of the DNA marker.

Further, nucleases such as restriction endonuclease do not cleave host DNA which is methylated. So, we wondered if Cas3/I-C displays such a trait. To test this, we used a linear plasmid (pQE2) isolated from *E. coli* as methylated substrate (*E. coli* methylates DNA) and a PCR amplified gene as non-methylated DNA. When these substrates were incubated with Cas3/I-C, we found that the nuclease activity does not differentiate between methylated (pQE2) and unmethylated (PCR amplicon) forms as substrates (Figure 2.1B).

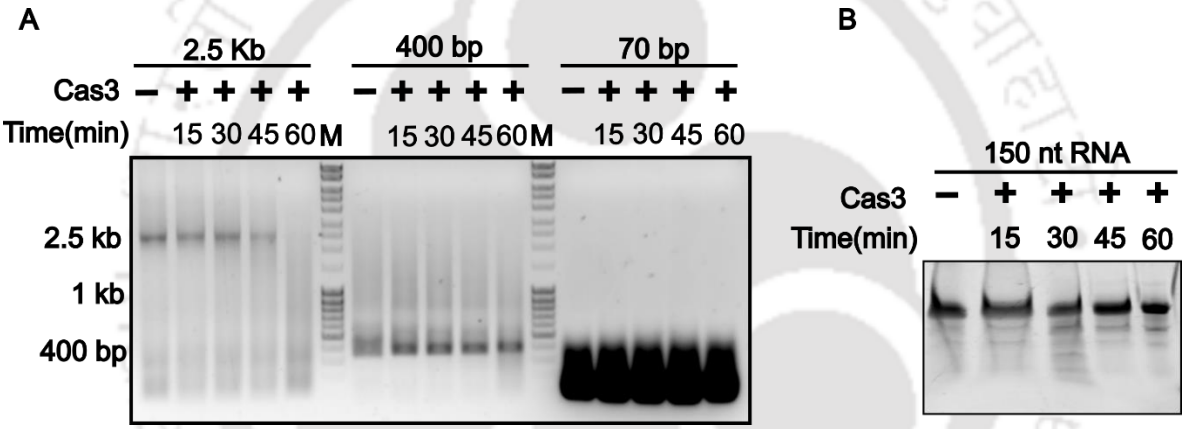

Figure 2. 2: DNase/RNase activity depends on the size of the substrate.

(**A**) DNase and (**B**) RNase activities were assessed on 0.8 % agarose and 12 % denaturing PAGE, respectively. To test the nuclease activity, 500 nM Cas3/I-C was incubated with 100 nM of 2.5 kb or 400 bp DNA. Similarly, 1.5 µM Cas3/I-C was used against 300 nM of 70 bp DNA. RNA concentration was 50 nM. Cas3/I-C preferred large DNA substrates (2.5 kb), whereas smaller DNA (400 and 70 bp) and RNA (150 nt) substrates were not completely cleaved even after 60 min of incubation. M denotes the position of the DNA marker.

Our experiments suggested that Cas3/I-C lacks intrinsic sequence specificity towards DNA. Prompted by the lack of apparent specificity for the target, we questioned whether Cas3/I-C shows any preference for the length of the target. We utilized DNA substrates of various lengths—70 bp, 400 bp and 2.5 kb. We found that while Cas3/I-C is active against 2.5

kb substrate, it is barely active against 400 bp and 70 bp substrates, respectively. In line with this, though it acts on long RNA, no nuclease activity was discernible with a small RNA fragment (150 nt). These suggest that despite the apparent preference for longer DNA, Cas3/I-C exists as a promiscuous nuclease with no apparent intrinsic specificity for the target (Figure 2.2 A-B). We also tested

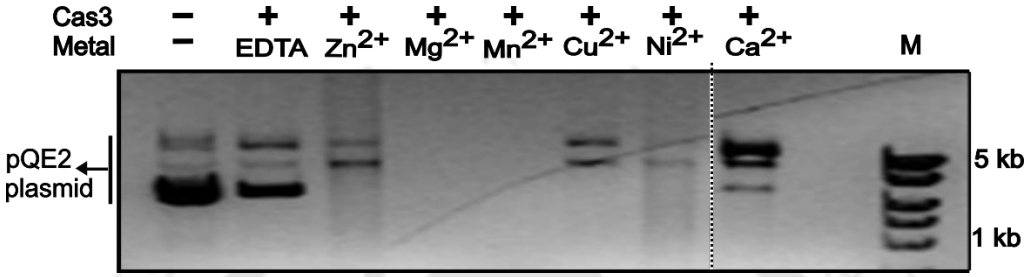

Figure 2. 3: Cas3/I-C mediated DNA cleavage is dependent on the divalent metal ion.

The addition of 500 nM Cas3/I-C in the presence of Mg^{2+} and Mn^{2+} in the reaction mixture showed prominent cleavage, whereas Ni^{2+}, Zn^{2+}, Cu^{2+}, and Ca^{2+} were less effective to a varying extent. EDTA inhibited the cleavage drastically. DNA marker (M) positions are shown on the right. The dotted line indicates a discontinuity in the gel for clarity. Note: All the samples were visualized using 0.8% agarose gel stained with ethidium bromide.

Having tested nuclease activity on DNA substrates, we also checked if Cas3/I-C can cleave RNA. To test that, we purified total RNA containing 23S, 16S and 5S rRNA from *B. halodurans* and incubated it with Cas3/I-C for several time points (0-60 mins) in the presence and absence of Mg^{2+} metal ion (Figure 2.4A). We observed that Cas3/I-C cleaves rRNA in both conditions; however, the cleavage was more efficient in the absence of Mg^{2+}. It is known that Mg^{2+} stabilizes RNA folding. Hence, in the absence of Mg^{2+}, the destabilization of the RNA structure might be responsible for prominent cleavage (Figure 2.4A). To negate the presence of metal ion already bound to Cas3/I-C, we added EDTA in the reaction and found that cleavage was inhibited. Additionally, the presence of co-purified tightly bound Fe^{2+} in the available structure of Cas3/I-E from *T. fusca* suggests that Cas3/I-C might have co-purified with a metal ion (Figure 2.4C) (Huo et al., 2014b). Mutation in the conserved aspartic acid residue (D48) leads to nuclease inactive Cas3/I-C (Figure 2.4B).

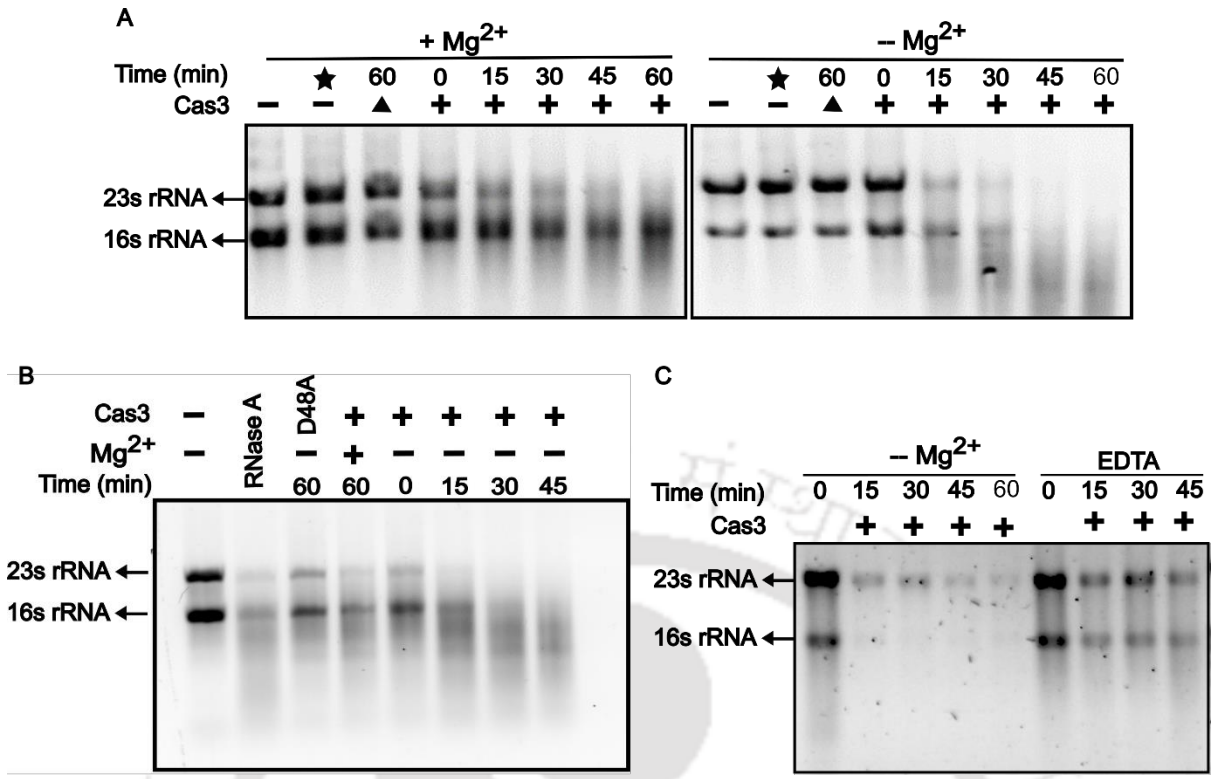

Figure 2. 4: Cas3/I-C RNase activity.

(A) 16S and 23S rRNAs isolated from *B. halodurans* were used as substrates for Cas3/I-C in the presence and absence of Mg^{2+}, and the reaction was monitored for several time points (0-60 mins). The absence of Mg^{2+} made rRNAs more susceptible to cleavage. Since Mg^{2+} is required for RNA folding, this enhanced cleavage is presumably due to the destabilization of RNA structure in the absence of Mg^{2+}. 'Triangle' denotes Cas3/I-C, which was rendered nuclease inactive after heat treatment at 80 ºC for 20 mins. The Star symbol signifies that rRNA was incubated in the reaction buffer for 60 mins without the addition of Cas3/I-C to negate the possibility of nuclease contamination.

(B) 16S and 23S rRNAs were used as a substrate for Cas3/I-C WT and Cas3/I-C nuclease mutant (D48A). We have incubated rRNA substrate for various time points. The nuclease domain mutation D48A abrogated nuclease activity. rRNA was incubated with RNase A for 60 mins.

(C) To test metal-dependent cleavage of RNA, 16S and 23S rRNA were used as substrates for Cas3/IC in the presence of 10 mM EDTA, and cleavage was monitored for several time points. The presence of EDTA inhibited rRNA cleavage, which suggests that Cas3/I-C co-purifies with a bound metal ion.

Note: All the samples were visualized using 0.8% agarose gel stained with ethidium bromide.

2.3.2 Characterization of the helicase activity of Cas3/I-C

Having characterized the HD nuclease activity, we set out to assess the DExD/H domain-mediated helicase activity using a non-specific DNA substrate. We utilized DNA containing 3′ and 5′ overhang substrates to assess the directionality of helicase activity (Figure 2.5A). We observed that Cas3/I-C unwinds only that DNA, which possesses a 3′ overhang, suggesting that it translocates in a 3′-5′ direction in the presence of ATP and Mg^{2+}. These suggest that Cas3/I-C is a generic 3′-5′ helicase with no apparent sequence specificity (Figure 2.5 B and C). Mutation in the conserved aspartic acid residue in the Walker B motif of Cas3/I-C inhibits DNA unwinding (Figure 2.5D).

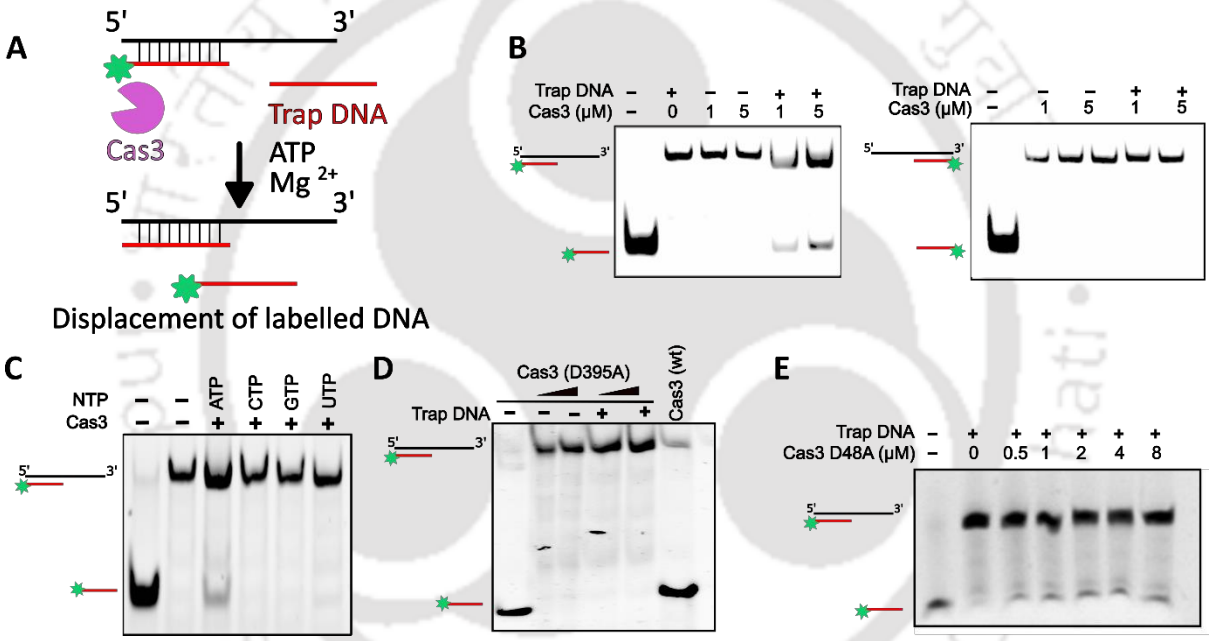

Figure 2. 5: Helicase activity of Cas3/I-C.

(A). Schematic representation of helicase unwinding assay. A partial duplex DNA was incubated with Cas3/I-C in the presence of ATP and Mg^{2+}. A trap DNA with the same sequence as 6-FAM labelled short strand was added to avoid reannealing of the displaced labelled strand. The displaced 6-FAM labelled strand can be visualized using a native PAGE.

(B). DNA unwinding was tested using a fluorescently labelled partial DNA duplex. Displacement of 36 nt was observed when a substrate containing 3′ overhang was used. About 15% native PAGE was used to analyse the results.

(C). Cas3/I-C unwinds DNA utilizing ATP as the sole energy source. No unwinding was observed in the presence of other nucleotide triphosphates (CTP, GTP, and UTP).

(D). Cas3/I-C helicase domain mutant (D395A) was unable to unwind partial duplex. We have used 250 nM and 500 nM Cas3/I-C (D395A) both with and without trap DNA. The concentration of Cas3/I-C (wt) was maintained at 500 nM.

(E). Cas3/I-C nuclease domain mutant (D48A) was able to unwind partial DNA duplex. An increasing concentration of D48A, as mentioned in the figure, was used.

2.3.3 Co-operation between Nuclease and Helicase domains of Cas3/I-C

To check if nuclease and helicase domains co-operate during nucleic acid degradation, we analysed the cleavage of the plasmid (pQE2) in the presence and absence of ATP. Since ATP is required for helicase activity, the effect of the presence of ATP can be reflected in the nuclease activity. Time-dependent cleavage of pQE2 showed that the nuclease activity is enhanced when ATP was added to the reaction. This suggests a functional overlap between the domains of Cas3/I-C (Figure 2.6A).

Additionally, we also calculated the rate of ATP hydrolysis by Cas3/I-C in the presence of several nucleic acids (variants of DNA and RNA). Cas3/I-C helicase domain shows an intrinsic ATP hydrolysis which is evident from the increase in inorganic phosphate concentration in the absence of nucleic acids (Figure 2.6B). Surprisingly, the ATP hydrolysis rate is enhanced when DNA (both single and double-stranded) was added along with metal ion. However, we could not observe any change when RNA was used as a substrate. Overall results suggest the co-operation between nuclease and helicase domains of Cas3/I-C.

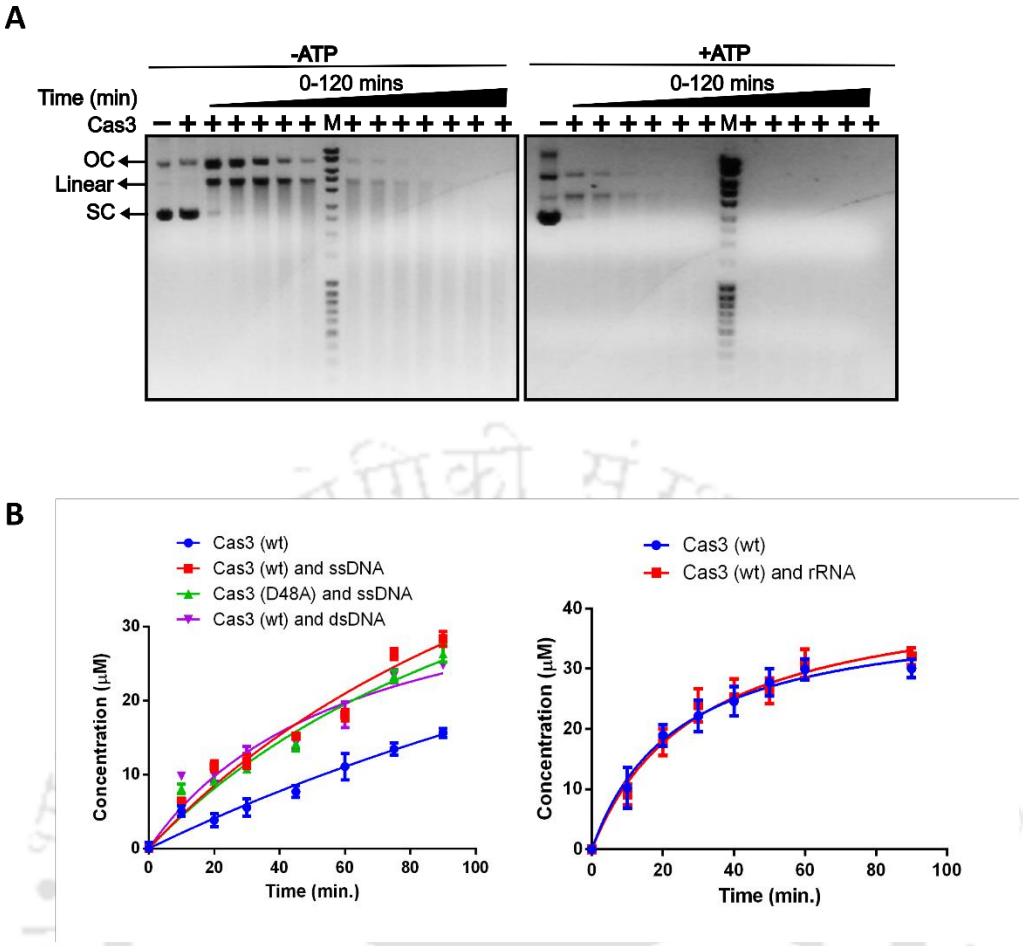

Figure 2. 6: Co-operation of nuclease and helicase domain during DNA cleavage.

(A). Time-dependent (0-120 mins) cleavage of plasmid DNA (pQE2) by Cas3/I-C shows that nuclease activity is enhanced in the presence of ATP, suggesting functional overlap between the DEXD/H helicase and HD nuclease domains. The size of the DNA markers is shown in lane M.

(B). The rate of ATP hydrolysis is enhanced in the presence of nucleic acids. Cas3/I-C (wt) is shown to hydrolyse ATP in the absence of any nucleic acid substrate. The addition of M13mp18 (ssDNA) and 70 bp partial duplex DNA with 3' overhang (dsDNA) enhanced ATP hydrolysis significantly as seen in the case of both Cas3/I-C (wt) and nuclease inactive Cas3/I-C (D48A). In comparison, the ATP hydrolysis rate remained unchanged when rRNAs (16S and 23S) were used as substrates with Cas3/I-C (wt). Error bar represents standard deviation measured from three independent trials.

2.4 Discussion

Precise recognition and cleavage of target DNA is the hallmark of the CRISPR-Cas system. In Class 2 systems, for recognition and cleavage of target DNA, a single multi-domain protein is sufficient, whereas, in type I systems, recognition and cleavage are separated between Cascade complex and Cas3, respectively. Previously published reports (Beloglazova et al., 2011; Gong et al., 2014; Huo et al., 2014a; Ivancic-Bace et al., 2013; Sinkunas et al., 2011; Sinkunas et al., 2015) and data from the current study suggest that Cas3 is a promiscuous nuclease *in vitro*. Unlike many restriction endonucleases, which specifically cleave non-methylated phage DNA and spare methylated host genome, Cas3/I-C has shown no such discrimination. Overall, the lack of sequence specificity might have a deleterious effect on cell growth. However, such an effect has not been observed *in vivo,* suggesting a separate mechanism imparting sequence specificity to Cas3, which we have addressed in chapter 4.

In Cas3/I-C, the requirement of HD-nuclease and DExD/H-helicase for DNA cleavage suggests that the concerted domain–domain crosstalk plays a pivotal role in regulating the nuclease activity. Strikingly, both HD-nuclease and DExD/H-helicase domains require single-stranded DNA substrates, which are transiently formed during highly orchestrated events such as replication and transcription. Therefore, Cas3 is unlikely to have access to its preferred single-stranded substrates most of the time, which may preclude any inadvertent self-targeting. Unlike plasmid DNA, small double-stranded DNA (<400 bp) is not cleaved, strengthening this conjecture. Due to the typical *in vitro* setting by which plasmid DNA is isolated by employing denaturation and renaturation cycles as well as given its large size, it is likely that some regions retain a single-stranded state even after renaturation. Moreover, pQE2 is found to be negatively supercoiled, which may also produce a single-stranded region. These single-stranded regions could become the substrate for Cas3/I-C mediated nicking and its subsequent loading points. Whereas shorter DNA fragment is likely to have efficient renaturation and this can pre-empt Cas3/I-C loading.

The Cas3/I-C helicase activity is similar to Cas3 from the type I system (Cas3/I-E) – requiring a 3’ overhang to load and unwind DNA in a 3’ to 5’ direction. The presence of the Q and Walker B motif in the helicase core of Cas3 allows only ATP to bind (Figure 2.7). Subsequently, ATP is hydrolysed to ADP. Later, ADP is released, leading to conformational changes in the helicase domain, and this conformational re-orientation is responsible for the unwinding of DNA.

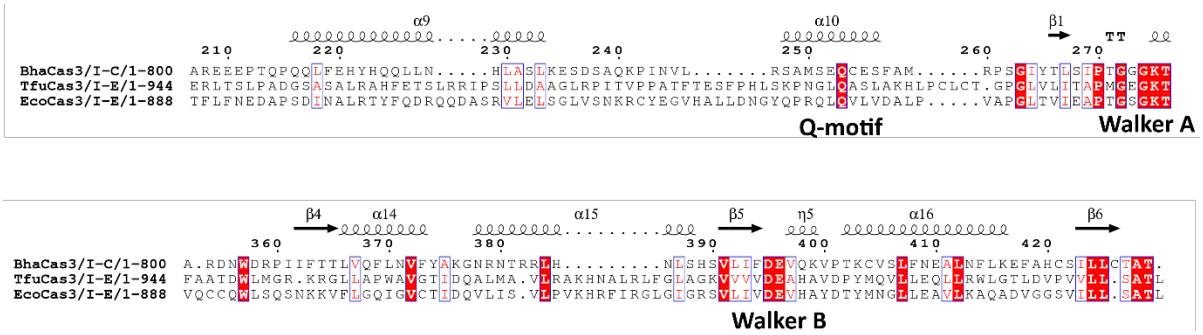

Figure 2. 7 Multiple sequence alignment of Cas3:

Multiple sequence alignment of Cas3/I-C with Cas3/I-E shows conserved motifs in the helicase domain. Q-motif is required for ATP binding, Walker A motif binds to phosphate and Walker B motif contains Mg^{2+} binding aspartic acid.

The current chapter has characterized Cas3/I-C, majorly focusing on its nuclease and helicase domains, and showed a co-operation between these domains during target cleavage. Apart from these two domains, Cas3 also possess a C-terminal domain (CTD). A CTD is usually found in helicases where they are typically involved in protein-protein interaction or contain an RNA recognition motif (RRM). However, the role of CTD has not been studied in Cas3 irrespective of CRISPR subtypes. We have tried to decipher its role in chapter 3.

2.5 Summary

In this chapter, we have accomplished the biochemical characterization of the nuclease and helicase domains of Cas3/I-C. We assessed the nuclease activity against several nucleic acid substrates (DNA and RNA). We showed that Cas3/I-C is a single-strand DNA (ssDNA) specific nuclease and that the DNA cleavage requires metal ions (Mg^{2+} and Mn^{2+}). Apart from the nuclease domain, Cas3/I-C also possesses a helicase domain. In order to check the effect of the active helicase motor on the nuclease activity, we carried out nuclease cleavage in the presence of ATP. Interestingly, with the addition of ATP, the nuclease activity was enhanced, suggesting an interplay between the two domains. Further, we observed that Cas3/I-C nuclease activity does not discriminate between methylated and unmethylated DNA. Apart from cleaving DNA, Cas3/I-C is also effective against RNA substrates. Active site mutation in the nuclease domain makes Cas3/I-C inactive against both DNA and RNA. Apart from being a nuclease, Cas3/I-C is also an active helicase whose motor activity helps in unwinding double-

stranded DNA during the interference stage. Cas3/I-C unwinds DNA in a 3' to 5' direction, requiring ATP and metal ion (Mg^{2+}). It was found that other NTPs (GTP, CTP, and TTP) could not stimulate the unwinding of DNA. Active site mutations in the helicase domain of Cas3/I-C stalls DNA unwinding. We also found that the presence of DNA stimulates ATP hydrolysis, whereas RNA showed no such apparent effect. With these observations, we characterized Cas3/I-C and proceeded further to understand the interference mechanism in chapter 3.

Chapter 3: Molecular Mechanism of Target Degradation by Effector Nuclease

The work presented in this chapter is published

Nimkar, S., and Anand, B. (2020). Cas3/I-C mediated target DNA recognition and cleavage during CRISPR interference are independent of the composition and architecture of Cascade surveillance complex. Nucleic Acids Res. 48, 2486-2501.

3.1 Introduction

Having studied the biochemical properties of Cas3/I-C *in vitro*, we sought to understand the interference mechanism in the type I-C CRISPR-Cas system in detail. Cas3/I-C works in co-operation with other Cas proteins (Hayes et al., 2016; Xiao et al., 2018), and hence in order to appreciate the mechanism of CRISPR interference in detail, it was obligatory to investigate target degradation in the presence of the Cascade complex. Results from the previous chapter suggest Cas3/I-C is a non-specific nuclease. So, it was interesting to investigate whether Cas3/I-C attains sequence specificity in the presence of the Cascade complex.

As mentioned earlier, type I systems are diverse, and Cas3 also shows variations in its domain architecture (Makarova et al., 2020). A peculiar example is type I-A, where Cas3 is split into a standalone helicase (Cas3') and a nuclease (Cas3") (Koonin et al., 2017a; Makarova and Koonin, 2015; Makarova et al., 2020). However, the functional role played by these Cas proteins is conserved and efficient (Majumdar and Terns, 2019; Plagens et al., 2012; Plagens et al., 2014). To check the effect of a domain split in type I-C CRISPR interference, we aimed to express the nuclease and helicase domains separately and measure the extent of interference. This study would deliver information about domains interplay in Cas3/I-C. On the other hand, the Cas3 also possess a C-terminal domain (CTD) (Elmore et al., 2015; Gong et al., 2014; Jackson et al., 2014b; Sinkunas et al., 2011; Sinkunas et al., 2015; van Erp et al., 2017; Xiao et al., 2018). Cas3 has been extensively studied in type I-E, and several atomic resolution structures are available (Beloglazova et al., 2011; Gong et al., 2014; Huo et al., 2014a; Mulepati and Bailey, 2011; Xiao et al., 2018). Studies have shown the interaction between the Cascade complex and Cas3 (Huo et al., 2014a; Xiao et al., 2018; Xiao et al., 2017a); however, the role of CTD was largely unknown. We set out to unravel the role of CTD in the present chapter. Alanine scanning mutagenesis of the conserved residues in Cas3/I-C CTD would reveal its functional role in interference.

Characterization of Cas3/I-C in chapter 2 suggested that it is a non-specific single-stranded nuclease. The generation of single-stranded DNA as a by-product of the R-loop provides Cas3 with a loading and target cleavage platform. Buoyed by these observations, we hypothesized that such a mechanism could be persistent among type I systems regardless of the variances in composition and architecture of Cascade. Hence, we devised a surrogate system to study the conservation of target cleavage between type I-E and I-C.

In vitro studies would provide mechanistic information about CRISPR interference; however, it would not mimic accurate physiological conditions present during target degradation. Cellular cytoplasm holds a repertoire of DNA binding proteins that can compete with the functioning of CRISPR proteins. Additionally, molecular crowding may impede the identification and cleavage of target DNA. We aimed to replicate true physiological conditions while addressing our questions, and hence *in vivo* systems were designed and utilized.

3.2 Material and Methods

3.2.1 Strain construction

For studies related to the integration of genes encoding Cascade/I-C (Cas5, Cas7, and Cas8) from *B. halodurans, E. coli* IYB5101 (Table 2) was used as a recipient strain. This strain harbours T7 RNA polymerase gene, which is under the control of araC, and hence it is inducible with arabinose. Genes encoding Cascade/I-C (Cas5, Cas7, and Cas8) were amplified as a polycistronic construct using *B. halodurans* C-125 genomic DNA as a template. This construct has a T7 promoter sequence upstream of the coding region. Subsequently, this construct was integrated into *E. coli* IYB5101 using pOSIP-CT (addgene #45981) to create the strain IC-I through one-step cloning and chromosomal integration of DNA (clonetegration), which uses phage integrases. A double knockout of *cas3* (Δ*cas3*) and *hns* (Δ*hns*) was created in *E. coli* K-12 BW25113 (Table 2) using λ Red recombineering to create strain IC-2. The list of strains employed for this study is presented in Table 2.

3.2.2 Molecular cloning

B. halodurans C-125 strain was procured from The Microbial Type Culture Collection and Gene Bank (MTCC), India. Gene encoding Cas3/I-C (BH0336) and all its mutants (pD48A, pQ253A, pD395A, pK742A, pK743A, pQ745A, pQ746A, and pY747A) were cloned in 1R (addgene #29664) at SspI restriction site by employing Gibson Assembly. Point mutations were created in Cas3/I-C using PCR-based mutagenesis. The 1R vector encodes Strep II tagged protein. Genes encoding Cascade/I-C (Cas5, Cas7, Cas8 and Cas11) were amplified as a polycistronic construct and inserted into the 1R expression vector to produce pCascade/I-C. The CRISPR array containing seven copies of identical repeat-spacer units preceded and succeeded by the T7 promoter and terminator sequences, respectively, was

commercially synthesized (GenScript) and inserted into 13S-R (addgene # 48328) to produce pCRISPR/I-C. 13S-R and 1R are compatible vectors and can be used for the purification of the Cascade/I-C complex. Target DNA sequence was inserted into the pUC19 vector using KpnI and HindIII restriction sites. All cloned constructs were verified by Sanger sequencing (Table 3).

3.2.3 Assay for nuclease activity

In order to assess the nuclease activity of Cas3/I-C in the presence of Cascade/I-C, target DNA and Cascade/I-C were incubated 1 µM of Cas3/I-C in the buffer containing 20 mM Tris-Cl (pH 8.0), 60 mM NaCl, 1 mM dithiothreitol (DTT), 10 mM $MgCl_2$ and 1 mM ATP. Proteinase K was added to release residual Cascade bound DNA fraction. After the incubation, samples containing unlabelled DNA were directly loaded onto 20% (w/v) native polyacrylamide gel and electrophoresed in 1X TBE at 4 °C. DNA bands were visualized in the gel documentation system (Bio-Rad) after staining with EtBr. 6-FAM labelled DNA samples were analysed on 20% (w/v) denaturing PAGE containing 8 M urea and directly visualized without any post-staining.

3.2.4 Assay for CRISPR interference

E. coli IC-1 harbouring pWT-1 and pCRISPR/I-C (Table 2) was grown in LB broth supplemented with 25 µg/ml kanamycin, 50 µg/ml spectinomycin, 25 µg/ml chloramphenicol, 0.2 % L-arabinose, and 0.02 mM IPTG at 37 °C until OD_{600} reached 0.3. Cells were harvested at 4 °C by centrifuging at 2700 × g. Harvested cells were made chemically competent using 0.1 M calcium chloride and transformed with pT1/pNT1 (see Table 3). Cells were grown overnight at 37 °C on LB agar supplemented with 25 µg/ml kanamycin, 50 µg/ml spectinomycin, 25 µg/ml chloramphenicol, 50 µg/ml ampicillin, 0.2% L-arabinose, and 0.1 mM IPTG. The number of colonies obtained was counted, and transformation efficiency was calculated using the equation mentioned below.

Transformation efficiency = (No. of colonies x Dilution factor) / (µg of DNA)

Transformation efficiency for Cas3/I-C variants (pHD, pHL, pCas3ΔCTD, pD48A, pQ253A, pD395A, pK742A, pK743A, pQ745A, pQ746A and pY747A; *vide* Table 3) was calculated using similar protocol.

E. coli IC-2 harbouring pWT-2 was cultured in LB broth supplemented with 50 µg/ml ampicillin and 20 µM IPTG at 37 °C. Cells were harvested when OD_{600} reached 0.3. Subsequently, harvested cells were made chemically competent, as mentioned above. These cells were transformed with pRSF-11/pST-KT (Table 3). Cells were grown overnight in LB agar with 25 µg/ml kanamycin and 50 µg/ml ampicillin. The transformation efficiencies for other variants of Cas3/I-C (pD48A, pQ253A, pD395A, pHD, pHL, pHD + pHL, pHD + pHLΔCTD) and pCas3/I-E were determined as mentioned above.

3.2.5 Homology modelling

A model of Cas3/I-C was prepared using the I-TASSER server (Zhang, 2008). Cas3/I-C (BH0336) sequence was uploaded on the I-TASSER server, and Cas3/I-C was modelled using Cas3/I-E from *T. fusca* [PDB ID: 4QQX and 6C66] as template structures. Cas3/I-C from *B. halodurans* and Cas3/I-E from *T. fusca* share 38% sequence identity. The model with the highest C-score was chosen for further analysis.

