Understanding the defence mechanism of CRISPR-Cas subtype I-C of *Leptospira interrogans*

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

by

Vineet Anand

Under the supervision of

Prof. Manish Kumar

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

TH-3470_166106104

Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

Guwahati-781039, Assam, India

Declaration

I do hereby declare that the matter embodied in this thesis entitled "**Understanding the defence mechanism of CRISPR-Cas subtype I-C of *Leptospira interrogans***" is the result of work carried out under the supervision of Prof. Manish Kumar, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati. In keeping with the general practice of reporting scientific observations, due acknowledgment has been made wherever the work described is based on the findings of other investigators. The work illustrated in this thesis has not been submitted elsewhere for any other degree.

30/9/24

Vineet Anand

Roll no. 166106104

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

TH-3470_166106104

Indian Institute of Technology Guwahati
Department of Biosciences and Bioengineering
Guwahati-781039, Assam, India

Certificate

It is certified that the work described in this thesis entitled "**Understanding the defence mechanism of CRISPR-Cas subtype I-C of *Leptospira interrogans***" by **Mr. Vineet Anand (Roll no. 166106104)** 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.

Manish Kumar
30/09/2024

Prof. Manish Kumar
(Thesis supervisor)
Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India

TH-3470_166106104

I dedicate this thesis to

my parents, relatives, friends and teachers.

TH-3470_166106104

Acknowledgment

In moments of profound emotion, words often fall short in conveying the depth of thanks and appreciation. The wisdom, commitments, and efforts of numerous individuals have been the driving force behind this research endeavour. Many have generously contributed their insights and offered valuable feedback. It fills me with pride and joy to recognize the invaluable contributions of those who have been inspirational, supportive, and instrumental throughout this journey, enriching me with essential knowledge that led to success in my endeavours. Their collective influence is evident in my work, and I am sincerely grateful to each one of them.

First and foremost, I humbly acknowledge the grace of God, who has blessed me with opportunities and guided my instincts in pursuing my career path.

I am deeply grateful to Prof. Manish Kumar, my thesis supervisor, for affording me the opportunity to pursue my aspirations. Under his mentorship, I have been privileged to access a wealth of knowledge and guidance that surpasses any other. The invaluable scientific insights and life lessons I have gained will undoubtedly benefit my future pursuits.

I wish to extend my gratitude to the members of my doctoral committee, namely Prof. Shankar Prasad Kanaujia (Chairperson), Dr. Rajkumar P. Thummer, and Prof. Lal Mohan Kundu, for their impartial insights and invaluable suggestions that have propelled the completion of my thesis. Additionally, I am thankful for their invaluable guidance provided during annual progress seminars.

I would like to acknowledge external thesis examiners Prof. Anirban Banerjee (IIT Bombay) and Prof. Prof. Mathieu Picardeau (Institute of Pasteur, New Caledonia, France) for their constructive suggestions.

I am grateful to Prof. Kannan Pakshirajan, Prof. Latha Rangan, and Prof. Rakhi Chaturvedi from the Department of Biosciences and Bioengineering at IIT Guwahati for their support and guidance as esteemed Heads of the Department throughout my PhD journey.

I am indebted to the Department of Biosciences and Bioengineering and the Central Instrumentation Facility at IIT Guwahati for furnishing the essential research infrastructure crucial for achieving the objectives of my PhD thesis. My heartfelt gratitude goes out to all the Department of Biosciences and Bioengineering staff members for their indispensable logistical

TH-3470_166106104

support, which facilitated the progression of my research endeavors. Furthermore, I extend my appreciation to IIT Guwahati, the Ministry of Human Resources Development, India, and the Department of Biotechnology, Government of India, for their financial support. Department of Biotechnology (DBT) and Department of Science and Technology (DST) for providing research grand to our laboratory.

I extend my deepest appreciation to all my lab colleagues: Dr. Karukriti Ghosh, Dr. Bhuvan Dixit, Dr. Aman Prakash, Dr. Anusua Dhara, Saswat Hota, Surbhi Kumari, Pratibha Koundal, Prattusha Khan, and Shibam Dey, for their invaluable support and insightful feedback.

I am also grateful to Dr. Jaideep Moitra (Geenove Biopharmaceutical, Pune; B.Tech project supervisor), Prof. Sanjay Gupta (JIIT, NOIDA; M.Tech project supervisor), and Prof. Sudha Srivastava (JIIT, NOIDA; project supervisior) for invaluable guidance during my bachelors and master's degree.

I want to express my gratitude to my amazing friends Rutav Mehta, Siddharth Nimkar, Kamal Shokeen, Suvankar Ghosh, Pratap Chandra, Adhiraj Nath, Ratan Kumar, Sayan Saha, Suraj Kumar Mandal, Monika Chandravanshi, Pratik Das Gupta, Gayatri, Kalyan Ghosh, Ritu Tripathi, Arpana Gupta, and Navneeth Sriram who stood by me through both the highs and lows, facing challenges together. Their care, love, support, constructive criticism, and shared knowledge have been a privilege to receive.

Finally, I extend my heartfelt gratitude to my family for being my unwavering support throughout my PhD journey. I am truly fortunate to have such a courageous mother whose unwavering support and encouragement, despite her ailments, propelled me to complete my PhD. The countless personal sacrifices made by my family have been instrumental in reaching this milestone in life.

Additionally, I express my appreciation to all others whom I may have inadvertently omitted.

Vineet Anand

TH-3470_166106104

Table of contents

TH-3470_166106104

TH-3470_166106104

TH-3470_166106104

TH-3470_166106104

TH-3470_166106104

TH-3470_166106104

List of figures

TH-3470_166106104

TH-3470_166106104

TH-3470_166106104

List of tables

TH-3470_166106104

Abbreviations

Amp:	Ampicillin
BLAST:	Basic Local Alignment Search Tool
bp:	Base pair
CRISPR-Cas:	Clustered Regularly Interspaced Short Palindromic Repeats and associated genes
cDNA:	Complementary DNA
crRNA:	CRISPR RNA
Csm:	Cas subtype Mtube
Cmr:	Cas module RAMP
CFU:	Colony forming unit
Chl:	Chloramphenicol
DNA:	Deoxyribonuleic acid
DNase:	Deoxyribonuclease
DSB:	Double stranded break
dsDNA:	Double stranded deoxyribonucleic acid
DTT:	Dithiothreitol
EDTA:	Ethylenediamine tetraacetic acid
EMJH:	Ellinghausen McCullough-Johnson-Harris
EtBr:	Ethidium bromide
FAM:	Fluorescein amidite
g:	gram
HRP:	Horse Radish Peroxidase
h:	hour

TH-3470_166106104

IHF:	Integration Host Factor
IPTG:	Isopropyl ß-D-1-thiogalactopyranoside
kb:	kilo base pair
kDa:	kilo Dalton
Kan:	kanamycin
LIC:	*Leptospira interrogans* sv. Copenhageni
MSA:	Multiple Sequence Alignment
min:	minute
mM:	Milli molar
NTA:	Nitrilotriacetic acid
NHEJ:	Non-homologous end joining
NCBI:	National Center for Biotechnology Information
ORF:	Open Reading Frame
OD:	Optical density
PCR:	Polymerase Chain Reaction
PDB:	Protein Data Bank
PAM:	Protospacer Adjacent Motif
Pre-crRNA:	Precursor CRISPR RNA
PBS:	Phosphate buffered saline
RT-PCR:	Reverse Transcriptase-Polymerase Chain Reaction
RPM:	Rotation Per Minute
RNA:	Ribonucleic acid
µg:	Micro gram

TH-3470_166106104

RMSD:	Root Mean Square Deviation
RRM:	RNA recognition motif
RAMP:	Repeat associated mysterious protein
SDS-PAGE:	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
sv.:	Serovar
SEC:	Size-exclusion chromatography
SYBR:	Synergy brand
ssDNA:	Single stranded DNA
TAE:	Tris-acetate EDTA
TBE:	Tris-borate EDTA

TH-3470_166106104

Abstract

Leptospira interrogans is a bacterium that causes Leptospirosis. However, its pathogenicity is poorly understood due to the lack of effective conventional genetic manipulation tools. *L. interrogans* serovar (sv.) Copenhageni encodes two CRISPR-Cas systems - I-B and I-C in its genome. The presence of the CRISPR-Cas system in the genome of pathogenic *Leptospira* has been hypothesized as the reason for low success in genetic manipulation and, hence, deciphering its virulent gene function. The CRISPR-Cas I-B locus in *L. interrogans* has been extensively studied. However, CRISPR-Cas I-C lacks an array element, making it an interesting subject to investigate its Cas protein's role in CRISPR biology. The CRISPR-Cas I-C locus encodes all sets of proteins required for adaptation (LinCas1C, LinCas2C, and LinCas4C), maturation (LinCas5C), and interference (LinCas7C, LinCas8C, LinCas3C, and LinCas3C′). Moreover, the genes encoding these proteins are transcriptionally active. The nucleotide sequence and encoded amino acids of the *cas* genes show low sequence similarity with subtype I-B and are phylogenetically distantly related. The adaptation Cas protein (LinCas1C, LinCas2C, and LinCas4C) exhibits metal-ion-dependent DNase and metal-ion-independent RNase activity under *in vitro* conditions. However, neither rLinCas1C nor LinCas2C demonstrates any activity towards small DNA oligos, except for rLinCas4C. The crystal structure analysis of rLinCas2C shows that it adopts a dimeric conformation, and each subunit exhibits the characteristic ferredoxin fold. The crystal structure of rLinCas2C indicates its existence in a catalytically inactive conformational state. The substitution mutation of active site residues (Tyr7, Asp8, Arg33, and Phe39) of rLinCas2C with Ala and deletion mutation of RNA recognition loop (Asp60, Lys62, Thr63, and Asp64) resulted in compromised DNase activity. The catalytic activity observed in adaptation Cas proteins (LinCas1C and LinCas4C) of sub-type I-C suggests their potential involvement in recognizing/processing foreign nucleic acid and integrating it as a spacer into the CRISPR array. While the catalytic function of LinCas2C might be necessary for biological processes distinct from the CRISPR-Cas-associated function. The rLinCas5C, a maturation Cas protein, exhibits DNase activity and is contingent on the large size of the DNA substrates (≥3kb), as it displayed no activity on short ssDNA and dsDNA oligos (≤36 nt). This study also revealed that within subtype I-B, which possesses a complete set of Cas proteins and array elements, the mature crRNA-rLinCas6 complex is protected from the endoribonuclease activity of rLinCas5C. *In silico* analysis discovered that among Cas5C orthologs, LinCas5C exclusively

TH-3470_166106104

contains β-sheet insertion in the C-terminal, and Phe141 is one of the catalytic triad equivalents to a more conserved histidine residue. Deleting additional β-sheet insertions and substituting Phe141 with His141 in LinCas5C resulted in the gain of nuclease activity compared to the wild-type. The rLinCas5C exhibits non-canonical cleavage activity on the non-cognate repeat RNAs and pre-crRNAs. This raises a question about CRISPR-Cas I-C interference capability. To address this, a plasmid interference assay was conducted, which revealed that the Cas proteins of CRISPR-Cas I-C of *Leptospira* involved in the interference could functionally intervene in the foreign plasmids carrying predicted protospacer adjacent motifs in the spacers. Overall, this study provides a comprehensive understanding of the *Leptospira* CRISPR-Cas I-C adaptation and maturation Cas proteins' biochemical activity and sheds light on the interference capability of Cas proteins in a heterologous host (*E. coli*). These findings lay the groundwork for leveraging the endogenous CRISPR-Cas systems of *L. interrogans* for genome editing in the near future.

TH-3470_166106104

Chapter 1

Review of literature

1. Defence system in prokaryotes

Bacteria are consistently confronted by bacteriophages, the most prevalent microbes on earth. In response, bacteria have evolved numerous immune systems to counteract bacteriophage infections. A dynamic interplay ensues as bacteriophages can develop resistance mechanisms, engaging in an evolutionary arms race with their hosts. The swift pace of evolution becomes imperative for the survival of bacterial cells (Bikard & Marraffini, 2012). Bacteria have evolved many defence mechanisms to prevent or lessen the effects of biotic assaults. Understanding these barriers is critical for several reasons. They provide insights into bacterial biology by shedding light on the environmental constraints that bacteria faced in the past and the strategies that evolved to solve them. These mechanisms are phylogenetically widespread and have an impact on the physiology of numerous bacterial species (Wein & Sorek, 2022).

1.1 Classes of bacterial defence.

Bacteria have numerous defence systems against rivals, phages, and predators. These systems operate at various spatial scales, ranging from molecular and cellular defences to those that require bacteria to function (Smith et al., 2023).

1.2 Defence at multi-cellular level.

Clonal groupings of bacteria frequently collaborate, thus, collectively surviving challenges that would kill single cells (Granato et al., 2023). The production of biofilms is the most common example of multicellular protection in bacteria. Biofilms are at the root of many chronic diseases and frequently form in response to drugs and competition from other strains (Cornforth & Foster, 2013; Lories et al., 2020; Oliveira et al., 2015). Another collective defensive approach that bacteria use is maintaining long-term population variability in particular phenotypes (for example, growth phase) so that all individual cells in the colony do not suffer equally when conditions deteriorate. This type of phenotypic variability is linked to clinical antibiotic tolerance (Arnoldini et al., 2014; Łapińska et al., 2022). This could also be a way for bacteria to resist toxins from competition.

1.3 Defence at a cellular level.

Most hazardous substances must first penetrate a cell before they can cause detriment, and bacterial membranes are crucial in preventing this invasion. Membrane surface structures, for example, lipopolysaccharides and curli fibers, protect by blocking phage or toxin-binding sites or shielding the cell against mechanical abuse (Dy et al., 2014; Vidakovic et al., 2018)**.** When the cell membrane fails to prevent the entry of hazardous molecules, bacteria can force them back out by efflux pumps. Efflux pumps are a varied set of bacterial membrane transport proteins with various substrate specificities and physiological functions (Saier Jr et al., 1998; Webber & Piddock, 2003).

1.4 Defence at the molecular level.

Molecular level defence is typically performed at post-translational and genetic modifications; for example, enzymatic methylation of bacterial ribosomes can prevent various types of antibiotics from binding to their target (Blair et al., 2015). Antibiotics, other competitive toxins, or phages frequently cause oxidative DNA damage in bacteria (Dong et al., 2015; Fan et al., 2018), which is compensated via the base excision repair and nucleotide excision repair systems, which are both highly conserved and ancient processes (Kisker et al., 2013; Wozniak & Simmons, 2022). Besides chromosomal repair, certain species have RNA ligases capable of repairing 16S ribosomal RNA damage produced by ribotoxic bacteriocins (Maviza et al., 2022). Similarly, filamentous phage extrusion can damage *E. coli's* inner membrane, although the production of membrane-binding phage shock proteins controls proton leakage and preserves the proton-motive force (Joly et al., 2010). Bacteria have a variety of mechanisms for degrading hazardous substances. β-lactamases are ancient proteins that hydrolyze the ring structure of β-lactam antibiotics like penicillin (Ross et al., 2019). Restriction-modification systems encode restriction endonucleases, which bind to and cleave phage and other foreign DNA at specified recognition sites. Target modification also plays a critical role here, but it is directed toward host DNA: recognition sequences on host DNA are modified to protect them from degradation, whereas the endonuclease destroys unmodified phage DNA (Ofir et al., 2018; Tock & Dryden, 2005) enabling innate protection against a subset of phages. The degradation of exogenous nucleic acids is achieved via complexity in CRISPR-Cas systems, which offer bacteria adaptive immunity against phages with previously encountered genetic signatures (Smith et al., 2023).

TH-3470_166106104

2. Innate and adaptative immune response in prokaryotes

The key differentiation lies in the fact that innate immune systems are present from birth and lack memory of previous infections, rendering innate immunity non-specific. In contrast, adaptive immune systems utilize past infections to mount defences against phages and other threats (Owen et al., 2013). The presence of numerous unidentified genes within the defence island implies its potential role in countering phages. Notably, genes associated with the restriction-modification (R-M) system, the toxin-antitoxin system, and abortive infection are frequently situated in defence islands (Rostøl & Marraffini, 2019). The clustering of these defence genes suggests functional cooperation among defence mechanisms, although the specific relationships between these systems remain unclear. Moreover, the connection between defence genes and mobilomes strongly indicates that horizontal gene transfer (HGT) significantly contributes to the evolution of these islands. The subsequent section delves into a detailed exploration of the diversity observed in prokaryotic defensive systems (Bernheim & Sorek, 2020; Labrie et al., 2010). Immune systems are typically classified into two main types: innate and adaptive immune systems.

2.1 Innate immunity.

2.1.1 Disable phage adsorption.

The invasion of phages initiates with the recognition of a specific surface receptor molecule, typically a protein, polysaccharide, or lipopolysaccharide (LPS), leading to the attachment of a phage particle to the bacterial cell wall. Subsequently, the phage genome is introduced into the bacteria. In response, bacteria employ a phage-specific strategy to thwart the entry of viral DNA, involving modifications to surface receptors. These receptor molecules must not only be present on the surface but also accessible to the phage particle. Bacteria attempt to alter the molecular structure of the receptor through mutations or by shielding it with additional barriers (Labrie et al., 2010). Alterations in the *lamB* gene, responsible for the phage receptor, can hinder phage absorption, resulting in robust resistance, as illustrated by early studies on the lambda coliphage (Clement et al., 1983).

2.1.2 Restriction modification system.

An RM system comprises two enzyme activities designed to identify and respond to a specific DNA sequence. One enzyme cuts the sequence when it lacks methylation, while the other attaches methyl groups to safeguard it against cleavage. When foreign, non-methylated DNA, such as that from invading phages, enters the cell, it is selectively cleaved. This mechanism provides a level of immunity to infection, the extent of which varies based on the specific RM system and the nature of the phage involved (Korona et al., 1993; Korona & Levin, 1993). The effectiveness of RM systems in conferring immunity is limited. Generally, an RM system identifies only a limited set of DNA segments within a particular phage, and each of these segments has the possibility of being methylated by the bacterial modification enzyme before undergoing cleavage. Consequently, the phage may occasionally succeed in initiating a productive infection and generating a substantial number of progeny phages. Additionally, when a phage survives an infection cycle, the DNA of its descendants becomes methylated. Subsequent infections with the same bacterial strain allow the progeny to evade the RM system (Arber, 1965; Korona & Levin, 1993).

2.1.3. Toxin-Antitoxin systems.

Bacterial Toxin-Antitoxin (TA) systems are prevalent genetic elements that serve as defenders against invading phages and plasmids. These systems consist of a toxin, disrupting a crucial cellular process and inducing growth arrest, and a corresponding antitoxin, which counteracts the toxin's activity under normal growth conditions. The specific action of the toxin varies among bacteria. It may function as a DNase or an RNase, inhibiting DNA replication or protein synthesis (Hayes & Van Melderen, 2011; Page & Peti, 2016). *Leptospira interrogans* is the sole organism known to harbor three toxin-antitoxin systems, namely *chpIK*, *mazEF*, and *vapBC*. In the VapBC toxin-antitoxin system, VapC exhibits toxicity towards *E. coli* cells, with VapB acting as the counteractive agent against VapC's harmful effects. Similarly, ChpK proves toxic to *E. coli*, and ChpI serves as the antidote to neutralize ChpK's toxicity. The discovery of mazEF was facilitated through homology searching (Picardeau et al., 2001; Zhang et al., 2004). The toxin-antitoxin (TA) system within *Pectobacterium atrosepticum* consists of an endoribonuclease named ToxN, serving as the toxin. To counteract the harmful effects of ToxN, a non-coding RNA called ToxI, functioning as the antitoxin, inhibits ToxN activity by forming a binding interaction with it. In the event of a phage attack, ToxN's ribonuclease

TH-3470_166106104

activity leads to the destruction of both host and phage RNA. Consequently, the compromised RNA renders both the host and phage incapable of survival, thereby halting the progression of the infection (Blower et al., 2011).

2.1.4. Argonaute.

Argonautes (Ago) were initially identified in RNA interference (RNAi) processes in eukaryotes (eAgo) (Bohmert et al., 1998). Subsequently, homologs of eAgo proteins were found in bacteria and archaea (pAgo) (Shujuan Jin et al., 2021). The structural composition of pAgo indicates the presence of PIWI, PAZ, and MID domains, responsible for nucleic acid binding and cleavage (Wang et al., 2008). Initially, the functional role of pAgo was not well understood, but their localization in defence islands suggested a potential involvement in defence mechanisms (Makarova et al., 2009). Experimental investigations later confirmed that pAgo participates in the recognition and nuclease-mediated cleavage of foreign genetic material guided by small single-stranded RNA/DNA (Makarova et al., 2009; Swarts et al., 2014; Yuan et al., 2005). In certain species, a nuclease-inactive form of pAgo, in conjunction with microRNAs, operates to suppress the translation of target mRNA without causing its cleavage (Hutvagner & Simard, 2008). For instance, *Thermus thermophilus* (TtAgo) employs DNA-guided DNA interference to protect against foreign nucleic acids (Swarts et al., 2014).

2.2 Adaptative immunity

2.2.1 CRISPR-Cas system.

Prokaryotic clustered regularly interspaced short palindromic repeats (CRISPR) and Cas (CRISPR associated) proteins encompass heritable adaptive immune systems to complement innate immunity strategies like receptor masking, restriction-modification (R-M) systems, DNA interference, bacteriophage exclusion, and abortive infection, which can be crossed easily by rapidly evolving viruses (Babu et al., 2011; Jansen et al., 2002).

2.2.1.1 Discovery of CRISPR-Cas immune systems.

CRISPR loci were discovered in 1987 after sequencing the *iap* gene of *Escherichia coli* (Ishino et al., 1987). A cluster of unique repeating sequences was discovered downstream of the *iap* gene, but its significance could not be determined until after the accumulation of prokaryote genomes in GenBank (Ishino et al., 1987). Francisco Mojica of the University of

Alicante (Spain) discovered comparable structures in the archaeal genome of *Haloferax mediterranei* in 1995, which marked a key step in understanding the biological role of CRISPR loci (Mojica et al., 1993). Mojica noticed similarities between the elements he described in archaea and previously discovered DNA repeats in bacterial genomes, and he was one of the first scientists to hypothesise that these unusual loci include fragments of foreign DNA and are, in fact, part of bacteria and archaea's immune system (F. J. Mojica et al., 2005). According to the prokaryotic immune system theory, virus DNA pieces separated by short palindromic repeats and clustered in intergenic regions represent a library of potentially hazardous genetic material (Isaev et al., 2021). Initially, it was anticipated that such a system would operate via the RNA interference mechanism. However, it was experimentally proven for the first time in the publication of Marraffini and Sontheimer that the actual target of the immune system of prokaryotes was foreign DNA (Marraffini & Sontheimer, 2008).

To check CRISPR-based immunity, *Streptococcus thermophilus* was infected with phages, and fewer phages were able to infect, which might be connected with the presence of spacer sequences in CRISPR loci (Barrangou et al., 2007). To validate the involvement of CRISPR-Cas loci in antiphage defence, *S. thermophilus* was infected with phage, and the CRISPR locus of the phage resistant bacteria was investigated, which revealed one or more novel spacer sequences that matched a region of the infecting phage's genome (Deveau et al., 2008). This demonstrated that new spacers are acquired during infection, resulting in resistance or immunity to invading phages. Another study discovered that the Cas ribonucleoprotein complex, which contains RNase, is responsible for the formation of mature crRNAs in *E. coli*, which was blocked by mutation of catalytic residues of the RNase or disruption of the complex (Deveau et al., 2008). Another study found that the *Staphylococcus epidermidis* CRISPR locus contains a spacer sequence that matches a portion of the *nickase* gene of *Staphylococcal* plasmids, inhibiting conjugation of such plasmids (F. J. Mojica et al., 2005).

2.2.1.2 Overview of CRISPR-Cas system.

The CRISPR-Cas system is exclusive to prokaryotes, being identified in over 50% of sequenced bacteria and over 80% of sequenced archaea (Grissa et al., 2007; Makarova & Koonin, 2015). CRISPR loci are made up of a succession of direct repeats (about 20-50 base pairs) that are conserved within the loci and are separated by unique spacer sequences **(Fig. 1.1)** (Jansen et al., 2002; Makarova, Wolf, Iranzo, et al., 2020). The sequence and length of

repeat sequences in various CRISPR loci can vary. In addition, the number of repeat-spacer units in a CRISPR locus varies greatly between organisms (Jansen et al., 2002). Comparative examinations of sequences near CRISPR loci found a (A+T)-rich 'leader sequence' that acts as a promoter element for CRISPR transcription (Plagens et al., 2015; Pougach et al., 2010). Bioinformatic investigations have revealed that CRISPR loci are flanked by a significant number of extremely different *cas* genes in addition to the leader sequence (Plagens et al., 2015; Rousseau et al., 2009).

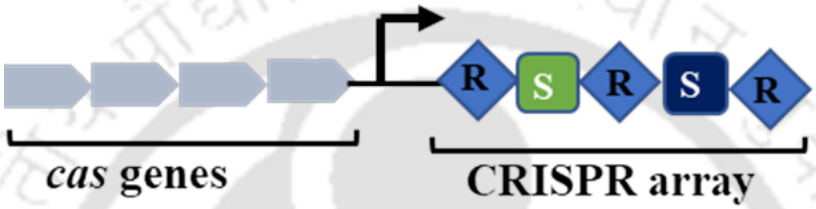


Figure 1.1. Organization of a typical CRISPR-Cas locus in a bacterial or archaeal genome. The CRISPR-Cas system is essentially comprised of sets of *cas* genes followed by a CRISPR array. Positioned upstream of the CRISPR array is a leader sequence, housing the promoter responsible for transcribing the CRISPR array. CRISPR array contains identical direct repeat (R) interspaced by a unique spacer (S) sequence. The arrow denotes the transcription start site (Makarova et al., 2013).

2.2.1.3 Classification of CRISPR-Cas system.

Regardless of sharing functional similarities, the CRISPR-Cas system demonstrates remarkable diversity in Cas proteins, gene compositions, and genomic loci architecture (Barrangou et al., 2007; Sternberg et al., 2016). The extensive variation within the CRISPR-Cas system poses a significant challenge to its categorization and classification. As genomic and metagenomic data continue to grow, understanding of CRISPR-Cas system diversity frequently expands. To address this expansion, various classification schemes have been proposed (Koonin et al., 2017; Makarova & Koonin, 2015). Given the absence of shared genes across CRISPR-Cas systems, a computational approach is employed for classification. This involves the identification of signature genes, comparison of genomic loci and composition, sequence similarity-based clustering, phylogenetic analysis, neighborhood analysis, and consideration of experimental evidence (Makarova, Wolf, Iranzo, et al., 2020). The roles played by Cas proteins serve as the basis for grouping all *cas* genes into three distinct functional

TH-3470_166106104

modules. The first module, the adaptation module, encompasses crucial *cas* genes involved in the spacer integration process, including a CRISPR array and three genes: *cas1*, *cas2*, and *cas4*. The expression module groups genes responsible for crRNA processing, while the interference module contains genes encoding subunits of the effector complex. The CRISPR/Cas system has been divided into two classes, Class 1 and Class 2, based on the architecture of the effector complex. Class 1 has multi-subunit effector complexes, while Class 2 has individual protein effector modules. Further, the Class 1 CRISPR/Cas system has been subdivided into types I, III, and IV, and Class 2 has also been subdivided into types II, V, and VI. **(Fig. 1.2)** (Koonin et al., 2017; Makarova & Koonin, 2015).

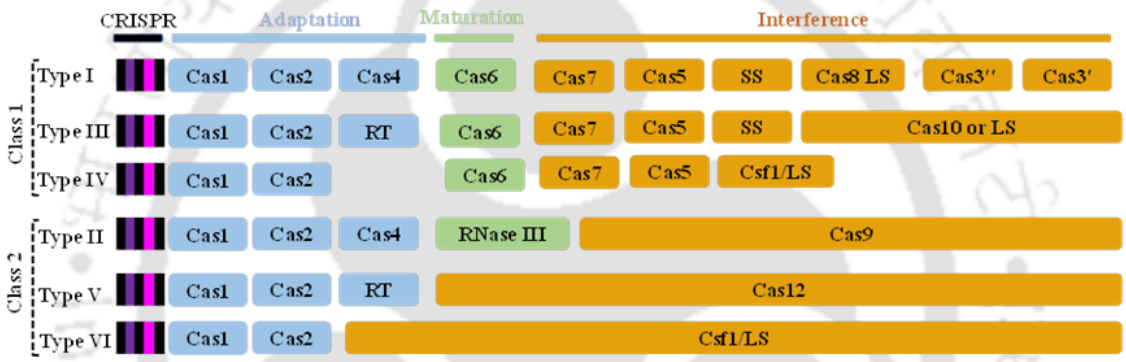


Figure 1.2. Schematic representation of the universal organization of class 1 and class 2 CRISPR-Cas system. Class 1 CRISPR–Cas systems feature interference modules comprised of several Cas proteins, forming a complex that binds to crRNA and cooperatively carries out target binding and processing. In contrast, Class 2 systems employ a single, multidomain crRNA-binding protein that serves a functionally equivalent role to the entire effector complex (interference complex) found in Class 1. In the CRISPR array, repeats are denoted by black color, while spacers are indicated by pink and purple color. Blue, green, and orange represent the adaptation, maturation, and interference modules, respectively. [adapted from (Makarova, Wolf, Iranzo, et al., 2020)].

The CRISPR-Cas type I system can be divided into nine major subtypes based on the presence of a signature gene. The signature protein of the type I system is the Cas3 protein, which contains an N-terminal exonuclease domain and a C-terminal helicase domain (Brouns et al., 2008; Pougach et al., 2010; Sinkunas et al., 2013). In some type, I systems, the endonuclease Cas3 nuclease and helicase domains are encoded by separate genes (*cas3* and *cas3'*, respectively), but in each case, they are thought to participate in degrading foreign nucleic acids (Brouns et al., 2008; Garneau et al., 2010; Han & Krauss, 2009; Mulepati & Bailey, 2011). Nine subtypes have been established within the type 1 CRISPR-Cas system,

TH-3470_166106104

delineated by variations in both the genomic organization of CRISPR-Cas loci and the sequence diversity of Cas8 proteins (Makarova et al., 2013). The type II CRISPR systems consist of four *cas* genes, one of which is *cas9* (signature gene), a large protein that includes both a RuvC-like nuclease domain and an HNH nuclease domain. The type III system has Cas10 as a signature protein, which assembles into a Cascade-like interference complex for target search and destruction. The type IV system is found as a cluster of *cas* genes but without an associated CRISPR array. Type V system contains a Cas9 like nuclease, either Cpf1, C2c1, or C2c3. Type VI systems have Csf1, a large protein with two predicted HEPN (higher eukaryotes and prokaryotes nucleotide-binding) RNase domains (Makarova et al., 2013).

Class 1 systems employ RNA-guided surveillance complexes, which are made up of a variety of subunits organised around a single crRNA molecule. The adaptation proteins Cas1, Cas2, and Cas4 are consistently found in all subtypes of the CRISPR-Cas system. However, Cas4, crucial for prespacer processing, is notably absent in types I-E and I-F, while in type I-G, it is fused with Cas1 (Almendros et al., 2019; Makarova, Wolf, Iranzo, et al., 2020). Furthermore, the maturation of crRNA involves Cas6, but this protein is lacking in several subtypes. For instance, in type I-C, Cas5 takes on the function of Cas6 **(Fig. 1.3)** (Nam, Haitjema, et al., 2012; Punetha et al., 2014).

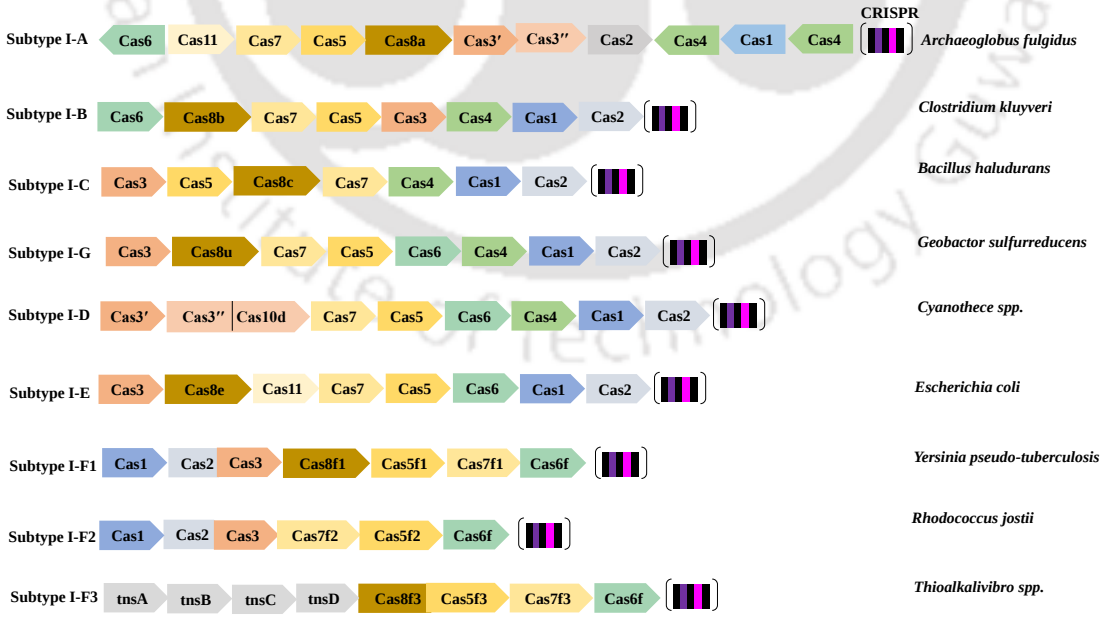


Figure 1.3. Schematic illustration of typical CRISPR-Cas loci of type I subtypes. The type 1 CRISPR-Cas system has been further categorized into nine subtypes based on the genomic structure of CRISPR-Cas loci and the diversity observed in the sequence of Cas8 proteins. On the left side, CRISPR subtypes are delineated, followed by *cas* genes, CRISPR array, and a

TH-3470_166106104

representative organism is depicted. Homologous genes share identical colors. In Type I-A, Cas3' and Cas3" correspond to DExD/H box helicase and HD-nuclease, respectively. In other subtypes, the fusion of helicase-nuclease is present [adapted from (Makarova, Wolf, Iranzo, et al., 2020)].

2.2.1.4 The three phases of CRISPR-Cas based immunity.

The CRISPR-Cas process includes three functionally distinct stages:

(i) In the CRISPR-Cas adaption phase, a segment of foreign DNA (protospacer) is cleaved and inserted as spacers into the chromosomal CRISPR locus by a set of specialized Cas proteins **(Fig. 1.4)** (Babu et al., 2011).

(ii) In the CRISPR-Cas expression and maturation, the spacers inserted into the CRISPR locus are transcribed into precursor-crRNA (pre-crRNA), which is then processed into mature crRNAs, each of which contains a single spacer flanked by a portion of the repeat sequence **(Fig. 1.4)** (Brouns et al., 2008).

(iii) CRISPR interference, the crRNAs bind to Cas endonucleases, forming an RNA-protein Cascade interference complex which is guided by crRNAs to the targeted foreign nucleic acid to which the complex binds and then degrades it **(Fig. 1.4)** (Garneau et al., 2010).

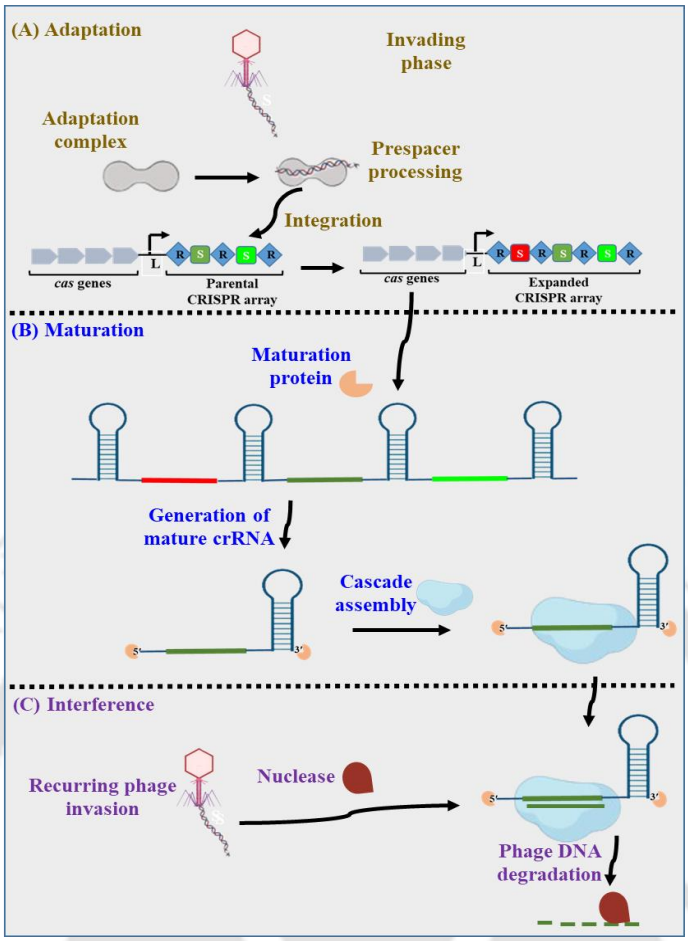


Figure 1.4. Schematic overview of three phases of CRISPR-Cas based immunity. Adaptation, maturation and interference phases of CRISPR-Cas mediated immunity. (A) Adaptation: The adaptation complex (Cas1-Cas2-Cas4) identifies a Protospacer Adjacent Motif (PAM) in the phage DNA or prespacer. The prespacer is then trimmed and integrated into the leader proximal region of the CRISPR array. The nucleases involved in prespacer trimming may vary depending on the CRISPR-Cas type. S represents spacers and R represents repeats. (B) Maturation: The initiation of CRISPR array expression begins in the leader region (L), giving rise to a precursor CRISPR RNA (pre-crRNA) transcript with a stem-loop structure. Cas proteins, such as Cas6 or Cas5, selectively cleave the pre-crRNA at a designated site, producing the ultimate mature CRISPR RNA (crRNA). This mature crRNA then assembles into a Cascade complex. (C) Interference: In the conclusive phase, the invading phage DNA is recognized, prompting the recruitment of an effector nuclease. This nuclease, in turn, facilitates the cleavage of the phage genome, completing the interference process [Adapted from (Donohoue et al., 2018)].

2.2.1.5 Adaptation phase of CRISPR-Cas system.

The insertion of DNA fragments (spacers) from invaders into host CRISPR loci is a critical first step in CRISPR-Cas-mediated defence. This "adaptation" or "spacer acquisition" stage gives heritable immunity to afflicted organisms against future encounters with specific invaders (Sternberg et al., 2016). Notably, new spacers are selectively incorporated at the leader end of the CRISPR array. First, the foreign DNA should be known to be a potential target for spacer acquisition. Second, an appropriate length of sequence (usually 30-40 bp, based on the subtype of CRISPR-Cas system) (Barrangou & Van Der Oost, 2013) must be obtained from the foreign DNA. Finally, the acquired sequence must be included in the CRISPR array as a new spacer, and the neighboring repeat sequence has to be replicated. Although the components and prerequisites of the spacer acquisition machinery differ between organisms and CRISPR-Cas subtypes, many components appear to be universally conserved and required by all CRISPR-Cas subtypes (Amitai & Sorek, 2016). These components are the Cas proteins Cas1 and Cas2 and, within the CRISPR array locus, the leader sequence and the first CRISPR repeat. Cas1 and Cas2 are often encoded in the same operon (Makarova et al., 2013), and together they form a structurally stable protein complex (Cas1-Cas2) and facilitate spacer acquisition (Nuñez et al., 2014a; Wang et al., 2015). The type I-E CRISPR-Cas system in *E. coli* revealed the participation of the Cas1-Cas2 complex in the adaptation stage for the first time in 2012, demonstrating that Cas1 and Cas2 are the only Cas proteins necessary for successful spacer acquisition. Typically, the Cas1-Cas2 complex integrates new spacers sandwiched between the CRISPR array's first repeat and leader sequence (Nuñez, Lee, et al., 2015). The sequence of the leaders is a long AT-rich sequence that is positioned directly upstream of the CRISPR array, which normally includes both the promoter for CRISPR RNA (crRNA) expression and the recognition sequence for spacer insertion (Pougach et al., 2010; Wei et al., 2015; Yosef et al., 2012).

2.2.1.5.1 Mechanism of spacer acquisition.

The junction between the leader sequence and the first CRISPR repeat is the optimal location for new spacer integration, and the minimal sequence required for integration is merely a brief segment at the 3′ end of the leader and a single repeat unit (Wei et al., 2015; Yosef et al., 2012). As a result of the preference for integration at this junction, spacers are inserted into the CRISPR array towards the leader sequence end of the array, resulting in a chronologically

ordered array with the most recently acquired spacer being the spacer closest to the leader sequence (Barrangou et al., 2007; Pourcel et al., 2005). Recent studies on the adaptation stage in *E. coli* type I-E CRISPR-Cas system have revealed that during spacer acquisition, the Cas1–Cas2 complex undergoes conformation change upon binding to a protospacer such that each of the Cas1 dimers rotates in opposing directions around the Cas2 dimer, exposing the surface of a protein that binds to the central duplex region of the protospacer and orienting the two Tyr22 residues bracketing the duplex forming a wedge that terminates the double-stranded DNA (dsDNA) region of 23 bp and splays apart the ends of the DNA on each side of the duplex (Cady et al., 2012). Furthermore, a catalytic pocket is also formed in one monomer of each Cas1 dimer, enabling accurate positioning of the PAM-complementary sequence in the 3′ single-stranded DNA (ssDNA) overhang within catalytically active site of Cas1 subunit as it threads into one of the monomers of each Cas1 dimer through an arginine-rich channel, where it is specifically recognized and cleaved (Connerton & Hooton, 2015). This resulted in a protospacer intermediate with a precise length of 33 nucleotides, consisting of 23 bp dsDNA core region and two splayed 5-nucleotide ssDNA overhangs with 3′-OH groups. The central segment of the 23 bp binds to the surface of the Cas2 dimer and is stabilized by the interactions formed between a set of two arginine residues (arginine clamp) in the Cas1–Cas2 hetero-complex and the phosphate groups in the phosphodiester backbone of the protospacer **(Fig. 1.5A).** This reveals the structural basis for spacer size determination and PAM recognition (Amitai & Sorek, 2016).

In type I-E, type I-F, and type I-B CRISPR–Cas systems where the CRISPR locus already contains a spacer targeting a particular phage or plasmid, the Cascade and Cas3 interference machinery which possibly generates cleavage points or nicks producing the substrate for acquisition by the Cas1–Cas2 complex (Beloglazova et al., 2008; Connerton & Hooton, 2015) leading to the acquisition of additional spacers from the same foreign element much more rapidly and efficiently than normal acquisition (Rouillon et al., 2013). This positive feedback between existing spacers and the acquisition of new spacers is termed primed acquisition or priming. Priming occurs for both types of spacers- active spacers, which trigger interference, and inactive, mismatch-containing spacers, which cannot elicit interference (Rouillon et al., 2013). In Type I-E CRISPR-Cas systems, primed spacers are obtained from a similar strand as the priming protospacer. However, in type I-B and Type I-F CRISPR-Cas

systems, primed spacers are acquired from both strands on either side of the priming protospacer (Amitai & Sorek, 2016).

Cas1-Cas2 complex, besides acquiring the spacer sequence from exogenous DNA, also acts as an integrase in the *E. coli* type I-E CRISPR-Cas system. The protospacer integrates into the CRISPR array in a two-step mechanism that resembles retroviral integration and DNA transposition (Mojica et al., 2009). In the first step, the protospacer 3'-OH group carries out a nucleophile attack on the 5'end of the first repeat, thus beginning spacer integration by forming a branched intermediate where a single strand of the protospacer is ligated to a single strand of the CRISPR array (Amitai & Sorek, 2016). The OH group at the 3' end on the other protospacer strand generates a second nucleophile attack on the 5' phosphate end of the opposite DNA strand of the repeat, which is adjacent to the leader sequence. The product of this reaction is an expanded CRISPR array with a duplicated repeat and a new spacer **(Fig. 1.5B)** (Amitai & Sorek, 2016). The single-stranded DNA gaps generated at the repeat sequences are filled and repaired by DNA polymerase and a DNA ligase **(Fig. 1.5B)**, but the specific proteins that carry out these tasks have not yet been identified (Connerton & Hooton, 2015). In the in vitro assay, spacer integration by the Cas1–Cas2 complex was possible when the protospacer was a dsDNA molecule with either blunt ends or 3′-overhangs but not when the protospacer was ssDNA molecule; this is consistent with the structure of protospacer-bound Cas1–Cas2 complexes. In the Cas1–Cas2–protospacer complex, there are two Cas1 monomers without protospacer DNA in their catalytic sites, which are possibly responsible for binding to the CRISPR array during spacer integration (Beloglazova et al., 2008).

TH-3470_166106104

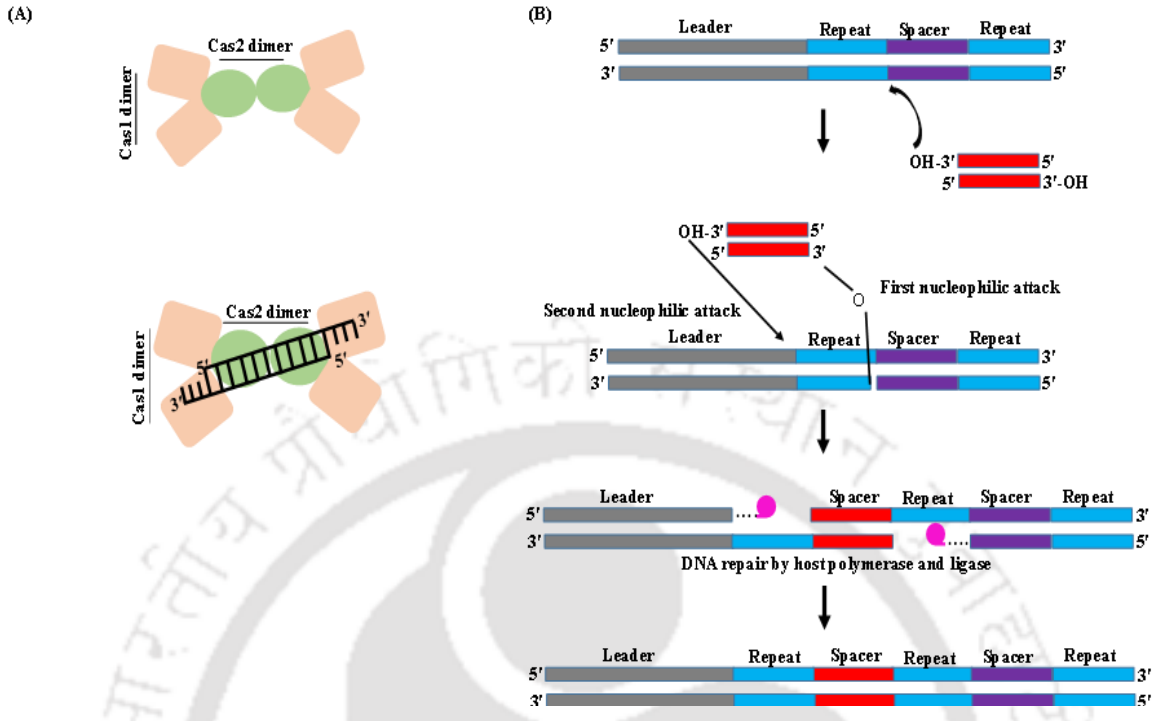


Figure 1.5. Schematic representation of the adaptation phase. (A) The schematic illustration of protospacer binding by Cas1-Cas2 heteroprotein complex, comprising two Cas1 dimers and one Cas2 dimer. The protospacer, containing double-stranded DNA with single-stranded DNA overhangs, is positioned in a top arch formed by all six subunits. (B) The mechanism for protospacer integration into the CRISPR array involves the 3′-OH group of the protospacer initiating a nucleophilic attack on the 5′ end of the first repeat. This action forms a branched intermediate where a single strand of the protospacer becomes ligated to a single strand of the CRISPR array, thereby initiating spacer acquisition. Subsequently, the 3′-OH group on the other protospacer strand performs a second nucleophilic attack on the 5′ end of the opposing DNA strand of the repeat, adjacent to the leader sequence. This results in the expansion of the CRISPR array with a new spacer and a duplicated repeat. Any single-stranded DNA gaps generated at the repeat sequences are filled and repaired by enzymes that have not yet been fully characterized. [adapted from (Amitai & Sorek, 2016)].

2.2.1.5.2 Self vs non-self-recognition.

Self-targeting poses a potential threat for two main reasons. Firstly, adaptation proteins may assimilate the host genome sequence, leading to the targeting of the host's genome (Wei et al., 2015; Yosef et al., 2012). Secondly, the CRISPR array-derived crRNA can pair with the corresponding spacer in the host genome through sequence complementarity, resulting in cell death (Bikard & Marraffini, 2012). Consequently, the CRISPR system must distinguish

TH-3470_166106104

between self and non-self DNA during both interference and adaptation phases. To prevent such an autoimmune response, the CRISPR system employs multiple mechanisms that favor specificity towards foreign nucleic acid. A protospacer-adjacent motif (PAM) is found upstream or downstream of the protospacer in the foreign DNA for type I, Type II, and type V CRISPR-Cas systems (Mendoza & Trinh, 2018). The PAM is a short (2-5 nucleotide) sequence that is required for target DNA breakage during the interference stage (Bolotin et al., 2005; Mojica et al., 2009; Shah et al., 2013). Spacers are preferentially picked during spacer acquisition from protospacers that have a corresponding PAM (Mojica et al., 2009). Although the Cas1-Cas2 complex is sufficient to mediate PAM-dependent spacer acquisition in Type I CRISPR-Cas systems (Yosef et al., 2012), Cas9 is also required for PAM recognition in type II systems (Wei et al., 2015).

2.2.1.5.3 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 an accessory protein known as Cas4. Cas4 plays a crucial role in the adaptation process, as indicated by various studies (Almendros et al., 2019; Kieper et al., 2018; Lee et al., 2019; Makarova, Wolf, Iranzo, et al., 2020). Cas4 shares homology with RecB and possesses four conserved cysteine residues responsible for coordinating an iron-sulfur cluster (Zhang et al., 2012)(Zhang et al., 2012). The protein's nuclease activity is essential for selecting prespacers with the correct Protospacer Adjacent Motif (PAM) and trimming prespacers at the PAM-proximal end before integration. Recent findings in *Bacillus halodurans* (Type I-C) demonstrated a direct interaction between Cas4 and Cas1-Cas2 (Lee et al., 2019). Cryo-Electron Microscopy (Cryo-EM) analysis of the Cas4/I-C structure revealed a complex with a 2:4:2 stoichiometry ($(Cas4)_2$-$(Cas1)_4$-$(Cas2)_2$) (Lee et al., 2019). Notably, a majority of observed particles in Cryo-EM contained a singular Cas4 copy in the complex, denoted as $(Cas4)_1$-$(Cas1)_4$-$(Cas2)_2$. In this configuration, Cas4 is positioned near the PAM region in unprocessed prespacers, suggesting its role in orienting prespacers at the leader-repeat-spacer junction within the CRISPR array (Lee et al., 2019).

2.2.1.5.4 Prime adaptation.

The mutation rate within foreign nucleic acid is notably high, particularly in the PAM and the 'seed' sequence, where a 7-13 nucleotide-long PAM proximal region in prespacer, crucial for interference. Mutations in these regions can result in compromised CRISPR

TH-3470_166106104

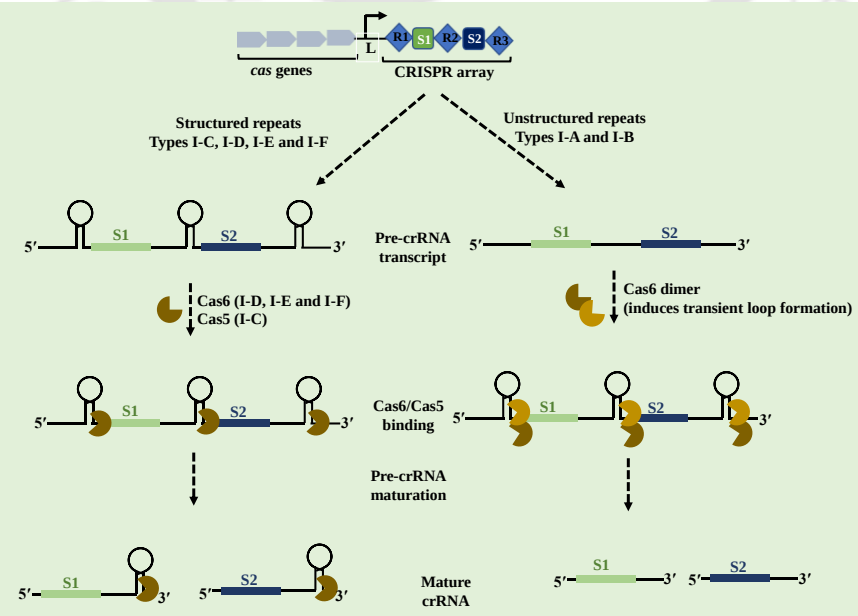
interference (Deveau et al., 2008; Semenova et al., 2011). To counteract the risk of foreign nucleic acid 'escape,' the CRISPR system has evolved a mechanism to enhance spacer acquisition frequency by using existing spacers as bait (Garrett et al., 2020). During this process, multiple spacers are acquired from the priming region of the interference complex, eliciting a robust immune response from the newly acquired spacers (Garrett et al., 2020; Hochstrasser et al., 2014; Sternberg et al., 2016). Moreover, the natural fusion of Cas2 and Cas3 (Cas2-3) in type I-F elucidates the evolutionary advantage of overlapping adaptation and interference stages (Makarova, Wolf, Iranzo, et al., 2020). Primed adaptation commences with the interference complex identifying target nucleic acid. Upon binding to the target DNA, the interference complex recruits Cas3, an active nuclease-helicase, which unwinds and degrades the target DNA into smaller fragments. In *E. coli*, the *in vitro* activity of Cas3 generates fragments of approximately 30 to 100 nt, which may anneal to form suitable prespacers for the Cas1-Cas2 complex (Künne et al., 2016). In a separate context, the conformation of the Cascade complex after binding to cognate or mutated target DNA dictates the prevalence of interference or adaptation, respectively. Mutations in the PAM or 'seed' region lead the Cascade complex to adopt a conformation unfavorable for interference. Instead, Cas1-Cas2 and Cas3 are recruited, forming a primed adaptation complex that facilitates the acquisition of new spacers (Hayes et al., 2016).

2.2.1.6 Maturation phase of CRISPR-Cas system.

A mature crRNA is required for an active CRISPR-Cas effector complex, where it guides the recognition of cognate targets for destruction. Due to the palindromic nature of the repeat sequence, the precursor crRNA (pre-crRNA) adopts a stem-loop structure that binds to invading nucleic acids sequence specifically (Haurwitz et al., 2010; Nam, Haitjema, et al., 2012). Unlike the CRISPR-Cas9 effector, which necessitates both a crRNA and a tracrRNA (trans-acting crRNA) for its activity, all Class 1 complexes feature only a solitary crRNA molecule (Liu & Doudna, 2020). An RNA recognition motif (RRM)-containing protein, such as Cas5d in type I-C systems or a Cas6 homolog for rest type I systems, recognizes and catalyzes crRNA maturation (Li et al., 2013; Nam, Haitjema, et al., 2012; Przybilski et al., 2011; Punetha et al., 2014; Sashital et al., 2011). As a result, Cas6/Cas5d enzymes have evolved diverse ways to handle the barriers of RNA binding and catalysis. All the stems and loop lengths vary widely in size and sequence, serving as the foundation for Cas6-specific binding.

Cas6 proteins recognize and cleave the pre-crRNAs in a single turnover, establishing a stable hairpin structure in each repeat (Haurwitz et al., 2010; Sashital et al., 2011; Sternberg et al., 2016). Similarly, repeats of type I-C from *Bacillus halodurans* create persistent stem-loop structures, which recognized by a unique Cas5 variant (Garside et al., 2012; Koo et al., 2013; Nam, Haitjema, et al., 2012; Punetha et al., 2014) [called as Cas5d, where the prefix "d" alludes to "Dvulg" because a Cas6 protein is missing from this system]. A 3 bp stem and a tetraloop in the hairpin region were identified as a minimum structural requirement for Cas5d recognition and cleavage (Nam, Haitjema, et al., 2012).

In Types I-A (Cas6a), I-B (Cas6b), I-D (Cas6d), I-E (Cas6e), and I-F (Cas6f), crRNA processing generates a mature crRNA containing a single spacer sequence with fragments of the CRISPR repeat on either side. The remaining repeat sequence at the 5′ end of the mature crRNA is known as the 5′ handle or 5′ tag, and in most type I systems, it is 8 nt in length **(Fig. 1.6)** (Brouns et al., 2008; Haurwitz et al., 2010). The leftover repeat segment at the 3′end of the crRNA, which is more variable in length and structure, is sometimes known as the 3′ handle or 3′ stem-loop. Intriguingly, Cas5d has evolved to serve as the dedicated endoribonuclease of type I-C systems, where it generates mature crRNAs with an 11-nt 5′ handle **(Fig. 1.6)** (Plagens et al., 2014; Wiedenheft et al., 2011). Cas5 proteins from other type I systems are non-catalytic and serve as structural subunits in interference complexes. In the type I-A system, further processing by an unknown trimming nuclease removes 3′ portions of the crRNA (Carte et al., 2010; Koo et al., 2013).



TH-3470_166106104

Figure 1.6. Processing of pre-crRNA in Class 1 CRISPR-Cas system. The entire Type I systems feature a combination of structured and unstructured repeat sequences. On the left side, the CRISPR array undergoes transcription, resulting in a lengthy pre-crRNA transcript. The Cas6-family protein or Cas5 (I-C) identifies the stem-loop structure and processes the crRNA by cleaving at the base of the loop. This processing generates a mature crRNA containing a full spacer sequence and a partial repeat region. The interaction of the crRNA with Cas6/Cas5(I-C) facilitates the assembly of additional Cascade proteins. On the right side, the non-palindromic repeats in I-A and I-B undergo a conformational change induced by Cas6, forming a hairpin-like structure. Cas6 recognizes and cleaves within these repeats. Notably, Cas6 functions as a multi-turnover enzyme in this context and is not involved in the formation of the Cascade complex. [adapted from (Hochstrasser & Doudna, 2015)].

Despite limited sequence homology among members of the Cas6 family, numerous crystal structures of Cas6 reveal a shared overall fold and specific structural characteristics crucial for crRNA binding (Hochstrasser & Doudna, 2015). Cas6 enzymes are composed solely of two repeat-associated mysterious protein (RAMP) domains, which adopt ferredoxin-like or RNA recognition motif (RRM) folds, a common trait among Cas proteins (Daniel H. Haft et al., 2005; Hochstrasser & Doudna, 2015; Reeks, Sokolowski, et al., 2013). In comparison, catalytic Cas5 enzymes possess an N-terminal RRM fold, but their C-terminal domain is smaller, comprising only three antiparallel β-strands, two of which form a β-hairpin, along with disordered regions (Hochstrasser & Doudna, 2015). Interestingly, the active site of Cas5C is situated differently from that of Cas6 enzymes, suggesting independent evolution of their catalytic centers (Hochstrasser & Doudna, 2015). Additionally, Cas5C proteins feature a G-loop in their RAMP domain that diverges from the consensus sequence of the Cas6 G-loop (Jackson et al., 2014).