3.3 Results

3.3.1 The interplay between helicase and nuclease domains of Cas3/I-C

It is not clear whether the nuclease and helicase domains of Cas3/I-C co-operate during interference *in vivo*. To address this, we utilized *E. coli* IYB5101 as a surrogate host (IC-1, Table 2) for porting the type I-C system (CRISPR/I-C and Cascade/I-C) from *B. halodurans* (Figure 3.1A). We introduced the protospacer sequence abutting the targeting PAM sequence (TTC) into the pUC19 vector, which acted as target plasmid (T1). On the other hand, an empty pUC19 vector lacking protospacer was used as the non-target plasmid (NT1). In the absence of Cas3/I-C, the transformation efficiency of both T1 and NT1 was comparable (Figure 3.1A). When Cas3/I-C was introduced via a plasmid-borne construct, the transformation efficiency of NT1 was retained; however, that of T1 was drastically reduced (Figure 3.1A). This suggests that the type I-C system is able to discriminate between the target and the non-target substrates *in vivo*. In addition, we observed that CRISPR interference is not active in those strains under the following conditions: (i) Cas3/I-C lacks CTD; (ii) separate nuclease and helicase domains are produced; and (iii) a mutation in HD (D48A) or DExD/H (D395A) domain that inhibits nuclease and helicase activity, respectively (Figure 3.1B).

However, co-expression of standalone HD and HL domains results in the functional assembly of the interference complex, as evidenced by the reduction in transformation efficiency (Figure 3.1B). This suggests that the interplay between these domains is crucial for CRISPR interference in the type I-C system.

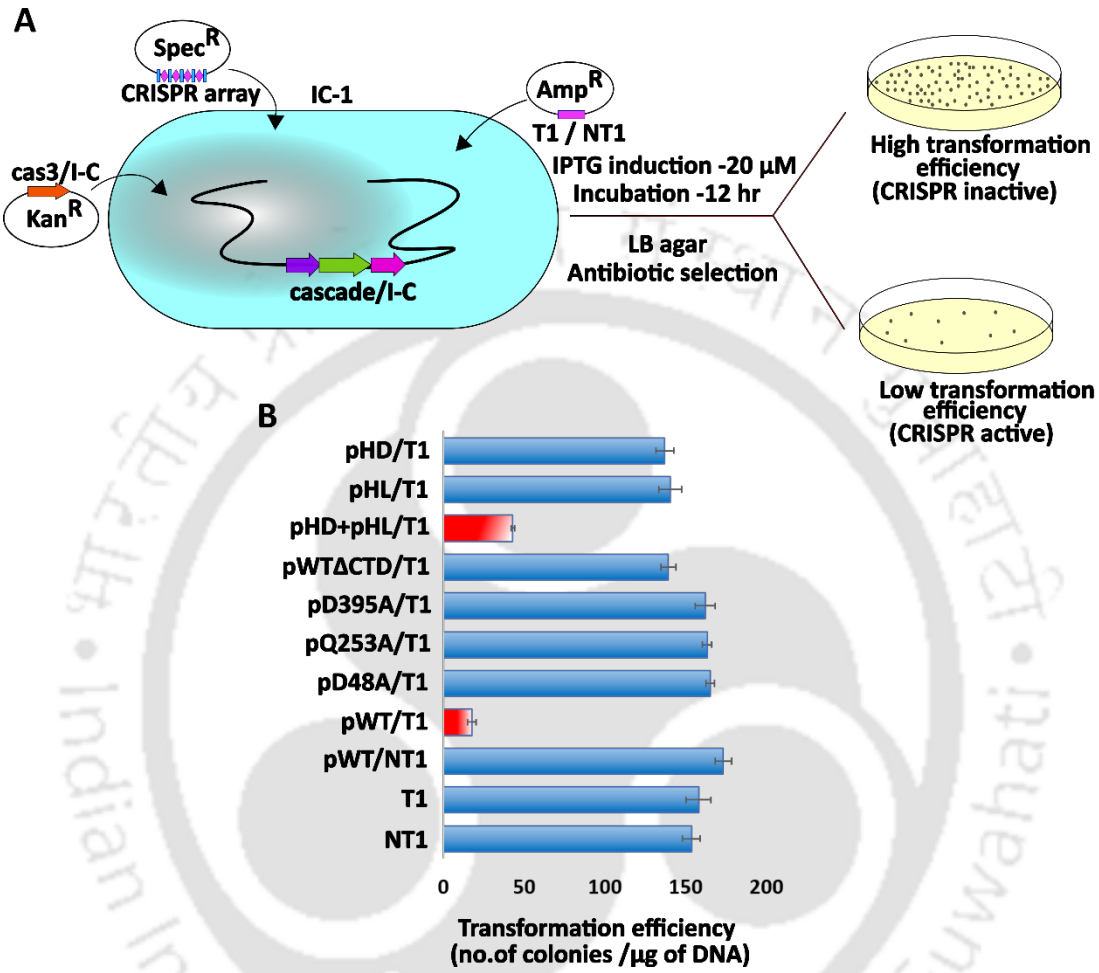

Figure 3. 1: Inter-domain interaction in Cas3/I-C influences CRISPR interference.

(A). An *E. coli* IYB5101 strain harbouring genes encoding Cascade/I-C (IC-1) was used as a surrogate for *in vivo* experiments. Cas3/I-C and crRNA were expressed through IPTG inducible vectors having kanamycin and spectinomycin antibiotic resistance markers, respectively. The target sequence was inserted into the pUC19 vector (T1), whereas an empty pUC19 vector was used as a non-target (NT1). Low transformation efficiency indicates functional CRISPR interference.

(B). Transformation efficiencies of the strain IC-1 against the target (T1) and the non-target (NT1) for wild-type and variants of Cas3/I-C are shown. In the presence of Cas3/I-C wild-type (pWT) and co-expressed construct of HD and DExD/H domain (pHD+pHL), there was a significant reduction in transformation efficiency (indicated by a red bar). Interference was rendered ineffective when the HD domain mutant (pD48A), Helicase domain mutant (pQ253A and pD395A), CTD deletion mutant (pWTΔCTD), standalone nuclease domain (pHD) and

standalone helicase domain (pHL) were used. Error bar represents standard deviation measured from three independent trials. The boundary of the domain variants are as follows: HD domain (pHD) 1–248; DExD/H domain (pHL) 249–800; WT without CTD (pWTΔCTD) 1–709.

3.3.2 A constellation of highly conserved residues in the C-terminal domain of Cas3/I-C regulates target cleavage

Apart from the two major functional domains—nuclease and helicase—Cas3/I-C has a CTD, whose function is poorly defined (Figure 3.2A). Intrigued by the fact that CTD is crucial for CRISPR interference (Figure 3.1B), we set out to probe its role in more detail. Previously determined crystal structure of Cas3/I-E with target DNA [PDB ID: 4QQX] shows that CTD caps the two juxtaposed RecA domains (that are part of the DExD/H domain) such that the target DNA is sandwiched between CTD and RecA domains (Figure 3.2A). Sequence alignment of Cas3/I-C orthologs showed high conservation of K742, Q745 and Y747, and these conserved residues were absent in Cas3/I-E, suggesting a functional significance for these residues in type I-C (Figure 3.2B and 3.3A). Since the structure of Cas3/I-C is not determined, we employed the Cas3/I-E structure [PDB ID: 4QQX] as a template and made a homology model for Cas3/I-C, which suggested that K742, K743, Q745, Q746, and Y747 that are part of CTD could stabilize the interaction of the target DNA with the helicase core (Figure 3.2A). Based on this, we hypothesized that CTD might be involved in stabilizing the interaction with target DNA so that Cas3/I-C does not dissociate from DNA during its translocation. We performed alanine-scanning mutagenesis of the aforementioned residues and assessed for the nuclease activity to test this hypothesis. Toward this, we designed a 100 bp target substrate harbouring a TTC PAM and performed the assay in the presence of Cascade/I-C and Cas3/I-C (Figure 3.2D and 3.2E). As compared to the wild-type, mutants harbouring K743A, Q745A, and Y747A were found to be inactive (Figure 3.2E).

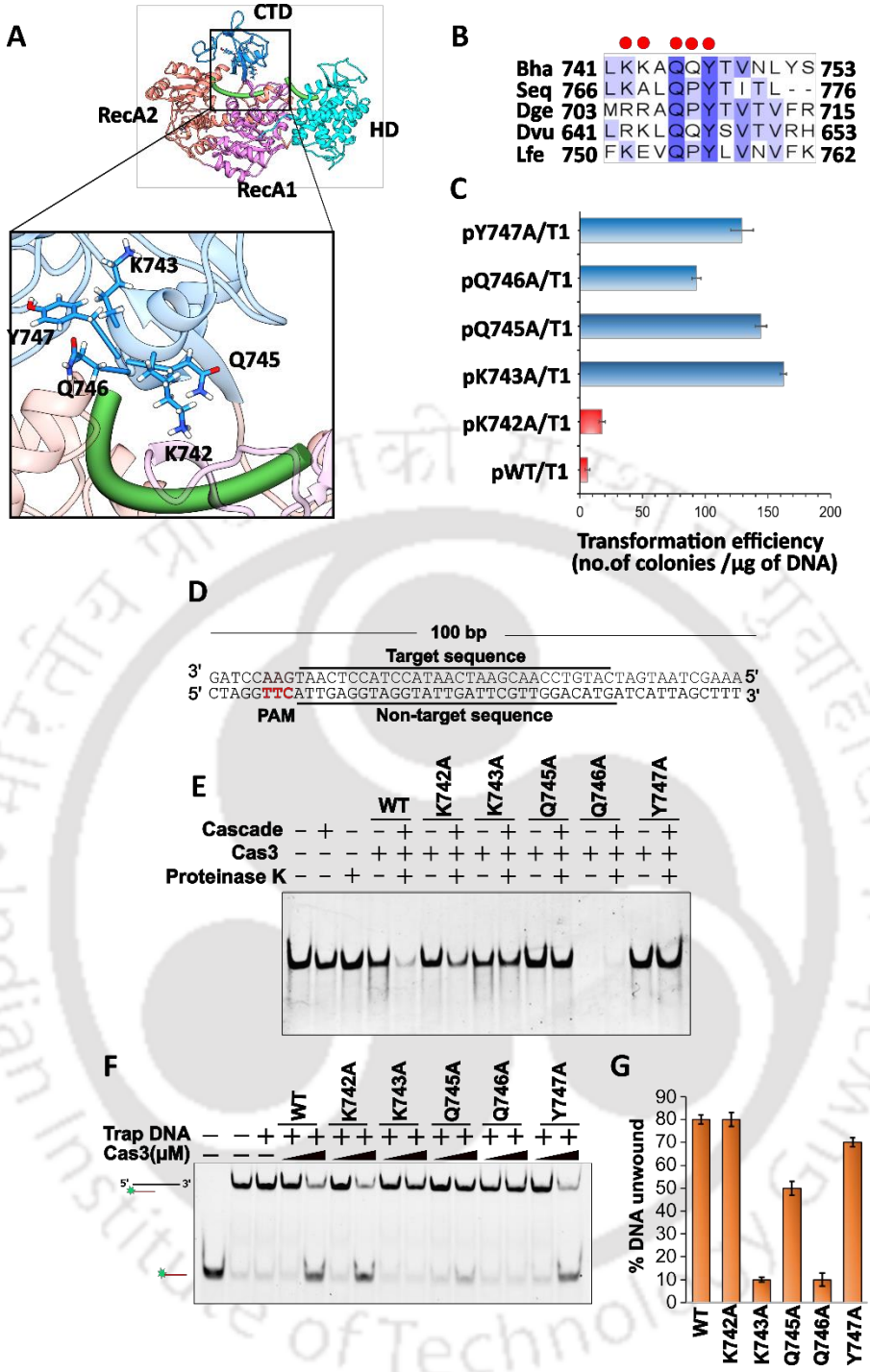

Figure 3. 2: Deciphering the role of Cas3/I-C CTD.

(A). Homology model of Cas3/I-C from *B. halodurans* based on the crystal structure of Cas3/I-E [PDB ID: 4QQX] shows the organization of nuclease (HD in cyan), DExD/H helicase (RecA1 in pink and RecA2 in orange) and CTD (in blue) domains. DNA is shown as a rod in green. Amino acids (K742, K743, Q745, Q746 and Y747) from the CTD domain that are possibly interacting with target DNA are shown.

(B). Highly conserved residues in the CTD domain from the type I-C system are shown. Red dots indicate those residues from *B. halodurans* that are chosen for alanine-scanning mutagenesis.

(C). CTD mutants were tested for their role in CRISPR interference using the *E. coli* IC-1 strain as the surrogate host. Wild-type and CTD mutants were expressed through an IPTG inducible 1R vector (referred to as pWT, pK742A and so on). Transformation efficiency was measured against pUC19 harbouring the target sequence with functional PAM TTC (referred to as pT1). Error bar represents standard deviation measured from three independent trials.

(D). A 100 bp target DNA harbouring functional PAM sequence TTC is depicted. Target sequence represents the strand, which is complementary to crRNA.

(E). CTD point mutants were tested for their nuclease activity. Cascade/I-C was incubated with target DNA in the reaction mixture containing Mg^{2+} and ATP to form R-loop. Wild-type or CTD mutant was added to the reaction, and the cleavage was monitored. Proteinase K was added to release residual Cascade bound DNA fraction. DNA was analysed using 20% native PAGE.

(F). Fluorescently labelled partial duplex DNA (with a 3' overhang) was used to test the helicase activity of Cas3/I-C CTD mutants. 500 nM of partial duplex DNA was incubated with 50 μM of trap DNA and 1 μM of Cas3/I-C and CTD mutants. ATP was added to 1 mM concentration. Trap DNA was provided in excess to avoid re-annealing of the displaced strand.

(G). The percentage of the DNA duplex unwound in the presence of Cas3/I-C and CTD mutants in 15 mins, as observed in (F), is shown. 100% unwinding signifies complete unwinding of DNA duplex, whereas 0% signifies no unwinding. DNA band intensities from (E) were quantified using Image Lab software (Bio-Rad). Error bar represents standard deviation calculated from two independent trials.

The effect of these CTD mutations on CRISPR interference *in vivo* was tested in *E. coli* IC-1. This strain was transformed with pCas3/I-C harbouring the desired mutation in the CTD, and the transformation efficiency was measured against pUC19 harbouring the target sequence. We anticipated that the active Cas3/I-C would target the pUC19 resulting in reduced transformation efficiency, whereas the inactive Cas3/I-C would result in target evasion leading to high transformation efficiency. We observed that K743A, Q745A, Q746A and Y747A in CTD produced high transformants compared to WT, suggesting that these mutations indeed render the CRISPR interference inactive (Figure 3.2C). Intriguingly, though Q746A showed nuclease activity in the absence and presence of Cascade/I-C *in vitro*, there was no appreciable CRISPR interference *in vivo* (Figure 3.2C).

Figure 3. 3: Comparison of Cas3/I-E and Cas3/I-C.

(A). Multiple sequence alignment corresponding to CTD of Cas3/I-C from select organisms is shown. The sequences were aligned using MUSCLE. The motif "KKAQQY" that shows high conservation among the orthologs is shown within the black box. The location of these residues in the homology model of Cas3/I-C suggests a possible role to stabilize the target DNA channel. Residues in the alignment are highlighted based on the extent of conservation using Jalview

(B). Cas3 sequences from type I-E systems were aligned using MUSCLE. In the type I-E system, the motif "KKAQQY" is absent, suggesting some differences in the functionality of CTD between type I-C and I-E.

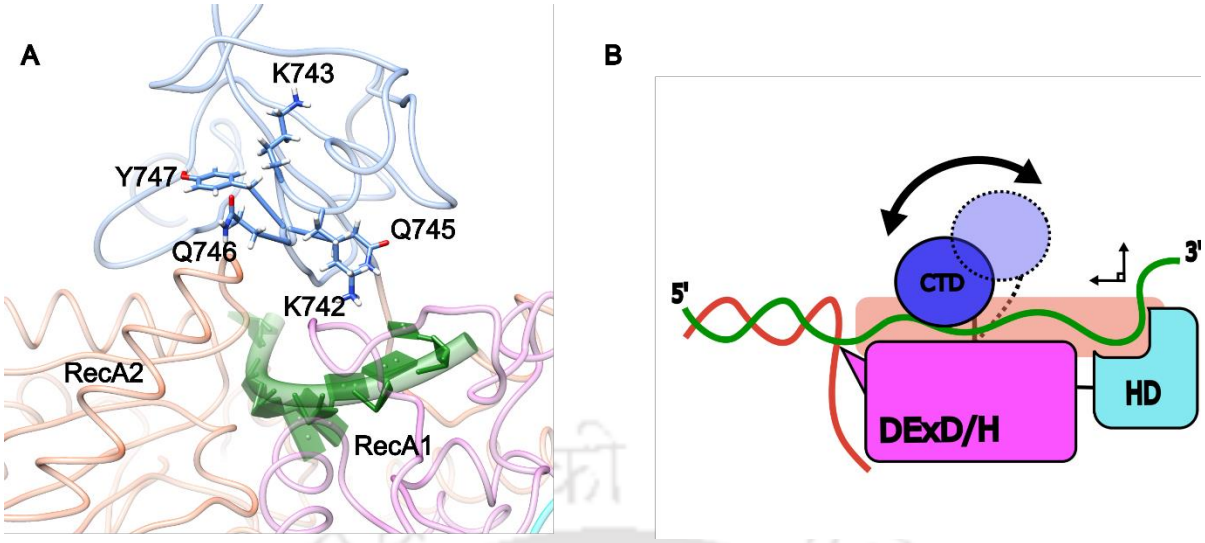

Figure 3. 4: Homology model of Cas3/I-C CTD based on Cas3/I-E structure

(A). Cas3/I-C was modelled using Cas3/I-E [PDB ID: 4QQX] as a template using I-TASSER. Residues selected for mutations in this study are indicated in the figure. The position of these mutations suggests their role in locking CTD after Cas3/I-C is bound to target DNA during interference.

(B). The modelled structure of Cas3/I-C, as well as available Cas3/I-E structures, suggest that the uncharacterized CTD rests on top of the DExD/H domain, providing a channel for ssDNA to enter, thereby stabilizing the interaction between DNA and Cas3/I-C. CTD (shown in blue) is linked to the helicase domain with the help of a long flexible linker. Since mutations in the CTD domain impede interference mechanism and CTD undergoes conformational changes (2), we speculate that the linker region allows CTD to move towards and away from the helicase domain (indicated by a double-headed arrow). This is likely to facilitate the bolting of Cas3/I-C onto the non-target strand. Upon recruitment, Cas3/I-C HD domain nicks the displaced non-target strand and subsequently, the ssDNA strand is fed continuously to the nuclease domain via a directed channel (indicated by a pale brown cylinder) formed at the interface between helicase and CTD domains.

3.3.3 Evolutionary conservation of target recognition and cleavage mechanism between type I-C and type I-E

Buoyed by the fact that it is the generation of single-stranded DNA as a by-product of the R-loop that provides a platform for Cas3/I-C loading and target cleavage, we hypothesized that such mechanism could be pervasive among type I system irrespective of the differences in composition and architecture of Cascade. To test this, we chose Cascade/I-E, which comprises five Cas subunits—as against Cascade/I-C, which comprises three subunits only. Despite these differences, both Cascades facilitate the formation of R-loop, and therefore, we anticipated that in such a scenario, it is possible to complement the Cascade/I-E with Cas3/I-C in order to promote the target cleavage *in vivo*. To test this, we created an *E. coli* strain (IC-2, Table 2) that harbours a spacer that targets pRSF-11 (T2, Table 3) and Cascade/I-E. In IC-2, the innate Cas3/I-E was deleted, and therefore the intrinsic CRISPR/I-E machinery can recognize the target, but it is deficient in target cleavage (Figure 3.5A). When IC-2 was challenged with T2, more transformants were observed (Figure 3.5B), suggesting that target cleavage is impaired in the absence of Cas3/I-E. When strain IC-2 was complemented with plasmid-borne Cas3/I-E (pCas3/I-E), there was a drastic reduction in transformation efficiency, which suggests the restoration of functional CRISPR interference.

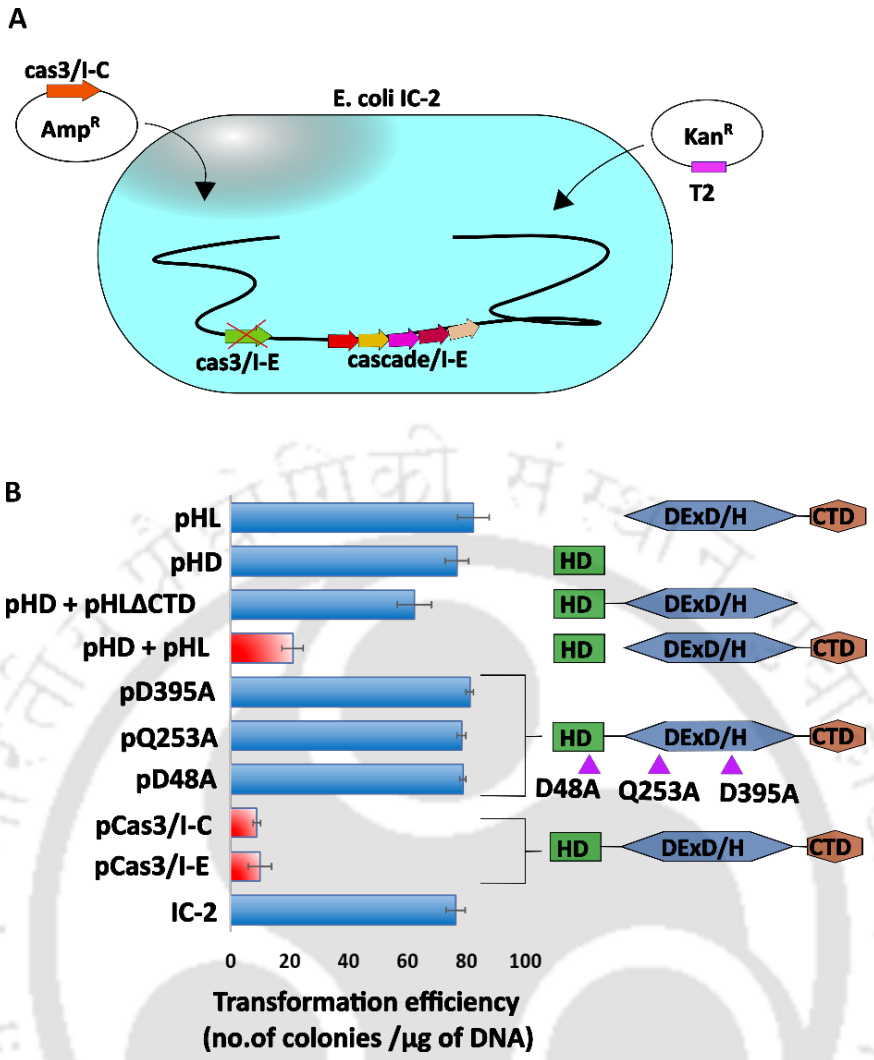

Figure 3. 5: Cas3/I-C supplants Cas3/I-E in type I-E CRISPR-Cas system.

(A). *E. coli* K-12 harbouring Cas3/I-E null mutant (IC-2) was used. IC-2 harbours spacer sequence targeting plasmid pRSF-11 (pT2). Cas3/I-C was heterologously expressed through IPTG inducible vector having an ampicillin selection marker. Transformation efficiency was calculated after induction with 20 µM IPTG and 12 h incubation.

(B). Cas3/I-C and its variants were tested for interference activity using *E. coli* IC-2. Heterologous expression of Cas3/I-C (pCas3/I-C) and Cas3/I-E (pCas3/I-E) showed reduced transformation efficiency (indicated as a red bar). When the HD (pHD) and helicase (pHL) domains of Cas3/I-C were co-expressed as two separate constructs, the transformation efficiency mirrored that of wild-type (indicated as a red bar). However, HD nuclease mutant (pD48A), helicase mutants (pQ253A and pD395A) and co-expressed constructs of HD and helicase without CTD (pHD + pHLΔCTD), standalone nuclease domain (pHD) and standalone helicase domain (pHL) showed high transformation efficiency. Error bar represents standard deviation measured from three independent trials. Purple triangles represent the location of the point mutations. The boundary of the domain variants are as follows: HD domain (pHD) 1-248; DExD/H domain (pHL) 249-800; pHD+pHLΔCTD 1–709.

Interestingly, a similar phenotype was noted when plasmid-borne Cas3/I-C (pCas3/I-C) complemented IC-2, suggesting that T2 is efficiently targeted by Cas3/I-C (Figure 3.5B). To probe this further, the HD (pHD) and DExD/H (pHL) domains of Cas3/I-C were co-expressed as independent domains for complementation *in vivo*. Remarkably, this strain showed less transformation efficiency that is comparable to Cas3/I-C WT, suggesting that these domains interact and produce functional Cas3/I-C *in vivo* during CRISPR interference. It is to be noted that HD and DExD/H domains, when expressed separately, failed to restore CRISPR interference, and it was only when they were co-expressed, T2 was targeted (Figure 3.5B). In line with this, a pCas3/I-C harbouring point mutation in lieu of catalytic residue in HD (pD48A) or Helicase domain (pQ253A and pD395A) is found to be inactive (Figure 3.5B). This suggests that despite the difference in the architecture of Cascade/I-C and Cascade/I-E, the Cas3-mediated target cleavage mechanism seems to be conserved between type I-C and type I-E.

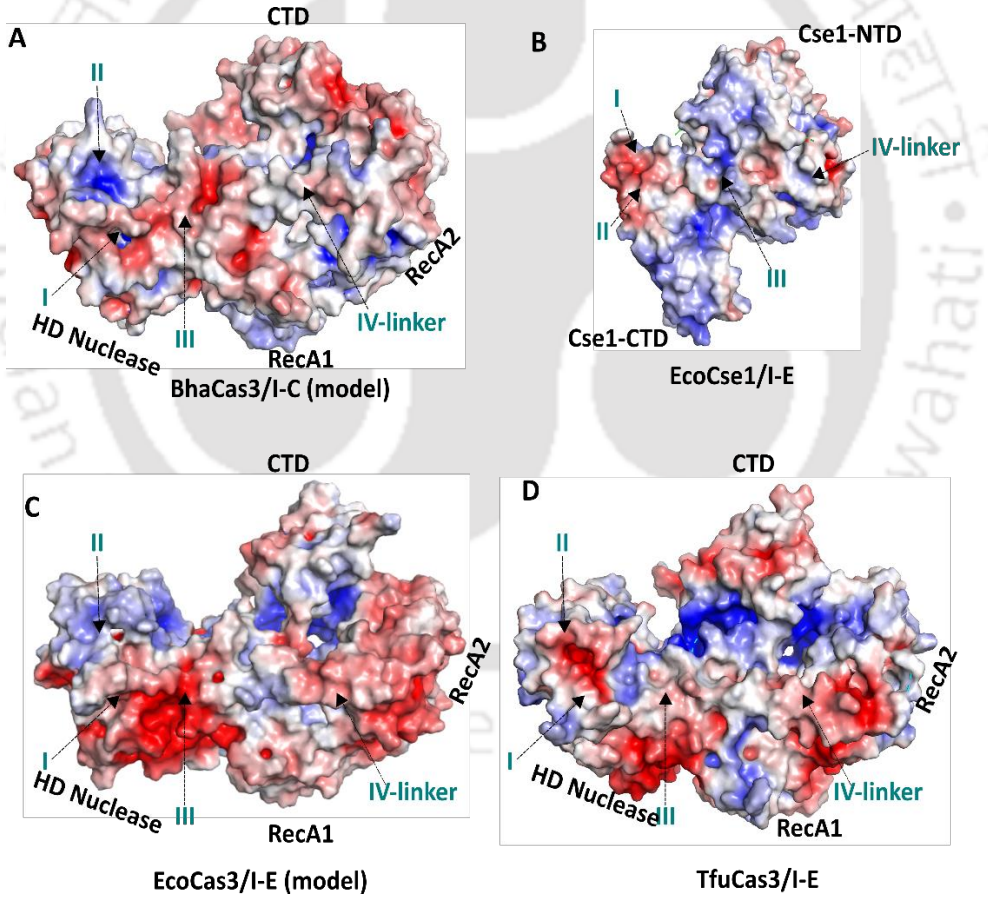

Figure 3. 6: Comparison of Cascade interacting interface between Cas3/I-E and Cas3/I-C.

Four interfaces on BhaCas3/I-C (A) and EcoCas3/I-E (C), which were modelled using Cas3/I-E (PDB ID: 6C66) as a template using I-TASSER, are shown. Blue represents a surface with basic residues, and red represents a surface with acidic residues. These interfaces are identified

as contact points between TfuCas3/I-E (D) and Cse1 subunit of TfuCascade/I-E. The surface electrostatics of EcoCse1/I-E (PDB ID: 5H9F) (B) is shown along with expected Cas3 contact interfaces. Surface charges on BhaCas3/I-C (A) and EcoCas3/I-E (C) shows similarities at interfaces I, II and III. Unlike TfuCas3/I-E (D) and EcoCse1/I-E (C), the charge compatibility between BhaCas3/I-C (A) and EcoCse1/I-E (C) (mainly interfaces II and III) explains why there is a hetero-assembly of BhaCas3/I-C and EcoCascade/I-E *in vivo* in our study

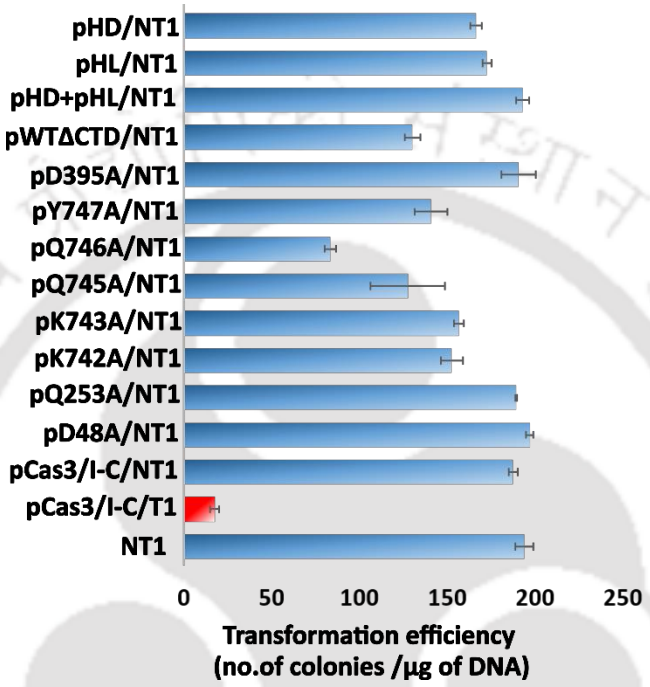

Figure 3. 7: CRISPR interference against non-targeting plasmid in *E. coli* IC-1.

E. coli IC-1 was used as a surrogate for Cas3/I-C interference assay. The target sequence was inserted in the pUC19 vector (T1), whereas an empty pUC19 vector was used as a non-target (NT1). Cas3/I-C and crRNA were expressed through compatible IPTG inducible vectors. High transformation efficiency indicates non-functional CRISPR interference. In the presence of Cas3/IC (pCas3/I-C), T1 (shown as a red bar) showed a reduction in transformation efficiency, whereas NT1 showed higher transformation efficiency. Similarly, Cas3/I-C mutants (pD48A, pQ253A, pK742A, pK743A, pQ745A, pQ746A, pY747A, pD395A, pWTΔCTD, pHD+pHL, pHL and pHD) were unable to target NT1.

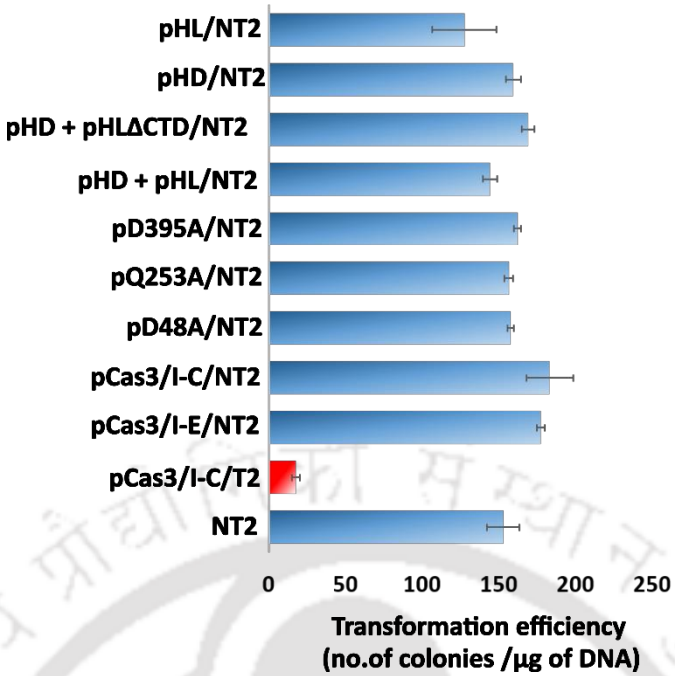

Figure 3. 8: CRISPR interference against non-targeting plasmid in *E. coli* IC-2

E. coli IC-2 harbouring a spacer that targets pRSF-11 plasmid (T2) was used for *in vivo* studies. Plasmid pST-KT, lacking spacer sequence, was used as a non-target plasmid (NT2). Cas3/I-C and its variants were expressed using IPTG inducible pQE2 plasmid, and subsequently, transformation efficiency was estimated. Cas3/I-C (pCas3/I-C), Cas3/I-E (pCas3/I-E), HD nuclease mutant (pD48A), helicase mutant (pQ253A, and pD395A), nuclease-helicase split (pHD + pHL), nuclease-13helicase split without CTD (pHD + pHLΔCTD), nuclease domain (pHD) and helicase domain (pHL)were all found to be inactive against NT2. However, as anticipated, Cas3/I-C (pCas3/I-C) was active against T2 (shown as a red bar).

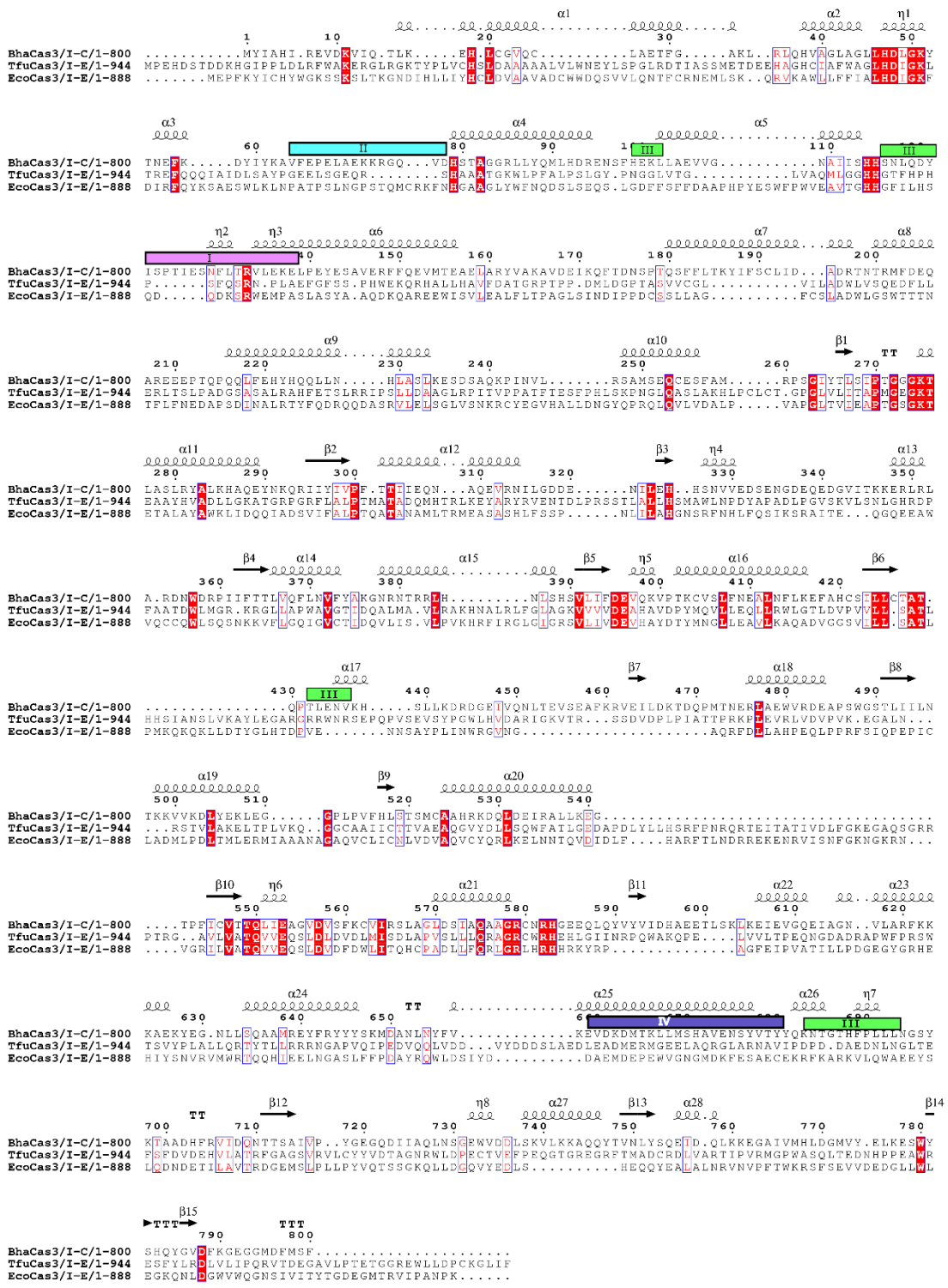

Figure 3. 9: Sequence conservation among Cas3 from type I-C and type I-E

Multiple sequence alignment of BhaCas3/I-C from *B. halodurans* with Cas3 from type I-E shows conservation in the functional motifs. Amino acids showing complete conservation are boxed in red, whereas partial conservation is boxed in blue. Cas3-Cascade interfaces (I-IV) are shown in coloured bars above the alignment. The secondary structural assignment is based on the homology model of BhaCas3/I-C.

3.4 Discussion

In this chapter, we have designed an *E. coli* surrogate system and studied CRISPR interference *in vivo*. Taken together, results from chapter 2 (Figure 2.1) and chapter 3 (Figure 3.1B and 3.7) suggest that the functional helicase and nuclease domains are crucial for CRISPR interference. This is evident from the abrogation of interference activity when active site mutants of nuclease and helicase domains were transformed in the *E. coli* IC-1 strain (Figure 3.1B). As shown in Chapter 1, type I CRISPR systems are diverse and further divided into seven subtypes. Out of these seven subtypes, type I-A CRISPR systems are frequently found in archaea and occurring sporadically in bacteria (Makarova et al., 2011b; Makarova et al., 2015). Cas3/I-A is split and naturally expressed as a standalone helicase (Cas3') and nuclease (Cas3") domains (Haft et al., 2005; Makarova et al., 2011b; Makarova et al., 2015). Moreover, Cas3'/I-A and Cas3"/I-A forms interference complex in type I-A system (Majumdar et al., 2015; Plagens et al., 2012; Plagens et al., 2014), whereas Cas3 in other subtypes is recruited after the Cascade complex has identified and bound to the target DNA (Gong et al., 2014; Hayes et al., 2016; Hochstrasser et al., 2014; Rutkauskas et al., 2015; van Erp et al., 2017; Wiedenheft et al., 2011; Xiao et al., 2018; Xiao et al., 2017a). Interestingly, like type I-A the CRISPR interference in the type I-C system was still active when the helicase (including CTD) and the nuclease domains of Cas3/I-C were deliberately expressed separately (Figure 3.1B). This suggests that the crosstalk between the nuclease and helicase domain is crucial for CRISPR interference.

Apart from two major domains, Cas3 also contains a C-terminal domain (CTD) whose function was poorly defined. Usually, helicase CTDs are involved in protein-protein binding or facilitate the recognition of specific DNA sequences (Buttner et al., 2007; Karow and Klostermeier, 2010; Luo et al., 2008). However, the information about the role of Cas3 CTD during CRISPR interference was limited. Typically, Cas3/I-E is expressed as a standalone protein; however, sequence analysis of the Cas3 from a few type I-E systems (For example, *Streptomyces sp*. SPB78 (Accession Number: ZP_07272643.1), *Streptomyces griseus* (Accession Number: YP_001825054), and in *Catenulispora acidiphila* DSM 44928 (Accession Number: YP_003114638)) shows that Cas3 is C-terminally fused to Cas8 (Hochstrasser et al., 2014; Westra et al., 2012b). The presence of these fusion proteins suggests Cas3 interacts with the Cas8 subunit of Cascade during target cleavage (Hochstrasser et al., 2014). Additionally, when the C-terminal tail of Cas3/I-E was synthetically fused to the N-

terminus of Cas8, the resulting interference complex was active and cleaved target DNA suggesting the role of CTD in interaction with the Cascade complex (Westra et al., 2012b). However, such interaction was not observed when individually purified Cas3 and Cas8 were mixed and analysed using size-exclusion chromatography (Gong et al., 2014). In another study, deletion of CTD reduces the affinity between Cas3 and Cascade complex; however, affinity was restored to wild-type level when CTD was expressed *in trans* (Huo et al., 2014a). In the present thesis, we analysed and compared the sequences in substantially studied Cas3/I-E and poorly-studied Cas3/I-C, and we could not detect any conservation in the CTD region. However, when only Cas3/I-C sequences were aligned, we could observe some conserved amino acids (K742, K743, Q745, Q746, and Y747; Figure 3.3A-B). Modelled structure of Cas3/I-C and available Cas3 structures show that CTD sits on top of two juxtaposed RecA domains and may provide a ssDNA channel, thereby stabilizing DNA-Cas3 interaction (Figure 3.4). The impairment in target cleavage activity when CTD residues are mutated is likely due to the opening of the ssDNA channel leading to destabilization of protein-DNA interaction (Figure 3.4).

The fact that Cas3/I-C—albeit exhibiting differences such as the presence of highly conserved residues in CTD—shows functional similarities with respect to Cas3/I-E has raised a question of how this functional equivalence could exist in harmony with Cascade/I-C—which exhibits compositional and architectural differences with Cascade/I-E. Strikingly, the complementation of Cas3/I-C in lieu of Cas3/I-E in the type I-E system *in vivo* shows that the target recognition and cleavage mechanism are conserved between type I-C and I-E (Figure 3.5). When we compared the homology model of Cas3/I-C from *B. halodurans* (BhaCas3/I-C) (Figure 3.2A and 3.6) and the homology model of Cas3/I-E from *E. coli* (EcoCas3/I-E)—both are modelled on the basis of the recent cryo-EM structure of Cas3/I-E from *T. fusca* (TfuCas3/I-E)—we could discern that BhaCas3/I-C shows a significant sequence and structural similarities with EcoCas3/I-E, if not complete, in interfaces I–IV (Figure 3.6 and 3.9). Based on this analysis, we attribute the reported incompatibility for hetero-assembly between TfuCas3/I-E and EcoCascade/I-E (Xiao et al., 2018) to the following factors: (i) the reconstitution of hetero-assembly was attempted *in vitro*. Therefore, it is quite likely that the lack of a cellular milieu could be a possible deterrent for hetero-assembly. (ii) TfuCas3/I-E and EcoCascade/I-E are not thermo-compatible. The former is thermophilic, and the latter is mesophilic. They have undergone different adaptations attuned to the thermal conditions. (iii) In our study, types I-C and I-E are not coexisting as composite CRISPR-Cas systems within the same organism. This

precludes any selection pressure to avert cross-over Cascade–Cas3 interaction between type I-C and I-E. Thus, our data suggest that despite the compositional and architectural differences seen between Cascade/I-C and Cascade/I-E, Cas3 recruitment and target cleavage mechanism seem to be conserved, at least in I-C and I-E systems, if not all type I.

3.5 Summary

In this chapter, we have characterized Cas3/I-C *in vivo* in the presence of Cascade complex and legitimate target plasmid DNA. It was observed that helicase and nuclease domains of Cas3/I-C co-operate during target cleavage and mutations in either of them renders Cas3/I-C inactive. Like type I-A, splitting of Cas3/I-C domains lead to functional interference complex. We have shown that an accessory CTD is crucial for interference. Unlike Cas3/I-E, CTD in Cas3/I-C shown sequence conservation, and mutation in these conserved residues leads to defective interference. Further, we have shown that the target identification and cleavage mechanism remain conserved across type I CRISPR-Cas subtypes irrespective of Cascade composition and architecture.

Chapter 4: Functional Insights into the Interference Mechanism

The work presented in this chapter is published

Nimkar, S., and Anand, B. (2020). Cas3/I-C mediated target DNA recognition and cleavage during CRISPR interference are independent of the composition and architecture of Cascade surveillance complex. Nucleic Acids Res. 48, 2486-2501.

4.1 Introduction

Having addressed a few interference related questions, we further expanded our understanding of interference by studying the molecular mechanism of Cas3/I-C mediated target degradation. We tried to unravel the molecular events from Cas3/I-C loading to the final step, where the target DNA is cleaved. After the Cascade complex locates the target DNA, the spatial rearrangements at the binding site lead to the formation of a DNA:RNA hybrid called R-loop (Hayes et al., 2016; Pausch et al., 2017; van Erp et al., 2017; Xiao et al., 2018; Xiao et al., 2017a). Formation of R-loop is a pre-requisite for Cas3 to load and cleave the bonafide target DNA. However, *in vitro* Cas3/I-C data presented in Chapter 2 and published information on Cas3/I-E (Sinkunas et al., 2011; Sinkunas et al., 2015) suggest that the Cas3 lacks sequence specificity. Additionally, as shown in chapter 3, when Cas3/I-C was expressed along with Cascade/I-E, the interference complex was active and efficiently degraded the target DNA. Based on these outcomes, we hypothesized that the ssDNA region created as a result of R-loop formation might act as a loading point for Cas3/I-C. Once Cas3/I-C is loaded onto the ssDNA region, target cleavage is activated through cooperation between the helicase core and the nuclease domain.

Next, we set out to understand the effect of Cas3 and Cascade interaction during interference (Hayes et al., 2016; Rutkauskas et al., 2015; Xiao et al., 2018; Xiao et al., 2017a). In type I-E, Cas3 interacts with Cascade through a combination of steric hindrance and electrostatic attraction (Xiao et al., 2018). After Cas3 is recruited, two possible mechanisms can be followed for target degradation; (1) Translocation – where Cas3, while unwinding duplex DNA, leaves Cascade complex and move towards 5' end cleaving DNA in the process; (2) Reeling – where Cas3 remain bound to Cascade and unwinds DNA by reeling/pulling (Loeff et al., 2018). In the current chapter, using fluorescence-based experiments, we have tried to understand which of these two mechanisms is followed by the type I-C interference complex.

On the other hand, while unwinding target DNA, Cas3 may encounter several 'roadblocks' in the form of DNA binding proteins, transcription factors or polymerases. Some helicases are known to displace protein roadblock; for example, PcrA displaces RecA protein bound to DNA during DNA replication (Park et al., 2010). It was interesting to understand whether the active helicase core of Cas3/I-C is able to displace such roadblocks and its implications on type I-C interference.

4.2 Materials and methods

4.2.1 Molecular cloning and protein purification:

For molecular cloning, please refer to Chapter 3, section 3.2.2.

E. coli BL21 (DE3) harbouring pCascade/I-C and pCRISPR/I-C (Table 3) was grown in LB broth supplemented with 50 µg/ml kanamycin and 100 µg/ml spectinomycin at 37 °C till the OD_{600} was equal to 0.6. Cascade/I-C expression was induced with 0.2 mM IPTG and cells were allowed to grow overnight at 25 °C. Cells were harvested and re-suspended in binding buffer (20 mM Tris-Cl (pH 7.5), 150 mM NaCl, 10% glycerol, 1 mM PMSF and 6 mM β-mercaptoethanol). Cells were lysed using a cell disruptor (20 kpsi), and cell debris was removed by centrifugation at 4 °C and 16000 × g for 30 min. After cell lysis, DNase I (50 µg/ml) was added to reduce the viscosity. The clarified supernatant was passed through a pre-equilibrated 5 ml StrepTrap HP (GE Healthcare) column. After washing with 10 CV of binding buffer, Cascade/I-C was eluted in a binding buffer containing 2.5 mM desthiobiotin. Samples were pooled up and passed through the HiTrap Heparin HP column (GE Healthcare) and eluted with a linear gradient of 0.15–2 M NaCl in the binding buffer to remove additional impurities. Later, proteins were further purified through HiLoad 16/600 Superdex 200 prep grade column (GE Healthcare) and eluted in buffer containing 20 mM Tris-Cl (pH 7.5), 150 mM NaCl, 10% glycerol and 6 mM β-mercaptoethanol. Concentrated samples were flash-frozen in liquid nitrogen and stored at −80 °C until further use.

For purification of Cas3/I-C and its mutants, please refer to chapter 2, section 2.2.2

4.2.2 Electrophoretic mobility shift assay

Single-stranded DNA (Table 1) with or without 6-FAM labels were purchased from Integrated DNA Technologies, Inc (IDT) and gel purified to remove truncated DNA fragments. A 100 bp target DNA construct was generated by annealing the oligonucleotides having 3′ complementarity and subjecting them to extension in a PCR using Pfu DNA polymerase. About 50 ng of target DNA was incubated with 500 nM of Cascade/I-C complex in the buffer containing 20 mM Tris-Cl (pH 8.0), 150 mM NaCl and 1 mM DTT at 37 °C for 30–60 min to allow R-loop formation. Post incubation, one set of the sample was treated with 1 mg/ml proteinase-K to test the release of target DNA bound to Cascade/I-C. Samples were directly

loaded onto 20% (w/v) native polyacrylamide gel and electrophoresed in 1X TBE at 4 °C. While 6-FAM labelled constructs were directly visualized under UV in a gel documentation system (Bio-Rad), unlabelled constructs were visualized after staining with ethidium bromide (EtBr).

4.2.3 Assay for nuclease activity

In order to assess the nuclease activity of Cas3/I-C in the presence of Cascade/I-C, target DNA and Cascade/I-C were incubated with varied concentration of Cas3/I-C in the buffer containing 20 mM Tris-Cl (pH 8.0), 60 mM NaCl, 1 mM dithiothreitol (DTT), 10 mM $MgCl_2$ and 1 mM ATP. After the incubation, samples containing unlabelled DNA were directly loaded onto 20% (w/v) native polyacrylamide gel and electrophoresed in 1X TBE at 4 °C. DNA bands were visualized in the gel documentation system (Bio-Rad) after staining with EtBr. 6-FAM labelled DNA samples were analysed on 20% (w/v) denaturing PAGE containing 8 M urea and directly visualized without any post-staining.

To check the cleavage in the presence of single-stranded DNA binding (SSB) protein, 5′ 6-FAM labelled 60 nt ssDNA was incubated with 0.1–1 μM of SSB, and the reaction was incubated at room temperature to allow SSB to bind to DNA. Later, 500 nM Cas3/I-C along with 1 mM ATP was added to the reaction. To check metal ion dependency, 2 mM of various metal salts and 500 nM Cas3/I-C were used. The reactions were carried out at 37 °C for 10–60 min. Cleavage products were visualized by either 0.8% agarose gel electrophoresis or 10–15% (w/v) 8 M urea denaturing PAGE

4.2.4 Nuclease activity on bubble DNA construct

Unlabelled single-stranded DNAs (100 nt) having complementary terminal ends were purchased from IDT and PAGE purified to remove truncated DNA fragments (Dloop-AAG-FP and Dloop-TTC-RP; refer Table 1). These oligonucleotides (1 μM) were annealed and incubated with increasing concentration of Cas3/I-C (0.2–1 μM) in the reaction mixture containing 20 mM Tris-Cl (pH 8.0), 150 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$ and 1 mM ATP at 37 °C for 120 min. Samples were directly loaded onto 20% (w/v) native polyacrylamide gel and electrophoresed in 1X TBE at 4 °C. Gels were post-stained with ethidium bromide (EtBr), and DNA bands were visualized in the gel documentation system.

4.2.5 Nuclease activity on biotin labelled DNA construct

Biotin-labelled DNA construct with 5′ 6-FAM label was designed and purchased from Bioserve Biotechnologies (India) Pvt. Ltd., and PAGE-purified to remove free 6-FAM and truncated oligonucleotides (ssDNA-biotin; refer Table 1). To form double-stranded target DNA, both the strands were annealed by heating at 95 °C followed by gradual cooling. dsDNA or ssDNA was pre-incubated with streptavidin to form a roadblock in the molar ratio 2:3 (DNA: Streptavidin). Wherever mentioned, 500 nM of Cascade/I-C was added to form R-loop. Subsequently, Cas3/I-C was added in the reaction mixture containing 20 mM Tris-Cl (pH 8.0), 150 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$ and 1 mM ATP/ADP/AMP-PNP. Samples were analysed on 20% (w/v) denaturing PAGE containing 8 M urea and directly visualized in the gel documentation system.

4.2.6 Fluorescence quenching assay

DNA strand with 5′ biotin and 6-FAM at 5th nucleotide (NTS) and Iowa Black FQ (IB) at 29th nucleotide was ordered from IDT. All the DNA constructs were PAGE purified to remove truncated fragments and annealed by heating at 95 °C followed by gradual cooling. About 100 nM of annealed dsDNA was pre-incubated with 200 nM streptavidin and 500 nM Cascade/I-C. Cas3/I-C at increasing concentration (0–5 μM) was added in the reaction mixture containing 20 mM Tris-Cl (pH 8.0), 150 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$ and 1 mM ATP/ADP/AMP-PNP, and fluorescence intensity was immediately measured using FluoroMax®-4 (Horiba Scientific). For time-dependent measurement, 500 nM of Cas3/I-C was added in the reaction mixture mentioned above, and fluorescence intensity was measured for 120 min.

4.2.7 Anisotropy measurements

The biotin labelled ssDNA construct with 5′ 6-FAM mentioned previously (Table 1) was pre-incubated with streptavidin in the molar ratio 2:3 (DNA: Streptavidin). Cas3/I-C was added in the reaction mixture containing 20 mM Tris-Cl (pH 8.0), 150 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$ and 1 mM ATP/ADP/AMP-PNP, and anisotropy values were measured using FluoroMax®-4 (Horiba Scientific). To assess Cas3/I-C binding, anisotropy readings were

taken immediately after adding Cas3/I-C (0–10 μM). For time-dependent measurement of anisotropy, 500 nM of Cas3/I-C was added, and readings were noted for 90 min

4.3 Results

4.3.1 Cascade/I-C purification and target DNA binding

We began our study with the purification of Cascade/I-C from *B. halodurans* (Figure 4.1A) and tested its binding efficiency to the various target DNAs. Electromobility shift in the 100 bp 6-FAM labelled target DNA showed Cascade/I-C binding (Figure 4.1B), which was further confirmed through size-exclusion chromatography (Figure 4.1C). In type I-E, the PAM sequence is crucial for target binding. Mutations in the PAM region weaken the binding of Cascade and thus lead to impaired interference. We checked the binding affinity of Cascade/I-C with three different PAM containing target DNA (TTC, TAC and ATG) (Figure 4.1D) and calculated binding affinity from the band intensities as observed in Figure 4.1D. We observed that Cascade/I-C shows slight variations in the binding affinity; however, these differences were not significant (Figure 4.1D). This suggests that the binding of Cascade/I-C to target DNA does not depend on the PAM sequence; however, PAM seems to be essential for Cas3/I-C mediated target degradation (Figure 4.3A).

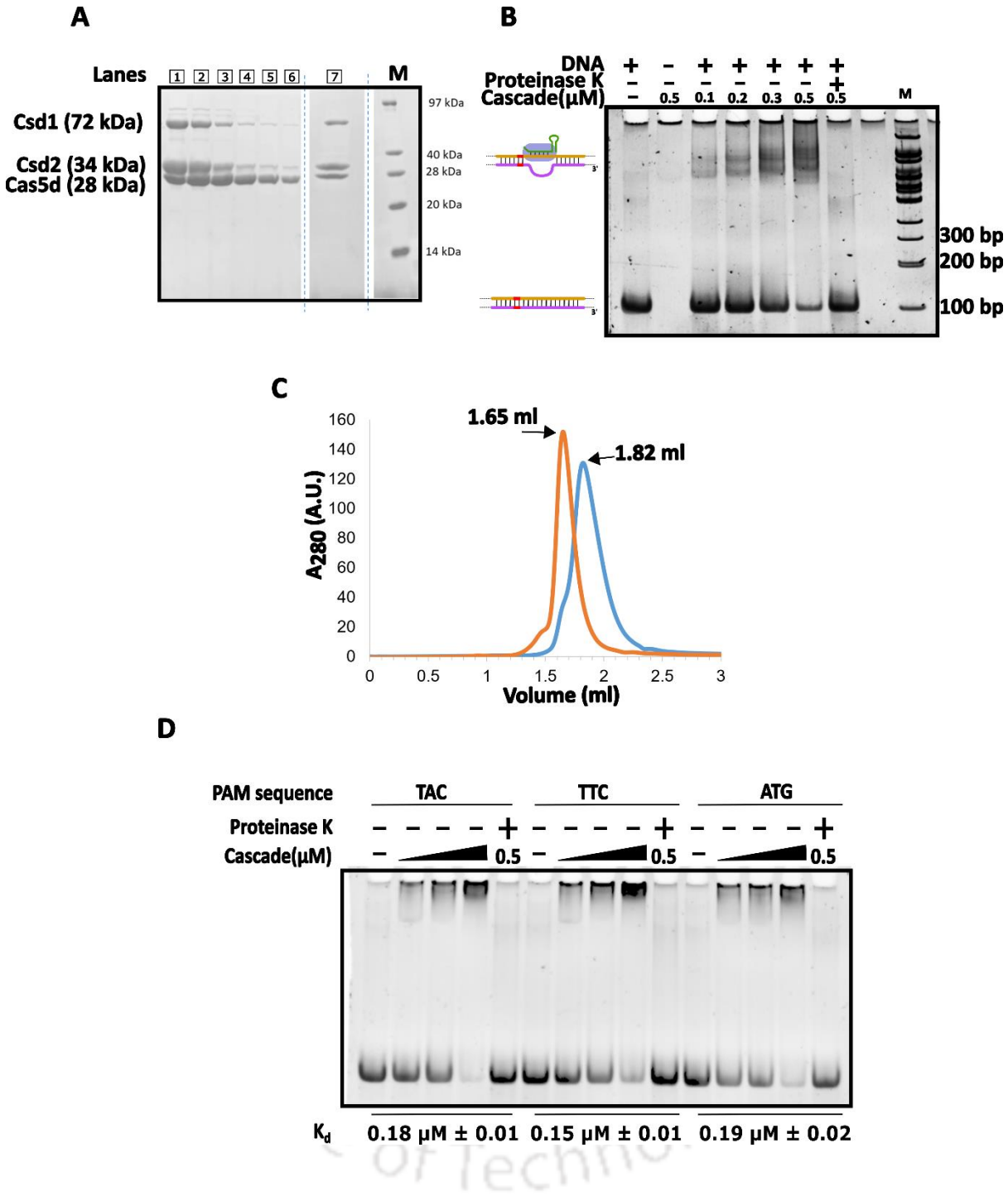

Figure 4. 1: Characterization of target DNA interaction with Cascade/I-C

(A). Purification of the Cascade/I-C complex was achieved by co-expressing Cas proteins and CRISPR array. Csd1 (Cas8c), Csd2 (Cas7) and Cas5d (Cas5) can be observed in 15 % SDS-PAGE, which shows the affinity-purified (lanes 1-6) and SEC-purified fractions of Cascade/I-C complex (lane 7). Protein marker (M) is shown on the right. The dotted line indicates a discontinuity in the gel for the purpose of clarity.

(B). A 100 bp target DNA that is complementary to crRNA was designed with a 6-FAM label at the 5'end of the non-target strand. In the presence of an increasing concentration of Cascade/I-C, a change in the migration of DNA was observed. Bound DNA fraction was released after the addition of proteinase K, which confirms the formation of the target DNA-Cascade/I-C complex. DNA marker (M) positions are shown on the right.

(C). DNA bound and free Cascade/I-C showed a significant difference in mobility in analytical size exclusion chromatography. Blue profile corresponds to free Cascade/I-C, and orange profile corresponds to DNA bound Cascade/I-C.

(D). Target DNAs with varied PAM sequences (TAC, TTC, and ATG) showed an altered binding affinity towards Cascade/I-C (0-500 nM). The canonical PAM sequence (TTC) showed a slightly stronger affinity (K_d=0.15μM) than other sequences; however, it is not significant. Standard deviation is calculated after measuring band intensities from two independent experiments.

Note: Samples in figures 4.1B and 4.1D were visualized using 15 % native PAGE.

4.3.2 Resection of short double-stranded linear DNA by Cas3/I-C necessitates the presence of Cascade/I-C

Piqued by the observation that Cas3/I-C lacks intrinsic target specificity *in vitro* and retains one *in vivo*, we set out to resolve this apparent paradox. We hypothesized that the absence of 'intrinsic' target specificity could be compensated by a mechanism that introduces 'induced' target specificity. To test our hypothesis, we designed a 100 bp DNA substrate with 'TTC' PAM at the 5′ end (Figure 3.2D and 4.2A). When Cas3/I-C alone was introduced, no apparent cleavage was noted (Figure 4.2A and 4.2C). This was surprising; however, it was consistent with our earlier observation (Figure 2.2A) that nuclease activity is attenuated for short DNA constructs (<400 bp). Led by the requirement of Cascade/I-C for CRISPR interference *in vivo*, we introduced the Cascade/I-C into the reaction. Remarkably, the target DNA—which was initially refractory to nuclease activity of Cas3/I-C—was cleaved in the presence of Cascade/I-C (Figure 4.2C). Intrigued by this, we asked what was bestowed by the Cascade/I-C on Cas3/I-C to target the DNA? Based on previous reports in type I-E (Jore et al., 2011; Rutkauskas et al., 2015; Sinkunas et al., 2013), we interpreted that the Cascade/I-C binding to target DNA ensues the generation of a DNA–RNA hybrid and a single-stranded DNA referred to as R-loop. Since Cas3/I-C requires 3′-overhang for loading (Figure 2.5), we hypothesized that the duplex nature of the target DNA perhaps abrogates the DNA binding. Therefore, it is likely that the ssDNA that was formed as a consequence of the R-loop could

then provide a platform for Cas3/I-C loading (Figure 4.2A and 4.2C). To test this, we designed a 100 bp bubble DNA construct with no base pairing in the central region, which mimics the R-loop in the absence of Cascade/I-C (Figure 4.2B). Remarkably, we found that Cas3/I-C sliced the bubble DNA even in the absence of Cascade/I-C, suggesting that it is indeed the single-stranded region that is becoming the loading point for Cas3/I-C (Figure 4.2D and 4.3B). Additionally, we also found the presence of small DNA fragment (~75–85 nt) when ATP was not included in the reaction that hints at the ATP independent nicking of the target DNA (Figure 4.3B)

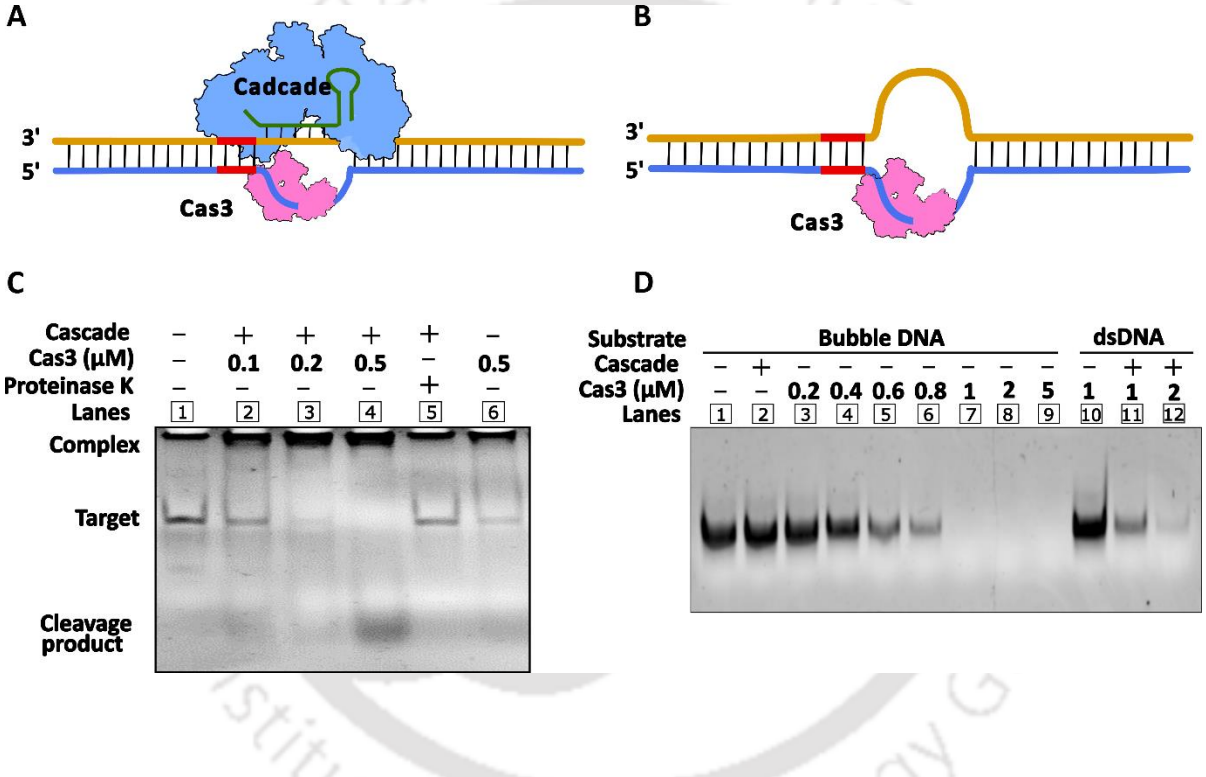

Figure 4. 2: Cascade/I-C provides a single-stranded DNA loop for Cas3/I-C binding.

(A). A schematic model suggests that Cascade/I-C binding to target DNA provides an ssDNA platform for Cas3 binding. The location of PAM (TTC in red) in the 100 bp construct is indicated.

(B). A 100 bp bubble DNA construct that lacks base pairing in the centre was designed with PAM (TTC in red). The single-stranded DNA loop is suggested to facilitate Cas3/I-C binding in the absence of Cascade/I-C.

(C). DNA cleavage was tested on 20% native PAGE with increasing concentration of Cas3/I-C in the presence of Cascade (lanes 2–4). Proteinase K was added to release bound DNA (lane 5). No cleavage was noticed in the absence of Cascade (lane 6).

(D). Bubble DNA was cleaved in the absence of Cascade/I-C (lanes 3–9), whereas dsDNA was cleaved only in the presence of Cascade/I-C (lanes 11–12). DNA was visualized using a 20% native PAGE.