In contrast to Class 1 systems, Class 2 systems employ a single-subunit multidomain effector nuclease and, in certain instances, non-CRISPR host proteins for crRNA maturation (Deltcheva et al., 2011b). Additionally, in type II and V-B systems, a trans-encoded small RNA known as trans-activating crRNA (tracrRNA) plays a crucial role during the maturation process. For example, in type II-A, crRNA processing relies on the tracrRNA:crRNA duplex; however, a functional interference complex can still form in the absence of matured crRNA (Deltcheva et al., 2011b; Makarova, Wolf, Iranzo, et al., 2020). Similarly, type V and VI systems also necessitate an effector nuclease for maturation and interference. Cas12a serves as the effector nuclease in type V-A systems and facilitates tracrRNA-independent maturation of crRNA (Dong et al., 2016). A short complementary untranslated RNA (scoutRNA) is essential

TH-3470_166106104

for maturation in types V-C and V-D (Harrington et al., 2020). Conversely, in type VI, tracrRNA is absent, and Cas13, the effector nuclease, is required for crRNA maturation (Abudayyeh et al., 2016; Makarova, Wolf, Shmakov, et al., 2020).

2.2.1.7 Interference phase of CRISPR-Cas system.

CRISPR-associated complex for antiviral defence (Cascade) of type 1 CRISPR-Cas systems is encoded by multiple Cas proteins. Type I CRISPR-Cas systems are geared towards identifying homologous segments of double-stranded DNA in phages or plasmids for degradation. This targeting mechanism unfolds in two pivotal phases: first, the surveillance complex identifies a complementary target within foreign DNA, and second, the target is cleaved by Cas3 (Hochstrasser et al., 2014; Mulepati & Bailey, 2013). Cascade was first characterise in *E. coli* type I-E CRISPR-Cas effector complex (Brouns et al., 2008) and comprises Cas8e, Cas7, Cas5, and Cas6e, as well as an RNA component, the 61-nt crRNA. Cascade shapes a seahorse-like structural design with a stoichiometry of $Cas8e_1$-$Cas11_2$-$Cas7_6$-$Cas5_1$-$Cas6e_1$ (Jore et al., 2011). As previously stated, pre-crRNA in *E. coli* is processed by the Cas6e ribonuclease and remains tightly bound to the 3′ repeat region of each crRNA after processing (Jore et al., 2011). Simultaneously, six Cas7 proteins run along the crRNA's guide region (spacer), forming the backbone filament, with the uppermost subunit interacting with Cas6. These Cas7 proteins have a right-hand-like shape (Mulepati et al., 2014), and the subunits are linked via interactions between the "thumb" and "fingers" domains of one subunit and the adjacent subunits (Jackson et al., 2014). Cas5 has a right-hand-like shape but lacks the "fingers" domain, allowing Cas5 to cap the Cas7 filament at the 5′ end of the crRNA by interacting with Cas7 subunits. Two Cas11 proteins form a dimer in the seahorse-like complex that does not directly bind with crRNA but serves as a bridge between the complex's head and tail. One Cas11 subunit, together with Cas5, contacts Cas8 at the complex's bottom, while the other Cas11 subunit interacts with Cas6 at the complex top (Jackson et al., 2014).

The developed strategies for avoiding self-CRISPR targeting in these systems involve a sequence immediately surrounding the protospacer, PAM, which is fundamentally required for Cascade to detect genuine DNA targets. The L1 loop, positioned within the N-terminal domain of Cas8, directly interacts with the PAM in the *E. coli* type I-E system, mediating selective binding of Cascade to the PAM-containing DNA target (Sashital et al., 2011). Further structural investigation revealed a glutamine wedge, a glycine loop, and a lysine finger were

TH-3470_166106104

essential for PAM identification by Cas8 and specified PAM interaction with the target strand (Hayes et al., 2016)(Hayes et al., 2016). A target DNA is determined by genuine PAM recognition and base pairing between a crRNA and the corresponding protospacer (Zheng et al., 2020).

The crRNA base pairs with the protospacer within the seed region and then throughout the matching sequences, finally dislodging the non-target strand and producing a full R-loop. The initial dsDNA binding, on the other hand, causes a substantial conformational shift in the Cascade, which may lead to the recruitment of Cas3 to the Cas8 docking site (Hochstrasser et al., 2014; Wei et al., 2015; Westra et al., 2012). Cas3 has an N-terminal HD phosphohydrolase domain and a C-terminal superfamily 2 helicase domain (Makarova et al., 2006). When the crRNA displaces the non-target strand, it is exposed and given over to the Cas3 nuclease for nicking and sequential degradation in the 3′ to 5′ direction **(Fig. 1.7)** (Gong et al., 2014; Sinkunas et al., 2013). Complete degradation of target DNA into ssDNA might not result in this reaction. Cas3 nuclease domain or other host nucleases, with the help of the Cas3 ATP-dependent helicase domain, leads to unwinding of the DNA target and its degradation (Brouns et al., 2008; Mulepati et al., 2014).

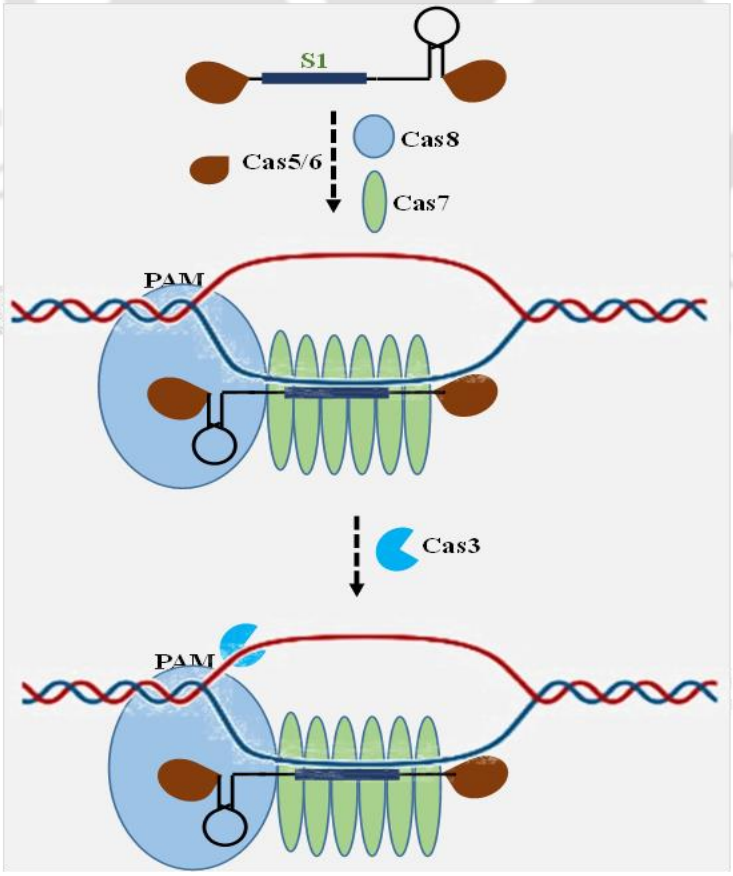


Figure 1.7. Schematic depiction of the maturation and interference by CRISPR-Cas type 1. In type 1, Cas5C/Cas6 proteins are responsible for processing pre-crRNA into mature crRNA, with Cas5C/Cas6 (depicted in brown) remaining bound at both ends. This serves as a scaffold for the assembly of the Cascade complex. The interference complex resembles a seahorse architecture, comprising various effector Cas proteins. Six Cas7 proteins (illustrated in green) assemble around the spacer, forming a helical backbone. A single Cas8 protein (shown in blue) caps the 5′ end of the mature crRNA. Cas8 proteins identify the PAM sequence in the invading nucleic acids. Subsequently, the crRNA's base pairing with the target invading nucleic acids induces R-loop formation. Cas3 binds to the single-stranded DNA of the non-target strand and cleaves the foreign nucleic acid in a 3′ to 5′ direction, employing its nuclease and helicase activities [adapted from (Li et al., 2013; Liu & Doudna, 2020)].

The type III Interference complex exhibits a Cascade-like structure (referred to as Csm for III-A and Cmr for III-B) and shares significant architectural and compositional similarities with the type I interference complex (Hochstrasser et al., 2014; Jackson et al., 2014). However, unlike type I, type III systems have the capability to target both invading RNA and DNA of mobile genetic elements (Deng et al., 2013). Following crRNA maturation, Cas5 binds to a 5′ repeat-derived handle, serving as a scaffold for the assembly of other Cas proteins. The backbone of the complex is formed by Cas7 family proteins (Csm3 and Csm5 in III-A, and Cmr1, Cmr4, and Cmr6 in III-B), which grip onto 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) constitute large and small subunits, respectively (Osawa et al., 2015; Taylor et al., 2015). Upon binding to a nascent target RNA through crRNA-mediated complementary base pairing, the Cas7 family subunits of the interference complex cleave the target RNA after every sixth base (Osawa et al., 2015; Taylor et al., 2015). In addition to cleaving the target, when bound to its target, Cas10 also catalyzes the conversion of ATP to cyclic adenylate. This cyclic adenylate serves as an activator for RNase Csm6, which subsequently initiates the degradation of nearby RNA transcripts that are nonspecific in nature (Kazlauskiene et al., 2017).

In type II systems, the interference complex comprises Cas9, crRNA, and a tracrRNA. Cas9 exhibits a bilobed structure consisting of an alpha-helical recognition lobe (REC) and a nuclease lobe (NUC), connected via a linker housing an arginine-rich helix. The positively charged bridge helix interacts directly with the crRNA. Within the NUC lobe, there are HNH and RuvC domains responsible for cleaving the target DNA during interference (Jiang & Doudna, 2017). The functional role of the effector nuclease Cas12 varies across type V systems. Cas12a (type V-A) operates independently of tracrRNA, while active Cas12b (type

TH-3470_166106104

V-B) relies on tracrRNA for its function. Similar to Cas9, Cas12 possesses a bilobed structure comprising REC and NUC lobes (Dong et al., 2016; Makarova, Wolf, Iranzo, et al., 2020). However, it contains only a single catalytic RuvC domain. When the target binding is perfect, it ensures the precise positioning and activation of the RuvC domain, leading to cleavage of both the target and non-target strands (Yang et al., 2016). The effector nuclease Cas13 is a protein containing High-Efficiency Prokaryotic Nuclease (HEPN) domains, specialized for cleaving single-stranded RNA. Similar to Cas9 and Cas12, it exhibits a bilobed structure, with two HEPN domains located in the nuclease (NUC) lobe. Upon binding to target ssRNA, the HEPN domains become activated, leading to the cleavage of both the target ssRNA and collateral RNA (Abudayyeh et al., 2016; Anantharaman et al., 2013).

2.2.1.8 Phage-based defence against the CRISPR-Cas system.

According to the co-evolution theory, when an organism develops a new method to beat a parasite and prevent extinction, the parasite responds by developing appropriate countermeasures to evade the host organism's resistance (Pawluk et al., 2018; Safari et al., 2020). Bacteria prevent phage invasion by stopping phage adsorption (i.e., suppressing phage receptor function), restricting phage DNA entrance, and cleaving phage nucleic acids (e.g., through the restriction-modification system). As a result, phages developed new combat techniques that included revising receptor-binding proteins and removing R-M system recognition sites from their genome (Labrie et al., 2010; Roberts et al., 2003; Samson et al., 2013; Xu et al., 2010). The CRISPR-Cas systems in prokaryotic cells elicited customised responses from phages to re-establish the required circumstances for bacteria and archaea to proliferate. Initially, single nucleotide changes or deletions in the conserved PAM or seed area along the protospacer allowed phages to survive (Deveau et al., 2008; Labrie et al., 2010; Semenova et al., 2011). In this environment, phages apprised their DNA to avoid being targeted by the CRISPR-Cas immune system (Samson et al., 2013), and anti-CRISPR proteins (Acrs) appeared (Pawluk et al., 2018). Anti-CRISPRs are small proteins (around 12-193 amino acids) that provide phages with an efficient and potent means of suppressing the CRISPR-Cas immune system (Liu et al., 2020). At present, 30 different families of anti-CRISPR proteins for type I and type V-A systems have been discovered (Q. Liu et al., 2020). Thirteen of them exerted an inhibitory activity on CRISPR-Cas, either by preventing the Cascade/Cas-crRNA

TH-3470_166106104

complex from joining the substrate DNA or by impeding DNA cleavage after inactivating the Cas effector (Q. Liu et al., 2020).

Anti-CRISPR-Cas proteins such as AcrIE1 and AcrIF3 inhibit type I immune activity by inhibiting the Cas3 helicase-nuclease complex from recruiting Cas3 helicase-nuclease. Cascade-crRNA binds the substrate DNA stably without causing DNA degradation when AcrIE1 or AcrIF3 is expressed. As a result, AcrIE1 and AcrIF3 can influence gene transcription (Luo et al., 2015; Marino et al., 2020). AcrIF1 and AcrIF8 target the Cas7f subunits, whereas AcrIF10 binds Cas5f-Cas8f. AcrI with comparable inhibitory actions can bind identical Cascade subunits. In type I-F, CRISPR-Cas AcrI proteins positioned on Cas8f tail (e.g., AcrIF10 and AcrIF2), mimic the DNA and prevent PAM recognition by the Cascade-crRNA complex. AcrVA1, AcrVA4, and AcrVA5 use different approaches to eliminate Cas12a-crRNA in type V CRISPR-Cas systems. AcrVA1 is similar to AcrIF10 in that it competes with crRNA for access to PAM. AcrVA1 does, however, exhibit RNase activity, which results in crRNA truncation. AcrVA4, conversely, causes structural alterations in Cas12a proteins to prevent DNA cleavage. Finally, AcrVA5 functions as an enzyme that, via a covalent alteration on Cas12a, permanently eliminates CRISPR-Cas12a nuclease activity (Marino et al., 2020; Yu & Marchisio, 2020).

2.2.1.9 Application of CRISPR-Cas system.

Prior to the advent of the CRISPR/Cas system, also recognized as the third-generation genome editing tool, gene-editing tools like ZFNs (Zinc Finger Nucleases) and TALENs (Transcription activator-like effector nucleases) were commonly employed (Ramalingam et al., 2014). However, these two technologies had several limitations, including limited editability, a high off-target rate, elevated cytotoxicity, increased costs, prolonged time consumption, and heightened labour demands. In comparison to TALENs and ZFNs, the CRISPR/Cas system stands out for its simpler design, lower cost, enhanced targeting efficiency, reduced off-target rate, decreased cytotoxicity, and the ability to edit numerous genes *in vitro* or *in vivo* (Bharathkumar et al., 2022). Due to these advantages, the CRISPR/Cas system has emerged as a powerful genome editing tool in the realm of molecular biology. Moreover, with the ongoing optimization of cleavage elements in the CRISPR/Cas9 system, it is anticipated that CRISPR technology, based on CRISPR/Cas9, may ultimately take precedence in the future of gene editing, even though TALEN technology is currently more prevalent in clinical applications

TH-3470_166106104

(Huang et al., 2022). CRISPR-Cas systems operate based on the spacer sequence present in the crRNA, providing guidance and sequence specificity (Makarova & Koonin, 2015). Upon binding to the target sequence adjacent to the 2-5 nucleotide PAM, the Cas nuclease initiates cleavage of the target nucleic acid (Hayes et al., 2016). This means that achieving sequence-specific cleavage at any location near the PAM is feasible using CRISPR-Cas systems by designing a unique crRNA or engineered guide RNA (gRNA) with an appropriate spacer sequence (Marraffini & Sontheimer, 2008). The application scope of CRISPR/Cas technology is extensive, encompassing research, development, and diagnostics. Its primary utility lies in modifying target genes (DNA/RNA) within the host as a genetic tool. Additionally, CRISPR/Cas technology serves diverse functions in Genome library screening, Transcriptional regulation (activation/repression), epigenetic modification, live cell imaging, and therapeutic applications (Bharathkumar et al., 2022). Significant strides in gene-editing technology in recent years have led to remarkable progress in enhancing human health. Gene editors are currently being utilized in active clinical trials to address a range of human disorders, including HIV, cancer, and hematologic conditions, showcasing the impactful achievements in this field.

While CRISPR/Cas9 stands as a promising method, its recent identification and application in humans pose challenges for integration into clinical studies (Morshedzadeh et al., 2024). Primary difficulties include immunogenicity, off-target effects, mutations, delivery techniques, and ethical considerations associated with CRISPR/Cas9 technology (Morshedzadeh et al., 2024). The Cpf1 system, in the CRISPR family, emerges as a potent tool for genome modification, distinct from the Cas9 system (Zetsche et al., 2015). Notably, Cpf1 is a singular RNA-guided endonuclease which recognizes a PAM resulting in staggered cuts distal to the PAM site (Zetsche et al., 2015). Additionally, Cpf1 exhibits dual functionality, capable of processing its own crRNA and cleaving the target DNA (Safari et al., 2019). Another distinctive feature is that the DNA cleavage in the CRISPR/Cpf1 system is guided solely by crRNA, eliminating the need for tracrRNA (Zetsche et al., 2015). The CRISPR/Cpf1 system's adeptness in processing crRNA with direct repeats makes it particularly well-suited for multiplex gene editing (Chen et al., 2023).

2.2.1.10 Webservers to predict CRISPR-Cas system in prokaryotic genome.

Despite experimental confirmation that CRISPR-Cas systems primarily target phages or plasmids, the specific targets of the majority of spacers remain unidentified (Barrangou et

al., 2007). For instance, among the 926 spacers identified for *E. coli* and *Salmonella*, Touchon and Rocha were only able to predict the likely targets of 8%, and a parallel study discovered the targets of 12% of spacers (Díez-Villaseñor et al., 2010; Touchon & Rocha, 2010). Several factors contribute to the difficulty in identifying crRNA targets. This challenge stems, in part, from the limited number of studies dedicated to exploring the sequences of phages in relation to their abundance and genetic diversity (Hatfull & Hendrix, 2011; Krupovic et al., 2011). Additionally, accessing phage sequences in databases like GenBank is often challenging, but a wealth of such sequences exists in viral metagenome or virome studies (Culley, 2011). CRISPRTarget predict the most likely targets of CRISPR RNAs. CRISPRTarget, can be utilized to identify targets in newly sequenced genomic or metagenomic data (Biswas et al., 2013).

CRISPROne systematically investigates genomic locations of CRISPR-Cas systems in both complete and draft bacterial genomes to quantify the frequency of co-occurrence between CRISPR arrays and *cas* genes. Additionally, it explores the potential reasons behind the presence of isolated *cas* loci using representative species. Within the isolated CRISPRs lacking companion *cas* genes, a subset was identified as highly suspicious due to their absence of spacer diversity, making them unlikely to be authentic CRISPRs. These suspicious CRISPRs, lacking spacer diversity and thus less likely to be genuine, were termed false-CRISPR elements. It also highlighted the possibility of confusion between some tandem repeats and CRISPRs, particularly those containing structures resembling "repeat-spacer" configurations (Zhang & Ye, 2017).

Cas protein identification can be accomplished using either BLAST or Hidden Markov Model (HMM) searches, as exemplified in tools like CasFinder. The precise identification of CRISPR-Cas systems poses a challenge during the analysis of new genomes. Accessible and regularly updated online databases prove to be invaluable resources for exploring the diversity of CRISPR-Cas systems. The pioneering CRISPRFinder program, with its associated database CRISPRdb, was the first website dedicated to these structures. Recently, a new tool and database, CRISPRCasFinder, have been developed, combining the functionalities of CRISPRFinder and CasFinder to identify both CRISPR arrays and *cas* genes in submitted sequences (Grissa et al., 2007; Pourcel et al., 2005; Pourcel et al., 2020). CRISPRDetect is a program suitable for local or online use in genome sequence analysis (Biswas et al., 2016),

TH-3470_166106104

while CRISPRdisco facilitates the discovery and analysis of CRISPR-Cas systems through local program installation (Crawley et al., 2018). CRISPRCasdb provides a detailed description of the database constructed with CRISPRCasFinder.

3. Leptospira interrogans

Leptospira, a pathogenic spirochete, is a causative agent for the zoonotic disease known as Leptospirosis, a significant global health concern (Adler & de la Peña Moctezuma, 2010; Bharti et al., 2003). Over the past decade, outbreaks linked to sporting events, adventure tourism, and natural disasters have highlighted how leptospirosis can pose a significant public health threat in unconventional settings (Bharti et al., 2003; Haake & Levett, 2015; Santos et al., 2010). Despite this, leptospirosis remains a neglected disease, disproportionately affecting impoverished populations in developing countries and tropical regions (Santos et al., 2010). Historically endemic among subsistence farmers, leptospirosis has also become a major issue in urban slums, where poor sanitation fosters conditions for rat-borne transmission (Reis et al., 2008; Riley et al., 2007). An estimated 1 million cases occur globally each year, resulting in nearly 60,000 deaths (Costa et al., 2015). In India, fewer than 10,000 cases are reported annually, while the estimated total ranges between 0.1 and 1 million. Most cases typically come from Gujarat, Maharashtra, Tamil Nadu, and Kerala (Agrawal et al., 2018; Kumar, 2013).

These spirochetes are helically coiled and exhibit motility, measuring approximately 6-25 μm in length and 0.1 to 0.2 μm in diameter (Bharti et al., 2003). *Leptospira* are mobility and obligate aerobes capable of surviving in diverse environments, such as soils, mud, swamps, streams, and rivers, as well as within the organs and tissues of living or deceased animals and even diluted milk (Adler & de la Peña Moctezuma, 2010; Bharti et al., 2003; Faine, 1999). *Leptospira* can be distinguished from other spirochaetes based on their unique hook or question mark-shaped ends (Li et al., 2001). Under *in vitro* conditions, *Leptospira* exhibits slow growth on solid and liquid media supplemented with long-chain fatty acids, ammonium salts, vitamins B1, and B12, typically at temperatures ranging from 28 to 30 °C. The most commonly employed medium for cultivation is the Ellinghausen-McCullough/Johnson-Harris (EMJH) medium, which includes oleic acid, bovine serum albumin, and polysorbate (Adler & de la Peña Moctezuma, 2010; Faine, 1999; Murray et al., 2015). Leptospires are not facultative intracellular organisms; they are seldom seen within host cells and typically only pass through them transiently when crossing cell monolayers *in vitro*. Infection results in prolonged

TH-3470_166106104

leptospiraemia until the host's acquired immune response becomes effective, which generally occurs one to two weeks after exposure (Barocchi et al., 2002).

3.1 The cycle of leptospiral infection.

Leptospira transmission occurs via direct contact with infected animals and bodily fluids, primarily urine, or indirectly through contact with contaminated water (Bharti et al., 2003; Sharma, 2008). *Leptospira* resides in the proximal renal tubules of the kidneys in both carriers and infected mammals, leading to intermittent shedding through urine over months or even a lifetime (Faine, 1999). Primary hosts include rats, mice, and moles, while a wide range of other mammals, such as dogs, deer, rabbits, hedgehogs, cows, sheep, raccoons, opossums, skunks, and certain marine mammals, serve as secondary hosts (Faine, 1999). Humans are accidental hosts for *Leptospira spp*., which can cause acute and sometimes fatal infections (Ko et al., 2009). Occupations involving frequent animal contact, such as farmers, abattoir workers, butchers, veterinarians, pet traders, hunters, and rodent control workers, are particularly susceptible to leptospirosis (Hartskeerl et al., 2011; Musso & La Scola, 2013).

Leptospira enters the body through cuts in the skin or mucous membranes of the eyes, nose, and throat, with an incubation period of approximately 2 to 20 days. Clinical symptoms vary depending on the *Leptospira* strain or serovar involved and the inoculum size, which varies among individuals based on age and health (Adler & de la Peña Moctezuma, 2010; Ko et al., 2009). In most human cases, leptospirosis symptoms are mild and self-limiting. However, the more severe Weil's syndrome presents with multi-organ complications, including jaundice, hepatic and renal dysfunction, cardiovascular collapse, pulmonary haemorrhage, and meningitis (Bharti et al., 2003). Additional symptoms encompass fever, nausea, headache, myalgia, chills, and skin rashes (Faine, 1999; Vinetz, 2000). Leptospirosis is often misdiagnosed in humans due to its symptom overlap with other illnesses, such as dengue fever, influenza, hepatic disease, and Hantavirus infections, leading to the underestimation of its occurrence **(Fig. 1.8)** (Hartskeerl et al., 2011; Victoriano et al., 2009).

TH-3470_166106104

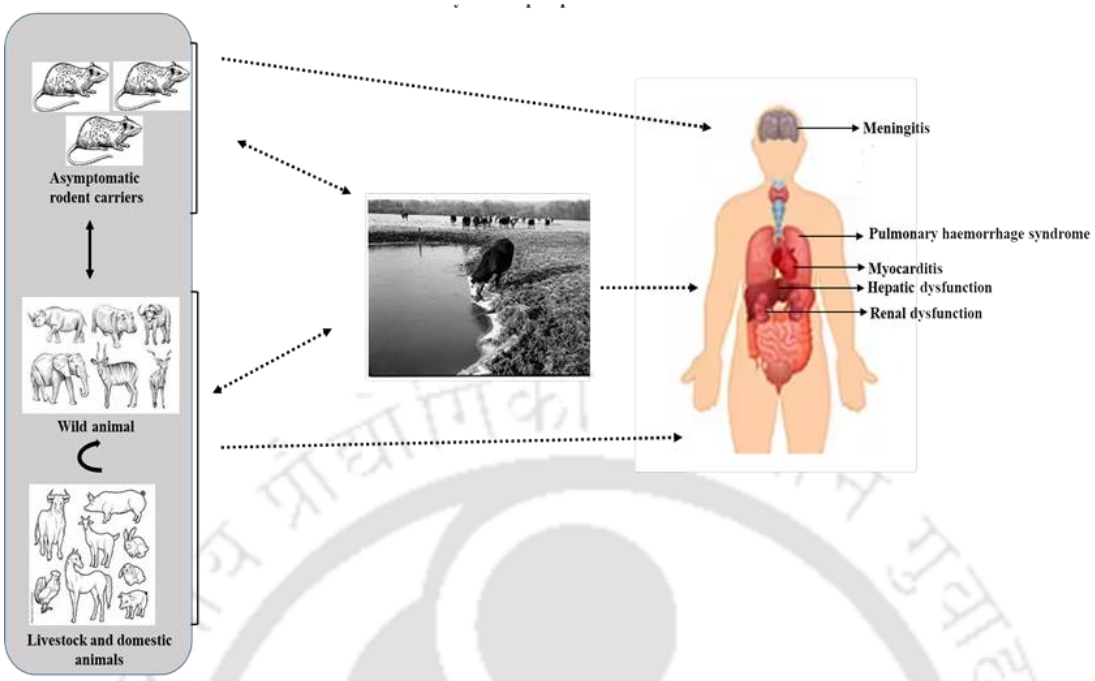


Figure 1.8. The cycle of leptospiral infection. Mammals excrete leptospiral pathogens through their urine, which serves as a reservoir for the transmission of pathogens. These pathogens are sustained in wild (sylvatic) and domestic environments through transmission among various rodents. Infection in these reservoirs results in a chronic, asymptomatic carriage of the pathogens. Leptospires can next infect livestock, domestic animals, and wildlife, leading to a manifestation of the spectrum of disease and their carrier. The sustained presence of leptospirosis in these populations is attributed to exposure to rodent reservoirs or communication within animal herds. In humans, Leptospirosis is transmitted either by direct contact with reservoir animals or through exposure to environmental surface water or soil contaminated with their urine. Leptospires can enter the body through abraded skin or mucous membranes, enter the bloodstream, and spread throughout the tissues. During the initial "leptospiraemic" phase, infection causes an acute febrile illness and progresses during the subsequent "immune" phase, causing severe multisystem manifestations such as hepatic dysfunction, jaundice, acute renal failure, pulmonary haemorrhage syndrome, myocarditis, and meningoencephalitis. While the immune reaction ultimately eradicates the infections, leptospires can survive for long periods in immune-privileged locations such as the renal tubules the anterior chamber, and the vitreous humor of the eye. This persistence might result in urine shedding weeks after the sickness has resolved and uveitis months after exposure. It is vital to emphasise that humans are unintentional hosts who do not excrete enough leptospires to act as reservoirs for subsequent transmission [(adapted from (Ko et al., 2009)].

3.2 Classification of Leptospira.

Initially, morphological characteristics served as the basis for classifying organisms within the *Leptospira* genus (Picardeau, 2013, 2020). Subsequently, the genus was categorized into pathogenic and non-pathogenic species, Initially, morphological characteristics served as the basis for classifying organisms within the *Leptospira* genus (Picardeau, 2013, 2020). Subsequently, the genus was categorized into pathogenic and non-pathogenic species. There are more than 68 *Leptospira* species; however (Arent et al., 2022; Caimi & Ruybal, 2020; Vincent et al., 2019), when considering serological classification based on agglutinating lipopolysaccharide antigens, *Leptospira* spp. are divided into 20 serogroups, encompassing over 300 serovars (Picardeau, 2013, 2020). Pathogenic *Leptospira* serovars exhibit host preferences, such as serovar Hardjo for cattle, Canicola for dogs, and Icterohaemorrhagiae for rats (Picardeau, 2017). However, these associations are not absolute, and the underlying genetic mechanisms for host specialization remain unclear. DNA-DNA hybridization (DDH) and phylogenetic analysis of 16S rRNA have identified three distinct clades within the *Leptospira* genus (Picardeau, 2017). One clade encompasses pathogens capable of infecting and causing diseases in both humans and animals, including species like *L. interrogans, L. noguchii, L. weilii, L. kirschneri, L. santarosai, L. mayottensis, L. kmetyi,* and *L. borgpetersenii*. Another group consists of intermediates isolated from humans and animals that may induce moderate symptoms of leptospirosis, such as *L. inadai, L. broomi, L. fainei, L. licerasiae,* and *L. wolffi*. The third clade comprises seven saprophytic species that cannot cause disease, including *L. biflexa, L. meyeri, L. wolbachii, L. terpstrae, L. vanthielii,* and *L. yanagawae* (Picardeau, 2017).

3.3 Defense systems in *Leptospira*.

Leptospira employ various defense mechanisms to counteract threats. Among these, the restriction-modification (R-M) system is notable. The most prevalent R-M systems are type IV, type I, and type II. Type III R-M is absent from all the examined strains (Petakh et al., 2024). Unlike innate immune systems that modify DNA based on fixed interactions with defense proteins in phage genomes, prokaryotes also utilize adaptive immune systems like CRISPR-Cas. For *Leptospira*, the dominant CRISPR types are CRISPR type I (which is present in most pathogenic species and includes subtypes I-B, I-C, and I-E) and CRISPR type III (Petakh et al., 2024). Among the antiphage systems, PrrC, an anticodon nuclease targeting tRNALys, is the most common in *L. interrogans* strains. PrrC supports the R-M system by

TH-3470_166106104

degrading tRNALys and inhibiting phage replication (Huiting & Bondy-Denomy, 2023; Kaufmann et al., 1986; Sirotkin et al., 1978). It is categorized as an abortive infection system due to its interference with the host's translation machinery (Huiting & Bondy-Denomy, 2023; Penner et al., 1995). The Shedu antiphage system is also observed in half of the Canicola strains. This system features a single protein, SduA, which functions as a nuclease with a conserved DUF4263 domain from the PD-(D/E)XK nuclease superfamily (Gu et al.). The PD-T4-7 system, found in 37.5% of Grippotyphosa strains, is a single-gene system operating through an abortive infection mechanism (Vassallo et al., 2022). Among 22,803 complete RefSeq genomes, PD-T4-7 is identified in 155 (0.68%). The Dsr system, also present in 25% of Grippotyphosa strains, comprises two subsystems (Petakh et al., 2024). Recently, an *in silico* study identified four vapBC loci in the genome of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130, which were numbered based on their genomic organization (Lopes et al., 2014). Of these loci, only one designated VapBC-3 had been characterized (Damiano et al., 2024; Lopes et al., 2014). Consequently, the remaining three loci were classified as putative toxin-antitoxin (TA) modules and required further investigation (Damiano et al., 2024). The frequently observed defense system is Borvo, a single-gene antiphage system identified through both bioinformatic prediction and experimental validation (Millman et al., 2022)[41]. Although Borvo is likely an abortive infection mechanism, the exact molecular mechanism by which it operates has not yet been elucidated (Millman et al., 2022; Stokar-Avihail et al., 2023).

3.4 Genomic Differences in Leptospira Species in Evolution from Saprophyte to Infectious Pathogen.

Leptospira strains exhibit genomic diversity with sizes ranging from 3.8 to 4.6 Mb and GC content between 35-45%, underscoring the considerable variability within the genus (Fouts et al., 2016; Xu et al., 2016). Recent insights into the evolutionary trajectory of *Leptospira* spp. have been gleaned through the reconstruction of gene gain and loss over time (Xu et al., 2016). The predominant trend in the evolution from a saprophytic progenitor to intermediate and pathogenic forms involves the gain or loss of ancestral genes, particularly those encoding proteins involved in metabolism and signal transduction systems. Horizontal gene transfer and strain-specific duplications have played pivotal roles in the adaptive strategies of pathogenic *Leptospira spp.* to diverse hosts (Picardeau, 2017). Computational biology analyses have unveiled a shared evolutionary relationship among all known Leptospira species (Fouts et al.,

TH-3470_166106104

2016). Pathogenic *Leptospira spp.* demonstrate distinctive genomic pathways associated with sialic acid and cobalamin (B12) biosynthesis, recognized as bacterial virulence factors. Additionally, they exhibit pathogen-specific porphyrin metabolism, indicative of their adaptation to the mammalian system. Notably, the CRISPR-Cas system, a defence mechanism against foreign nucleic acids, is prevalent in infectious members of the *Leptospira* genus (Fouts et al., 2016).

3.5 Genetic manipulation in Leptospira.

Genetic manipulation in pathogenic *Leptospira* species is still in its early stage, and the reason for *Leptospira's* genetic inaccessibility, possibly involving the CRISPR-Cas system, remains unknown. Homologous recombination is the common technique for targeted genetic knockout or random insertion (Croda et al., 2008). Transcription activator-like effectors (TALE) have also been employed for exact gene silencing in both saprophytic and pathogenic strains, albeit at high cost and with labor-intensive efforts (Pappas et al., 2015). The progress of studying genetic manipulation in pathogenic *Leptospira* has been hindered due to the lack of naturally occurring plasmids and the slow growth of the bacteria in both solid and liquid media (Bulach et al., 2006; Nascimento et al., 2004; Picardeau et al., 2008). The lack of plasmids in the genomes of pathogenic leptospires highlights the necessity to create a replicative vector with a wide host range, suitable for genetic experimentation purposes. Upon identifying the bacteriophage LE1 replicating as a circular plasmid in *L. biflexa*, a shuttle vector was formulated for *E. coli-L. biflexa* (Saint Girons et al., 1990). This vector incorporates the LE1 origin of replication and antibiotic resistance markers (Girons et al., 2000). Through electroporation, this DNA can be introduced to leptospires and replicate within saprophytic *L. biflexa*, but not in pathogenic species (Girons et al., 2000; Ko et al., 2009). A recent display of conjugation between *E. coli* and *Leptospira spp.* utilizing the RP4 plasmid indicates a potential alternative approach for introducing DNA into leptospires (Picardeau, 2008). Significant advancements have been achieved in the genetic manipulation of *Leptospira spp*, with the development of a shuttle vector, pMaORI (Pappas et al., 2015). This vector can replicate within saprophytic, intermediate, and pathogenic leptospires. The construction of the shuttle vector involved the insertion of a 2.9-kb DNA segment, encompassing the parA, parB, rep genes, aadA cassette, ori R6K, and oriT RK2/RP4 into pMAT vector backbone (Pappas et al., 2015).

TH-3470_166106104

Recently, CRISPR interference techniques have been employed for genetic manipulation of *Leptospira.* The recognition of a 30 bp protospacer adjacent motif (PAM) by Cas9, followed by RNA-DNA Watson and Crick base pairing, results in DNA cleavage. This process creates double-strand breaks (DSBs) that must be repaired for the cell to survive (Lieber, 2010). Most bacteria cannot repair double-strand breaks (DSBs) without a template for recombination, and Cas9 cleavage has been shown to be lethal to *Leptospira* (Fernandes et al., 2019; Fernandes et al., 2021). The first strategy to mitigate this lethality in both saprophytic and pathogenic strains involved expressing a catalytically inactive Cas9 (dCas9) along with a single-guide RNA (sgRNA) designed to pair with the coding strand of the target gene (Fernandes et al., 2019; Fernandes et al., 2021). The dCas9–sgRNA complex acts as a physical barrier, obstructing RNA polymerase elongation and leading to gene silencing instead of gene disruption (Zheng et al., 2017). The second strategy is to employ bacterial DSBs: nonhomologous end-joining (NHEJ) system. Certain bacteria, including *Mycobacterium tuberculosis, M. smegmatis, Pseudomonas aeruginosa, Bacillus anthracis*, and *B. subtilis*, express a more streamlined version of this pathway, consisting of two essential proteins: an ATP-dependent ligase (LigD) and a DNA end-binding protein (Ku) (Matthews & Simmons, 2014). In this context, the NHEJ system of *M. tuberculosis* and *M. smegmatis* has been demonstrated to function in some bacteria (Zheng et al., 2017). However, the use of a heterologously expressed NHEJ system in *Leptospira* spp. has not yet been investigated. The lethality caused by double-strand breaks (DSBs) in *L. biflexa* was overcome by the simultaneous expression of the *S. pyogenes* CRISPR/Cas9 and M. tuberculosis NHEJ systems. This resulted in null phenotypes for LipL32, consistent with the indel mutations observed in the LipL32 genes (Fernandes & Nascimento, 2022). The CRISPRi (dCas9) approach was employed to generate mutants of the key proteins LipL32, LipL41, LipL21, and OmpL1 in *L. interrogans* serovar Copenhageni. Silencing LipL32 resulted in increased bacterial virulence, while LipL41 silencing led to a slight attenuation of acute disease symptoms. Silencing LipL21 rendered the leptospires incapable of causing acute disease, whereas silencing OmpL1 was lethal to both saprophytic and pathogenic Leptospira, highlighting its essential role in the basic biology of the bacteria (Fernandes et al., 2023).

The novel CRISPR prime editing (PE) strategy was first introduced to edit human cells and has since been applied to various model organisms, including mouse embryos, organoid lines, plants, and *E. coli* (Geurts et al., 2021; Shuai Jin et al., 2021; Y. Liu et al., 2020; Park et

al., 2021; Tong et al., 2021). CRISPR-PE is the first precise genome-editing technology capable of performing deletions, insertions, and base substitutions without creating double-strand breaks (DSBs) (Zhao et al., 2023). This is achieved by using a nickase Cas9 (Cas9n) that cuts only a single strand of DNA. In CRISPR-PE, Cas9n is fused with an engineered reverse transcriptase (RT), and the system employs a modified sgRNA, known as prime editing guide RNA (PEgRNA), which contains a primer-binding sequence (PBS) and a reverse transcription template (RTT) (Geurts et al., 2021). The PEgRNA not only enables target recognition through protospacer base-pairing but also guides the incorporation of the desired edits into the DNA target as specified by the RTT sequence (Fernandes et al., 2024). The use of CRISPR-PE in both saprophytic (mutation in β-galactosidase) and pathogenic *Leptospira* (mutation in LipL32) species has enabled the introduction of deletions or insertions into target DNA with remarkable precision, down to a single nucleotide. This marks not only the first application of the innovative CRISPR-PE technique to develop knock-out strains across different species and serovars of *Leptospira* spp., but also the first instance of this tool being used in spirochetal species (Fernandes et al., 2024).

However, applying the endogenous CRISPR-Cas system of pathogenic *Leptospira* strains as a genome editing tool remains a future challenge, making it crucial to understand *Leptospira's* CRISPR-Cas system, as it may eventually enable genetic manipulation in these strains.

4. Research gap

L. interrogans sv. Copenhageni genome encodes two CRISPR-Cas system (subtype I-B and I-C) (Makarova & Koonin, 2015; Xiao et al., 2019). The *cas* genes of subtype I-B are aligned in one direction and are located on either side of the CRISPR array in two different operons (Dixit et al., 2016). Adaptation Cas proteins (LinCas1B, LinCas2B, and LinCas4B) of CRISPR-Cas I-B exhibit metal-ion dependent DNase activity on a variety of DNA substrates. LinCas1B and LinCas2B are inert on small oligo (23 and 50 mer) whereas LinCas4B completely degrade them (Dixit, Anand, et al., 2021b; Dixit et al., 2016; Dixit, Prakash, et al., 2021). LinCas6B cleaves its cognate repeat RNA upstream to 8 nt from its 3′ end in a single turnover mode (Prakash & Kumar, 2021). Additionaly, LinCas6B canonical processing its cognate pre-crRNA transcript into mature cRNA under *in vitro* conditions (Prakash & Kumar, 2021). LinCas7B exhibits modulated activity against DNA and RNA substrates in the presence

TH-3470_166106104

of Mg^{2+} ions. Consistent with its established role in CRISPR-Cas immunity, LinCas7B selectively binds to mature crRNA exclusively when Mg^{2+} ions are present. In the absence of divalent metal ions, LinCas7B functions as a non-specific RNase. Notably, among proteins categorized within the Cas7 family, LinCas7B stands out for its Mg^{2+} ion-dependent DNase activity, showcasing its unique characteristics (Hussain & Kumar, 2022). CRISPR-Cas I-B interference machinery confers PAM-dependent target plasmid interference in the heterologous host (*E. coli*). LinCas11B co-translates within LinCas8B ORF exhibits non-specific affinity for nucleic acids and is indispensable for the LinCascade-mediated target DNA degradation (Hussain et al., 2023).

CRISPR-Cas systems encompass two essential genetic elements: the CRISPR array and *cas* genes responsible for encoding Cas proteins (Makarova & Koonin, 2015). Conventionally, the *cas* genes are situated in the vicinity of CRISPR loci (Daniel H Haft et al., 2005). Nevertheless, in an intriguing departure from this norm, the CRISPR-Cas I-C system identified in *L. interrogans* sv. Copenhageni contradicts this expectation by lacking a CRISPR array element (Makarova et al., 2011b). Hence, the efficacy of a CRISPR-Cas system devoid of a CRISPR array in an organism raises questions about its functionality. Consequently, investigating the functional activity of each *cas* genes becomes crucial. This study aims to explore whether maturation and interference *cas* genes can cleavage non-cognate CRISPR arrays for immunity. Additionally, the inquiry extends to whether endogenous CRISPR-Cas types I (I-B and I-C) can be harnessed for genome editing purposes.

****

TH-3470_166106104

Chapter 2

Deciphering the architecture of CRISPR-Cas subtype I-C of *Leptospira interrogans* serovar Copenhageni strain Fiocruz L1-130 and comparison with subtype I-B

The PAM prediction portion in this chapter is partially adapted from the published article and reprinted with author's permission. Hussain, M. S., **Anand, V**., & Kumar, M. (2023). Functional PAM sequence for DNA interference by CRISPR-Cas IB system of *Leptospira interrogans* and the role of LinCas11b encoded within lincas8b. International Journal of Biological Macromolecules, 237, 124086.

Abstract

The CRISPR-Cas system has captivated attention for its remarkable ability to identify and eliminate foreign nucleic acids precisely. In line with a previous report and the CRISPR-CasFinder webserver, it has been determined that the genome of *Leptospira interrogans* serovar Copenhageni contains two subtypes of CRISPR-Cas systems (I-B and I-C), accompanied by two CRISPR arrays (CRISPR-2 and -3). Remarkably, subtype I-C is barren of an array element, and concurrently, CRISPR-3 lacks a *cas* operon. This prompts inquiries into the underlying rationale for retaining *cas* genes in the absence of the characteristic array and, conversely, maintaining a CRISPR array without the typical *cas* operon. This study conducted an *in silico* analysis of repeat and spacer sequences, predicting the protospacer adjacent motif using spacers of both CRISPRs. Notably, CRISPR-Cas I-C nucleotide and protein exhibit lower similarity and are phylogenetically distant from I-B. The transcript of *cas* genes in CRISPR-Cas I-C was detected under *in vitro* conditions.

TH-3470_166106104

2.1 Introduction

The CRISPR-Cas system, recognized as a form of bacterial adaptive immune system, has garnered considerable interest for its extraordinary capacity to accurately identify and eradicate foreign nucleic acids (Makarova & Koonin, 2015). There are numerous bacteria and archaea which have been documented to possess more than two CRISPR-Cas systems. For instance, *Pyrococcus furiosus* (type I-A, I-G, and III-B) (Majumdar et al., 2015), *Pseudomonas aeruginosa* (type I-C, I-E, and I-F) (Gagaletsios et al., 2022), *Clostridium species* (type I-B, II-B, II-C, and III-B) (Louwen et al., 2014), and *Bacillus species* (type I-B and I-C) are known to have multiple CRISPR-Cas systems (Makarova et al., 2011b). In the ongoing evolutionary competition with CRISPR-Cas; the phages, and other genetic elements have developed varied strategies to hinder or bypass immunity. Organisms that possess multiple CRISPR-Cas systems use different effector complexes to target and destroy diverse foreign nucleic acids (Smith et al., 2023; Zahedipour et al., 2024).

The genomes of *L. interrogans* and *L. kirschneri* encode CRISPR-Cas subtypes I-B and I-C. In contrast, the genome of *L. noguchii* carries CRISPR-Cas subtypes I-C and I-E. Additionally, the genomes of *L. alstonii*, *L. weilii*, and *L. santarosai* harbor CRISPR-Cas subtypes I-B and I-E. (Makarova et al., 2011b; Xiao et al., 2019). *L. interrogans* serovar (sv.) Copenhageni has CRISPR-Cas subtypes I-B and I-C, along with two CRISPR arrays (CRISPR-2; LIC_Cr^2 and -3; LIC_Cr^3) in its genome (Makarova et al., 2011b). The *cas* genes of CRISPR-Cas I-B are oriented in a single direction, positioned on both sides of the LIC-Cr^2 array within two distinct operons, whereas CRISPR-Cas I-C is devoid of CRISPR array (Dixit et al., 2016; Xiao et al., 2019). LIC_Cr^3, however, does not have a *cas* operon in its vicinity, and it can be considered an orphan CRISPR. CRISPR array contains fragments of invading genetic elements called spacers that recognise the target nucleic acids through base pairing, thereby enabling immunity against foreign nucleic acids. To distinguish between self and non-self DNA, a conserved trinucleotide sequence, known as the protospacer adjacent (PAM) motif, is located adjacent to the target DNA (Gleditzsch et al., 2019). In this study, identification of spacers derived from infiltrating genetic elements (protospacer) in LIC_Cr^2 and LIC_Cr^3 was carried out and using upstream region of protospacer PAM was predicted.

TH-3470_166106104

Although it is known that *cas* genes in CRISPR-Cas I-B of *L. interrogans* sv. Copenhageni and CRISPR-Cas I-C of sv. Linhai is transcriptionally active (Dixit et al., 2016; Xiao et al., 2019); there is no information about the transcription of CRISPR-Cas I-C *cas* genes of sv. Copenhageni. The *cas* genes in CRISPR-Cas I-C of sv. Copenhageni was found to be transcriptionally active. To understand similarities and evolutionary relatedness among *cas* genes and protein sequences of both subtypes (I-B and I-C). This study examines the similarities and evolutionary relationships between *cas* genes and protein sequences of both subtypes (I-B and I-C). The findings indicate that CRISPR-Cas I-C nucleotide and protein are less similar and phylogenetically distantly related to I-B.

2.2 Results

2.2.1 CRISPR-Cas elements of L. interrogans sv. Copenhageni.

A prior study indicates *L. interrogans* sv. Copenhageni carries CRISPR-Cas I-B and I-C. The genome includes two CRISPR arrays, LIC_Cr^2 and LIC_Cr^3 (Makarova et al., 2011b; Prakash & Kumar, 2022). Array LIC_Cr^2 array is positioned between the two *cas* I-B operons (Dixit et al., 2016). Array LIC_Cr^3 is located away from both subtypes and thus can be considered as an orphan CRISPR. To investigate the presence of additional CRISPR-Cas elements in the *L. interrogans* sv. Copenhageni genome, we conducted a genome analysis using the CRISPR-CasFinder webserver. The CRISPR-CasFinder web server predicted a total of 11 CRISPRs, out of which two were confirmed (LIC_Cr^2 and LIC_Cr^3), and nine were putative CRISPRs (CRISPR-1, -4, -5, -6, -7, -8, -9, -10, and -11) based on number of spacer present within the array **(Fig. 2.1A)**. The numbering of the CRISPR array is based on its position within the genome. The CRISPR-CasFinder webserver distinguishes between confirmed and putative CRISPRs based on the number of spacer sequences present. An array with more than one spacer is considered confirmed, while an array with only one spacer is considered putative. LIC_Cr^2 is composed of 3 spacers (34, 39, and 37 bp) and 4 repeats (36 bp) (Dixit et al., 2016). In contrast, the LIC_Cr^3 has 4 spacers (41 bp) and 5 repeats (28 bp) sequences **(Fig. 2.1B)**.

TH-3470_166106104

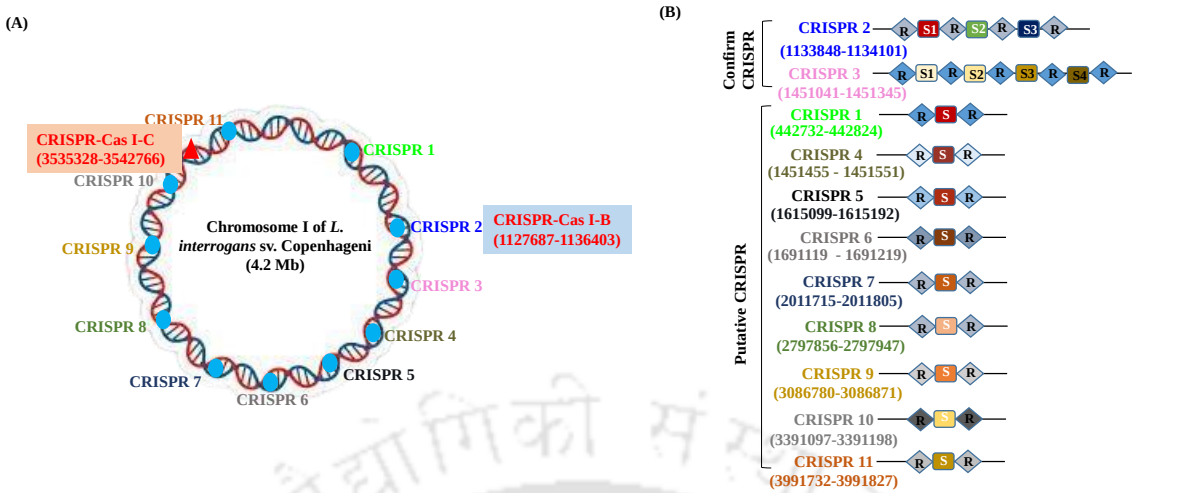


Figure 2.1. Schematic representation of CRISPR-Cas elements within the *L. interrogans* genome. (A) Schematically representation of CRISPR-Cas elements within chromosome I of *L. interrogans* sv. Copenhageni. CRISPR arrays are depicted as sky blue circles, and CRISPR-Cas systems are represented by red triangles. (B) Predicted CRISPR arrays within the genome of *L. interrogans* sv. Copenhageni. CRISPR-2 and -3 are confirmed CRISPR, while others are putative CRISPRs. The repeat (R) is symbolized by a rhombus, and the spacer (S) is represented by a rectangle. Genome coordinates are provided in brackets. CRISPR arrays 1-11 exhibit distinct repeat and spacer sequences and each is uniquely colored for clarity.

The transcription direction of array LIC_Cr^2 was experimentally shown in the direction of the *cas* operon in CRISPR-Cas I-B (Prakash & Kumar, 2021). However, there is no information about the direction of transcription for LIC_Cr^3. As per CRISPR-CasFinder web server, the predicted transcription of array LIC_Cr^3 takes place using the sense strand as a template. In arrays, repeats generally exhibit uniform length and sequence, although minor differences may occur specifically at the end of an array (Lillestøl et al., 2009; F. J. M. Mojica et al., 2005). Variation in terminal repeats has been documented previously in the CRISPR array of *Streptococcus thermophilus* (Horvath et al., 2008) and *E. coli* (Touchon & Rocha, 2010). Likewise, variations were noted among the repeats of array LIC_Cr^2, specifically within its terminal repeat and more prominently toward the 3′ end (Prakash & Kumar, 2021). This instigated to investigate the repeat sequence of LIC_Cr^3. As the transcription direction of LIC_Cr^3 remains experimentally undetermined and has solely been predicted *in silico*, we conducted a thorough examination of both the sense and anti-sense strands of LIC_Cr^3 to explore similarities among the repeat sequences.

The repeats sequence of LIC_Cr^3 from sense strand were aligned and variations among the repeats were seen towards 5′ terminal **(Fig. 2.2A)**. Variations were specifically observed at

2, 5-8, 14, 17, 20, 24, and 25th position in the repeat nucleotide sequence. In contrast, within the repeats sequence of LIC_Cr^3 from anti-sense strand variations was observed towards the 3′ terminal. Differences were precisely observed at the 4, 5, 9, 12, 21-24, and 27th positions in the repeat nucleotide sequence **(Fig. 2.2B)**. Overall, a 65 percent similarity was noted among the repeat sequences of LIC_Cr^3. Consequently, to investigate the cleavage site by the CRISPR-Cas I-C maturation Cas protein within the repeat, a consensus repeat sequence derived from both sense and anti-sense strands of LIC_Cr^3 was selected.

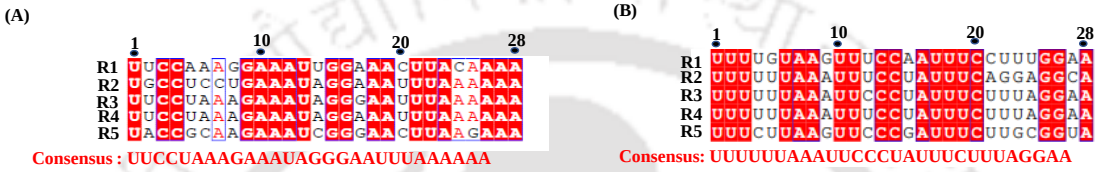


Figure 2.2. Analysis of repeat sequence of orphan LIC_Cr^3. (A) Alignment of LIC_Cr^3 sense strand. Alignment of LIC_Cr^3 anti-sense strand. Sense and antisense strands were predicted using CRISPR-CasFinder webserver. Alignment of repeats was performed using the Clustal Omega webserver using the default parameter. R1, R2, R3, R4, and R5 depict repeat sequences.

Spacer sequences are predominantly distinct within the CRISPR array. Numerous spacers have been identified by aligning with sequences that originate from extrachromosomal sources such as phages, plasmids, and other transferable elements (F. J. M. Mojica et al., 2005; Pourcel et al., 2005). Importantly, these spacers confer sequence-specific immunity against foreign nucleic acids (Barrangou et al., 2007; Deveau et al., 2010). Using the CRISPRTarget web server (Biswas et al., 2013) the origin of spacers within LIC_Cr^3 were identifed (Table 2.1).

Table 2.1. Exotic genetic elements (protospacer) from which spacers of array LIC_Cr^3 were acquired.

CRISPR	SPACER	Protospacer (5'-3')	Invading phages strain (coordinates)
3	1	ATAACCGTCAATTATAATTTCGTTGTAGCTTATAGATAATTCTTTTAAC	IMGVR_UViG_3300010938_000117 (26328-26292)
	2	TTTAGGAATTGTTTCAAGTCTGTTTCCATT	IMGVR_UViG_2541047037_000001 (30045-30006)

TH-3470_166106104

		TAAAGATAACACCTG TAAA	
	3	TTTGGGAAGTGTTTG AAGTTTATTCCATTCT ATCGACAGTTCTTTT AGT	IMGVR_UViG_2541047037_000001 (30114-30074)
	4	TTACTTTAATATTTTA CTTTTATTTTTGATTA AATATTATTCTTTTTA A	IMGVR_UViG_3300009182_000299 (87-124)

2.2.2 PAM prediction using foreign genetic elements invaded in Leptospira.

Cas proteins have a vital role in breaking down foreign DNA into shorter segments, which are then incorporated into the CRISPR array as new spacers (Nuñez et al., 2014a). The Cascade complex uses a two-factor authentication process to distinguish between self and non-self DNA. The first factor involves the mature crRNA, which must be robustly complementary to the target sequence within the invading genome. The second factor is the presence of a PAM located immediately adjacent to the protospacer (Gleditzsch et al., 2019; Leenay & Beisel, 2017; Marraffini & Sontheimer, 2010a). Furthermore, the PAM plays a crucial role in the CRISPR adaptation phase. In *E. coli* subtype I-E, the Cas1-Cas2 complex actively examines foreign nucleic acid to identify the presence of the PAM sequence (Rollie et al., 2015). Similarly, in *Bacillus halodurans*, the Cas1-Cas2-Cas4 complex plays an active role in scrutinizing foreign nucleic acid for the presence of the PAM sequence (Lee et al., 2019; Shiimori et al., 2018). Subsequently, adaptation complex removes the PAM adjacent to the protospacer and incorporates them as a new spacer towards the leader-repeat end of the CRISPR array (Lee et al., 2019; Rollie et al., 2015; Shiimori et al., 2018). In the type I CRISPR-Cas system, the PAM is positioned upstream of the protospacer. The CRISPRTarget webserver (Biswas et al., 2013) aligns spacers with the protospacer and provides the eight-nucleotide sequences upstream and downstream of the flanking region of the protospacer (Table 2.2).

Table 2.2. Origin of spacers within array LIC_Cr[3]. Eight nucleotides upstream of invading genetic elements (protospacers) is depicted using the CRISPRTarget webserver.