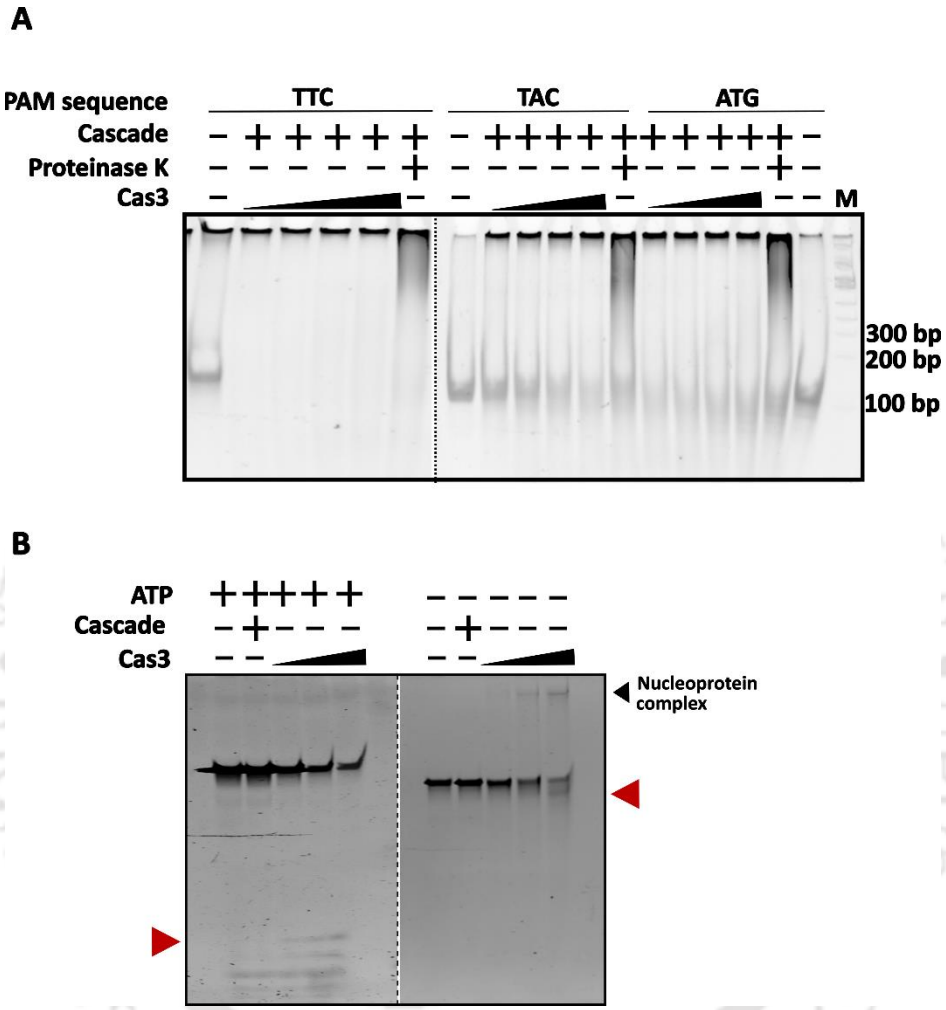

Figure 4. 3: Characterization of target DNA interaction with Cascade/I-C and Cas3/I-C.

(A). Interference was checked on target DNA with varying PAM sequences. Target DNA with TTC containing PAM was cleaved by Cas3/I-C (0-500 nM), whereas TAC and ATG were non-functional. The dotted line indicates two separate gels, which are merged for comparison. Cascade/I-C was used at 300 nM. DNA marker (M) positions are shown on the right.

(B). 100 bp Bubble DNA was incubated with an increasing concentration of Cas3/I-C (0-500 nM) in the presence and absence of 1mM ATP. The cleavage products are indicated with a red triangle. In the absence of ATP, we observed a shifted DNA band (nucleoprotein complex) which is absent when ATP is added to the reaction. Additionally, the appearance of a small DNA fragment (indicated by a red triangle) in the absence of ATP suggests a nick in the target DNA. The dotted line indicates two separate gels, which are merged for comparison.

Note: Samples in figures 4.3A and 4.3B were visualized using 15 % native PAGE.

4.3.3 Cas3/I-C cleaves target via a reeling mechanism

The interaction between Cascade/I-C and Cas3/I-C facilitates the target DNA cleavage. After recruitment by Cascade/I-C, the helicase activity of Cas3/I-C may allow it to translocate on the ssDNA, unwinding the target DNA in 3′ to 5′ direction followed by cleavage via nuclease domain, eventually leading to the dissociation from Cascade/I-C. An alternative possibility is to remain associated with Cascade/I-C and reel in the target DNA, which may result in the formation of a loop in the target strand. In order to test how Cas3/I-C unwinds double-stranded target DNA upon recruitment by Cascade/I-C, we designed a 100 bp dsDNA with a fluorophore (6-FAM) and a quencher (Iowa Black FQ) (Figure 4.4). The non-target strand (NTS) was tagged with biotin at the 5′ end and 6-FAM at the 5th nucleotide from the 5′ end. On target strand (TS), Iowa Black FQ (IB) was introduced proximal to the PAM at the 29th nucleotide from 3′ end (Figure 4.4). Biotin was introduced at 5′ end with the presumption that Cas3/I-C would halt on encountering the biotin-streptavidin block, thus sustaining 6-FAM quenching by Iowa Black FQ for a sufficiently long duration for measurements. Since 6-FAM and IB were far apart, the fluorescence of 6-FAM was not quenched. However, if Cas3/I-C reeled in the DNA, 6-FAM would come in proximity with IB, which in turn would quench the fluorescence (Figure 4.4). Target DNA was saturated with streptavidin and Cascade/I-C before adding Cas3/I-C, and fluorescence intensity was measured to estimate the extent of quenching.

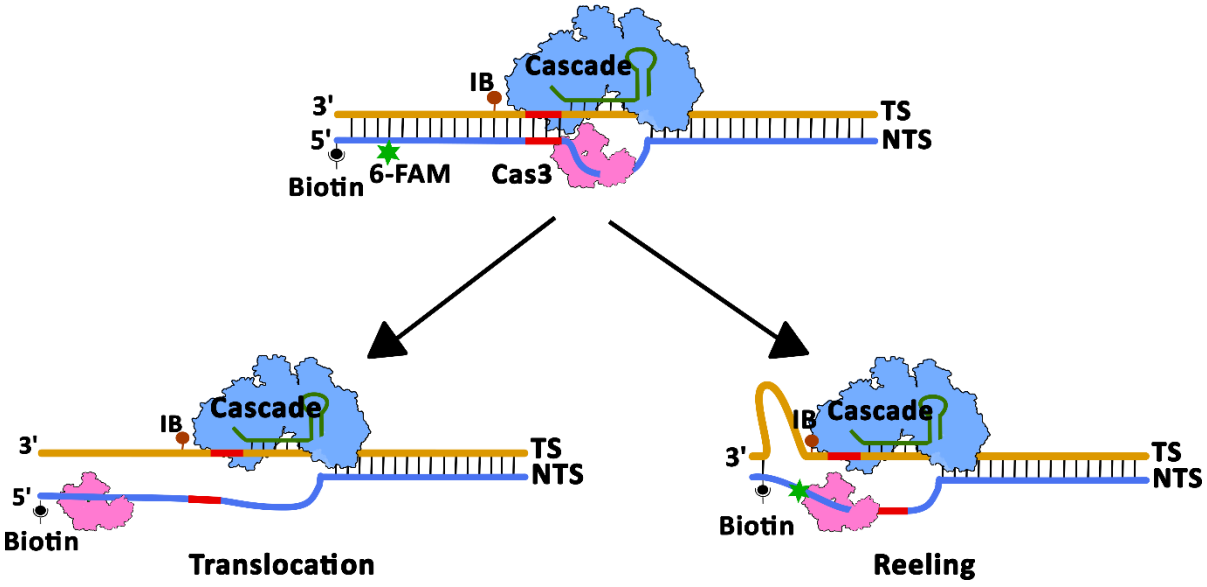

Figure 4. 4: Reeling and Translocation modes of target degradation.

Schematic representation of 100 bp DNA substrates with biotin (black dot) at 5′ end of the non-target strand (NTS), 6-FAM (shown as a green star) at 5th nucleotide, Iowa Black® FQ (shown

as a brown dot) at 29th nucleotide on target strand (TS) and PAM sequence TTC (depicted in red). Cas3-Cascade interaction is not stable in the translocation mode, and Cas3 would translocate towards the 5' end. Cas3-Cascade interaction is stable in the reeling mode, and thus Cas3 would pull the DNA leading to fluorescence quenching by Iowa black quencher.

Interestingly, we observed significant fluorescence quenching with increasing Cas3/I-C concentrations (0–5 μM) when Cascade/I-C was present. On the other hand, quenching was not observed when the reaction did not contain Cascade/I-C (dsDNA), or only a non-target strand (ssDNA) was present (Figure 4.5A). Moreover, fluorescence quenching was observed only when ATP was supplied to Cas3/I-C in addition to Cascade/I-C (Figure 4.5A and 4.5B). Taken together, our result suggests that Cas3/I-C remain bound to Cascade/I-C and pulls in 6-FAM towards IB, which leads to fluorescence quenching.

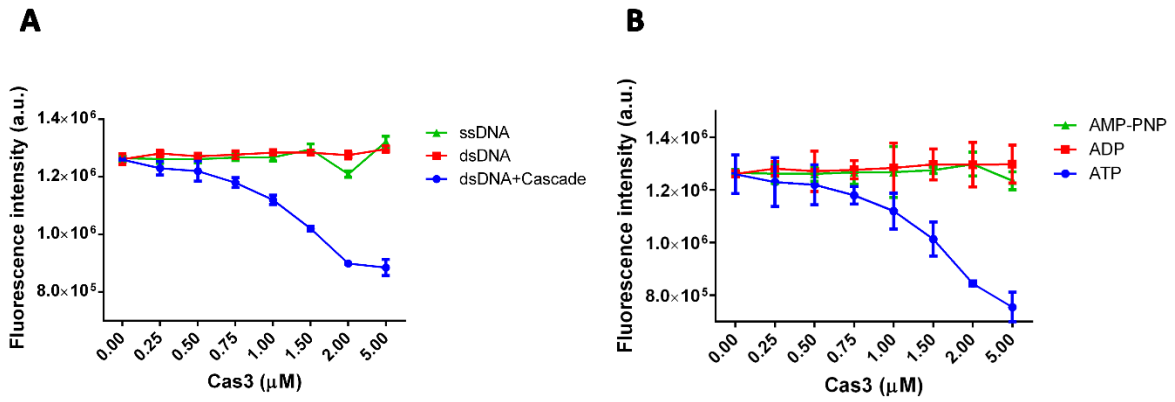

Figure 4. 5: Cascade/I-C and Cas3/I-C form a stable complex during interference.

(A). The substrate mentioned in figure 4.4 was incubated with or without Cascade/I-C (1 μM) and increasing the concentration of Cas3/I-C (0–5 μM). A significant decline in fluorescence intensity is evident when both Cascade/I-C and Cas3/I-C were present. There was no apparent quenching when dsDNA and ssDNA were used in the absence of Cascade/I-C.

(B). Fluorescence quenching was observed in the presence of ATP but not when ADP and AMP-PNP were used.

4.3.4 Stalling the helicase motor of Cas3/I-C stimulates the nuclease activity

Bacterial cytoplasm holds a repertoire of DNA-binding proteins, and therefore, Cas3/I-C, during its processive nuclease–helicase activity, is expected to encounter proteins such as single-stranded DNA binding protein (SSB), RNA polymerases etc. In order to decipher what would happen if Cas3/I-C encounters such roadblocks, we intentionally introduced a stalling site in the target DNA. The stalling site comprises a biotin labelled nucleotide that binds to streptavidin, which is expected to block the movement of the Cas3/I-C helicase motor. We labelled the 5′ end of the non-target strand (NTS) with FAM and the 12th nucleotide from 5′ end with biotin (Figure 4.6). Prior to the addition of Cas3/I-C, target DNA was incubated with Streptavidin and Cascade/I-C.

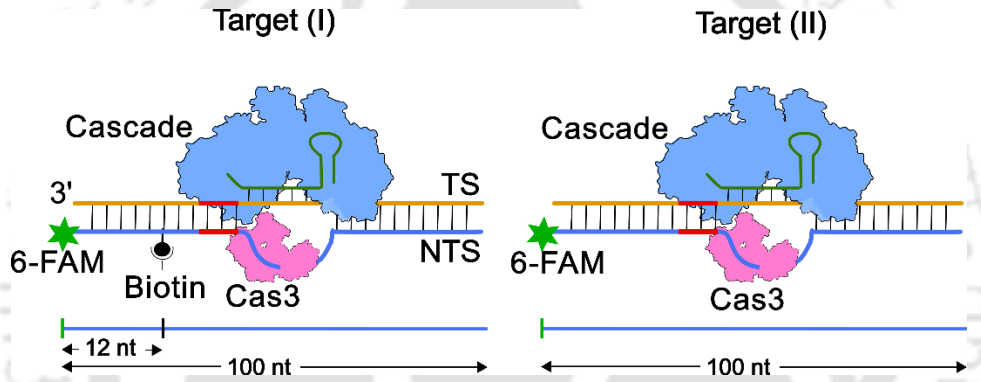

Figure 4. 6: Target design for stalling helicase motor

Schematic representation of target DNA (100 bp) with PAM (TTC in red) end-labelled with 6-FAM at 5′ end of the non-target strand and biotinylated at 12th nucleotide (Target I). A similar target DNA without biotin (Target II) is also represented.

Cas3/I-C was introduced with ATP/ADP/AMP-PNP, and cleavage was monitored using denaturing PAGE (Figure 4.7). We could discern the accumulation of a prominent ~55–60 nt band in the absence of ATP (lanes 10, 11, 13, 14 and 16 in Figure 4.7). Since the helicase motor is inactive in the absence of ATP, we presume it to be a single-stranded nick generated by Cas3/I-C nuclease upon binding to NTS. Apart from the above observation, the presence of cleavage product (~55–60 nt) in the absence of nucleotide suggests that Cas3/I-C induced initial nick is independent of the nucleotide-bound state (Figure 4.7). Interestingly, in the presence of ATP, we observed a prominent band at ~40 nt; however, the band corresponding to ~55–60 nt was not perceptible (Figure 4.7). On the contrary, ~40 nt band was not visible

when the biotin-streptavidin block was absent, even in the presence of ATP; however, we could spot the ~55–60 nt band (Figure 4.7). This suggests that Cas3/I-C introduces a nick in the single-stranded region of NTS irrespective of the nucleotide-bound state.

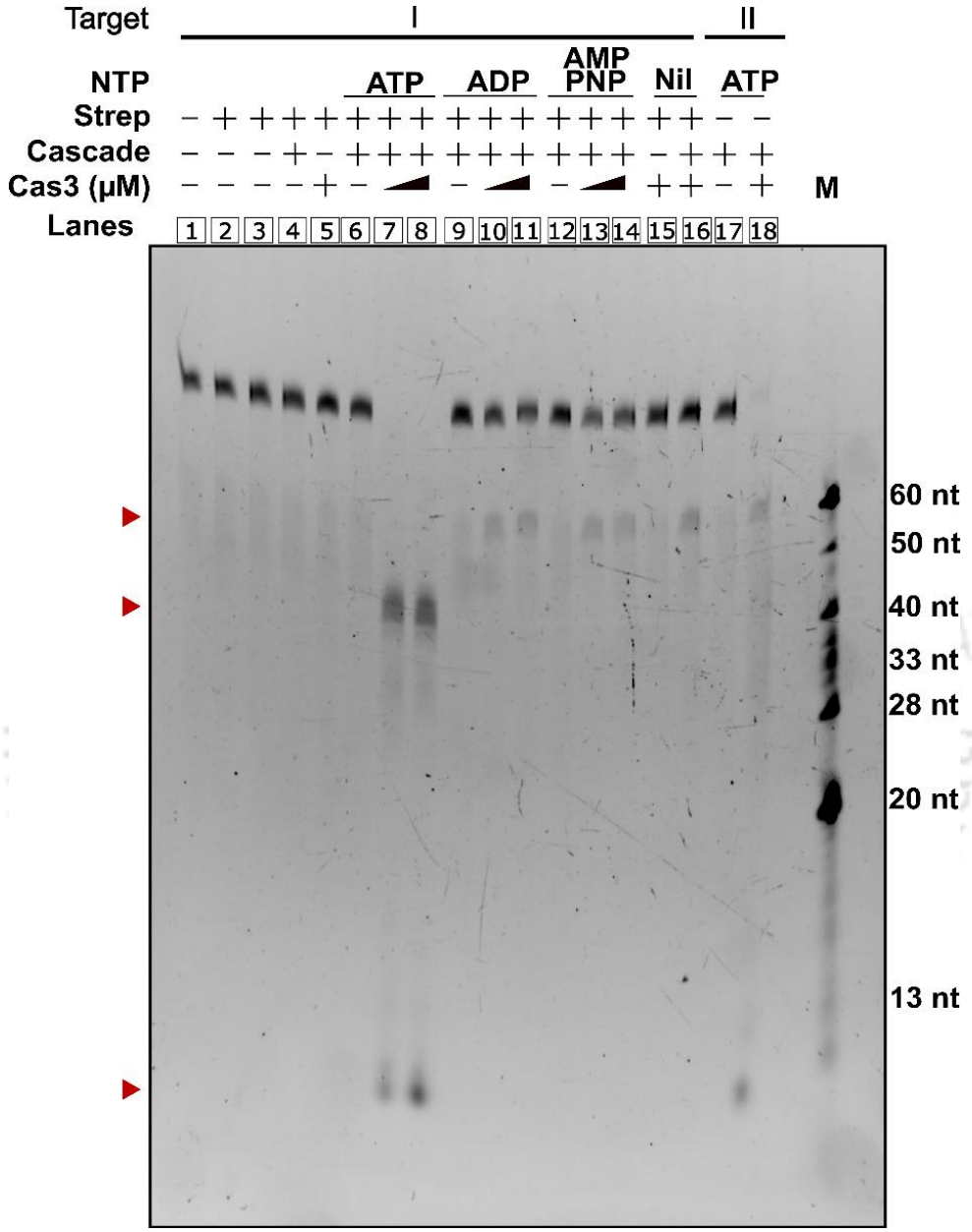

Figure 4. 7: Stalling the translocation of helicase motor stimulates cleavage.

Target (I) and (II) mentioned in figure 4.6 were incubated with streptavidin and Cascade/I-C to form an interference complex. Cleavage was initiated by addition of Cas3 (0.2 and 0.5 μM) and 1 mM ATP/ADP/AMP-PNP. Prominent cleavages of the target (I) and (II) are indicated by a red triangle (~60, 40 and 10 nt). The size of the DNA markers is shown in lane M. A 20% denaturing PAGE was used to assess the cleavage of DNA.

To account for the appearance of ~40 nt band specifically in the presence of ATP and biotin-streptavidin block, we hypothesized that the helicase motor of Cas3/I-C, in the presence of Cascade/I-C and ATP, would have reeled in the NTS from the loading point and a biotin-streptavidin roadblock could have triggered cleavage by the trailing nuclease domain. To test our hypothesis, we intentionally introduced a similar stalling site in single-stranded DNA labelled with 6-FAM at 5′-end. Two constructs were made in which the biotin was labelled at 12th (Target A) and 20th (Target B) nucleotide positions, respectively (Figure 4.8A and 4.9A).

We measured the change in fluorescence anisotropy of substrates mentioned above upon the addition of Cas3/I-C (Figure 4.8B, 4.9B, 4.10A and 4.10B). Initially, we observed a slight increase in anisotropy for biotin-labelled DNA, suggesting early binding of Cas3/I-C to ssDNA. With time, we could discern a significant decline in anisotropy value in the absence or when ATP/ADP/AMP-PNP was present; however, the decay was steep when ATP was present (Figure 4.8B and 4.9B). Remarkably, in the absence of biotin, the anisotropy decay is not steep (Figure 4.8B and 4.9B). This decrease in anisotropy value with time suggests that DNA is fragmented, possibly due to Cas3/I-C induced cleavage. In order to understand such cleavage, Cas3/I-C was mixed with ATP/ADP/AMP-PNP and bands were analysed using denaturing PAGE (Figure 4.8C and 4.9C). In line with the above observations, two prominent bands at the top corresponding to ~30 and 40 nt for Target A (lanes 4, 5, 7, 8 and 11 in Figure 4.8C) and ~45 and 55 nt for Target B (lanes 4, 5, 7, 8 and 11 in Figure 4.9C), respectively, were perceptible in the presence of ATP/ADP/AMP-PNP. For Target A, we posit that the ~30 nt band could arise by cleavage due to biotin block, and the other ~40 nt could emerge by cleavage when a second Cas3/I-C gets stalled by the preceding Cas3/I-C. A similar scenario can be envisaged for Target B as well. Whereas in the absence of a biotin-streptavidin block, we could not observe any cleavage (Figure 4.10E). Interestingly, we also observed small sized bands (<10 nt), which suggests that few Cas3/I-C could trespass the biotin-streptavidin block and possibly get stalled by 6-FAM at the 5′ end (Figure 4.8C and 4.9C).

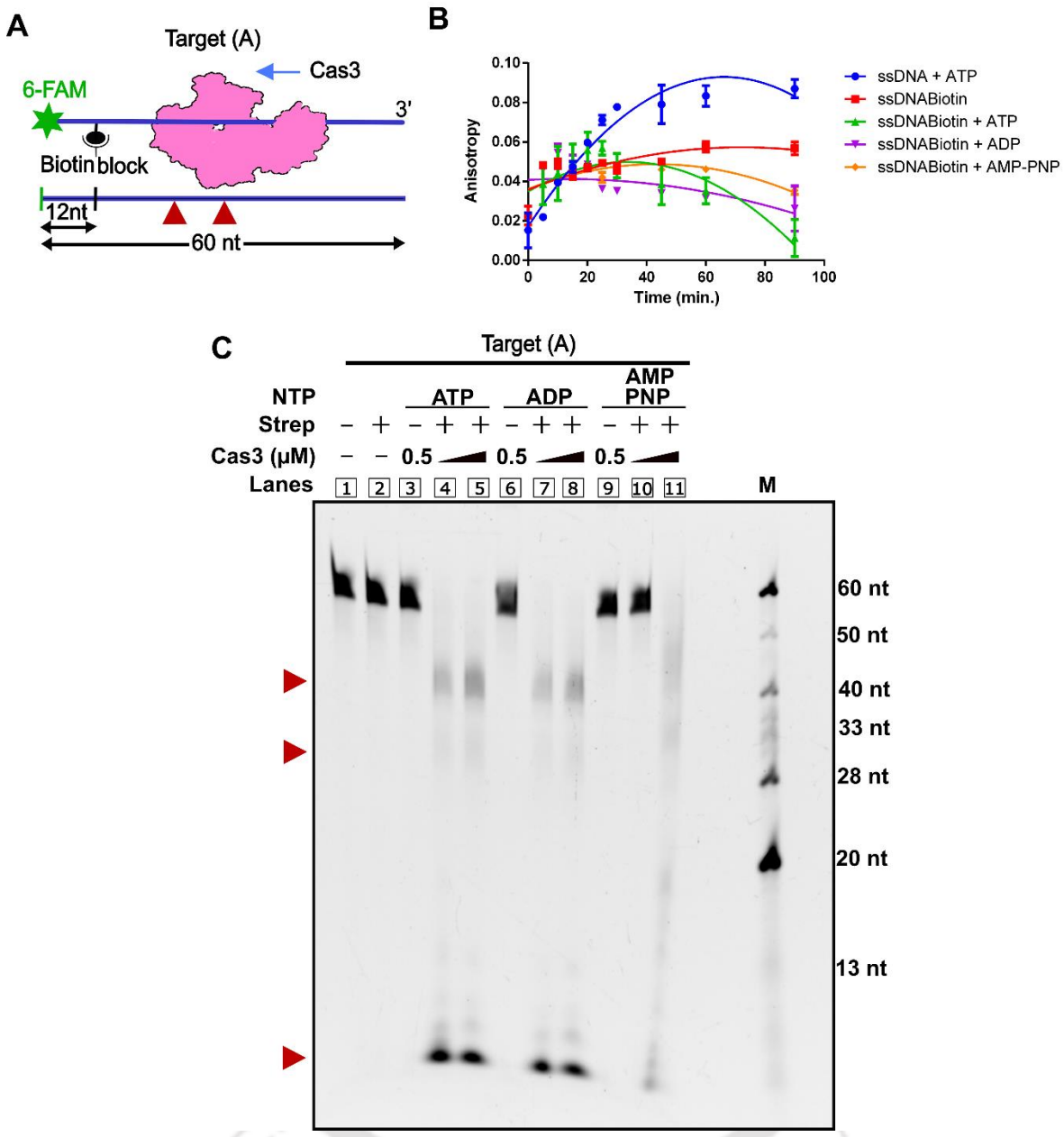

Figure 4. 8: Roadblock in the translocation of Cas3/I-C stimulates cleavage

(A). A schematic representation of 60 nt 5′ 6-FAM labelled ssDNA with biotin at 12th nt (Target A) is shown.

(B). The ssDNA mentioned above was incubated with 200 nM of Cas3/I-C for several time points, and anisotropy measurements were recorded. With time, the decrease in anisotropy values was observed for ssDNA with a biotin roadblock.

(C). The ssDNA was pre-incubated with streptavidin before the addition of Cas3/I-C. Prominent DNA cleavage products were observed in the presence of ATP at ~40 nt in target A, indicated with a red triangle. In the presence of AMP-PNP, a higher Cas3 concentration was required for the cleavage. The size of the DNA markers is shown in lane M. A 20% denaturing PAGE was used to assess the cleavage of DNA.

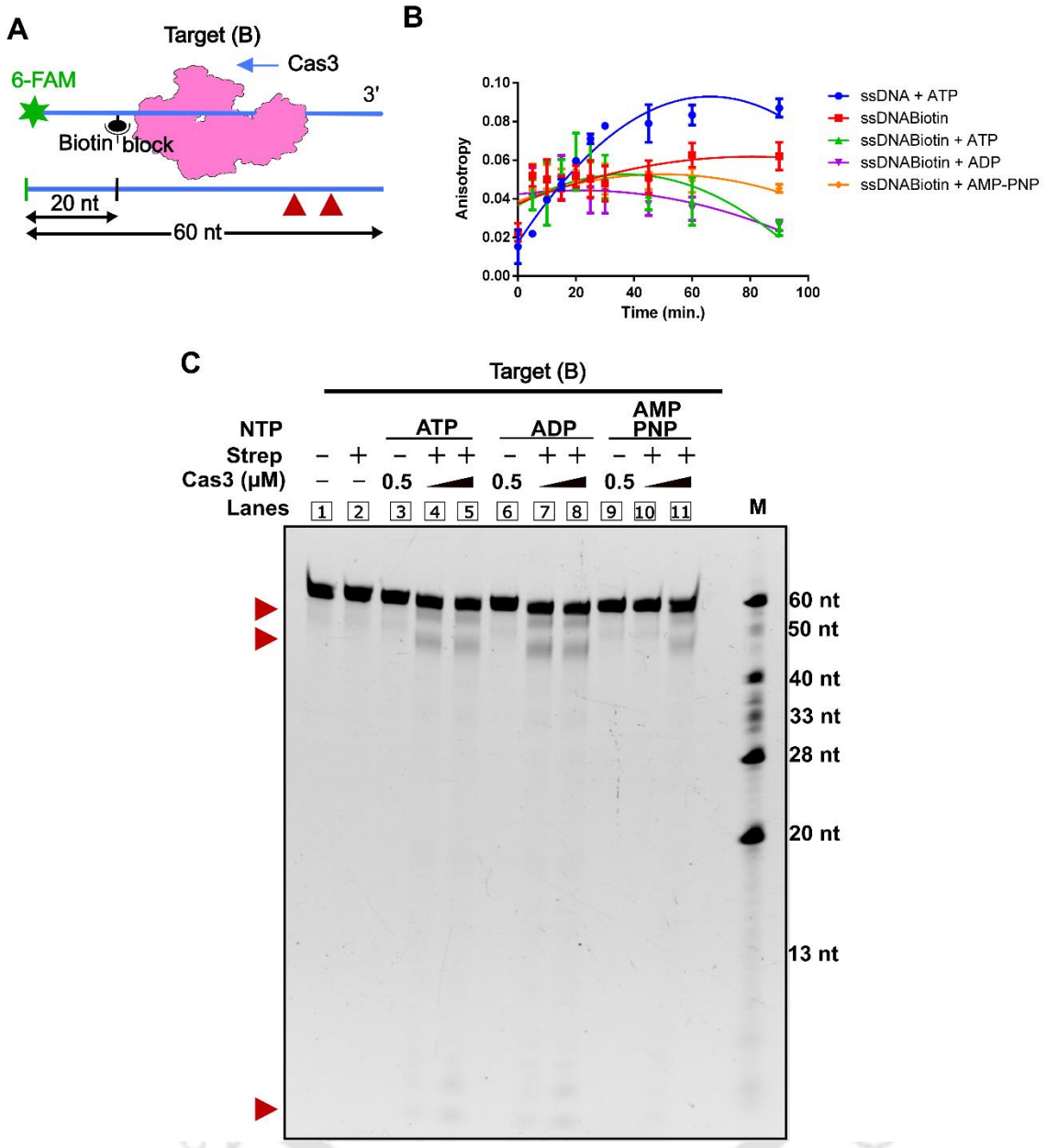

Figure 4. 9: Roadblock in the translocation of Cas3/I-C stimulates cleavage.

(A). A schematic representation of 60 nt 5′ 6-FAM labelled ssDNA with biotin at 20th nucleotide (Target B) is shown.

(B). The ssDNA mentioned above was incubated with 200 nM of Cas3/I-C for several time points, and anisotropy measurements were recorded. With time, the decrease in anisotropy values was observed for ssDNA with a biotin roadblock.

(C). ssDNA was pre-incubated with streptavidin before the addition of Cas3/I-C. Prominent DNA cleavage products were observed in the presence of ATP at ~50 nt in target B, indicated with a red triangle. In the presence of AMP-PNP, a higher Cas3 concentration was required for the cleavage. The size of the DNA markers is shown in lane M. A 20% denaturing PAGE was used to assess the cleavage of DNA.

Since the biotin-streptavidin block is highly stable and not present inside the bacterial cell, we used a single-stranded DNA binding protein (SSB) to create a roadblock. We incubated 60 nt 5' 6-FAM labelled with SSB at room temperature to allow SSB to bind to ssDNA. Later, we added Cas3/I-C along with ATP in the reaction. We could observe multiple bands on denaturing PAGE. Since SSB lacks sequence specificity and binds to ssDNA at random positions, the presence of multiple DNA bands suggests Cas3/I-C might have stalled at multiple locations, leading to cleavage by the nuclease domain of Cas3/I-C (Figure 4.10C). Altogether the data suggest that when the helicase motor of Cas3/I-C encounters a roadblock en route, the nuclease activity associated with the HD domain is stimulated.

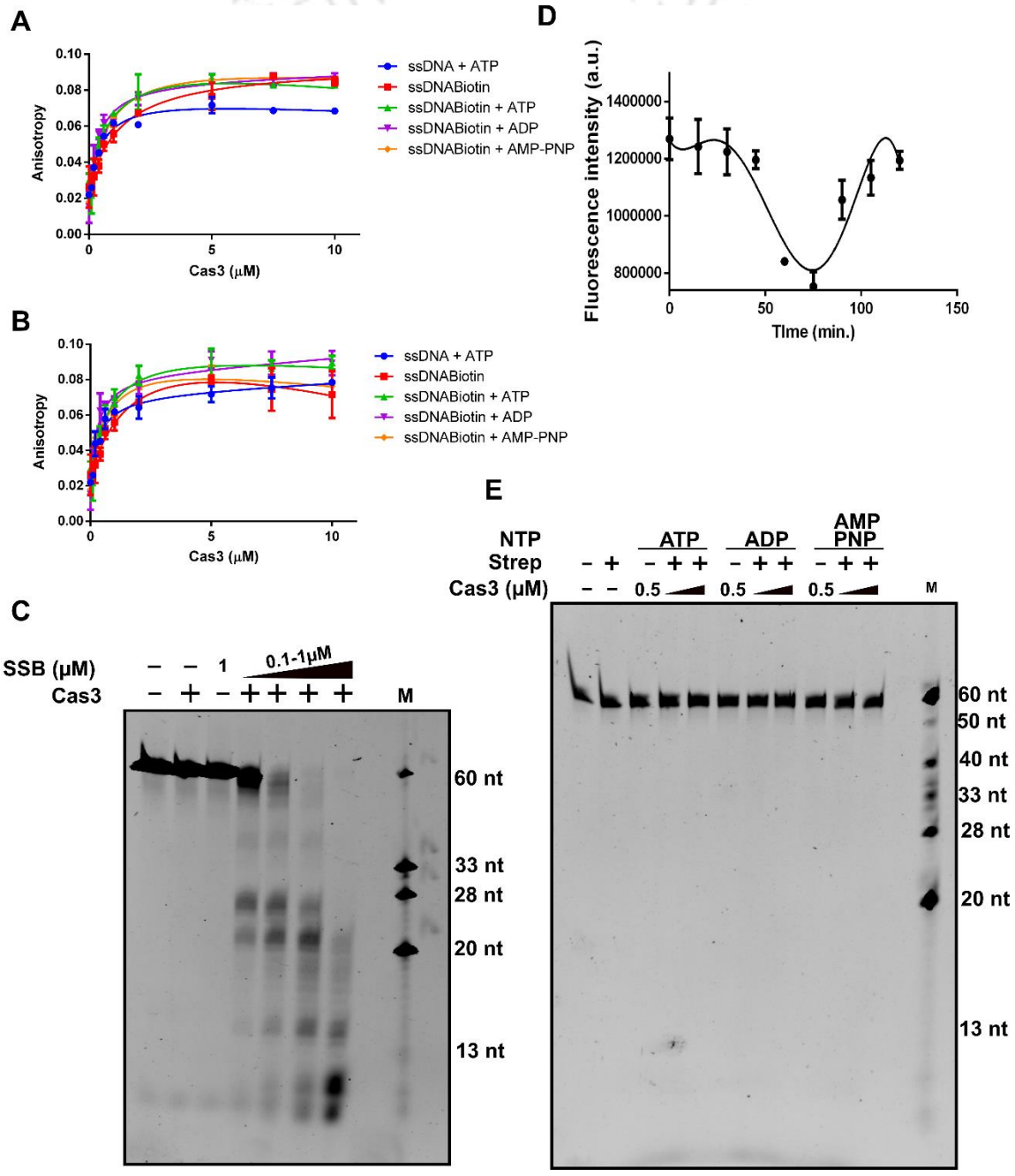

Figure 4. 10: Target degradation with and without Cas3/I-C stalling.

(A)-(B). A roadblock was created using biotin-streptavidin conjugate. An increase in anisotropy was observed when the ssDNA mentioned above (Target A and Target B in figure 4.8A and 4.9A) was incubated with increasing Cas3/I-C concentration, which suggests that Cas3/I-C binds to ssDNA irrespective of the nucleotide.

(C). Since the biotin-streptavidin block is highly stable and not present usually inside the bacterial cell, we used single-stranded DNA binding protein (SSB) to create a roadblock. A 60 nt ssDNA with 5' 6-FAM label was incubated with single-stranded DNA binding protein (SSB, 0.1-1 μM) and later 500 nM Cas3/I-C along with 1 mM ATP was added. Multiple cleavage bands were observed with increasing concentration of SSB, suggesting cleavage by Cas3/I-C at several locations on the DNA. Since SSB binds to ssDNA non-specifically at multiple locations, it might act as a roadblock for Cas3/I-C, which in turn favours cleavage. DNA marker (M) positions are shown on the right. Samples were visualized using a 20% 8 M urea denaturing PAGE.

(D). DNA construct showed in Figure 4.4 was incubated with 500 nM of Cas3/I-C, and fluorescence intensity was measured over a period of 120 mins. As expected from the reeling mechanism, Cas3/IC fluorescence intensity showed a sharp decline for 60 mins due to quenching of 6-FAM by Iowa Black FQ. Subsequently, there was an increase in fluorescence intensity, suggesting cleavage and release of NTS.

(E). ssDNA without a biotin block was incubated with Cas3/I-C (250 nM and 500 nM) and 1 mM ATP/ADP/AMPPNP for 30 mins. No cleavage was observed, suggesting that in the absence of a blockade, Cas3/I-C translocate without exhibiting apparent nuclease activity. DNA marker (M) positions are shown on the right. Samples were visualized using a 20% 8M urea denaturing PAGE.

4.4 Discussion

A significant outcome of this chapter underlines the molecular events that occur during interference in the type I-C CRISPR-Cas system. During the interference stage, the Cascade complex using sequence complementarity of crRNA identifies 30-35 nt in the protospacer sequence adjacent to PAM on the invading MGE. After the target is identified, duplex DNA is directionally (PAM proximal end to PAM distal end) unwound (Rutkauskas et al., 2015; Xiao et al., 2017a), and a DNA:RNA hybrid called R-loop is formed (Gasiunas et al., 2014; Hochstrasser et al., 2014; Huo et al., 2014a; Ivančić-Baće et al., 2012; O'Brien et al., 2020; Rutkauskas et al., 2015; Wiedenheft et al., 2011; Xiao et al., 2018; Xiao et al., 2017a). Formation of R-loop leads to displacement and emergence of an ssDNA region on the non-target strand (Figure 1.11) (Redding et al., 2015; Rutkauskas et al., 2015; Szczelkun et al.,

2014; van Erp et al., 2017; Westra et al., 2012b; Wiedenheft et al., 2011; Xiao et al., 2017a). Biochemical studies on type I-E CRISPR systems in *E. coli*, *Thermobaculum terrenum,* and *T. fusca* provides detailed information on how the Cascade complex locates and forms R-loop and how Cas3 is recruited for final cleavage. In type I-E, Cas3 directly interacts with Cas8, which is evident from the hydrogen-deuterium exchange and structural studies on Cas3 and Cascade (van Erp et al., 2017; Xiao et al., 2018). However, all this information was limited to type I-E systems. Our results suggest that the single-stranded region is enough for Cas3/I-C to cleave DNA and hints towards an important role played by the Cascade complex during interference. The displacement of the non-target strand by the Cascade complex provides a loading point for Cas3/I-C. Once Cas3/I-C is loaded, the target DNA is cleaved by the active cooperation of helicase and nuclease domains. However, these findings do not refute the significance of Cas3-Cascade interaction. As shown in type I-E, Cas3-Cascade interaction may help in directing its nuclease activity exclusively towards CRISPR immunity (Rutkauskas et al., 2015; Sashital et al., 2012; Sinkunas et al., 2013), thus preventing non-specific cleavage of the host genome. Moreover, the recently discovered type IV CRISPR-Cas system (Figure 1.2) lacks the conserved adaptation module and Cas3-like effector nuclease, and it was observed that these systems co-exist with other CRISPR types (Pinilla-Redondo et al., 2020). It was found that type IV crosstalk with co-existing type I-E CRISPR system and believed to share Cas proteins (Pinilla-Redondo et al., 2020). Due to the differences in CRISPR architecture, a strong interaction between co-existing Cascade/type-IV and Cas3/I-E is unlikely; however, Cascade-like complex in type IV system may borrow Cas3/I-E by displacing non-target strand and subsequently allowing Cas3 loading. Taken together, our data suggest that the recruitment of Cas3 is not solely contingent on the Cas3-Cascade complex, which is in part driven by the single-stranded region created by R-loop formation.

The R-loop has no free 3′ end for helicase loading, and therefore threading the ssDNA into the helicase domain creates a topological problem. This can be overcome if the nuclease activity of Cas3/I-C produces a free 3′ end. Nonetheless, the homology model of Cas3/I-C based on the crystal structure of Cas3/I-E suggests that the active site of HD nuclease is oriented toward the helicase domain, and therefore, it is unlikely to access the ssDNA directly. Consequently, it appears that the ssDNA has to enter the HD active site via the helicase domain only. However, our data suggest that upon recruitment to the target site, Cas3/I-C nicks non-target strand at ~30–35 nt from PAM (Figure 4.7). This initial nick in the absence of ATP points toward the direct binding of the nuclease domain at the PAM distal region of the R-loop,

which then renders a loading point for the helicase domain (Loeff et al., 2018). This also concurs with the interpretation from the recent Cryo-EM structure of Cascade/I-E and Cas3/I-E (Xiao et al., 2018). Our data further suggest that though Cas3/I-C lacks DNA unwinding activity in the absence of ATP, it can translocate on the ssDNA (Figure 4.8 and 4.9). In the presence of ATP, it seems that Cas3/I-C unwinds the target DNA as well as channelizes the displaced ssDNA toward the active site of the HD nuclease through the helicase motor activity. This implies that the DNA is unlikely to reach the HD nuclease in the absence of the helicase motor activity. This is in line with the current understanding of Cas3/I-E mediated cleavage and explains the helicase domain requirement for CRISPR interference (Figures 4.7 and 3.1) (Xiao et al., 2018; Xiao et al., 2017a).

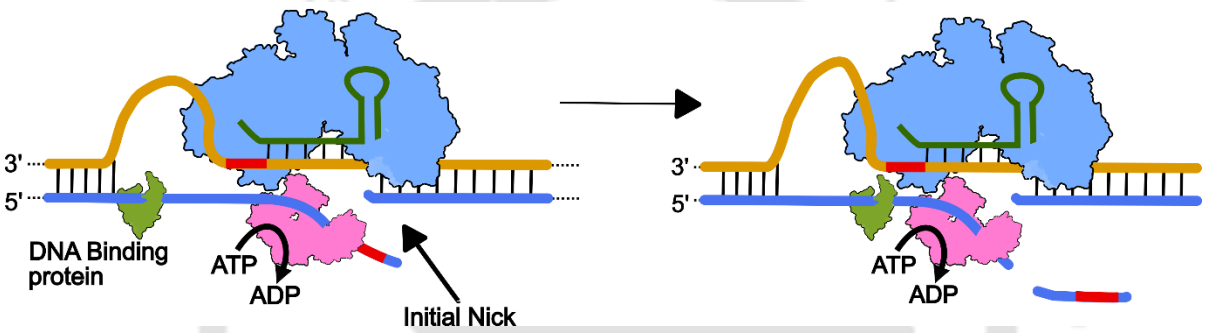

Figure 4. 11: Schematic representation of target degradation by type I-C interference.

The conformational changes in the Cascade complex lead to the recruitment of Cas3 towards the target location. Cas3 nicks the displaced single-stranded non-target strand (bottom strand) before loading. Subsequently, the helicase core of Cas3 reels the non-target strand while remaining bound to the Cascade complex. This movement may produce a bulge on the target strand (top strand). During the reeling, Cas3 may encounter roadblocks leading to cleavage.

Presumably, we posit that the incessant movement of DNA toward the HD nuclease would not provide enough residence time for the alignment of catalytic residues. This is likely to impede the nuclease activity as long as the helicase motor is active (Figures 3.4B, 4.10D and 4.10E). Therefore, whenever there is a blockade for helicase movement, it is likely that the catalytic residues of HD nuclease have sufficient residence time to be suitably oriented toward the scissile bond (Figures 4.7, 3.4B). This appears to be an ingenious strategy to time the cleavage. Our data on Cas3/I-C are in agreement with the recent finding where stalling of Cas3/I-E translocation at roadblocks stimulates nuclease activity that is otherwise reported to be sparse (Dillard et al., 2018; Loeff et al., 2018). After the Cas3/I-C loading onto the single-stranded region, we envisage two modes of helicase motor activity: (i) translocation—Cas3/I-

C is mobile with respect to the Cascade:target-DNA complex and (ii) reeling—Cas3/I-C is stationary with respect to Cascade:target-DNA. In *E. coli,* Cas3/I-E is known to leave the Cascade complex behind. In this process, the Cas3/I-E moves towards 5' end cleaving ssDNA generated by active helicase domain and no re-association of Cas3/I-E with Cascade complex was observed (Mulepati and Bailey, 2013; Redding et al., 2015). In contrast, Cas3/I-E in *T. fusca* remains associated with the Cascade complex, and the helicase domain pulls the single-stranded non-target strand of DNA (Dillard et al., 2018). However, in line with this previous report on Cas3/I-E ((Dillard et al., 2018; Loeff et al., 2018)), our data showed that the cleavage pattern concurs with a DNA reeling mechanism wherein Cas3/I-C has an intimate interaction with Cascade/I-C during DNA unwinding (Figure 4.5 and 4.10D). In this setup—where the mobility of the first Cas3/I-C molecule is locked due to interaction with Cascade/I-C—a caravan consisting of multiple Cas3/I-C that can load onto the R-loop is likely to collide with each other as well as with the one that interacts with Cascade/I-C to inflict multiple cleavages (Figure 4.8 and 4.9). While the rationale for adopting reeling over translocation is not abundantly clear, it is tempting to speculate that reeling allows Cas3/I-C to inflict cleavage proximal to the interference effector complex, thus exhibiting an apparent target site-specificity.

4.5 Summary

In this chapter, we have shown that the ssDNA region created as a result of R-loop formation by the Cascade complex acts as a loading point for Cas3/I-C. Cas3/I-C essentially nicks the single-stranded non-target strand at ~ 30 nt proximal to PAM before loading and translocate in 5' to 3' direction utilizing energy from ATP. It was found that in the absence of ATP, Cas3/I-C nicks target DNA; however, it does not translocate or cleave DNA further. Our fluorescence quenching data suggests that Cas3-Cascade interaction is stable in type I-C system, and as a result, Cas3/I-C exhibits a reeling mechanism during interference. Further, if the Cas3/I-C helicase motor encounters a roadblock, the nuclease activity associated with the nuclease domain is stimulated. Overall, we have probed the molecular events that occur during interference in the type I-C CRISPR-Cas system, which provided new insights on the specificity of molecular interactions.

Chapter 5: Deciphering the importance of CRISPR Interference in Primed Adaptation

5.1 Introduction

Despite being an elegant system, which targets invading MGEs with high specificity, the CRISPR-Cas system has its limitations. The replication machinery of phages is error-prone and frequently inserts random mutations in its genome during replication. The occurrence of mutation in the PAM and 'seed' region of the protospacers leads to defective interference, and consequently, the MGEs escape/avoid the defences of the CRISPR-Cas system. To win the evolutionary arms race, bacteria have conceived a mechanism to acquires multiple spacers from mutant MGEs. The acquisition of multiple spacers and their subsequent use during interference results in a robust immune response against phages. This is achieved through an alternative spacer acquisition mechanism called primed adaptation (Datsenko et al., 2012; Nussenzweig et al., 2019). The primed adaptation is probed broadly in type I systems and only recently in type II systems.

Most of the understanding of primed adaptation in type I is attributed to studies in type I-E (Datsenko et al., 2012; Dillard et al., 2018; Fineran et al., 2014; Kuznedelov et al., 2016; Savitskaya et al., 2013; Semenova et al., 2016; Staals et al., 2016a; Swarts et al., 2012; Xue et al., 2015; Xue et al., 2016) and I-F (Richter et al., 2014; Staals et al., 2016a) with a few reports on I-B (Li et al., 2014) and I-C (Rao et al., 2017). Depending on the CRISPR-Cas type, a few differences in the acquisition mechanism have been observed; for example, there is a strand bias between target and non-target strands during primed adaptation. In type I-B, 70% of the acquired spacers during primed adaptation is from the target strand, whereas in type I-E, it is 90%. Moreover, Cas4 is absent in type I-E, and Cas1-Cas2 associate with Cas3 to form a Primed Adaptation Complex (PAC) (Dillard et al., 2018). Such a phenomenon has not been observed in other CRISPR types yet. In type I-F, Cas3 is naturally fused to Cas2, which hints towards the direct association of Cas3 with the adaptation complex (Makarova et al., 2020). Thus, Cas3 being a nuclease, plays a crucial role in spacer generation. Mechanistic understanding of primed adaptation is still in the nascent stage. Several models have been proposed to explain the primed adaptation mechanism; however, a thorough understanding is yet to emerge on the underlying process (Semenova et al., 2016; Xue et al., 2015; Xue et al., 2016). In this chapter, we set out to study the molecular role of Cas3/I-C during primed adaptation in depth. In chapter 3, we have shown that CTD is essential for interference. Since interference and adaptation are linked, we found it interesting to understand the role of Cas3/I-

C CTD in primed adaptation. Lastly, we tested the PAM dependent transition of the Interference complex to the Adaptation complex.

5.2 Materials and Methods

5.2.1 Molecular cloning:

B. halodurans C-125 strain was procured from The Microbial Type Culture Collection and Gene Bank (MTCC), India. Gene encoding Cas3/I-C (BH0336) and its mutants (pD48A, pQ253A, pD395A, pK742A, pK743A, pQ745A, pQ746A, and pY747A) were cloned in 1R (Addgene #29664) at the SspI restriction site by employing the Gibson Assembly method. Point mutations were created in Cas3/I-C using PCR-based mutagenesis. The 1R vector encodes Strep II tagged protein. Genes encoding Cas4-1-2 complex as a single operon was cloned in 13S-R (Addgene #48328) at the SspI restriction site. 13S-R plasmid (CDF origin) is compatible with 1R (ColE1 origin), and these are used to co-express multiple proteins. Target sequence, complementary to crRNA sequence, with and without PAM, is inserted in pUC19 vector at HindIII and EcoRI restriction site.

5.2.2 Strain construction

E. coli IC-1 strain (Refer chapter 3, section 3.2.1) was used to integrate the CRISPR/I-C array using phage-186 integrase from pOSIP-KO (Addgene # 45895) to generate *E. coli* IC-3 strain. Please refer to section 3.2.1 for the detailed procedure.

5.2.3 Spacer acquisition assay

E. coli IC-3 strain was transformed with a plasmid containing Cas3/I-C (pWT/I-C) (Refer table 2) and a plasmid harbouring Cas4-1-2 (pCas412/I-C) and grown in LB broth containing 25 µg/ml kanamycin, 50 µg/ml spectinomycin, 25 µg/ml chloramphenicol, 0.2 % L-arabinose, and 0.02 mM IPTG at 37 °C until OD_{600} reached 0.3. Cells were harvested at 4 °C by centrifuging at 2700 × g. Harvested cells were made chemically competent using 0.1 M calcium chloride and transformed with target plasmid (pT-PAM/pT-mPAM, refer to table 3). Cells were grown overnight at 37 °C on LB broth supplemented with 25 µg/ml kanamycin, 50

µg/ml spectinomycin, 25 µg/ml chloramphenicol, 50 µg/ml ampicillin, 0.2 % L-arabinose, and 0.1 mM IPTG. Two cycles of growth and induction were performed for all the *in vivo* integration-related experiments unless mentioned otherwise. Cells were grown for 16 hours at 37 °C between each cycle, and 1 % inoculum from the previous cycle was transferred to fresh LB broth medium and growth was continued with the antibiotics and inducers mentioned above for 16 hours. Cells were harvested after two cycles unless mentioned otherwise. Cells were washed with water and then resuspended in PCR grade water. In order to take an equal amount of template for PCR, water volume was adjusted for each sample to maintain OD_{600} at 0.1. Array expansion was observed using PCR, and results were observed on 0.8 % agarose gel. DNA band intensities were measured using Image Lab Bio-Rad software, and relative spacer acquisition was calculated using the equation given below, where A′ is percentage relative acquisition and A is acquisition in presence of Cas3/I-C (wt),

$$\text{Aquisition (A)} = \frac{\text{Intensity of Expanded array (E)}}{\text{Intensity of Expanded array (E)} + \text{Intensity of Parent array(P)}}$$

$$A' = \left(\frac{E}{E+P}\right)/A * 100$$

5.2.4 Nuclease assay

For the procedure, please refer to section 4.2.4 in chapter 4. The concentration of Cascade/I-C, Cas4-1-2/I-C and Cas3/I-C was 1 µM each. Target DNA was incubated for 60 mins before visualizing on 20% denaturing Urea PAGE.

5.2.5 Fluorescence quenching assay

For a detailed method, please refer to section 4.2.7 in chapter 4. The concentration of Cascade/I-C, Cas4-1-2 and Cas3/I-C was 1 µM if mentioned otherwise.

5.3 Results

5.3.1 The presence of Cas3 enhances the frequency of spacer acquisition

Cas3 is indispensable for CRISPR interference, and in several subtypes, its association with the Adaptation complex makes it crucial for spacer uptake (Kunne et al., 2016; Li et al., 2014; Ramachandran and Bailey, 2016; Staals et al., 2016a). We set out to explore the importance of Cas3/I-C in primed adaptation using *E. coli* as a surrogate host. To test this, we used *E. coli* IC-3 strain (refer to table 2), which comprises of all the Cascade/I-C subunits (Cas5, Cas7, Cas8 and Cas11) and repeat-spacer units (CRISPR/I-C array) from *B. halodurans* integrated into a distant genomic locus using phage-186 integrase (Figure 5.1A). In order to negate the effect of intrinsic type I-E CRISPR-Cas machinery, Cascade/I-E was deleted using λ-red mediated homologous recombination. Next, we introduced a protospacer sequence abutting a cognate PAM (TTC) into a pUC19 plasmid, which acts as a target plasmid pT-PAM (Figure 5.1A). On the other hand, a mutated PAM (ATG or TAC) was inserted into pUC19, termed as pT-mPAM (Figure 5.1A). Components of the Adaptation complex (Cas1, Cas2 and Cas4) and Cas3/I-C (and its variants) were expressed using a compatible IPTG inducible expression vector (Figure 5.1A). Once all the plasmid except pT-PAM/pT-mPAM were transformed into *E. Coli* IC-3 strain, Cas proteins were nominally expressed using 20 μM IPTG. To this strain, target plasmids were transformed accordingly, and CRISPR array expansion was observed using a PCR based method (Figure 5.1B).

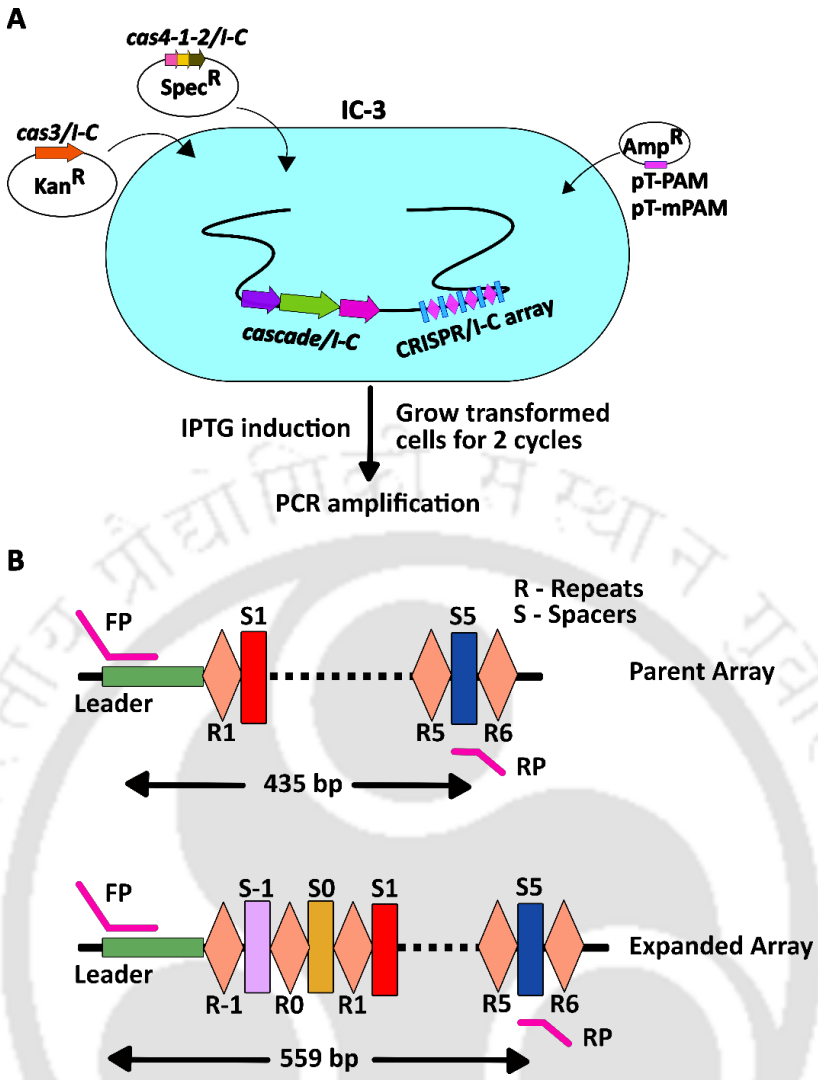

Figure 5. 1: Spacer acquisition assay.

(A). An E. coli IYB5101 strain harbouring genes encoding Cascade/I-C and CRISPR/I-C array (IC-3) was used as a surrogate for *in vivo* experiments. Cas3/I-C (Kanamycin) and Cas4-1-2/I-C (Spectinomycin) were expressed using IPTG inducible plasmids. The target sequence was inserted into the pUC19 vector with either cognate PAM (TTC, pT-PAM) or mutated PAM (TAC, ATG, pT-mPAM). After transforming the target plasmid, cells were grown at 37° C for two cycles and harvested. The expansion of the array was observed using PCR.

(B). Spacer acquisition was measured using PCR based array expansion assay. FP (Forward primer) and RP (Reverse primer) correspond to the leader region and the 5th spacer, respectively. After PCR, in the parent array, 435 bp fragment is observed, while an increased fragment size is seen in the expanded array. The bands were visualised using 0.8% agarose gel electrophoresis.

When *E. coli* IC-3 strain, harbouring Cas3/I-C (wt), Cascade/I-C and Cas4-1-2/I-C, was transformed with pT-mPAM, we could observe an increase in the frequency of spacer acquisition (Figure 5.2, lane 6). On the contrary, the spacer acquisition frequency was significantly less in the absence of Cas3/I-C, suggesting functional overlap between interference and adaptation machinery (Figure 5.2, lane 4). Surprisingly, we could observe array expansion even in the absence of the target plasmid, which could be attributed to the acquisition of spacers from already existing plasmids (Cas3/I-C and Cas4-1-2/I-C) through naïve adaptation (Figure 5.2, lane 5). As expected, Cas4, Cas1 and Cas2 are crucial for adaptation (Figure 5.2, lane 3 and 14).

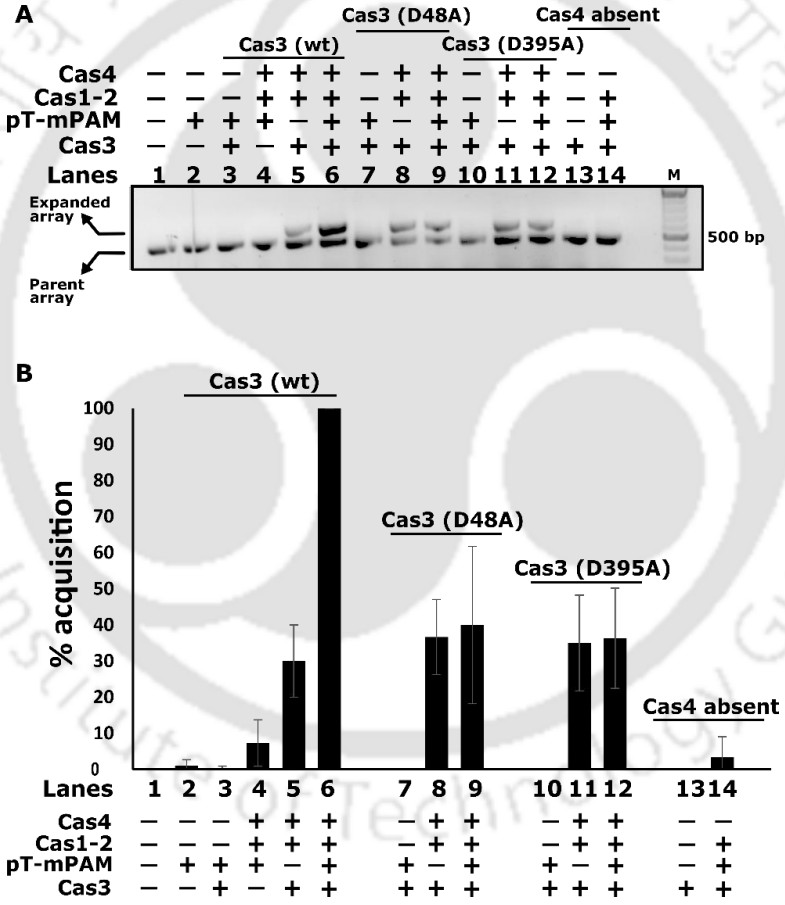

Figure 5. 2: Primed adaptation with Cas3/I-C variants.

(A). PCR products from a representative spacer acquisition assay (out of three independent trials) using *E. coli* IC-3 strain are shown in the above figure. The parent array corresponds to 435 bp fragment, and the expanded array corresponds to 559 bp. Array expansion was observed in the presence of Cas3/I-C (wt), Cas3/I-C D48A and Cas3/I-C D395A (Lanes 3-12). Cas4/I-C was removed in lane 14 to assess its importance in primed adaptation. M indicates the DNA marker. Samples were analysed using 0.8 % agarose gel.

(B). The figure shows the percentage acquisition measured from three independent trials, as shown in figure 5.2A. Band intensities of parent and expanded array were measured using Bio-Rad Image Lab software. The relative expansion was calculated using the equation in method section 5.2.3. Error bar represents standard deviation measured from three independent trials.

In type I-E, Cas3 associates with the Cas1-2 complex to form a Primed Adaptation Complex (PAC) (Redding et al., 2015). PAC is known to translocate on the target DNA, displacing protein roadblocks *en route*. However, it was not clear if the nuclease domain of Cas3/I-E cleaves the target DNA forming substrates for the Cas1-Cas2 complex. To test the importance of nuclease and helicase domains during primed adaptation, Cas3/I-C (WT) was replaced with its nuclease (D48A) (Figure 5.2, lane 7, 8 and 9) and helicase mutants (D395A) (Figure 5.2, lane 10, 11 and 12), successively, and array expansion was assessed. Interestingly, the frequency of acquisition was reduced, suggesting that the nuclease and helicase activities are required for the primed adaptation in type I-C (Figure 5.2).

5.3.2 PAM sequence acts as a gate between interference and adaptation

Having established the cooperation between interference and adaptation complexes in type I-C, we set out to study the effect of the alternate PAM sequence on Primed Adaptation. The protospacer sequences with three different PAM were inserted in the pUC19 plasmid to generate pT-PAM (TTC) and pT-mPAM (TAC and ATG), and these plasmids were transformed into *E. coli* IC-3 strain harbouring Cas3/I-C, Cascade/I-C and Cas4-1-2/I-C. We observed that when a non-target plasmid (without protospacer sequence) was transformed, the acquisition frequency was in line with the naïve adaptation process (Figure 5.3, lane 2). When TTC was used as PAM, the adaptation frequency was similar in the presence and absence of Cas3, suggesting that the acquisition frequency is similar to naïve adaptation. Interestingly, when the PAM was mutated to ATG or TAC, a significant increase in acquisition frequency was observed (Figure 5.3, lanes 8 and 11). In order to assess the fold difference between naïve and primed adaptation in type I-C, we used ATG PAM in the absence of Cas3 and cells were grown for four cycles instead of two cycles. We observed that even after four cycles, the adaptation frequency of naïve adaptation was significantly lower than the frequency observed during primed adaptation (Figure, lane 12). On the other hand, deviation from the cognate TTC PAM leads to increased spacer acquisition, suggesting that it is primed adaptation.

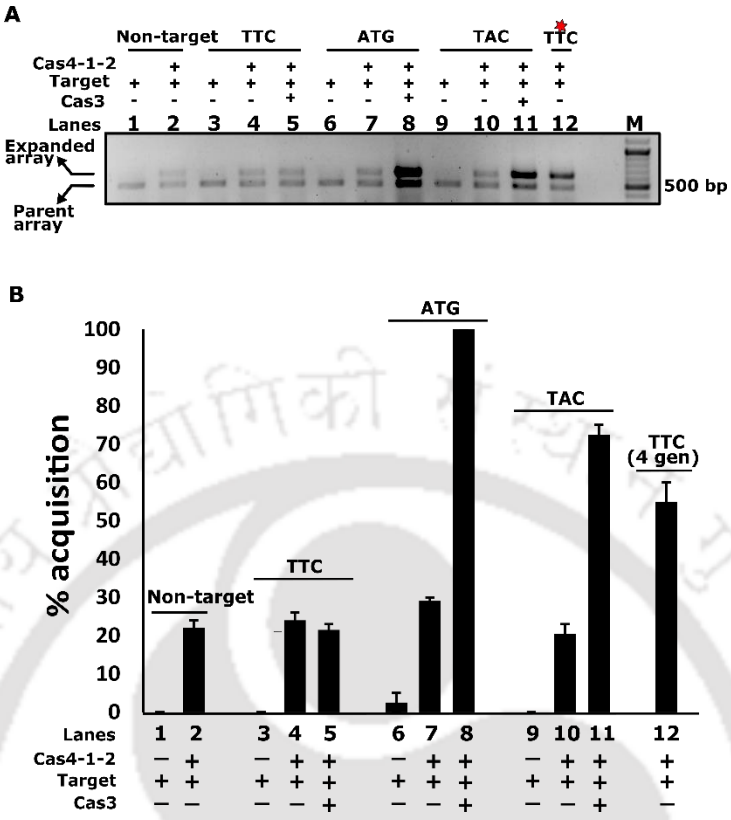

Figure 5. 3: Effect of PAM sequence on primed adaptation.

(A). *E. coli* IC-3 was transformed with pT-PAM (TTC) and pT-mPAM (TAC and ATG), and spacer acquisition was observed using PCR. The frequency of spacer acquisition for non-target plasmid (lanes 2) was similar to the frequency observed when Cas3/I-C is absent (lanes 4, 7 and 10), suggesting that is naïve adaptation. When TTC was used as PAM in the presence of Cas3/I-C, the frequency of acquisition was similar to naïve adaptation, whereas, when the PAM sequence was mutated to TAC or ATG, we could observe a significant rise in spacer acquisition. To compare naïve and primed adaptation, the target plasmid with ATG PAM was transformed in the absence of Cas3/I-C (lane 12), and cells were grown for four cycles instead of two cycles (marked as a red star). Surprisingly, even after four generations, the frequency of acquisition was lower than that of primed adaptation. Samples were analysed using 0.8 % agarose gel. This is a representative spacer acquisition assay out of three independent trials.

(B). The figure shows the percentage acquisition from three independent trials, as shown in the figure above. Error bar represents standard deviation measured from three independent trials. The band intensities were measured using Image Lab software and plotted as a bar graph.

Based on our experiments and previously reported observations (Semenova et al., 2016), we hypothesized that in the presence of non-cognate PAM (ATG and TAC), target DNA

is not rapidly cleaved by Cas3/I-C, providing adequate time for adaptation machinery to select and uptake spacers. To test this hypothesis, we used a 100 bp 6-FAM labelled (non-target strand) DNA construct, as shown in Figure 5.4A, with cognate 'TTC' and non-cognate 'ATG' PAM sequence. We observed cleavage after the addition of type I-C Cas proteins. When 'TTC' was used as PAM, the target DNA was degraded in the presence of Cas3/I-C and Cascade/I-C (Lanes 5, 7, 11 and 13 in Figure 5.4B), whereas the target DNA containing 'ATG' PAM sequence was comparatively intact (Lanes 5, 7, 11 and 13 in Figure 5.4C). Moreover, we could also observe multiple bands in the presence of Cas4-2-1/I-C (Lanes 4 and 10 in Figure 5.4B-C). It has been shown that Cas4, in association with Cas1-2, cleaves PAM containing prespacer, ensuring spacer fidelity (Lee et al., 2019; Lee et al., 2018). Presumably, the presence of multiple PAM sequences on the target DNA (Figure 5.4D) may favour nuclease cleavage by the Cas4-2-1 complex.