CRISPR	SPACER	PAM-Protospacer (5'-3')	Invading phages strain (coordinates)
		TCTCAATAAAGTTTTTCACGGGGTGACGAATGTTT TCCCCCT	Ga0123519_10003459 (22167-22136)
		TTTGATGAAGGGGGAAAACATTCGTCACCCCGTG AAAAACTT	IMGVR_UViG_2531839053_000001 (22136-22167)

TH-3470_166106104

2	1	TTTAATGAAGGGCGAGAACATTCGTCATCCTGTGAAAAACTT	IMGVR_UViG_2541047030_000003 (12676-50382)
		TTTAATGAAGGGCGAGAACATTCGTCATCCTGTGAAAAACTT	IMGVR_UViG_2541047030_000002 (19810-19779)
		TTTAATGAAGGGCGAGAACATTCGTCATCCTGTGAAAAACTT	IMGVR_UViG_2541047029_000002 (19822-19791)
		TTTAATGAAGGGCGAGAACATTCGTCATCCTGTGAAAAACTT	IMGVR_UViG_2534681715_000001 (16951-16920)
	2	No hits	-
	3	TTTAATGAAGGGCGAGAACATTCGTCATCCTGTGAAAAACTT	IMGVR_UViG_3300025691_000035 (12213-12239)
3	1	ATAACCGTCAATTATAATTTCGTTGTAGCTTATAGATAATTCTTTTAAC	IMGVR_UViG_3300010938_000117 (26328-26292)
		CTCCTAATAACCTGTTGTTTCGTATTCGCTTACATATGATTCTTTTAAA	IMGVR_UViG_3300029349_000026 (8814-8778)
		CTTTAAACGAATTATGTCATCTTTTGCGCTTAAAAATAATTCTTTTAAT	IMGVR_UViG_3300010289_000015 (4311-4339)
		GTTAGGAAGAGTTGTAAGTTGGTTAGCATTTAAATACAATTTTTGCAAA	*Leptospira interrogans* serovar Bataviae strain 1548 plasmid p3 (34936-34976)
		ATTACTTAGTGTTGTAATTAAGGTAGCGTTTAAAGATAATTTTTGTAAA	IMGVR_UViG_3300014492_000006 (7620-7658)
			IMGVR_UViG_3300014499_000007 (21043-21081)
			IMGVR_UViG_3300014838_000020 (38924-38962)
			IMGVR_UViG_3300017697_001055 (3283-3246)
			IMGVR_UViG_3300007540_001321 (2795-2832)
		AGGAGGCAGAGTTGTAATTTCGTTCGCGCTTAAATATAATTCTTTTAAA	IMGVR_UViG_2541047037_000001 (29976-29936)
	2	TTTAGGAATTGTTTCAAGTCTGTTTCCATTTAAAGATAACACCTGTAAA	IMGVR_UViG_2541047037_000001 (30045-30006)
	3	TTTGGGAAGTGTTTGAAGTTTATTCCATTCTATCGACAGTTCTTTTAGT	IMGVR_UViG_2541047037_000001 (30114-30074)
		TTTGGGAAGTGTTTGAAGTTTATTCCATTCTATCGACAGTTCTTTTAGT	IMGVR_UViG_2541047037_000001 (30114-30074)
		TTTAAGCTGAGTTTCAAGCTTATTCCTATCTATAGCCAGTTCTTTCTGT	IMGVR_UViG_3300008611_000021 (71156-71122)
			IMGVR_UViG_7000000099_000029 (57721-57687)
			IMGVR_UViG_3300007096_000002 (71449-71415)
			UGV-GENOME-0352847 (55068-55032)
			IMGVR_UViG_7000000422_000021 (25221-25255)
			UGV-GENOME-0352542 (11200-11236)
			DTR_380873 (11314-11348)
			UGV-GENOME-0350548 (23062-23098)
		CAGGCTGATTGATTAAAGTTTATTCGATTCTTTCCAAAGCTCTTTTACA	IMGVR_UViG_3300013307_000016 (135933-135896)
		TTGCGGTAATATTTTGAGTTGATTTCTGCTTAAATATAATTCTTTTAAG	IMGVR_UViG_2541047037_000001 (30183-30143)
			IMGVR_UViG_GVMAG-M-3300025860-25_000001 (27519-27554)
			IMGVR_UViG_3300025138_002215 (7773-7807)
			IMGVR_UViG_3300001567_000012 (3900-3862)
		TTACTTTAATATTTTACTTTTATTTTTGATTAAATATTATTCTTTTTAA	IMGVR_UViG_3300009182_000299 (87-124)
		ACGAGATGATATTCCGAGTTTAATTCCACTTTTAAATAATTCTTTTAAG	IMGVR_UViG_3300006793_002264 (6315-6275)
			IMGVR_UViG_3300025066_000658 (892-932)
			IMGVR_UViG_3300025103_000041 (29383-29343)
			IMGVR_UViG_3300025133_000754 (15438-15478)
			IMGVR_UViG_3300010150_000270 (4913-4873)

	4	TAGCTGCTTCGTTTTCATTTGGCTTATGATTAAATA TAATTCTTTCAAG	IMGVR_UViG_3300009071_000436 (6977-6940)
			IMGVR_UViG_3300025849_000937 (5493-5530)
		AGAGCTTTTTCATTTGATTTTATTTCGGCTTCAATA GAATTCTTTGAAT	IMGVR_UViG_3300027977_000172 (7785-7750)
			IMGVR_UViG_3300028044_000207 (285-250)
			IMGVR_UViG_3300028581_000270 (14443-14478)
			IMGVR_UViG_3300028557_000246 (18127-18162)
			IMGVR_UViG_3300028553_000043 (47156-47191)
			IMGVR_UViG_3300027730_000101 (1170-1135)
			IMGVR_UViG_3300027970_000032 (46865-46900)
		TATGTTCTATATTTTGTGTTGCTTTCCTCATGAATA TAATTGTTTCATA	Norman_145_NODE_2_length_145303_cov_286.136298_1 (2229-2259)
			IMGVR_UViG_3300006802_002588 (210-5246)
			IMGVR_UViG_3300007236_000347 (1979-1943)
			IMGVR_UViG_3300025810_000011 (29440-29476)
			IMGVR_UViG_3300025803_000225 (12099-12135)
			IMGVR_UViG_3300025769_000360 (31675-31711)
			IMGVR_UViG_3300025759_001977 (8078-8114)
		TGTAAAATATATTTTTAGAAGATTTATGTTTATATAA AATTCTTTATAA	IMGVR_UViG_3300006147_000029 (1317-1281)
			IMGVR_UViG_3300006145_000023 (19209-19173)
			IMGVR_UViG_3300006217_000088 (29601-29637)
			IMGVR_UViG_3300006149_000100 (16930-16894)
		TTGCGGTAATATTTTGAGTTGATTTCTGCTTAAATA TAATTCTTTTAAG	IMGVR_UViG_2541047037_000001 (30183-30143)
			IMGVR_UViG_GVMAG-M-3300025860-25_000001 (27519-27554)
			IMGVR_UViG_3300025138_002215 (7773-7807)
			IMGVR_UViG_3300001567_000012 (3900-3862)

Using WebLogo webserver (Crooks et al., 2004) an eight nucleotide sequence upstream of the protospacers were aligned to determine consensus trinucleotides or PAM. “TGA” and “ATA” were predicted PAM employing spacers of LIC_Cr^{2} **(Fig. 2.3A)**. Likewise, “TTA”, “TTT”, “GAA”, “TCT”, and “GCT” were predicted PAM using spacers of LIC_Cr^{3} **(Fig. 2.3B)**.

(A)

TH-3470_166106104

Figure 2.3 PAM prediction using foreign genetic elements invaded in *Leptospira*. Web logo representation of eight nucleotides upstream of exotic nucleic acid sequence from which LIC_Cr^2 and LIC_Cr^3 spacers are acquired. (A) ATA and TGA are the predicted PAM using spacers of LIC_Cr^2. (B) TTA, TTT, GAA, TCT, and GCT are the predicted PAM using spacers of LIC_Cr^3.

It is noteworthy that the prior analysis of *Leptospira's* CRISPR-Cas systems revealed a moderately conserved PAM, specifically "TAC" for subtype I-B and "TTC" for I-E. However, no apparent motifs were detected in the downstream region of protospacers (Xiao et al., 2019).

2.2.3 Comparison of CRISPR-Cas I-C and I-B.

CRISPROne is one of the renowned webservers to predict presence of CRISPR-Cas system within prokaryotic genome (Zhang & Ye, 2017). Based on homology search CRISPROne webserver predicts Cas proteins involved in CRISPR-Cas based immunity. Additionally, CRISPROne also provider genome coordinates of each *cas* genes present in *cas* loci. Using CRISPROne webserver and the data retrieved from the earlier report (Makarova & Koonin, 2015), architecture of CRISPR-Cas I-C in *L. interrogans* sv. Copenhageni was predicted. The CRISPR-Cas I-C of *L. interrogans* sv. Copenhageni predicted to have have eight *cas* genes which are named as follows: *Lincas3C', Lincas3C, Lincas5C, Lincas8C, Lincas7C, Lincas4C, Lincas1C*, and *Lincas2C*. The coordinates of these genes start from *Lincas3C'* (3535328) and end at *Lincas2C* (3542766) **(Fig. 2.4)**. There is a 4 bp overlap between *Lincas5C and Lincas8C*. Additionally, there are 2, 139, 254, and 2 bp intergenic regions between

TH-3470_166106104

Lincas8C-Lincas7C, *Lincas7C-Lincas4C*, *Lincas4C-Lincas1C*, and *Lincas1C-Lincas2C*, respectively **(Fig. 2.4).** In the CRISPR-Cas I-C systems of *Bacillus halodurans* and *Streptococcus pyogenes*, the *cas3C* helicase and nuclease domain is encoded in one open reading frame (ORF). On the other hand, in *L. interrogans*, the predicted *Lincas3C* helicase domain and nuclease domain are encoded by two separate ORFs **(Fig. 2.4)**.

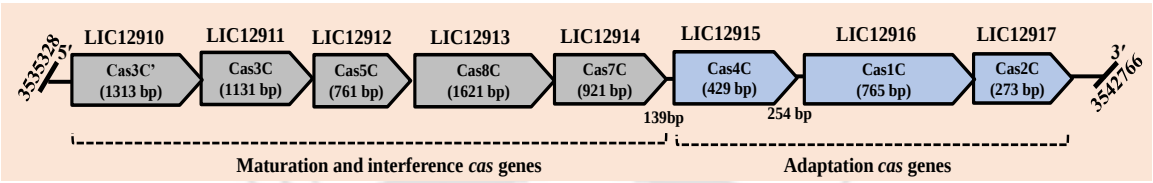


Figure 2.4. The predicted architecture of CRISPR-Cas I-C locus of *L. interrogans* sv. Copenhageni. Architecture of CRISPR-Cas I-C was predicted using CRISPROne webserver. Coordinates of the CRISPR-Cas I-C are mentioned at each end of the *cas* locus. *Lincas1, Lincas2, and Lincas4 are* anticipated adaptation *cas* genes, *Lincas5C* is predicted to play a maturation role, while *Lincas7, Lincas8, and Lincas3C* are forecasted interference *cas genes* within the CRISPR-Cas I-C.

As earlier stated *L. interrogans* sv. Copenhageni also possess CRISPR-Cas I-B (Makarova & Koonin, 2015). Therefore, it was interesting to explore the similarity and evolutionary connections within *cas* genes and protein sequences of both subtypes. Each set of *cas* gene and protein sequences of CRISPR-Cas I-B and I-C were aligned to examine similarity between them using the EMBOSS and BLAST web server enlisted in Table 2.3. CRISPR-Cas I-C and I-B share 30-40% *cas* gene sequence and 2-40% protein sequence similarity.

Table 2.3. Comparing the *cas* genes and protein sequences of CRISPR-Cas I-C with I-B.

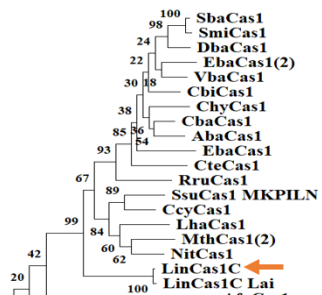
***cas* genes of subtype I-C (Locus tag)**	***cas* genes of subtype I-B (Locus tag)**	**Nucleotide sequence similarity in percent**	**Protein sequence similarity in percent**
Lin*cas*1C (LIC12917)	Lin*cas*1B (LIC10942)	40.3	25.4
Lin*cas*2C (LIC12916)	Lin*cas*2B (LIC10941)	41.4	47.0
Lin*cas*4C (LIC12915)	Lin*cas*4B (LIC10943)	33.6	17.0
Lin*cas*5C (LIC12912)	Lin*cas*5B (LIC10935)	40.7	2.4
Lin*cas*7C	Lin*cas*7B	39.3	18.0

TH-3470_166106104

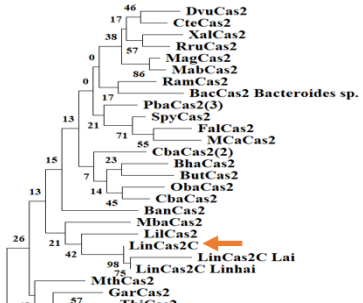
(LIC12914)	(LIC10936)		
Lin*cas*8C (LIC12913)	Lin*cas*8B (LIC10937)	39.5	14.1
Lin*cas*3C (LIC12911)	Lin*cas*3B (LIC10938)	34.9	8.4
Lin*cas*3C′ (LIC12910)		30.3	12.1

Due to the low sequence similarity between *cas* genes and protein sequences of CRISPR-Cas I-C and I-B, an analysis of their evolutionary relationship was undertaken. Cas protein orthologs with more than 20 percent similarity were considered for constructing a phylogenetic tree, regardless of their subtypes. The lineage of predicted adaptation Cas protein in CRISPR-Cas I-C is distantly related to I-B within the same serovar (Copenhageni) as both are placed in different branch in phylogentic tree. However, the lineage of predicted adaptation Cas protein in sv. Copenhageni CRISPR-Cas I-C is closely associated with I-C of sv. Lai as they are clustered in single texa **(Fig. 2.5A, 2.5B, and 2.5C).** The anticipated LinCas5C and LinCas5B are positioned at distinct branch in the phylogenetic tree, indicating a distant relationship between their lineages. Despite this, the predicted LinCas5C and LinCas5C_Lai lineages are closely related, as evidenced by their clustering within a single taxon **(Fig. 2.5 D)**. The predicted lineage of interference Cas proteins in CRISPR-Cas I-C is distantly related to I-B, with both being clustered on different branches. Although, the lineage of predicted interference Cas proteins of CRISPR-Cas I-C in sv. Copenhageni is closely associated with I-C of sv. Lai, as both are clustered within a single taxon **(Fig. 2.5E, 2.5F, and 2.5G).** Overall, it can be concluded that the Cas proteins of CRISPR-Cas I-C in *L. interrogans* sv. Copenhageni is are distantly related to I-B and closely related to I-C of sv. Lai.

(A)



(B)



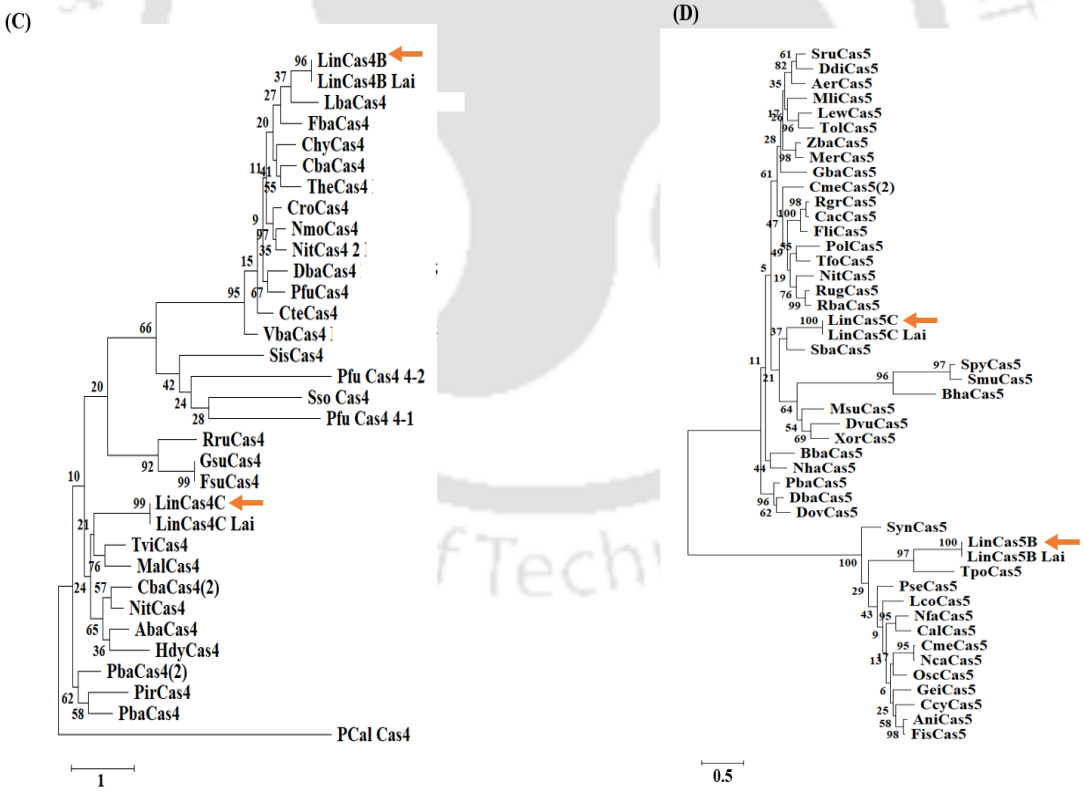

(C)
LinCas4B
LinCas4B Lai
LbaCas4
FbaCas4
ChyCas4
CbaCas4
TheCas4
CroCas4
NmoCas4
NitCas4 2
DbaCas4
PfuCas4
CteCas4
VbaCas4
SisCas4
Pfu Cas4 4-2
Sso Cas4
Pfu Cas4 4-1
RruCas4
GsuCas4
FsuCas4
LinCas4C
LinCas4C Lai
TviCas4
MalCas4
CbaCas4(2)
NitCas4
AbaCas4
HdyCas4
PbaCas4(2)
PirCas4
PbaCas4
PCal Cas4
1
(D)
SruCas5
DdiCas5
AerCas5
MliCas5
LewCas5
TolCas5
ZbaCas5
MerCas5
GbaCas5
CmeCas5(2)
RgrCas5
CacCas5
FliCas5
PolCas5
TfoCas5
NitCas5
RugCas5
RbaCas5
LinCas5C
LinCas5C Lai
SbaCas5
SpyCas5
SmuCas5
BhaCas5
MsuCas5
DvuCas5
XorCas5
BbaCas5
NhaCas5
PbaCas5
DbaCas5
DovCas5
SynCas5
LinCas5B
LinCas5B Lai
TpoCas5
PseCas5
LcoCas5
NfaCas5
CalCas5
CmeCas5
NcaCas5
OscCas5
GeiCas5
CcyCas5
AniCas5
FisCas5
0.5
(E)
MreCas7
ScoCas7
FisCas7
DolCas7
Dvicas7
CbaCas7(2)
LepCas7
PseCas7
CgoCas7
CbaCas7
GbaCas7
NitCas7
SacCas7
AbaCas7
MyxCas
LinCas7B
LinCas7B Lai
TpoCas7
DbaCas7
BhaCas7
(F)
NbaCas7
PbaCas8
CbrCas8
NbaCas8
DbaCas8(2)
DbaCas8
GarCas8
TliCas8
TlaCas8
TniCas8
GbaCas8
NbaCas8(2)
ZbaCas8
MerCas8
(G)
GalCas3
Sbacas3
CniCas3
CniCas3(2)
GbaCas3
TliCas3
McaCas3
BbaCas3
FerCas3
ThiCas3
AsuCas3
HbaCas3
HalCas3
DbaCas3(2)
CcoCas3
XanCas3
NhaCas5
FbaCas3
NbaCas3
SpaCas3
LinCas3C
LinCas3C Lai

Figure 2.5. Evolutionary relationship between Cas proteins of CRISPR-Cas I-C with its orthologs. The phylogenetic tree of predicted Cas proteins in CRISPR-Cas I-C with their orthologs were constructed using Mega11 software taking 1000 bootstraps replicates and Maximum likelihood algorithms. The scale of each tree is mentioned at the bottom. Phylogenetic analysis of LinCas1C (A), LinCas2C (B), LinCas4C (C), LinCas5C (D), LinCas7C (E), LinCas8C (F) and LinCas3C helicase and nuclease (G) with their orthologs. Orange arrow demarcated CRISPR-Cas I-C and I-B Cas protein. Spy (*Streptococcus pyogenes*), Xor (*Xanthomonas oryzae*), Bha (*Bacillus haludurans*), Dvu (*Desulfovibro vulgaris*), Sso (*Sulfolobus solfataricus*), Neu (*Nitrosomonas europaea*), Mth (*Moorella thermoacetica*), Rru (*Rhodospirillum rubrum*), Cte (*Chlorobaculum tepidum*), Syn (*Synechocystis sp*), Fal (*Frankia alni*), Ban (*Bifidobacterium animalis*), Aba (*Armatimonadetes bacterium*), Pba (*Phycisphaerales bacterium*), Mag (*Magnetospirillum*), Gar (*Geobacter argillaceus*), Ram (*Ruficoccus amylovorans*), Gba (*Gammaproteobacteria bacterium*), Mba (*Myxococcales bacterium*), Mab (*Magnetospirillum aberrantis*), Pol (*Polaromonas sp*.), Bac (*Bacteroides sp*.), Aba (*Alphaproteobacteria bacterium*), Thi (*Thiohalocapsa sp*.), Pel (*Pelomonas sp.* P7), But (*Butyrivibrio sp*.), Fal (*Filifactor alocis*), Cba (*Clostridia bacterium*), Oba (*Oscillospiraceae bacterium*), Bba (*Burkholderiales bacterium*), Lmo (*Lamprobacter modestohalophilus*), Cba (*Clostridiales bacterium*), Gba (*Gammaproteobacteria bacterium*), Xba (*Xanthomonadales bacterium*), MCa (*Macrococcus carouselicus*), Chy (*Candidatus hydrogenedentes*), Nau (*Novipirellula aureliae*), Com (*Candidatus omnitrophica*), Nga (*Novipirellula galeiformis*), Dba (*Dehalococcoidia bacterium*), Aco (*Aminobacterium colombiense*), Pba (*Planctomycetes bacterium*), Rub (*Rubinisphaera sp*), Syn' (*Synergistaceae bacterium*), Pla (*Planktothricoides sp.*), Apy (*Aminivibrio pyruvatiphilus*), Zav (*Zavarzinella sp*.), Pba' (*Phycisphaerales bacterium*), Cva (*Chloracidobacterium validum*), Tro (*Thiocapsa roseopersicina*), Cac (*Candidatus accumulibacter*), Fba (*Firmicutes bacterium*), Mth (*Moorella thermoacetica*), Rru (*Rhodospirillum rubrum*), Cte (*Chlorobaculum tepidum*), Gsu (*Geobacter sulfurreducens*), Tth (*Thermus thermophilus*), Mhu (*Methanospirillum hungatei*), Mth (*Methanothermobacter thermautotrophicus*), Fnu (*Fusobacterium nucleatum*), Mxa (*Myxococcus xanthus*), Mru (*Methanobrevibacter ruminantium*), Ssu (*Spirulina subsalsa*), Nit (*Nitrospira sp*.), Eba (*Elusimicrobia bacterium*), Ccy (*Coleofasciculaceae cyanobacterium* SM2_1_6), Cbi (*Candidatus Bipolaricaulis sibiricus*), Sba (*Syntrophaceae bacterium*), Cba

TH-3470_166106104

(*Chloroflexi bacterium*), Aba (*Acidobacteria bacterium*), Eba (*Eubacteriales bacterium*), Vba (*Verrucomicrobia bacterium*), Dba (*Deltaproteobacteria bacterium*), Lha (*Longimonas halophila*), Smi (*Smithella sp*. SCADC), Bbo (*Brevibacillus borstelensis*), Nex (*Nannocystis exedens*), The (*Thermomicrobium sp*.), Apy (*Aminivibrio pyruvatiphilus*), Bbo (*Brevibacillus borstelensis*), Afe (*Acidibacillus ferrooxidans*), Tso (*Tautonia sociabilis*), Nmo (*Nitrospira moscoviensis*), Dmu (*Desulfococcus multivorans*), Dfe (*Desulforamulus ferrireducens*), Smo (*Synechococcus moorigangaii*), Rru (*Rhodospirillum rubrum*), Cba (*Chloroflexi bacterium*), Pir (Pirellula sp.), Tvi (Thiocystis violacea), Pba (Planctomycetaceae bacterium), Mal (*Methylotuvimicrobium alcaliphilum*), Aba (*Armatimonadetes bacterium*), Hdy (*Hydrococcus sp.*), Dba (*Deltaproteobacteria bacterium*), Vba (*Verrucomicrobia bacterium*), Nmo (*Nitrospira moscoviensis*), Cte (*Candidatus tectomicrobia*), Pfu (*Polyangium fumosum*), Cba (*Chloroflexi bacterium*), The (*Thermomicrobium sp*.), Lba (*Leptospiraceae bacterium*), Cma (*Candidatus macondimonas diazotrophica*), Azo (*Azoarcus sp.* Aa7), Rgr (*Rhodocyclus gracilis*), Cac (*Candidatus accumulibacter* sp.), Pro (*Propionivibrio sp*.), Rba (*Rhodocyclaceae bacterium*), Cco (*Candidatus competibacteraceae*), Dag (*Dechloromonas agitate*), Rgr (*Rhodocyclus gracilis*), Cme (*Candidatus methylophosphatis*), Den (*Denitromonas sp*.), Rba (*Rhodocyclaceae bacterium*), ZBa (*Zoogloeaceae bacterium*), Nba (*Nitrospinae bacterium*), Pba (*Proteobacteria bacterium*), Gba (*Gammaproteobacteria bacterium*), Cbr (*Candidatus brocadia*), Lep (*Leptonema illini*), Tli (*Thiothrix litoralis*), Tla (*Thiothrix lacustris*), Nba (*Nitrospirae bacterium*), Zba (*Zetaproteobacteria bacterium*), Mer (*Mariprofundus erugo*), Gar (*Geobacter argillaceus*), Tni (*Thiothrix nivea*), Vba (*Verrucomicrobia bacterium*), Cro (*Candidatus rokubacteria bacterium*), Tpo (*Thalassoglobus polymorphus*), Sco (*Synechococcus*), Scy (*Scytonema sp*.), Xan (*Xanthomonadaceae bacterium*), Mca (*Methylococcus capsulatus*), Cni (*Candidatus nitrotoga fabula*), Gba (*Gammaproteobacteria bacterium*), Nba (*Nitrospirae bacterium*), Nha (*Nitrosococcus halophilus*), Gal (*Gallionella sp*.), Cni (*Candidatus nitrotoga sp*.), Sba (*Sulfuricellaceae bacterium*), Spa (*Salinispira pacifica*), Hba (*Halothiobacillaceae bacterium*), Asu (*Acidithiobacillus sulfuriphilus*), Hal (*Halothiobacillus sp*.), Bba (*Burkholderiales bacterium*), Fer (*Ferrovum sp*.), Cro (*Candidatus rokubacteria bacterium*), Des (*Desulfobacca sp*.), Riv (*Rivularia sp*.), Sym (*Symploca sp*.), Osc (*Oscillatoria sp*.), Lep (*Leptolyngbya sp*.), Ali (*Aphanocapsa lilacina*), Nos (*Nostoc sp*. MS1), Cal (*Calothrix sp*.), Kim (*Kaiparowitsia implicata*) and Tel (*Thermosynechococcus elongatus*).

2.2.4 Transcript analysis of CRISPR-Cas I-C cas genes.

The CRISPROne webserver was employed to predict *cas* genes in CRISPR-Cas I-C of *L. interrogans* sv. Copenhageni. Subsequently, it was interesting to examine the transcription of these *cas* genes in an *in vitro* cultured *L. interrogans* sv. Copenhageni. To investigate the transcript of *cas* genes in CRISPR-Cas I-C **(Fig. 2.6A)**, RT-PCR was performed using cDNA obtained from *in vitro* cultivated sv. Copenhageni as a template. The analysis revealed that the adaptation *cas* genes of the CRISPR-Cas I-C (*Lincas1C, Lincas2C,* and *Lincas4C)* undergo transcription in an *in vitro* condition **(Fig. 2.6B).** Nevertheless, transcription of maturation and

TH-3470_166106104

interference full-length *cas* genes were not detectable **(Fig. 2.6C)**. Occasionally, the intricate folding of RNA molecules may hinder the successful synthesis of full-length cDNA. This situation can result in the creation of truncated cDNA (DAS et al., 2001). Therefore, partial gene amplifying primers were designed to amplify 300 bp towards the 3′ end of the maturation and interference *cas* genes. Using partial gene amplifying primers, transcripts of maturation and interference *cas* genes were detectable **(Fig. 2.6D)**.

Previous studies have shown that the adaptation *cas* genes of CRISPR-Cas I-B in sv. Copenhageni is transcribed from a single promoter (Dixit et al., 2016). To confirm the polycistronic transcription of adaptation *cas* genes in CRISPR-Cas I-C, various primer combinations were tested **(Fig. 2.6A)**. Transcript was detected exclusively when the primer set of *Lincas1C:*forward-*Lincas2C*:reverse was utilized, while no transcript was detected with the use of others primer set. Thus, evaluation of the transcriptional organization of CRISPR-Cas I-C adaptation genes revealed that *Lincas1C* and *Lincas2C* are transcribed as a polycistronic unit controlled by a single promoter **(Fig. 2.6E).**

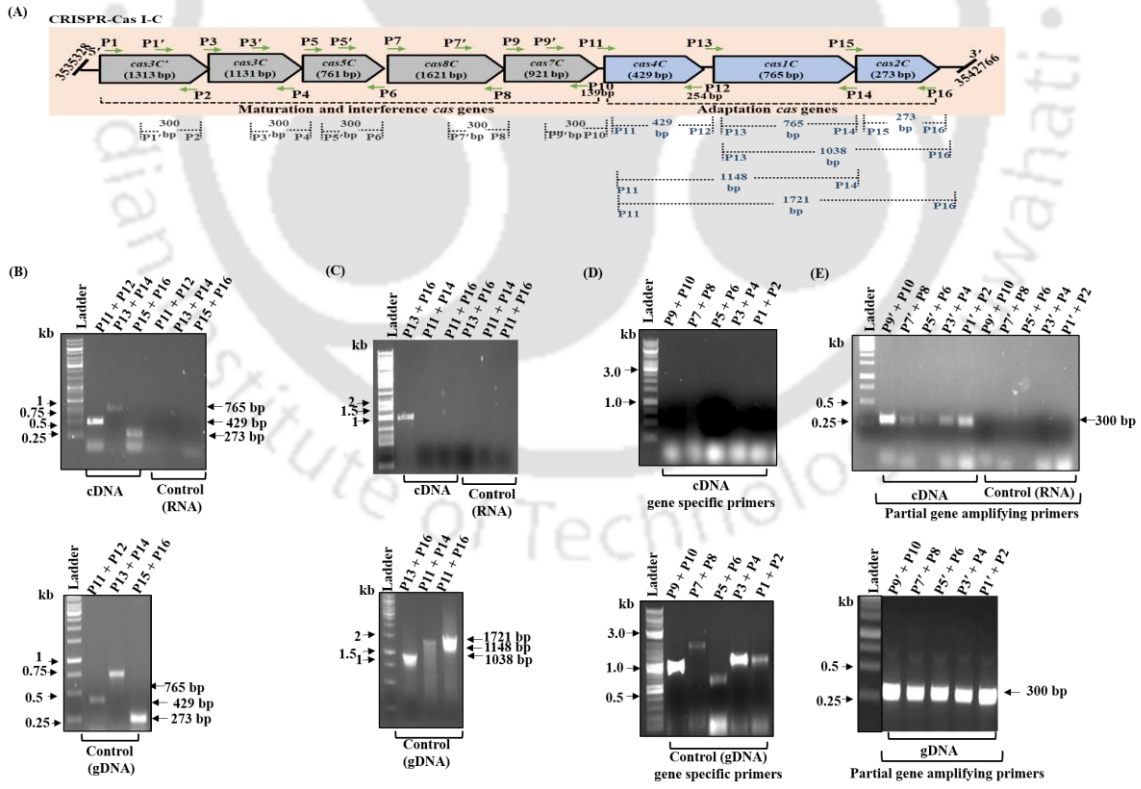


Figure 2.6. Transcript analysis of *cas* genes in CRISPR-Cas I-C of *L. interrogans* sv. Copenhageni. (A) Schematic depiction of various primer sets used for transcript analysis. P1: *Lincas3C'* forward, P2: *Lincas3C'* reverse, P3: *Lincas3C* forward, P4:

TH-3470_166106104

Lincas3C reverse, P5: *Lincas5C* forward, P6: *Lincas5C* reverse, P7: *Lincas8C* forward. P8: *Lincas8C* reverse, P9: *Lincas7C* forward, P10: *Lincas7C* reverse, P11: *Lincas4C* forward, P12: *Lincas4C* reverse, P13: *Lincas1C* forward, P14: *Lincas1C* reverse, P15: *Lincas2C* forward and P16: *Lincas2C* reverse. (') primer represents partial gene amplifying primers. (B) Transcript analysis of adaptation *cas* genes (*Lincas1C, Lincas2C*, and *Lincas4C*), (C) maturation (*Lincas5*) and interference *cas* genes (*Lincas7C, Lincas8C, Lincas3C*, and *Lincas3C'*) using sv. Copenhageni cDNA as a template and gene specific primers. (D) Detection of short transcripts of maturation and interference *cas* genes using partial gene amplifying primers. (E) Polycistronic transcription of adaptation cas genes using various primer combinations as depicted in Fig. A. The PCR with sv. Copenhageni total RNA (cDNA synthesized without reverse transcriptase) serves as RNA control in the experiment. gDNA: genomic DNA.

2.3 Discussion

Prokaryotes host a variety of evolved parasites, including plasmids, viruses, and transposons (van Beljouw et al., 2023). The sole identified form of adaptive immunity in bacteria and archaea is the CRISPR-Cas system (Makarova et al., 2011b). The CRISPR-Cas system acknowledged as a type of bacterial adaptive immune system, has attracted significant attention due to its remarkable ability to precisely detect and eliminate foreign nucleic acids (Koonin et al., 2017). In the continuous evolutionary struggle against CRISPR-Cas, phages, and other genetic elements have evolved diverse strategies to impede or circumvent the host's immune response (Zahedipour et al., 2024). Several bacteria and archaea carry multiple CRISPR-Cas systems, enabling the formation of diverse effector complexes. The presence of diverse effector complexes facilitates the efficient destruction of numerous foreign genetic elements (Smith et al., 2023; Zahedipour et al., 2024). It's important to note that *P. furiosus* has three effector complexes - I-A, I-G, and III-B - that share a single adaptation module associated with type I-G. There is a functional overlap observed among the three CRISPR effector complexes, and crRNAs originating from any of the seven CRISPR arrays can associate and operate efficiently with any of the three effector complexes (Garrett et al., 2020; Majumdar et al., 2015). It is possible that *L. interrogans* sv. Copenhageni's CRISPR-Cas I-C may share LIC_Cr^2, which is located between CRISPR-Cas I-B *cas* operons, or use an orphan LIC-Cr^3 to facilitate CRISPR-Cas based immunity. This can help in eliminating foreign nucleic acids by using different effector complexes.

CRISPR-Cas-mediated immunity relies on the CRISPR array element to recognize and dismantle foreign nucleic acids. The mature crRNA, in association with various Cas protein

TH-3470_166106104

families, forms the Cascade (CRISPR-associated complex for antiviral defence), which recognises and disrupts foreign nucleic acids. Cascade employs diverse mechanisms to distinguish between self and non-self DNA (Gleditzsch et al., 2019; Hochstrasser et al., 2014). CRISPR interference is hindered when, in addition to base pairing with the spacer sequence, the 5′ handle also base pairs with the protospacer-flanking sequence of the target DNA. Self-targeting of the CRISPR locus is naturally avoided, as it would result in complete base pairing of the 5′ handle of the crRNA with the CRISPR repeat sequence from which it is transcribed (Marraffini & Sontheimer, 2010b). The presence of a PAM sequence also serves as an important criterion to discriminate between self and non-self DNA (Marraffini & Sontheimer, 2010b). Cascade complexes can effectively identify PAM sequences in exotic nucleic acids by examining extended DNA sequences (Gleditzsch et al., 2019; Leenay & Beisel, 2017; Marraffini & Sontheimer, 2010a). Likewise, the predicted PAM of *L. interrogans* sv. Copenhageni might help Cascade complex in distinguishing between self and non-self nucleic acid.

In general, CRISPR-Cas subtype I-C includes adaptation (*cas1C*, *cas2C*, and *cas4C*), maturation (*cas5C*), and interference (*cas3C*, *cas7C*, and *cas8C*) *cas* genes (Makarova & Koonin, 2015). CRISPR-Cas I-C Cascade is considered a minimal Cascade as it contains only three *cas* genes in its operon: *cas5C*, *cas7C*, and *cas8C* (Makarova & Koonin, 2015). The subtype I-C Cascade uses Cas5C to process the crRNA instead of Cas6 (Makarova, Wolf, Iranzo, et al., 2020; Makarova et al., 2013). Similarly, predicted CRISPR-Cas I-C architecture of *L. interrogans* sv. Copenhageni includes adaptation (*Lincas1C*, *Lincas2C*, and *Lincas4C*), maturation (*Lincas5C*), and interference (*Lincas3C*, *Lincas7C*, and *Lincas8C*) *cas* genes. CRISPR-Cas I-C of *L. interrogans* sv. Copenhageni also encodes minimal Cascade as that of its orthologs. In contrast to the typical CRISPR-Cas I-C system, the *L. interrogans* sv. Copenhageni CRISPR-Cas I-C, *Lincas3C* helicase, and nuclease domains are encoded in two separate ORF, resembling the organization *cas3* in CRISPR-Cas I-A and I-D (Makarova & Koonin, 2015).

The prevailing belief is that *cas* genes are typically in close proximity to CRISPR loci (Daniel H Haft et al., 2005). However, the CRISPR-Cas I-C system in *L. interrogans* sv. Copenhageni deviates from this expectation as it lacks a CRISPR array element. CRISPRs that are in isolation without nearby *cas* genes have been identified in several species, including

TH-3470_166106104

Listeria monocytogenes (Mandin et al., 2007), *Aggregatibacter actinomycetemcomitans* (Jorth & Whiteley, 2012), and *Enterococcus faecalis* (Hullahalli et al., 2015). In *L. monocytogenes*, isolated CRISPRs were noted to be expressed but not processed into mature crRNAs (Makarova & Koonin, 2015; Zhang & Ye, 2017). In *Yersinia pestis* and *Streptococcus thermophilus*, isolated arrays have been shown to capture new spacers (Barrangou et al., 2007; Pourcel et al., 2005). In *S. islandicus* type III CRISPR-Cas systems, which do not possess CRISPR arrays, have been demonstrated to employ trans crRNA located next to type I-A loci (Deng et al., 2013; Garrett et al., 2011; Shah & Garrett, 2011). *Haloferax volcanii*, harboring three CRISPR loci featuring nearly identical repeat sequences. The expression of all three CRISPR loci occurred, resulting in the generation of mature crRNA. Nonetheless, it was observed that not all crRNAs possessed the capability to effectively initiate interference (Maier et al., 2013). Cas loci found in isolation serve either as vestiges of CRISPR-Cas systems lacking the immunity function or operate in conjunction with distant CRISPR(s) within the same genome (Zhang & Ye, 2017). Overall, it is still unclear which of the isolated arrays are functionally active and, specifically, whether they can complement sets of *cas* genes that are devoid of arrays (Shmakov et al., 2020).

The transcription of *cas* genes is commonly described as polycistronic. Usually, *cas1* and *cas2* are encoded in same operon (Amitai & Sorek, 2016; Makarova et al., 2013). In *L. interrogans* sv. Copenhageni CRISPR-Cas I-B, two clusters of genes adaptation (*Lincas4B, Lincas2B, Lincas1B)* and maturation, interference (*Lincas6B, Lincas3B, Lincas8B, Lincas7B, Lincas5B*) are transcribed under different operons (Dixit et al., 2016). However, in the CRISPR-Cas I-C, *Lincas1C* and *Lincas2C* share a common promoter, while the transcription of *Lincas4C* is not regulated by this promoter. Full-length transcripts were not discernible for the maturation and interference *cas* genes within the CRISPR-Cas I-C; however, short transcripts were detected. In contrast, a full-length transcript for the maturation and interference module was detected in the CRISPR-Cas I-C of sv. Linhai. (Xiao et al., 2019).

2.4 Material and Methods.

2.4.1 Bacterial strain and culture media.

TH-3470_166106104

L. interrogans sv. Copenhageni strain Fiocruz L1-130, was sourced from the Regional Medical Research Centre in Port Blair, Andaman and Nicobar Islands, India. *L. interrogans* sv. Copenhageni was cultivated in EMJH (Ellinghausen-McCullough-Johnson-Harris) medium at a temperature of 29°C. To isolate genomic DNA and RNA of *L. interrogans* sv. Copenhageni, 10 ml of a 7-day-old culture was used.

2.4.2 CRISPR-Cas I-B and I-C sequence analysis.

The CRISPR-Cas system and CRISPR array in *L. interrogans* sv. Copenhageni genome was determined using the CRISPR-CasFinder webserver (Grissa et al., 2007; Pourcel et al., 2005). The genome sequence of *L. interrogans* sv. Copenhageni, which comprises two chromosomes, is publicly accessible on the NCBI database (accession no. ASM207349v2). The genome sequence was loaded onto the CRISPR-CasFinder webserver and the default parameters were used. The set of *cas* genes and its encoded protein sequences in the I-B and I-C locus of *Leptospira* were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/). The genetic architecture of CRISPR-Cas I-C was created based on the documented *cas* gene coordinates (Makarova et al., 2011b) and using the CRISPROne webserver (Zhang & Ye, 2017). A comparison of the set of *cas* genes and its encoded protein sequences in the I-B and I-C locus of *Leptospira* was conducted using BLAST (Korf et al., 2003) and the EMBOSS webserver (Rice et al., 2000).

2.4.3 Phylogenetic analysis.

Cas endonuclease amino acid sequences used in phylogenetic tree construction were retrieved from the UniProt database (https://www.uniprot.org/). The phylogenetic tree was constructed using Mega11 software via the Maximum Likelihood algorithm with 1000 bootstrap replicates (Tamura et al., 2021).

2.4.4 Complimentary DNA synthesis and Reverse transcriptase PCR (RT-PCR).

The total RNA from *L. interrogans* sv. Copenhageni was isolated using a trizol reagent (Thermo Fisher, cat. No. 15596026) following the manufacturer's protocol. The total RNA (2 µg) was then treated with DNase I (NEB, cat. No. M0303S) to exclude the possibility of genomic DNA contamination. 1 µg of total RNA (DNase treated) was used to synthesize cDNA using random hexamer provided in Verso cDNA synthesis kit (Thermo Fisher, cat. No. AB1453A) and as per manufacturer’s protocol. Additionally, cDNA was synthesized using 1

TH-3470_166106104

µg of total RNA (DNase treated) without adding the reverse transcriptase which serve as an RNA control. For RT-PCR, CRISPR-Cas I-C adaptation *cas* gene-specific primers were used with cDNA as a template. In contrast, partial gene amplifying primers were employed to amplify maturation and interference *cas* genes.

2.4.5 PAM prediction.

The CRISPRTarget webserver (Biswas et al., 2013) was used to predict PAM. The LIC_Cr^2 and LIC_Cr^3 spacer sequences were fed separately into the webserver to detect the protospacer and the eight nucleotides upstream flanking region of the protospacer. The eight-nucleotide flank sequence of each protospacer was aligned using the WebLogo webserver (Crooks et al., 2004) to identify the consensus trinucleotide or PAM.

Table 2.4. Oligos used in Chapter 2.

DNA oligos	Sequence (5'-3')
Lincas1C: F	CTAGCTAGCTTGATCAGAAAAGCTCAAT
Lincas1C: R	CCGCTCGAGTTATTTCCATATAAAAGGAA

TH-3470_166106104

Lincas2C: F	CGCGGATCCCATGTTTATCATTGTATGTTACGACGT
Lincas2C: R	GCGTCGACTTAAAAATCAAGAATGTTAGAAACTCC
Lincas4C: F	CTAGCTAGCATGTTTAGTAAATCAATAAAACTGT
Lincas4C: R	CCGCTCGAGTTAATATCCGCCAGAGTCTT
Partial gene amplifying primers	
Lincas5C: F	GACAGTATTACAAAATTTGAAGAG
Lincas5C: R	TCATAGTCTCGCTCCATTGGT
Lincas7C: F	TTTTCCGAAGAAGATTTAGAATTATTTTG
Lincas7C: R	TTAGGAAACTTTTTTGATTAATTCCAC
Lincas8C: F	GAAAGATTACAGGAAAGAGCAAATC
Lincas8C: R	TTATTCTCCTTGTGATAGTTCTTC
Lincas3C: F	AAAAATTTCAAACTAATCGACAAAGAC
Lincas3C: R	TTAAATGATTAAATCTTGATCATTCGG
Lincas3C': F	GTATTACAAACAAAATGGATCGAGTG
Lincas3C': R	CTAAGCTTCATCCAATATGATTACG

****

Chapter 3

Biochemical Characterization of Cas Proteins Involved in Adaptation Phase in CRISPR-Cas Subtype I-C of *L. interrogans* sv. Copenhageni Strain Fiocruz L1-130

This is chapter is partially adapted from the published article and reprinted with author's permission. **Anand, V**., Prabhakaran, H. S., Gogoi, P., Kanaujia, S. P., & Kumar, M. (2022). Structural and functional characterization of Cas2 of CRISPR-Cas subtype IC lacking the CRISPR component. Frontiers in Molecular Biosciences, 9, 988569. (PDB ID: 7F84).

Abstract

The pathogenic *Leptospira interrogans* serovars Copenhageni and Lai encodes CRISPR-Cas subtypes I-B and I-C. However, CRISPR-Cas subtype I-C lacks a CRISPR array. In the CRISPR-Cas I-C, the gene size of *Lincas1C* reported in the NCBI database is 765 bp, which is inconsistent with the documented protein size of 336 amino acids. However, a reverse transcription-polymerase chain reaction has revealed the existence of a *Lincas1C* transcript of 765 bp, suggesting that *Lincas1C* is 765 bp (254 amino acids). Multiple sequence alignment and modeled structure of LinCas1C displayed a shorter N-terminal region than its orthologs; however, it is comparable with LinCas1C_Lai. In this study, the amino acids E78, H146, and E162 of LinCas1C are predicted to constitute the core nucleolytic site. In addition, the adaptation Cas proteins in CRISPR-Cas I-C of *Leptospira* were cloned, overexpressed, and purified. We observed that due to a natural frameshift mutation in the rLinCas2C_Lai, it expresses truncated protein but retains its nuclease activity. The adaptation Cas protein in CRISPR-Cas I-C of sv. Copenhageni and rLinCas2C_Lai of sv. Lai exhibited metal-dependent DNase and metal-independent RNase activities. Moreover, rLinCas1C and rLinCas2C exhibited inertness towards single-stranded oligos, while rLinCas4C displayed sequence-specific cleavage activity on single-stranded oligos. The crystal structure of rLinCas2C obtained at the resolution of 2.60 Å revealed that the protein is in an apostate conformation and contains N- (1-71 amino acids) and C-terminal (72-90 amino acids) regions, with the former possessing a ferredoxin fold. Substitution of the conserved residues (Tyr7, Asp8, Arg33, and Phe39) with alanine and deletion of Loop L2 resulted in compromised DNase activity. On the other hand, only selective rLinCas2C mutants showed a moderate reduction in RNase activity.

TH-3470_166106104

3.1 Introduction

Cas1 and Cas2 are universal Cas proteins found in all CRISPR-Cas systems and are necessary for spacer integration (Koonin et al., 2017). The Cas1-Cas2 protein complex helps insert new spacers in the microbial genome between the leader sequence and the first repeat of the CRISPR array using a two-step transesterification reaction (Nunez et al., 2015; Rollie et al., 2015). Cas1 functions as a nuclease and engages metal ions to cleave double stranded DNA, single stranded DNA, and branched DNA substrates (Babu et al., 2011; He et al., 2018; Wiedenheft et al., 2009). *L. interrogans* Cas1B (LinCas1B), *Archaeoglobus fulgidus* Cas1 (AfuCas1), *Pseudomonas aeruginosa* (PaeCas1), *Riemerella anatipestifer* (RanCas1), and *E. coli* Cas1 (EcoCas1) exhibit metal-ion dependent DNase activity on a wide variety of DNA substrates (Babu et al., 2011; Dixit, Prakash, et al., 2021; He et al., 2018; Wiedenheft et al., 2009). Furthermore, it has been found that both AfuCas1 and EcoCas1 demonstrate RNase activity that is dependent on metal ions (Babu et al., 2011; Kim et al., 2013).

Cas2 of *E. coli* (EcoCas2) aids the non-catalytic acquisition of exotic nucleic acids (protospacers) into the CRISPR array. The mutation of Cas2's active site does not prevent the acquisition of spacers by the heterohexameric complex of Cas1-Cas2 in *E. coli* (Lee et al., 2019; Nuñez et al., 2014b; Rollie et al., 2015). EcoCas2 acts non-catalytically as a yardstick to measure the length of the protospacer (Nunez et al., 2015). Cas2 proteins function as core metallonucleases and have been linked to the virulence process in *Legionella pneumophil* (Gunderson et al., 2015). However, the catalytic role of Cas2 in CRISPR biology is not yet fully understood, despite the structural and functional characterization of several Cas2 orthologs (Beloglazova et al., 2008; Jung et al., 2016; Ka et al., 2014; Kwon et al., 2012; Nam, Ding, et al., 2012; Samai et al., 2010). It has also been suggested that Cas2 may play a role in inducing morphological changes in *E. coli*, indicating that its catalytic activity could potentially be harnessed for other biological processes beyond the scope of CRISPR-Cas (Wang et al., 2019).

The tertiary structure of Cas2 from various organisms, including *Sulfolobus solfataricus* Cas2 (SsoCas2), *Bacillus halodurans* Cas2 (BhaCas2), *Streptococcus pyogenes* Cas2 (SpyCas2), *Desulfovibrio vulgaris* Cas2 (DvuCas2), and *Thermococcus onnurineus* Cas2 (TonCas2), consists of N- and C-terminal regions (Beloglazova et al., 2008; Jung et al., 2016; Ka et al., 2014; Nam, Ding, et al., 2012; Samai et al., 2010). The former region has a ferredoxin

TH-3470_166106104

(βαββαβ) fold, while Cas2 proteins form dimers through the interaction of their C-terminal β5 strands (Beloglazova et al., 2008; Jung et al., 2016; Ka et al., 2014; Nam, Ding, et al., 2012; Samai et al., 2010). In the SsoCas2 dimer, a pair of conserved aspartate residues (Asp10) are involved in catalytic activity (Beloglazova et al., 2008).

Type-specific ancillary *cas* genes have been integrated into the adaptation module in CRISPR-Cas systems. Cas4 is a widely distributed ancillary protein in type I, II, and V CRISPR-Cas systems (Hudaiberdiev et al., 2017). Cas4 usually exists as an independent protein, but certain systems, such as type I-G and V-B, encode Cas4/1 fusion protein (Li et al., 2014). An *in vivo* study on *Haloarcula hispanica* type I-B CRISPR system suggests that deleting the *cas4* gene prevents the uptake of new spacers (Li et al., 2014). In addition, Cas4 has been found to exhibit metal-ion dependent DNase activity, as observed in *L. interrogans* sv. Copenhageni (LinCas4B) (Dixit, Anand, et al., 2021b). Moreover, in *Sulfolobus solfataricus*, Cas4 proteins, such as Sso0001Cas4 and Sso1391Cas4, exhibit various DNA processing abilities such as unwinding, exonuclease, and endonuclease activities (Lemak et al., 2013; Zhang et al., 2012). In *Bacillus halodurans* type I-C, BhaCas4 forms a higher-order complex with BhaCas1-BhaCas2 in the presence of a double-stranded DNA (dsDNA) substrate with $BhaCas4_2$-$BhaCas1_4$-$BhaCas2_2$ stoichiometry (Lee et al., 2019). The BhaCas4-BhaCas1-BhaCas2 complex is highly specific for PAM sequences and cleaves precisely upstream of the PAM in single-stranded DNA (Lee et al., 2019). *Bacillus halodurans* Cas4 (BhaCas4) recognizes the PAM sequence with high specificity and sequesters the overhang away from the active site of Cas1 (Dhingra et al., 2022). In *Synechocystis*, Cas4 collaborates with Cas1 to process the prespacer in a PAM-dependent manner. Following prespacer processing, the Cas1-prespacer complex exhibits an increased affinity for Cas2. Subsequently, Cas1 dissociates from Cas4 and forms a Cas1-Cas2-prespacer ternary complex for prespacer integration (Wu et al., 2021).

In the *Leptospira* genome (accession no. ASM207349v2), an inconsistency was found between the reported nucleotide size (765 bp) and protein sequence size (336 amino acids) of LinCas1C in the NCBI database. The RT-PCR analysis confirmed that LinCas1C has a precise size of 765 bp and corresponds to 255 amino acids. Therefore, for subsequent study, LinCas1C of 765 bp was considered. The *cas2C* gene of another pathogenic serovar Lai (*Lincas2C_Lai*; LA0683), encodes only the first 58 amino acids due to a natural frameshift mutation. Therefore, it was fascinating to explore whether LinCas2C_Lai could demonstrate nuclease activity

TH-3470_166106104

despite expressing truncated proteins. Consequently, we further scrutinized the biochemical properties of rLinCas2C_Lai and juxtaposed its activity with that of rLinCas2C.

L. interrogans adaptation Cas proteins, possess metal-ion dependent DNase activity on various DNA substrates and metal-ion independent RNase activity on *luciferase* mRNA substrate. Additionally, rLinCas1C and rLinCas2C showed no activity towards ss-oligos. However, rLinCas4C demonstrated cleavage activity on the ss-oligo sequence specifically. The crystal structure of rLinCas2C was determined and found to exist in the dimeric form with the characteristic N-terminal ferredoxin fold (βαββαβ). The structure was further compared with its homologs. This is the first report on the crystal structure and function of CRISPR-Cas lacking array elements from spirochetes.

3.2 Results

3.2.1 Transcript analysis of Lincas1C.

Unexpectedly, a notable discrepancy in the NCBI database regarding the gene size of *Lincas1C* was discovered. The database states that the *LinCas1C* gene sequence is 765 bp long, while the corresponding protein sequence is said to consist of 336 amino acids. This inconsistency raises concerns about the accuracy of the data. Subsequently, the information about the LinCas1C locus in the Uniprot database was examined, which is known for its meticulously curated sequences, but unfortunately, we could not find any information (LIC12916). To determine the precise size of the LinCas1C transcript, primers that could amplify sequences of different lengths were designed (765, 785, 805, and 1002 bp) **(Fig. 3.1A)**. Interestingly, when employing sv. Copenhageni cDNA as the template in RT-PCR, only the 765 bp amplicon was witnessed **(Fig. 3.1B)**. Previous studies have also reported the Cas1C transcript of sv. Linhai (*Lincas1C_Linhai*) to be 765 bp long using RT-PCR (Xiao et al., 2019). Additionally, the Uniprot database lists the lengths of other Cas1 proteins of *L. interrogans*, such as LinCas1B sv. Copenhageni (LIC10942) and Cas1C of sv. Lai (LinCas1C_Lai; LA0684), as 284 and 255 amino acids, respectively. The length of LinCas1B and LinCas1C_Lai is similar to the length of the LinCas1C nucleotide sequence documented in the NCBI database. Based on our findings and the previous report, we have selected a 765 bp sequence, which corresponds to a protein comprising 265 amino acids, for this study.

TH-3470_166106104

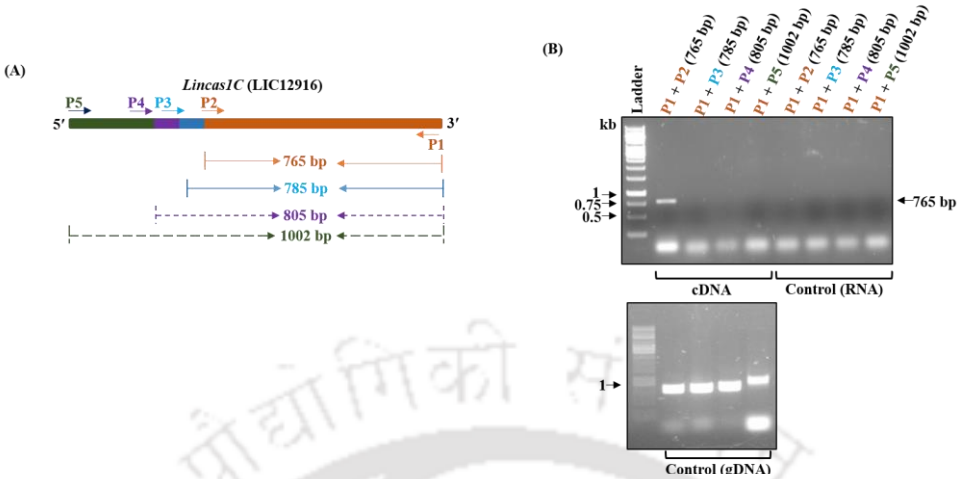


Figure 3.1. Transcript analysis of LinCas1C by RT-PCR. Determination of *Lincas1C* transcript size using sv. Copenhageni complementary DNA (cDNA) as a template. (A) Schematic representation of primer sets constructed to amplify transcripts (765, 785, 805, and 1002 bp) of the *Lincas1C*. (B) A transcript of 765 bp was detected for *Lincas1C* using sv. Copenhageni cDNA as a template in RT-PCR. Complementary DNA template synthesized from total RNA of sv. Copenhageni using random hexamers.

3.2.2 In silico characterization of LinCas1C.

A Multiple Sequence Alignment (MSA) was conducted to determine the similarity between LinCas1C and its orthologs. The MSA analysis unveiled that LinCas1C is 32 residues shorter at the N-terminal compared to LinCas1B (CRISPR-Cas I-B of sv. Copenhageni; LIC10942). LinCas1C protein sequence length is comparable to LinCas1C (LA0684) of sv. Lai CRISPR-Cas IC (LinCas2C_Lai). Compared to other orthologs, LinCas1C is approximately 80-90 amino acids shorter toward the N-terminal. The N-terminal of *Leptospira* LinCas1 (LinCas1C, LinCas1C_Lai, and LinCas1B) is generally shorter than their orthologs **(Fig. 3.2A)**.

The disparity in size between LinCas1C and its orthologs prompted to investigate its structural conservation. The tertiary structure of LinCas1C was modeled using AfuCas1 as a template **(Fig. 3.2B)** and compared to its closest orthologs. The LinCas1C lacks two beta-strands and a helix at its N-terminal, in contrast to the LinCas1B modeled structure **(Fig. 3.2C)**. When comparing the modeled structure of LinCas1C to that of LinCas1C_Lai, it was found that both have a similar overall tertiary structure **(Fig. 3.2D)**. However, LinCas1C is missing

TH-3470_166106104

eight beta-strands and a helix at its N-terminal, compared to BhaCas1 and AfuCas1 **(Fig. 3.2E)**. In AfuCas1 and *Pseudomonas aeruginosa* Cas1 (PaeCas1), the N-terminal domain has a significant contribution in the formation of homodimer (Kim et al., 2013; Wiedenheft et al., 2009). Despite having a shorter N-terminal domain, LinCas1B reportedly forms a homodimer through *in silico* analysis (Dixit, Prakash, et al., 2021). Similarly, the homodimer of LinCas1C was modeled to identify the specific residues responsible for its dimer formation **(Fig. 3.2F)**. The LinCas1C homodimer demonstrates interactions between Lys4 and Glu13 of one protomer with Lys8, and Ser10 of another protomer, at a distance of ≥3.5 Å **(Fig. 3.2F)**. These residues exhibit conservation within LinCas1C and LinCas1C_Lai and are hypothesized to contribute in the homodimer formation.

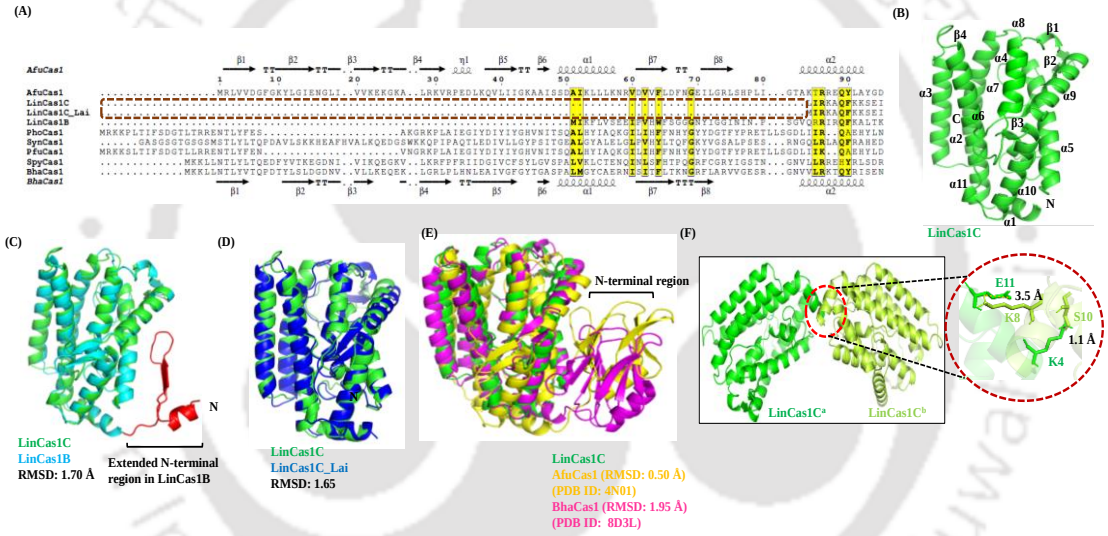


Figure 3.2. Multiple sequence alignment, tertiary structure prediction of LinCas1C, and comparison with its orthologs. (A) Multiple sequence alignment (MSA) of LinCas1C with its orthologs. MSA was performed using Clustal Omega and decorated using an EsPript 3 webserver. AfuCas1: *Archaeoglobus fulgidus* Cas1 (Uniprot ID: O28401), PhoCas1: *Pyrococcus horikoshii* Cas1 (Uniprot ID: O58938), SynCas1: *Synechocystis* Cas1 (Uniprot ID: A0A068N1P8), PfuCas1: *Pyrococcus furiosus* Cas1 (Uniprot ID: Q8U1T7), SpyCas1: *Streptococcus pyogenes* Cas1 (Uniprot ID: A0A660A7M1) and BhaCas1: *Bacillus halodurans* Cas1 (Uniprot ID: Q9KFX9). (B) LinCas1C modeled tertiary structure (255 amino acids), generated by AlfaFold using *Archaeoglobus fulgidus* Cas1 as a template (PDB ID: 4N06). (C) Structural comparison of LinCas1C with LinCas1B (RMSD: 1.70). (D) LinCas1C_Lai (RMSD: 1.65). (E) BhaCas1 (RMSD: 1.95), AfuCas1 (RMSD: 0.50). (F) Modeled homodimer of LinCas1C using AfuCas1 homodimer (PDB ID: 4N06) as a template. LinCas1C^a and LinCas1C^b are the two protomers of LinCas1C. Red circle represents residues involved in homodimer formation.

TH-3470_166106104

The α5-α7 region in AfuCas1 contains potential nucleolytic residues (E168, H239, and E254) (Kim et al., 2013). It was observed that LinCas1C (E78, H142, and E162) and LinCas1B (E108, H176, and E191) have similar residues mapping in their modeled structures, which is consistent with the putative nucleolytic residues of AfuCas1 **(Fig. 3.3A).** It is worth noting that the nucleolytic sites of Cas1 variants (LinCas1C, LinCas1B, and AfuCas1) mostly cluster with positively charged residues **(Fig. 3.3B, 3.3C, and 3.3D).** When the nucleolytic core residues in AfuCas1 (E168, H239, and E254) and LinCas1B (E108, H176, and E191) were substituted with Ala, DNase activity was abolished (Dixit, Prakash, et al., 2021; Kim et al., 2013). The study on AfuCas1 (Kim et al., 2013) and LinCas1B (Dixit, Prakash, et al., 2021) has revealed that their nucleolytic sites are involved in metal-dependent endonuclease activity. Similarly, LinCas1C also contains a nucleolytic site in its tertiary structure, which prompted us to investigate its nuclease activity on different nucleic acid substrates.

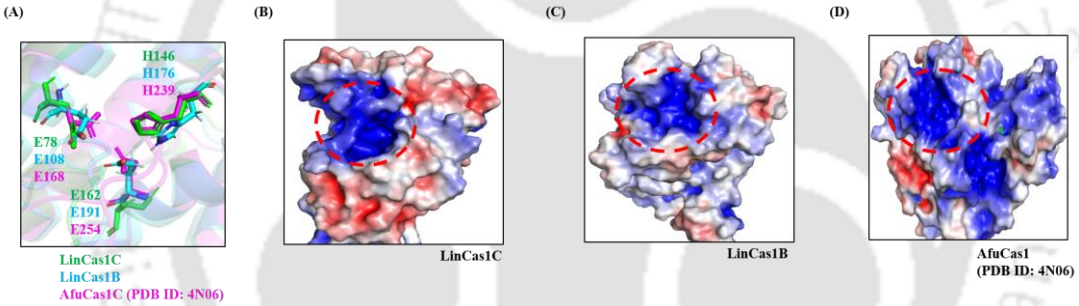


Figure 3.3. Illustration of the predicted nucleolytic core site of LinCas1C. (A) Tertiary structure alignment of modeled LinCas1C and LinCas1B with the available structure of AfuCas1 (PDB ID: 4N06) to predict nucleolytic core site. Residues E108, H176, and E191 are predicted to form a nucleolytic core site of LinCas1C. (B, C & D) Surface electrostatic potential map of LinCas1C to map positive charged amino acids at nucleolytic core site. The positive and negative charges are blue and red, respectively (scale 5 kT/e to +5 kT/e for red and blue, respectively).

3.2.3 Oligomeric state of rLinCas1C, rLinCas2C and rLinCas2C_Lai in solution.

To study the adaptation Cas proteins in CRISPR-Cas I-C of *L. interrogans* sv. Copenhageni, *Lincas1C*, *Lincas2C*, and *Lincas4C* genes were cloned. In addition, *LinCas2C_Lai* of sv. Lai was cloned. Thereafter, these proteins were overexpressed and the corresponding recombinant proteins were purified using Ni-NTA affinity chromatography **(Fig. 3.4)**.

TH-3470_166106104

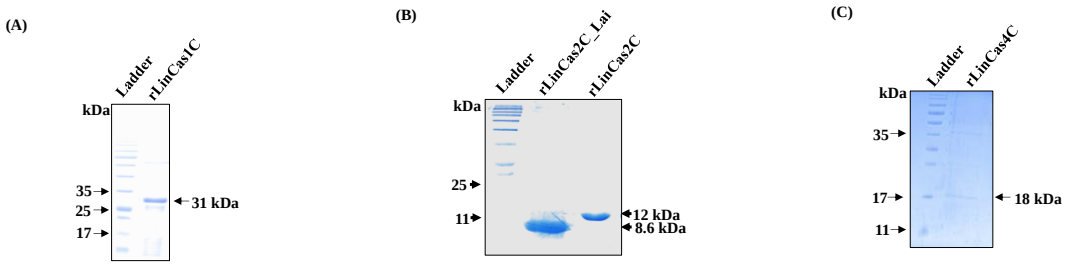


Figure 3.4. Cloning, overexpression, and purification of adaptation Cas proteins of *L. interrogans* CRISPR-Cas I-C. 15 % SDS-PAGE analysis of purified rLinCas1C (yield: 2 mg/L) (A), rLinCas2C (yield: 5 mg/L), rLinCas2C_Lai (yield: 1 mg/L) (B) and rLinCas4C (yield: 0.1 mg/L) (C).

To analyze the oligomeric state of purified rLinCas1C and rLinCas2C proteins in solution size-exclusion chromatography (SEC) was performed. The rLinCas1C was found to exist in a monomeric state **(Fig. 3.5A)**. LinCas1C has a shorter N-terminal region in comparison to its homologous, and the modeled homodimer indicates that only two residues are involved in the homodimer formation, it suggests that there may be a weak interaction between the two protomers. Therefore, LinCas1C eluted in a monomeric state. The rLinCas2C was found to exist in both dimeric (28 kDa) and monomeric (15 kDa) states in solution **(Fig. 3.5B)**. However, due to a natural frameshift mutation, rLinCas2C_Lai was purified in a truncated form (8.6 kDa) that lacked the C-terminal region essential for dimer formation. Despite this, rLinCas2C_Lai could self-assemble to a trimeric (34 kDa) form, in addition to the expected monomeric state (12 kDa), instead of forming a dimer during SEC **(Fig. 3.5B)**. Unfortunately, the low yield of protein prevented the SEC of rLinCas4C despite multiple attempts.

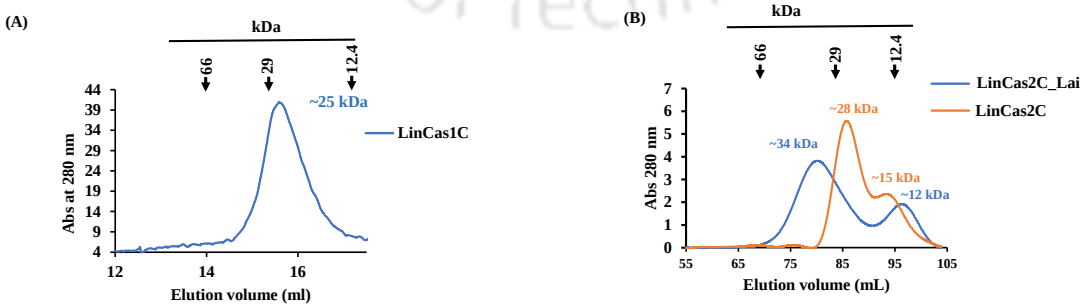


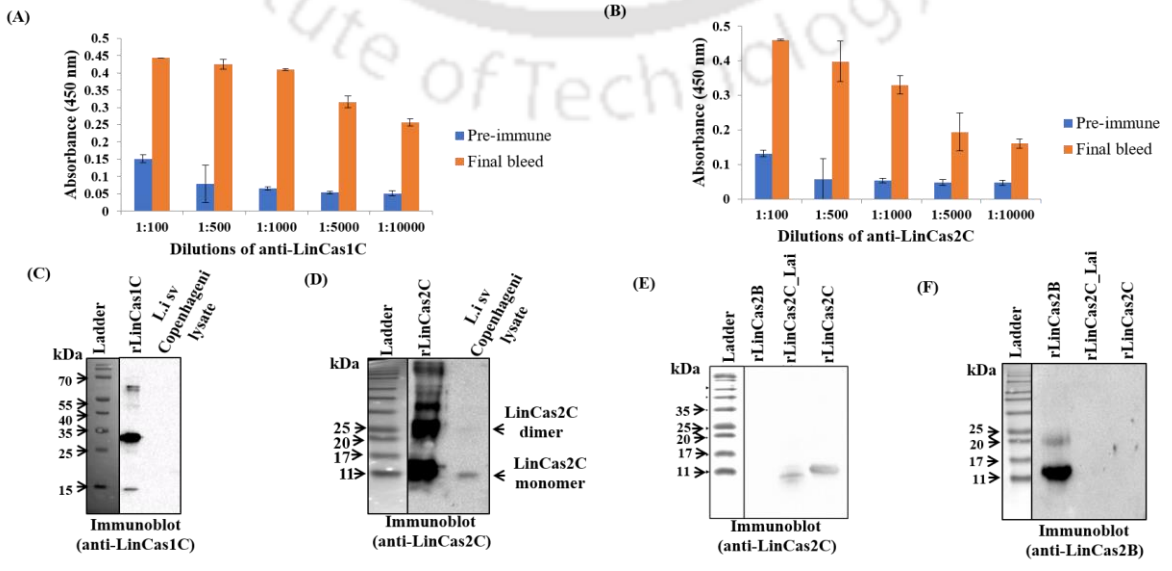
Figure 3.5. Oligomeric state of rLinCas1C, rLinCas2C and rLinCas2C_Lai in solution. (A) Size-exclusion chromatography of rLinCas1C. The chromatogram

TH-3470_166106104

shows that the rLinCas1C eluted at the monomeric size (approximately 25 kDa). SEC650 high-resolution column (10 mm×300 mm, catalog no. 7801650) was used for conducting SEC using an NGC chromatography system (BioRad). (B) Chromatogram shows the rLinCas2C eluted at the dimeric (approx. 28 kDa) and monomeric size (approx.15 kDa), whereas rLinCas2C_Lai eluted at higher oligomeric size (approx. 34 kDa) along with monomeric size (approx. 12 kDa). Superdex 200 increase column (GE Healthcare, catalog. no.28-9909-44) on AKTA prime plus (GE Healthcare).

3.2.4 Native detection of LinCas1C and LinCas2C and cross-antigenic verification.

The detection of the transcript of adaptation *cas* genes in the CRISPR-Cas I-C has led to an investigation of the expression of these adaptation Cas proteins in an *in vitro* grown sv. Copenhageni. To study the expression of native LinCas1C and LinCas2C in an *in vitro* grown sv. Copenhageni, polyclonal antibodies against rLinCas1C and rLinCas2C were raised in mice. Anti-rLinCas1C and rLinCas2C shows optimal titer at 1:1000 dilution **(Fig. 3.6A and 3.6B)**. The expression of native LinCas1C was not detectable in the *L. interrogans* sv. Copenhageni cell lysate (~5 × 10^9 cells) possibly due to its lower expression level **(Fig. 3.6C).** However, monomeric forms of native LinCas2C were identified through immunoblot in the lysate of *L. interrogans* sv. Copenhageni (~5 × 10^9 cells), including a faintly observed dimeric form of native LinCas2C **(Fig. 3.6B)**. A phylogenetic analysis of LinCas2C and its orthologs revealed that LinCas2C is distantly related to LinCas2B but closely related to LinCas2C_Lai (discussed in Chapter 2). This raises the question whether anti-LinCas2C antibodies could detect LinCas2B or LinCas2C_Lai or vice versa. Recombinant LinCas2C and rLinCas2B were found not to cross-react with each other **(Fig. 3.6E).** Nevertheless, anti-LinCas2C could cross-react with rLinCas2C_Lai **(Fig. 3.6F)**, which supports the phylogenetic study.



TH-3470_166106104

Figure 3.6. Native detection of LinCas1C and LinCas2C and cross-antigenic verification. Titer determination of polyclonal antibody generation against rLinCas1C (A), and rLinCas2C (B). Detection of Native expression of LinCas1C (C) and LinCas2C (D) in *L. interrogans* sv. Copenhageni (L.i sv) by immunoblot. *Leptospira* lysates were probed with anti-rLinCas1C or anti-rLinCas2C (1:1000 dilution). (E) Immunoblot to detect rLinCas2B and rLinCas2C_Lai with anti-LinCas2C. (F) Immunoblot to detect rLinCas2C and rLinCas2C_Lai with anti-LinCas2B (1:1000 dilution). Secondary antibody HRP-conjugated anti-mouse secondary antibodies at 1:5000.

3.2.5 Nuclease activity of rLinCas1C on double-stranded DNA substrate.

The conserved nucleolytic core site in LinCas1C as compared to its orthologs driven to explore its nuclease activity on various DNA substrates. To determine the most effective concentration of rLinCas1C for complete degradation of dsDNA substrate, we tested its nuclease activity at various protein concentrations (ranging from 1-15 µM) and found that optimal cleavage occurred at a concentration of 15 µM of rLinCas1C in the presence of Mg^{2+} ions after an hour of incubation **(Fig. 3.6A)**. This cleavage was observed to be more effective with Mg^{2+} ions, compared to Mn^{2+} and Ni^{2+} ions. However, no nuclease activity was detected in the presence of Ca^{2+} and Fe^{2+} ions. This suggests that rLinCas1C requires a divalent Mg^{2+} ion as a cofactor for optimal DNase activity and possesses endo-deoxyribonuclease activity. We also found that adding a metal-ion chelating agent (EDTA) completely abolished the rLinCas1C nuclease activity, emphasizing the importance of specific divalent ions for DNase activity **(Fig. 3.7B)**.

The effect of pH on the DNase activity of rLinCas1C was determined within a pH range of 3.0 to 11.0. It was observed that the most significant nuclease activity occurred at pH 8.0, while no activity was seen at pH 9.0 to 11.0. At lower (3.0 - 6.0) and higher (9.0 -10.0) pH values, strong nucleic acid binding activity was observed, but it does not have any physiological relevance **(Fig. 3.7C)**.

The DNase activity of rLinCas1C was tested on linear dsDNA substrate. In the absence of Mg^{2+} ions, rLinCas1C showed DNA binding activity, but when Mg^{2+} ions were present, its DNase activity was found to be deficient **(Fig. 3.7D)**. This is consistent with AfuCa1, which was reported to be inactive towards linear double-stranded DNA at a concentration of 0.1 µM, but exhibited DNA binding activity at higher concentrations (Kim et al., 2013). Notably, the DNase activity of rLinCas1C on circular dsDNA was similar to that of LinCas1B (Dixit, Prakash, et al., 2021).

TH-3470_166106104

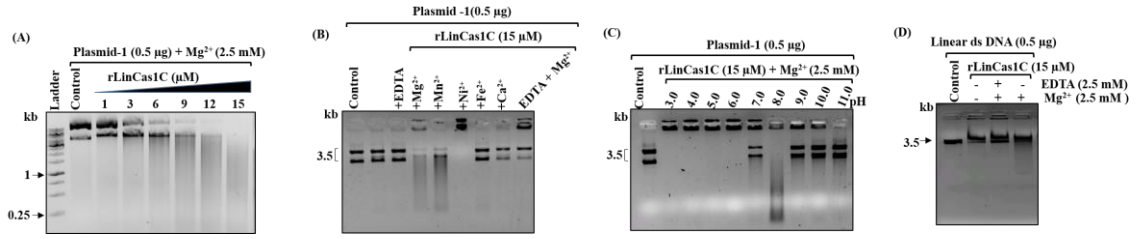


Figure 3.7. DNase activity of rLinCas1C on double-stranded DNA substrates. Nuclease activity was accomplished for an hour at 37°C. (A) Concentration-dependent DNase activity of rLinCas1C on plasmid DNA substrate. (B) Metal-ion dependent activity of rLinCas1C in the presence of different divalent metal-ions (2.5 mM $MgCl_2$, $MnSO_4$, $NiSO_4$, $FeCl_2$, and $CaCl_2$). (C) DNase activity of rLinCas1C evaluated at different pH (3-11). (D) DNase activity of rLinCas1C on linear dsDNA in the presence or absence of meat-ion. Plasmid-1 (Circular dsDNA substrate): pTZ57R/T plasmid-1;500 ng, linear dsDNA: linearized pTZ57R/T plasmid;500 ng, divalent metal-ion: 2.5 mM, EDTA: 2.5 mM, Protein ladder: pre-stained protein ladder (Himedia, cat. No. MBT092) and DNA Ladder: GeneRuler 1 kb DNA Ladder (Thermo scientific, cat. No. SM0311). Nuclease reaction products were analysed on ethidium bromide-stained 1% agarose gel.

3.2.6 Nuclease activity of rLinCas1C on single-stranded DNA and RNA substrate.

The purpose of the study was to investigate the nuclease activity of rLinCas1C on different types of DNA and RNA substrates. The nuclease activity of the rLinCas1C on the dsDNA substrate instigated to explore its activity on ssDNA and RNA substrates to comprehend its variability towards different nucleic acid substrates. The results showed that rLinCas1C exhibited DNase activity on linear ss-DNA (M13mp18) and circular ss-DNA (ΦX174) in the presence of Mg^{2+} ions (**Fig. 3.8A and 3.8B)**, but did not show any activity towards small 50- and 23-mer oligos **(Fig. 3.8C)**. These findings are consistent with LinCas1B, which degrades viral ss-DNA and is inert towards small oligos (Dixit, Prakash, et al., 2021). Furthermore, rLinCas1C showed ribonuclease activity on the *luciferase* mRNA substrate, independent of Mg^{2+} **(Fig. 3.8D)**.

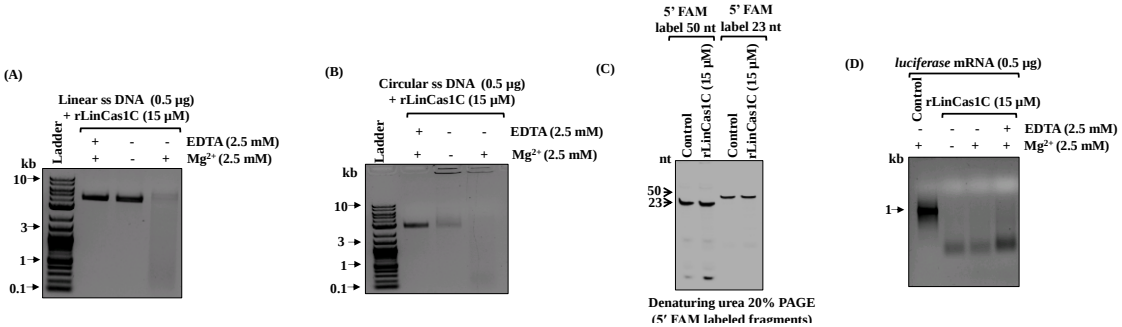


Figure 3.8. Nuclease activity of rLinCas1C on single-stranded DNA and RNA substrates. Nuclease activity was accomplished for an hour at 37°C. (A) DNase activity of rLinCas1C on linear ssDNA substrate (M13mp18; 6.4kb, 500 ng) and (B) circular ss DNA substrate (ΦX174; 5 kb, 500 ng) in the presence or absence of metal-ion. (C) The DNase activity of rLinCas1C on 5' FAM labeled ss oligos (50 and 23-mer). (D) RNase activity of rLinCas1C on *luciferase* mRNA substrate (500 ng). rLinCas1C: 15µM, EDTA: 2.5 mM, $MgCl_2$: 2.5 mM and Ladder: GeneRuler 1 kb DNA Ladder (Thermo scientific, cat. No. SM0311). Nuclease reaction products (A), (B), and (D) were analyzed on ethidium bromide stained 1% agarose gel. Reaction products of (C) were analyzed on 8 M 20% denaturation PAGE, and the FAM-labeled product was visualized directly.

3.2.7 Cas2C of the CRISPR-Cas I-C locus is present in the genomes of L. interrogans sv. Copenhageni and sv. Lai.

The CRISPR-Cas I-C locus of sv. Copenhageni and Lai contains the *Lincas2C* gene (ORF id: LIC12917) which is 273 base pairs long and is located between the nucleotide coordinate 3535328 - 3542766 **(Fig. 3.9A)**. In a previous study, researchers analyzed the CRISPR-Cas I-C locus of the pathogenic spirochete *L. interrogans* sv. Lai using an *in silico* approach and reported that the *Lincas2C_Lai* gene (ORF id: LA0683) was missing at nucleotides coordinate 686432- 693873 (Xiao et al., 2019). However, the NCBI genome database of *Leptospira* predicts that the *Lincas2C_Lai* gene (LA0683) is 272 bp in size, with a natural deletion of one nucleotide (adenine108th), which may result in partial ORF translation. Thus, Xiao and co-workers (Xiao et al., 2019) reported that in *L. interrogans* sv. Lai genome, *Lincas2C_Lai* (LA0683) might encode for truncated (58 amino acids) and inactive LinCas2C_Lai (Xiao et al., 2019). We have described the partial reading frame (177 of 272 bp) of the CRISPR I-C locus of *L. interrogans* sv. Lai **(Fig. 3.9B)**.

The CRISPR-Cas I-C locus in *Leptospira* lacks the CRISPR array, which is necessary for RNA-mediated interference of foreign nucleic acids and makes the role of Cas2 in CRISPR

biology uncertain in the absence of an array. In phylogenetic analysis, it was revealed that among Cas2 of *Leptospira*, the lineage of LinCas2C is closely related to LinCas2C_Linhai and LinCas2C_Lai. LinCas2B is a well-studied Cas protein of *Leptospira* from our research group (Dixit et al., 2016). However, LinCas2B and LinCas2C proteins grouped separately in the phylogenetic tree analysis (discussed in Chapter 2). This led us to investigate whether the LinCas2C nuclease property differs from that of LinCas2B. Both LinCas2B and LinCas2C_Linhai have been characterized to have metal-dependent DNase activity and are transcriptionally active under in vitro growth conditions (Dixit et al., 2016; Xiao et al., 2019). However, the role of LinCas2C is not yet fully understood.

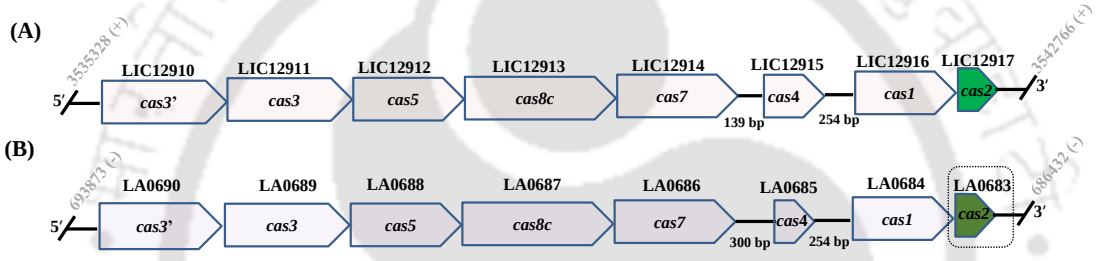


Figure 3.9. Schematic representation of CRISPR-Cas I-C locus of *L. interrogans*. (A and B) Schematic representation of the architecture of CRISPR-Cas I-C of sv. Copenhageni and Lai, respectively, highlight *Lincas2C* ORF (green). Genomic coordinates are mentioned at the ends. + and – represent sense and anti-sense strands, respectively.

3.2.8 Recombinant LinCas2C and LinCas2C_Lai nuclease activity on double-stranded DNA.

The Cas2C enzyme was purified and used to study its nuclease activity on various DNA substrates. The LinCas2B_Lai was not included in the analysis as it had a similar sequence to the well-studied LinCas2B. To optimize the cleavage of dsDNA (plasmid DNA; 0.5 µg), we conducted a nuclease assay with increasing concentrations (5-25 µM) of each rLinCas2C and rLinCas2C_Lai. We found that around 25 µM of each enzyme could completely cleave the DNA within an hour at 37°C **(Fig. 3.10A and 3.10B)**. Moreover, the study showed that both Cas2C enzymes are endodeoxyribonucleases.

To better understand their role in nuclease activity, we substituted metal ions in the nuclease reaction. The results showed that both nucleases had a higher affinity for Mg^{2+} as a cofactor for their DNase activity. Mn^{2+} and Fe^{2+} followed next in terms of preference. Both nucleases showed reduced or no DNA cleavage activity in the presence of Ca^{2+}, Cu^{2+}, and Zn^{2+}

TH-3470_166106104

cofactors **(Fig. 3.10C and 3.10D)**. The agarose gel electrophoresis of the reaction products further revealed a shift in DNA mobility, possibly due to the preserved DNA binding property of LinCas2C.

In this study, we also investigated whether the nucleotide sequence of the substrate influences the activity of the rLinCas2C nuclease. The DNase activity of rLinCas2C was tested on two circular dsDNA plasmids and showed non-specific DNase activity that was similar to that of LinCas2B and LinCas2C_Linhai **(Fig. 3.10E)** (Dixit et al., 2016; Xiao et al., 2019). The DNase activity of rLinCas2C required divalent metal ions, and the optimum pH range between 7.0 to 9.0. The nuclease activity decreased at pH 10.0 to 11.0, and the enzyme showed a moderate binding affinity for DNA. At lower pH (3.0 to 6.0) enzyme showed a high binding affinity for DNA **(Fig. 3.10F)**. The pH-dependent DNase activity of rLinCas2C was similar to that of LinCas2B (Dixit et al., 2016). Furthermore, it is suggested that BhaCas2 assumes a metal-bound catalytically active conformation at the optimum pH (Nam, Ding, et al., 2012).

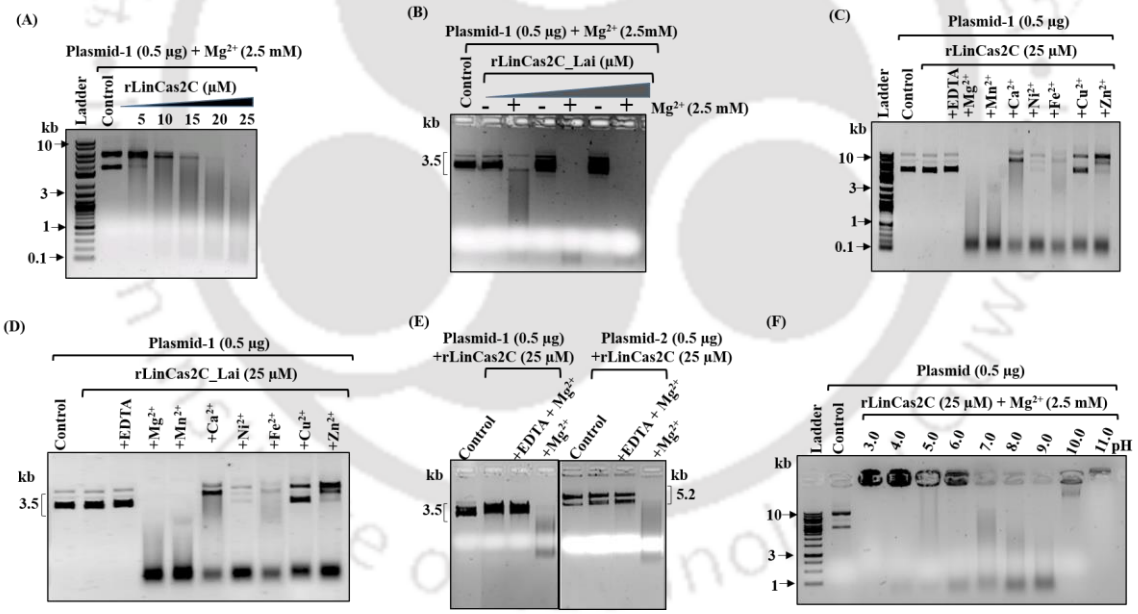


Figure 3.10. DNase activity of rLinCas2C and LinCas2C_Lai on plasmid DNA. DNase activity reaction was carried out at 37ºC for an hour. Concentration-dependent DNase activity of rLinCas2C (A) and (B) rLinCas2C_Lai on plasmid-1 substrate (5.3 kb pET28a vector, 0.5 µg) in the presence of Mg^{2+} ion. DNase activity of rLinCas2C (C) and LinCas2C_Lai (D) in the presence of different divalent metal ions on plasmid exemplifies its optimum activity in Mg^{2+} and Mn^{2+} ions. (E) The substrate specificity of rLinCas2C on two different plasmid substrates. Substrate plasmid-1 (3.5 kb pTZ57R/T vector, 0.5 µg) and plasmid-2 (5.2 kb pET28a vector, 0.5 µg) were employed for DNase activity. (F) DNase activity of rLinCas2C at different pH. The optimum activity was observed at pH 8.0 and 9.0. DNA ladder: 2 log DNA ladder (NEB). rLinCas2C: 25

TH-3470_166106104

µM, EDTA: 2.5 mM, Mg^{2+} and others divalent metal: 2.5 mM. Reaction products were analyzed on 2% agarose gel.

3.2.9 Recombinant LinCas2C nuclease activity on single-stranded DNA and RNA.

It was interesting to investigate whether rLinCas2C can degrade ssDNA and RNA. Previous studies have shown that LinCas2B and BhaCas2 were inactive against short DNA oligos (28-32-mer) (Dixit et al., 2016; Nam, Ding, et al., 2012). The results of nuclease activity suggested that rLinCas2C was incapable of cleaving short DNA oligos (23- and 50-mer), even in the presence of a cofactor **(Fig. 3.11A).** However, both rLinCas2C and rLinCas2C_Lai demonstrated DNase activity on viral ssDNA (linear M13mp18 and ΦX174) in the presence of divalent metal ion **(Fig. 3.11B, 3.11C, 3.11D and 3.11E).** The activity of rLinCas2C on viral ssDNA was similar to the activity reported for LinCas2B (Dixit et al., 2016). Moreover, both rLinCas2C and rLinCas2C_Lai can cleave mRNA transcripts of the *luciferase* gene without requiring divalent metal ions **(Fig. 3.11F and 3.11G)**.

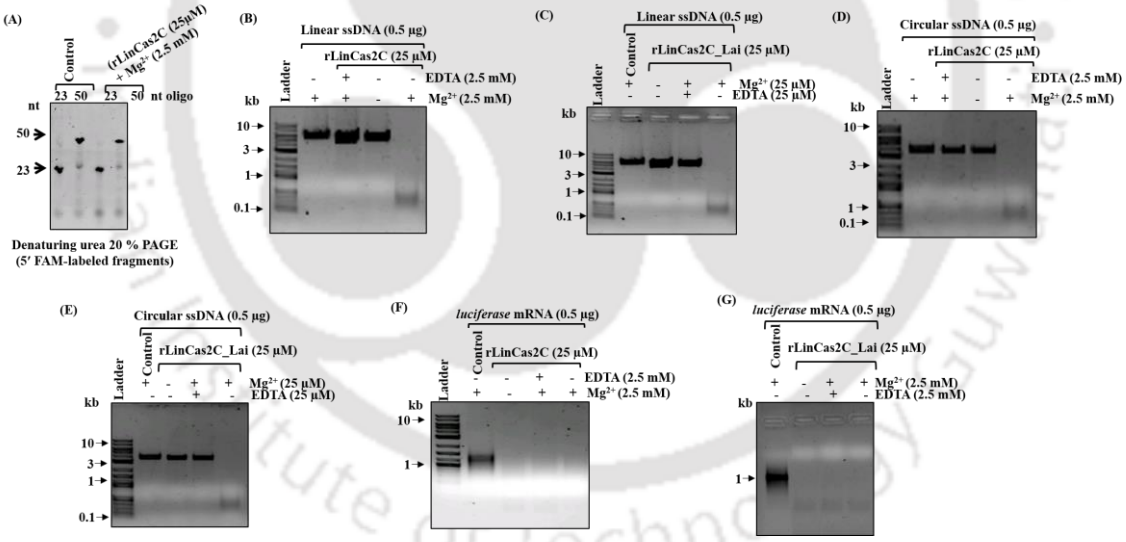


Figure 3.11. Nuclease activity of rLinCas2C on single-stranded DNA and RNA. The nuclease activity reaction was carried out at 37ºC for an hour. (A) DNase activity of rLinCas2C on synthesized single-stranded linear DNA (oligo-1: 23-mer 0.4 µM, oligo-2: 50-mer 0.4 µM). The nuclease reaction product was analysed on 8 M 12% urea-PAGE. DNase activity of rLinCas2C (B) and LinCas2C_Lai (C) on linear single-stranded DNA (0.5 µg of 6.4 kb M13mp18). DNase activity of rLinCas2C (D) and LinCas2C_Lai (E) on circular single-stranded DNA (3.6 kb ΦX174, 0.5 µg). RNase activity of rLinCas2C (F) and LinCas2C_Lai (G) on *luciferase* mRNA (0.5 µg). Complete degradation of RNA was observed even in the absence of Mg^{2+} ions. (DNA ladder: 2 log DNA ladder (NEB). rLinCas2C: 25 µM and Mg^{2+}: 2.5 mM. The nuclease

TH-3470_166106104

reaction products of Fig. (B), (C), (D), (F) and (G) were analysed on EtBr stained 2% agarose gel.

3.2.10 Overall structure of rLinCas2C.

The crystal structure of rLinCas2C was shown to contain the N-terminal ferredoxin domain (βαββαβ), a characteristic of nuclease protein. Overall, the LinCas2C is made up of three α-helices (α1-α3) and five anti-parallel β-strands (β1-β5) **(Fig.3.12A)**. Data collection and refinement statistics of rLinCas2C is enlisted in Table 3.1. This is consistent with the structure of other Cas2 orthologs listed in Table 3.2. Cas2 orthologs have two loops named L1 and L2 that connect β1-α1 and α2-β4, respectively. It is speculated that these loops recognize DNA and RNA substrates. Structure superimposition of rLinCas2C over modeled rLinCas2B shows that the DNA binding loop L1 is shorter, while the RNA binding loop L2 is comparable in size **(Fig. 3.12B)**. The modeled three-dimensional structure of rLinCas2C_Lai reveals that it has two α-helices (α1 and α2) and three anti-parallel β-strands (β1-β3). However, unlike conventional Cas2 proteins, two β-strands at the C-terminus, β4 and β5, are missing in rLinCas2C_Lai **(Fig. 3.12C)**. Structural superimposition of rLinCas2C_Lai reveals that it has an identical DNA binding loop L1 as rLinCas2, but a shorter loop L2 **(Fig. 3.12D)**. However, shorter loop L2 may be the possible reason behind rLinCas2C_Lai displaying activity despite expressing the truncated protein.

After conducting a structural homology search using the web server DALI (Holm, 2022), the closest homologs to rLinCas2C were BhaCas2 and SpyCas2 based on scores of selected parameters, including Z-score and root mean square deviation (Table 3.2). To further analyze the putative DNA and RNA binding loops of rLinCas2C, its crystal structure was superimposed with the structures of Cas2 orthologs (SpyCas2C, BhaCas2C, DvuCas2C, and SsoCas2) **(Fig. 3.12E)**. Results showed that Cas2 proteins (SpyCas2C, BhaCas2C, and DvuCas2) have identical L1 and L2 loop sizes to rLinCas2C. Additionally, to understand the rLinCas2C_Lai divergence in the putative substrate-binding loop, its modeled structure was superimposed with the SpyCas2, BhaCas2, DvuCas2, and SsoCas2. The putative loop L1 size of rLinCas2C_Lai aligns with SpyCas2, BhaCas2, and DvuCas2 but not SsoCas2. However, the putative loop L2 of rLinCas2C_Lai was smaller than its orthologs **(Fig. 3.12F-3.12I)**.

A comparison of LinCas2C with its related genes revealed differences in the amino acid sequence of the L1 and L2 loops. LinCas2C_Lai, which is 31% dissimilar to LinCas2C, has a few different conserved residues (His8, Pro37, Phe38, Leu39, Trp44, Asn54, and Lys57)

TH-3470_166106104

compared to their counterparts in LinCas2C (Asp8, Ser37, Val38, Phe39, Leu44, Asp64, and Arg67) **(Fig. 3.11J)**. Another Cas2C paralog, LinCas2C_Linhai, has a 3% difference at the N- (Pro32 and His47) and C-terminal region (Ile78, Glu91, Glu92, Pro93, Ile94, Ile95, and Leu96) compared to LinCas2C **(Fig. 3.11J)**.

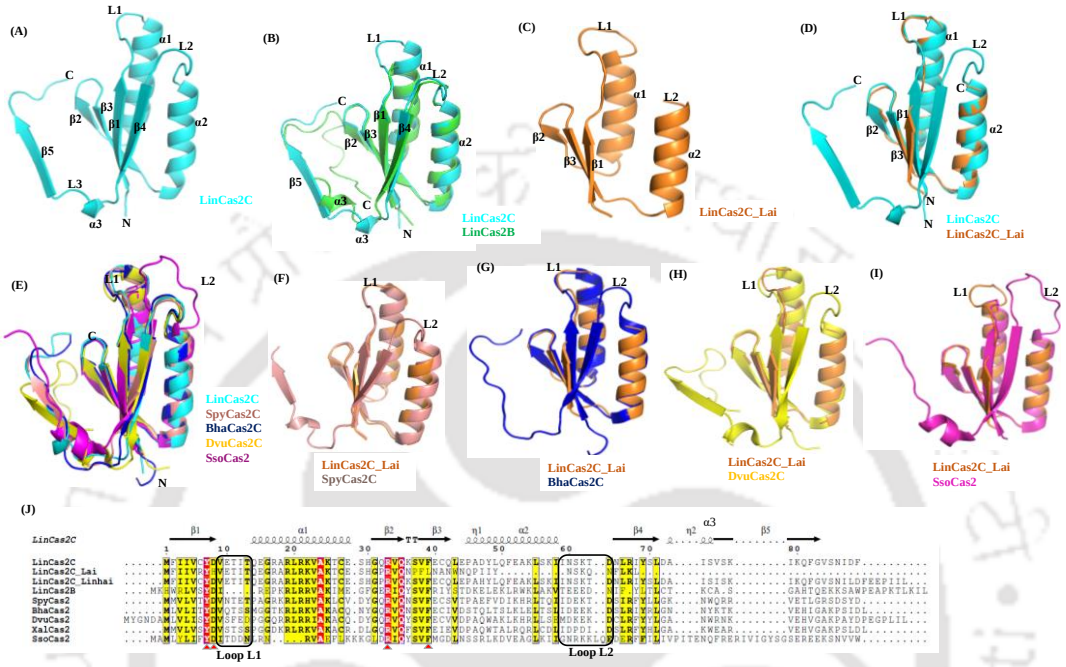


Figure 3.12. The crystal structure of rLinCas2C and its correlation with various orthologs. (A) The crystal structure of rLinCas2C is represented as a cartoon model. All the secondary structural elements, along with the N and C-termini, are labeled for clarity. To map the putative substrate-binding loop (L1 and L2), rLinCas2C structure was correlated with LinCas2B; rmsd: 0.8 (B). (C) Modeled structure of LinCas2C_Lai obtained by the homology modeling via I-TASSER program using the template of SpyCas2 from *Streptococcus pyogenes* serotype M1 (4QR0) with rmsd of 0.6 Å. Structural correlation of LinCas2C with modeleld rLinCas2C_Lai; rmsd: 1.0 (D), SpyCas2; rmsd: 0.9, BhaCas2; rmsd: 0.6, DvuCas2; rmsd: 0.9, SsoCas2; rmsd: 1.4 Å. (E). Structural correlation of modeled LinCas2C_Lai with with SpyCas2 (F), BhaCas2; rmsd: 0.7 Å (G), DvuCas2; rmsd: 1.7 Å (H), SsoCas2; rmsd: 0.8 Å (I). (J) Multiple sequence alignment of rLinCas2C with its orthologs. Two putative substrate-binding loops, L1 (DNA) and L2 (RNA), and secondary structure elements, are labeled. The secondary structural elements on top of the alignment are given according to the rLinCas2C. XalCas2: *Xanthomonas albilineans* (D2UG58), BhaCas2: *Bacillus halodurans* (Q9KFX8), DvuCas2: *Desulfovibrio vulgaris* (Q72WF4), SsoCas2: *Sulfolobus solfataricus* (Q97YC2), SpyCas2: *Streptococcus pyogenes* (Q99YS8). Loop L1 and L2 are marked with rectangles. Red triangles highlight conserved residues.

TH-3470_166106104

Table 3.1. Data collection and refinement statistics of rLinCas2C. The values in parenthesis are for the last resolution shell.

Parameters	rLinCas2C
Wavelength (Å)	1.5418
Temperature (K)	100
Space group	*I*422
Unit-cell parameters (Å, °)	*a*=*b*=103.16, *c*=97.30, *α*=*β*=*γ*=90.00
Resolution (Å)	72.94 - 2.60 (2.72 - 2.60)
No. of observed reflections	140737 (17109)
No. of unique reflections	8390 (1011)
Mn(I) CC (1/2)	0.998 (0.959)
Completeness (%)	100 (100)
V_M ($Å^3$ Da^{-1})	3.19
Solvent content (%)	61.44
Mosaicity (°)	0.60
Mean I/σ(I)	12.3 (4.8)
R_{merge}† (%)	15.8 (50.4)
R_{pim} (%)	5.6 (17.8)
R_{meas} (%)	16.8 (53.5)
Multiplicity	16.8 (16.9)
R_{work}/R_{free} (%)	20.87/26.96
Protein model	
No. of subunits in ASU	2
Protein atoms	1428
Water molecules	51
Other molecules	1
Deviation from ideal geometry	
Bond length (Å)	0.013
Bond angles (°)	1.869
Average *B*-factor ($Å^2$)	
Protein atoms	27.39
Water molecules	33.52
Other molecules	63.98
Ramachandran plot (%)	
Favored	96.53
Allowed	3.47
PDB id	**7F84**

Table 3.2. List of structural homologs of LinCas2C.

Protein	PDB id	rmsd (Å)	Z-Score	Reference
BhaCas2	4ES1	1.2	14.7	(Nam, Ding, et al., 2012)
SpyCas2	4QR2	1.2	14.6	(Ka et al., 2014)
XalCas2	5H1O	1.6	13.8	(Ka et al., 2017)

TH-3470_166106104

TthCas2	1ZPW	1.8	11.3	(Seto et al., 2003)
DvuCas2	3OQ2	1.9	13.4	(Samai et al., 2010)
PhoCas2	6K2E	1.9	10.4	-
TonCas2	5G4D	1.9	10.5	(Jung et al., 2016)
PfuCas2	2I0X	2.0	10.1	-
TdeCas2	6JHZ	2.2	9.3	-
SpyCas2	5ZYF	2.5	9.7	(Ka et al., 2014)
SsoCas2	2I8E	2.5	9.3	(Beloglazova et al., 2008)
EcoCas2	4MAK	-	-	-

The crystal structure of Cas2 orthologs enlisted in Table 3.2 agrees with the asymmetric unit of the rLinCas2C crystal, which contains a dimeric assembly of two protein subunits **(Fig. 3.13A)** (Beloglazova et al., 2008; Ka et al., 2014; Samai et al., 2010). In SpyCas2, upon dimerization, the buried surface area is 2793-2856 $Å^2$, forming 29-32 hydrogen bonds between the two protomers (Ka et al., 2014). In rLinCas2C, two protomers form 33 hydrogen bonds upon dimerization with a buried surface area of 3430 $Å^2$ (out of 9920 $Å^2$), identified by PDBePISA (Velankar, Best, et al., 2010) and WinCoot (Corradi, 2015). The rLinCas2C dimer is characterized by a specific interaction between two protomers. This interaction involves the β5 strand (6 residues) of one protomer, which forms hydrogen bonds with the β4 strand (5 residues) of the other protomer. Additional residues at the dimeric interface are listed in Table 3.3. Some residues, namely Asp8, Asp64, Leu66, and Arg67, are conserved in LinCas2C orthologs. However, rLinCas2C_Lai lacks β4 and β5 strands and still appears to exist in a trimeric state when in solution.

The trimeric structure of LinCas2C_Lai was predicted by generating a symmetry mate. Upon trimeric structural analysis, it was found that Arg6, His8, Gln35, and Asn36 of one protomer interact with Gln35, Asn36, Arg6, and His8 of the second protomer, respectively. Furthermore, Arg17 of the third protomer interacts with Ser29 of the second protomer at a distance of ≤3.5 Å. This interaction is the probable reason behind the self-assembly of rLinCas2C_Lai as a trimer in solution.

The interface of rLinCas2C and DNA demonstrates that each protomer of the nuclease cleaves the two complementary strands of DNA **(Fig. 3.13B)**. The residues of rLinCas2C that interact with DNA primarily come from the L1 loop, L2 loop, and α1 region. This is consistent with the heterocomplex Cas1-Cas2-dsDNA structure of *E. coli*, which reveals that the L1 loop of Cas2 interacts with dsDNA dsDNA (Nuñez, Harrington, et al., 2015). The surface electrostatic potential mapping of rLinCas2C revealed that the putative nucleic acid substrate-

TH-3470_166106104

binding loop (L1 and L2) and the α1 region have a positive charge **(Fig. 3.13C)**. Table 3.4 lists the residues of rLinCas2C that interact with dsDNA, and among these interacting residues, seven residues (Agr17, Arg21, Arg33, Lys36, and Lys62) are positively charged.

The obtained crystal structure of rLinCas2C (in dimeric form) suggests that it is in a catalytically inactive conformational state, as the distance between the conserved Asp8 residue of each protomer is 11.0 Å (compared to 6.5 Å in SsoCas2) **(Fig. 3.13D)**. The distance of 11.0 Å is considered too far to coordinate a single Mg^{2+} ion of the protein. Similarly, the protomers of SpyCas2 (11.4 Å), BhaCas2 (10.6 Å), DvuCas2 (15.4 Å), and HpyVapD (12.6 Å) measured uneven distance between the conserved equivalent aspartate residue **(Fig. 3.13E)** (Beloglazova et al., 2008; Bertelsen et al., 2021; Ka et al., 2014; Nam, Ding, et al., 2012; Samai et al., 2010).

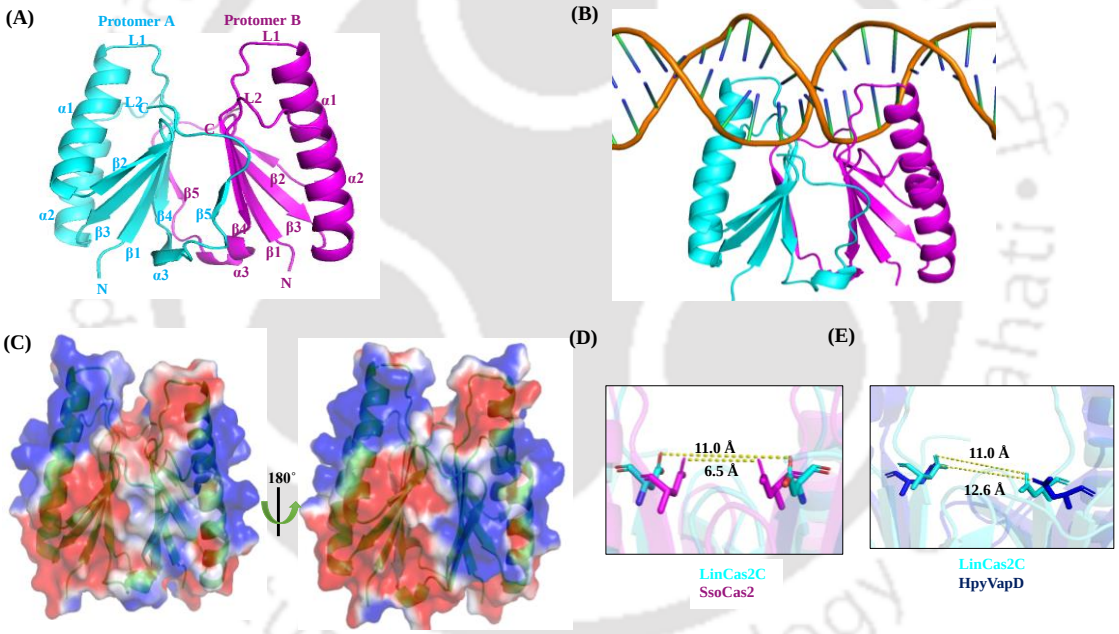


Figure 3.13. The dimeric interface of rLinCas2C crystal structure. (A) The crystal structure shows a dimeric form of rLinCas2C, where one protomer is shown in cyan and another in magenta. (B) Putative DNA binding region on LinCas2C. LinCas2C bound DNA model was proposed using the template of *Synechocystis* Cas1-Cas2/prespacer binary complex structure (PDB id:7CR6). (C) Surface electrostatic potential map of rLinCas2C. The positive and negative charges are blue and red, respectively (scale 5 kT/e to +5 kT/e for red and blue, respectively). (D) Comparison of the distance between Asp residues side chain of the two protomers of rLinCas2C and SsoCas2 and (E) rLinCas2C-HpyVapD.

Table 3.3. Residues of rLinCas2C protomer A interacting with protomer B (≤ 3.5 Å).

Protomer A	Residues	Protomer B	Residues	No. of H-bond
β-1	Asp8	Turn	Lys36	1
β-2	Gln35	β-1, β-4	Cys6, Asn65	1,2
Turn	Lys36	β-1	Asp8	1
α-2	Glu52	β-5	Lys80	1
Loop 2	Thr63	Turn	Asn86	1
Loop 2	Asp64	β-5	Val84	1
β-4	Asn65	β-2, β-4	Gln35, Val84	2,1
	Leu66	β-5	Phe82	2
	Arg67	β-3, β-5	Glu40, Gln81	2,1
	Ile68	β-5	Lys80	2
	Ser70		Lys78, Lys80	1,1
α-3	Ser75	β-4	Asp72	2
β-5	Ser77	β-4	Ser70	2
	Lys78		Ser70	1
	Lys80		Ile68	2
	Val84	Loop2, β-4	Asp64, Asn65	1,1
	Phe82	β-4	Leu66	2
	Gln81		Asn65, Arg67	1,1

Table 3.4. Residues of rLinCas2C protomers interacting with ds-DNA (distance of ≤ 3.5 Å).

Region	Protomer A	Protomer B
β1	Asp8	Asp8
Loop 1	Val9	Val9
	Glu10	Glu10
	Thr11	Thr11
	Ile12	Ile12
	Thr13	-
α1	Gln14	Gln14
	Arg17	Arg17
	Leu20	Leu20
	Arg21	-
β2	Arg33	Arg33
Turn	Lys36	Lys36
Loop 2	Asn60	Asn60
	Lys62	Lys62
	Thr63	Thr63
	Asp64	Asp64
	Asn86	-

3.2.11 Recombinant LinCas2C mutants and their activity.

A comparative analysis of LinCas2C showed significant similarity with SpyCas2, XabCas2, BhaCas2, DvuCas2, and SsoCas2. The maximum sequence similarity of LinCas2C was found with SpyCas2, which was 45% similar and had a query coverage of 98%. LinCas2C was 41% similar to XabCas2, 39% similar to BhaCas2, 37% similar to DvuCas2, and 30% similar to SsoCas2. The study identified conserved residues and motifs in LinCas2C, including Tyr7, Asp8, Ala24, Arg33, Gln35, Leu55, Leu71, RVQ, and SVF **(Fig 3.12J)**. A previous study showed that the mutation of Asp10 in LinCas2B caused a loss in its DNase activity but did not affect its RNase activity (Dixit et al., 2016). A model of LinCas2C was proposed, with a metal-ion binding to specific residues. The residues Tyr7 and Asp8 were identified as putative metal-binding residues of LinCas2C. Residues Arg33 and Phe39, located near metal-binding and active site grooves, were conserved among Cas2 homologs **(Fig. 3.14A)** (Beloglazova et al., 2008). In this study, additional site-directed mutations were made in rLinCas2C at one or more sites that are likely to be involved in nuclease activity. The purified recombinant LinCas2C mutant variants were obtained **(Fig. 3.14B)**.

In SsoCas2, the residues Arg31 and Phe37 are crucial for nuclease activity homologs (Beloglazova et al., 2008). In rLinCas2C, both single (rLinCas2C^{Y7A}) and double (rLinCas2C$^{Y7A+D8A}$) mutants showed reduced DNase activity. As a result, a partial nick in DNA caused a change in the plasmid's conformation **(Fig. 3.14C)**. Additional mutations were made to rLinCasC, including Arg33 and Phe39, resulting in rLinCas2C$^{R33A+F39A}$ and rLinCas2C$^{Y7A+D8A+R33A+F39A}$, which completely abolished the DNase activity. Another mutant construct (rLinCas2C$^{\Delta L2}$) with L2 deletion was generated to explore the role of loop L2 in rLinCas2 and to analyse the DNase and RNase activity. To our surprise, rLinCas2C$^{\Delta L2}$ exhibited a complete loss of DNase activity **(Fig. 3.14C)**. This contradicts an earlier report (Beloglazova et al., 2008) that suggested that the loop L1 was responsible for DNA substrate recognition. To resolve this inconsistency, rLinCas2C was docked with random DNA, indicating that DNA may interact with rLinCas2C at multiple sites, including the residues Asp60, Lys62, Thr63, and Asp64, which form the loop L2 **(Fig. 3.14D)**.

Using a kit-based fluorogenic RNA substrate, the RNase activity of rLinCas2C was measured and compared to its mutants or rLinCas2C_Lai. The mutant rLinCas2CY7A showed RNase activity very similar to rLinCas2C, whereas the activity of other mutants (rLinCas2C$^{Y7A+D8A}$, rLinCas2C$^{R33A+F39A}$, rLinCas2C$^{Y7A+D8A+R33A+F39A}$, and rLinCas2C$^{\Delta L2}$) was

TH-3470_166106104

moderately reduced. The mutant rLinCas2C$^{Y7A+D8A+R33A+F39A}$ was found to have the most adversely affected RNase activity; however, none of the mutants showed a complete abolition in RNase activity. After heat denaturation, the RNase activity of rLinCas2C and rLinCas2C_Lai was abolished, indicating that the protein is free of RNase contaminant and activity is dependent on protein conformation **(Fig. 3.14E)**. The RNase assay suggested that the residues involved in RNA cleavage are different from those involved in DNA cleavage, making LinCas2C metal-independent RNase.

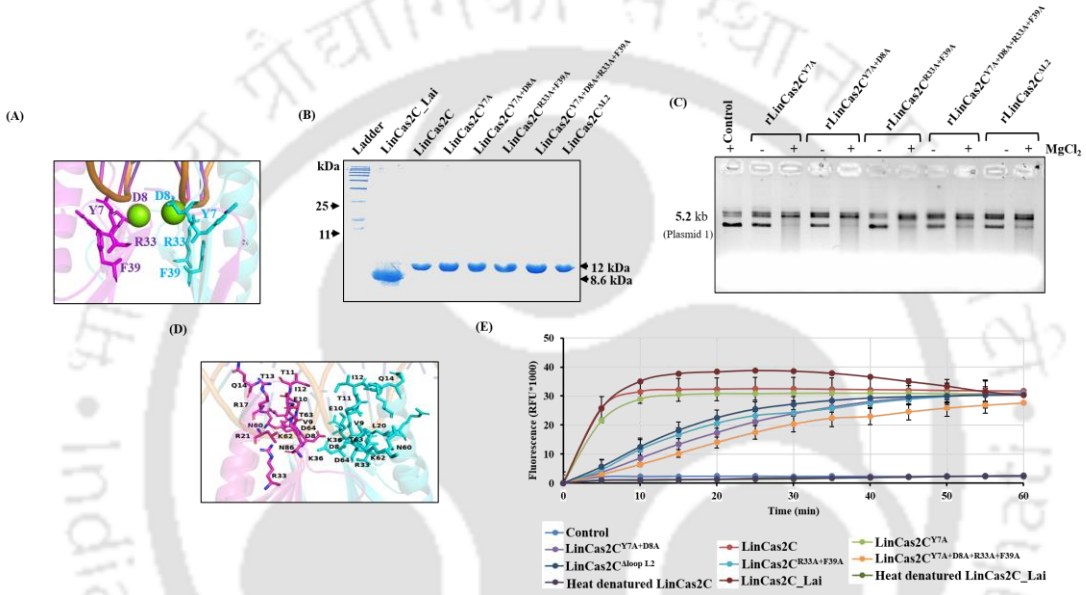


Figure 3.14. Generation of rLinCas2C mutants and assessing its nuclease activity. (A) Purified rLinCas2C or its variant and LinCas2C_Lai resolved on 15% polyacrylamide gel. (B) Proposed model of LinCas2C with Mg^{2+} ion. LinCas2C amino acid residues interacting with metal ions and putative active site are shown in stick form. (C) The nuclease activity of rLinCas2C mutants was evaluated on plasmid-1 (0.5 µg. (D) Interaction of rLinCas2C with non-specific DNA was analyzed by the NPDock webserver. Amino acid residues of rLinCas2C interacting with DNA at a distance of ≤ 3.5 Å are labeled. (E) RNase activity of rLinCas2C or its mutant variants and the rLinCas2C_Lai was quantified using fluorescently labeled RNA substrate. A fluorescent RNA substrate (10 pmol) was incubated with rLinCas2C or its mutants (25 µM) at 37˚C, and fluorescent was recorded at 5 min intervals for one hour.