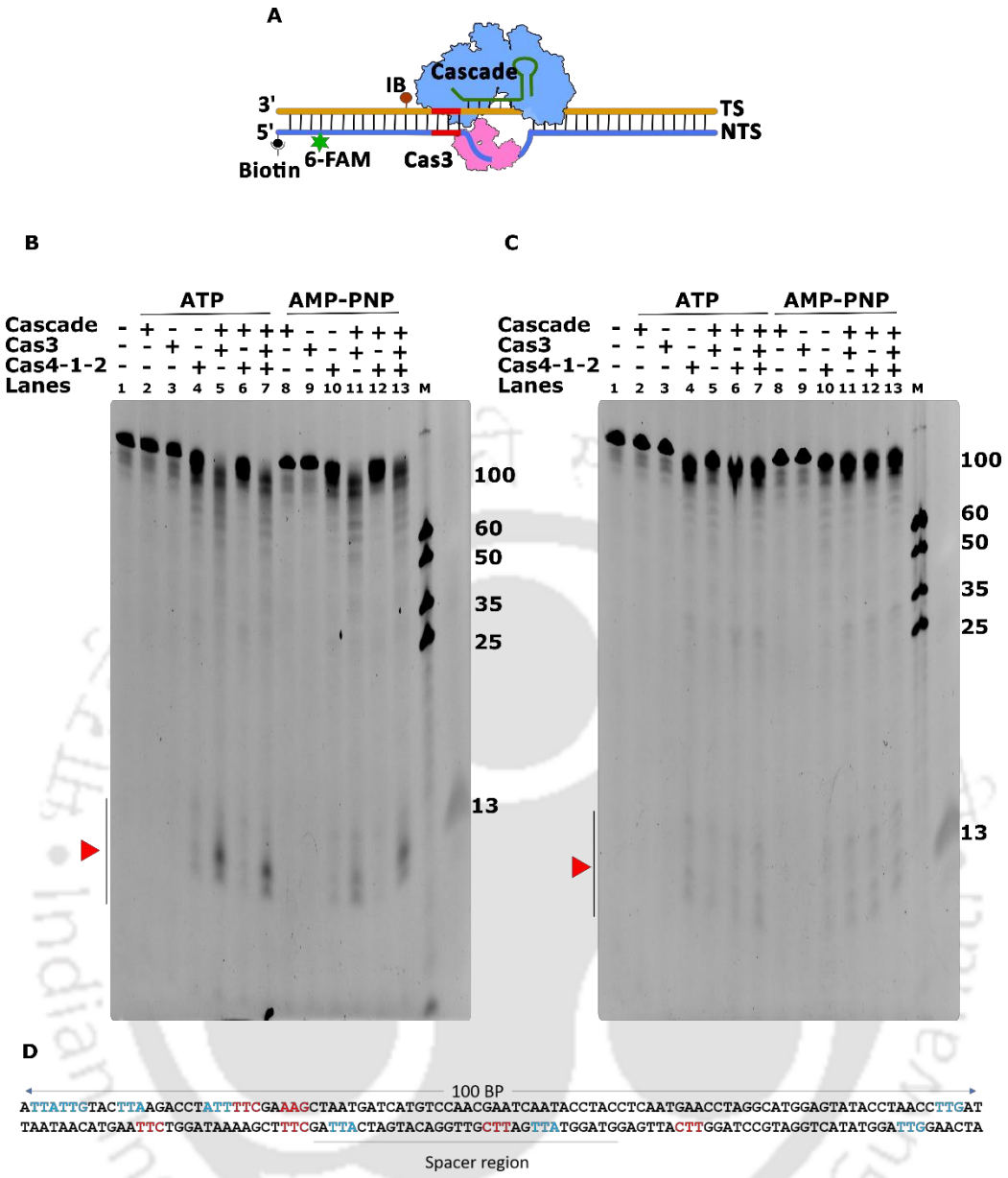

Figure 5. 4: Rapid degradation of target DNA by interference machinery.

(A). Schematic representation of 100 bp DNA substrates (also mentioned in Figure 4.4) with biotin (black dot) at 5′ end of the non-target strand (NTS), 6-FAM (shown as a green star) at 5th nucleotide, Iowa Black® FQ (shown as a brown dot) at 29th nucleotide on target strand (TS) and PAM sequence TTC (depicted in red).

(B). 100 bp DNA harbouring 'TTC' PAM sequence was incubated with Cas3/I-C, Cas4-1-2/I-C and Cascade/I-C for 60 mins in various combinations, and ATP/AMP-PNP was added to the reaction. The red triangle indicates cleavage products. M denotes 6-FAM labelled DNA markers.

(C). 100 bp DNA harbouring 'ATG' PAM sequence was incubated with Cas3/I-C, Cas4-1-2/I-C and Cascade/I-C for 60 mins in various combinations, and ATP/AMP-PNP was added to the reaction.

(D). The sequence of target DNA indicating the presence of PAM (red) or PAM-like (blue) regions triggering nuclease cleavage by the Cas4-1-2 complex.

Note: Samples were analysed using 20% denaturing 8M urea polyacrylamide gel.

5.3.3 Cas3/I-C CTD is essential for primed adaptation

Having established the importance of nuclease and helicase domains during primed adaptation, we asked if the Cas3 CTD is crucial for spacer acquisition. In chapter 3, we have shown that CTD is essential for interference and point mutations on the conserved residues of CTD lead to impaired target cleavage (Figure 3.2C and 3.2E). We transformed the *E. coli* IC-3 strain with Cas3/I-C CTD mutants along with the Cas4-2-1 complex. Apart from wild-type Cas3/I-C and Cas3/I-C K742A (Figure 5.5 lanes 3-4), we observed that the acquisition frequency for all other mutants (K743A, Q745A, Q746A, Y747A and ΔCTD) corresponds to naïve adaptation, suggesting the importance of CTD during adaptation (Figure 5.5, lanes 5-9).

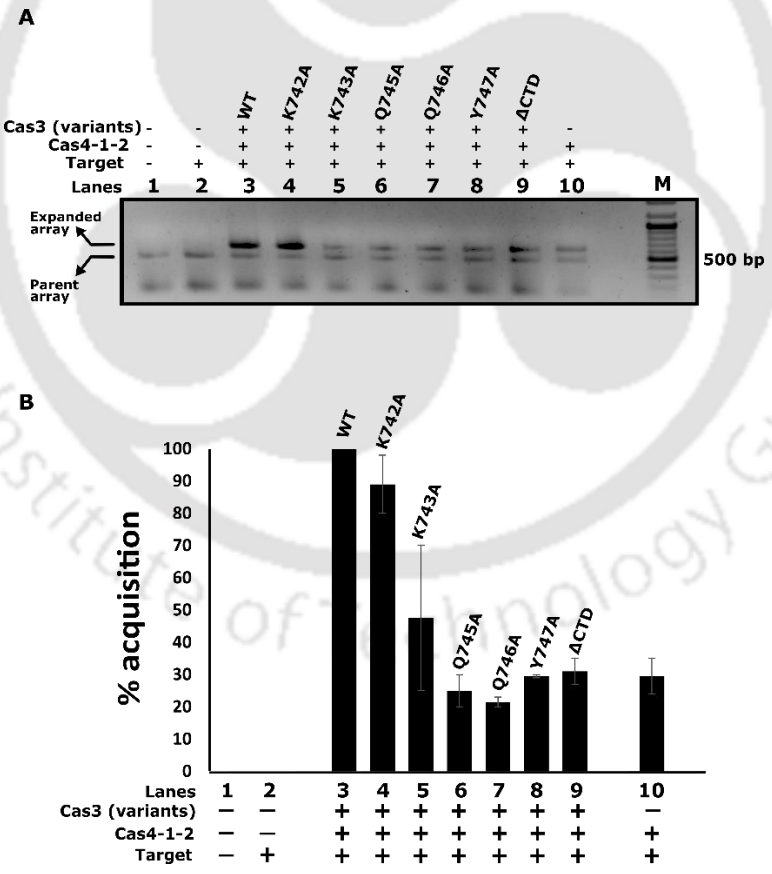

Figure 5. 5: Role of Cas3/I-C CTD in primed adaptation.

(A). Cas3/I-C (wt) and CTD mutants were tested for their involvement in primed adaptation. Except for Cas3/I-C K742A, other CTD mutants seem to be crucial for primed adaptation. This is a representative experiment out of three independent trials.

(B). The figure represents percentage acquisition from three independent trials, as shown in figure 5.4A. Error bar represents standard deviation measured from three independent trials

5.3.4 Minimal interaction between adaptation and interference complex drives primed adaptation.

There are two possible mechanisms of spacer uptake that have been observed in the type I-E system in *E. coli*. According to the first mechanism, in the type I-E system, the PAM containing fragments (~ 30-40 nt) generated as a result of nuclease degradation by Cas3 are uptaken by the Cas1-2 complex and integrated into the CRISPR locus (Kunne et al., 2016). This model may or may not require direct interaction of the Cas1-2 complex with the interference complex. On the contrary, another model in type I-E system suggests that mutations in PAM aborts interference and promotes forming a primed adaptation complex (PAC) containing adaptation proteins and Cas3 (Redding et al., 2015; Xue et al., 2016). Here, the adaptation complex along with Cas3 moves on the target DNA, selecting new spacers for integration (Redding et al., 2015). On the other hand, the adaptation complex in the type I-C system harbours Cas4, which is absent in the type I-E system. Hence, we asked which of the two possible mechanisms is followed in the type I-C system where the composition of the adaptation complex is different from the type I-E system. In order to check this, we employed 100 bp 6-FAM labelled target DNA containing mutated "ATG" PAM sequence (Figure 5.6A). As expected, when Cascade/I-C and Cas4-1-2/I-C were added, we could observe an increase in fluorescence anisotropy suggesting efficient binding, while Cas3/I-C could not bind to dsDNA (Figure 5.6B).

We first incubated the target DNA with Cascade, followed by other proteins (Figure 5.6C) and measured a change in fluorescence anisotropy. As expected, the addition of Cas3 showed an increase in anisotropy, suggesting efficient binding. On the other hand, the addition of the Cas4-2-1 complex exhibited no change in anisotropy, suggesting a lack of direct interaction between the Cascade and the Cas4-1-2 complex. Next, to test if the Cas4-1-2 complex interacts with Cas3, we first incubated Cascade-target DNA with Cas3 and then Cas4-2-1 was added in an increasing concentration. Surprisingly, we could not observe any significant change in fluorescence anisotropy, suggesting an apparent lack of interaction between the Cas3 and Cas4-1-2 complex. Additionally, we reversed the order of interaction by incubating Cascade bound target DNA with Cas4-1-2 complex followed by an increasing

concentration of Cas3 and observed an increase in fluorescence anisotropy. As mentioned above, there is no direct interaction between Cascade and Cas4-2-1, and hence this increase in fluorescence anisotropy is attributed to the binding of Cas3 to the Cascade complex. To further probe the molecular interaction, we first incubated the target DNA with the Cas4-1-2 complex and later added either Cascade or Cas3, followed by Cascade. Surprisingly we could not observe a significant change in fluorescence anisotropy, possibly due to the protection of the Cascade binding site by the Cas4-1-2 complex (Figure 5.6 D).

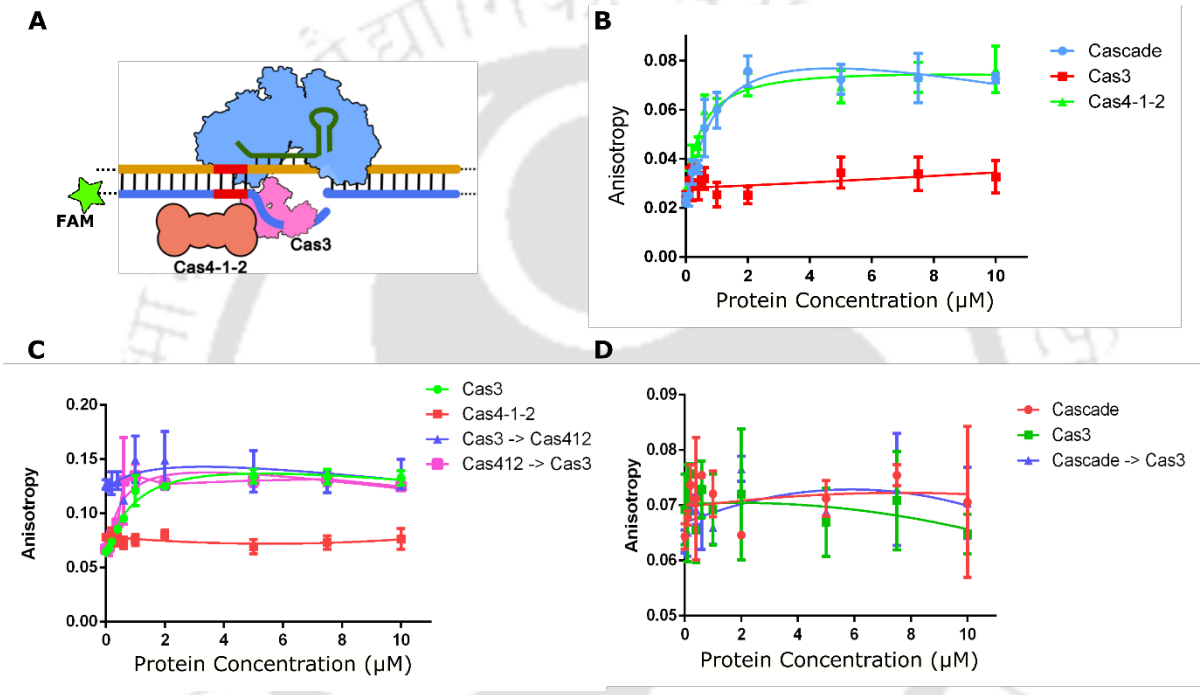

Figure 5. 6: Sequential binding of effector Cas proteins to target DNA.

(A). A schematic representation of 100 bp FAM labelled target DNA ('ATG' PAM), representing the probable interaction between Cas4-1-2/I-C and Cas3/I-C. The green star represents 6-FAM. Target Strand and non-target are depicted in yellow and blue, respectively. PAM is represented in Red colour.

(B). The target DNA mentioned above was incubated with an increasing concentration of Cascade/Cas3/Cas4-1-2 complex, and fluorescence anisotropy measurements were recorded.

(C). The target DNA was first incubated with 1 μM of Cascade, and Cas3/Cas4-1-2 was added later. The symbol '->' represents the sequential addition of a given protein to Cascade bound target DNA. Cas3 -> Cas4-1-2 represent the addition of 1 μM Cas3, followed by an increasing concentration of Cas4-1-2. Similarly, Cas412 -> Cas3 represents the addition of 1 μM Cas4-1-2, followed by an increasing concentration of Cas3.

(D). The target DNA was first incubated with 1 μM of Cas4-1-2, and Cas3/Cascade was added later. The symbol '->' represents the sequential addition of protein to Cas4-1-2 bound target DNA. Cascade -> Cas3 represents the addition of 1 μM Cascade, followed by an increasing concentration of Cas3.

Note: All the experiments were performed three times, and standard deviations are represented as bars. The x-axis represents the concentration (in μM) of indicated protein(s) or complex.

5.3.5 Cas4-1-2/I-C retards target DNA reeling by Cas3/I-C

In chapter 4, we have shown that Cas3/I-C degrades target DNA via the reeling mechanism (Figure 4.5). Therefore, we asked whether the transient association between Cas3/I-C and Cas4-1-2/I-C alters target DNA reeling. To address this, we used a 100 bp dsDNA with a fluorophore (6-FAM) and a quencher (Iowa Black FQ) (Figure 4.4 and 5.7A). Biotin was introduced at the 5′ end with the presumption that Cas3/I-C would halt on encountering the biotin-streptavidin block, thus sustaining 6-FAM quenching by Iowa Black FQ for a sufficiently long duration for measurements. Since 6-FAM and IB were far apart, the fluorescence of 6-FAM was not quenched. However, if Cas3/I-C reeled the DNA, 6-FAM would come in proximity with IB, which would quench the fluorescence. Target DNA was saturated with streptavidin and Cascade/I-C before adding Cas3/I-C, and fluorescence intensity in the presence of ATP was measured to estimate the extent of quenching. As expected, we could observe a significant fluorescence quenching when Cascade/I-C, followed by Cas3/I-C, was added to the reaction (purple curve in Figure 5.7B). Interestingly, we could observe only a slight decrease in fluorescence intensity when the Cas4-1-2 complex was present along with Cascade/I-C and Cas3/I-C (black and orange curve in Figure 5.7B). This reduction in fluorescence quenching may be either due to DNA cleavage by Cas3/I-C or due to blocking of Cas3/I-C movement. There was no quenching when only Cascade/Cas3/Cas4-1-2 or Cas4-1-2, followed by Cas3, was present. On the other hand, when ATP was not present, we could not observe any fluorescence quenching (Figure 5.7C).

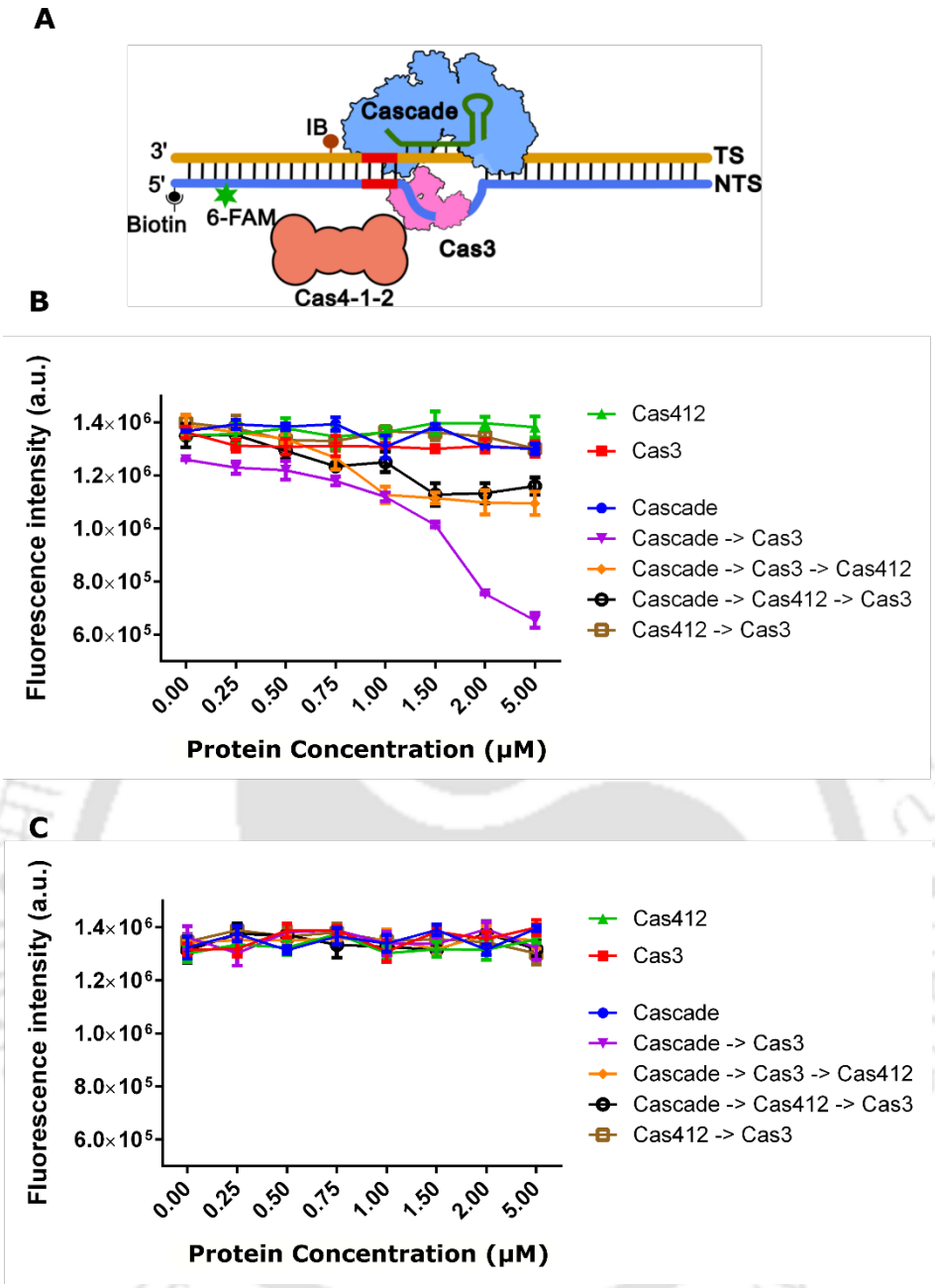

Figure 5. 7: DNA reeling in the presence of Cascade, Cas3 and Cas4-1-2.

(A). A schematic representation of 100 bp FAM, Iowa Black (IB) and Biotin labelled target DNA with 'TTC' PAM. TS – Target Strand, NTS – Non-target strand.

(B). The DNA substrate mentioned above was incubated only with Cascade/Cas3/Cas4-1-2 or Cascade, followed by Cas3 and Cas4-1-2, and fluorescence intensity was measured at 516 nm. ATP concentration used was 1 mM. The symbol '->' represents the sequential addition of proteins to Cascade/Cas4-1-2 bound target DNA. Cascade -> Cas3 -> Cas4-1-2 represent the addition of 1 μM Cascade, followed by 1 μM Cas3 and an increasing concentration of Cas4-1-2. Similarly, Cas412 -> Cas3 represents the addition of 1 μM Cas4-1-2, followed by an increasing concentration of Cas3.

(C). Fluorescence quenching was measured with the substrate mentioned in (b) in the absence of ATP.

5.4 Discussion

The crosstalk between interference and adaptation is known in several CRISPR types. In an ideal condition, when there is a perfect target-DNA:crRNA matching, the interference machinery diverts towards the target-degradation pathway. However, when the target degradation is delayed due to target-DNA:crRNA mismatching, it becomes essential to channel the resources to acquire new spacers using available defective interference machinery (Datsenko et al., 2012; Kunne et al., 2016; Nussenzweig et al., 2019; Semenova et al., 2016; Staals et al., 2016a; Xue et al., 2015). Subsequently, the newly acquired spacers can efficiently guide the interference complex during target degradation. To understand such a phenomenon in type I-C system, we have designed an *E. coli* surrogate system and utilized it to understand primed adaptation *in vivo*. In line with the previous reports (Datsenko et al., 2012; Kunne et al., 2016; Nussenzweig et al., 2019; Rao et al., 2017; Semenova et al., 2016; Staals et al., 2016a; Xue et al., 2015), the presence of Cas3/I-C assists in increasing the acquisition frequency through its interaction with the adaptation complex. Surprisingly, we could detect an expansion of the CRISPR array even in the absence of either Cas3 or the target DNA, which is likely due to spacer acquisition from already existing plasmids (Yosef et al., 2012). Since there is no priming of interference complex in the absence of Cas3 or target DNA, the spacer acquisition under such condition can happen through a naïve adaptation mechanism (RecBCD mediated acquisition in *E. coli*) (Ivancic-Bace et al., 2015; Levy et al., 2015).

Cas2/I-F and Cas3/I-F are fused in type I-F system, and the complex Cas2-3/I-F has long provided a link between interference and adaptation machinery (Fagerlund et al., 2017; Koonin et al., 2017a; Makarova et al., 2020). In Pectobacterium *atrosepticum* (type I-F), the Cas1/I-F, Cas2-Cas3/I-F assemble to form a ~ 400 kDa complex with four Cas1/I-F subunits and two Cas2-3/I-F subunits (Fagerlund et al., 2017). Interestingly, in this system, the integrase activity of Cas1/I-F was essential, and it was independent of the nuclease and helicase activity of Cas3/I-F (Fagerlund et al., 2017). Moreover, in the type I-F system in *P. aeruginosa*, Cas1/I-F also acts as a repressor of the nuclease activity of Cas2-3/I-F (Rollins et al., 2017). Our data on type I-C suggest that both the nuclease and helicase activities of Cas3/I-C are essential for primed adaptation. The available cryo-EM structure of the Cas4-1-2 complex from *B.*

halodurans shows an asymmetrically placed Cas4 subunit (Lee et al., 2019). Hence it is likely that the PAM proximal end of the protospacer may be trimmed by asymmetrically placed Cas4 subunit in Cas4-1-2 complex while the nuclease domain of Cas3/I-C processes the PAM distal end. Additionally, the deletion of Cas3/I-C CTD also leads to impaired spacer acquisition. Similarly, point mutations on the conserved amino acid residues on the CTD do not support primed adaptation, probably due to the destabilization of Cas3-DNA interaction (Figure 3.4).

In chapter 4, we have shown that Cas3/I-C remains attached to target-bound Cascade molecule and effectively reels DNA which was evident from fluorescence quenching (Figure 4.5). However, we observed that the fluorescence quenching was significantly reduced when Cas4-1-2/I-C was also present in the reaction. Additionally, we observed that Cas4-1-2/I-C cleaved the target DNA even in the absence of Cascade/I-C and Cas3/I-C (Lanes 4 and 10 in Figure 5.4B-C). Hence, the apparent decrease in fluorescence quenching in the presence of Cas4-1-2/I-C may be attributed to the cleavage of target DNA, forbidding reeling by Cas3/I-C.

Although primed adaptation has been established, the fundamental question of how the spacers are generated by cooperation between interference and adaptation complexes is poorly understood. Understanding how Cascade recruits either Cas3 for DNA degradation or Cas4-1-2 for spacer acquisition is likely to elucidate this mechanism. At present, there are two plausible models to explain this cooperation: (1) According to the first model, the mismatch between target-DNA and crRNA or mutated PAM induces an 'open' conformation in the Cas8 subunit of the Cascade complex (Blosser et al., 2015; Redding et al., 2015; Xue et al., 2015; Xue et al., 2016). This model explains the Cas1-2 dependent recruitment of Cas3 at the target location leading to spacer acquisition. (2) According to an alternate model, the PAM and 'seed' sequence-dependent conformational changes in the interference complex are not prerequisites for switching between interference and adaptation (Semenova et al., 2016). Here, both interference and adaptation are the results of the same target-DNA destruction pathway. When fully and partially matching targets were present on the same DNA molecule, it was observed that the frequency of acquisition for a fully matched target was at least ten times more than the mismatched target (Semenova et al., 2016). On the contrary, when matching and mismatching targets were on separate DNA molecule, the mismatched target DNA was a preferred substrate for spacer acquisition (Semenova et al., 2016). These discrepancies in the protospacer preference can be explained by considering the rate of target degradation. When both the targets were on the same DNA, the rate of target degradation was the same for both matched and mismatched targets that led to similar spacer acquisition. When matched and mismatched

targets were present on separate DNA molecule, the target DNA was cleaved at a higher rate, making it difficult for adaptation machinery to harvest spacers while the mismatched target was available for sufficient time, allowing easy spacer selection. In this study, we have used both cognate as well as mutated PAM sequences. The spacer acquisition pathway is favoured when the PAM region is mutated, which is in line with the previous reports. Based on previous reports on *Legionella pnuemophila* and *E. coli* (Rao et al., 2017; Semenova et al., 2011), the spacer uptake mechanism in *B. halodurans* is likely to follow the second model of primed adaptation; however, since the primed adaptation mechanism shows differences within the same CRISPR subtype, this will prime the future studies to resolve the basis for such differences.

5.5 Summary:

Most of our understandings about primed adaptation belong to type I-E and I-F systems and some sporadic pieces of information from type I-B and I-C. In this chapter, we have shown that the presence of Cas3/I-C enhances the frequency of spacer acquisition by functioning in cooperation with Cascade and adaptation complex. We showed that, unlike the type I-F system, the nuclease and helicase activities of Cas3/I-C are required most likely for translocation during spacer selection and uptake from the target DNA and processing of the PAM distal end. Deletion and mutation in Cas3/I-C CTD destabilize the interaction of Cas3 with DNA, and thus, the spacer acquisition mechanism follows RecBCD mediated naïve adaptation. Apart from protein components, PAM sequence mismatch is also responsible for the apparent increase in spacer acquisition frequency. We have also shown that the interaction between type I-C adaptation and interference proteins seems to be transient.

Chapter 6: Conclusion, Future Directions and Applications

6.1 Conclusion

The phrase "survival of the fittest" is the hallmark of the Darwinian theory of evolution. This statement applies to all living organisms, including microscopic bacteria and phages. Phages/MGEs that are devoid of cytoplasm hijack bacterial system to replicate and propagate. As protection against invading phages, bacteria have developed and regularly improved several defence systems, of which the CRISPR-Cas system is the only identified adaptive immune system in prokaryotes (Bolotin et al., 2005; Jansen et al., 2002; Mojica et al., 2005; Pourcel et al., 2005). The CRISPR-Cas system acts against phages and other MGEs using the sequence specificity imparted by the crRNA present within the interference complex. As mentioned earlier, the CRISPR-Cas systems are highly diverse and yet perform a similar function, i.e. degradation of invading MGE (Koonin et al., 2017a; Makarova et al., 2020). The interference stage, where the invading nucleic acid molecule is cleaved, is mechanistically similar to RNA interference in eukaryotes and is the hallmark of the CRISPR-Cas system. Interestingly, the proteins involved in the interference stage also seem to govern the adaptation and the maturation stage, thus making interference unique. For example, Cas6 or Cas5 (type I-C) is a constituent of the interference complex that helps in the trimming and maturation of crRNA (Nam et al., 2012; Punetha et al., 2014). Similarly, Cas3 and Cascade interact with the adaptation complex to increase spacer acquisition frequency (Dillard et al., 2018; Rao et al., 2017; Richter et al., 2014; Xue et al., 2016). The present work has attempted to unravel the molecular mechanism of CRISPR interference in a minimalistic type I-C CRISPR-Cas system and shows the interplay between interference and adaptation. Focusing on understanding the functioning of the effector nuclease, we characterized Cas3/I-C in detail. Like Cas3/I-E, Cas3/I-C is also a metal-dependent ssDNA specific nuclease that unwinds duplex DNA in a 3' to 5' direction utilizing energy from ATP; however, these activities are independent of DNA sequence. The cooperation between nuclease and helicase domains was evident when the addition of ATP stimulated the nuclease cleavage of plasmid, and the addition of DNA increases the rate of ATP hydrolysis by the helicase domain. However, the knowledge about the kinetics of Cas3 loading, translocation, and degradation rate is poorly understood and therefore, it will be interesting to probe these mechanistic aspects.

We further tested the activities and interplay between nuclease and helicase domains *in vivo*, using an *E. coli* (IC-1 strain) surrogate system. This strain of *E. coli* contained a spacer for targeting pUC19 plasmid with the corresponding protospacer. The region harbouring

Cascade proteins on intrinsic type I-E CRISPR-Cas locus was deleted using λ-red recombinase to avoid activity overlap between type I-E and integrated type I-C system. Since *E. coli* is a model bacterium, having abundantly available genetic manipulation tools, the development of a surrogate system facilitated the study of type I-C system in an *in vivo* set-up. Additionally, the *E. coli* IC-1 strain was used to comprehend the contribution of Cas3 CTD during target cleavage. Interestingly, the sequence analysis of Cas3 CTD revealed the presence of conserved amino acids in type I-C, whereas we could not observe any conservation in type I-E, and hence we set out to understand the importance of these conserved residues. Our experimental data suggested that residues K742, K743, Q745, Q746 and Y747 are crucial for interference; however, the exact involvement of these residues in interference was not apparent initially. Our unsuccessful attempts to crystallize Cas3/I-C led us to adopt homology modelling as an alternative approach. Based on the crystal structure Cas3/I-E structure, Cas3/I-C was modelled, which suggested that the conserved residues in CTD bolts Cas3/I-C on target DNA and thus crucial for stabilization of Cas3-DNA interaction. Additionally, the Cryo-EM structure of the Cas3/I-E in complex with Cascade/I-E suggested that Cas3/I-E interacts with Cascade/I-E via a set of four interfaces (Xiao et al., 2018), and CTD seems to not play any role in the interaction (Xiao et al., 2018); however, since the structure represents one snapshot of the dynamic interactions, the possibility of Cas3-Cascade interaction in type I-C system via CTD cannot be ruled out. Future investigations that attempt to determine the high-resolution structure of type I-C interference complex will significantly advance our understanding of molecular interactions in this system.

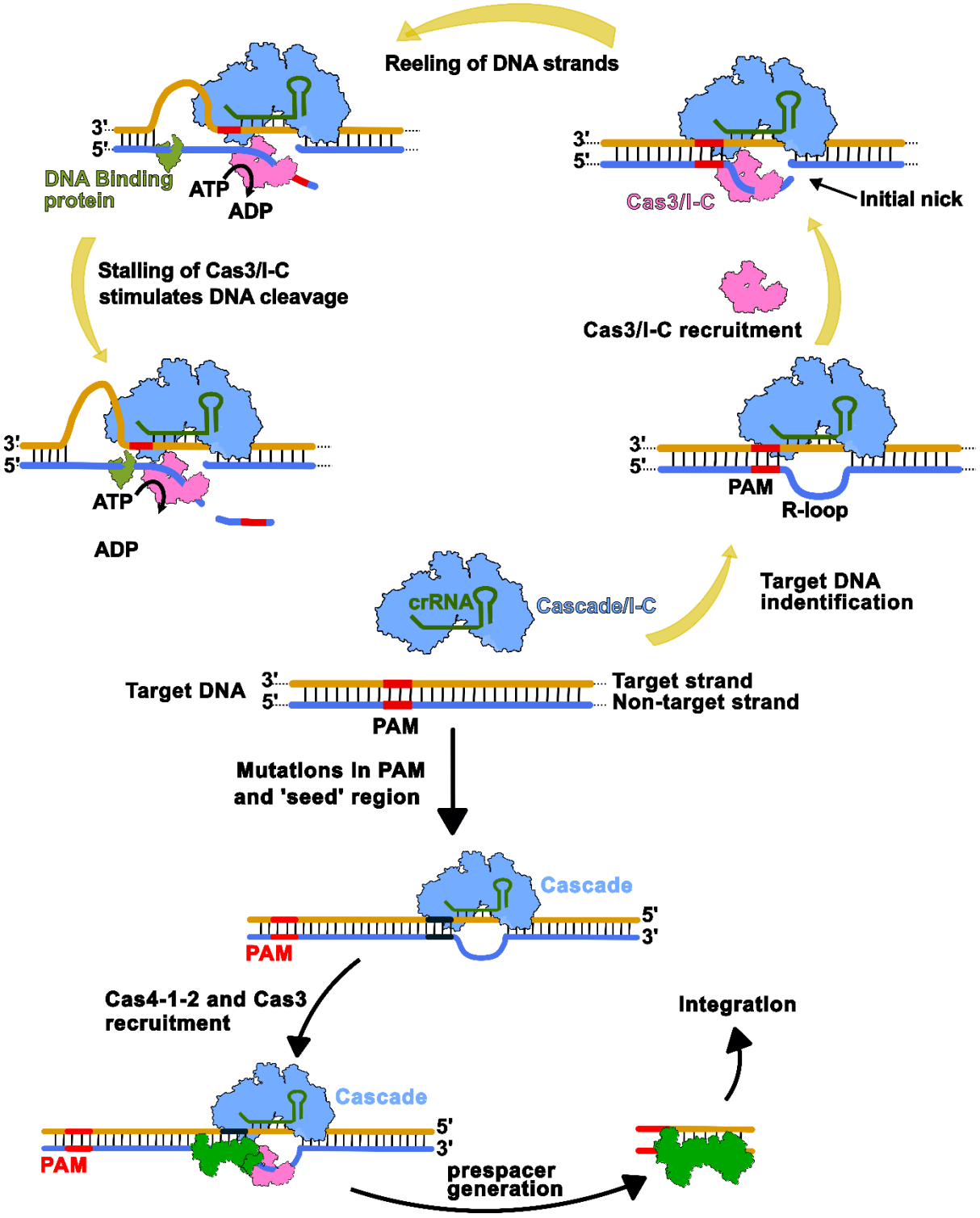

Figure 6. 1: A schematic model representing the possible scenarios leading to Cas3/I-C mediated target DNA cleavage or primed adaptation.