3.2.12 Nuclease activity of rLinCas4C on double-stranded DNA substrate.

Some of the CRISPR-Cas systems (I-A, I-B, I-C, I-D, I-U, and II-B types or subtypes) include an extra Cas4 protein that plays a functional role in the adaptation process of prespacer DNA maturation and the precise integration of genetic material (Lee et al., 2019; Makarova et al., 2015). LinCas4B from sv. Copenhageni CRISPR-Cas I-B demonstrates nuclease activity

TH-3470_166106104

on various DNA substrates (Dixit, Anand, et al., 2021b). It was thus interesting to conduct a comparative study of LinCas4C and LinCas4B of *Leptospira* to evaluate its similarity in nuclease activity. To assess the nuclease activity of rLinCas4C, a circular double-stranded DNA (dsDNA; plasmid) substrate was used with increasing concentrations of rLinCas4C (1-3 µM). The optimal cleavage of circular ds-plasmid DNA substrate by rLinCas4C (3 µM) was exclusively observed in the presence of Mn^{2+} and Fe^{2+} ions. whereas activity was not observed in the presence of Zn^{2+}, Ca^{2+}, Cu^{2+}, and Ni^{2+} ions **(Fig. 3.15A and 3.15B)**. The DNase activity of rLinCas4C was also evaluated in the presence at different pH range. The most pronounced activity was observed at pH 8.0 **(Fig. 3.15C)**. The versatility of rLinCas4C nuclease activity was also examined on linear ds-DNA substrate, which is a linearized plasmid. In the presence of Mn^{2+} ion, rLinCas4C cleaves linear ds-DNA substrate **(Fig. 3.15D)**. The DNase activity of rLinCas4C was found to be similar to that of the well-studied LinCas4B, *Pyrobaculum calidifontis* Cas4 (PcaCas4), and SsoCas4 (Dixit, Anand, et al., 2021b; Lemak et al., 2013).

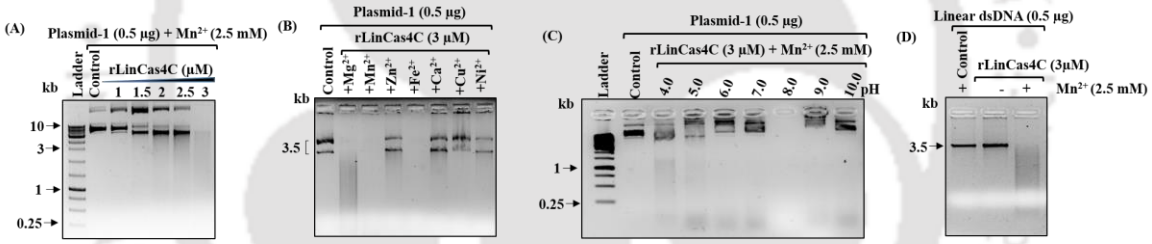


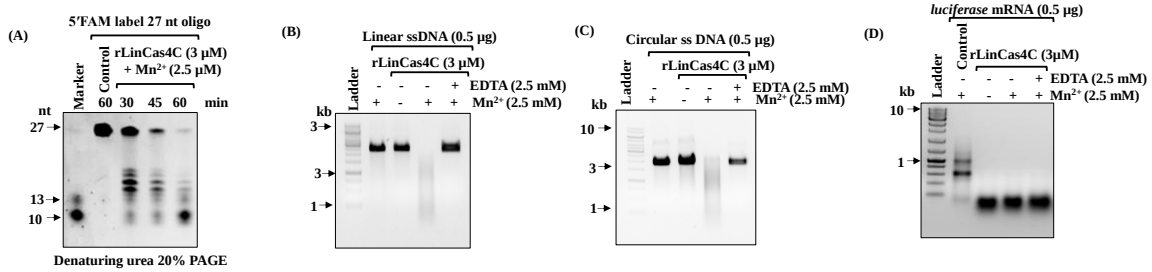
Figure 3.15. DNase activity of rLinCas4C on double-stranded DNA substrates. Nuclease activity was accomplished for an hour at 37°C. (A) Concentration-dependent DNase activity of rLinCas4C on plasmid DNA substrate. (B) Metal-ion dependent activity of rLinCas4C in the presence of different divalent metal-ions (2.5 mM $MgCl_2$, $MnSO_4$, $NiSO_4$, $FeCl_2$, and $CaCl_2$). (C) DNase activity of rLinCas1C evaluated at different pH (4-10). (D) DNase activity of rLinCas4C on linear dsDNA in the presence or absence of metal ion. Plasmid -1 (Circular dsDNA substrate): pTZ57R/T plasmid;500 ng, linear dsDNA: linearized pTZ57R/T plasmid;500 ng, monovalent metal-ion: 50-150 mM, divalent metal-ion: 2.5 mM, EDTA: 2.5 mM, Protein ladder: pre-stained protein ladder (Himedia, cat. No. MBT092) and DNA Ladder: GeneRuler 1 kb DNA Ladder (Thermo Scientific, cat. No. SM0311). Nuclease reaction products were analysed on ethidium bromide-stained 1% agarose gel.

TH-3470_166106104

3.2.13 Nuclease activity of rLinCas4C on single-stranded DNA and RNA substrate.

In previous studies, it has been shown that both PcaCas4 and SsoCas4 exhibit metal-dependent 5′-to-3′ exonuclease activity on single-stranded DNA substrates (Lemak et al., 2013; Lemak et al., 2014). In contrast, LinCas4B can completely degrade single-stranded oligos (Dixit, Anand, et al., 2021b). On the other hand, the BhaCas4 cleavage process is highly specific to a particular sequence, and it only processes the PAM end of the prespacer (Lee et al., 2019). Like in other type I systems where DnaQ-like exonucleases are associated with prespacer processing, it is likely that cellular non-specific exonucleases play a role in trimming the non-PAM end of the prespacer (Hu et al., 2021; Kieper et al., 2018).

In this study, a 27-nucleotide oligomer with a 5′ FAM label was designed to test the site-specific cleavage of a single-stranded oligomer by rLinCas4C. In contrast to LinCas4B, LinCas4C did not completely degrade single-stranded oligos. However, sequence-specific cleavage was observed **(Fig. 3.16A).** In the absence of Cas1 and Cas2, SsoCas4 and LinCas4B proteins have been shown to exhibit non-specific nuclease activities (Dixit, Anand, et al., 2021b; Lemak et al., 2013). In contrast, BhaCas4 acts as a sequence-specific endonuclease, in the presence of Cas1 and Cas2 (Dhingra et al., 2022). Furthermore, the nuclease activity of rLinCas4C was assessed on circular (ΦX174) and linear (M13mp18) single-stranded DNA substrates. It was discovered that rLinCas4C can cleave both circular and linear ss-DNA substrates in the presence of Mn^{2+} ions. However, in the absence of Mn^{2+} ions and in the presence of EDTA, the DNase activity was eliminated **(Fig. 3.16B and 3.16C).** Previous studies also suggest that rLinCas4B can cleave circular and linear ssDNA when Mn^{2+} ions are present (Dixit, Anand, et al., 2021b). Additionally, the nuclease activity of rLinCas4C was examined on a *luciferase* mRNA substrate. Interestingly, rLinCas4C exhibited RNase activity even in the absence of Mn^{2+} ions and in the presence of EDTA **(Fig. 3.16D)**. Henceforth, rLinCas4C exhibits both Mn^{2+} ions dependent DNase activity and Mn^{2+} ions independent RNase activity.



TH-3470_166106104

Figure 3.16. Nuclease activity of rLinCas1C on single-stranded DNA and RNA substrates. Nuclease activity was accomplished for an hour at 37°C unless specified. (A) The DNase activity of rLinCas4C on ss-oligo (27-mer) in the presence of metal ions. (B) DNase activity of rLinCas4C on linear ssDNA substrate (M13mp18; 6.4kb, 500 ng), and (C) circular ssDNA substrate (ΦX174; 5 kb, 500 ng) in the presence or absence of metal-ion. (D) RNase activity of rLinCas4C on *luciferase* mRNA substrate (500 ng). rLinCas4C: 3 µM, EDTA: 2.5 mM, $MnSO_4$: 2.5 mM and Ladder: GeneRuler 1 kb DNA Ladder (Thermo scientific, cat. No. SM0311). Reaction products of (C) were analysed on 8 M 20% denaturation PAGE, and the FAM-labeled product was visualised directly. Nuclease reaction products (B), (C), and (D) were analysed on ethidium bromide-stained 1% agarose gel.

3.2.14 LinCas1C-LinCas2C interacts to form heteroprotein complex.

Adaptation module proteins Cas1 and Cas2 are present in all CRISPR-Cas systems (Makarova et al., 2015). In *E. coli*, *Pyrococcus furious*, *Synechocystis*, *Geobacter sulfurreducens*, *Pectobacterium atrosepticum,* and *Bacillus halodurans* Cas1 and Cas2 interact to form heterocomplex and further facilitates spacer acquisition in the CRISPR array by two step transesterification reaction (Fagerlund et al., 2017; Hu et al., 2021; Lee et al., 2019; Nunez et al., 2015; Rollie et al., 2015; Tang et al., 2021; Wu et al., 2021). In Cas1-Cas2 heterocomplex generally N-terminal of Cas1 interacts with C-terminal of Cas2 to form heteroproteion complex (Fagerlund et al., 2017; Hu et al., 2021; Lee et al., 2019; Nunez et al., 2015; Rollie et al., 2015; Tang et al., 2021; Wu et al., 2021).

However, in *Pyrococcus furiosus* one of the residues at the C-terminal of Cas1 N276 interacts with E83 of Cas2, as charged reversal mutation of these residues abolishes Cas1-Cas2 interaction (Tang et al., 2021). LinCas1B and LinCas2B of *L. interrogans* also reported to form heteroprotein complex, despite LinCas1B having shorter N-terminal region (Dixit, Prakash, et al., 2021). Using PfuCas1-2C crystal structure, the LinCas1-2C heteroprotein complex was modeled. Interestingly in LinCas1-2C modeled hetero complex, the C-terminal of LinCas2C interacts with the C-terminal of LinCas1C, where Ser61 and Lys62 of LinCas2C form polar contact with Lys223 and Glu214 of LinCas1C, respectively, at a distance of ≥ 3.5 Å **(Fig. 3.17A)**; which was also witness in PfuCas1-2 heterocomplex.

In silico analysis unveiled two residues of LinCas1C interacting with LinCas2C, prompting further investigation into their *in vitro* interaction.

TH-3470_166106104

We have demonstrated LinCas1C-2C interaction via pull-down assay. His tagged LinCas2C and untagged LinCas1C were co-expressed (LinCas1C-His:LinCas2C) and purified by affinity chromatography. Here, untagged Lincas1C was expected to co-elutes along withHis tagged LinCas2C. To check whether untagged LinCas1C got co-eluted along with His tagged LinCas2C, immunoblot was performed using anti-rLinCas1C. Immunoblot revealed untagged LinCas1C got co-eluted along with His tagged LinCas2C, indicating LinCas1C and LinCas2C interact with each to form a hetero-protein complex **(Fig. 3.17B)**. Untagged LinCas1C do not bing with N-NTA beads nor to LinCas6B **(Fig. 3.17B).** LinCas1C and LinCas2C interaction provide strong evidence regarding their involved in spacer acquisition during adaptation of CRISPR based immunity. The rLinCas1C and rLinCas2C interaction was also studied via spectroscopic method. LinCas1C contains two tryptophan residues, whereas LinCas2C lacks tryptophan, so it was speculated that the intrinsic fluorescence of LinCas1C tryptophan increases or decreases upon interaction with LinCas2C. Interestingly, the intrinsic fluorescence of rLinCas1C was found to decrease upon the increasing concentration of rLinCas2C, indicating an interaction between both proteins **(Fig. 3.17C).** These results conclude rLinCas1C and rLinCas2C interact to form a hetero-protein complex.

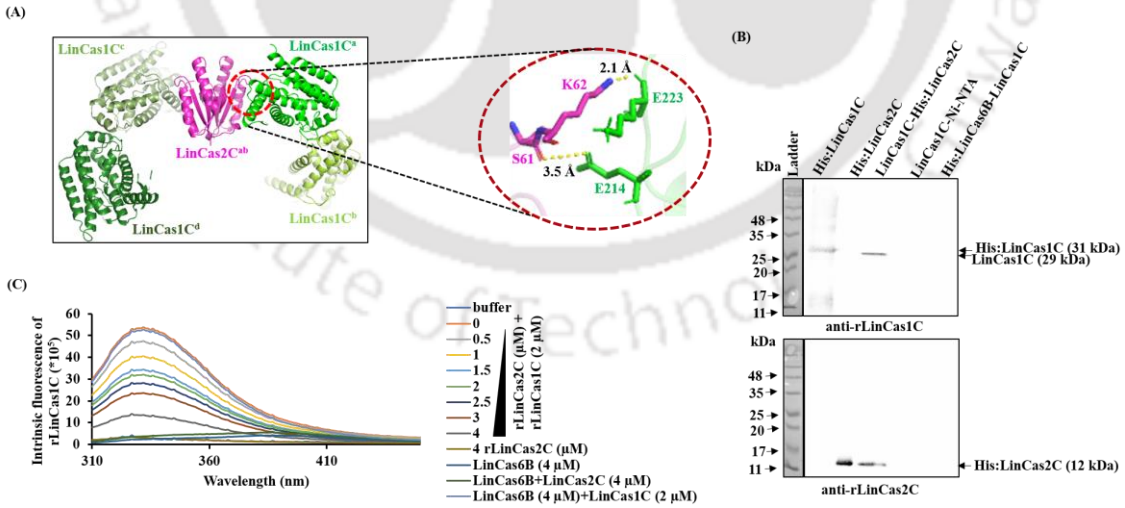


Figure 3.17. The rLinCas1C and rLinCas2C interaction analysis. (A) *In silico* LinCas1C-LinCas2C interaction analysis. LinCas1C and LinCas2C hetero-protein complex was modeled using *Pyrococcus furiosus* Cas1 and Cas2 crystal structure as a template (PDB ID: 7EI1). The interacting residues were measured at a distance of ≤

TH-3470_166106104

3.5 Å. (B) Interaction analysis of rLinCas1C and rLinCas2C via pull-down assay. LinCas1C and His:LinCas2C were co-expressed, solubilized, bound to Ni-NTA beads, and eluted using 250 mM imidazole. LinCas1C or LinCas1C-His:LinCas6 were expressed, solubilized, bound to Ni-NTA beads, and used as a negative control to ensure LinCas1C do not interact with Ni-NTA beads and LinCas6. Immunoblot to confirm the presence of LinCas1C in the elutes (upper panel). Immunoblot to confirm the presence of His:LinCas2C in the elutes (lower panel). Primary antibody (anti-rLinCas1C and anti-rLinCas2C): 1:1000 dilution. Secondary antibody HRP-conjugated anti-mouse secondary antibodies at 1:5000 dilution. (C) Interaction analysis of rLinCas1C and rLinCas2C by spectroscopic method. rLinCas1C (2 µM) mixed with an increasing concentration of rLinCas2C (0.5 to 4 µM). The fluorescence of tryptophan was recorded at 295 nm excitation and 310–570 nm emission spectra in FluoroMax-4 (HORIBA). LinCas6 and BSA (4 µM) were used as a negative control.

3.2.15 To study naive adaptation using CRISPR-Cas I-C adaptation machinery in a heterologous host.

The aim of the investigation was to study the acquisition of spacers by introducing the adaptation module from *Leptospira* CRISPR-Cas I-C into *E. coli* BL21cells, which does not have a native CRISPR-Cas system. To conduct this study, plasmids (pHis.*Lincas1-2C*, pET28a.*Lincas4C*, and pCDF.*LIC_Cr2* array) were used to transform *E. coli* BL21 cells. The spacer acquisition assay, conducted in these cells, showed no increase in the size of the parental CRISPR array (564 bp) from *Leptospira* CRISPR-Cas I-C adaptation module. This lack of expansion was attributed to the absence of insertion of a new repeat-spacer unit (70-74 bp) at the leader-repeat junction **(Fig. 3.17A)**. Additionally, no evidence of spacer acquisition was observed in the *E. coli* BL21 CRISPR array, which maintained an approximate size of 400 bp **(Fig. 3.17B).** Nevertheless, employing a comparable methodology, spacer acquisition was demonstrated in the *E. coli* CRISPR array through the utilization of *E. coli* Cas1 and Cas2 enzymes. This resulted in an expansion of the CRISPR array's size. The implication is that unlike LinCas1C, LinCas2C, and LinCas4C, *E. coli* Cas1 and Cas2 are proficient in acquiring and integrating new spacers into the *E. coli* CRISPR array (Dixit, 2021; Yosef et al., 2012). In the *E. coli* type I-E system, spacer integration by the Cas1-Cas2 complex is directed to the leader-proximal repeat by a host factor, the integration host factor (IHF), which binds to a consensus site in the CRISPR leader sequence to induce DNA bending. This allows the Cas1-Cas2 integrase to catalyze the first integration reaction at the leader-repeat junction (Nuñez et al., 2016). In the *S. islandicus* type I-A CRISPR-Cas system, a robust positive regulator (Csa3) of the adaptation system was identified (Liu et al., 2015). Csa3a binds to the cas1 promoter,

TH-3470_166106104

and overexpression of Csa3a elevates Cas1 protein expression, driving the integration of spacers in the CRISPR loci (Liu et al., 2015). In the *P. atrosepticum* type I-F system, the CRP (cAMP receptor protein)-cAMP activates the cas1 promoter in response to glucose levels, driving the expression of the entire *cas* operon (Patterson et al., 2015). The upregulation of *cas* expression within this subtype correlates with increases in both interference and adaptation, fostering an efficient defense response against horizontally acquired genetic elements (Patterson et al., 2015). CRP also increases cas gene transcription in *T. thermophilus* type I-E and type III-A systems (Agari et al., 2010). In vivo spacer acquisition studies (in *E. coli*) revealed that the type I-F system from *P. aeruginosa* requires an intact effector (Cys) complex and a crRNA besides Cas1 and Cas2 for naive adaptation (Vorontsova et al., 2015). Therefore, identifying host factors in *L. interrogans* that are linked to CRISPR adaptation would be advantageous. This could help us develop an effective system to illustrate *L. interrogans* spacer acquisition using a heterologous host.

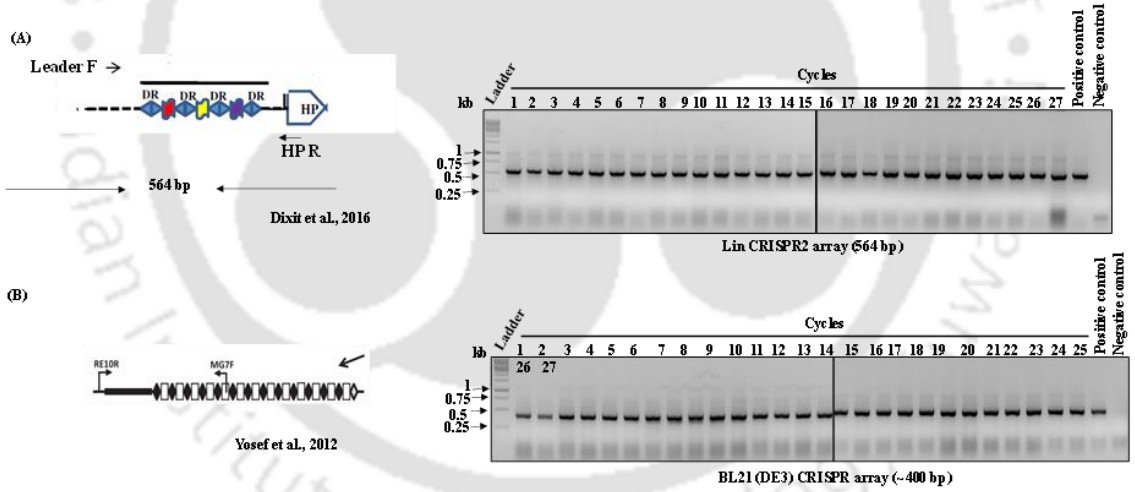


Figure 3.18. Naive adaptation using CRISPR-Cas I-C adaptation module in the heterologous host (*E. coli*). In this experiment pHis-Lin*cas1-2C,* pET28*cas4C,* and pCDF.LIC_Cr^2 was used to study adaptation. All the plasmid was co-transformed in *E. coli* BL21 (DE3) cells and cyclic induction was given using IPTG (1 mM) for 24 hours each cycle. (A) Expansion in CRISPR array of LinCRISPR2. (B) Expansion of *E. coli* BL21 (DE3) CRISPR array. Analysed on 2% agarose gel. Expansion in the CRISPR array was not observed.

TH-3470_166106104

3.3 Discussion

This study sheds light on the adaptation Cas proteins of *L. interrogans* sv. Copenhageni. Cas1 and Cas2 self-assemble to form a heterohexameric complex and facilitate the integration of spacers by dual transesterification reactions. These reactions take place via the nucleophilic attack of the 3'-hydroxyl on both strands of a double-stranded prespacer substrate at the phosphodiester backbone within the CRISPR array (Babu et al., 2011; Lee et al., 2019). Additional Cas proteins, such as Cas4, are essential for adaptation in some CRISPR-Cas systems. Cas4 is present in type I, II, and V systems (Hudaiberdiev et al., 2017). *In vivo* experiments proved that deleting Cas4 reduces the efficiency of adaptation and leads to the integration of non-functional spacers from regions that lack a proper PAM (Li et al., 2014; Liu et al., 2015). The adaptation module in such cases is not limited to Cas1 and Cas2 but comprises the Cas1 protein complexed with Cas2 and Cas4. Two Cas1 dimers encircle a Cas2 dimer at the center, with a single Cas4 subunit connecting to each Cas1 dimer on one end. The complex has a stoichiometry of $Cas4_2$:$Cas1_4$:$Cas2_2$ (Dhingra et al., 2022).

BhaCas1, AfuCas1, and EcoCas1 exist in a dimeric state, and Cas1's N-terminal region is involved in dimerization (Kim et al., 2013; Lee et al., 2019; Nuñez et al., 2014a). The dimerization of the N-terminal domain (NTD) involves three secondary structural elements: β1, β6, and β8, through polar and non-polar interactions (Kim et al., 1997). Remarkably, *interrogans* sv. Copenhageni's LinCas1 has a shorter N-terminal region compared to its orthologs and exists in a monomeric state in the solution. To integrate a spacer, Cas1 and Cas2 proteins form a complex, and in a heteroprotein complex, Cas1 and Cas2 exist in a dimeric conformation (Nuñez et al., 2014a). Therefore, LinCas1C may interact with Cas1 proteins from other subtypes to form a dimer or work with an auxiliary protein to assist in spacer integration.

Recombinant LinCas1C protein that has been adapted to have DNase activity on different types of DNA substrates, such as circular double-stranded DNA, linear and circular single-stranded DNA. This DNase activity is dependent on metal ions. LinCas1B, *Riemerella anatipestifer* Cas1 (RanCas1), and PaeCas1 have also been found to exhibit metal-ion-dependent DNase activity on DNA substrates (Dixit, Prakash, et al., 2021; Wiedenheft et al., 2009). AfuCas1, like LinCas1C, can bind to linear double-stranded DNA substrates (Kim et al., 2013). Both LinCas1C and LinCas1B cannot cleave oligos of 36- and 50-mer, which is consistent with the behavior of *E. coli* Cas1 (EcoCas1) (Nuñez et al., 2014a). However, the

TH-3470_166106104

substrate specificity of most Cas1 proteins, such as LinCas1B, AfuCas1, PaeCas1, EcoCas1, and RanCas1, is not yet clear.

The recombinant forms of LinCas2C and the naturally truncated LinCas2C_Lai have been found to possess nuclease activity on various DNA substrates, including circular double-stranded DNA, linear and circular single-stranded DNA. This activity is dependent on the presence of divalent metal ions and the pH of the reaction. However, these nucleases were inactive towards small DNA oligos measuring 23-50-mer. In a recent study, LinCas2C_Linhai prefers Mg^{2+} to exhibit nuclease activity (Xiao et al., 2019). However, *Legionella pneumophila* Cas2 (LpnCas2) and *Thermus thermophilus* Cas2 (TthCas2) demonstrated nuclease activity in the presence of Mn^{2+} (Gunderson et al., 2015; Nam, Ding, et al., 2012). The DNase activity of rLinCas2C and rLinCas2C_Lai was consistent with the other reported Cas2 proteins, including BhaCas2 (Nam, Ding, et al., 2012), SpyCas2 (Ka et al., 2014), XorCas2 (Makarova et al., 2011b), and LinCas2B (Dixit et al., 2016). Similar to LinCas2B activity, rLinCas2C was inactive towards single-stranded short oligos (Dixit et al., 2016). The recombinant Cas2 nucleases (LinCas2C and LinCas2C_Lai) of the two sv. of *Leptospira* are divalent metal-independent RNase. In contrast, Sso8090Cas2 homologs from *Sulfolobus solfataricus*, TmaCas2 of *Thermotoga maritima*, MthCas2 of *Methanobacterium thermoautotrophicum*, AfuCas2 of *Archaeoglobus fulgidus*, LpnCas2, and NeuCas2 of *Nitrosomonas europea* exhibited metal-dependent RNase activity (Beloglazova et al., 2008). Detection of nuclease activity in rLinCas2C_Lai suggests that the conserved residues at the N-terminal are more involved in nucleic acid catalysis. The metal-independent RNase activity rLinCas2C indicates its additional role beyond CRISPR biology.

A recent study found that the virulence-associated protein D (VapD) of toxin-antitoxin systems (TA) possesses a ribonuclease fold similar to Cas2 proteins (Bertelsen et al., 2021). VapD toxins are metal-independent nucleases that can moderate gene expression by breaking down specific, stable RNAs. VapD is part of a toxin-antitoxin (TA) system with other homologs such as RelE, MazF, and VapC VapC (Bertelsen et al., 2021; Goeders & Van Melderen, 2014; Kwon et al., 2012). These homologs function as RNases. The bacterial CRISPR-Cas immunity systems may have evolved from a primordial vapXD-type TA system, as the structural similarity of VapD with Cas2 supports this notion (Makarova et al., 2013). The structure analysis of rLinCas2C revealed that it exists as a dimeric and apostate conformation, with each subunit of the protein containing the signature ferredoxin fold. The

TH-3470_166106104

structure of rLinCas2C confirms the evolutionary conservation of the VapD/Cas2-like ribonuclease protein fold (Bertelsen et al., 2021; Shaibullah et al., 2023). In TonCas2, each protomer's C-terminal β5 region interacts with the other molecule's β4, forming a β-sheet of five strands in both subunits (Jung et al., 2016). The structure of rLinCas2C explains how the catalytic aspartate plays a crucial role in limiting conformational freedom. It is possible that each site may contain a symmetrically bound metal ion since divalent cations take part in catalysis for DNA.

The function of rLinCas2C's metal-independent RNase activity is similar to that of HpyVapD from *H. pylori* and *Burkholderia pseudomallei* Cas2/VapD but different from that of SsoCas2 (Kwon et al., 2012; Shaibullah et al., 2023). In SsoCas2, the phosphodiester cleavage is initiated by the coordination between Mg^{2+} and two Asp10 residues from two dimer subunits (Beloglazova et al., 2008). However, for HpyVapD or rLinCas2C, the distance between the side chains of two instances of Asp7 or Asp8 is greater than 10 Å, indicating that they are unable to coordinate the metal ion (Kwon et al., 2012). It is noteworthy that HpyVapD demonstrates ribonuclease activity even without the addition of metal ions. However, based on studies of HpyVapD and comparisons with Cas2, two aspartate residues (Asp7 and Asp76) have been suggested as strong candidates for the catalytic site of VapD (Kwon et al., 2012).

The closest homolog of LinCas2C, SpyCas2 and BhaCas2 is a metal- and pH-dependent dsDNase (Ka et al., 2014). Mutagenesis of SsoCas2 (SSO1404) identified six residues essential for RNase activity: Tyr9, Asp10, Arg17, Arg19, Arg31, and Phe37. The study also suggested that Asp10 might be the principal catalytic residue (Beloglazova et al., 2008). However, in DvuCas2, neither Tyr13 nor Phe45 played a catalytic role due to their buried location (Samai et al., 2010). In most RNases, catalysis is dependent on two or three acidic residues that are conserved. These residues coordinate with one or two metal cations, which act as a catalyst to activate a water molecule that acts as a nucleophile to hydrolyze the phosphodiester bond. Alternatively, these residues stabilize the transition state in cleavage reactions (Worrall & Luisi, 2007). The Cas2 family's nuclease activity varies due to the structural difference at its catalytic site. In LinCas2C, substitution of conserved residues Tyr7, Asp8, Arg33, and Phe39 with Ala and loop L2 deletion mutants, leads to abolishment of DNase activity. However, moderate reduction of RNase activity was evident in selected mutants.

LinCas4C, another adaptation Cas protein of CRISPR-Cas I-C, exhibits metal-ion dependent DNase activity on circular ds-DNA substrate similar to LinCas4B (Dixit, Anand, et

TH-3470_166106104

al., 2021b). SsoCas4 (Sso0001 and Sso1391) and PcaCas4 cleave short oligos but not circular ds-DNA (Lemak et al., 2013; Zhang et al., 2012). rLinCas4C cleaves viral ssDNA substrates similar to LinCas4B, PcaCas4, Sso0001, and Sso1391(Dixit, Anand, et al., 2021b; Lemak et al., 2013). LinCas4B completely degrades oligos (Dixit, Anand, et al., 2021b), while LinCas4C cleaves sequence specifically.

The EcoCas1E identifies the adjacent PAM sequence in the protospacer, which is bound by the Cas1-Cas2 nucleoprotein complex. It then processes the protospacer at the PAM before integrating it into the CRISPR loci (Makarova et al., 2015; Nuñez et al., 2014a; Rollie et al., 2015; Wang et al., 2015; Xiao et al., 2017). BhaCas4 promotes the integration of the prespacer at the non-PAM end, while the BhaCas1-2-4 complex mediates the polarized integration of the protospacer at the leader-repeat junction (Dhingra et al., 2022). BhaCas4 protospacer processing requires both BhaCas1 and BhaCas2 and is dependent on the BhaCas4 active site but not on the BhaCas1 active site (Lee et al., 2019). LinCas4C can cleave single-stranded sequence specifically. Sequence specific cleavage of oligos by LinCas4C provides strong evidence of its involvement in prespacer processing and assisting in the adaptation phase of CRISPR-Cas-based immunity.

The recombinant LinCas1C, LinCas2C, and LinCas4C display RNase activity that is independent of metal-ions, in contrary AfuCas1B and EcoCas1E exhibit RNase activity that depends on metal-ions (Babu et al., 2011; Kim et al., 2013). Another Cas protein, LinCas2B, has been reported to have nonspecific RNase activity that does not depend on metal-ions (Dixit et al., 2016). Cas nucleases such as BhaCas5C have been shown to have DNase activity that depends on metal-ions and RNase activity that does not depend on metal-ions (Punetha et al., 2014). The simultaneous occurrence of RNA and DNA hydrolysis, whether in the presence or absence of Mg^{2+} ions, suggests the presence of a singular active site exhibiting variable target selectivity. When DNA is not in close proximity, the low affinity of Cas5C for metal-ions makes it an RNase, enabling it to assist in ribonuclease activity. However, when DNA is close by, the affinity for metal-ions increases, making it a deoxyribonuclease (Punetha et al., 2014). LinCas1C and LinCas4C may share a similar mechanism in terms of DNase activity that depends on metal-ions and RNase activity that does not depend on metal-ions.

DNA and RNA can undergo cleavage at either the 5′ or 3′ position of a scissile phosphate bond. However, the structures and catalytic mechanisms of RNA and DNA

TH-3470_166106104

nucleases are complex and varied. Nucleases use water, (deoxy)ribose, inorganic phosphate, or the sidechains of Ser, Tyr, or His as a nucleophile. Additionally, nuclease activities are strictly regulated by stringent substrate specificity. Catalysis may or may not require metal ions (Gerlt et al., 1983; Yang et al., 2006). Metal-ion independent phosphodiester hydrolysis is generally based on general acid-base catalysis. Here, the 2'-OH of ribose makes an intramolecular nucleophilic attack on the adjacent 3'-phosphate, breaking the RNA backbone and generating 2′,3′-cyclic phosphate and 5′-OH products (Yang, 2011b). DNA lacks a 2′-OH group, so, in this case, the nucleophile is more likely to be a water molecule. The metal-bound water molecule is involved in nucleophile activation. A nucleophile attacking the scissile phosphate from the 3′ side breaks the 5′ O-P bond, producing 3′-phosphate and 5′-OH (Dupureur, 2008; Yang, 2011b).

Recombinants LinCas1C, LinCas2C, and LinCas4C show different preferences for divalent metal ions when carrying out DNase activity. LinCas1C and LinCas2C prefer Mg^{2+} ions, while LinCas4C prefers Mn^{2+} ions. The preference for divalent metal ions is based on their hydration properties and specific ligand requirements (Yang, 2011b). Mg^{2+} and Fe^{2+} ions tend to have six inner-sphere ligands arranged in an octahedral configuration. Meanwhile, Ca^{2+} and Zn^{2+} ions can also have octahedral coordination, but Zn^{2+} ions frequently have four ligands in a tetrahedron, and Ca^{2+} ions have seven, eight, or even nine ligands. Nucleases usually favor Mg^{2+} ions because of their abundance, solubility, and redox stability when compared with Mn^{2+}, Fe^{2+}, and Cu^{2+} ions. Mg^{2+} ions are smaller than Ca^{2+} ions and have more rigid coordination geometry compared to the transition metals Fe^{2+}, Cu^{2+}, Ni^{2+}, and Zn^{2+} (Maguire & Cowan, 2002; Yang, 2011b). The relaxation of substrate specificity observed in metalloenzymes with Mn^{2+} is a common phenomenon, possibly attributed to Mn^{2+} being a transition element with less rigid coordination requirements compared to Mg^{2+}. Numerous enzymes remain active even with just a trace of either Mg^{2+} or Mn^{2+} (Brautigam & Steitz, 1998; Pingoud et al., 2009). Like other divalent cations, Fe^{2+} also plays a role in enzymatic catalysis, such as nonheme iron enzymes. Similar to Mg^{2+}, Fe^{2+} also forms six inner-sphere ligands arranged in an octahedral configuration (Harding, 1999).

Evidence suggests that Cas proteins have a role in various cellular functions. For instance, deleting the *cas1* gene in *E. coli* leads to abnormally elongated, sensitive cells to DNA damage and have weakened chromosome segregation. This was observed after treating the cells with mitomycin C (Babu et al., 2011). Similarly, deleting *cas1* in *H. volcanii* reduces cell

TH-3470_166106104

survival, implying the CRISPR-Cas immune protein's role in DNA repair (Wörtz et al., 2022). The mechanism behind these findings is most likely based on Cas1's ability to cleave branched DNA substrates commonly present during DNA repair and recombination (Babu et al., 2011; Rollie et al., 2015). Moreover, Cas1 has been shown to associate with several repair proteins in *E. coli*, lending credence to the idea that *E. coli* CRISPR-Cas system components play a part in DNA repair (Babu et al., 2011). In the archaeon *Sulfolobus solfataricus*, Csa3a plays a crucial role in regulating the expression of CRISPR adaptation genes (Liu et al., 2015). Similarly, in *Sulfolobus islandicus*, Csa3a has been found to stimulate the expression of adaptation *cas* genes, CRISPR RNA, and DNA repair genes such as helicase *herA*, nuclease *nurA*, and *dpo2* (Liu et al., 2017). The Csa3a protein helps activate and regulate DNA repair genes (Liu et al., 2017; Z. Liu et al., 2020). It is thought that Cas2 orthologs' nuclease properties may enable them to degrade exotic phage DNA or transcripts or inhibit translation by globally cleaving mRNA. Cas2 proteins may use their VapD-like fold to modulate bacterial cell growth and survival during infection by utilizing an intrinsic metal-independent ribonuclease activity (Bertelsen et al., 2021; Gunderson et al., 2015). Moreover, the Cas4 protein family has similarities with RecB and AddB nucleases, which play a vital role in DNA repair mechanisms (Zhang et al., 2012). The endonuclease activity of LinCas1C and LinCas4C may provide an advantage with an additional role other than typical CRISPR adaptation. The current work on deciphering shared protein structure-function relationships between bacterial defence systems is building an exciting future. The global inhibition of translation by mRNA cleavage may be a fundamental principle in the biological role of Cas2 proteins, as reported for TA systems, including RelBE, MazEF, PemIK, and ChpBIK (Beloglazova et al., 2008; Gerdes et al., 2005; Masuda et al., 1993; Zhang & Inouye, 2009; Zhang et al., 2005).

3.4 Materials and Methods.

3.4.1 Bioinformatics analysis.

The CRISPR-Cas I-C adaptation *cas* genes present in *L. interrogans* sv. Copenhageni and Lai were obtained from NCBI. The 3D atomic coordinates of the Cas1 and Cas2 orthologs were downloaded from the Protein Data Bank (PDB) (Berman et al., 2002). The phylogenetic tree LinCas2C and its orthologs were constructed as discussed in Chapter 2. The 3D structures of LinCas1C, LinCas1B, LinCas1C_Lai, LinCas2C_Lai, and LinCas2B were predicted using the programs I-TASSER (Yang & Zhang, 2015) and AlphaFold (Jumper et al., 2021). The

TH-3470_166106104

models were refined, and the energy was minimized using the ModRefiner web server (Xu & Zhang, 2011). Clustal Omega was used to conduct a multiple sequence alignment (Sievers et al., 2011) with the default set of parameters and decorated using the web tool ESPript for better visual effect (Gouet et al., 2003). The program NPDock was used to perform molecular docking of LinCas2C with a non-specific dsDNA (Tuszynska et al., 2015). The protein tertiary structure of Cas1 and Cas2 orthologs was retrieved from the PDB database (Velankar, Alhroub, et al., 2010). The program PyMOL was used to generate the superimposition of structures and figures (DeLano, 2002). The buried surface area of the LinCas2C dimer was calculated using the webserver PDBePISA (Velankar, Best, et al., 2010). The polar contacts between LinCas2C protomers and LinCas2C-DNA interface were identified using PyMOL within a distance of 3.5 Å. LinCas2C with Mg^{2+} ion was modeled using *Enterococcus faecalis* Cas1-Cas2/prespacer ternary complex as a template (PDB id: 5XVP).

3.4.2 Complimentary DNA synthesis and reverse transcriptase PCR (RT-PCR).

The cDNA was synthesised using in *vitro grown L. interrogans* sv. Copenhageni as discussed in chapter 2. Finally, RT-PCR was performed using four different pairs of primers, each capable of amplifying a different length of *Lincas1C* transcript (765, 885, 905, and 1002 bp).

3.4.3 Nucleic acid isolation and cloning.

The spirochete *L. interrogans* serovars Copenhageni or Lai were cultured in Ellinghausen-McCullough-Johnson-Harris (EMJH) media at 29ºC with 1× enrichment media (Difco) and 5-fluorouracil (100 µg/mL). After incubating for seven days, the culture was sub-cultured successively. Genomic DNA of *L. interrogans* sv. Copenhageni and Lai were isolated from a 7-day-old culture containing approximately 1×10^8 cells per ml using the QIAamp DNA Blood Mini Kit (Qiagen) following the manufacturer's protocol.

E. coli strains DH5α and BL21 (DE3) were grown in Luria Bertani (LB, Himedia) broth or agar for cloning, transformation, and expression. The open reading frame (ORF) of LIC12916 (*Lincas1C*, 765 bp), LIC12917 (*Lincas2C*, 273 bp), and LIC12915 (*Lincas4C*, 429 bp) were amplified using the genomic DNA templates of *L. interrogans* serovars Copenhageni. On the other hand, LA0683 (*Lincas2C_Lai*, 272 bp) was amplified using the genomic DNA templates of *L. interrogans* sv. Lai. *Lincas1C* and *Lincas4C* were cloned in the pET28a expression vector (*Nhe*I-*Xho*I) (Novagen), while *Lincas2C* and *Lincas2C_Lai* genes were

TH-3470_166106104

cloned in the pCDF-1b expression vector (*BamH*I-*Sal*I) (Novagen). Clones were confirmed by double digestion and sequencing of plasmids.

3.4.4 Protein overexpression and purification.

The genes (*LinCas1C*, *LinCas2C*, *LinCas2C_Lai*, and *LinCas4C)* ligated in the expression plasmids were transformed into *E. coli* BL21 (DE3) cells, where they were overexpressed using one mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 hours at 37ºC. LinCas1C, LinCas2C, and LinCas4C were purified using native conditions. However, rLinCas2C_Lai remained insoluble in native conditions. As a result, rLinCas2C_Lai was purified using the hybrid method described before (Dixit, Anand, et al., 2021a; Dixit et al., 2016; Prakash & Kumar, 2021). The yield of rLinCas1C, rLinCas2C, rLinCas2C_Lai, and rLinCas4C purification were 2, 5, 0.40, and 0.1 mg per liter, respectively, and the purified proteins were stored at -20˚C till further use.

3.4.5 Size-exclusion chromatography.

Size-exclusion chromatography was carried out on rLinCas2C and rLinCas2C_Lai using a Superdex 200 increase column (GE Healthcare, catalogue no. 28-9909-44) on an AKTA Prime Plus (GE Healthcare). For rLinCas1C, a high-resolution column (10 mm × 300 mm, catalogue no. 7801650) was used with a plate number of SEC650. The column was equilibrated with an equilibration buffer (50 mM Tris-Cl pH 8.0 and 150 mM NaCl) and then calibrated with standard proteins of known molecular mass. The standard proteins used were β-amylase (200 kDa), alcohol dehydrogenase (158 kDa), albumin (66 kDa), carbonic anhydrase (29 kDa), and cytochrome C (12.4 kDa) (Sigma, catalogue no. MWGF-200). Approximately 200 µg of each protein were resolved in SEC.

3.4.6 Generation of polyclonal antibodies against rlinCas1C and rLinCas2C.

Polyclonal antibodies against pure rLincas1C and rLinCas2C were generated in BALB/c mice aged 4-6 weeks. Each mouse was subcutaneously injected with around 20 µg of each protein emulsified in Freund's complete adjuvant (FCA; catalogue no. sc-3727; Santa Cruz Biotechnology) for primary immunization. Two booster doses of each antigen emulsified in Freund's incomplete adjuvant (FIA; catalogue no. 3726; Santa Cruz Biotechnology) were administered via subcutaneous injection after 14 and 20 days of primary immunization. After 30 days of primary immunization, blood was collected from each mouse by retro-orbital bleeding. Serums were pooled from the collected blood through centrifugation at 3000 rpm for 5 minutes, and the titers of the antibodies were estimated through ELISA. These immunization

TH-3470_166106104

experiments were conducted at the Department of Veterinary Microbiology, College of Veterinary Science, Assam Agriculture University Guwahati.

3.4.7 Immunoblotting.

Recombinant proteins (50 ng) each of (LinCas1C, LinCas2B, LinCas2C, and LinCas2C_Lai) along with *Leptospira* lysate (~1×10^9 cells) were separated using SDS-PAGE and transferred onto a nitrocellulose membrane (Himedia) at 100V for 60 minutes in chilled transfer buffer (2.4 g Tris-base and 11.16 g glycine dissolved in 900 mL distilled water and 100 mL Methanol). The membrane was blocked with 5% skim milk in TBS solution (Tris buffer saline solution containing 8 g NaCl and 20 mL of 1 M Tris-Cl pH 7.5 per liter of distilled water) for 2 hours at room temperature on a shaker. Subsequently, the blocking solution was removed, and the membrane was probed with 10 mL of 1:1000 diluted anti-rLinCas1C or anti-rLinCas2C or anti-rLinCas2B in TBST (2% skim milk) for 2 hours at room temperature on a shaker. The membrane was washed thrice with TBST for 10 min each and was probed with 1:10000 anti-mouse IgG HRPO conjugated secondary antibody in TBST (2% skim milk) for 1 hour at room temperature on a shaker. The membrane was further washed thrice with TBST, and the enzymatic activity of HRP was determined by adding a mixture (1:1) of Peroxide Solution A and Luminol substrate Solution B (Thermo Scientific Pierce ECL Western Blotting Substrate). Finally, chemiluminescence was observed in the Chemidoc (Biorad XRS+) at different exposure times.

3.4.8 Nuclease activity assay.

The study investigated the nuclease activity of adaptation Cas proteins on various substrates (DNA and RNA). The RNA transcript of the luciferase gene was synthesized using the HiScribe T7 high-yield RNA synthesis kit (NEB) following the manufacturer's protocol. The plasmid was obtained from a 5 mL overnight grown culture of *E. coli* DH5α cells using a mini-prep kit (Thermo Scientific). The substrates used for nuclease activity of adaptation Cas proteins were circular double-stranded (ds) plasmid DNA (pET28a, pTZ57/R/T, 500 ng), circular single-stranded (ss) DNA (M13mp18, 500 ng), linear ssDNA (Φx174 genome, 500 ng), 23-, -27 -36 and 50-mer nucleotides (400 nM), and firefly *luciferase* mRNA (500 ng) (Table 3S1). Each substrate was independently incubated with proteins in a total reaction volume of 25 µl of nuclease buffer (25 mM Tris-HCl pH 8.0, 100 mM KCl, and 2.5 mM $MgCl_2$ or $MnCl_2$) for an hour at 37°C. DNase activity dependence for divalent metal ions (2.5 mM) was determined by substituting various divalent metal ions ($MgCl_2$, $MnSO_4$, $CaCl_2$, $NiSO_4$,

TH-3470_166106104

$FeSO_4$, $CuSO_4$, and $ZnSO_4$). All reaction products were separated on ethidium bromide-stained 2% (w/v) agarose gel electrophoresis. The nuclease reaction containing 23-, 27-, 36-, and 50-mer nucleotides was assessed on 8 M 20% urea-PAGE.

3.4.9 Site-directed mutagenesis.

Mutant variants of rLinCas2C were created using the Q5 site-directed mutagenesis kit (NEB). The template plasmid pCDF_LIC12917 and primers listed in Table 3S1 were used to generate the mutants. In rLinCas2C, potential residues that could be involved in nuclease activities were substituted with Ala at one or multiple sites to create different mutant variants. The mutant variants include rLinCas2C^{Y7A}, rLinCas2C$^{Y7A+D8A}$, rLinCas2C$^{R33A+F39A}$, and rLinCas2C$^{Y7A+D8A+R33A+F39A}$. In one of the mutant variants called rLinCas2C$^{\Delta L2}$, residues framing the loop L2 were deleted. The generated constructs were sequenced before overexpression, purification, and characterization of proteins.

The RNase activity of rLinCas2C, rLinCas2C_Lai, and the mutant variants of rLinCas2C was analyzed using the RNaseAlert kit (Integrated DNA technology, IDT; Cat # 11-02-01-02). The RNaseAlert kit contains synthetic RNA oligo substrate labeled with fluorescein and a quencher at the end. When cleaved by an RNase, the substrate fluoresces green (490 nm excitation and 520 nm emission) and can be measured using a fluorometer. RNase activity was performed in black flat-bottom 96-well plates (Invitrogen) at 37 °C. Fluorogenic RNA substrate (10 pmol) was incubated with rLinCas2C, its mutant variants, and LinCas2C_Lai (25 µM) in a total of 100 µL reaction buffer (25 mM Tris-Cl pH 8.0 and 100 mM KCl). Fluorescence was measured at 5-minute intervals for 60 minutes using the Infinite M200Pro plate reader (Tecan).

3.4.10 Crystallization, data collection, and structure determination.

The protein rLinCas2C (5 mg/mL) was purified and screened for initial crystal hits using crystallization conditions from Hampton Research. This was done using the hanging-drop vapor-diffusion method at a temperature of 4°C. Eventually, diffraction-quality crystals of rLinCas2C were obtained in a solution containing 0.2 M sodium citrate tribasic dihydrate pH 5.6, 5% 2-propanol, 20% polyethylene glycol (PEG) 4000, and 0.2% low melting agarose (LMA). X-ray intensity diffraction data were collected using the home source Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) and R-Axis IV++ imaging-plate detector available at the central instrument facility (CIF) of the Indian Institute of Technology Guwahati, India at a temperature of -173°C. The crystal-to-detector distance was

TH-3470_166106104

maintained at 170 mm. The diffraction data were processed and scaled using the programs iMosflm (Battye et al., 2011) and Aimless (Evans & Murshudov, 2013) embedded in the CCP4 package (Winn et al., 2011). The intensities were converted to structure factors using the module ctruncate available in the CCP4 package. A summary of X-ray intensity data collection and processing statistics is provided in **Table 3.1**. Initial phases of the protein rLinCas2C were determined employing the molecular replacement method using the crystal structure of SpyCas2 (PDB id: 4QR0) from *Streptococcus pyogenes* having a sequence identity (query coverage) of 45 (98)% as a search model using the program Phaser (McCoy et al., 2007). To calculate the R_{free}, 5% of the total reflections were kept aside as a test data set (Brünger, 1992). The atomic model building and iterative cycles of structural parameters refinement were carried out using Coot (Emsley et al., 2010) and Refmac5 (Murshudov et al., 2011), respectively. The structural quality of the final refined model was validated using the PROCHECK program (Laskowski et al., 1993) and MolProbity program (Chen et al., 2010). The details of the structure refinement and validation of the final structure models are provided in Table 3.1. The three-dimensional atomic coordinates of the protein LinCas2 have been deposited in the RCSB Protein Data Bank (Berman et al., 2000).

3.4.11 Naive adaptation in a heterologous host.

pHis.*Lincas1-2C*, pET28a.*Lincas4C* and pTZ57R/T-LIC_Cr2 plasmid constructs were co-transformed in E. coli BL21 cells. Cyclic induction was given using IPTG (1 mM) for 16 hours at 37˚C. After 25 cycles of induction, 25 µl of TE buffer was added to the cell pellet and heated at 95 ˚C for 20 min to initiate cell lysis and plasmid DNA to get released. After heating, the lysate was spun at 13000 rpm for 20 min, and the supernatant was collected in a 0.5 ml centrifuge tube. One microliter of the supernatant was used in PCR to examine the expansion in the CRISPR array.

3.4.12 LinCas1C and LinCas2C interaction analysis.

3.4.12.1 In silico approach. LinCas1C-LinCas2C hetero-protein complex was modeled employing *Pyrococcus furiosus* Cas1 and Cas2 crystal structure as a template (PDB ID: 7EI1). Amino acids residues participate in interacting at the distance of ≤3.5Å was analysed using PyMol software.

3.4.12.2 Spectroscopic method. Recombinant LinCas1C (2 µM) and increasing concentration of rLinCas2C (0.5 to 4 µM) were mixed in a buffer (containing 25 mM Tris-Cl pH 8.0 and 100 mM NaCl) and incubated at 4˚C for an hour to form a hetero-protein complex as previously

TH-3470_166106104

described (Nunez et al., 2015; Rollie et al., 2015). Thereafter, the fluorescence of tryptophan (present in LinCas1C, not in LinCas2C) was recorded at 295 nm excitation and 310–570 nm emission spectra in FluoroMax-4 (HORIBA).

3.4.12.3 Pull-down assay. *Lincas1C* (LIC12916), it was cloned in pET28a vector (*Nhe*I-*Xho*I), which add His tag at the C-terminal of recombinant protein. To get untagged LinCas1C an stop codon was included in reverse primer. *Lincas2C* (LIC12917) was cloned in pCDF-1b vector (*BamH*I-*Sal*I). Untagged LinCas1C (in pET28a)-His:LinCas2C (pCDF1-b) was co-expressed in *E. coli* BL21 (DE3) cells using IPTG (1 mM). Co-expressed cells were lysed using native lysis buffer (25 mM Tris-Cl pH 8.0, 100 mM NaCl, 1 % triton-X 100, and 10 % glycerol) and centrifuged at 13000 rpm for 20 min. The supernatant was allowed to bind pre-equilibrated Ni-NTA beads for 1 hour at 4˚C. Post-binding Ni-NTA beads were washed using native wash buffer (25 mM Tris-Cl pH 8.0 and 100 mM NaCl). After washing, the protein was eluted using native elution buffer (25 mM Tris-Cl pH 8.0, 100 mM NaCl, 250 mM imidazole, and 10 % glycerol). Presence of LinCas1C and rLinCas2C in LinCas1C-His:LinCas2C (co-expressed) elutes were confirmed by immunoblot using anti-rLinCas1C and anti-rLinCas2C (1:1000), respectively. LinCas1C-His:LinCas6 was used as a negative control (LinCas6; LIC10939 Cas6 of subtype I-B).

Table 3.5. Oligos used in Chapter 3.

DNA oligos	Sequence (5'-3')	Purpose
LinCas1C: F (*Nhe*I) and R (*Xho*I)	F: CGCGGATCCCATGTTTATCATTGTATGTTACGACGT	Cloning of Lin*cas*1C
	R: GCGTCGACTTAAAAATCAAGAATGTTAGAAACTCC	
LinCas2C: F (*BamHI*) and R (*Sal*I)	F: CGCGGATCCCATGTTTATCATTGTATGTTACGACGT	Cloning of Lin*cas*1C
	R: GCGTCGACTTAAAAATCAAGAATGTTAGAAACTCC	
LinCas4C:F (*Nhe*I) and R (*Xho*I)	F: CTAGCTAGCATGTTTAGTAAATCAATAAAACTGT	Cloning of Lin*cas4*C
	R: CCGCTCGAGTTAATATCCGCCAGAGTCTT	
LinCas1C^{785}:F	F: GTTCCAACTCCGGTAACGTT TTG	Lin*cas*1C transcript analysis
LinCas1C^{805}:F	F: GTGGAAAATTTCAAGGACGAC	
LinCas1C^{1002}:F	F: CCGGGATCCATCACTCAAGAAGGTTTATATC	
LinCas2C^{Y7A}:F and R	F: CATTGTATGTgcaGACGTAGAGACGATTAC	LinCas2C ^{Y7A} mutant
	R: ATAAACATTTAAAAATCAAGAATGTTAG	
	F: CATTGTATGTgcagccGTAGAGACGATTACCC	

TH-3470_166106104

LinCas2C$^{Y7A+D8A}$F and R	R: ATAAACATTTAAAAATCAAGAATGTTAG	LinCas2C $^{Y7A+D8A}$ mutant
LinCas2C$^{R33A+Y39A}$F and R	F: atccgttgcaGAATGCCAACTGGAACCAG	LinCas2C $^{R33A+Y39A}$ mutant
	R: ttttgaactgcTTGGCCATGGCTTTCGCA	
LinCas2C$^{\Delta L2}$:F and R	F: AATCTTAGAATCTATTCTCTCG	LinCas2C $^{\Delta L2}$ mutant
	R: TATAATTTTAGAAAGTTTTGCTTC	
1	CTAGGATCCCATGATTAAATTTTTAGTTTCCGAAGA	36-mer oligo
2	TTTTTTTTTTTTTTATTATCTGAGGGTTTAATCTTATTAATCTCTTACTA	50-mer oligo
3	CCCCCCCCCCTGACCCCCCCCCCCCCC	27-mer oligo
4	CCGAACTTTCAATTCTATAAGAG	23-mer oligo

****

Chapter 4

Biochemical characterization of Cas proteins Involved in Maturation and Interference phase in CRISPR-Cas I-C of *L. interrogans* sv. Copenhageni strain Fiocruz L1-130.

This is chapter is partially adapted from the published article and reprinted with author's permission. **Anand, V**., Prabhakaran, H. S., Prakash, A., Hussain, M. S., & Kumar, M. (2023). Differential processing of CRISPR RNA by LinCas5c and LinCas6 of *Leptospira*. Biochimica et Biophysica Acta (BBA)-General Subjects, 1867(12), 130469.

Abstract

L. interrogans, the subtype I-C locus lacks an array component essential for assembling an interference complex. Thus, the reason for sustaining the expense of a cluster of *cas* genes (I-C) is obscure. Type I-C (previously Dvulg) is the only CRISPR subtype that engages Cas5C, a Cas5 variant, to process precursor CRISPR RNA (pre-crRNA). In this study, thus, the rLinCas5C and its mutant variants were cloned, expressed, and purified. The purified rLinCas5C is illustrated as metal-independent cleavage of repeat RNA and pre-crRNA of subtype I-B or orphan CRISPR array. However, the Cas6-bound mature crRNA of subtype I-B fends off from the rLinCas5C activity. In addition, rLinCas5C holds metal and size-dependent DNase activity. The bioinformatics analysis of LinCas5C inferred that it belongs to the subgroup Cas5C-B. Substitution of Phe141 with a more conserved His residue and deletion of unique (β1′-β2′) insertions usher a gain of rLinCas5C activity over nucleic acid. The results uncover the functional diversity of Cas5C ribonucleases and infer an incognito auxiliary role in the absence of a cognate CRISPR array. This study also affirms the functionality of subtype I-C interference machinery, suggesting its potential as an endogenous genetic manipulation tool for future applications.

TH-3470_166106104

4.1 Introduction

In the CRISPR-Cas type-I, most of its subtypes (I-A, I-B, I-D, I-E, and I-F) engage Cas6 to process pre-crRNA (Jackson et al., 2014), while Cas5 (catalytic inactive) operates as a structural subunit in interference complexes (Jackson et al., 2014; Zhao et al., 2014). On the other hand, type I-C (previously Dvulg), which lacks the Cas6, engages Cas5C, a Cas5 variant (catalytic active), to process pre-crRNA (Ebihara et al., 2006; Makarova et al., 2011a). The processing of the pre-crRNA is not restrained to Cas6 and Cas5C; it involves very diverse families of proteins, including Cas12, Cas13, and endogenous RNase III, depending on the type of CRISPR-Cas system (Behler & Hess, 2020; Charpentier et al., 2015b; Deltcheva et al., 2011a; Makarova, Wolf, Iranzo, et al., 2020). The active Cas5 variants in type I-C have been formerly denoted as Cas5d in diverse microorganisms, including *Bacillus halodurans* (BhaCas5C) (Nam, Haitjema, et al., 2012), *Desulfovibrio vulgaris* (DvuCas5C) (Hochstrasser et al., 2016), *Mannheimia succiniciproducens* (MsuCas5C) (Garside et al., 2012), *Streptococcus pyogenes* (SpyCas5C) (Koo et al., 2013), *Streptococcus mutans* (SmuCas5C) (Lemak et al., 2021), *Thermus thermophilus* (TthCas5C) (Garside et al., 2012), and *Xanthomonas oryzae* (XorCas5C) (Koo et al., 2013). Catalytic active Cas5C (BhaCas5C, DvuCas5C, MsuCas5C, SpyCas5C, and XorCas5C) cleave the cognate repeat RNA segment upstream of the 11-nt from the 3′ -ends of pre-crRNA in a sequence-specific canonical mode to render the mature crRNAs with an 11-nt 5′-handle (Garside et al., 2012; Hochstrasser & Doudna, 2015; Hochstrasser et al., 2016; Koo et al., 2013; Nam, Haitjema, et al., 2012).