Cascade/I-C (shown in blue) through base pairing between crRNA (indicated as stem and loop structure in green) and target DNA locates the invading foreign genetic element, thereby forming an R-loop. The ensuing conformational change in the large subunit of Cascade/I-C facilitates its binding to Cas3/I-C. This is followed by nucleotide independent single-stranded nick on the non-target strand by HD domain. This promotes the loading of the helicase domain

of Cas3/I-C onto the ssDNA of the non-target strand. Based on the longevity of Cascade/I-C–Cas3/I-C interaction, we propose that Cas3/I-C remains attached to Cascade/I-C as a stable complex and reels in the non-target strand (this leads to DNA looping on the target strand), utilizing energy from ATP. Subsequently, Cas3/I-C may hit a roadblock (e.g. a DNA-binding protein), which can trigger further DNA cleavage. Here, the cleavage site is likely to be in proximity to the Cascade/I-C. On the contrary, the mutation in the PAM or 'seed' region may lead to an alternate Cascade/I-C confirmation, which recruits Cas4-1-2/I-C complex (coloured green) and Cas3/I-C. Spacers may be acquired from an upstream cognate PAM containing a portion of DNA leading to primed adaptation.

Having characterized Cas3/I-C, we tried to understand the molecular events that occur during target cleavage. Interestingly, we found that the single-stranded region (displaced non-target strand, Figure 6.1 and 4.2) formed as a result of R-loop formation by Cascade: target-DNA binding is sufficient for Cas3/I-C to load and degrade DNA in the presence of ATP; however, ATP is not required for Cas3/I-C loading. However, these findings do not rebut the importance of Cas3-Cascade interaction during interference. The interaction between Cas3 and Cascade is believed to direct the nuclease activity of Cas3 towards a specific target. The interaction between Cas3 and Cascade can be weak or strong depending on the CRISPR type and the species of bacteria/archaea. For example, this interaction is weak in *E. coli* but is observed to be very strong in *T. fusca* (Loeff et al., 2018). Interestingly both the *E. coli* and *T. fusca* harbour type I-E system and yet exhibit these differences. While studying the prevalence of Cas3-Cascade interaction in type I-C system during interference, we observed that the contact between Cas3 and Cascade in *B. halodurans* is strong, and thus we observed that type I-C follows the reeling mechanism of target degradation similar to *T. fusca* (Loeff et al., 2018). It will be interesting to see if a similar mechanism is followed in other CRISPR types, including the type I-C system in other bacteria. During reeling, Cas3/I-C may encounter protein roadblocks that can impede its translocation. The stalling of the movement of Cas3/I-C on the target DNA stimulates the nuclease cleavage, which is in line with Cas3/I-E from *T. fusca* (Dillard et al., 2018). We presume that the stalling of Cas3/I-C movement provides enough residence time for the nuclease domain to cleave the DNA.

The identification and degradation of target DNA rely on the sequence of the protospacer. However, phages and MGEs are known to mutate rapidly, which can lead to impaired interference. The CRISPR system compensates for the mutated target by acquiring more spacers abutting the Cascade priming region. Using *E. coli* as a surrogate system, we have

shown that all the three domains of Cas3/I-C – nuclease, helicase and CTD – are essential for primed adaptation where PAM acts as a switch between interference and adaptation pathway. Overall, the significant outcomes of the thesis underline the crucial events that occur during the target cleavage and cross-talk between interference and the adaptation stage.

6.1.1 Future scope and application:

In the present study, we have attempted to uncover the molecular events leading to the target recognition and cleavage mechanism in the type I-C CRISPR-Cas system. However, several questions remain open. At present, the type I-C CRISPR-Cas system lacks high-resolution structural information, and hence, it is challenging to study its functional mechanism at atomic details. Determining molecular structures of proteins in the type I-C system would help answer questions related to the domain conformation, protein interactions, and PAM identification. On the other hand, we have shown that primed adaptation is prevalent in the type I-C system. However, the precise molecular mechanism is yet not known. It will be exciting to understand the cross-talk between interference and adaptation despite differences in CRISPR architecture.

The current research on type-I CRISPR-Cas system unlocks several opportunities in the development of novel applications in the field of therapeutics, bio-processing, medical diagnosis and many more. Class 2 CRISPR proteins (Cas9, Cas12 or Cas13) have shown their potential as a genetic manipulation tool recently and are already being used for genome editing. However, it has also been found that the efficiency of these proteins is not significant and the usage of these proteins leads often leads to off-site targeting. Although type I systems work in a similar mechanism, their use was limited due to the presence of a multi-subunit effector complex, unlike a single subunit in Cas9/Cas12/Cas13. However, recently, harnessing type I systems for genome editing has been shown to be effective in prokaryotes (Hidalgo-Cantabrana et al., 2019; Xu et al., 2021). The spacer length in the type I system varies from a minimum of ~25 nt to as long as ~36 nt, which is significantly higher than that of Cas9 (~20 nt). A longer spacer length can potentially decrease the off-site effect, increasing the overall efficiency of the system.

References

Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., and Block, S.M. (2005). Direct observation of base-pair stepping by RNA polymerase. Nature *438*, 460-465.

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L.*, et al.* (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science *353*, aaf5573.

Almendros, C., Nobrega, F.L., McKenzie, R.E., and Brouns, S.J.J. (2019). Cas4-Cas1 fusions drive efficient PAM selection and control CRISPR adaptation. Nucleic Acids Res *47*, 5223-5230.

Amitai, G., and Sorek, R. (2016). CRISPR–Cas adaptation: insights into the mechanism of action. Nature Reviews Microbiology *14*, 67-76.

Anantharaman, V., Makarova, K.S., Burroughs, A.M., Koonin, E.V., and Aravind, L. (2013). Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. Biol Direct *8*, 15.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569-573.

Andersson, A.F., and Banfield, J.F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. Science *320*, 1047-1050.

Aravind, L., Watanabe, H., Lipman, D.J., and Koonin, E.V. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proc Natl Acad Sci U S A *97*, 11319-11324.

Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, U. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. Nucleic Acids Res *42*, 7884-7893.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A.*, et al.* (2011). A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol Microbiol *79*, 484-502.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Barrangou, R., and Horvath, P. (2017). A decade of discovery: CRISPR functions and applications. Nat Microbiol *2*, 17092.

Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., and Yakunin, A.F. (2011). Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. EMBO J *30*, 4616-4627.

Bickle, T.A., and Kruger, D.H. (1993). Biology of DNA restriction. Microbiol Rev *57*, 434-450.

Bikard, D., Hatoum-Aslan, A., Mucida, D., and Marraffini, L.A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. Cell Host Microbe *12*, 177-186.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res *41*, 7429-7437.

Blosser, T.R., Loeff, L., Westra, E.R., Vlot, M., Kunne, T., Sobota, M., Dekker, C., Brouns, S.J.J., and Joo, C. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. Mol Cell *58*, 60-70.

Blower, T.R., Pei, X.Y., Short, F.L., Fineran, P.C., Humphreys, D.P., Luisi, B.F., and Salmond, G.P. (2011). A processed noncoding RNA regulates an altruistic bacterial antiviral system. Nat Struct Mol Biol *18*, 185-190.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J *17*, 170-180.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551-2561.

Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M.F., Hidalgo-Reyes, Y., Wiedenheft, B., Maxwell, K.L., and Davidson, A.R. (2015). Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature *526*, 136-139.

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., and Davidson, A.R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature *493*, 429-432.

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

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

Burstein, D., Sun, C.L., Brown, C.T., Sharon, I., Anantharaman, K., Probst, A.J., Thomas, B.C., and Banfield, J.F. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. Nat Commun *7*, 10613.

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

Buttner, K., Nehring, S., and Hopfner, K.P. (2007). Structural basis for DNA duplex separation by a superfamily-2 helicase. Nat Struct Mol Biol *14*, 647-652.

Cagan, R.L., and Ready, D.F. (1989). Notch is required for successive cell decisions in the developing Drosophila retina. Genes Dev *3*, 1099-1112.

Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev *22*, 3489-3496.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642-655.

Castillo, J.A., Secaira-Morocho, H., Maldonado, S., and Sarmiento, K.N. (2020). Diversity and Evolutionary Dynamics of Antiphage Defense Systems in Ralstonia solanacearum Species Complex. Front Microbiol *11*, 961.

Charpentier, E., Richter, H., van der Oost, J., and White, M.F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. FEMS Microbiol Rev *39*, 428-441.

Chen, J.S., Ma, E., Harrington, L.B., Da Costa, M., Tian, X., Palefsky, J.M., and Doudna, J.A. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science *360*, 436-439.

Chopin, M.C., Chopin, A., and Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. Curr Opin Microbiol *8*, 473-479.

Chowdhury, S., Carter, J., Rollins, M.F., Golden, S.M., Jackson, R.N., Hoffmann, C., Nosaka, L., Bondy-Denomy, J., Maxwell, K.L., Davidson, A.R.*, et al.* (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. Cell *169*, 47-57 e11.

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., and Zhang, F. (2017). RNA editing with CRISPR-Cas13. Science *358*, 1019-1027.

Cumby, N., Edwards, A.M., Davidson, A.R., and Maxwell, K.L. (2012). The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. J Bacteriol *194*, 5012-5019.

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

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun *3*, 945.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Deng, L., Garrett, R.A., Shah, S.A., Peng, X., and She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. Mol Microbiol *87*, 1088-1099.

Depardieu, F., Didier, J.P., Bernheim, A., Sherlock, A., Molina, H., Duclos, B., and Bikard, D. (2016). A Eukaryotic-like Serine/Threonine Kinase Protects Staphylococci against Phages. Cell Host Microbe *20*, 471-481.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol *190*, 1390-1400.

Díez-Villaseñor, C., Guzmán, N.M., Almendros, C., García-Martínez, J., and Mojica, F.J.M. (2013). CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. RNA Biology *10*, 792-802.

Dillard, K.E., Brown, M.W., Johnson, N.V., Xiao, Y., Dolan, A., Hernandez, E., Dahlhauser, S.D., Kim, Y., Myler, L.R., Anslyn, E.V., *et al.* (2018). Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. Cell *175*, 934-946 e915.

Dillingham, M.S., and Kowalczykowski, S.C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol Mol Biol Rev *72*, 642-671, Table of Contents.

Dong, Guo, M., Wang, S., Zhu, Y., Wang, S., Xiong, Z., Yang, J., Xu, Z., and Huang, Z. (2017). Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. Nature *546*, 436-439.

Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., *et al.* (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. Nature *532*, 522-526.

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., and Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. Science *359*.

Drabavicius, G., Sinkunas, T., Silanskas, A., Gasiunas, G., Venclovas, C., and Siksnys, V. (2018). DnaQ exonuclease-like domain of Cas2 promotes spacer integration in a type I-E CRISPR-Cas system. EMBO Rep *19*.

Dugar, G., Herbig, A., Forstner, K.U., Heidrich, N., Reinhardt, R., Nieselt, K., and Sharma, C.M. (2013). High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. PLoS Genet *9*, e1003495.

Durmaz, E., and Klaenhammer, T.R. (2007). Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected Lactococcus lactis. J Bacteriol *189*, 1417-1425.

Dy, R.L., Richter, C., Salmond, G.P., and Fineran, P.C. (2014). Remarkable Mechanisms in Microbes to Resist Phage Infections. Annu Rev Virol *1*, 307-331.

East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H., Tjian, R., and Doudna, J.A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature *538*, 270-273.

Edgar, R., and Qimron, U. (2010). The Escherichia coli CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. J Bacteriol *192*, 6291-6294.

Elmore, J., Deighan, T., Westpheling, J., Terns, R.M., and Terns, M.P. (2015). DNA targeting by the type I-G and type I-A CRISPR-Cas systems of Pyrococcus furiosus. Nucleic Acids Res *43*, 10353-10363.

Elmore, J.R., Sheppard, N.F., Ramia, N., Deighan, T., Li, H., Terns, R.M., and Terns, M.P. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. Genes Dev *30*, 447-459.

Engelberg-Kulka, H., Hazan, R., and Amitai, S. (2005). mazEF: a chromosomal toxin-antitoxin module that triggers programmed cell death in bacteria. J Cell Sci *118*, 4327-4332.

Erdmann, S., and Garrett, R.A. (2012). Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Molecular Microbiology *85*, 1044-1056.

Ershova, A.S., Karyagina, A.S., Vasiliev, M.O., Lyashchuk, A.M., Lunin, V.G., Spirin, S.A., and Alexeevski, A.V. (2012). Solitary restriction endonucleases in prokaryotic genomes. Nucleic Acids Res *40*, 10107-10115.

Estrella, M.A., Kuo, F.T., and Bailey, S. (2016). RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. Genes Dev *30*, 460-470.

Fagerlund, R.D., Wilkinson, M.E., Klykov, O., Barendregt, A., Pearce, F.G., Kieper, S.N., Maxwell, H.W.R., Capolupo, A., Heck, A.J.R., Krause, K.L.*, et al.* (2017). Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. Proc Natl Acad Sci U S A *114*, E5122-E5128.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science *281*, 237.

Filipowicz, W. (2005). RNAi: the nuts and bolts of the RISC machine. Cell *122*, 17-20.

Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S., and Salmond, G.P. (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc Natl Acad Sci U S A *106*, 894-899.

Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H., and Brouns, S.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proc Natl Acad Sci U S A *111*, E1629-1638.

Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature *532*, 517-521.

Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. Virus Research *117*, 5-16.

Gao, P., Yang, H., Rajashankar, K.R., Huang, Z., and Patel, D.J. (2016). Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. Cell Res *26*, 901-913.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.

Garside, E.L., Schellenberg, M.J., Gesner, E.M., Bonanno, J.B., Sauder, J.M., Burley, S.K., Almo, S.C., Mehta, G., and MacMillan, A.M. (2012). Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. RNA *18*, 2020-2028.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci USA *109*, E2579-E2586.

Gasiunas, G., Sinkunas, T., and Siksnys, V. (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. Cell Mol Life Sci *71*, 449-465.

Gesner, E.M., Schellenberg, M.J., Garside, E.L., George, M.M., and Macmillan, A.M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. Nat Struct Mol Biol *18*, 688-692.

Gleditzsch, D., Muller-Esparza, H., Pausch, P., Sharma, K., Dwarakanath, S., Urlaub, H., Bange, G., and Randau, L. (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. Nucleic Acids Res *44*, 5872-5882.

Goldberg, G.W., Jiang, W., Bikard, D., and Marraffini, L.A. (2014). Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. Nature *514*, 633-637.

Gong, B., Shin, M., Sun, J., Jung, C.H., Bolt, E.L., van der Oost, J., and Kim, J.S. (2014). Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. Proc Natl Acad Sci USA *111*, 16359-16364.

Goren, M.G., Yosef, I., Auster, O., and Qimron, U. (2012). Experimental definition of a clustered regularly interspaced short palindromic duplicon in Escherichia coli. J Mol Biol *423*, 14-16.

Goryanin, II, Kudryavtseva, A.A., Balabanov, V.P., Biryukova, V.S., Manukhov, I.V., and Zavilgelsky, G.B. (2018). Antirestriction activities of KlcA (RP4) and ArdB (R64) proteins. FEMS Microbiol Lett *365*.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics *8*, 172.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol *1*, e60.

Hale, C., Kleppe, K., Terns, R.M., and Terns, M.P. (2008). Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. Rna *14*, 2572-2579.

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M.*, et al.* (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Mol Cell *45*, 292-302.

Hammond, S.M. (2005). Dicing and slicing: the core machinery of the RNA interference pathway. FEBS Lett *579*, 5822-5829.

Hampton, H.G., Watson, B.N.J., and Fineran, P.C. (2020). The arms race between bacteria and their phage foes. Nature *577*, 327-336.

Harrington, L.B., Burstein, D., Chen, J.S., Paez-Espino, D., Ma, E., Witte, I.P., Cofsky, J.C., Kyrpides, N.C., Banfield, J.F., and Doudna, J.A. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science *362*, 839-842.

Harrington, L.B., Ma, E., Chen, J.S., Witte, I.P., Gertz, D., Paez-Espino, D., Al-Shayeb, B., Kyrpides, N.C., Burstein, D., Banfield, J.F.*, et al.* (2020). A scoutRNA Is Required for Some Type V CRISPR-Cas Systems. Mol Cell *79*, 416-424 e415.

Hatoum-Aslan, A., Maniv, I., and Marraffini, L.A. (2011). Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. Proc Natl Acad Sci U S A *108*, 21218-21222.

Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W., and Marraffini, L.A. (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. J Biol Chem *288*, 27888-27897.

Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J.A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science *329*, 1355-1358.

Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B., Rajashankar, K., Bailey, S., Wiedenheft, B., and Ke, A. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. Nature *530*, 499-503.

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature *519*, 199-202.

Heler, R., Wright, A.V., Vucelja, M., Bikard, D., Doudna, J.A., and Marraffini, L.A. (2017). Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR-Cas Immune Response. Mol Cell *65*, 168-175.

Hidalgo-Cantabrana, C., Goh, Y.J., and Barrangou, R. (2019). Characterization and Repurposing of Type I and Type II CRISPR-Cas Systems in Bacteria. J Mol Biol *431*, 21-33.

Hille, F., and Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. Philos Trans R Soc Lond B Biol Sci *371*.

Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F.*, et al.* (2016). Structure and Engineering of Francisella novicida Cas9. Cell *164*, 950-961.

Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc Natl Acad Sci U S A *111*, 6618-6623.

Hochstrasser, M.L., Taylor, D.W., Kornfeld, J.E., Nogales, E., and Doudna, J.A. (2016). DNA Targeting by a Minimal CRISPR RNA-Guided Cascade. Mol Cell *63*, 840-851.

Hooton, S.P., and Connerton, I.F. (2014). Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. Front Microbiol *5*, 744.

Hoyland-Kroghsbo, N.M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-Denomy, J., and Bassler, B.L. (2017). Quorum sensing controls the Pseudomonas aeruginosa CRISPR-Cas adaptive immune system. Proc Natl Acad Sci U S A *114*, 131-135.

Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M.D., Jr., Zhou, S., Rajashankar, K., Kurinov, I.*, et al.* (2014a). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nat Struct Mol Biol *21*, 771-777.

Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M.D., Zhou, S., Rajashankar, K., Kurinov, I.*, et al.* (2014b). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nature Structural and Molecular Biology *21*, 771-777.

Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol *9*, 22-32.

Hynes, A.P., Villion, M., and Moineau, S. (2014). Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. Nat Commun *5*, 4399.

Ishino, Y., Krupovic, M., and Forterre, P. (2018). History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. J Bacteriol *200*.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol *169*, 5429-5433.

Ivančić-Baće, I., Al Howard, J., and Bolt, E.L. (2012). Tuning in to interference: R-loops and cascade complexes in CRISPR immunity, pp. 607-616.

Ivancic-Bace, I., Cass, S.D., Wearne, S.J., and Bolt, E.L. (2015). Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity. Nucleic Acids Res *43*, 10821-10830.

Ivancic-Bace, I., Radovcic, M., Bockor, L., Howard, J.L., and Bolt, E.L. (2013). Cas3 stimulates runaway replication of a ColE1 plasmid in Escherichia coli and antagonises RNaseHI. RNA Biol *10*, 770-778.

Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014a). Structural

biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science *345*, 1473-1479.

Jackson, R.N., Lavin, M., Carter, J., and Wiedenheft, B. (2014b). Fitting CRISPR-associated Cas3 into the helicase family tree. Curr Opin Struct Biol *24*, 106-114.

Jackson, R.N., and Wiedenheft, B. (2015). A Conserved Structural Chassis for Mounting Versatile CRISPR RNA-Guided Immune Responses. Mol Cell *58*, 722-728.

Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol *43*, 1565-1575.

Jia, N., Jones, R., Yang, G., Ouerfelli, O., and Patel, D.J. (2019). CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise cA4 Cleavage with ApA>p Formation Terminating RNase Activity. Mol Cell *75*, 944-956 e946.

Jiang, F., and Doudna, J.A. (2015). The structural biology of CRISPR-Cas systems, pp. 100-111.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science *351*, 867-871.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol *31*, 233-239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012a). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012b). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (New York, NY) *337*, 816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. eLife *2*, e00471-e00471.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S.*, et al.* (2014a). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science *343*, 1247997.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S.*, et al.* (2014b). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science (New York, NY) *343*, 1247997-1247997.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R.*, et al.* (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol *18*, 529-536.

Karow, A.R., and Klostermeier, D. (2010). A structural model for the DEAD box helicase YxiN in solution: localization of the RNA binding domain. J Mol Biol *402*, 629-637.

Karvelis, T., Bigelyte, G., Young, J.K., Hou, Z., Zedaveinyte, R., Budre, K., Paulraj, S., Djukanovic, V., Gasior, S., Silanskas, A.*, et al.* (2020). PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage. Nucleic Acids Res *48*, 5016-5023.

Kazlauskiene, M., Kostiuk, G., Venclovas, C., Tamulaitis, G., and Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science *357*, 605-609.

Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, C., and Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. Mol Cell *62*, 295-306.

Kieper, S.N., Almendros, C., Behler, J., McKenzie, R.E., Nobrega, F.L., Haagsma, A.C., Vink, J.N.A., Hess, W.R., and Brouns, S.J.J. (2018). Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. Cell Rep *22*, 3377-3384.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A *93*, 1156-1160.

Kim, Y.G., and Chandrasegaran, S. (1994). Chimeric restriction endonuclease. Proc Natl Acad Sci U S A *91*, 883-887.

Kim, Y.G., Shi, Y., Berg, J.M., and Chandrasegaran, S. (1997). Site-specific cleavage of DNA-RNA hybrids by zinc finger/FokI cleavage domain fusions. Gene *203*, 43-49.

Klompe, S.E., and Sternberg, S.H. (2018). Harnessing "A Billion Years of Experimentation": The Ongoing Exploration and Exploitation of CRISPR-Cas Immune Systems. CRISPR J *1*, 141-158.

Knott, G.J., East-Seletsky, A., Cofsky, J.C., Holton, J.M., Charles, E., O'Connell, M.R., and Doudna, J.A. (2017). Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme. Nat Struct Mol Biol *24*, 825-833.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420-424.

Komor, A.C., Zhao, K.T., Packer, M.S., Gaudelli, N.M., Waterbury, A.L., Koblan, L.W., Kim, Y.B., Badran, A.H., and Liu, D.R. (2017). Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. Sci Adv *3*, eaao4774.

Koo, Y., Ka, D., Kim, E.J., Suh, N., and Bae, E. (2013). Conservation and variability in the structure and function of the Cas5d endoribonuclease in the CRISPR-mediated microbial immune system. J Mol Biol *425*, 3799-3810.

Koonin, E.V., Makarova, K.S., and Zhang, F. (2017a). Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol *37*, 67-78.

Koonin, E.V., Makarova, K.S., and Zhang, F. (2017b). Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol *37*, 67-78.

Kruger, D.H., and Bickle, T.A. (1983). Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. Microbiol Rev *47*, 345-360.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol *8*, R61.

Kunne, T., Kieper, S.N., Bannenberg, J.W., Vogel, A.I., Miellet, W.R., Klein, M., Depken, M., Suarez-Diez, M., and Brouns, S.J. (2016). Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. Mol Cell *63*, 852-864.

Kuznedelov, K., Mekler, V., Lemak, S., Tokmina-Lukaszewska, M., Datsenko, K.A., Jain, I., Savitskaya, E., Mallon, J., Shmakov, S., Bothner, B., *et al.* (2016). Altered stoichiometry Escherichia coli Cascade complexes with shortened CRISPR RNA spacers are capable of interference and primed adaptation. Nucleic Acids Res *44*, 10849-10861.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nat Rev Microbiol *8*, 317-327.

Lee, H., Dhingra, Y., and Sashital, D.G. (2019). The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. Elife *8*.

Lee, H., Zhou, Y., Taylor, D.W., and Sashital, D.G. (2018). Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. Mol Cell *70*, 48-59 e45.

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.

Li, M., Wang, R., Zhao, D., and Xiang, H. (2014). Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process. Nucleic Acids Res *42*, 2483-2492.

Lillestol, R.K., Shah, S.A., Brugger, K., Redder, P., Phan, H., Christiansen, J., and Garrett, R.A. (2009). CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. Mol Microbiol *72*, 259-272.

Liu, L., Chen, P., Wang, M., Li, X., Wang, J., Yin, M., and Wang, Y. (2017a). C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism. Mol Cell *65*, 310-322.

Liu, M., Deora, R., Doulatov, S.R., Gingery, M., Eiserling, F.A., Preston, A., Maskell, D.J., Simons, R.W., Cotter, P.A., Parkhill, J.*, et al.* (2002). Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. Science *295*, 2091-2094.

Liu, T., Liu, Z., Ye, Q., Pan, S., Wang, X., Li, Y., Peng, W., Liang, Y., She, Q., and Peng, N. (2017b). Coupling transcriptional activation of CRISPR-Cas system and DNA repair genes by Csa3a in Sulfolobus islandicus. Nucleic Acids Res *45*, 8978-8992.

Loeff, L., Brouns, S.J.J., and Joo, C. (2018). Repetitive DNA Reeling by the Cascade-Cas3 Complex in Nucleotide Unwinding Steps. Mol Cell *70*, 385-394 e383.

Loenen, W.A., and Raleigh, E.A. (2014). The other face of restriction: modification-dependent enzymes. Nucleic Acids Res *42*, 56-69.

Luo, D., Xu, T., Hunke, C., Gruber, G., Vasudevan, S.G., and Lescar, J. (2008). Crystal structure of the NS3 protease-helicase from dengue virus. J Virol *82*, 173-183.

Luo, M.L., Jackson, R.N., Denny, S.R., Tokmina-Lukaszewska, M., Maksimchuk, K.R., Lin, W., Bothner, B., Wiedenheft, B., and Beisel, C.L. (2016). The CRISPR RNA-guided surveillance complex in Escherichia coli accommodates extended RNA spacers. Nucleic Acids Res *44*, 7385-7394.

Majsec, K., Bolt, E.L., and Ivancic-Bace, I. (2016). Cas3 is a limiting factor for CRISPR-Cas immunity in Escherichia coli cells lacking H-NS. BMC Microbiol *16*, 28.

Majumdar, S., and Terns, M.P. (2019). CRISPR RNA-guided DNA cleavage by reconstituted Type I-A immune effector complexes. Extremophiles *23*, 19-33.

Majumdar, S., Zhao, P., Pfister, N.T., Compton, M., Olson, S., Glover, C.V., 3rd, Wells, L., Graveley, B.R., Terns, R.M., and Terns, M.P. (2015). Three CRISPR-Cas immune effector complexes coexist in Pyrococcus furiosus. Rna *21*, 1147-1158.

Makarova, K.S., Aravind, L., Grishin, N.V., Rogozin, I.B., and Koonin, E.V. (2002). A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Res *30*, 482-496.

Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E.V. (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol Direct *6*, 38.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct *1*, 7.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F.*, et al.* (2011b). Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol *9*, 467-477.

Makarova, K.S., and Koonin, E.V. (2015). Annotation and Classification of CRISPR-Cas Systems. Methods in molecular biology *1311*, 47-75.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H.*, et al.* (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol *13*, 722-736.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P.*, et al.* (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol *18*, 67-83.

Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2018). Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? CRISPR J *1*, 325-336.

Makarova, K.S., Wolf, Y.I., van der Oost, J., and Koonin, E.V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct *4*, 29.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843-1845.

Marraffini, L.A., and Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568-571.

McBride, T.M., Schwartz, E.A., Kumar, A., Taylor, D.W., Fineran, P.C., and Fagerlund, R.D. (2020). Diverse CRISPR-Cas Complexes Require Independent Translation of Small and Large Subunits from a Single Gene. Mol Cell *80*, 971-979 e977.

McGinn, J., and Marraffini, L.A. (2016). CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. Mol Cell *64*, 616-623.

Medina-Aparicio, L., Rebollar-Flores, J.E., Gallego-Hernandez, A.L., Vazquez, A., Olvera, L., Gutierrez-Rios, R.M., Calva, E., and Hernandez-Lucas, I. (2011). The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in Salmonella enterica serovar Typhi. J Bacteriol *193*, 2396-2407.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. (2005). Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev *19*, 2837-2848.

Modell, J.W., Jiang, W., and Marraffini, L.A. (2017). CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. Nature *544*, 101-104.

Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., and van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science *353*, aad5147.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733-740.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol *60*, 174-182.

Mojica, F.J., Ferrer, C., Juez, G., and Rodriguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax

mediterranei and Haloferax volcanii and could be involved in replicon partitioning. Mol Microbiol *17*, 85-93.

Molina, R., Stella, S., Feng, M., Sofos, N., Jauniskis, V., Pozdnyakova, I., Lopez-Mendez, B., She, Q., and Montoya, G. (2019). Structure of Csx1-cOA4 complex reveals the basis of RNA decay in Type III-B CRISPR-Cas. Nat Commun *10*, 4302.

Moreno-Mateos, M.A., Fernandez, J.P., Rouet, R., Vejnar, C.E., Lane, M.A., Mis, E., Khokha, M.K., Doudna, J.A., and Giraldez, A.J. (2017). CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. Nat Commun *8*, 2024.

Mulepati, S., and Bailey, S. (2011). Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). J Biol Chem *286*, 31896-31903.

Mulepati, S., and Bailey, S. (2013). In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. J Biol Chem *288*, 22184-22192.

Mulepati, S., Heroux, A., and Bailey, S. (2014). Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science *345*, 1479-1484.

Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M.P., and Ke, A. (2012). Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. Structure *20*, 1574-1584.

Niewoehner, O., Garcia-Doval, C., Rostol, J.T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L.A., and Jinek, M. (2017). Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. Nature *548*, 543-548.

Nishimasu, H., Cong, L., Yan, W.X., Ran, F.A., Zetsche, B., Li, Y., Kurabayashi, A., Ishitani, R., Zhang, F., and Nureki, O. (2015). Crystal Structure of Staphylococcus aureus Cas9. Cell *162*, 1113-1126.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935-949.

Nunez, J.K., Bai, L., Harrington, L.B., Hinder, T.L., and Doudna, J.A. (2016). CRISPR Immunological Memory Requires a Host Factor for Specificity. Mol Cell *62*, 824-833.

Nunez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N., and Doudna, J.A. (2015a). Foreign DNA capture during CRISPR-Cas adaptive immunity. Nature *527*, 535-538.

Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. Nat Struct Mol Biol *21*, 528-534.

Nunez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015b). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature *519*, 193-198.

Nussenzweig, P.M., McGinn, J., and Marraffini, L.A. (2019). Cas9 Cleavage of Viral Genomes Primes the Acquisition of New Immunological Memories. Cell Host Microbe *26*, 515-526 e516.

O'Brien, R.E., Santos, I.C., Wrapp, D., Bravo, J.P.K., Schwartz, E.A., Brodbelt, J.S., and Taylor, D.W. (2020). Structural basis for assembly of non-canonical small subunits into type I-C Cascade. Nat Commun *11*, 5931.

O'Connell, M.R. (2019). Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR-Cas Systems. J Mol Biol *431*, 66-87.

Osawa, T., Inanaga, H., and Numata, T. (2013). Crystal structure of the Cmr2-Cmr3 subcomplex in the CRISPR-Cas RNA silencing effector complex. J Mol Biol *425*, 3811-3823.

Osawa, T., Inanaga, H., Sato, C., and Numata, T. (2015). Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog. Mol Cell *58*, 418-430.

Ozcan, A., Pausch, P., Linden, A., Wulf, A., Schuhle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G., and Randau, L. (2019). Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. Nat Microbiol *4*, 89-96.

Paez-Espino, D., Morovic, W., Sun, C.L., Thomas, B.C., Ueda, K.-i., Stahl, B., Barrangou, R., and Banfield, J.F. (2013). Strong bias in the bacterial CRISPR elements that confer immunity to phage. Nature Communications *4*, 1430.

Page, R., and Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. Nat Chem Biol *12*, 208-214.

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.

Patterson, A.G., Jackson, S.A., Taylor, C., Evans, G.B., Salmond, G.P.C., Przybilski, R., Staals, R.H.J., and Fineran, P.C. (2016). Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems. Mol Cell *64*, 1102-1108.

Pausch, P., Muller-Esparza, H., Gleditzsch, D., Altegoer, F., Randau, L., and Bange, G. (2017). Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. Mol Cell *67*, 622-632 e624.

Pawluk, A., Bondy-Denomy, J., Cheung, V.H., Maxwell, K.L., and Davidson, A.R. (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa. mBio *5*, e00896.

Pawluk, A., Staals, R.H., Taylor, C., Watson, B.N., Saha, S., Fineran, P.C., Maxwell, K.L., and Davidson, A.R. (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. Nat Microbiol *1*, 16085.

Peng, R., Xu, Y., Zhu, T., Li, N., Qi, J., Chai, Y., Wu, M., Zhang, X., Shi, Y., Wang, P., *et al.* (2017). Alternate binding modes of anti-CRISPR viral suppressors AcrF1/2 to Csy surveillance complex revealed by cryo-EM structures. Cell Res *27*, 853-864.

Peng, W., Feng, M., Feng, X., Liang, Y.X., and She, Q. (2015). An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference. Nucleic Acids Res *43*, 406-417.

Pingoud, A., Wilson, G.G., and Wende, W. (2016). Type II restriction endonucleases - a historical perspective and more. Nucleic Acids Res *44*, 8011.

Pinilla-Redondo, R., Mayo-Munoz, D., Russel, J., Garrett, R.A., Randau, L., Sorensen, S.J., and Shah, S.A. (2020). Type IV CRISPR-Cas systems are highly diverse and involved in competition between plasmids. Nucleic Acids Res *48*, 2000-2012.

Plagens, A., Tjaden, B., Hagemann, A., Randau, L., and Hensel, R. (2012). Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon Thermoproteus tenax. J Bacteriol *194*, 2491-2500.

Plagens, A., Tripp, V., Daume, M., Sharma, K., Klingl, A., Hrle, A., Conti, E., Urlaub, H., and Randau, L. (2014). In vitro assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. Nucleic Acids Res *42*, 5125-5138.

Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L., and Severinov, K. (2010). Transcription, processing and function of CRISPR cassettes in Escherichia coli. Mol Microbiol *77*, 1367-1379.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology *151*, 653-663.

Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., and Wagner, R. (2010). Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol *75*, 1495-1512.

Punetha, A., Sivathanu, R., and Anand, B. (2014). Active site plasticity enables metal-dependent tuning of Cas5d nuclease activity in CRISPR-Cas type I-C system. Nucleic Acids Res *42*, 3846-3856.