Various Cas5C proteins (BhaCas5C, DvuCas5C, SpyCas5C, TthCas5C, MsuCas5C, SmuCas5C, and XorCas5C) have two-domain architectures consisting of an N-terminal RRM (RNA recognition motif) domain and a C-terminal β-sheet domain (Brouns et al., 2008; Garside et al., 2012; Hochstrasser et al., 2016; Koo et al., 2013; Lemak et al., 2021; Nam, Haitjema, et al., 2012). Comparatively, the Cas6 of *Sulfolobus solfataricus* (SsoCas6) and *Thermus thermophilus* (TthCas6) commonly have two sequential RRM domains or ferredoxin folds, each at the N- and C-terminus (Ebihara et al., 2006; Reeks, Sokolowski, et al., 2013). The Cas5C active site generally differs from Cas6, and the catalytic center may have evolved independently (Hochstrasser & Doudna, 2015). Moreover, DvuCas5C has been elegantly portrayed to recognize the 5′-handle of the repeat while Cas6 recognizes the 3′ stem-loop (Hochstrasser et al., 2016). Nevertheless, the innate endonuclease activity of Cas6 and Cas5C

TH-3470_166106104

fosters the cleavage of pre-crRNAs through a metal-independent and general acid-base mechanism centered on the catalytic triad (Tyr, His, and Lys) present at the N-terminal RRM domain (Charpentier et al., 2015a; Lemak et al., 2021; Prakash & Kumar, 2021; Punetha et al., 2014). In SmuCas5C/BhaCas5C, the catalytic His121/117 residue (broad base) deprotonates the nucleophile 2′-oxygen of the crRNA ribose, rendering 2′,3′-cyclic phosphate, and 5′-OH cleavage products. While the residue Tyr50/46 (general acid) protonates the leaving group 5′ hydroxyl oxygen (5′-OH), and the Lys120/116 stabilizes the developing negative charge of the transition state (Lemak et al., 2021; Nam, Haitjema, et al., 2012; Yang, 2011a). The Lys residue of the catalytic triad is conserved across the Cas5C homologs; however, Tyr and His residues are not strictly conserved (Hochstrasser & Doudna, 2015; Lemak et al., 2021). Catalytic Tyr residue can be superseded with His, Phe, or Leu, while the His residue can be superseded with Phe (Hochstrasser & Doudna, 2015; Nam, Haitjema, et al., 2012; Reeks, Naismith, et al., 2013).

In the CRISPR interference stage, the CRISPR-associated complex for antiviral defence (Cascade) complex binding to foreign nucleic acid initiates the formation of a localized gap in the target DNA, which is referred to as the R-loop. Within the R-loop, the displaced non-target strand provides a base for the attachment of the Cas3 nuclease, which ultimately cleaves the DNA target (Liu & Doudna, 2020). PAM components play a crucial role in identifying authentic targets within a model studied in *E. coli* (Semenova et al., 2011). Surveillance complexes of Cas proteins exhibit effective scanning capabilities across extended foreign nucleic acids sequences in search of PAM. The Cas8 protein specifically identify and attach to the PAM sequence. The seed sequences in close proximity to these PAM elements are scrutinized for complementarity with the crRNA spacer, prompting complete base pairing and subsequent interference (Semenova et al., 2011; Sternberg et al., 2014).

In this study, we sought to generate the recombinant Cas5C (rLinCas5C) and its mutant variants of the *L. interrogans* serovar (sv.) Copenhageni. We dissected the rLinCas5C RNase activity on repeat and precursor transcript sequences of CRISPRs that belong to *L. interrogans* sv. Copenhageni. The catalytic activity of rLinCas5C was also probed on the DNA substrates of different sizes and conformations. Conservation and variability among Cas5C homologs were also represented through multiple sequence alignment and LinCas5C modeled structure

TH-3470_166106104

analysis. Employing CRISPR-Cas I-C maturation and interference module and predicted PAM interference has been studied via plasmid interference.

4.2 Results

4.2.1 Pure LinCas5C exhibits non-canonical cleavage activity on the repeat RNAs.

The existence of the subtype I-C locus in *L. interrogans* sv. Copenhageni, without a CRISPR array, instigated us to apprehend the role of LinCas5C in CRISPR biology. The gene (*LIC12912*) encoding for LinCas5C was cloned and overexpressed in *E. coli* BL21-AI. The overexpressed recombinant protein (rLinCas5C, 28 kDa) was purified using Ni-NTA affinity chromatography **(Fig. 4.1A)**. The size-exclusion chromatography of rLinCas5C illustrated its existence in a monomeric state (28 kDa) **(Fig. 4.1B).** However, an oligomeric state of pure Cas5C proteins has variably been reported. BhaCas5C exists in the monomeric state of 26.9 kDa (Nam, Haitjema, et al., 2012), while SmuCas5C is in the dimeric state of 61.2 kDa (Lemak et al., 2021). Interestingly, the crystal structure of BhaCas5C is arranged into two protein molecules in an asymmetrical unit (Nam, Haitjema, et al., 2012).

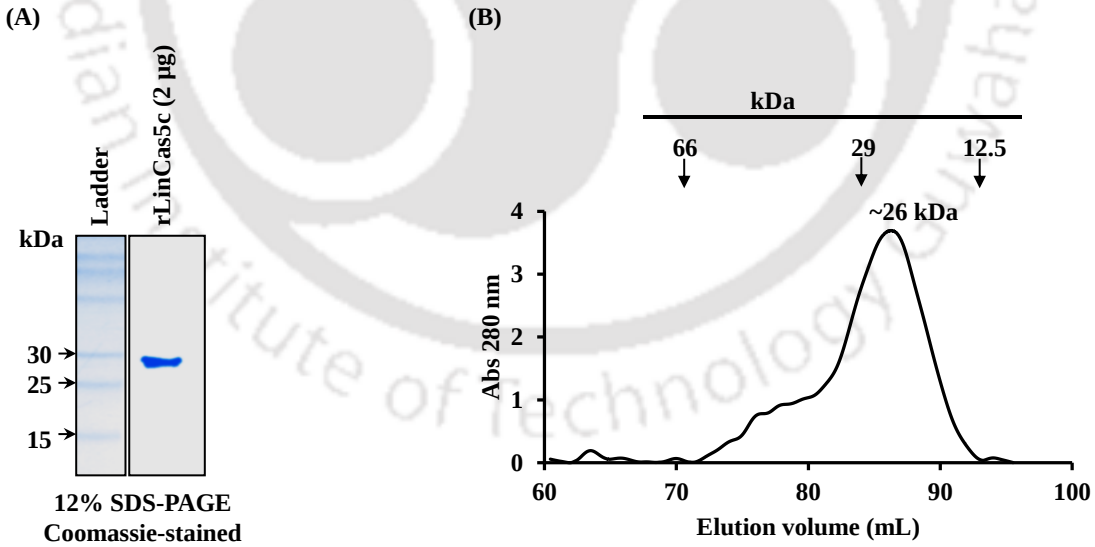


Figure 4.1. Purification and size-exclusion chromatography of rLinCas5C. (A) Ni-NTA affinity chromatography purified rLinCas5C, resolved on 12% SDS-PAGE (PAA), and stained with Coomassie Brilliant Blue. (B) Size-exclusion chromatography of rLinCas5C. The elution profile demonstrates rLinCas5C existence in a monomeric

TH-3470_166106104

state (approx. 26 kDa). Superdex 200 increase column (GE Healthcare, catalogue. no.28-9909-44) on an NGC chromatography system (BioRad).

The molecular facts of CRISPR transcription, orientation, and repeat-spacer boundaries are prerequisites for comprehending its processing by Cas5C. This study sought to investigate if rLinCas5C can be operated to cleave the repeat RNA of two independent CRISPR arrays (LIC_Cr^{2} and LIC_Cr^{3}) of *Leptospira*. The consensus sequence of repeat RNA of array LIC_Cr^{2+} described before (Prakash & Kumar, 2021). Similarly, sense (+) and antisense (-) repeat RNA consensus sequence of array LIC_Cr^{3} was assessed (discussed in Chapter 2). The predicted secondary structures of each repeat RNA consensus sequence (LIC_Cr^{2+}, LIC_Cr^{3+}, and LIC_Cr^{3-}) exhibited stable folding with a distinct hairpin-stem loop **(Fig. 4.2A; left, middle and right panel)**. The rLinCas5C (3 µM) was incubated with a 5′ FAM-labeled sense repeat RNA (36-nt, 250 nM) of array LIC_Cr^{2+} in the absence of Mg^{2+} ions. After an hour of nuclease reaction, the reaction products were resolved by electrophoresis on denaturing polyacrylamide gel. As previously reported (Prakash & Kumar, 2021), rLinCas6 canonically processed its cognate repeat RNA of array LIC_Cr^{2+} generating 28-nt 5′ FAM-labeled RNA fragment. In contrast, rLinCas5C non-canonically processed the LIC_Cr^{2+} repeat RNA substrates, generating a prominent 7-nt long 5′ FAM-labeled RNA fragments and two faint bands of sizes 3- and 2-nt **(Fig. 4.2B).** The cleavage sites between the bases (U2-G3, G3-A4, and A7-U8) of the repeat RNA sequence of array LIC_Cr^{2+} is shown **(Fig. 4.2A; left panel and Fig. 4.2B)**. On the gel, a prominent intensity of the 7-nt fragment suggested one cleavage site (A7-U8) as the most predominant for rLinCas5C on 5′ FAM-labeled sense repeat RNA (36-nt) of LIC_Cr^{2+}.

The catalytic activity of rLinCas5C on the 5′ FAM-labeled sense repeat RNA (28-nt) of array LIC_Cr^{3+} produced one prominent cleavage product of size 6-nt along with five different size fragments (5′ FAM-labeled) of varied band intensity (22<13<2<4-nt and 3-nt) **(Fig. 4.2C)**. In LIC_Cr^{3+} repeat RNA, the cleavage sites between the multiple bases (A6-A7, U2-C3, C3-C4, C4-U5, U13-A14, and U22-A23) were detected **(Fig. 4.2A; middle panel and Fig. 4.2C)**. On the other hand, a single prominent band of 8-nt along with four different size fragments (5′ FAM-labeled) of similar intensity (6, 4, 3, and 2-nt) was detected in the cleavage reaction of 5′ FAM-labeled antisense repeat RNA of array LIC_Cr^{3-} **(Fig. 4.2D)**. In this case, the predominant cleavage site was between A8-A9 bases of the antisense repeat RNA of array LIC_Cr^{3-} **(Fig. 4.2A; right panel and Fig. 4.2D)**. Interestingly, rLinCas6 remained inert

TH-3470_166106104

against the non-cognate repeat RNA of array LIC_Cr^{3+} **(Fig. 4.2C)** and LIC_Cr^{3-} **(Fig. 4.2D)**. Given assays signify that rLinCas5C non-canonically cleaves the repeat RNAs.

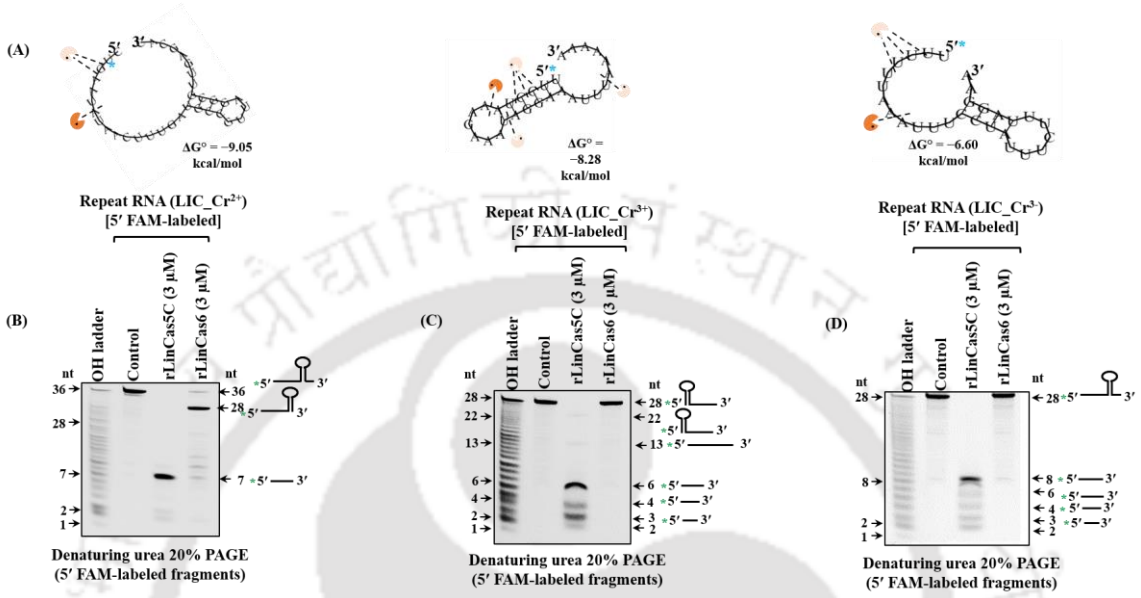


Figure 4.2. RNase activity of rLinCas5C and rLinCas6 on repeat RNAs. (A) The secondary structure of repeat RNAs. The *in silico* predicted repeat RNA folding of array LIC_Cr^{2+}, LIC_Cr^{3+}, and LIC_Cr^{3-} (left, middle, and right panels, respectively) at 37°C. The minimum free energy (ΔG°) of predicted repeat RNA folds and the location of cleavage sites by rLinCas5C are indicated. Most and less preferred cleavage sites are demarcated by dark and faded color (red) emojis, respectively. (B) *In vitro,* cleavage of sense repeat RNA of array LIC_Cr^{2+}. The rLinCas5C or rLinCas6 (3 μM) was incubated with 5' FAM-labeled sense repeat RNA substrate (250 nM). Under similar reaction conditions, an equivalent amount of RNA substrate was incubated without the rLinCas5C as a control. The 5′ FAM-labeled fragments were observed on the gel without staining. Cleavage sites were mapped within the predicted folding of repeat RNA shown above the gel image. The dashed line in Figure 1A shows the cleavage site based on the fragments observed on denaturing urea PAGE. Unprocessed and processed RNA fragments visible on the gel are also indicated by the line diagrams on the right side of the gel image. Green asterisks demarcate 5′ FAM labeling of repeat RNA. (C) *In vitro,* cleavage of sense repeat RNA of array LIC_Cr^{3+}. (D) *In vitro,* cleavage of antisense repeat RNA of array LIC_Cr^{3-}. An alkaline hydrolysis ladder of the repeat RNAs was generated, and the reaction products were resolved by electrophoresis on denaturing urea 20% polyacrylamide gel (PAGE). Superscripts "+" and "−" on LIC_$Cr^{2/3}$ correspond to sense and antisense strands, respectively. All reactions, including controls, were incubated at 37°C for 1 h.

TH-3470_166106104

To rule out the possibility of Mg^{2+} ions contamination during protein purification, we have assessed the rLinCas5C activity on one of the RNA substrates (LIC_Cr^{2+} repeat RNA) in the presence or absence of EDTA **(Fig. 4.3)**. Under both conditions, rLinCas5C demonstrated similar endonuclease activity, suggesting its RNase activity to be Mg^{2+} ion-independent.

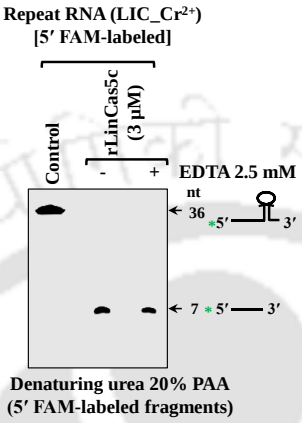


Figure 4.3. The rLinCas5C Mg^{2+} ion-independent RNase activity. The 5' FAM-labeled sense repeat RNA substrate (LIC_Cr^{2+}, 250 nM) was incubated with rLinCas5C (3 µM) in the presence or absence of EDTA for an hour at 37 ˚C. Post completion, the reactions were resolved on denaturing urea polyacrylamide (PAA, 20%) gel, and fluorescently labeled fragments were visualized.

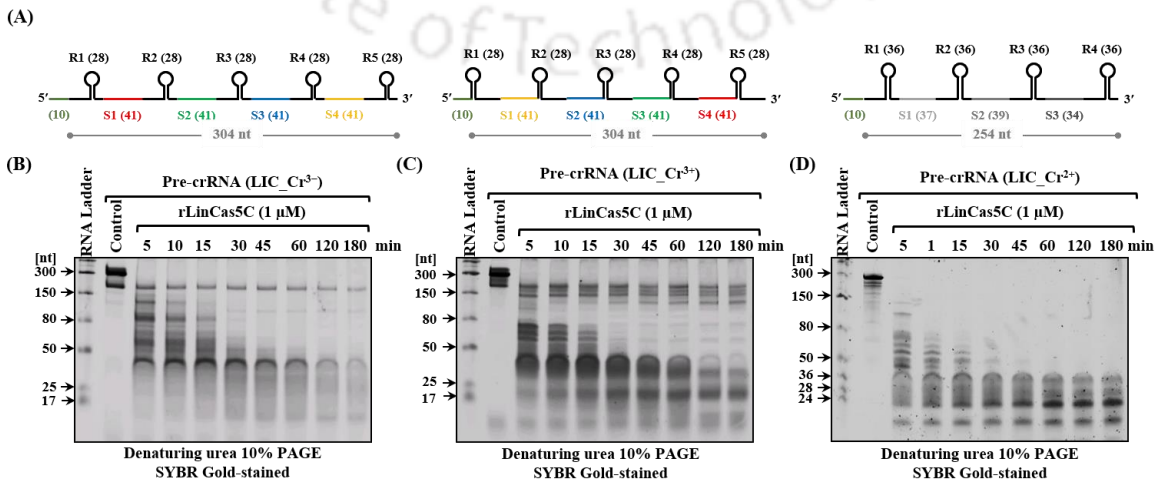
4.2.2 Pure LinCas5C processes the pre-crRNAs.

The rLinCas5C mediated cleavage of 5′ FAM-labeled repeat RNA (28-nt) of array LIC_Cr^{3-} (antisense) displayed a single fragment (8-nt) on the gel. In a recent study, it has been recommended that an adjacent spacer sequence may impact the cleavage of repeat in pre-crRNA of *Methanococcus maripaludis* by Cas6 (MmaCas6) (Richter et al., 2013). Based on the findings of MmaCas6 cleavage efficiencies on different assortments of spacer-repeat-spacer RNAs, we were inquisitive in addressing if spacers in pre-crRNA may impact the processing of repeats to generate mature crRNAs. We thus explored the rLinCas5C efficiency in the processing of precursor crRNA transcripts. Moreover, with the findings of multiple cleavage sites within sense (+) repeat RNA of arrays (LIC_$Cr^{2+/3+}$), it was also interesting to evaluate the activity of rLinCas5C on their full-length pre-crRNA. Thus, the full-length pre-crRNA (unlabeled) was synthesized *in vitro* and taken up as substrates of rLinCas5C in the time-dependent (5-180 min) assay.

TH-3470_166106104

For clarity, the full-length sense or antisense pre-crRNA with its repeats (R) and spacers (S) in arrays LIC_Cr^{3-} (304-nt), LIC_Cr^{3+} (304-nt), and LIC_Cr^{2+} (254-nt) are schematically portrayed **(Fig. 4.4A; left, middle and right panels)**. We conjectured that a single site-specific cleavage of repeat RNA might generate functionally mature crRNAs, each consisting of an entire spacer flanked by partial repeat sequences as formerly documented for LinCas6 (Sokolowski et al., 2014). Hence, rLinCas5C (1 µM) was permitted to act upon antisense pre-crRNA (LIC_Cr^{3-}) for 5-180 min. The reaction products at specific periods were resolved on the urea PAGE. At a 5-15 min cleavage time, two distinguished processed fragments (~80 and 36 nt) were detected on staining with SYBR-Gold. An increase in cleavage time (30-180 min) led to the further ~80 nt fragment processing. The intensity of ~36-nt fragment augmented in 5-30 min of the assay; however, it came to be reduced in 45-180 min cleavage time (**Fig. 4.4B**). In the given assay, thus, pure rLinCas5C, in addition to cleaving between the A8-A9 base of the repeat sequence (LIC_Cr^{3-}), some non-specific cleavage within the spacers (41-nt each) was also detected.

Similarly, pre-crRNA of sense array LIC_Cr^{3+}, when treated with rLinCas5C, yielded a fragment (69-nt) of mature crRNA length in the early 5-15 min of incubation and later at extended reaction time (30-180 min) processed to a short length (< 50-nt) (**Fig. 4.4C)**. Thus, although the cleavage pattern of sense (+) and antisense (-) pre-crRNA of array LIC_Cr^{3} by the rLinCas5C displayed disparity, a prominent band (69-nt) for constituting mature crRNAs could not be detected on denaturing PAGE. Similarly, the processing of pre-crRNA of sense array LIC_Cr^{2+} by rLinCas5C also erred in yielding mature crRNAs (70-75-nt) (**Fig. 4.4D)**. Increasing the reaction time beyond an hour (60-180 min) did not oversee further enhancement in the size (< 30-nt) of cleavage products.



TH-3470_166106104

Figure 4.4. RNase activity of rLinCas5C on the pre-crRNAs. (A) Schematic representation of pre-crRNAs of LIC_Cr^{3-}(left panel), LIC_Cr^{3+} (middle panel), and LIC_Cr^{2+} (right panel). The black lines denote repeats (R) within each pre-crRNA. The orange lines represent an additional 10-nucleotide fragment (vector-derived) co-transcribed at the 5′ end of each pre-crRNA. The unique colored lines represent spacers (S). (B) *In vitro* cleavage of repeat RNA of LIC_Cr^{3-}, The rLinCas5C (1 µM) was incubated (5-180 min) with pre-crRNA of LIC_Cr^{3-} (~100 nM), and then the reactants were resolved on denaturing urea 10% PAA. The gel was imaged after staining in SYBR Gold. (C) *In vitro* cleavage of pre-crRNA of LIC_Cr^{3+}. (D) *In vitro* cleavage of pre-crRNA of LIC_Cr^{2+}. A Low range ss-RNA (300, 150, 80, and 50 nt) ladder (NEB #N0364S) and a mixture of three 5' fluorescent-labeled RNA fragments (36, 28, and 24 nt) or microRNA ladder were used as size markers. All reactions, including controls, were incubated at 37°C.

Interestingly, rLinCas6 remained inactive against non-cognate antisense **(Fig. 4.5A)** and sense pre-crRNA (304-nt) **(Fig. 4.5B)** of array LIC_Cr^{3}. In contrast, a prior study suggests (Prakash & Kumar, 2021) that the mature crRNAs could be generated by the endonuclease activity of rLinCas6 on pre-crRNA of the array LIC_Cr^{2+} **(Fig. 4.5C)**. The cleavage rate for the array by rLinCas5C indicates that the LIC_Cr^{2+} was processed more efficiently than the LIC_Cr^{3}. Altogether, the rLinCas5C (subtype I-C) displayed a remarkable difference in the CRISPR array processing.

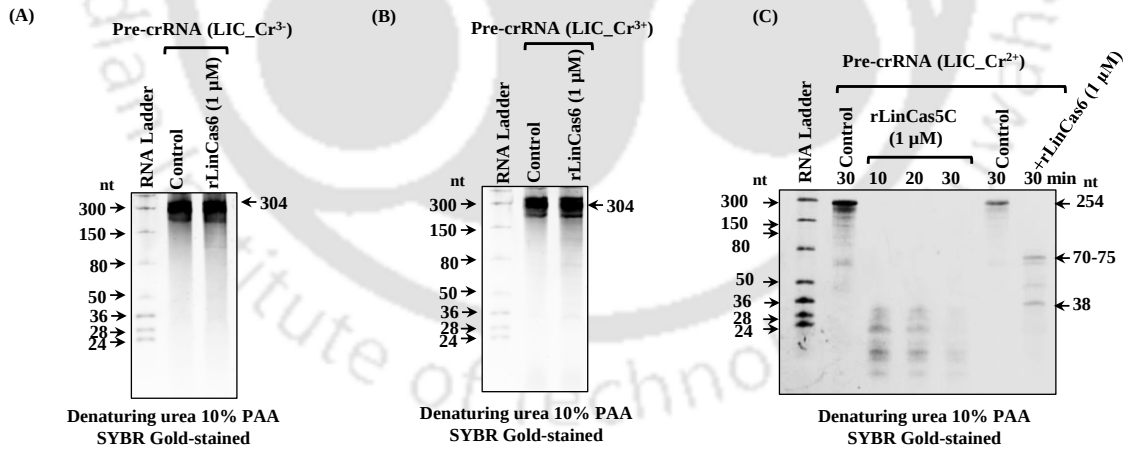
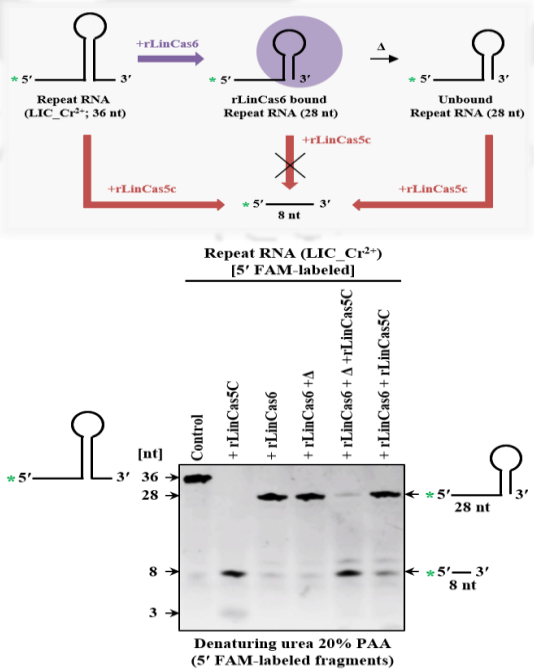


Figure 4.5. Nuclease activity of rLinCas5C and rLinCas6 on the pre-CrRNA. (A) *In vitro* cleavage of repeat RNA of array LIC_Cr^{3-} by rLinCas6. The rLinCas6 (1 µM) was incubated (30 min) with pre-crRNA of array LIC_Cr^{3-} (~100 nM) and was resolved on denaturing urea 10% PAA. The gel was imaged after staining in SYBR Gold. (C) *In vitro* cleavage of pre-crRNA of array LIC_Cr^{3+} by rLinCas6. (C) *In vitro* cleavage of pre-crRNA of array LIC_Cr^{2+} by rLinCas5C and rLinCas6. A Low range ss-RNA (300, 150, 80, and 50 nt) ladder (NEB #N0364S) and a mixture of three 5' fluorescent-labeled RNA fragments (36, 28, and 24 nt) or microRNA ladder were used as size markers. All reactions, including controls, were incubated at 37°C.

TH-3470_166106104

4.2.3 LinCas6 protects mature crRNA of CRISPR I-B from LinCas5C.

We have earlier illustrated that rLinCas6 of subtype I-B remains bound to cleaved repeat RNA after processing the repeat RNA of array LIC_Cr^2 (Prakash & Kumar, 2021). However, the role of rLinCas6 adherence to the mature crRNA during the interference stage of CRISPR-Cas immunity is obscure. Such adherence nudged us to address if bound rLinCas6 can fend its cognate repeat from the rLinCas5C endoribonuclease activity of *Leptospira*. With this intention, the activity of rLinCas5C on rLinCas6 bound repeat RNA fragment (28-nt) of array LIC_Cr^2 was first scrutinized. The repeat RNA fragment (28-nt) pre-bound with rLinCas6 (nucleoprotein complex) was acquired by incubating the repeat RNA (36-nt sense strand) of array LIC_Cr^2 with rLinCas6. In another independent reaction to set free the rLinCas6 from the processed repeat RNA (28-nt), the nucleoprotein complex was heat-denatured (95°C, 5 min). The activity of rLinCas5C was then assessed on repeat RNA fragments (28-nt) developed in the unbound and bound state with rLinCas6. The rLinCas5C cleaved the unbound repeat RNA fragment (28-nt) from 5′ ends into fragments of 7-nt size **(Fig. 4.6)**. The cleavage activity of rLinCas5C on the unbound repeat RNA (28-nt) fragment agreed with that conducted on full-length repeat RNA (36-nt). On the other hand, the activity of rLinCas5C was compromised on a repeat RNA fragment (28-nt) bound to rLinCas6 **(Fig. 4.6).**



TH-3470_166106104

Figure 4.6. The activity of rLinCas5C on rLinCas6 bound repeat RNA. The rLinCas6 (3 µM) was incubated with 5' FAM-labelled sense repeat RNA (250 nM) of LIC_Cr[2] (36-nt) at 37°C for 30 min. One reaction was heat-denatured (Δ) at 95°C for 5 min, and the other was left untreated at room temperature. Then rLinCas5C (3 µM) was added to each reaction and incubated at 37°C for 1 h. Additional RNase reactions without any protein (control) and with rLinCas5C were set against the full-length repeat RNA for 1 h at 37°C. Each sample was then resolved by denaturing urea 20% PAA, and the gel was imaged without staining. MicroRNA ladder or established length of cleavage products was deemed to mark the size of ribonucleotide in the gel. A schematic diagram of RNA fragments on the right side of the gel image is delineated.

The next question addressed was whether, in the presence of rLinCas5C and under *in vitro* conditions, the biogenesis of mature crRNA from pre-crRNA (LIC_Cr^{2+}) can be achieved by rLinCas6 (subtype I-B). On denaturing urea PAGE analysis, the pure rLinCas6 (1 µM) could process the pre-crRNA (264-nt, ~100 nM) of array LIC_Cr^{2+} into mature crRNA (70-75-nt) **(Fig. 4.7A)**. In addition, a 38-nt RNA fragment was detected on denaturing urea PAGE from the cleavage of the pre-crRNA (LIC_Cr^{2+}) within the first repeat, as documented earlier (Prakash & Kumar, 2021). On the other hand, under similar reaction conditions, an equimolar concentration of pure rLinCas5C (1 µM) after processing the pre-crRNA of the array LIC_Cr^{2+} generated multiple (six) RNA fragments **(Fig. 4.7A)**. However, with rLinCas6 and rLinCas5C (1 µM each), a mature crRNA could be yielded from the pre-crRNA (LIC_Cr^{2+}). This result indicates that rLinCas6 has a higher affinity towards its cognate pre-crRNA, and hence, rLinCas6 fends pre-crRNA from undesirable processing by LinCas5C.

Our earlier study documented the binding of rLinCas6 to processed cognate repeat RNA and mature crRNA by an electrophoretic mobility shift assay (Prakash & Kumar, 2021). We thus conjectured that the rLinCas6 binding to the mature crRNA may also fend from the catalytic activity of rLinCas5C, as observed for rLinCas6 bound repeat RNA **(Fig. 4.6)**. The rLinCas6 (1 µM) was permitted to bind to the 5′ FAM-labeled mature crRNA (70-nt, 250 nM) of the LIC_Cr^{2+} array for 30 min at 37°C for further validation. Next, half of the reaction was allowed to undergo heat treatment (95°C, 5 min) to denature the rLinCas6 and abrogate its binding affinity to mature crRNA. After that, the rLinCas5C (1 µM) was incubated with the heat-treated or untreated reaction for 30 min at 37°C, and the reaction product was resolved on denaturing urea PAGE. Analysis of reaction products detected the activity of rLinCas5C to be compromised on mature crRNA bound with rLinCas6 **(Fig. 4.7B)**. However, rLinCas5C was adept in cleaving the mature crRNA that had been dissociated from rLinCas6 denaturation.

TH-3470_166106104

This observation led us to propose that LinCas6 in the subtype I-B may fend the crRNA from co-existing endoribonucleases like LinCas5C in the *Leptospira*.

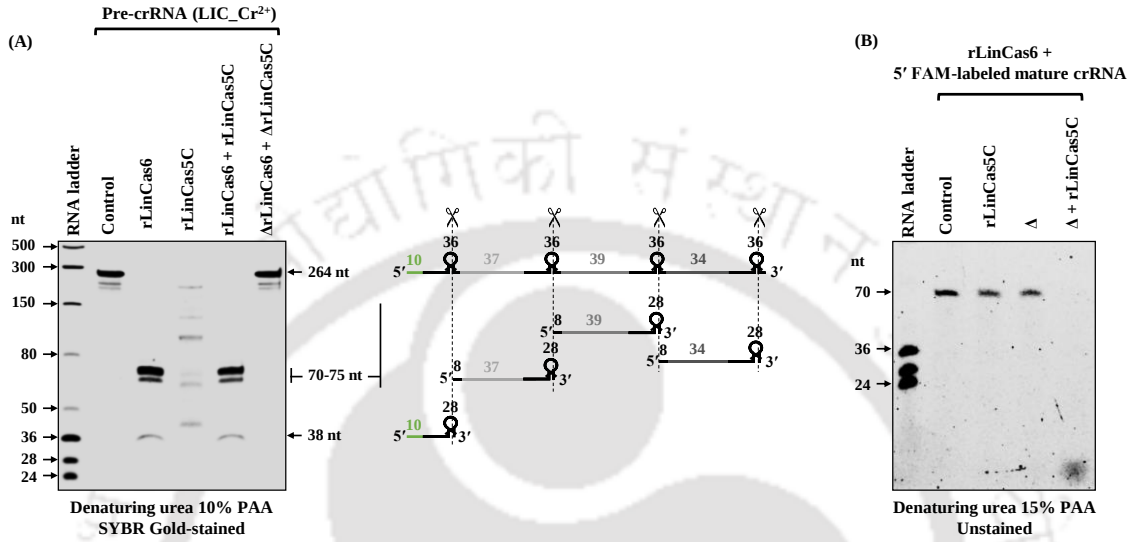


Figure 4.7. The rLinCas6-mediated generation of mature crRNA and its protection from rLinCas5C. (A) The activity of rLinCas6 and rLinCas5C on the unlabeled pre-crRNA of LIC_Cr^{2+}. Pre-crRNA (100 nM) was incubated (30 min at 37°C) with rLinCas6 (1 µM), rLinCas5C (1 µM), or with both rLinCas6 and rLinCas5C (1 µM each). An equivalent amount of pre-crRNA was incubated without protein (control) under similar reaction conditions. Heat-denatured (95°C, 5 min) forms of rLinCas6 and rLinCas5C (1 µM each) as another control of the assay were also incubated with an equivalent amount of pre-crRNA. Reactions, including control, were resolved on a denaturing 10% polyacrylamide gel and visualized after staining with SYBR Gold (left panel). Processed RNA fragments from pre-crRNA by rLinCas6 were represented graphically (right panel) (B) The activity of rLinCas5C on free or rLinCas6 bound mature crRNA. The rLinCas6 (1 µM) was incubated with 5′ fluorescents labeled mature crRNA for 30 min at 37°C to make the rLinCas6-crRNA complex. The rLinCas5C (1 µM) was incubated with an unheated or heated (95°C, 5 min) rLinCas6-crRNA nucleoprotein complex for 30 min at 37°C. Reactions were resolved by denaturing 15% PAA and visualized without staining. The red arrowhead indicates the 5′ fluorescent labeled RNA fragment generated from the freed mature crRNA in the presence of rLinCas5C.

4.2.4 Structural modeling and MSA of LinCas5C.

RNA recognition motif (RRM) is a typical feature among various endoribonucleases, including Cas6 and Cas5C family proteins (Koo et al., 2013; Nam, Haitjema, et al., 2012;

TH-3470_166106104

Samai et al., 2010). Cas5C proteins contain a single N-terminal RRM domain and a C-terminal β-sheet domain consisting of three antiparallel β-strands, whereas Cas6 proteins have two sequential RRM domains in tandem at N- and C-terminus (Charpentier et al., 2015a; Koo et al., 2013; Lemak et al., 2021). Based on sequence analysis, the Cas5C proteins are categorized into two subgroups: Cas5C-A and Cas5C-B (Nam, Haitjema, et al., 2012). Cas5C-A includes Cas5C proteins (BhaCas5C, SpyCas5C, and SmuCas5C) with an extended C-terminus (~30 residues). The members of the subgroup Cas5C-B (XorCas5C, DvuCas5C, and MsuCas5C) contain an α-helical insertion (~23 residues) between two short β-strands (e.g., β4 - β5 in XorCas5C) of the N-terminal region (Koo et al., 2013). In addition, evolutionarily, Cas5C proteins are clubbed into three (I, II, and III) major monophyletic groups (Lemak et al., 2021). Group I encompasses the subgroup Cas5C-B proteins having an insertion in the N-terminal RRM domain, whereas groups II and III include the subgroup Cas5C-A proteins encompassing an extended C-terminus (Lemak et al., 2021).

The modeled tertiary structure of LinCas5C consists of an N-terminal RRM-like domain and a modified C-terminal β-sheet domain **(Fig. 4.8A).** LinCas5C modeled structure displays similarity to the previously reported Cas5C structures, including XorCas5C (PDB ID: 3VZI) (Koo et al., 2013), MsuCas5C (PDB ID: 3KG4) (Garside et al., 2012), BhaCas5C (PDB ID: 4F3M) (Nam, Haitjema, et al., 2012), and SmuCas5C (PDB ID: 4R0J) (Lemak et al., 2021). Structural homology search using Dali web server (Holm, 2022) revealed XorCas5C (PDB ID: 3VZI chain A) as the closest topological homolog of modeled LinCas5C based on RMSD and Z-score parameters (RMSD: 1.1 Å, Z- score: 30). The XorCas5C RRM-like domain encloses a helical insertion (α1′- α2′) between β4 and β5 strands (previously specified between β3 and β4) (Koo et al., 2013). The modeled LinCas5C in the RRM-like domain also encloses a helical insertion (α1′- α2′- α3′) between β4 and β5 strands **(Fig. 4.8A)**. For clarity, the structural superimposition of XorCas5C and LinCas5C is shown **(Fig. 4.8B)**. In addition, the LinCas5C at the C-terminal unprecedently contains an antiparallel β-sheet (β1′- β2′) insertion between the β7 and β8 strands **(Fig. 4.8A and 4.8B)**. MSA analysis of LinCas5C with its orthologs revealed that it lacks the extra C-terminal residues **(Fig. 4.8C)**, like the Cas5C-B subgroup, including XorCas5C (Koo et al., 2013; Lemak et al., 2021). Overall, the modeled LinCas5C structure displays an excellent structural analogy with XorCas5C; hence, it can be proclaimed a member of the Cas5C-B subgroup.

TH-3470_166106104

In a recent review, Cas5C proteins are known to share substantial sequence conservation of the catalytic triad (Tyr, Lys, and His) (Lemak et al., 2021; Reeks, Naismith, et al., 2013) and are independent of their subgroup classification (Lemak et al., 2021; Reeks, Naismith, et al., 2013). The catalytic triad residues of BhaCas5C (Y46, K116, H117) and SmuCas5C (Y50, K120, H121) have been comprehensively studied for their endoribonuclease activity (Lemak et al., 2021; Nam, Haitjema, et al., 2012). In the catalytic triad of Cas5C (Y, K, H), the Tyr residue, is less critical for ribonuclease activity. Tyr residue is naturally substituted with His (MsuCas5C and XorCas5C) or Leu (LinCas5C) among Cas5C orthologs (Garside et al., 2012; Koo et al., 2013; Reeks, Naismith, et al., 2013). Similarly, in SmuCas5C, substituting Tyr50 with Ala50 retained its RNase activity (Lemak et al., 2021).

However, in the catalytic triad, the other two residues (Lys and His) of BhaCas5C (K116 and H117) and SmuCas5C (K120 and H121) are critical (Lemak et al., 2021; Nam, Haitjema, et al., 2012). In SmuCas5C and BhaCas5C, the single residue substitution mutation of either of the two conserved residues (Lys and His) with Ala resulted in the abolishment of ribonuclease activity. Similarly, in XorCas5C catalytic triad (H44, K134, and H135), the mutation of His135 with Ala135 drastically reduced the RNase activity (Koo et al., 2013). In this study, the LinCas5C catalytic triad (L51, K140, and F141) was predicted based on multiple sequence alignment (MSA) with other available Cas5C orthologs **(Fig. 4.8C).** MSA in this study is also in line with elsewhere reported elegant review (Reeks, Naismith, et al., 2013), it is observed (Reeks, Naismith, et al., 2013) that in Cas5C (GenBank ID:EKQ47929) of another strain of *L. interrogans* (*L. interrogans* strain 2002000623), the conventional Tyr and His residues of the catalytic triad is replaced with the Leu46 and Phe136, respectively. However, the Leu and Phe residues are non-dissociable hydrophobic amino acids whose catalytic properties are questionable in LinCas5C (Ribeiro et al., 2020).

TH-3470_166106104

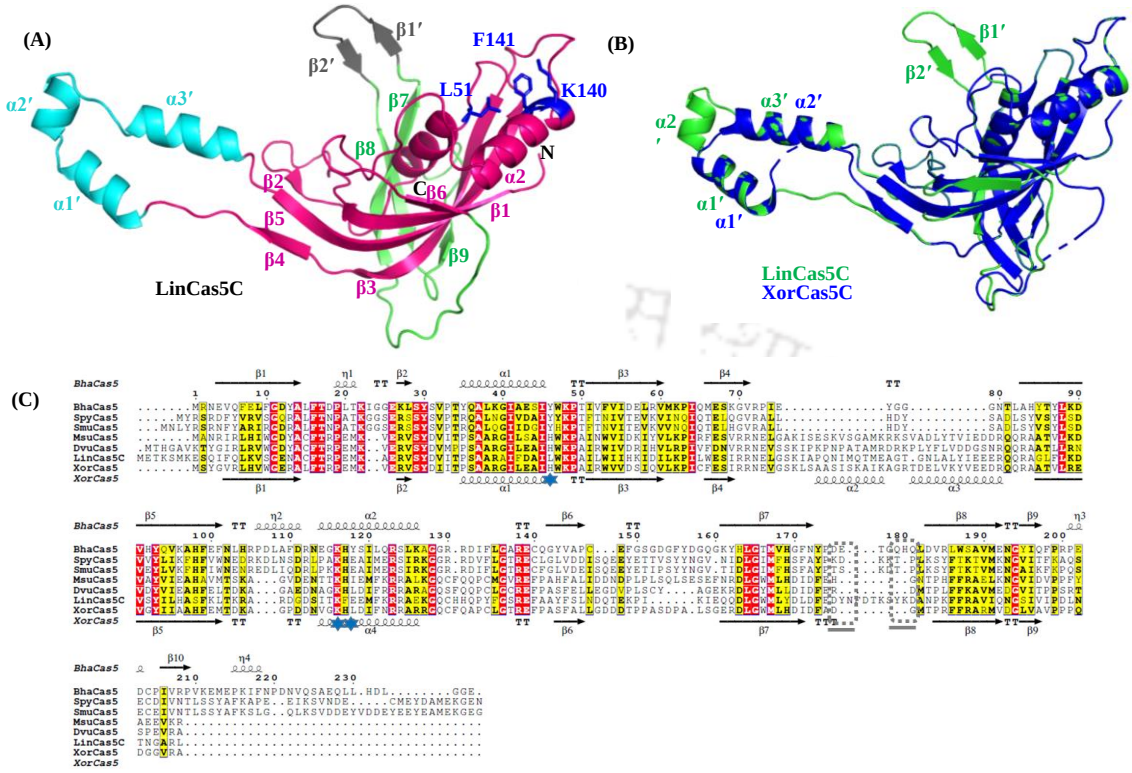


Figure 4.8. Structure prediction and multiple sequence alignment of LinCas5C. (A) The modeled tertiary structure of LinCas5C. The predicted 3D model of LinCas5C is shown as a ribbon diagram with secondary structure elements labeled. The N-terminal RRM-like and C-terminal β-sheet domains are indicated in magenta and green, respectively. The α-helical insertion (α1'- α2'- α3') in the N-terminal RRM-like domain- a characteristic of the Cas5C-B subgroup is colored cyan. The unique insert of two-stranded antiparallel β-sheet (β1'- β2') between β7 and β8 in the C-terminal β-sheet domain is shown in grey. The *in silico* predicted catalytic triad residues (L51, K140, and F141) are labeled as blue sticks. (B) The structural superimposition of LinCas5C (green) onto XorCas5C (Cas5C of *X. oryzae*; PBD ID: 3VZI; blue). (C) Multiple sequence alignment (MSA) of LinCas5C with its orthologs. MSA was performed using the Clustal Omega server with default parameters and visualized by ESPript v3. The Cas5C orthologs from various organisms used for MSA include BhaCas5C (*B. halodurans;* Q9KFY3*),* SpyCas5C (*S. pyogenes*; UniProt ID: Q99YS3), SmuCas5C (*Streptococcus mutans*; Q8DSL7), MsuCas5C (*M. succiniciproducens*; Q65TW5), DvuCas5C (*Desulfovibrio vulgaris*; Q72WF9), and XorCas5C (*X. oryzae*; A0A0J9X178) and LinCas5C (*L. interrogans*; Q72NC2). The corresponding subgroup (Cas5C-A or Cas5CB) under which the different Cas5C orthologs falls has been indicated at the left of the alignment. The secondary structure elements are inferred from BhaCas5C (PDB ID: 4F3M; top) and XorCas5C (PDB ID: 3VZI; bottom) structures. The LinCas5C's secondary structure elements based on its predicted 3D model are also depicted at the bottom of the alignment and color-coded as in (A). The residue number indicated on top is as per BhaCas5C. The conserved residues are in

TH-3470_166106104

white font highlighted with red, whereas semi-identical residues are in red font within unfilled blue boxes. The predicted catalytic triad residues L51 (light blue star), K140 (black oval), and F141 (yellow triangle) in LinCas5C correspond to the conserved catalytic triad (Y46, K116, and H117) of BhaCas5C. The characteristic insertion in the N-terminal RRM-like domain and absence of extended sequence at the C-terminal in subgroup Cas5C-B proteins (MsuCas5C, DvuCas5C, LinCas5C, and XorCas5C) are shown in green and orange boxes, respectively.

The predicted catalytic triad of LinCas5C (L51, K140 and F141), when mapped with its closest homologs BhaCas5 (Y46, K116 and H117) and XorCas5 (H44, K134 and H135) demonstrated conformational alignment in the triad **(Fig. 4.9).** In LinCas5C, the existence of Phe141 residue as one of the predicted catalytic triads (L51, K140, and F141) and its structural correlation with H117 of BhaCas5C and H135 of XorCas5C boosted us to evaluate the impact of the presence of Phe141 on LinCas5C activity. Therefore, a mutant variant of LinCas5C (rLinCas5C^{F141H}) was generated wherein the Phe141 was substituted with His residue. In addition, a deletion mutant (rLinCas5C$^{\Delta\beta}$) for β-sheet insertion (β1′- β2′) was created to uncover its role in the LinCas5C activity.

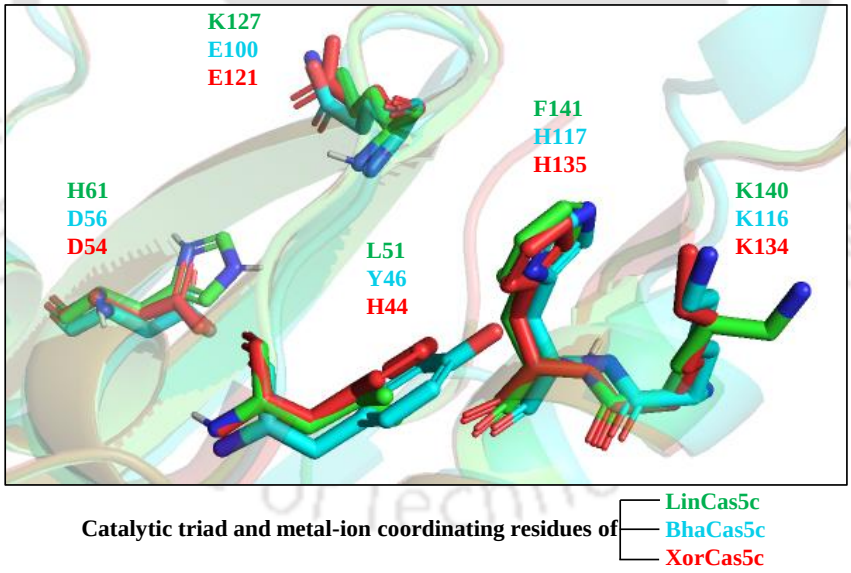


Figure 4.9. Illustration of predicted catalytic triad and metal-ion coordinating residues of LinCas5C. Modeled LinCas5C was superimposed with the available structure of XorCas5C and BhaCas5C. The mapped catalytic triad (L51, K140, and F141) together with metal-ion coordinating residues (H61 and K127) of LinCas5C is demonstrated and compared with XorCas5C and BhaCas5C.

4.2.5 LinCas5C mutant variants exhibit gain of function in RNA cleavage.

The mutant variants of LinCas5C (rLinCas5C$^{\Delta\beta}$ and rLinCas5C^{F141H}) were purified as described for rLinCas5C. A quality check of the purified rLinCas5C mutant variants (2 µg each) was accomplished on SDS-polyacrylamide gel before using it for downstream assays (**Fig. 4.10A**). To measure the effect of mutation on RNase activity, purified proteins (rLinCas5C, rLinCas5C$^{\Delta\beta}$, and rLinCas5C^{F141H}) at an equimolar concentration (3 µM) were individually incubated with the 5′ FAM-labeled sense repeat RNA (LIC_Cr^{2+}, 36-nt). The resulting products were resolved by electrophoresis on a urea polyacrylamide gel. Consistent with the earlier experiment, rLinCas5C cleavage activity yielded a major intense band (7-nt) and a faint band of 3-nt size **(Fig. 4.10B)**. However, under similar reaction conditions, the rLinCas5C mutant variants (rLinCas5C$^{\Delta\beta}$ and rLinCas5C^{F141H}) at equimolar concentration yielded only 3-nt product from the 5′ FAM-labeled sense repeat RNA (LIC_Cr^{2+}, 36-nt). Similarly, the rLinCas5C mutant variants displayed a gain of activity on the 5′ FAM-labeled sense and antisense repeat RNA (28-nt) of array LIC_Cr^{3} **(Fig. 4.10C and Fig. 4.10D)**. Interestingly, the mutant variant rLinCas5C^{F141H} displayed stronger activity than rLinCas5C$^{\Delta\beta}$ on the 5′ FAM-labeled repeat RNA of arrays LIC_Cr^{2+} and LIC_Cr^{3-} **(Fig. 4.10B and Fig. 4.10D)**.

The ribonuclease activity of rLinCas5C variants was also quantitively analyzed on a commercially available model fluorogenic RNA substrate (Integrated DNA technology, IDT Cat #11-02-01-02). The rate of ribonuclease activity of mutant variants (rLinCas5C$^{\Delta\beta}$ and rLinCas5C^{F141H}) showed a gain of function than the equivalent concentration (0.5 µM) of rLinCas5C **(Fig. 4.10E)**. It is known that the residue His, in general, is a stronger base than Phe (Kessler & Raja, 2019) and thus, the substitution of Phe141 with a more conserved His141 residue might have accounted for the gain in the activity of rLinCas5C^{F141H}. However, for LinCas5C$^{\Delta\beta}$, at this stage, it is unclear about the mechanism of gain of nuclease activity. The tertiary structure of LinCas5C$^{\Delta\beta}$ (modeled structure) showed more conservation (RMSD: 0.8 Å and Z-score: 29.2) with XorCas5C than LinCas5C (1.1 Å, and 30). Evidently, one may hypothesize alterations in the conformation of rLinCas5C$^{\Delta\beta}$ tertiary structure as one of the possible causes for the gain of nuclease activity. Perhaps in the LinCas5C, the insertion of β1′ and β2′ by a unique evolutionary adaptation may offer an added advantage of regulated and

TH-3470_166106104

controlled nuclease activity. The crystal structure of LinCas5C in the near future may provide better insight into the role of additional β-sheet insertion in the regulation of nuclease activity.

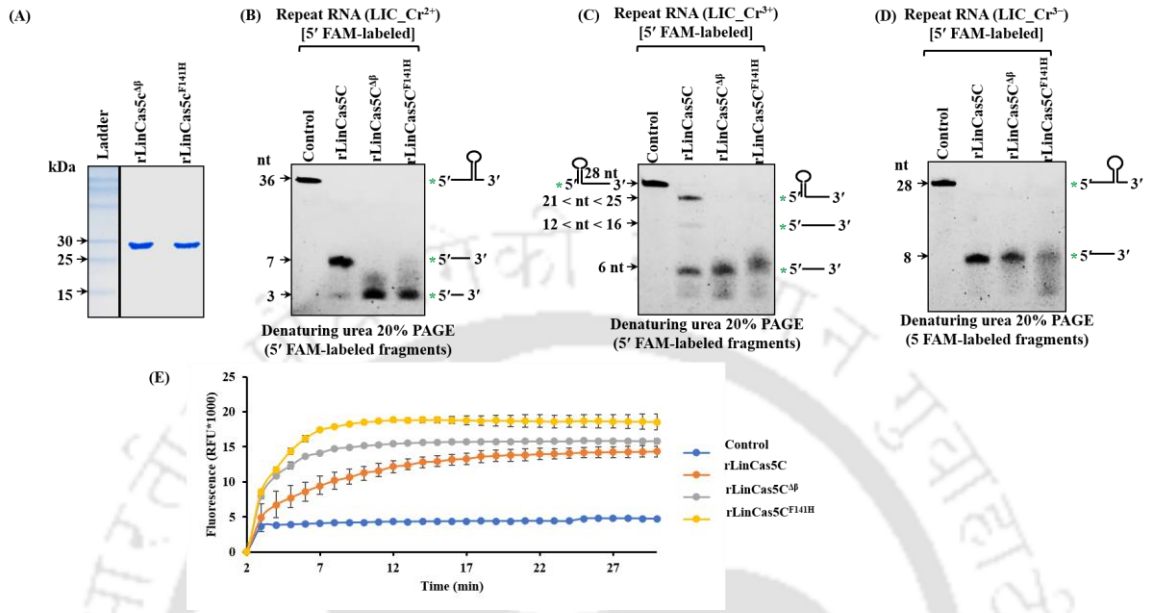


Figure 4.10. Comparison of RNase activity between rLinCas5C and its mutant variants. (A) Ni-NTA affinity chromatography-purified recombinant LinCas5C mutant's variants (rLinCas5C$^{\Delta\beta}$, and rLinCas5C^{F141H}) are resolved on 12% SDS-PAGE (PAA) and stained with Coomassie Blue. RNase activity of rLinCas5C and its mutant variants (rLinCas5C$^{\Delta\beta}$ and rLinCas5C^{F141H}) on 5' FAM-labeled sense repeat RNA of (B) LIC_Cr^{2+}, (C) LIC_Cr^{3+} and (D) LIC_Cr^{3-}. All reactions, including controls, were incubated at 37°C for 1 h and resolved on denaturing urea PAGE (20% PAA). (D) RNase activity of rLinCas5C and its variants was quantified using fluorogenic RNA substrate. A fluorogenic RNA substrate (10 pmol) was incubated with rLinCas5C or its variants (0.5 μM) at 37°C, and the fluorescence was recorded at 2 min intervals for one hour. The data shown here represents Mean ± Standard Error Mean (SEM) from three different experiments performed in duplicates.

4.2.6 LinCas5C activity on DNA substrates.

Recombinant LinCas5C specific cleavage on RNA substrate instigated us to assess its activity and binding on 5′ FAM-labeled antisense repeat DNA (36-nt) of array LIC_Cr^{2}. The reaction products of 5′ FAM-labeled antisense repeat DNA substrate were resolved on denaturing urea PAGE or native PAGE for assessing cleavage activity and binding activity of rLinCas5C, respectively. The rLinCas5C displayed neither cleavage (**Fig. 4.11A)** nor any binding affinity on 5′ FAM-labeled antisense repeat DNA (**Fig. 4.11B)** in the presence or

TH-3470_166106104

absence of Mg^{2+}. Similarly, rLinCas5C was inert on duplex DNA oligo (ds-repeat DNA, 36 bp) generated by annealing sense and antisense repeat DNA of LIC_Cr^2 **(Fig. 4.11C)**.

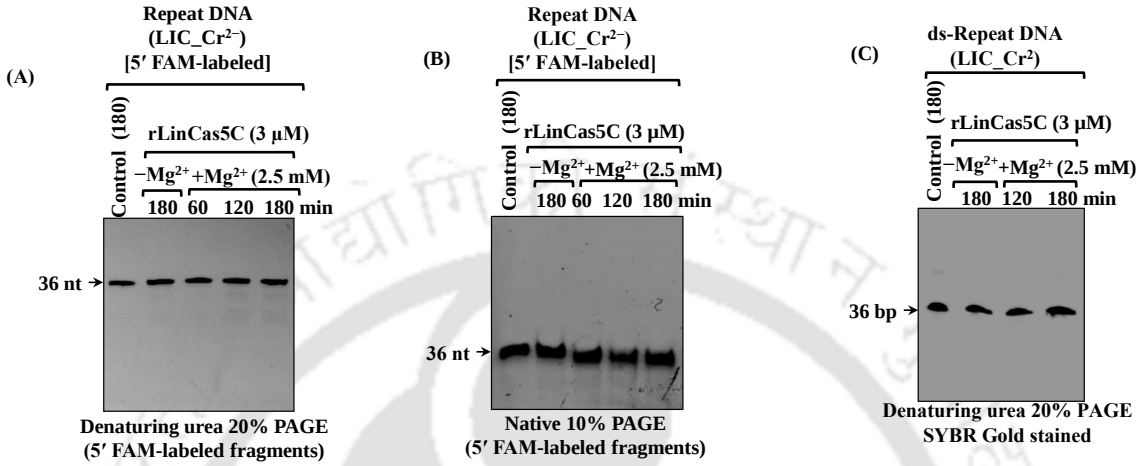


Figure 4.11. The nuclease activity of rLinCas5C on small DNA substrates. (A & B) The activity of rLinCas5C on antisense repeat DNA of array LIC_Cr^{2-}. The 5' FAM-labeled antisense repeat DNA (0.5 µM) was incubated with rLinCas5C (3 µM) at time intervals of 60, 120, and 180 min in the presence or absence of Mg^{2+} ions as indicated. Reactions were resolved on 8 M urea 20 % PAGE (A) and 10% native-PAGE (B) followed by visualization at 490 nm. (C) The activity of rLinCas5C (3 µM) on duplex repeat DNA (LIC_Cr^2; 36 bp, 0.5 µM) and reaction products were resolved on 8 M urea 20 % PAGE and visualized after SYBR Gold staining.