Ramachandran, A., and Bailey, S. (2016). Memory Upgrade: Insights into Primed Adaptation by CRISPR-Cas Immune Systems. Mol Cell *64*, 641-642.

Ramachandran, A., Summerville, L., Learn, B.A., DeBell, L., and Bailey, S. (2020). Processing and integration of functionally oriented prespacers in the Escherichia coli CRISPR system depends on bacterial host exonucleases. J Biol Chem *295*, 3403-3414.

Ramalingam, S., Annaluru, N., Kandavelou, K., and Chandrasegaran, S. (2014). TALEN-mediated generation and genetic correction of disease-specific human induced pluripotent stem cells. Curr Gene Ther *14*, 461-472.

Rao, C., Chin, D., and Ensminger, A.W. (2017). Priming in a permissive type I-C CRISPR-Cas system reveals distinct dynamics of spacer acquisition and loss. Rna *23*, 1525-1538.

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

Rauch, B.J., Silvis, M.R., Hultquist, J.F., Waters, C.S., McGregor, M.J., Krogan, N.J., and Bondy-Denomy, J. (2017). Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell *168*, 150-158 e110.

Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B., Doudna, J.A., and Greene, E.C. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. Cell *163*, 854-865.

Reeks, J., Graham, S., Anderson, L., Liu, H., White, M.F., and Naismith, J.H. (2013). Structure of the archaeal Cascade subunit Csa5: relating the small subunits of CRISPR effector complexes. RNA Biol *10*, 762-769.

Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H., and Fineran, P.C. (2014). Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res *42*, 8516-8526.

Richter, H., Lange, S.J., Backofen, R., and Randau, L. (2013). Comparative analysis of Cas6b processing and CRISPR RNA stability. RNA Biol *10*, 700-707.

Richter, H., Zoephel, J., Schermuly, J., Maticzka, D., Backofen, R., and Randau, L. (2012). Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. Nucleic Acids Res *40*, 9887-9896.

Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S., Dryden, D.T., Dybvig, K.*, et al.* (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res *31*, 1805-1812.

Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2007). REBASE--enzymes and genes for DNA restriction and modification. Nucleic Acids Res *35*, D269-270.

Rollie, C., Graham, S., Rouillon, C., and White, M.F. (2018). Prespacer processing and specific integration in a Type I-A CRISPR system. Nucleic Acids Res *46*, 1007-1020.

Rollie, C., Schneider, S., Brinkmann, A.S., Bolt, E.L., and White, M.F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. Elife *4*.

Rollins, M.F., Chowdhury, S., Carter, J., Golden, S.M., Wilkinson, R.A., Bondy-Denomy, J., Lander, G.C., and Wiedenheft, B. (2017). Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. Proc Natl Acad Sci U S A *114*, E5113-E5121.

Rollins, M.F., Schuman, J.T., Paulus, K., Bukhari, H.S., and Wiedenheft, B. (2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. Nucleic Acids Res *43*, 2216-2222.

Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V., and Seidel, R. (2015). Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. Cell Rep *10*, 1534-1543.

Safari, F., Zare, K., Negahdaripour, M., Barekati-Mowahed, M., and Ghasemi, Y. (2019). CRISPR Cpf1 proteins: structure, function and implications for genome editing. Cell Biosci *9*, 36.

Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A., and Marraffini, L.A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. Cell *161*, 1164-1174.

Samson, J.E., Magadan, A.H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. Nat Rev Microbiol *11*, 675-687.

Sashital, D.G., Jinek, M., and Doudna, J.A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nat Struct Mol Biol *18*, 680-687.

Sashital, D.G., Wiedenheft, B., and Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol Cell *46*, 606-615.

Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., and Severinov, K. (2013). High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. RNA Biol *10*, 716-725.

Scadden, A.D. (2005). The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. Nat Struct Mol Biol *12*, 489-496.

Scholl, D., Adhya, S., and Merril, C. (2005). Escherichia coli K1's capsule is a barrier to bacteriophage T7. Appl Environ Microbiol *71*, 4872-4874.

Seed, K.D., Lazinski, D.W., Calderwood, S.B., and Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature *494*, 489-491.

Sefcikova, J., Roth, M., Yu, G., and Li, H. (2017). Cas6 processes tight and relaxed repeat RNA via multiple mechanisms: A hypothesis. Bioessays *39*.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci U S A *108*, 10098-10103.

Semenova, E., Savitskaya, E., Musharova, O., Strotskaya, A., Vorontsova, D., Datsenko, K.A., Logacheva, M.D., and Severinov, K. (2016). Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex. Proc Natl Acad Sci U S A *113*, 7626-7631.

Shao, Y., and Li, H. (2013). Recognition and cleavage of a nonstructured CRISPR RNA by its processing endoribonuclease Cas6. Structure *21*, 385-393.

Shao, Y., Richter, H., Sun, S., Sharma, K., Urlaub, H., Randau, L., and Li, H. (2016). A Non-Stem-Loop CRISPR RNA Is Processed by Dual Binding Cas6. Structure *24*, 547-554.

Sheng, G., Zhao, H., Wang, J., Rao, Y., Tian, W., Swarts, D.C., van der Oost, J., Patel, D.J., and Wang, Y. (2014). Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. Proc Natl Acad Sci U S A *111*, 652-657.

Shin, J., Jiang, F., Liu, J.J., Bray, N.L., Rauch, B.J., Baik, S.H., Nogales, E., Bondy-Denomy, J., Corn, J.E., and Doudna, J.A. (2017). Disabling Cas9 by an anti-CRISPR DNA mimic. Sci Adv *3*, e1701620.

Shipman, S.L., Nivala, J., Macklis, J.D., and Church, G.M. (2016). Molecular recordings by directed CRISPR spacer acquisition. Science *353*, aaf1175.

Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K.*, et al.* (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol Cell *60*, 385-397.

Shmakov, S., Savitskaya, E., Semenova, E., Logacheva, M.D., Datsenko, K.A., and Severinov, K. (2014). Pervasive generation of oppositely oriented spacers during CRISPR adaptation. Nucleic Acids Res *42*, 5907-5916.

Shmakov, S.A., Sitnik, V., Makarova, K.S., Wolf, Y.I., Severinov, K.V., and Koonin, E.V. (2017). The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes. mBio *8*, e01397-01317.

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J *30*, 1335-1342.

Sinkunas, T., Gasiunas, G., and Siksnys, V. (2015). Cas3 nuclease-helicase activity assays. Methods Mol Biol *1311*, 277-291.

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. EMBO J *32*, 385-394.

Smargon, A.A., Cox, D.B.T., Pyzocha, N.K., Zheng, K., Slaymaker, I.M., Gootenberg, J.S., Abudayyeh, O.A., Essletzbichler, P., Shmakov, S., Makarova, K.S., *et al.* (2017). Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. Mol Cell *65*, 618-630 e617.

Sontheimer, E.J. (2005). Assembly and function of RNA silencing complexes. Nat Rev Mol Cell Biol *6*, 127-138.

Staals, R.H., Jackson, S.A., Biswas, A., Brouns, S.J., Brown, C.M., and Fineran, P.C. (2016a). Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. Nat Commun *7*, 12853.

Staals, R.H., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K., *et al.* (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. Mol Cell *56*, 518-530.

Staals, R.H.J., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K., *et al.* (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. Mol Cell *52*, 135-145.

Staals, R.H.J., Jackson, S.A., Biswas, A., Brouns, S.J.J., Brown, C.M., and Fineran, P.C. (2016b). Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR–Cas system. Nature Communications *7*, 12853.

Stern, A., Keren, L., Wurtzel, O., Amitai, G., and Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? Trends Genet *26*, 335-340.

Sternberg, S.H., LaFrance, B., Kaplan, M., and Doudna, J.A. (2015). Conformational control of DNA target cleavage by CRISPR-Cas9. Nature *527*, 110-113.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62-67.

Sternberg, Samuel H., Richter, H., Charpentier, E., and Qimron, U. (2016). Adaptation in CRISPR-Cas Systems. Molecular Cell *61*, 797-808.

Swarts, D.C., Hegge, J.W., Hinojo, I., Shiimori, M., Ellis, M.A., Dumrongkulraksa, J., Terns, R.M., Terns, M.P., and van der Oost, J. (2015). Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA. Nucleic Acids Res *43*, 5120-5129.

Swarts, D.C., Jore, M.M., Westra, E.R., Zhu, Y., Janssen, J.H., Snijders, A.P., Wang, Y., Patel, D.J., Berenguer, J., Brouns, S.J.J., *et al.* (2014). DNA-guided DNA interference by a prokaryotic Argonaute. Nature *507*, 258-261.

Swarts, D.C., Mosterd, C., van Passel, M.W., and Brouns, S.J. (2012). CRISPR interference directs strand specific spacer acquisition. PLoS One *7*, e35888.

Swinton, D., Hattman, S., Crain, P.F., Cheng, C.S., Smith, D.L., and McCloskey, J.A. (1983). Purification and characterization of the unusual deoxynucleoside, alpha-N-(9-beta-D-2'-deoxyribofuranosylpurin-6-yl)glycinamide, specified by the phage Mu modification function. Proc Natl Acad Sci U S A *80*, 7400-7404.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc Natl Acad Sci USA *111*, 9798-9803.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123-132.

Talansky, B.E., and Gordon, J.W. (1988). Cleavage characteristics of mouse embryos inseminated and cultured after zona pellucida drilling. Gamete Res *21*, 277-287.

Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, C., Nwokeoji, A.O., Dickman, M.J., Horvath, P., and Siksnys, V. (2014). Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus. Mol Cell *56*, 506-517.

Tamulaitis, G., Venclovas, C., and Siksnys, V. (2017). Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. Trends Microbiol *25*, 49-61.

Tang, G. (2005). siRNA and miRNA: an insight into RISCs. Trends Biochem Sci *30*, 106-114.

Taylor, D.W., Zhu, Y., Staals, R.H., Kornfeld, J.E., Shinkai, A., van der Oost, J., Nogales, E., and Doudna, J.A. (2015). Structural biology. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. Science *348*, 581-585.

Tock, M.R., and Dryden, D.T. (2005). The biology of restriction and anti-restriction. Curr Opin Microbiol *8*, 466-472.

Unterholzner, S.J., Poppenberger, B., and Rozhon, W. (2013). Toxin-antitoxin systems: Biology, identification, and application. Mob Genet Elements *3*, e26219.

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol *12*, 479-492.

van Erp, P.B.G., Patterson, A., Kant, R., Berry, L., Golden, S.M., Forsman, B.L., Carter, J., Jackson, R.N., Bothner, B., and Wiedenheft, B. (2017). Conformational Dynamics of DNA Binding and Cas3 Recruitment by the CRISPR RNA-Guided Cascade Complex. ACS Chem Biol *13*, 481-490.

van Houte, S., Ekroth, A.K.E., Broniewski, J.M., Chabas, H., Ashby, B., Bondy-Denomy, J., Gandon, S., Boots, M., Paterson, S., Buckling, A.*, et al.* (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. Nature *532*, 385-388.

Vasu, K., and Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev *77*, 53-72.

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., and Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. Cell *163*, 840-853.

Wang, R., and Li, H. (2012). The mysterious RAMP proteins and their roles in small RNA-based immunity. Protein Sci *21*, 463-470.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D.J. (2008). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature *456*, 921-926.

Warren, R.A. (1980). Modified bases in bacteriophage DNAs. Annu Rev Microbiol *34*, 137-158.

Wei, Y., Terns, R.M., and Terns, M.P. (2015). Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. Genes Dev *29*, 356-361.

Westra, E.R., Nilges, B., van Erp, P.B., van der Oost, J., Dame, R.T., and Brouns, S.J. (2012a). Cascade-mediated binding and bending of negatively supercoiled DNA. RNA Biol *9*, 1134-1138.

Westra, E.R., Pul, U., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L.*, et al.* (2010). H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. Mol Microbiol *77*, 1380-1393.

Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K., and Brouns, S.J. (2013). Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. PLoS Genet *9*, e1003742.

Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K.*, et al.* (2012b). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol Cell *46*, 595-605.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature *477*, 486-489.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331-338.

Wilkinson, M., Drabavicius, G., Silanskas, A., Gasiunas, G., Siksnys, V., and Wigley, D.B. (2019). Structure of the DNA-Bound Spacer Capture Complex of a Type II CRISPR-Cas System. Mol Cell *75*, 90-101 e105.

Willkomm, S., Zander, A., Grohmann, D., and Restle, T. (2016). Mechanistic Insights into Archaeal and Human Argonaute Substrate Binding and Cleavage Properties. PLoS One *11*, e0164695.

Wright, A.V., and Doudna, J.A. (2016). Protecting genome integrity during CRISPR immune adaptation. Nat Struct Mol Biol *23*, 876-883.

Wright, A.V., Liu, J.J., Knott, G.J., Doxzen, K.W., Nogales, E., and Doudna, J.A. (2017). Structures of the CRISPR genome integration complex. Science *357*, 1113-1118.

Wright, A.V., Nunez, J.K., and Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell *164*, 29-44.

Xiao, Y., Luo, M., Dolan, A.E., Liao, M., and Ke, A. (2018). Structure basis for RNA-guided DNA degradation by Cascade and Cas3. Science *361*.

Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M., and Ke, A. (2017a). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. Cell *170*, 48-60.e11.

Xiao, Y., Ng, S., Nam, K.H., and Ke, A. (2017b). How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. Nature *550*, 137-141.

Xu, Z., Li, Y., Li, M., Xiang, H., and Yan, A. (2021). Harnessing the type I CRISPR-Cas systems for genome editing in prokaryotes. Environ Microbiol *23*, 542-558.

Xue, C., Seetharam, A.S., Musharova, O., Severinov, K., Brouns, S.J., Severin, A.J., and Sashital, D.G. (2015). CRISPR interference and priming varies with individual spacer sequences. Nucleic Acids Res *43*, 10831-10847.

Xue, C., Whitis, N.R., and Sashital, D.G. (2016). Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. Mol Cell *64*, 826-834.

Yamano, T., Zetsche, B., Ishitani, R., Zhang, F., Nishimasu, H., and Nureki, O. (2017). Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1. Mol Cell *67*, 633-645 e633.

Yan, W.X., Chong, S., Zhang, H., Makarova, K.S., Koonin, E.V., Cheng, D.R., and Scott, D.A. (2018). Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. Mol Cell *70*, 327-339 e325.

Yan, W.X., Hunnewell, P., Alfonse, L.E., Carte, J.M., Keston-Smith, E., Sothiselvam, S., Garrity, A.J., Chong, S., Makarova, K.S., Koonin, E.V.*, et al.* (2019). Functionally diverse type V CRISPR-Cas systems. Science *363*, 88-91.

Yang, H., Gao, P., Rajashankar, K.R., and Patel, D.J. (2016). PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. Cell *167*, 1814-1828 e1812.

Yoganand, K.N., Muralidharan, M., Nimkar, S., and Anand, B. (2019). Fidelity of prespacer capture and processing is governed by the PAM-mediated interactions of Cas1-2 adaptation complex in CRISPR-Cas type I-E system. J Biol Chem *294*, 20039-20053.

Yoganand, K.N., Sivathanu, R., Nimkar, S., and Anand, B. (2017). Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. Nucleic Acids Res *45*, 367-381.

Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Research *40*, 5569-5576.

Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T., and Patel, D.J. (2005). Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol Cell *19*, 405-419.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A.*, et al.* (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759-771.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad, N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S.*, et al.* (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat Biotechnol *35*, 31-34.

Zhang, B., Ye, Y., Ye, W., Perculija, V., Jiang, H., Chen, Y., Li, Y., Chen, J., Lin, J., Wang, S.*, et al.* (2019). Two HEPN domains dictate CRISPR RNA maturation and target cleavage in Cas13d. Nat Commun *10*, 2544.

Zhang, C., Konermann, S., Brideau, N.J., Lotfy, P., Wu, X., Novick, S.J., Strutzenberg, T., Griffin, P.R., Hsu, P.D., and Lyumkis, D. (2018). Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d. Cell *175*, 212-223 e217.

Zhang, J., Kasciukovic, T., and White, M.F. (2012). The CRISPR Associated Protein Cas4 Is a 5′ to 3′ DNA Exonuclease with an Iron-Sulfur Cluster. PLOS ONE *7*, e47232.

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

Zhang, Y., Heidrich, N., Ampattu, B.J., Gunderson, C.W., Seifert, H.S., Schoen, C., Vogel, J., and Sontheimer, E.J. (2013). Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. Mol Cell *50*, 488-503.

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. Nature *515*, 147-150.

Appendix

1. Table 1. Oligonucleotide and other DNA sequences used in this study:

Oligo name	Sequence (5' to 3')	Description
Cas3-pQE2-FP	ATGCCATATGATGTACATTGCCCATATTCGA	Amplification of gene encoding Cas3 from *B. halodurans* with restriction sites NdeI and KpnI for pQE2 plasmid.
Cas3-pQE2-RP	GGTACCTTAAAACGACATAAAGTCCAT	
Cas3-LIC-FP	TACTTCCAATCCAATGCAATGTACATTGCCCATATTCGA	Amplification of gene encoding Cas3 from *B. halodurans* with LIC sites for p1R plasmid.
Cas3-LIC-RP	TTATCCACTTCCAATGTTATTATTAAAACGACATAAAGT CCAT	
Cas3-D48A	GTGTATTTCCCTAGAGCGTGGAGGAGCCCCG	Along with Cas3-pQE2 and Cas3-LIC primers, these primers were used for PCR based site-directed mutagenesis.
Cas3-D395A	GCCATTCCGTCTTAATTTTTGCTGAAGTGCAGAAAGTAC CG	
Cas3-Q253A	TGGCAAATGACTCGCACGCCTCAGACATAGCAGACC	
Cas3-K742A	GACGACTTATCAAAGGTGTTGGCAAAGGCGCAGCAGTA TACAG	
Cas3-K743A	GTTGACGACTTATCAAAGGTGTTGAAAGCGGCGCAGCA GTATA	
Cas3-Q745A	CAAAGGTGTTGAAAAAGGCGGCGCAGTATACAGTCAAC CTTT	
Cas3-Q746A	AAGGTGTTGAAAAAGGCGCAGGCGTATACAGTCAACCT TTATTC	
Cas3-Y747A	GTGTTGAAAAAGGCGCAGCAGGCTACAGTCAACCTTTA TTCACA	
Cas3'-FP	GGAATTCCATATGGCTATGTCTGAGCAGTGCGAGTCATT T	Amplification of helicase and C-terminal domain of Cas3 from *B. halodurans* (Amino acid 249-800) with restriction sites NdeI and KpnI for insertion in pQE2 plasmid
Cas3"-RP	ATACGGGGTACCTTAAGACCGCAGTACATTAATTGGTTT CTGCGC	Amplification of nuclease domain of Cas3 from *B. halodurans* (Amino acid 1-248) with restriction sites NdeI and KpnI for insertion in pQE2 plasmid.
Δ*hns*-FP	ATTATTACCTCAACAAACCACCCCAATATAAGTTTGAGA TTACTACAATGTATGAATATCCTCCTTAGTT	For deletion of *hns* gene from Δcas3 strain of *E. coli K-12*
Δ*hns*-RP	GATTTTAAGCAAGTGCAATCTACAAAAGATTATTGCTTG ATCAGGAAATCTGTAGGCTGGAGCTGCTTCG	
Cas5-FP	TACTTCCAATCCAATGCAATGAGAAACGAAGTCCAATTT GAGCTATTT	Amplification of *cas5*, *cas8c* and *cas7* as a single construct from *B. halodurans* with LIC site for p1R
Cas7-RP	TTATCCACTTCCAATGTTATTATTACTGGCCATCAATCA CTT	

Δ*cas3*-FP	AGCCCGCTGATATCATCGATAATACTAAAAAAACAGGG AGGCTATTAATGGGCGCGCCTACCTGTGACGG	For deletion of *cas3* (*ygcB*) gene from *E. coli K-12*
Δ*cas3*-RP	ATCGTCATTGATAACAATCATTCCCGAAGTTATTTGGGA TTTGCAGGGATAACTTCATTTAAATGGCGCG	
F1-T7-FP	ATTCGAGCTCGGTACCCGGGGATCCTAATACGACTCACT ATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGA AATAATTTTGTT	For integration of *cascade/I-C*, using clonetegration into P21 *attB* site of IYB5101. C1-T7-FP was used to insert T7 promoter upstream of *cascade/I-C*. Other oligonucleotides were used to amplify *cascade* from *B. halodurans*
F2-*cas5*	AGCTCAAATTGGACTTCGTTTCTCATGGTATATCTCCTTC TTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTTA TCCGCTCACA	
F3-*cas5*	TTTAACTTTAAGAAGGAGATATACCATGAGAAACGAAG TCCAATTTGAGCT	
F4-*cas7*-RP	CATGCATCTCGAGGCATGCCTGCAGTTACTGGCCATCAA TCACTTCAACA	
Helicase-act-long	AGTGACTTCTGAGGTGTAGGTTACTCACGCTGCTAAGAA TGTAGTTTAAGGTGCGTGAGTCCGTGTGTAG	For helicase activity assays. Helicase-act-long was annealed to either of the two 6-FAM labelled oligonucleotides to form 3' or 5' overhang.
Helicase-act-3'short	TTAGCAGCGTGAGTAACCTACACCTCAGAAGTCACT-6-FAM3'	
Helicase-act-5'short	5'-6-FAM-CTACACACGGACTCACGCACCTTAAACTACATTC	
F5-Target-TTC-FP	AAACAGCTATGACCATGATTACGCCAAGCTTTTCTAACT CCATCCATAACTAAGCAACCT	These oligonucleotides have 3' complementarity. 100 bp fragment was generated using these oligonucleotides in a PCR reaction and inserted in pUC19 vector to generate Target plasmid (T1)
F6-Target-RP	CAGTGAATTCGAGCTCGGTACCCGGGGATCCTACTAGT ACAGGTTGCTTAGTTATGGATG	
Target-TTC-FP	TACTTCCAATCCAATGCAGGTACCATCGAAAGTAACTCC ATCCATAACTAAGCAACCTGT	These oligonucleotides have 3' complementarity. 100 bp DNA construct was generated using PCR
Target-TAC-FP	TACTTCCAATCCAATGCAGGTACCATCGAATGTAACTCC ATCCATAACTAAGCAACCTGT	
Target-ATG-FP	TACTTCCAATCCAATGCAGGTACCATCGATACTAACTCC ATCCATAACTAAGCAACCTGT	
Targert-RP	TAATAACATTGGAAGTGGATAAAAGCTTTCGATTACTA GTACAGGTTGCTTAGTTATGGA	
Dloop-AAG-FP	TACTTCCAATCCAATGCAGGTACCATCGAAAGTAACTCC ATCCATAACTAAGCAACCTGTACTAGTAATCGAAAGCTT TTATGGACTTCCAATGTTATTA	These oligonucleotides were annealed to generate 100 bp construct for DNA loop related experiment.
Dloop-TTC-RP	TAATAACATTGGAAGTCCATAAAAGCTTTCGAAATGAT CATGTCCAACGAATCAATACCTACCTCAATCTTTCGATG GTACCTGCATTGGATTGGAAGTA	
ssDNA-biotin_12	ATGAAGGTT-Biotin-AGGTATACTCCATGCCTAGGTTCATTGAGGTAGGTATTG ATTCGTTGGACA-6-FAM-3'	ssDNA target with biotin at 12th position
ssDNA	ATGAAGGTTAGGTATACTCCATGCCTAGGTTCATTGAGG TAGGTATTGATTCGTTGGACA-6-FAM-3'	ssDNA target without biotin
ssDNA-biotin_20	ATGAAGGTTAGGTATACTCC-Biotin-ATGCCTAGGTTCATTGAGGTAGGTATTGATTCGTTGGAC A-6-FAM-3'	ssDNA target with biotin at 20th position
CRISPR/I-C	TAATACGACTCACTATAGGGTCGCACTCTTCATGGGTGC GTGGATTGAAATATTGAGGTAGGTATTGATTCGTTGGAC ATGATCATGTCGCACTCTTCATGGGTGCGTGGATTGAAA TATTGAGGTAGGTATTGATTCGTTGGACATGATCATGTC GCACTCTTCATGGGTGCGTGGATTGAAATATTGAGGTAG	DNA sequence encoding type I-C CRISPR array having 7 copies of same repeat-spacer units with T7

	GTATTGATTCGTTGGACATGATCATGTCGCACTCTTCAT GGGTGCGTGGATTGAAATATTGAGGTAGGTATTGATTC GTTGGACATGATCATGTCGCACTCTTCATGGGTGCGTGG ATTGAAATATTGAGGTAGGTATTGATTCGTTGGACATGA TCATGTCGCACTCTTCATGGGTGCGTGGATTGAAATATT GAGGTAGGTATTGATTCGTTGGACATGATCATGTCGCAC TCTTCATGGGTGCGTGGATTGAAATATTGAGGTAGGTAT TGATTCGTTGGACATGATCATGTCGCACTCTTCATGGGT GCGTGGATTGAAATATTGAGGTAGGTATTGATTCGTTGG ACATGATCATGTCGCACTCTTCATGGGTGCGTGGATTGA AATCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCT GCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGG GGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGA	promoter and T7 terminator.

2. Table 2. List of strains used in this study:

Strain	Genotype	Source
E. coli BW25113	Δ(*araD-araB*)567, Δ*lacZ4787*(::rrnB-3), λ-, rph-1, Δ(*rhaD-rhaB*)568, *hsdR514*	CGSC#: 7636
E. coli IYB5101 (Wt)	F-Δ(*araD-araB*)567 Δ*lacZ4787* (::rrnB-3) λ- rph-1 Δ(*rhaD-rhaB*)568 *hsdR514 araB*::T7-RNAp-*tetA*, TetR	Kind gift from Prof. Udi Qimron
E. coli JW1225-2	F-, Δ(*araD-araB*)567, Δ*lacZ4787*(::rrnB-3), λ-, *Δhns-746::kan*, rph-1, Δ(*rhaD-rhaB*)568, *hsdR514,* Kan^R	CGSC#: 9111
E. coli BL21-AI	F- *ompT hsdsB*(rB-, mB-) *gal dcm araB*:: T7-RNAp-*tetA*, Tet^R	Invitrogen
E. coli BL21(DE3)	F- *ompT hsdsB*(rB-, mB-) *gal dcm* λ(DE3)	
E. coli TOP10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80*lacZ*ΔM15 Δ *lacX74 recA1 araD139* Δ(*araleu*)7697 *galU galK rpsL endA1 nupG*, Str^R	Invitrogen
E. coli IC-1	IYB5101 *Δcse1-cas6* P21::*cas5-cas7*, Cam^R	This study
E. coli IC-2	JW1225-2 *Δhns Δcas3*, Cam^R	This study
E. coli IC-3	IYB5101 *Δcse1-cas6* P21::*cas5-cas7*, 186::Array/I-C, Cam^R	This study

3. Table 3: List of plasmids used in this study

Plasmid name	Description	Source
pKD46	*ori* R101, *repA101ts*, Amp^R, *araC*, expresses λ Red genes (*gam-bet-exo*) under the control of arabinose inducible promoter (P_{araBAD}).	CGSC #7739
pOSIP-CT	*ori* Rγ, *ori* pUC, Cam^R, *attP* P21, *ccdB*, λ (*cI857*) encodes P21 integrase under the control of λ promoter (λ pR).	Addgene #45981
pUC19 / pNT1	*ori* PBR322, Amp^R, for gene insertion under the control of lac promoter.	New England Biolabs
pQE2	*ori* ColE1, Amp^R, expresses gene of interest to synthesize N-terminal 6xHis tagged protein under the IPTG inducible T5 promoter.	
pST-KT / pNT2	*ori* PBR322, Kan^R	Kind gift from Dr. VK Nandicoori

pRSF-11 / pT2	*ori* RSF, Kan^R, cloning and expression vector with T7 inducible promoter	
pET StrepII TEV LIC cloning vector (p1R)	*ori* pMB1, Kan^R, *lacI*, expresses the gene of interest to synthesize N-terminal StrepII tagged protein under the control IPTG inducible promoter (PT7lac).	Addgene #29664 (Scott Gradia)
pET StrepII TEV co-transformation cloning vector (p13SR)	*ori* CloDF13, Spc^R, *lacI*, expresses the gene of interest to synthesize N-terminal StrepII tagged protein under the control IPTG inducible promoter (PT7lac).	Addgene #48328 (Scott Gradia)
pCas3/I-E	pQE2 inserted with *cas3* gene from *Escherichia coli* encoding N-terminal 6xHis tagged protein. Ampicillin resistance plasmid. Amp^R	This study
pWT-1/pCas3/I-C	p1R inserted with *cas*3 gene from *B. halodurans* encoding N-terminal Strep-II tagged protein. Kanamycin resistance plasmid (Kan^R)	This study
pWT-2/pCas3/I-C	pQE2 inserted with *cas3* gene from *B. Halodurans* encoding N-terminal 6xHis tagged protein. Ampicillin resistance plasmid (Amp^R)	This study
pHD	p1R inserted with HD domain (nuclease) of Cas3/I-C encoding N-terminal Strep-II tagged protein. Kan^R	This study
pHL	p1R inserted with DExD/H domain (helicase) of Cas3/I-C encoding N-terminal Strep-II tagged protein. Kan^R	This study
pCas3ΔCTD	p1R inserted with Nuclease and Helicase domain of Cas3/I-C without CTD, encoding N-terminal Strep-II tagged protein. Kan^R	This study
pCascade/I-C	p1R inserted with Cascade proteins (Cas5, Cas8 and Cas7) encoding N-terminal Strep-II tagged protein. Kan^R	This study
pK742A	p1R inserted with Cas3/I-C with alanine mutation at K742 amino acid, encoding N-terminal Strep-II tagged protein. Kan^R	This study
pK743A	p1R inserted with Cas3/I-C with alanine mutation at K743 amino acid, encoding N-terminal Strep-II tagged protein. Kan^R	This study
pQ745A	p1R inserted with Cas3/I-C with alanine mutation at Q745 amino acid, encoding N-terminal Strep-II tagged protein. Kan^R	This study
pQ746A	p1R inserted with Cas3/I-C with alanine mutation at Q746 amino acid, encoding N-terminal Strep-II tagged protein. Kan^R	This study
pY747A	p1R inserted with Cas3/I-C with alanine mutation at Y747 amino acid, encoding N-terminal Strep-II tagged protein. Kan^R	This study
pD48A-1	p1R inserted with HD domain mutant (D48A) of Cas3/I-C encoding N-terminal Strep-II tagged protein. Kan^R	This study
pQ253A-1	p1R inserted with helicase domain mutant (Q253A) of Cas3/I-C encoding N-terminal Strep-II tagged protein. Kan^R	This study
pD395A-1	p1R inserted with helicase domain mutant (D395A) of Cas3/I-C encoding N-terminal Strep-II tagged protein. Kan^R	This study
pT1	Target DNA sequence with TTC PAM inserted in pUC19 vector	This study
pCRISPR/I-C	p13SR plasmid inserted with DNA sequence encoding CRISPR array having 7 copies of same repeat-spacer units. Spectinomycin resistance	This study
pD48A-2	pQE2 inserted with Cas3/I-C with alanine mutation at D48 amino acid, encoding N-terminal 6x His tagged protein. Amp^R	This study
pQ253A-2	pQE2 inserted with Cas3/I-C with alanine mutation at D253 amino acid, encoding N-terminal 6x His tagged protein. Amp^R	This study
pD395A-2	pQE2 inserted with Cas3/I-C with alanine mutation at D395 amino acid, encoding N-terminal 6x His tagged protein. Amp^R	This study
pHD + pHL	pQE2 inserted with Nuclease and Helicase domain of Cas3/I-C encoding a bicistronic cassette. Amp^R	This study
pHD + pHLΔCTD	pQE2 inserted with Nuclease and Helicase domain of Cas3/I-C without CTD encoding bicistronic cassette. Amp^R	This study
pCas4-1-2	13S-R inserted with Cas4-1-2 as a single operon. $Spec^R$	This study
pCas1-2	13S-R inserted with Cas1-2 as a single operon. $Spec^R$	This study

List of Publications

- **Nimkar, S.**, and Anand, B. (2020). Cas3/I-C mediated target DNA recognition and cleavage during CRISPR interference are independent of the composition and architecture of Cascade surveillance complex. Nucleic Acids Res. *48*, 2486-2501.
- Yoganand, K.N., Muralidharan, M., **Nimkar, S.**, and Anand, B. (2019). Fidelity of prespacer capture and processing is governed by the PAM-mediated interactions of Cas1-2 adaptation complex in CRISPR-Cas type I-E system. J. Biol. Chem. *294*, 20039-20053.
- Punetha A, Yoganand KNR, **Nimkar S**, Anand B (2018). Cutting it Right: Plasticity and Strategy of CRISPR RNA Specific Nucleases. Proc. Indian Natn. Sci. Acad. *84*: 455-477
- Yoganand, K.N., Sivathanu, R., **Nimkar, S.**, and Anand, B. (2017). Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. Nucleic Acids Res. *45*, 367-381.

Conferences and workshops attended

Poster presentations in conferences

- **Siddharth Nimkar** and B. Anand, "Investigating interference mechanism in type I-C CRISPR-Cas system". 9th RNA Group Meeting. 26-28 October 2017, Banaras Hindu University, Varanasi, India
- **Siddharth Nimkar** and B. Anand, "Characterising the role of Cas3 in CRISPR Interference". MCB 75: from Molecules to Organisms, 11-14 Dec. 2015, IISc Bangalore, India
- **Siddharth Nimkar** and B. Anand, "Characterising the role of Cas3 nuclease domain in CRISPR-Cas interference" 6th Annual Meeting of the Proteomics Society, India PS(I) and International Proteomics Conference, 7-9 Dec. 2014, IIT Bombay, India
- **Siddharth Nimkar** and B. Anand, "Characterising the role of Cas3 helicase in CRISPR-Cas interference". International Conference/WORKSHOP on "Recent Advances in Structure Biology & Drug Discovery. 9-11 October 2014, IIT Roorkee, Roorkee, India

Workshops attended

- CEM3DIP 2018 of macromolecular assemblies and cellular tomography, EMBO practical course. 18 - 29 March 2018, IIT Delhi, New Delhi, India
- Computational Biotechnology at the Nanoscale: CCP4-2016. 15-20 February 2016, Regional Centre for Biotechnology, Faridabad, India