The rLinCas5C in the presence of Mg^{2+} ions displayed the DNase activity on larger size circular ds-DNA (pTZ57R/T plasmid; 2886 bp) and circular ss-DNA (Φx174 DNA; 5386 bp) (**Fig. 4.12A and 4.12B**). Similarly, the mutant variants (rLinCas5C^{F141H} and rLinCas5C$^{\Delta\beta}$) also displayed metal-dependent endonuclease activity on the DNA substrates **(Fig. 4.12C, 4.12D, 4.12E, and 4.12F)**. The endonuclease activity of rLinCas5C^{F141H} and rLinCas5C$^{\Delta\beta}$ on the DNA substrates consistently demonstrated a gain of function similar to the activity on repeat RNA substrate. Using a well-studied Cas5C (BhaCas5C) tertiary structure as a template, the metal-ion coordinating residues were predicted for modeled LinCas5C and XorCas5C (PDB ID: 3VZI chain A) structure. The LinCas5C metal-ion coordinating residues (H61 and K127) plausibly coordinate with Mg^{2+} ions through a non-canonical site, unlike BhaCas5C (D56 and E100) and XorCas5C (D54 and E121) during the catalysis of DNA **(Fig. 4.9)**. Also, the proximity between the predicted metal-ion coordinating residues and the catalytic triad in the LinCas5C modeled and XorCas5C (PDB ID: 3VZI chain A) structure were compared to that

TH-3470_166106104

of BhaCas5C **(Fig. 4.9)**. The comparative result suggested the existence of predicted metal-ion coordinating residues and the catalytic triad of the LinCas5C model and XorCas5C in close proximity to that of BhaCas5C.

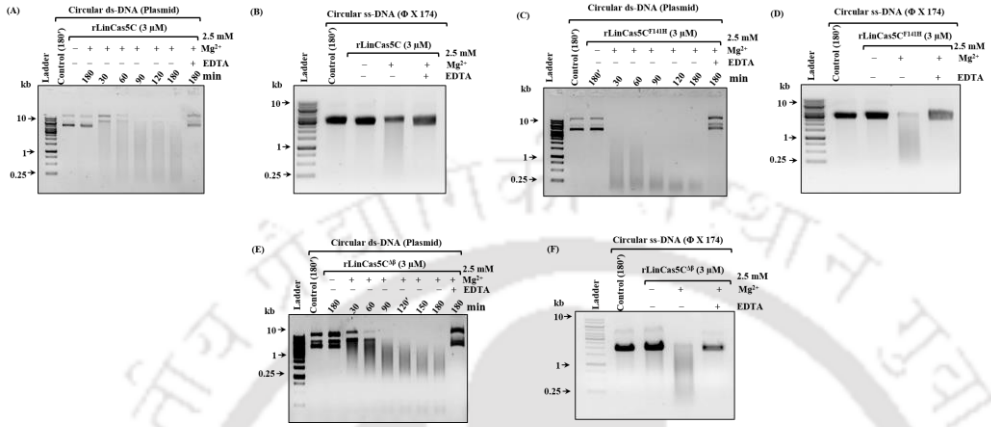


Figure 4.12. The nuclease activity of rLinCas5C on large DNA substrates. All the assays were performed using 3 μM enzymes at 37°C for 60 min or as indicated, and the reactions were resolved on ethidium bromide stained-1% agarose gel. The activity of rLinCas5C on (A) circular ds-DNA (pTZ57R/T; 2886 bp, 0.5 μg) and (B) circular ss-DNA (Φx174 virion DNA; 5386 bp, 0.5 μg). The activity of rLinCas5C^{F141H} on (C) circular ds-DNA (pTZ57R/T; 2886 bp) and (D) circular ss-DNA (Φx174 virion DNA; 5386 bp, 0.5 μg). The DNase activity of rLinCas5C$^{\Delta\beta}$ on (E) circular ds-DNA (pTZ57R/T; 2886 bp) and (F) circular ss-DNA (Φx174 virion DNA; 5386 bp, 0.5 μg).

5.2.7 Interference by CRISPR-Cas I-C through plasmid interference assay.

Above finding suggests rLinCas5C standalone non-canonically cleaves repeat RNA and is unable to generate mature CrRNA from pre-CrRNA of array LIC_Cr^2 and LIC_Cr^3. We hypothesized that LinCas5C, in alliance with other subtype-specific Cas proteins or host factors, may act as a specific endoribonuclease to set off crRNAs or Cascade I-C in the *Leptospira* CRISPR immunity. To test this hypothesis plasmid interference assay was conducted employing the maturation and interference module of CRISPR-Cas I-C. Protospacers were constructed using spacer 1 from array LIC_Cr^2, incorporating the predicted PAM sequences from both LIC_Cr^2 and LIC_Cr^3 at their 5′ ends (Table 4.1).

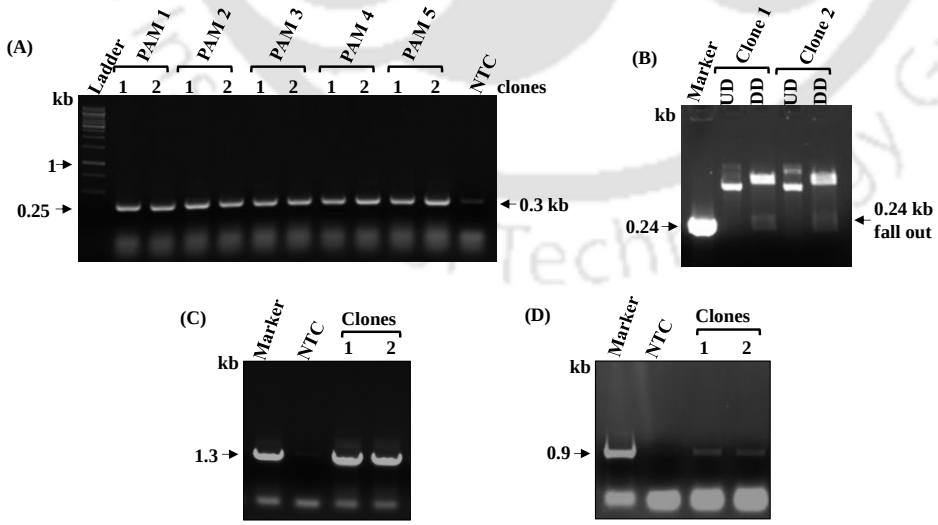
Table 4.1: Protospacers used to study plasmid interference.

	PAM 1	***5′ AGCTTATA***AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTTC***A 3′*** ***5′ TATG***AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT***TATA 3′***

TH-3470_166106104

LIC_CR^2	PAM2	***5' AGCTTTGA***AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTTC***A 3'*** ***5' TATG***AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT***TCAA 3'***
	PAM3	***5' AGCTTATG***AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTTC***A 3'*** ***5' TATG***AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT***CATA 3'***
	PCR2SP1	**5' *AGCTT*** AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTT***CA* 3'** **3' *A***TTTCCTAGGAAACTAGTTTTCTTAAGCAGGAACTAAA***GTAT* 5'**
LIC_CR3	PAM 1	***5' AGCTTTTA***AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTT***CA 3'*** ***5' TATG***AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT***TAAA 3'***
	PAM 2	***5' AGCTTTTT***AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTT***CA 3'*** ***5' TATG***AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT***AAAA 3'***
	PAM 3	***5' AGCTTGAA***AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTT***CA 3'*** ***5' TATG***AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT***TTCA 3'***
	PAM 4	***5' AGCTTTCT***AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTT***CA 3'*** ***5' TATG***AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT***AGAA 3'***
	PAM 5	***5' AGCTTGCT***AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTT***CA 3'*** ***5' TATG***AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT***AGCA 3'***

ATG and TAC were previously reported PAM for the CRISPR-Cas I-B locus of the genus *Leptospira* (Mendoza & Trinh, 2018; Xiao et al., 2019), therefore ATG PAM was used as a positive control and wild-type pT7Blue was served as a negative control in this study. The protospacers containing PAM predicted using spacers of array LIC_Cr^2 were previously cloned in pT7Blue vector (Hussain et al., 2023) and in this study PAM predicted using spacers of array LIC_Cr^3 were cloned in pT7Blue vector. These constructs were named as pPAM **(Fig. 4.13A)**. The array LIC_Cr^2 was cloned in pCDF-1b vector and the construct was named as pCr-array **(Fig. 4.13B)**. The interference module of CRISPR-Cas I-C was subsequently cloned into the pACAYDuet vector, utilizing two multiple cloning sites (MCS). In one MCS, *Lincas5-7C* was inserted, while in another MCS, *Lincas3'-3C* was integrated **(Fig. 4.13C and 4.13D).**



TH-3470_166106104

Figure 4.13. **Generation of tools to study plasmid interference by CRISPR-Cas I-C.** (A) Cloning of predicted PAM using spacers of LIC_Cr^3 in pT7blue vector (*Hind*III and *Nde*I). Clone confirmation using PCR employing recombinant pT7blue vector as a template. (B) Cloning of array LIC_Cr^2 in pCDF-1b vector (*BamH*I-*Sal*I). Clone confirmation using restriction digestion of recombinant plasmid. Cloning of interference module of CRISPR-Cas I-C in pACYCDuet vector. Clone confirmation using PCR employing recombinant pACYCDuet vector as a template with the aid of *Lincas7C* (C) and *Lincas3C* (D) primers. NTC: no template control.

To investigate interference mediated by the maturation and interference module of CRISPR-Cas I-C, *E. coli* BL21 (DE3) cells were co-transformed with three plasmids: p*Lincas3-7C*, pCr^2_array, and pEmpty or pCr2SP1 or pPAM. The transformed *E. coli* BL21 (DE3) cells were cultured until they reached the logarithmic growth phase. An equal number of cells from the test and control groups were then plated on LB agar supplemented with IPTG (0.05 mM) and antibiotics (ampicillin, chloramphenicol, and spectinomycin, each at a concentration of 100 µg/ml). The degradation of the pPAM plasmid by the maturation and interference module of CRISPR-Cas I-C led to the loss of ampicillin resistance, preventing the cells from growing on the LB agar plate supplemented with IPTG (0.05 mM) and the specified antibiotics. Consequently, the reduction in the number of bacterial colonies on the test plate was a result of the destruction of the pPAM plasmids (carrying the AmpR gene) by the CRISPR-Cas I-C interference module, indicating the presence of a CRISPR interference phenotype. For clarity schematically depicted **(Fig. 4.14).**

TH-3470_166106104

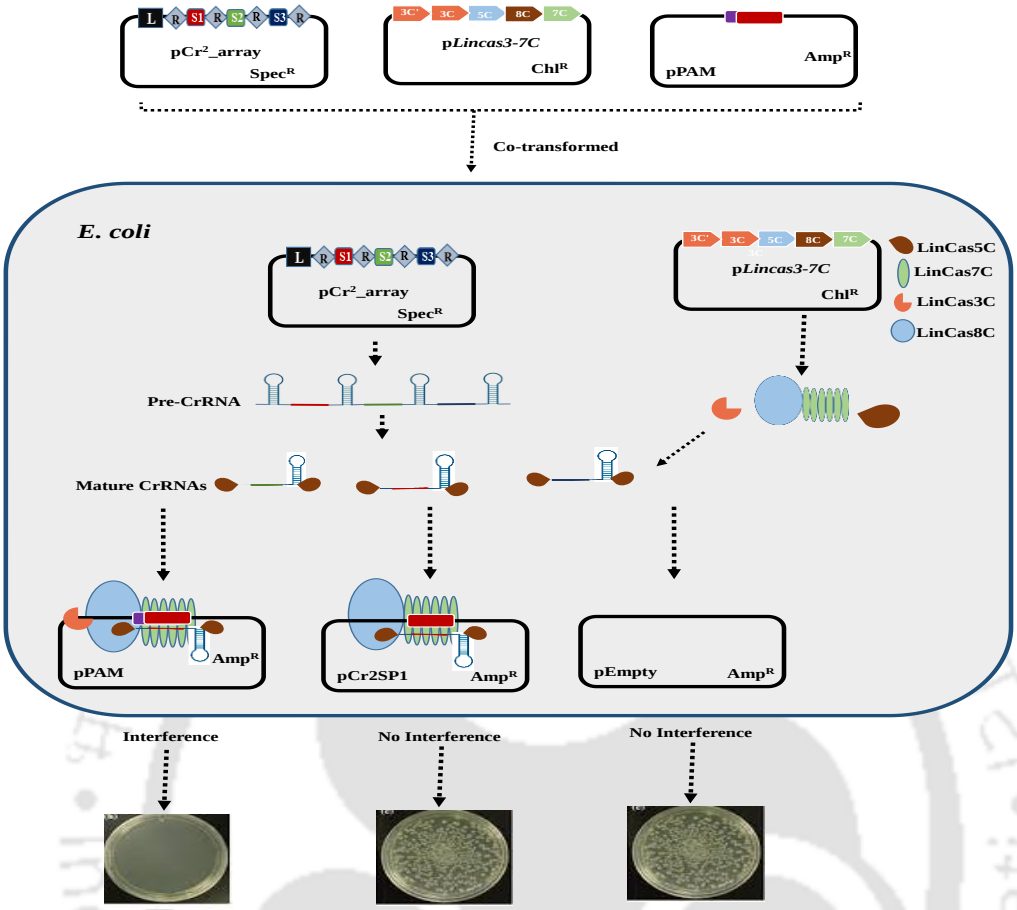


Figure 4.14. Schematic representation of plasmid interference by *L. interrogans* CRISPR-Cas I-C maturation and interference module. pCr2_array (Spec^R) harbors LIC_Cr^2 array which gets transcribed into Pre-CrRNA and further processed into mature CrRNA. p*Lincas3-7C* (Chl^R) harbors *cas* genes involved in the maturation and interference phase. pCr2SP1 (Amp^R) harbors the spacer1 of LIC_Cr^2 array without PAM, pPAM contains spacer 1 of LIC_Cr^2 along with the predicted PAM at its 5'end and pEmpty is empty plasmid. Mature crRNA will bind to the protospacer and recruit effector complex, thus leading to the degradation of the pPAM, and low transformants will be observed on LB agar plate containing spectinomycin, chloramphenicol, and ampicillin. The effector complex will bind to the protospacer in pCr2SP1 but will not be able to cause interference of plasmid due to lack of PAM whereas the effector complex will not bind to pEmpty as it lacks protospacer. Spec^R**:** Spectinomycin resistance, Chl^R**:** Chloramphenicol resistance, and Amp^R**:** Ampicillin resistance. L: leader, S: spacer, R: repeat, Red rectangle: protospacer, PAM is demarcated by a purple rectangle upstream of protospacer.

The degree of interference with plasmid DNA was quantified by calculating the percentage reduction in colony-forming units (CFU) in the test group compared to the control group. For PAM sequences TGA, ATA, and ATC (predicted using spacers from LIC_Cr^2), reductions of 75%, 69%, and 78% in CFU were observed, respectively **(Fig. 4.15A)**. For PAM

sequences TTA, TTT, GAA, TCT, and GCT (predicted using spacers from LIC_Cr3), reductions of 25%, 75%, 79%, 52%, and 53% in CFU were observed, respectively **(Fig. 4.15B)**.

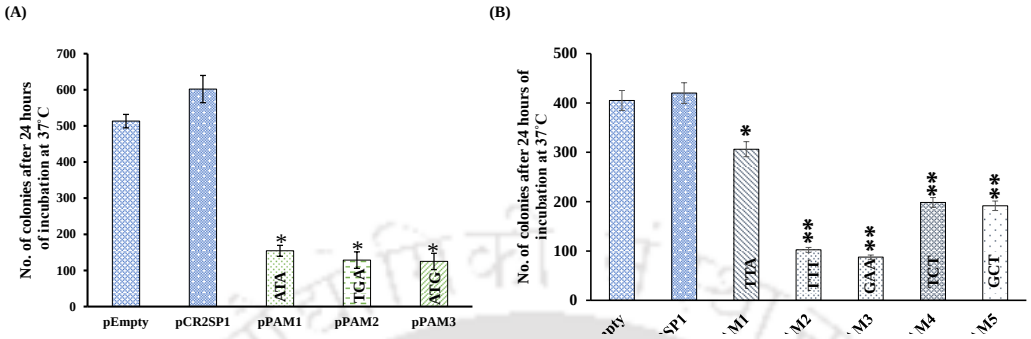


Figure 4.15. Plasmid interference by *L. interrogans* CRISPR-Cas I-C maturation and interference module. 5000 cells containing respective plasmids were plated on a LB agar plate containing antibiotics along with 0.05 mM IPTG and incubated at 37˚C for 24 hours. (A) Plasmid interference employing PAM predicted using spacers of array LIC_Cr^2 and (B) array LIC_Cr^3. Three independent experiments were performed. * Indicates the significance level p << 0.05 compared to pEmpty. pEmpty: empty pT7blue vector, pCR2SP1: pT7blue vector containing spacer1 of array LIC_Cr^2 without PAM.

4.3 Discussion

This study illustrates that the purified rLinCas5C from the subtype I-C locus, devoid of CRISPR element, is an active metal-independent endoribonuclease variant of Cas5C. The rLinCas5C possesses non-canonical cleavage activity on various RNA substrates, including repeat RNAs (sense and antisense), the pre-crRNAs (subtype I-B), and the orphan CRISPR array. The cleavage sites of rLinCas5C within different RNA substrates were not conserved. It denotes that the pure rLinCas5C may standalone not be efficient in generating functional crRNAs from pre-crRNA in *Leptospira*. The presence of other cognate Cas proteins may modulate the processivity of pre-crRNA by rLinCas5C. On the contrary, the catalytic active Cas5C protein family cleaves its cognate repeat RNA upstream of the 11-nt from the 3′-ends of pre-crRNA in a sequence-specific canonical approach to provide the mature crRNAs with an 11-nt 5′-handle (Garside et al., 2012; Hochstrasser & Doudna, 2015; Hochstrasser et al., 2016; Koo et al., 2013; Nam, Haitjema, et al., 2012).

Our *in vitro* RNase assays suggest that pure rLinCas5C may be ineffective in generating mature crRNAs from the array of subtype I-B locus or orphan CRISPR arrays. Controlled or regulated conduct of Cas proteins counts on many parameters like protein-protein interaction

TH-3470_166106104

encoded within the same locus or the presence of other accessory factors in the host, for example, integration host factor in *E. coli* (Nuñez et al., 2016). BhaCas5C alone had a weak affinity for pre-crRNA; however, the presence of Csd1 and Csd2 results in the formation of a stable complex, symbolizing that other Cascade elements may contribute to the selective pre-crRNA processing inside the cell (Nam, Haitjema, et al., 2012). Pure Cas nucleases of *Leptospira* (LinCas2) and *Sulfolobus solfataricus* (SsoCas2) display non-specific RNase activity *in vitro*. Nevertheless, the physiological function of Cas2 in the host is anticipated to be explicit in CRISPR biology (Anand et al., 2022; Beloglazova et al., 2008; Dixit et al., 2016). In *E. coli*, the role of the Cas1-Cas2 complex has been defined; it integrates new spacer DNA in the CRISPR locus of the host's genome (Rollie et al., 2015). The oligomerization of Cas1 and Cas2 of *E. coli* is vital to enforce the desired function. In addition, an integration host factor (IHF) was critical in the Cas1-Cas2 complex of *E. coli* to achieve efficient integration (Nuñez et al., 2016). Similarly, it is possible that LinCas5C, in coalition with other Cas proteins or host factors, may act as a specific endoribonuclease to set off crRNAs or Cascade I-C in the *Leptospira* CRISPR immunity. Thus, further study on LinCas5C and subtype-specific other Cas proteins is mandated to unveil subtype I-C's role in the *Leptospira*.

Interestingly, in the presence of rLinCas6, the catalytic activity of rLinCas5C is drastically reduced over repeat RNA or mature crRNA of the CRISPR I-B locus. The rLinCas6 after pre-crRNA processing is documented to remain bound with mature crRNA(Prakash & Kumar, 2021); however, the significance of such binding was obscure. This study reveals that the mature crRNA-rLinCas6 complex of subtype I-B fends from the endonuclease rLinCas5C. It is also conceivable that other Cas proteins involved in Cascade formation may fend off the mature crRNA degradation from co-existing endoribonucleases. In this study, heat denaturation (95°C for 5 min) of the pre-formed rLinCas6-crRNA nucleoprotein complex results in the dissociation of crRNA and rLinCas6. This makes the rLinCas5C accessible to crRNA for RNase activity. A bacterial genome may encode ≥1 CRISPR-Cas systems (Sokolowski et al., 2014); nevertheless, the Cas endoribonucleases (Cas6 in type I and III, Cas5C in type I-C systems) meant for processing and maturation of crRNA are specific to its cognate pre-crRNA. After that, other Cas proteins (Cas5, Cas7, Cas8, and Cas11) with or without Cas6 bind to mature crRNA and form a ribonucleoprotein complex (Cascade). Such binding of Cas proteins stabilizes the crRNA against cleavage by co-existing cellular RNases (Brendel et al., 2014; Peng et al., 2013).

TH-3470_166106104

Similarly, rLinCas6 (subtype I-B) processes the cognate pre-crRNA (LIC_Cr^{2+}) into mature crRNA (Prakash & Kumar, 2021) and forms a Cascade complex together with other Cas subunits (LinCas5, LinCas7, LinCas8b, and LinCas11b) (Hussain et al., 2023). Hence, the non-specific RNase activity of rLinCas5C in this study is shielded by the mature-crRNA-LinCas6 complex. However, the hypothesis about other Cas proteins of Cascade in fending crRNA must be tested before delineating any conclusion. This study also illustrates the role of divalent metals in the rLinCas5C non-specific endo-DNase activity on ss-DNA and ds-DNA. The DNase activity of rLinCas5C was also size-dependent as it was inert on short single-stranded and duplex DNA substrates. The DNase activity of rLinCas5C on ds-DNA and ss-DNA substrates agreed with BhaCas5C (Nam, Haitjema, et al., 2012).

On the contrary, XorCas5C has been shown to have a binding affinity for ds-DNA and ss-DNA substrates but no cleavage activity (Koo et al., 2013). This indicates that the ds-DNA cleavage activity of LinCas5C might be an evolutionary transformation that introduces the other feasible functions of LinCas5C in various stages of CRISPR biology. In a recent study, such non-specific DNase activity has been implied to confer a selective edge to bacteria by empowering defence against intruding phages that escape recognition (Vasu et al., 2012). The divalent metal-independent RNase and metal-dependent DNase activity of rLinCas5C indicate that the phosphodiester cleavage mechanism for RNA is different from DNA. In agreement, the Mg^{2+} ions tunable DNase/RNase activity of BhaCas5C has been reported elsewhere (Punetha et al., 2014). The simultaneous occurrence of RNA and DNA hydrolysis in the absence and presence of Mg^{2+} ions implies a single active site with variable target selectivity (Punetha et al., 2014). When the DNA is not in the vicinity of Cas5C, its low affinity for metal turns it into an RNase, allowing it to assist crRNA maturation. However, closeness to DNA, which may occur during the interference stage of CRISPR immunity, is expected to increase metal affinity, altering it into a DNase (Punetha et al., 2014). Thus, the predicted catalytic triad and metal-ion coordinating residues may account for the LinCas5C DNase activity. The modeled structure of LinCas5C aided us in categorizing it in the Cas5C-B subgroup. LinCas5C modeled structure encloses an N-terminal RRM-like domain with a helical region between β4 and β5 (α1′- α2′- α3′), a characteristic of all reported Cas5C-B subgroup endoribonucleases, and an additional β-sheet insertion (β1′-β2′) in the C-terminal β-sheet domain. *In silico* analysis predicted Phe141 as one of the catalytic triad residues (L51, K140, and F141) in modeled LinCas5C, equivalent to more conserved His residue in its orthologs. The deletion of additional

TH-3470_166106104

β-sheet (β1′-β2′) insertions in the LinCas5C and the substitution of Phe141 with His141 ushered gain of nuclease activity. In BhaCas5C, the residue Tyr46 (corresponds Leu51 in LinCas5C) was hypothesized to act as a base deprotonating the 2′-OH of G21 for inline nucleophilic attack on the scissile phosphate (Punetha et al., 2014). This hypothesis is also reinforced by the fact that replacing this 2′-OH with a deoxy derivative and/or changing this Tyr46 to Ala46 eliminates the cleavage (Carte et al., 2010; Nam, Haitjema, et al., 2012). In BhaCas5C, Lys116 (corresponds to Lys140 in LinCas5C) is a likely candidate to stabilize the negatively charged transition state, while His117 (corresponds to Phe141 in LinCas5C) may protonate the leaving group (Punetha et al., 2014; Raines, 1998). However, because the nucleophile is most likely a water molecule rather than an intrinsic 2′-OH group, the role of a nucleophile activator may be taken up by a Mg^{2+} ion during DNA hydrolysis (Yang et al., 2006). As a result, the active site that promotes RNA hydrolysis may also participate in DNA hydrolysis (Punetha et al., 2014). LinCas5C modeled structure may provide an added advantage of having a controlled RNase activity, an unprecedented evolutionary adaptation observed between the two β-strand (β1′ and β2′) and at one of the catalytic residue Phe141. Plasmid interference assay revealed CRISPR-Cas I-C exhibiting interference, thus LinCas5C in presence of other Cas protein is able to generate mature crRNA thus facilitating interference. Functioning of *Leptospira* interference module has also been previously studied using plasmid interference (Hussain et al., 2023). Interference has also been studied in *Bifidobacterium breve* by the means of plasmid interference employing predicted PAM (Han et al., 2024). This study, deliver strong evidence regarding the role of CRISPR-Cas I-C in providing immunity to *Leptospira* against foreign nucleic acids.

4.4 Materials and Methods.

4.4.1 Cloning and purification of recombinant LinCas5C.

The available genomic sequence of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 at NCBI was used for designing the primers (Table 4S1) and amplifying the Lin*cas5C* gene (*LIC12912*). The amplicon was ligated in the pET28-a vector at the *Nhe*I-*Xho*I restriction sites. The recombinant plasmid (pET28a-Lin*cas5C*) was outsourced to Bioserve for sequencing before the protein overexpression and purification. Recombinant proteins (rLinCas5C, rLinCas5C$^{\Delta\beta}$, and rLinCas5C^{F141H}) were overexpressed in *E. coli* BL21-AI using isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM) and L-arabinose (0.2 %) in LB broth

TH-3470_166106104

supplemented with 0.1 M sorbitol at 25°C for 8 h. Recombinant proteins were purified using Ni-NTA affinity chromatography.

4.4.2 Size-exclusion chromatography.

Size-exclusion chromatography was conducted using HiLoad™ 16/600 Superdex™ 200 pg column (GE Healthcare #28-9893-35) and was operated with an NGC chromatography system (BioRad). SEC was performed as described in Chapter 3.

4.4.3 Synthetic and in vitro synthesized RNA substrates.

Custom synthesized repeat RNA (sense and antisense LIC_Cr^3) substrates labeled (fluorescence) with 5′-FAM (6-fluorescein amidites) were purchased from Integrated DNA technologies (Table 4S1). The sense LIC_Cr^3 array transcription orientation was based on the program CRISPRCasdb (Pourcel et al., 2020) output. In addition, a pre-crRNA of array LIC_Cr^2 and its 5′ FAM-labeled sense repeat RNA is used, as described before (Prakash & Kumar, 2021). The pre-crRNA of array LIC_Cr^3 [sense (+) and antisense (-)] was synthesized *in vitro*, as reported previously (Prakash & Kumar, 2021). All *in vitro* synthesized pre-crRNA possesses a vector derived ten nucleotides (nt) sequences (5′-GGGAAAGCUU-3′) at its 5′-end.

4.4.4 RNase assay of rLinCas5 and rLinCas6 on repeat RNA.

The secondary structure of repeat RNA consensus with minimum free energy (MFE; ΔG°) was estimated at an experimental temperature (37°C) using an RNAalifold web server (Bernhart et al., 2008). RNase assays of rLinCas5C and its mutant variants were conducted in a nuclease buffer lacking Mg^{2+} ions (20 mM HEPES-KOH pH 8.0, 250 mM KCl, and 1 mM DTT). Briefly, the rLinCas5C or rLinCas6 (3 μM) was incubated with 5′ FAM-labeled repeat RNAs (250 nM) or unlabeled pre-crRNAs (100 nM) for an hour at 37°C, unless stated otherwise. Under similar reaction conditions as the control, an equivalent amount of substrates was incubated without rLinCas5C. The cleavage reactions were terminated, and the products were resolved by electrophoresis on denaturing urea (8 M) polyacrylamide gel (10-20%), as described previously (Prakash & Kumar, 2021; Sokolowski et al., 2014). The alkaline hydrolysis ladder was used to map the RNA cleavage sites. The hydrolyzed ladder was generated by incubating the sense repeat RNA of array LIC_Cr^2 and LIC_Cr$^{3+/-}$ with 1× alkaline hydrolysis buffer (1 mM EDTA, 50 mM sodium carbonate, pH 9.2) for 30 min at 95°C.

4.4.5 RNase assay of rLinCas5C and rLinCas6 on pre-crRNA and mature crRNA.

The activity of endoribonuclease (rLinCas6 and rLinCas5C) was explored on substrates (pre-crRNA and mature crRNA) in a nuclease buffer (20 mM HEPES-KOH pH 8.0, 250 mM

TH-3470_166106104

KCl, and 1 mM DTT) lacking Mg^{2+} ions. Unlabeled pre-crRNA of array LIC_Cr^{2+} (100 nM) was incubated with either pure rLinCas6 (1 µM) or rLinCas5C (1 µM) or both (1 µM each) for 30 min (unless stated otherwise) at 37°C. As controls of the assay, an equivalent amount of pre-crRNA was incubated without endoribonuclease (rLinCas5C or rLinCas6) or its heat-denatured variants (95°C, 5 min) (1 µM) under similar reaction conditions. The 5′ FAM-labeled mature crRNA (250 nM) of array LIC_Cr^{2+} was first incubated with rLinCas6 (1 µM) for 30 min at 37°C to dissect the RNase activity of rLinCas5C on mature crRNA. Then, rLinCas5C (1 µM) was incubated with an untreated or heat-denatured (95°C, 5 min) solution of rLinCas6-mature crRNA for 30 min at 37°C. Reaction products were resolved on denaturing polyacrylamide gel (10-15%) by electrophoresis and visualized with or without staining with SYBR-Gold.

4.4.6 RNase assay of rLinCas5C on fluorogenic RNA substrate.

RNase activity of rLinCas5C and its variants were quantified using the RNaseAlert kit (Integrated DNA technology, IDT; Cat #11-02-01-02). The RNaseAlert kit includes a synthetic RNA substrate which is fluorescence-quenched oligonucleotide probes. When cleaved by an RNase, the substrate tends to fluoresce green (490 nm excitation and 520 nm emission) and can be estimated by a fluorometer. RNase activity was conducted in black flat-bottom 96-well plates (Invitrogen) at 37°C. Fluorogenic RNA substrate (10 pmol) was incubated with rLinCas5C or its mutant variants (rLinCas5C^{F141H} and rLinCas5C$^{\Delta\beta}$, 0.5 µM) in a total reaction volume of 100 µL containing 1× reaction buffer (20 mM HEPES-KOH pH 8.0, 250 mM KCl, and 1 mM DTT) lacking Mg^{2+} ions. Fluorescence was estimated at every 5 min interval till 60 min using the Infinite M200Pro plate reader (Tecan).

4.4.7 DNase activity of rLinCas5C.

The DNase activity of rLinCas5C or its mutant variants (rLinCas5C^{F141H} and rLinCas5C$^{\Delta\beta}$) was assessed on DNA substrates of variable sizes. The DNA substrates of larger sizes were double-stranded circular plasmid (pTZ57R/T, 2.8 kb, 0.5 µg) and single-stranded circular DNA (Φx174 virion ss-DNA, 5.3 kb, 0.5 µg). The small-sized DNA oligos used as substrates were 5′ FAM-labeled repeat DNA of array LIC_Cr^{2-} (36-nt, 0.5 µM) and the 36 bp duplex repeat DNA of array LIC_Cr^{2} (0.5 µM). DNase assays were executed using rLinCas5C (3 µM) on the plasmid and viral DNA substrates for an hour at 37°C (or as indicated) in the cleavage buffer (20 mM HEPES-KOH, pH 8.0, and 100 mM KCl) supplemented with or without Mg^{2+} ion (10 mM). Reactions performed with large-size DNA substrates were stopped

TH-3470_166106104

using 6× DNA loading dye and were resolved by electrophoresis on 1% agarose gel stained with ethidium bromide. EDTA (ethylenediaminetetraacetic acid) was used as a divalent metal chelator at a final concentration of 10 mM. DNase activity executed on LIC_Cr^{2-} (36-nt) and LIC_Cr^{2} (36 bp) were either resolved on denaturation PAGE (20%) or native PAGE (10%) and visualized with or without SYBR-Gold staining. A protein storage buffer was utilised in place of endo-deoxyribonuclease in control reactions to exclude the possibility of nuclease contamination.

4.4.8 Bioinformatics analysis of LinCas5C and generation of its mutant variants.

The tertiary structure (3D) model of LinCas5C was predicted by AlphaFold (Jumper et al., 2021; Varadi et al., 2022) and decorated using PyMOL v2.4.1 (PyMOL molecular graphic system, Schrodinger, LLC, New York, NY, USA) (DeLano, 2002). The multiple sequence alignment (MSA) of LinCas5C (UniProt ID: Q72NC2) of *Leptospira* with its orthologs was conducted using Clustal Omega software (Thompson et al., 1994) under default parameters and visualized by the online tool ESPript v3 (Robert & Gouet, 2014). The Cas5C orthologs used for MSA were retrieved by Dali search (heuristic) (Holm, 2022) using the modeled LinCas5C structure as a query. The Cas5 orthologs used for MSA originated from diverse organisms, including *B. halodurans* (BhaCas5C; Q9KFY3), *Xanthomonas oryzae* (XorCas5C; WP_011257746.1), *Mannheimia succiniciproducens* (MsuCas5C; Q65TW5), *Streptococcus mutants* (SmuCas5C; Q8DSL7) and *Streptococcus pyogenes* (SpyCas5C; Q99YS3). The protein sequences were retrieved from the NCBI (National Centre for Biotechnology Information) or UniProtKB database (UniProt: A Hub for Protein Information, 2015). Based on MSA analysis of LinCas5C, the substitution of a single amino acid (Phe141 to His) or deletion of β-sheet insertion (β1′-β2′) of 11 residues (197^{th} - 207^{th}) was independently introduced in the vector (pET28a-Lin*cas*5C) using Q5 site-directed mutagenesis kit (#E0554S, NEB). The mutant variants obtained from mutagenesis were designated rLinCas5C^{F141H} (Phe141 to His141) and rLinCas5C$^{\Delta\beta}$ (β1′-β2′ deletion). Catalytic triad and metal coordinating residues were predicted using multiple sequence alignment and tertiary structure analysis. The crystal structure of XorCas5C (*Xanthomonas oryzae*; PDB ID: 3VZI), and BhaCas5C (*Bacillus halodurans*; PDB ID: 4F3M) coordinating residues of LinCas5C.

4.4.9 Protospacer construction.

To study interference protospacers were constructed using LIC_Cr^{2} and LIC_Cr^{3} spacers (5′-PAM-spacer1-3′). The positive and negative strands of each protospacers were

TH-3470_166106104

custom synthesised in such a way that after annealing cohesive ends (*Hind*III and *Nde*I) are generated to clone in pT7Blue plasmid.

4.4.10 Generation of tools for plasmid interference assay.

The interference module of CRISPR-Cas I-C was cloned in pACYCDuet vector. *Lincas7C-Lincas8C-Lincas5C* and *Lincas3C-Lincas3C′* ORFs were cloned at *BamH*I-*Xho*I *Sal*I-*Sac*I in the pACYCDuet vector, respectively. pACYCDuet vector containing CRISPR-Cas I-C interference module is named p*Lincas3-7C*. LIC_Cr^2 array was cloned in pCDF-1b vector (*BamH*I-*Sal*I) and named as pCr^2_array and pT7blue plasmid containing 5′PAM-LIC_Cr^2 spacer1 3′ is named as pPAM. pT7blue plasmid containing 5′LIC_Cr^2 spacer1 3′ is named as pCr2SP1 and Wild type pT7blue plasmid is named as pEmpty.

4.4.11 Plasmid interference assay.

The generated recombinant plasmids (p*Lincas3-7C*, pCr^2_array, and pPAM) were transformed in *E. coli* BL21 (DE3) cells along with empty pT7blue as a negative control. The transformed *E. coli* BL21 were cultured to mid-log phase in LB broth containing antibiotics (ampicillin, chloramphenicol, and spectinomycin). Half of these cultures were left uninduced, while the other half were induced with IPTG and L-arabinose for 4 hours at 37°C. Thereafter, assuming OD600 1 = 1×109 CFU/mL, 5000 cells of each bacterial strain were spread onto LB agar plates supplemented with antibiotics (ampicillin, chloramphenicol, and spectinomycin, 100 µg/ml each) and IPTG (0.05 mM). The culture plates were incubated at 37ºC for the next 24 hours and scored for growth by colony counting.

Table 4.2. Oligos used in Chapter 4.

DNA oligos	Sequence (5'-3')	Purpose
Lincas5C: F	CTAGCTAGCATGGAAACAAAAAGTATGAAGGAATC	Cloning in pET28a
Lincas5C: R	CCGCTCGAGTCATAGTCTCGCTCCATTGGT	
LIC_Cr^{3+}: F	CTAAAGCTTTTCCAAAGGAAATTGGAAACTT	Cloning in pTZ57R/T for pre-crRNA
LIC_Cr^{3+}: R	CCGGGTACCTTTCTTAAGTTCCCGATTTCTTGC	
LIC_Cr^{3-}: F	CCGGGTACCTTCCAAAGGAAATTGGAAACTT	
LIC_Cr^{3-}: R	CTAAAGCTTTTTCTTAAGTTCCCGATTTCTTGC	
Lincas5C^{F141H}: F	TATTACAAAACATGAAGAGATGTTTAAACGAAGA	Site-directed mutagenesis of pET28a-Lincas5C
Lincas5C^{F141H}: R	CTGTCCCATCTCTCGCT	
Lincas5C$^{\Delta\beta}$: F	GCCAATCCAAAATTTTTTCG	
Lincas5C$^{\Delta\beta}$: R	GTCTTCAAAATCAAGATCG	
RNA oligos	**Sequence (5′-3′) with 5′-FAM modification**	

TH-3470_166106104

Sense repeat RNA of LIC_Cr^2	CUGAAUAUAACUUUGAUGCCGUUAGGCGUUGAGCAC
Sense repeat RNA of LIC_Cr^3	UUCCUAAAGAAAUAGGGAAUUUAAAAAA
Antisense repeat RNA of LIC_Cr^3	UUUUUUAAAUUCCCUAUUUCUUUAGGAA

****

Chapter 5

Conclusions and future prospects

Leptospira interrogans is a pathogenic strain that causes leptospirosis. This disease is a significant public health challenge and due to various ailments in livestock, leads to economic loss (Costa et al., 2015; Ellis, 2015). Genetic manipulation in pathogenic *Leptospira* is difficult; therefore, the pathogenesis mechanism is poorly understood (Fernandes et al., 2019). It is believed that the presence of the CRISPR-Cas system may be responsible for this difficulty. The *L. interrogans* serovar (sv.) Copenhageni genome encodes two CRISPR-Cas system subtypes, I-B and I-C, and two CRISPR arrays (LIC_Cr^2 and LIC_Cr^3) (Makarova et al., 2013; Xiao et al., 2019). While the CRISPR-Cas I-B has been well studied (Dixit, Anand, et al., 2021b; Dixit et al., 2016; Dixit, Prakash, et al., 2021; Hussain et al., 2023; Hussain & Kumar, 2022; Prakash & Kumar, 2021, 2022), this study focuses on CRISPR-Cas I-C.

The arrangement of the CRISPR array is unique in *L. interrogans* CRISPR-Cas I-B. Unlike most studied CRISPR-Cas systems where the CRISPR array is conventionally located up or downstream of the *cas* genes operon, the LIC_Cr^2 array is situated between the two *cas* operons (Dixit et al., 2016). The LIC_Cr^3 array, on the other hand, is spatially distant from both subtypes (I-B and I-C), making it an orphan CRISPR array. The LIC_Cr^3 array comprises four spacers (41 bp) and four repeats (28 bp). Using spacers of the, LIC_Cr^2 and LIC_Cr^3 array, “TGA”, “ATA” and “TTA”, “TTT”, “GAA”, “TCT”, “GCT”, were predicted potential PAM, respectively.

The CRISPR-Cas I-C of sv. Copenhageni has eight *cas* genes but no CRISPR array nearby. The existence of the subtype I-C locus in *L. interrogans* sv. Copenhageni, without a CRISPR array, instigated us to apprehend the role of its Cas proteins in CRISPR biology. The set of *cas* gene and protein sequences in CRISPR-Cas I-C and I-B were aligned to assess their degree of similarity. CRISPR-Cas I-C and I-B have 30-40% similarity in *cas* gene sequence and 2-40% in protein sequence, and they are phylogenetically distinct. The *cas* genes in CRISPR-Cas I-C of sv. Copenhageni are transcriptionally active. Within the adaptation module, *Lincas1C* and *Lincas2C* share a common promoter for transcription. LinCas2C is also found to be translationally active *in vitro* grown *Leptospira*, but LinCas1C could hardly be detected due to minimal protein expression. In-depth proteome analysis of the *Haloferax volcanii* revealed that the Cas1 protein is present in low levels under standard growth conditions (Jevtić et al., 2019).

TH-3470_166106104

The adaptation module of *L. interrogans* sv. Copenhageni comprises LinCas1C, LinCas2C, and LinCas4C. Sequence and tertiary structure analysis of LinCas1C shows that it is approximately 80-90 amino acids shorter towards the N-terminal than its closest orthologs. Residues E78, H239, and E162 are predicted to form the nucleolytic core site in LinCas1C and might be involved in nuclease activity. Recombinant LinCas1C, an adaptation Cas protein of CRISPR-Cas I-C, demonstrated nuclease activity on diverse DNA substrates in a divalent metal- and pH-dependent manner. Recombinant LinCas1C also exhibited RNase activity independent of metal-ion on *luciferase* mRNA substrate. Furthermore, rLinCas1C was inert toward small DNA oligos (36-50-mer).

Cas2C of sv. Lai (LinCas2C_Lai) expresses truncated protein due to natural adenine (108th) deletion mutation (Xiao et al., 2019). The rLinCas2C and naturally truncated rLinCas2C_Lai demonstrated nuclease activity on diverse DNA substrates in a divalent metal- and pH-dependent manner. However, these nucleases were inert towards small DNA oligos (23-50-mer). The rLinCas2C and rLinCas2C_Lai exhibited cleavage of mRNA transcripts of *luciferase* gene independent of divalent metal ions. The rLinCas2C_Lai exhibited nuclease activity even after expressing the truncated protein. The structural analysis of rLinCas2C reveals that it adopts a dimeric conformation, and each subunit exhibits the characteristic ferredoxin fold. In LinCas2C, the dimeric interface β4 of one protomer interacts with β5 of another protomer. The structural analysis of rLinCas2C highlights the significance of the catalytic aspartate residue in constraining conformational flexibility. The separation between the conserved aspartate residues in each protomer plays a crucial role in metal-ion coordination. In the case of LinCas2C, this distance measures 11.0 Å, which appears to be insufficient for the coordination of a single Mg^{2+} ion. To enable these aspartates to bind a bridging metal, rLinCas2C would require a substantial conformational alteration in the β1 or ferredoxin fold region or even a complete reorientation of the dimeric structure. Thus, the crystal structure of rLinCas2C (dimeric form) indicates it is in a catalytically inactive conformational state. The rLinCas2C residues interacting with DNA are primarily from the loop L1 (putative DNA binding loop), L2 (putative DNA binding loop), and α1 regions. The L1 and L2 loops in LinCas2C were observed to be comparable to those in its closest homologs. In the modeled structure of LinCas2C_Lai, the L1 loop displays an identical conformation to its closest homologs, while the L2 loop was noted to be shorter in length. Alanine replacement mutation of active-site residue (Try7, Asp8, Arg33, and Phe39) and deletion of RNA recognition loop

TH-3470_166106104

(L2) resulted in compromised DNase activity; however, slightly reduced RNase activity was observed in selected mutants.

Recombinant LinCas4C, an adaptation Cas protein of CRISPR-Cas I-C, demonstrated nuclease activity on diverse DNA substrates in a divalent metal- and pH-dependent manner. Recombinant LinCas4C exhibited RNase activity independent of metal-ion on *luciferase* mRNA substrate. Additionally, rLinCas4C cleaves oligos sequence specifically. BhaCas4 cleaves prespacer highly specific to the PAM end (Lee et al., 2019). Analogously, the PAM-dependent cleavage of single-stranded oligos using LinCas4C can be further investigated. The cleavage of ss-DNA plays a crucial role in the pre-integration processing of the protospacer, which is bound by the Cas1-Cas2 complex before being incorporated into the CRISPR loci integration (Nuñez et al., 2016). The apparent inability of LinCas1C and LinCas2C to effectively act on short ss-DNAs might potentially be addressed by the enzymatic activity of LinCas4 in *Leptospira interrogans*. LinCas1C, LinCas2C and LinCas4C endonuclease activity may provide an advantage with an additional role other than typical CRISPR adaptation. *In vivo* naive adaptation using CRISPR-Cas I-C could not be detected in heterologous host (*E. coli*). Various host factors play pivotal roles in aiding the integration of new spacers by Cas1 and Cas2. In the Cas1-Cas2 complex of *E. coli*, the integration host factor (IHF) is crucial for achieving efficient integration (Nuñez et al., 2016). *Sulfolobus islandicus*, Csa3a and Csa3b, additional CRISPR factors serve as accessory elements in the adaptation process (Liu et al., 2022). Likewise, it is plausible that the *Leptospira* adaptation module requires an additional host factor to facilitate spacer integration. The identification of this host factor is essential for further *in vivo* or *in vitro* studies on adaptation in *Leptospira*.

CRISPR-Cas I-C maturation and interference machinery comprise LinCas5C, LinCas7C, LinCas8C, and LinCas3C. In general, the Cas5C protein in CRISPR-Cas I-C is catalytically active (Hochstrasser & Doudna, 2015). Catalytic active Cas5C cleaves the cognate repeat RNA segment upstream of the 11-nt from the 3′-ends of pre-crRNA in a sequence-specific canonical mode to render the mature crRNAs with an 11-nt 5′-handle (Garside et al., 2012; Hochstrasser & Doudna, 2015; Hochstrasser et al., 2016; Koo et al., 2013; Nam, Haitjema, et al., 2012). On the contrary, recombinant LinCas5C exhibits non-canonical cleavage activity on non-cognate repeat RNAs. In addition, standalone rLinCas5C is unable to generate mature crRNA from pre-crRNA from the array of LIC_Cr^{2} and LIC_Cr^{3}; thus, rLinCas5C (subtype I-C) displayed a remarkable difference in the CRISPR array processing. The presence of other cognate Cas

TH-3470_166106104

proteins may modulate the processivity of pre-crRNA by rLinCas5C. BhaCas5C alone exhibits a limited affinity for pre-crRNA. Nevertheless, the co-presence of Csd1 and Csd2 leads to the establishment of a stable complex. This implies that additional Cascade elements may play a role in the selective processing of pre-crRNA within the cellular environment (Nam, Haitjema, et al., 2012). This study discloses that the mature crRNA-rLinCas6 complex within subtype I-B is protected from the endoribonuclease activity of rLinCas5C. It is also plausible that additional Cas proteins participating in Cascade assembly could shield the mature crRNA from concurrent degradation by co-existing endoribonucleases. This study further elucidates the role of divalent metals in the non-specific endo-DNase activity of rLinCas5C on ssDNA and dsDNA substrates. The DNase activity of rLinCas5C was also contingent on the size of the DNA substrates, as it displayed no activity on short ssDNA and dsDNA oligomers. Modeled structure analysis of LinCas5C, categorizing it in the Cas5C-B subgroup. LinCas5C tertiary structure exclusively contains additional β-sheet insertion in the C-terminal β-sheet domain. *In silico* analysis predicted Phe141 as one of the catalytic triad residues (L51, K140, and F141) in modeled LinCas5C, equivalent to a more conserved histidine residue in its orthologs. The deletion of additional β-sheet insertions and the substitution of Phe141 with more conserved His141 in the LinCas5C accompanied the gain of nuclease activity. The ds-DNA cleavage activity observed in LinCas5C could represent an evolutionary adaptation, introducing additional potential functions of LinCas5C across different stages of CRISPR biology. A recent study suggests that this non-specific DNase activity may provide bacteria with a selective advantage by enhancing defence against invading phages that manage to evade recognition (Vasu et al., 2012).

This study raises questions about the effectiveness of CRISPR-Cas I-C's interference module due to the suboptimal processing of pre-crRNA by LinCas5C. However, the plasmid interference assay proved that CRISPR-Cas I-C can still cause interference, implying that when accompanied by other interference Cas proteins, LinCas5C can process pre-crRNA into mature crRNA, thus enabling interference. The study focuses on examining Cas proteins during the adaptation and maturation phase, and also explores the interference module of CRISPR-Cas I-C in a heterologous host. The findings pave the way for utilizing the endogenous CRISPR-Cas Type-I of *L. interrogans* for genome editing since interference machinery of both the subtypes is active. To implement this, a miniature CRISPR array must be designed, including the *Leptospira* leader sequence, repeat sequence, and a spacer targeting the desired gene. The

TH-3470_166106104

predicted trinucleotide PAM sequence needs to be identified in the target gene, and about 36 base pairs downstream of the PAM should be used as the spacer. When a miniature CRISPR array is transcribed, it forms a pre-crRNA, which is then processed into mature crRNA by endogenous Cas endonucleases during the maturation phase. The mature crRNA will bind to the Cas endonuclease, forming an interference complex that will base-pairs with the target gene in the *L. interrogans* genome. The interference complex will recognize the PAM sequence, leading to the degradation of the gene within the organism's own genome. Henceforth, it will be advantageous in genetic manipulation in pathogenic *Leptospira*.

References

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. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, *353*(6299), aaf5573.

Adler, B., & de la Peña Moctezuma, A. (2010). Leptospira and leptospirosis. *Veterinary microbiology*, *140*(3-4), 287-296.

Agrawal, S. K., Chaudhry, R., Gupta, N., Arif, N., & Bhadur, T. (2018). Decreasing trend of seroprevalence of leptospirosis at all India Institute of medical Sciences New Delhi: 2014–2018. *Journal of Family Medicine and Primary Care*, *7*(6), 1425-1428.

Almendros, C., Nobrega, F. L., McKenzie, R. E., & Brouns, S. J. J. (2019). Cas4–Cas1 fusions drive efficient PAM selection and control CRISPR adaptation. *Nucleic acids research*, *47*(10), 5223-5230.

Amitai, G., & Sorek, R. (2016). CRISPR–Cas adaptation: insights into the mechanism of action. *Nature Reviews Microbiology*, *14*(2), 67-76.

Anand, V., Prabhakaran, H. S., Gogoi, P., Kanaujia, S. P., & Kumar, M. (2022). Structural and functional characterization of Cas2 of CRISPR-Cas subtype IC lacking the CRISPR component. *Frontiers in Molecular Biosciences*, 953.

Anantharaman, V., Makarova, K. S., Burroughs, A. M., Koonin, E. V., & Aravind, L. (2013). Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biology direct*, *8*, 1-28.

Arber, W. (1965). Host specificity of DNA produced by Escherichia coli: V. The role of methionine in the production of host specificity. *Journal of molecular biology*, *11*(2), 247-256.

Arent, Z., Pardyak, L., Dubniewicz, K., Plachno, B., & Kotula-Balak, M. (2022). Leptospira taxonomy: then and now.

Arnoldini, M., Vizcarra, I. A., Peña-Miller, R., Stocker, N., Diard, M., Vogel, V., Beardmore, R. E., Hardt, W.-D., & Ackermann, M. (2014). Bistable expression of virulence genes in salmonella leads to the formation of an antibiotic-tolerant subpopulation. *PLoS biology*, *12*(8), e1001928.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., & Binkowski, A. (2011). A dual function of the CRISPR–Cas system in bacterial antivirus immunity and DNA repair. *Molecular microbiology*, *79*(2), 484-502.

Barocchi, M. A., Ko, A. I., Reis, M. G., McDonald, K. L., & Riley, L. W. (2002). Rapid translocation of polarized MDCK cell monolayers by Leptospira interrogans, an invasive but nonintracellular pathogen. *Infection and immunity*, *70*(12), 6926-6932.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709-1712.

Barrangou, R., & Van Der Oost, J. (2013). CRISPR-Cas systems. *RNA-Mediated Adaptative Immunity in Bacteria and Archaea. 101007978364234657 6th ed. Heidelberg SVB, editor*.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 271-281.

Behler, J., & Hess, W. R. (2020). Approaches to study CRISPR RNA biogenesis and the key players involved. *Methods*, *172*, 12-26.

Beloglazova, N., Brown, G., Zimmerman, M. D., Proudfoot, M., Makarova, K. S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., & Minor, W. (2008). A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *Journal of Biological Chemistry*, *283*(29), 20361-20371.

Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., & Jain, S. (2002). The protein data bank. *Acta Crystallographica Section D: Biological Crystallography*, *58*(6), 899-907.

Berman, H. M., Bhat, T. N., Bourne, P. E., Feng, Z., Gilliland, G., Weissig, H., & Westbrook, J. (2000). The Protein Data Bank and the challenge of structural genomics. *Nature structural biology*, *7*(11), 957-959.

Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R., & Stadler, P. F. (2008). RNAalifold: improved consensus structure prediction for RNA alignments. *BMC bioinformatics*, *9*, 1-13.

Bernheim, A., & Sorek, R. (2020). The pan-immune system of bacteria: antiviral defence as a community resource. *Nature Reviews Microbiology*, *18*(2), 113-119.

Bertelsen, M. B., Senissar, M., Nielsen, M. H., Bisiak, F., Cunha, M. V., Molinaro, A. L., Daines, D. A., & Brodersen, D. E. (2021). Structural basis for toxin inhibition in the VapXD toxin-antitoxin system. *Structure*, *29*(2), 139-150. e133.

Bharathkumar, N., Sunil, A., Meera, P., Aksah, S., Kannan, M., Saravanan, K. M., & Anand, T. (2022). CRISPR/Cas-Based modifications for therapeutic applications: A review. *Molecular biotechnology*, *64*(4), 355-372.

Bharti, A. R., Nally, J. E., Ricaldi, J. N., Matthias, M. A., Diaz, M. M., Lovett, M. A., Levett, P. N., Gilman, R. H., Willig, M. R., & Gotuzzo, E. (2003). Leptospirosis: a zoonotic disease of global importance. *The Lancet infectious diseases*, *3*(12), 757-771.

Bikard, D., & Marraffini, L. A. (2012). Innate and adaptive immunity in bacteria: mechanisms of programmed genetic variation to fight bacteriophages. *Current opinion in immunology*, *24*(1), 15-20.

Biswas, A., Gagnon, J. N., Brouns, S. J., Fineran, P. C., & Brown, C. M. (2013). CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA biology*, *10*(5), 817-827.

Biswas, A., Staals, R. H., Morales, S. E., Fineran, P. C., & Brown, C. M. (2016). CRISPRDetect: a flexible algorithm to define CRISPR arrays. *BMC genomics*, *17*(1), 1-14.

Blair, J. M., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. (2015). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, *13*(1), 42-51.

Blower, T. R., Pei, X. Y., Short, F. L., Fineran, P. C., Humphreys, D. P., Luisi, B. F., & Salmond, G. P. (2011). A processed noncoding RNA regulates an altruistic bacterial antiviral system. *Nature structural & molecular biology*, *18*(2), 185-190.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., & Benning, C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. *The EMBO journal*, *17*(1), 170-180.

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, *151*(8), 2551-2561.

Brautigam, C. A., & Steitz, T. A. (1998). Structural principles for the inhibition of the 3′-5′ exonuclease activity of Escherichia coli DNA polymerase I by phosphorothioates. *Journal of molecular biology*, *277*(2), 363-377.

Brendel, J., Stoll, B., Lange, S. J., Sharma, K., Lenz, C., Stachler, A.-E., Maier, L.-K., Richter, H., Nickel, L., & Schmitz, R. A. (2014). A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in Haloferax volcanii. *Journal of Biological Chemistry*, *289*(10), 7164-7177.

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., & Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, *321*(5891), 960-964.

Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, *355*(6359), 472-475.

Bulach, D. M., Zuerner, R. L., Wilson, P., Seemann, T., McGrath, A., Cullen, P. A., Davis, J., Johnson, M., Kuczek, E., & Alt, D. P. (2006). Genome reduction in Leptospira borgpetersenii reflects limited transmission potential. *Proceedings of the National Academy of Sciences*, *103*(39), 14560-14565.

Cady, K. C., Bondy-Denomy, J., Heussler, G. E., Davidson, A. R., & O'Toole, G. A. (2012). The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. *Journal of bacteriology*, *194*(21), 5728-5738.

Caimi, K., & Ruybal, P. (2020). Leptospira spp., a genus in the stage of diversity and genomic data expansion. *Infection, Genetics and Evolution*, *81*, 104241.

Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M., & Terns, M. P. (2010). Binding and cleavage of CRISPR RNA by Cas6. *Rna*, *16*(11), 2181-2188.

Charpentier, E., Richter, H., van der Oost, J., & White, M. F. (2015a). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS microbiology reviews*, *39*(3), 428-441.

Charpentier, E., Richter, H., van der Oost, J., & White, M. F. (2015b). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS microbiology reviews*, fuv023.

Chen, T., Chen, Z., Zhang, H., Li, Y., Yao, L., Zeng, B., & Zhang, Z. (2023). Development of a CRISPR/Cpf1 system for multiplex gene editing in Aspergillus oryzae. *Folia Microbiologica*, 1-10.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*(1), 12-21.

Clement, J.-M., Lepouce, E., Marchal, C., & Hofnung, M. (1983). Genetic study of a membrane protein: DNA sequence alterations due to 17 lamB point mutations affecting adsorption of phage lambda. *The EMBO journal*, *2*(1), 77-80.

Connerton, I. F., & Hooton, S. P. (2015). Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harbouring a CRISPR-like Cas4 protein. *Frontiers in Microbiology*.

Cornforth, D. M., & Foster, K. R. (2013). Competition sensing: the social side of bacterial stress responses. *Nature Reviews Microbiology*, *11*(4), 285-293.

Corradi, H. R. (2015). Using Crystallographic Data To Facilitate Students' Discovery of How Protein Models Are Produced An Activity Illustrating the Effect of Resolution on Model Quality. *Journal of Chemical Education*, *92*(12), 2117-2119.

Costa, F., Hagan, J. E., Calcagno, J., Kane, M., Torgerson, P., Martinez-Silveira, M. S., Stein, C., Abela-Ridder, B., & Ko, A. I. (2015). Global morbidity and mortality of leptospirosis: a systematic review. *PLoS neglected tropical diseases*, *9*(9), e0003898.

Crawley, A. B., Henriksen, J. R., & Barrangou, R. (2018). CRISPRdisco: an automated pipeline for the discovery and analysis of CRISPR-Cas systems. *The CRISPR Journal*, *1*(2), 171-181.

Croda, J., Figueira, C. P., Wunder Jr, E. A., Santos, C. S., Reis, M. G., Ko, A. I., & Picardeau, M. (2008). Targeted mutagenesis in pathogenic Leptospira species: disruption of the LigB gene does not affect virulence in animal models of leptospirosis. *Infection and immunity*, *76*(12), 5826-5833.

Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome research*, *14*(6), 1188-1190.

Culley, A. I. (2011). Virophages to viromes: a report from the frontier of viral oceanography. *Current Opinion in Virology*, *1*(1), 52-57.

Damiano, D. K., Azevedo, B. O., Fernandes, G. S., Teixeira, A. F., Gonçalves, V. M., Nascimento, A. L., & Lopes, A. P. (2024). The Toxin of VapBC-1 Toxin-Antitoxin Module from Leptospira interrogans Is a Ribonuclease That Does Not Arrest Bacterial Growth but Affects Cell Viability. *Microorganisms*, *12*(8), 1660.

DAS, M., HARVEY, I., CHU, L. L., SINHA, M., & PELLETIER, J. (2001). Full-length cDNAs: more than just reaching the ends. *Physiological genomics*, *6*(2), 57-80.

DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography*, *40*(1), 82-92.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011a). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011b). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602-607.

Deng, L., Garrett, R. A., Shah, S. A., Peng, X., & She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in S ulfolobus. *Molecular microbiology*, *87*(5), 1088-1099.

Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of bacteriology*, *190*(4), 1390-1400.

Deveau, H., Garneau, J. E., & Moineau, S. (2010). CRISPR/Cas System and Its Role in Phage-Bacteria Interactions. *Annual Review of Microbiology*, *64*(1), 475-493. https://doi.org/10.1146/annurev.micro.112408.134123

Dhingra, Y., Suresh, S. K., Juneja, P., & Sashital, D. G. (2022). PAM binding ensures orientational integration during Cas4-Cas1-Cas2-mediated CRISPR adaptation. *Molecular Cell*, *82*(22), 4353-4367. e4356.

Díez-Villaseñor, C., Almendros, C., García-Martínez, J., & Mojica, F. J. (2010). Diversity of CRISPR loci in Escherichia coli. *Microbiology*, *156*(5), 1351-1361.

Dixit, B. (2021). *Elucidation of adaptive stage of CRISPR-Cas subtype I-B defense system in Leptospira interrogans Copenhageni strain Fiocruz L1-130* Indian Institute of Technology Guwahati]. https://www.iitg.ac.in/lib/

Dixit, B., Anand, V., Hussain, M. S., & Kumar, M. (2021a). The CRISPR-associated Cas4 protein from Leptospira interrogans demonstrate versatile nuclease activity. *Current Research in Microbial Sciences*, 100040.

Dixit, B., Anand, V., Hussain, M. S., & Kumar, M. (2021b). The CRISPR-associated Cas4 protein from Leptospira interrogans demonstrate versatile nuclease activity. *Current Research in Microbial Sciences*, *2*, 100040.

Dixit, B., Ghosh, K. K., Fernandes, G., Kumar, P., Gogoi, P., & Kumar, M. (2016). Dual nuclease activity of a Cas2 protein in CRISPR–Cas subtype I-B of Leptospira interrogans. *FEBS letters*, *590*(7), 1002-1016.

Dixit, B., Prakash, A., Kumar, P., Gogoi, P., & Kumar, M. (2021). The core Cas1 protein of CRISPR-Cas IB in Leptospira shows metal-tunable nuclease activity. *Current Research in Microbial Sciences*, *2*, 100059.

Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., & Yang, D. (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature*, *532*(7600), 522-526.

Dong, T. G., Dong, S., Catalano, C., Moore, R., Liang, X., & Mekalanos, J. J. (2015). Generation of reactive oxygen species by lethal attacks from competing microbes. *Proceedings of the National Academy of Sciences*, *112*(7), 2181-2186.

Donohoue, P. D., Barrangou, R., & May, A. P. (2018). Advances in industrial biotechnology using CRISPR-Cas systems. *Trends in biotechnology*, *36*(2), 134-146.

Dupureur, C. M. (2008). Roles of metal ions in nucleases. *Current opinion in chemical biology*, *12*(2), 250-255.

Dy, R. L., Richter, C., Salmond, G. P., & Fineran, P. C. (2014). Remarkable mechanisms in microbes to resist phage infections. *Annual review of virology*, *1*, 307-331.

Ebihara, A., Yao, M., Masui, R., Tanaka, I., Yokoyama, S., & Kuramitsu, S. (2006). Crystal structure of hypothetical protein TTHB192 from Thermus thermophilus HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein science*, *15*(6), 1494-1499.

Ellis, W. A. (2015). Animal leptospirosis. *Leptospira and leptospirosis*, 99-137.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 486-501.

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, *69*(7), 1204-1214.

Fagerlund, R. D., Wilkinson, M. E., Klykov, O., Barendregt, A., Pearce, F. G., Kieper, S. N., Maxwell, H. W., Capolupo, A., Heck, A. J., & Krause, K. L. (2017). Spacer capture and integration by a type IF Cas1–Cas2-3 CRISPR adaptation complex. *Proceedings of the National Academy of Sciences*, *114*(26), E5122-E5128.

Faine, S. (1999). AB, Bolin C. *Perolat P. Leptospira and Leptospirosis. 2nd ed: Published by MediSci, Melbourne, Vic. Australia*.

Fan, X.-Y., Tang, B.-K., Xu, Y.-Y., Han, A.-X., Shi, K.-X., Wu, Y.-K., Ye, Y., Wei, M.-l., Niu, C., & Wong, K.-W. (2018). Oxidation of dCTP contributes to antibiotic lethality in stationary-phase mycobacteria. *Proceedings of the National Academy of Sciences*, *115*(9), 2210-2215.

Fernandes, L. G., & Nascimento, A. L. (2022). A novel breakthrough in Leptospira spp. mutagenesis: knockout by combination of CRISPR/Cas9 and non-homologous end-joining systems. *Frontiers in Microbiology*, *13*, 915382.

Fernandes, L. G., Teixeira, A. F., & Nascimento, A. L. (2023). Evaluation of Leptospira interrogans knockdown mutants for LipL32, LipL41, LipL21, and OmpL1 proteins. *Frontiers in Microbiology*, *14*, 1199660.

Fernandes, L. G. V., Guaman, L., Vasconcellos, S. A., Heinemann, M. B., Picardeau, M., & Nascimento, A. L. T. O. d. (2019). Gene silencing based on RNA-guided catalytically inactive Cas9 (dCas9): a new tool for genetic engineering in Leptospira. *Scientific reports*, *9*(1), 1839.

Fernandes, L. G. V., Hamond, C., Tibbs-Cortes, B. W., Putz, E. J., Olsen, S. C., Palmer, M. V., & Nally, J. E. (2024). CRISPR-prime editing, a versatile genetic tool to create specific mutations with a single nucleotide resolution in Leptospira. *mBio*, e01516-01524.

Fernandes, L. G. V., Hornsby, R., Nascimento, A. L. T. O. d., & Nally, J. (2021). Genetic manipulation of pathogenic Leptospira: CRISPR interference (CRISPRi)-mediated gene silencing and rapid mutant recovery at 37 C. *Scientific reports*, *11*(1), 1768.

Fouts, D. E., Matthias, M. A., Adhikarla, H., Adler, B., Amorim-Santos, L., Berg, D. E., Bulach, D., Buschiazzo, A., Chang, Y.-F., & Galloway, R. L. (2016). What makes a bacterial species pathogenic?: comparative genomic analysis of the genus Leptospira. *PLoS neglected tropical diseases*, *10*(2), e0004403.

Gagaletsios, L. A., Papagiannitsis, C. C., & Petinaki, E. (2022). Prevalence and analysis of CRISPR/Cas systems in Pseudomonas aeruginosa isolates from Greece. *Molecular Genetics and Genomics*, *297*(6), 1767-1776.

Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H., & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67-71.

Garrett, R. A., Vestergaard, G., & Shah, S. A. (2011). Archaeal CRISPR-based immune systems: exchangeable functional modules. *Trends in microbiology*, *19*(11), 549-556.

Garrett, S., Shiimori, M., Watts, E. A., Clark, L., Graveley, B. R., & Terns, M. P. (2020). Primed CRISPR DNA uptake in Pyrococcus furiosus. *Nucleic Acids Research*, *48*(11), 6120-6135.

Garside, E. L., Schellenberg, M. J., Gesner, E. M., Bonanno, J. B., Sauder, J. M., Burley, S. K., Almo, S. C., Mehta, G., & MacMillan, A. M. (2012). Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *Rna*, *18*(11), 2020-2028.

Gerdes, K., Christensen, S. K., & Løbner-Olesen, A. (2005). Prokaryotic toxin–antitoxin stress response loci. *Nature Reviews Microbiology*, *3*(5), 371-382. https://doi.org/10.1038/nrmicro1147

Gerlt, J. A., Coderre, J. A., & Mehdi, S. (1983). Oxygen chiral phosphate esters. *Adv Enzymol Relat Areas Mol Biol*, *55*, 291-380.

Geurts, M. H., de Poel, E., Pleguezuelos-Manzano, C., Oka, R., Carrillo, L., Andersson-Rolf, A., Boretto, M., Brunsveld, J. E., van Boxtel, R., & Beekman, J. M. (2021). Evaluating CRISPR-based prime editing for cancer modeling and CFTR repair in organoids. *Life science alliance*, *4*(10).

Girons, I. S., Bourhy, P., Ottone, C., Picardeau, M., Yelton, D., Hendrix, R. W., Glaser, P., & Charon, N. (2000). The LE1 bacteriophage replicates as a plasmid within Leptospira biflexa: construction of an L. biflexa-Escherichia coli shuttle vector. *Journal of bacteriology*, *182*(20), 5700-5705.

Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., & Randau, L. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA biology*, *16*(4), 504-517.

Goeders, N., & Van Melderen, L. (2014). Toxin-Antitoxin Systems as Multilevel Interaction Systems. *Toxins*, *6*(1), 304-324. https://www.mdpi.com/2072-6651/6/1/304

Gong, B., Shin, M., Sun, J., Jung, C.-H., Bolt, E. L., van der Oost, J., & Kim, J.-S. (2014). Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proceedings of the National Academy of Sciences*, *111*(46), 16359-16364.

Gouet, P., Robert, X., & Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic acids research*, *31*(13), 3320-3323.

Granato, E. T., Smith, W. P., & Foster, K. R. (2023). Collective protection against the type VI secretion system in bacteria. *The ISME Journal*, 1-11.

Grissa, I., Vergnaud, G., & Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic acids research*, *35*(suppl_2), W52-W57.

Gu, Y., Li, H., Deep, A., Enustun, E., Zhang, D., & Corbett, K. Bacterial Shedu immune nucleases share a common enzymatic core regulated by diverse sensor domains. bioRxiv(2023). *Google Scholar*.

Gunderson, F. F., Mallama, C. A., Fairbairn, S. G., & Cianciotto, N. P. (2015). Nuclease activity of Legionella pneumophila Cas2 promotes intracellular infection of amoebal host cells. *Infection and immunity*, *83*(3), 1008-1018.

Haake, D. A., & Levett, P. N. (2015). Leptospirosis in humans. *Leptospira and leptospirosis*, 65-97.

Haft, D. H., Selengut, J., Mongodin, E. F., & Nelson, K. E. (2005). A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *PLoS Computational Biology*, *1*(6), e60-e60. https://doi.org/10.1371/journal.pcbi.0010060

Haft, D. H., Selengut, J., Mongodin, E. F., & Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology*, *1*(6), e60.

Han, D., & Krauss, G. (2009). Characterization of the endonuclease SSO2001 from Sulfolobus solfataricus P2. *FEBS letters*, *583*(4), 771-776.

Han, X., Chang, L., Chen, H., Zhao, J., Tian, F., Ross, R. P., Stanton, C., van Sinderen, D., Chen, W., & Yang, B. (2024). Harnessing the endogenous Type IC CRISPR-Cas system for genome editing in Bifidobacterium breve. *Applied and Environmental Microbiology*, e02074-02023.

Harding, M. M. (1999). The geometry of metal–ligand interactions relevant to proteins. *Acta Crystallographica Section D: Biological Crystallography*, *55*(8), 1432-1443.

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. (2020). A scoutRNA is required for some type V CRISPR-Cas systems. *Molecular cell*, *79*(3), 416-424. e415.

Hartskeerl, R., Collares-Pereira, M., & Ellis, W. (2011). Emergence, control and re-emerging leptospirosis: dynamics of infection in the changing world. *Clinical microbiology and infection*, *17*(4), 494-501.

Hatfull, G. F., & Hendrix, R. W. (2011). Bacteriophages and their genomes. *Current Opinion in Virology*, *1*(4), 298-303.

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., & Doudna, J. A. (2010). Sequence-and structure-specific RNA processing by a CRISPR endonuclease. *Science*, *329*(5997), 1355-1358.

Hayes, F., & Van Melderen, L. (2011). Toxins-antitoxins: diversity, evolution and function. *Critical reviews in biochemistry and molecular biology*, *46*(5), 386-408.

Hayes, R. P., Xiao, Y., Ding, F., Van Erp, P. B., Rajashankar, K., Bailey, S., Wiedenheft, B., & Ke, A. (2016). Structural basis for promiscuous PAM recognition in type I–E Cascade from E. coli. *Nature*, *530*(7591), 499-503.

He, Y., Wang, M., Liu, M., Huang, L., Liu, C., Zhang, X., Yi, H., Cheng, A., Zhu, D., & Yang, Q. (2018). Cas1 and Cas2 from the type II-C CRISPR-Cas system of Riemerella anatipestifer are required for spacer acquisition. *Frontiers in cellular and infection microbiology*, *8*, 195.

Hochstrasser, M. L., & Doudna, J. A. (2015). Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends in biochemical sciences*, *40*(1), 58-66.

Hochstrasser, M. L., Taylor, D. W., Bhat, P., Guegler, C. K., Sternberg, S. H., Nogales, E., & Doudna, J. A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proceedings of the National Academy of Sciences*, *111*(18), 6618-6623.

Hochstrasser, M. L., Taylor, D. W., Kornfeld, J. E., Nogales, E., & Doudna, J. A. (2016). DNA targeting by a minimal CRISPR RNA-guided cascade. *Molecular cell*, *63*(5), 840-851.

Holm, L. (2022). Dali server: structural unification of protein families. *Nucleic Acids Research*.

Horvath, P., Romero, D. A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., & Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *Journal of bacteriology*, *190*(4), 1401-1412.

Hu, C., Almendros, C., Nam, K. H., Costa, A. R., Vink, J. N., Haagsma, A. C., Bagde, S. R., Brouns, S. J., & Ke, A. (2021). Mechanism for Cas4-assisted directional spacer acquisition in CRISPR–Cas. *Nature*, *598*(7881), 515-520.

Huang, J., Zhou, Y., Li, J., Lu, A., & Liang, C. (2022). CRISPR/Cas systems: Delivery and application in gene therapy. *Frontiers in Bioengineering and Biotechnology*, *10*, 942325.

Hudaiberdiev, S., Shmakov, S., Wolf, Y. I., Terns, M. P., Makarova, K. S., & Koonin, E. V. (2017). Phylogenomics of Cas4 family nucleases. *BMC evolutionary biology*, *17*(1), 1-14.

Huiting, E., & Bondy-Denomy, J. (2023). Defining the expanding mechanisms of phage-mediated activation of bacterial immunity. *Current opinion in microbiology*, *74*, 102325.

Hullahalli, K., Rodrigues, M., Schmidt, B. D., Li, X., Bhardwaj, P., & Palmer, K. L. (2015). Comparative analysis of the orphan CRISPR2 locus in 242 Enterococcus faecalis strains. *PloS one*, *10*(9), e0138890.

Hussain, M. S., Anand, V., & Kumar, M. (2023). Functional PAM sequence for DNA interference by CRISPR-Cas IB system of Leptospira interrogans and the role of LinCas11b encoded within lincas8b. *International Journal of Biological Macromolecules*, *237*, 124086.

Hussain, M. S., & Kumar, M. (2022). Assembly of Cas7 subunits of Leptospira on the mature crRNA of CRISPR-Cas IB is modulated by divalent ions. *Gene*, *818*, 146244.

Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. *Nature reviews Molecular cell biology*, *9*(1), 22-32.

Isaev, A. B., Musharova, O. S., & Severinov, K. V. (2021). Microbial arsenal of antiviral defenses–part I. *Biochemistry (Moscow)*, *86*, 319-337.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology*, *169*(12), 5429-5433.

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., & Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA–guided surveillance complex from Escherichia coli. *Science*, *345*(6203), 1473-1479.

Jansen, R., Embden, J. D. v., Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology*, *43*(6), 1565-1575.

Jevtić, Ž., Stoll, B., Pfeiffer, F., Sharma, K., Urlaub, H., Marchfelder, A., & Lenz, C. (2019). The Response of Haloferax volcanii to Salt and Temperature Stress: A Proteome Study by Label-Free Mass Spectrometry. *Proteomics*, *19*(20), 1800491.

Jiang, F., & Doudna, J. A. (2017). CRISPR–Cas9 structures and mechanisms. *Annual review of biophysics*, *46*, 505-529.

Jin, S., Lin, Q., Luo, Y., Zhu, Z., Liu, G., Li, Y., Chen, K., Qiu, J.-L., & Gao, C. (2021). Genome-wide specificity of prime editors in plants. *Nature Biotechnology*, *39*(10), 1292-1299.

Jin, S., Zhan, J., & Zhou, Y. (2021). Argonaute proteins: structures and their endonuclease activity. *Molecular Biology Reports*, *48*(5), 4837-4849.

Joly, N., Engl, C., Jovanovic, G., Huvet, M., Toni, T., Sheng, X., Stumpf, M. P., & Buck, M. (2010). Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology. *FEMS microbiology reviews*, *34*(5), 797-827.

Jore, M. M., Lundgren, M., Van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, Ü., Wurm, R., & Wagner, R. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology*, *18*(5), 529-536.

Jorth, P., & Whiteley, M. (2012). An evolutionary link between natural transformation and CRISPR adaptive immunity. *MBio*, *3*(5), 10.1128/mbio. 00309-00312.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., & Potapenko, A. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589.

Jung, T. Y., Park, K. H., An, Y., Schulga, A., Deyev, S., Jung, J. H., & Woo, E. J. (2016). Structural features of Cas2 from Thermococcus onnurineus in CRISPR-cas system type IV. *Protein Science*, *25*(10), 1890-1897.

Ka, D., Hong, S., Jeong, U., Jeong, M., Suh, N., Suh, J.-Y., & Bae, E. (2017). Structural and dynamic insights into the role of conformational switching in the nuclease activity of the Xanthomonas albilineans Cas2 in CRISPR-mediated adaptive immunity. *Structural Dynamics*, *4*(5), 054701.

Ka, D., Kim, D., Baek, G., & Bae, E. (2014). Structural and functional characterization of Streptococcus pyogenes Cas2 protein under different pH conditions. *Biochemical and biophysical research communications*, *451*(1), 152-157.

Kaufmann, G., David, M., Borasio, G. D., Teichmann, A., Paz, A., Amitsur, M., Green, R., & Snyder, L. (1986). Phage and host genetic determinants of the specific anticodon loop

cleavages in bacteriophage T4-infected Escherichia coli CTr5X. *Journal of molecular biology*, *188*(1), 15-22.

Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., & Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, *357*(6351), 605-609.

Kessler, A. T., & Raja, A. (2019). Biochemistry, histidine.

Kieper, S. N., Almendros, C., Behler, J., McKenzie, R. E., Nobrega, F. L., Haagsma, A. C., Vink, J. N., Hess, W. R., & Brouns, S. J. (2018). Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. *Cell reports*, *22*(13), 3377-3384.

Kim, J.-S., Kim, J., Cepek, K. L., Sharp, P. A., & Pabo, C. O. (1997). Design of TATA box-binding protein/zinc finger fusions for targeted regulation of gene expression. *Proceedings of the National Academy of Sciences*, *94*(8), 3616-3620.

Kim, T.-Y., Shin, M., Yen, L. H. T., & Kim, J.-S. (2013). Crystal structure of Cas1 from Archaeoglobus fulgidus and characterization of its nucleolytic activity. *Biochemical and biophysical research communications*, *441*(4), 720-725.

Kisker, C., Kuper, J., & Van Houten, B. (2013). Prokaryotic nucleotide excision repair. *Cold Spring Harbor perspectives in biology*, *5*(3), a012591.

Ko, A. I., Goarant, C., & Picardeau, M. (2009). Leptospira: the dawn of the molecular genetics era for an emerging zoonotic pathogen. *Nature Reviews Microbiology*, *7*(10), 736-747.

Koo, Y., Ka, D., Kim, E.-J., Suh, N., & Bae, E. (2013). Conservation and variability in the structure and function of the Cas5d endoribonuclease in the CRISPR-mediated microbial immune system. *Journal of molecular biology*, *425*(20), 3799-3810.

Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current opinion in microbiology*, *37*, 67-78.

Korf, I., Yandell, M., & Bedell, J. (2003). *Blast*. " O'Reilly Media, Inc.".

Korona, R., Korona, B., & Levin, B. R. (1993). Sensitivity of naturally occurring coliphages to type I and type II restriction and modification. *Microbiology*, *139*(6), 1283-1290.

Korona, R., & Levin, B. R. (1993). Phage-mediated selection and the evolution and maintenance of restriction-modification. *Evolution*, *47*(2), 556-575.

Krupovic, M., Prangishvili, D., Hendrix, R. W., & Bamford, D. H. (2011). Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiology and molecular biology reviews*, *75*(4), 610-635.

Kumar, S. S. (2013). Indian guidelines for the diagnosis and management of human leptospirosis. *Medicine Update*, *23*.

Künne, T., Kieper, S. N., Bannenberg, J. W., Vogel, A. I., Miellet, W. R., Klein, M., Depken, M., Suarez-Diez, M., & Brouns, S. J. (2016). Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. *Molecular cell*, *63*(5), 852-864.

Kwon, A. R., Kim, J. H., Park, S. J., Lee, K. Y., Min, Y. H., Im, H., Lee, I., Lee, K. Y., & Lee, B.-J. (2012). Structural and biochemical characterization of HP0315 from Helicobacter pylori as a VapD protein with an endoribonuclease activity. *Nucleic acids research*, *40*(9), 4216-4228.

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, *8*(5), 317-327.

Łapińska, U., Voliotis, M., Lee, K. K., Campey, A., Stone, M. R. L., Tuck, B., Phetsang, W., Zhang, B., Tsaneva-Atanasova, K., & Blaskovich, M. A. (2022). Fast bacterial growth reduces antibiotic accumulation and efficacy. *Elife*, *11*, e74062.

Lee, H., Dhingra, Y., & Sashital, D. G. (2019). The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. *Elife*, *8*, e44248.

Leenay, R. T., & Beisel, C. L. (2017). Deciphering, communicating, and engineering the CRISPR PAM. *Journal of molecular biology*, *429*(2), 177-191.

Lemak, S., Beloglazova, N., Nocek, B., Skarina, T., Flick, R., Brown, G., Popovic, A., Joachimiak, A., Savchenko, A., & Yakunin, A. F. (2013). Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S] cluster containing Cas4 nuclease SSO0001 from Sulfolobus solfataricus. *Journal of the American Chemical Society*, *135*(46), 17476-17487.

Lemak, S., Nocek, B., Beloglazova, N., Skarina, T., Flick, R., Brown, G., Joachimiak, A., Savchenko, A., & Yakunin, A. F. (2014). The CRISPR-associated Cas4 protein Pcal_0546 from Pyrobaculum calidifontis contains a [2Fe-2S] cluster: crystal structure and nuclease activity. *Nucleic acids research*, *42*(17), 11144-11155.

Lemak, S., Serbanescu, M. A., Khusnutdinova, A. N., Ruszkowski, M., Beloglazova, N., Xu, X., Brown, G., Cui, H., Tan, K., & Joachimiak, A. (2021). Structural and biochemical insights into CRISPR RNA processing by the Cas5c ribonuclease SMU1763 from Streptococcus mutans. *Journal of Biological Chemistry*, *297*(5).

Li, C., Motaleb, M., Sal, M., Goldstein, S., & Charon, N. (2001). Gyrations, rotations, periplasmic flagella: the biology of spirochete motility. *The spirochetes. Molecular and cellular Biology. Horizon Press, Norfolk, NK*, 11-22.

Li, M., Liu, H., Han, J., Liu, J., Wang, R., Zhao, D., Zhou, J., & Xiang, H. (2013). Characterization of CRISPR RNA biogenesis and Cas6 cleavage-mediated inhibition of a provirus in the haloarchaeon Haloferax mediterranei. *Journal of bacteriology*, *195*(4), 867-875.

Li, M., Wang, R., Zhao, D., & Xiang, H. (2014). Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic acids research*, *42*(4), 2483-2492.

Lieber, M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry*, *79*(1), 181-211.

Lillestøl, R. K., Shah, S. A., Brügger, K., Redder, P., Phan, H., Christiansen, J., & Garrett, R. A. (2009). CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. *Molecular microbiology*, *72*(1), 259-272.

Liu, Q., Zhang, H., & Huang, X. (2020). Anti-CRISPR proteins targeting the CRISPR-Cas system enrich the toolkit for genetic engineering. *The FEBS Journal*, *287*(4), 626-644.

Liu, T., Li, Y., Wang, X., Ye, Q., Li, H., Liang, Y., She, Q., & Peng, N. (2015). Transcriptional regulator-mediated activation of adaptation genes triggers CRISPR de novo spacer acquisition. *Nucleic acids research*, *43*(2), 1044-1055.

Liu, T., Liu, Z., Ye, Q., Pan, S., Wang, X., Li, Y., Peng, W., Liang, Y., She, Q., & Peng, N. (2017). Coupling transcriptional activation of CRISPR–Cas system and DNA repair genes by Csa3a in Sulfolobus islandicus. *Nucleic acids research*, *45*(15), 8978-8992.

Liu, T., Xu, Y., Wang, X., Ye, Q., Liu, Z., Zhang, Z., Liu, J., Yang, Y., Peng, X., & Peng, N. (2022). DNA Motifs and an Accessory CRISPR Factor Determine Cas1 Binding and Integration Activity in Sulfolobus islandicus. *International Journal of Molecular Sciences*, *23*(17), 10178.

Liu, T. Y., & Doudna, J. A. (2020). Chemistry of Class 1 CRISPR-Cas effectors: Binding, editing, and regulation. *Journal of Biological Chemistry*, *295*(42), 14473-14487.

Liu, Y., Li, X., He, S., Huang, S., Li, C., Chen, Y., Liu, Z., Huang, X., & Wang, X. (2020). Efficient generation of mouse models with the prime editing system. *Cell discovery*, *6*(1), 27.

Liu, Z., Sun, M., Liu, J., Liu, T., Ye, Q., Li, Y., & Peng, N. (2020). A CRISPR-associated factor Csa3a regulates DNA damage repair in Crenarchaeon Sulfolobus islandicus. *Nucleic acids research*, *48*(17), 9681-9693.

Lopes, A. P., Lopes, L. M., Fraga, T. R., Chura-Chambi, R. M., Sanson, A. L., Cheng, E., Nakajima, E., Morganti, L., & Martins, E. A. (2014). VapC from the leptospiral VapBC toxin-antitoxin module displays ribonuclease activity on the initiator tRNA. *PloS one*, *9*(7), e101678.

Lories, B., Roberfroid, S., Dieltjens, L., De Coster, D., Foster, K. R., & Steenackers, H. P. (2020). Biofilm bacteria use stress responses to detect and respond to competitors. *Current Biology*, *30*(7), 1231-1244. e1234.

Louwen, R., Staals, R. H., Endtz, H. P., van Baarlen, P., & van der Oost, J. (2014). The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiology and Molecular Biology Reviews*, *78*(1), 74-88.

Luo, M. L., Mullis, A. S., Leenay, R. T., & Beisel, C. L. (2015). Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic acids research*, *43*(1), 674-681.

Maguire, M. E., & Cowan, J. A. (2002). Magnesium chemistry and biochemistry. *Biometals*, *15*(3), 203-210.

Maier, L.-K., Lange, S. J., Stoll, B., Haas, K. A., Fischer, S. M., Fischer, E., Duchardt-Ferner, E., Wöhnert, J., Backofen, R., & Marchfelder, A. (2013). Essential requirements for the detection and degradation of invaders by the Haloferax volcanii CRISPR/Cas system IB. *RNA biology*, *10*(5), 865-874.

Majumdar, S., Zhao, P., Pfister, N. T., Compton, M., Olson, S., Glover, C. V., Wells, L., Graveley, B. R., Terns, R. M., & Terns, M. P. (2015). Three CRISPR-Cas immune effector complexes coexist in Pyrococcus furiosus. *Rna*, *21*(6), 1147-1158.

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., & 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. *Biology direct*, *1*(1), 1-26.

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. (2011a). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, *9*(6), 467.

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. (2011b). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, *9*(6), 467-477.

Makarova, K. S., & Koonin, E. V. (2015). Annotation and classification of CRISPR-Cas systems. *CRISPR: methods and protocols*, 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. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, *13*(11), 722-736.

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

Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2013). Comparative genomics of defense systems in archaea and bacteria. *Nucleic acids research*, *41*(8), 4360-4377.

Makarova, K. S., Wolf, Y. I., Shmakov, S. A., Liu, Y., Li, M., & Koonin, E. V. (2020). Unprecedented Diversity of Unique CRISPR-Cas-Related Systems and Cas1 Homologs in Asgard Archaea. *The CRISPR Journal*, *3*(3), 156-163.

Makarova, K. S., Wolf, Y. I., van der Oost, J., & 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. *Biology direct*, *4*(1), 1-15.

Mandin, P., Repoila, F., Vergassola, M., Geissmann, T., & Cossart, P. (2007). Identification of new noncoding RNAs in Listeria monocytogenes and prediction of mRNA targets. *Nucleic acids research*, *35*(3), 962-974.

Marino, N. D., Pinilla-Redondo, R., Csörgő, B., & Bondy-Denomy, J. (2020). Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nature methods*, *17*(5), 471-479.

Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *science*, *322*(5909), 1843-1845.

Marraffini, L. A., & Sontheimer, E. J. (2010a). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Reviews Genetics*, *11*(3), 181-190.

Marraffini, L. A., & Sontheimer, E. J. (2010b). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, *463*(7280), 568-571.

Masuda, Y., Miyakawa, K., Nishimura, Y., & Ohtsubo, E. (1993). chpA and chpB, Escherichia coli chromosomal homologs of the pem locus responsible for stable maintenance of plasmid R100. *Journal of Bacteriology*, *175*(21), 6850-6856.

Matthews, L. A., & Simmons, L. A. (2014). Bacterial nonhomologous end joining requires teamwork. *Journal of bacteriology*, *196*(19), 3363-3365.

Maviza, T. P., Zarechenskaia, A. S., Burmistrova, N. R., Tchoub, A. S., Dontsova, O. A., Sergiev, P. V., & Osterman, I. A. (2022). RtcB2-PrfH operon protects E. coli ATCC25922 strain from colicin E3 toxin. *International Journal of Molecular Sciences*, *23*(12), 6453.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of applied crystallography*, *40*(4), 658-674.

Mendoza, B. J., & Trinh, C. T. (2018). In silico processing of the complete CRISPR-Cas spacer space for identification of PAM sequences. *Biotechnology Journal*, *13*(9), 1700595.

Millman, A., Melamed, S., Leavitt, A., Doron, S., Bernheim, A., Hör, J., Garb, J., Bechon, N., Brandis, A., & Lopatina, A. (2022). An expanded arsenal of immune systems that protect bacteria from phages. *Cell host & microbe*, *30*(11), 1556-1569. e1555.

Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, *155*(3), 733-740.

Mojica, F. J., Díez-Villaseñor, C. s., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, *60*, 174-182.

Mojica, F. J., Juez, G., & Rodriguez-Valera, F. (1993). Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Molecular microbiology*, *9*(3), 613-621.

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution*, *60*(2), 174-182. https://doi.org/10.1007/s00239-004-0046-3

Morshedzadeh, F., Ghanei, M., Lotfi, M., Ghasemi, M., Ahmadi, M., Najari-Hanjani, P., Sharif, S., Mozaffari-Jovin, S., Peymani, M., & Abbaszadegan, M. R. (2024). An update on the application of CRISPR technology in clinical practice. *Molecular Biotechnology*, *66*(2), 179-197.

Mulepati, S., & Bailey, S. (2011). Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). *Journal of Biological Chemistry*, *286*(36), 31896-31903.

Mulepati, S., & Bailey, S. (2013). In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *Journal of Biological Chemistry*, *288*(31), 22184-22192.

Mulepati, S., Héroux, A., & Bailey, S. (2014). Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. *Science*, *345*(6203), 1479-1484.

Murray, P. R., Rosenthal, K. S., & Pfaller, M. A. (2015). *Medical microbiology*. Elsevier Health Sciences.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 355-367.

Musso, D., & La Scola, B. (2013). Laboratory diagnosis of leptospirosis: a challenge. *Journal of Microbiology, Immunology and Infection*, *46*(4), 245-252.

Nam, K. H., Ding, F., Haitjema, C., Huang, Q., DeLisa, M. P., & Ke, A. (2012). Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *Journal of Biological Chemistry*, *287*(43), 35943-35952.

Nam, K. H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M. P., & Ke, A. (2012). Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype IC/Dvulg CRISPR-Cas system. *Structure*, *20*(9), 1574-1584.

Nascimento, A., Ko, A. I., Martins, E. A. L., Monteiro-Vitorello, C. B., Ho, P. L., Haake, D., Verjovski-Almeida, S., Hartskeerl, R., Marques, M. d. V., & Oliveira, M. C. d. (2004). Comparative genomics of two Leptospira interrogans serovars reveals novel insights into physiology and pathogenesis. *Journal of bacteriology*, *186*(7), 2164-2172.

Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L., & Doudna, J. A. (2016). CRISPR immunological memory requires a host factor for specificity. *Molecular cell*, *62*(6), 824-833.

Nunez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N., & Doudna, J. A. (2015). Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature*, *527*(7579), 535-538.

Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N., & Doudna, J. A. (2015). Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature*, *527*(7579), 535-538.

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014a). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature structural & molecular biology*, *21*(6), 528-534.

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014b). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature structural & molecular biology*, *21*(6), 528.

Nuñez, J. K., Lee, A. S., Engelman, A., & Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature*, *519*(7542), 193-198.

Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., & Sorek, R. (2018). DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nature microbiology*, *3*(1), 90-98.

Oliveira, N. M., Martinez-Garcia, E., Xavier, J., Durham, W. M., Kolter, R., Kim, W., & Foster, K. R. (2015). Correction: Biofilm formation as a response to ecological competition. *PLoS biology*, *13*(8), e1002232.

Osawa, T., Inanaga, H., Sato, C., & Numata, T. (2015). Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog. *Molecular cell*, *58*(3), 418-430.

Owen, J. A., Punt, J., Stranford, S. A., & Jones, P. P. (2013). *Kuby immunology* (Vol. 27). WH Freeman New York.

Page, R., & Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nature chemical biology*, *12*(4), 208-214.

Pappas, C. J., Benaroudj, N., & Picardeau, M. (2015). A replicative plasmid vector allows efficient complementation of pathogenic Leptospira strains. *Applied and Environmental Microbiology*, *81*(9), 3176-3181.

Park, S.-J., Jeong, T. Y., Shin, S. K., Yoon, D. E., Lim, S.-Y., Kim, S. P., Choi, J., Lee, H., Hong, J.-I., & Ahn, J. (2021). Targeted mutagenesis in mouse cells and embryos using an enhanced prime editor. *Genome biology*, *22*(1), 170.

Pawluk, A., Davidson, A. R., & Maxwell, K. L. (2018). Anti-CRISPR: discovery, mechanism and function. *Nature Reviews Microbiology*, *16*(1), 12-17.

Peng, W., Li, H., Hallstrøm, S., Peng, N., Liang, Y. X., & She, Q. (2013). Genetic determinants of PAM-dependent DNA targeting and pre-crRNA processing in Sulfolobus islandicus. *RNA biology*, *10*(5), 738-748.

Penner, M., Morad, I., Snyder, L., & Kaufmann, G. (1995). Phage T4-coded Stp: double-edged effector of coupled DNA and tRNA-restriction systems. *Journal of molecular biology*, *249*(5), 857-868.

Petakh, P., Oksenych, V., Khovpey, Y., & Kamyshnyi, O. (2024). Comprehensive Analysis of Antiphage Defense Mechanisms: Serovar-Specific Patterns. *Antibiotics*, *13*(6), 522.

Picardeau, M. (2008). Conjugative transfer between Escherichia coli and Leptospira spp. as a new genetic tool. *Applied and Environmental Microbiology*, *74*(1), 319-322.

Picardeau, M. (2013). Diagnosis and epidemiology of leptospirosis. *Médecine et maladies infectieuses*, *43*(1), 1-9.

Picardeau, M. (2017). Virulence of the zoonotic agent of leptospirosis: still terra incognita? *Nature Reviews Microbiology*, *15*(5), 297-307.

Picardeau, M. (2020). Leptospira and leptospirosis. *Leptospira spp. Methods and Protocols*, 271-275.

Picardeau, M., Bulach, D. M., Bouchier, C., Zuerner, R. L., Zidane, N., Wilson, P. J., Creno, S., Kuczek, E. S., Bommezzadri, S., & Davis, J. C. (2008). Genome sequence of the saprophyte Leptospira biflexa provides insights into the evolution of Leptospira and the pathogenesis of leptospirosis. *PloS one*, *3*(2), e1607.

Picardeau, M., Ren, S., & Saint Girons, I. (2001). Killing effect and antitoxic activity of the Leptospira interrogans toxin-antitoxin system in Escherichia coli. *Journal of Bacteriology*, *183*(21), 6494-6497.

Pingoud, V., Wende, W., Friedhoff, P., Reuter, M., Alves, J., Jeltsch, A., Mones, L., Fuxreiter, M., & Pingoud, A. (2009). On the divalent metal ion dependence of DNA cleavage by restriction endonucleases of the EcoRI family. *Journal of molecular biology*, *393*(1), 140-160.

Plagens, A., Richter, H., Charpentier, E., & Randau, L. (2015). DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS microbiology reviews*, *39*(3), 442-463.

Plagens, A., Tripp, V., Daume, M., Sharma, K., Klingl, A., Hrle, A., Conti, E., Urlaub, H., & Randau, L. (2014). In vitro assembly and activity of an archaeal CRISPR-Cas type IA Cascade interference complex. *Nucleic acids research*, *42*(8), 5125-5138.

Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K. A., Djordjevic, M., Wanner, B. L., & Severinov, K. (2010). Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Molecular microbiology*, *77*(6), 1367-1379.

Pourcel, C., Salvignol, G., & 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*(3), 653-663.

Pourcel, C., Touchon, M., Villeriot, N., Vernadet, J.-P., Couvin, D., Toffano-Nioche, C., & Vergnaud, G. (2020). CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. *Nucleic acids research*, *48*(D1), D535-D544.

Prakash, A., & Kumar, M. (2021). Characterizing the transcripts of Leptospira CRISPR IB array and its processing with endoribonuclease LinCas6. *International Journal of Biological Macromolecules*, *182*, 785-795.

Prakash, A., & Kumar, M. (2022). Transcriptional analysis of CRISPR IB arrays of Leptospira interrogans serovar Lai and its processing by Cas6. *Frontiers in Microbiology*, *13*.

Przybilski, R., Richter, C., Gristwood, T., Clulow, J. S., Vercoe, R. B., & Fineran, P. C. (2011). Csy4 is responsible for CRISPR RNA processing in Pectobacterium atrosepticum. *RNA biology*, *8*(3), 517-528.

Punetha, A., Sivathanu, R., & Anand, B. (2014). Active site plasticity enables metal-dependent tuning of Cas5d nuclease activity in CRISPR-Cas type IC system. *Nucleic acids research*, *42*(6), 3846-3856.

Raines, R. T. (1998). Ribonuclease a. *Chemical reviews*, *98*(3), 1045-1066.

Ramalingam, S., Annaluru, N., Kandavelou, K., & Chandrasegaran, S. (2014). TALEN-mediated generation and genetic correction of disease-specific human induced pluripotent stem cells. *Current gene therapy*, *14*(6), 461-472.

Reeks, J., Naismith, J. H., & White, M. F. (2013). CRISPR interference: a structural perspective. *Biochemical Journal*, *453*(2), 155-166.

Reeks, J., Sokolowski, R. D., Graham, S., Liu, H., Naismith, J. H., & White, M. F. (2013). Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing. *Biochemical Journal*, *452*(2), 223-230.

Reis, R. B., Ribeiro, G. S., Felzemburgh, R. D., Santana, F. S., Mohr, S., Melendez, A. X., Queiroz, A., Santos, A. C., Ravines, R. R., & Tassinari, W. S. (2008). Impact of environment and social gradient on Leptospira infection in urban slums. *PLoS neglected tropical diseases*, *2*(4), e228.

Ribeiro, A. J., Tyzack, J. D., Borkakoti, N., Holliday, G. L., & Thornton, J. M. (2020). A global analysis of function and conservation of catalytic residues in enzymes. *Journal of Biological Chemistry*, *295*(2), 314-324.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends in genetics*, *16*(6), 276-277.

Richter, H., Lange, S. J., Backofen, R., & Randau, L. (2013). Comparative analysis of Cas6b processing and CRISPR RNA stability. *RNA biology*, *10*(5), 700-707.

Riley, L. W., Ko, A. I., Unger, A., & Reis, M. G. (2007). Slum health: diseases of neglected populations. *BMC international health and human rights*, *7*, 1-6.

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research*, *42*(W1), 320-324. https://doi.org/10.1093/nar/gku316

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S. K., Dryden, D. T., & Dybvig, K. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic acids research*, *31*(7), 1805-1812.

Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L., & White, M. F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife*, *4*, e08716.

Ross, B. D., Verster, A. J., Radey, M. C., Schmidtke, D. T., Pope, C. E., Hoffman, L. R., Hajjar, A. M., Peterson, S. B., Borenstein, E., & Mougous, J. D. (2019). Human gut bacteria contain acquired interbacterial defence systems. *Nature*, *575*(7781), 224-228.

Rostøl, J. T., & Marraffini, L. (2019). (Ph) ighting phages: how bacteria resist their parasites. *Cell host & microbe*, *25*(2), 184-194.

Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C. V., Spagnolo, L., & White, M. F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Molecular cell*, *52*(1), 124-134.

Rousseau, C., Gonnet, M., Le Romancer, M., & Nicolas, J. (2009). CRISPI: a CRISPR interactive database. *Bioinformatics*, *25*(24), 3317-3318.

Safari, F., Sharifi, M., Farajnia, S., Akbari, B., Karimi Baba Ahmadi, M., Negahdaripour, M., & Ghasemi, Y. (2020). The interaction of phages and bacteria: the co-evolutionary arms race. *Critical reviews in biotechnology*, *40*(2), 119-137.

Safari, F., Zare, K., Negahdaripour, M., Barekati-Mowahed, M., & Ghasemi, Y. (2019). CRISPR Cpf1 proteins: structure, function and implications for genome editing. *Cell & bioscience*, *9*, 1-21.

Saier Jr, M. H., Paulsen, I. T., Marek, K. s., Pao, S. S., Ronald, A. s., & Nikaido, H. (1998). Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. *The FASEB Journal*, *12*(3), 265-274.

Saint Girons, I., Margarita, D., Amouriaux, P., & Baranton, G. (1990). First isolation of bacteriophages for a spirochaete: potential genetic tools for Leptospira. *Research in microbiology*, *141*(9), 1131-1138.

Samai, P., Smith, P., & Shuman, S. (2010). Structure of a CRISPR-associated protein Cas2 from Desulfovibrio vulgaris. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *66*(12), 1552-1556.

Samson, J. E., Magadán, A. H., Sabri, M., & Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nature Reviews Microbiology*, *11*(10), 675-687.

Santos, C. S., Macedo, J. O., Bandeira, M., Chagas-Junior, A. D., McBride, A. J., McBride, F. W., Reis, M. G., & Athanazio, D. A. (2010). Different outcomes of experimental leptospiral infection in mouse strains with distinct genotypes. *Journal of medical microbiology*, *59*(9), 1101-1106.

Sashital, D. G., Jinek, M., & Doudna, J. A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nature structural & molecular biology*, *18*(6), 680-687.

Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., Van Der Oost, J., Brouns, S. J., & Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences*, *108*(25), 10098-10103.

Seto, A., Shirouzu, M., Terada, T., Murayama, K., Kuramitsu, S., & Yokoyama, S. (2003). Crystal structure of a hypothetical protein, TT1725, from Thermus thermophilus HB8 at 1.7 Å resolution. *Proteins: Structure, Function, and Bioinformatics*, *53*(3), 768-771.

Shah, S. A., Erdmann, S., Mojica, F. J., & Garrett, R. A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. *RNA biology*, *10*(5), 891-899.

Shah, S. A., & Garrett, R. A. (2011). CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. *Research in Microbiology*, *162*(1), 27-38.

Shaibullah, S., Shuhaimi, N., Ker, D.-S., Mohd-Sharif, N., Ho, K. L., Teh, A.-H., Waterman, J., Tang, T.-H., Wong, R.-R., & Nathan, S. (2023). Structural and functional analyses of Burkholderia pseudomallei BPSL1038 reveal a Cas-2/VapD nuclease sub-family. *Communications biology*, *6*(1), 920.

Shapiro, R. S., Chavez, A., & Collins, J. J. (2018). CRISPR-based genomic tools for the manipulation of genetically intractable microorganisms. *Nature Reviews Microbiology*, *16*(6), 333-339.

Sharma, M. (2008). Leptospirosis: epidemiology, diagnosis, and control. *J Infect. Dis. Antimicrob. Agents.*, *25*, 93-103.

Shiimori, M., Garrett, S. C., Graveley, B. R., & Terns, M. P. (2018). Cas4 nucleases define the PAM, length, and orientation of DNA fragments integrated at CRISPR loci. *Molecular cell*, *70*(5), 814-824. e816.

Shmakov, S. A., Utkina, I., Wolf, Y. I., Makarova, K. S., Severinov, K. V., & Koonin, E. V. (2020). CRISPR arrays away from cas genes. *The CRISPR Journal*, *3*(6), 535-549.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., & Söding, J. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, *7*(1), 539.

Sinkunas, T., Gasiunas, G., Waghmare, S. P., Dickman, M. J., Barrangou, R., Horvath, P., & Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *The EMBO journal*, *32*(3), 385-394.

Sirotkin, K., Cooley, W., Runnels, J., & Snyder, L. (1978). A role in true-late gene expression for the T4 bacteriophage 5′ polynucleotide kinase 3′ phosphatase. *Journal of molecular biology*, *123*(2), 221-233.

Smith, W. P., Wucher, B. R., Nadell, C. D., & Foster, K. R. (2023). Bacterial defences: mechanisms, evolution and antimicrobial resistance. *Nature Reviews Microbiology*, 1-16.

Sokolowski, R. D., Graham, S., & White, M. F. (2014). Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system. *Nucleic acids research*, *42*(10), 6532-6541.

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Biophysical Journal*, *106*(2), 695a.

Sternberg, S. H., Richter, H., Charpentier, E., & Qimron, U. (2016). Adaptation in CRISPR-Cas systems. *Molecular cell*, *61*(6), 797-808.

Stokar-Avihail, A., Fedorenko, T., Hör, J., Garb, J., Leavitt, A., Millman, A., Shulman, G., Wojtania, N., Melamed, S., & Amitai, G. (2023). Discovery of phage determinants that confer sensitivity to bacterial immune systems. *Cell*, *186*(9), 1863-1876. e1816.

Swarts, D. C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R. F., Koonin, E. V., Patel, D. J., & Van Der Oost, J. (2014). The evolutionary journey of Argonaute proteins. *Nature structural & molecular biology*, *21*(9), 743-753.

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, *38*(7), 3022-3027.

Tang, D., Li, H., Wu, C., Jia, T., He, H., Yao, S., Yu, Y., & Chen, Q. (2021). A distinct structure of Cas1–Cas2 complex provides insights into the mechanism for the longer spacer acquisition in Pyrococcus furiosus. *International Journal of Biological Macromolecules*, *183*, 379-386.

Taylor, D. W., Zhu, Y., Staals, R. H., Kornfeld, J. E., Shinkai, A., van der Oost, J., Nogales, E., & Doudna, J. A. (2015). Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science*, *348*(6234), 581-585.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*.

Tock, M. R., & Dryden, D. T. (2005). The biology of restriction and anti-restriction. *Current opinion in microbiology*, *8*(4), 466-472.

Tong, Y., Jørgensen, T. S., Whitford, C. M., Weber, T., & Lee, S. Y. (2021). A versatile genetic engineering toolkit for E. coli based on CRISPR-prime editing. *Nature communications*, *12*(1), 5206.

Touchon, M., & Rocha, E. P. (2010). The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella. *PloS one*, *5*(6), e11126.

Tuszynska, I., Magnus, M., Jonak, K., Dawson, W., & Bujnicki, J. M. (2015). NPDock: a web server for protein–nucleic acid docking. *Nucleic acids research*, *43*(W1), W425-W430.

van Beljouw, S. P., Sanders, J., Rodríguez-Molina, A., & Brouns, S. J. (2023). RNA-targeting CRISPR–Cas systems. *Nature Reviews Microbiology*, *21*(1), 21-34.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., & Laydon, A. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, *50*(D1), D439-D444.

Vassallo, C. N., Doering, C. R., Littlehale, M. L., Teodoro, G. I., & Laub, M. T. (2022). A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome. *Nature microbiology*, *7*(10), 1568-1579.

Vasu, K., Nagamalleswari, E., & Nagaraja, V. (2012). Promiscuous restriction is a cellular defense strategy that confers fi tness advantage to bacteria. https://doi.org/10.1073/pnas.1119226109

Velankar, S., Alhroub, Y., Alili, A., Best, C., Boutselakis, H. C., Caboche, S., Conroy, M. J., Dana, J. M., Van Ginkel, G., & Golovin, A. (2010). PDBe: protein data bank in Europe. *Nucleic acids research*, *39*(suppl_1), D402-D410.

Velankar, S., Best, C., Beuth, B., Boutselakis, C., Cobley, N., Sousa Da Silva, A., Dimitropoulos, D., Golovin, A., Hirshberg, M., & John, M. (2010). PDBe: protein data bank in Europe. *Nucleic acids research*, *38*(suppl_1), D308-D317.

Victoriano, A. F. B., Smythe, L. D., Gloriani-Barzaga, N., Cavinta, L. L., Kasai, T., Limpakarnjanarat, K., Ong, B. L., Gongal, G., Hall, J., & Coulombe, C. A. (2009). Leptospirosis in the Asia Pacific region. *BMC infectious diseases*, *9*(1), 1-9.

Vidakovic, L., Singh, P. K., Hartmann, R., Nadell, C. D., & Drescher, K. (2018). Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. *Nature microbiology*, *3*(1), 26-31.

Vincent, A. T., Schiettekatte, O., Goarant, C., Neela, V. K., Bernet, E., Thibeaux, R., Ismail, N., Mohd Khalid, M. K. N., Amran, F., & Masuzawa, T. (2019). Revisiting the taxonomy and evolution of pathogenicity of the genus Leptospira through the prism of genomics. *PLoS neglected tropical diseases*, *13*(5), e0007270.

Vinetz, J. M. (2000). Ten common questions about leptospirosis. *Infectious Diseases in Clinical Practice*, *9*(2), 59-65.

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., & Wang, Y. (2015). Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell*, *163*(4), 840-853.

Wang, L., Yu, X., Li, M., Sun, G., Zou, L., Li, T., Hou, L., Guo, Y., Shen, D., & Qu, D. (2019). Filamentation initiated by Cas2 and its association with the acquisition process in cells. *International journal of oral science*, *11*(3), 1-7.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., & Patel, D. J. (2008). Structure of the guide-strand-containing argonaute silencing complex. *Nature*, *456*(7219), 209-213.

Webber, M., & Piddock, L. (2003). The importance of efflux pumps in bacterial antibiotic resistance. *Journal of antimicrobial chemotherapy*, *51*(1), 9-11.

Wei, Y., Chesne, M. T., Terns, R. M., & Terns, M. P. (2015). Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. *Nucleic acids research*, *43*(3), 1749-1758.

Wein, T., & Sorek, R. (2022). Bacterial origins of human cell-autonomous innate immune mechanisms. *Nature Reviews Immunology*, *22*(10), 629-638.

Westra, E. R., van Erp, P. B., Künne, T., Wong, S. P., Staals, R. H., Seegers, C. L., Bollen, S., Jore, M. M., Semenova, E., & Severinov, K. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular cell*, *46*(5), 595-605.

Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., Westphal, W., Heck, A. J., Boekema, E. J., & Dickman, M. J. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings of the National Academy of Sciences*, *108*(25), 10092-10097.

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W., & Doudna, J. A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*, *17*(6), 904-912.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., & McCoy, A. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 235-242.

Worrall, J. A., & Luisi, B. F. (2007). Information available at cut rates: structure and mechanism of ribonucleases. *Current opinion in structural biology*, *17*(1), 128-137.

Wörtz, J., Smith, V., Fallmann, J., König, S., Thuraisingam, T., Walther, P., Urlaub, H., Stadler, P. F., Allers, T., & Hille, F. (2022). Cas1 and Fen1 Display Equivalent Functions During Archaeal DNA Repair. *Frontiers in microbiology*, *13*.

Wozniak, K. J., & Simmons, L. A. (2022). Bacterial DNA excision repair pathways. *Nature Reviews Microbiology*, *20*(8), 465-477.

Wu, C., Tang, D., Cheng, J., Hu, D., Yang, Z., Ma, X., He, H., Yao, S., Fu, T.-M., & Yu, Y. (2021). Mechanisms of spacer acquisition by sequential assembly of the adaptation module in Synechocystis. *Nucleic acids research*, *49*(5), 2973-2984.

Xiao, G., Yi, Y., Che, R., Zhang, Q., Imran, M., Khan, A., Yan, J., & Lin, X. a. (2019). Characterization of CRISPR-Cas systems in Leptospira reveals potential application of CRISPR in genotyping of Leptospira interrogans. *Apmis*, *127*(4), 202-216.

Xiao, Y., Ng, S., Nam, K. H., & Ke, A. (2017). How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature*, *550*(7674), 137-141.

Xu, D., & Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical journal*, *101*(10), 2525-2534.

Xu, T., Yao, F., Zhou, X., Deng, Z., & You, D. (2010). A novel host-specific restriction system associated with DNA backbone S-modification in Salmonella. *Nucleic acids research*, *38*(20), 7133-7141.

Xu, Y., Zhu, Y., Wang, Y., Chang, Y.-F., Zhang, Y., Jiang, X., Zhuang, X., Zhu, Y., Zhang, J., & Zeng, L. (2016). Whole genome sequencing revealed host adaptation-focused genomic plasticity of pathogenic Leptospira. *Scientific reports*, *6*(1), 20020.

Yang, H., Gao, P., Rajashankar, K. R., & Patel, D. J. (2016). PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease. *Cell*, *167*(7), 1814-1828. e1812.

Yang, J., & Zhang, Y. (2015). Protein structure and function prediction using I-TASSER. *Current protocols in bioinformatics*, *52*(1), 5.8. 1-5.8. 15.

Yang, W. (2011a). Nucleases: diversity of structure, function and mechanism. *Quarterly reviews of biophysics*, *44*(1), 1.

Yang, W. (2011b). Nucleases: diversity of structure, function and mechanism. *Quarterly reviews of biophysics*, *44*(1), 1-93.

Yang, W., Lee, J. Y., & Nowotny, M. (2006). Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. *Molecular cell*, *22*(1), 5-13.

Yosef, I., Goren, M. G., & Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic acids research*, *40*(12), 5569-5576.

Yu, L., & Marchisio, M. A. (2020). Types I and V Anti-CRISPR proteins: from phage defense to eukaryotic synthetic gene circuits. *Frontiers in Bioengineering and Biotechnology*, *8*, 575393.

Yuan, Y.-R., Pei, Y., Ma, J.-B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.-Y., Dauter, Z., Tuschl, T., & Patel, D. J. (2005). Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Molecular cell*, *19*(3), 405-419.

Zahedipour, F., Zahedipour, F., Zamani, P., Jaafari, M. R., & Sahebkar, A. (2024). Harnessing CRISPR technology for viral therapeutics and vaccines: from preclinical studies to clinical applications. *Virus Research*, *341*, 199314.

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. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*(3), 759-771.

Zhang, J., Kasciukovic, T., & White, M. F. (2012). The CRISPR associated protein Cas4 Is a 5′ to 3′ DNA exonuclease with an iron-sulfur cluster.

Zhang, Q., & Ye, Y. (2017). Not all predicted CRISPR–Cas systems are equal: isolated cas genes and classes of CRISPR like elements. *BMC bioinformatics*, *18*, 1-12.

Zhang, Y., & Inouye, M. (2009). The inhibitory mechanism of protein synthesis by YoeB, an Escherichia coli toxin. *Journal of Biological Chemistry*, *284*(11), 6627-6638.

Zhang, Y., Zhu, L., Zhang, J., & Inouye, M. (2005). Characterization of ChpBK, an mRNA interferase from Escherichia coli. *Journal of Biological Chemistry*, *280*(28), 26080-26088.

Zhang, Y. X., Guo, X. K., Wu, C., Bi, B., Ren, S. X., Wu, C. F., & Zhao, G. P. (2004). Characterization of a novel toxin-antitoxin module, VapBC, encoded by Leptospira interrogans chromosome. *Cell research*, *14*(3), 208-216.

Zhao, C., Shu, X., & Sun, B. (2017). Construction of a gene knockdown system based on catalytically inactive ("dead") Cas9 (dCas9) in Staphylococcus aureus. *Applied and Environmental Microbiology*, *83*(12), e00291-00217.

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., & Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. *Nature*, *515*(7525), 147-150.

Zhao, Z., Shang, P., Mohanraju, P., & Geijsen, N. (2023). Prime editing: advances and therapeutic applications. *Trends in Biotechnology*, *41*(8), 1000-1012.

Zheng, X., Li, S.-Y., Zhao, G.-P., & Wang, J. (2017). An efficient system for deletion of large DNA fragments in Escherichia coli via introduction of both Cas9 and the non-homologous end joining system from Mycobacterium smegmatis. *Biochemical and biophysical research communications*, *485*(4), 768-774.

Zheng, Y., Li, J., Wang, B., Han, J., Hao, Y., Wang, S., Ma, X., Yang, S., Ma, L., & Yi, L. (2020). Endogenous type I CRISPR-Cas: from foreign DNA defense to prokaryotic engineering. *Frontiers in bioengineering and biotechnology*, *8*, 62.

List of publications

Publications from thesis work

- **Anand, V.,** Prabhakaran, H. S., Gogoi, P., Kanaujia, S. P., & Kumar, M. (2022). Structural and functional characterization of Cas2 of CRISPR-Cas subtype IC lacking the CRISPR component. Frontiers in Molecular Biosciences, 9, 988569.
  PDB ID.: 7F84
- **Anand, V.,** Prabhakaran, H. S., Prakash, A., Hussain, M. S., & Kumar, M. (2023). Differential processing of CRISPR RNA by LinCas5c and LinCas6 of *Leptospira*. Biochimica et Biophysica Acta (BBA)-General Subjects, 1867(12), 130469.
- Hussain, M. S., **Anand, V**., & Kumar, M. (2023). Functional PAM sequence for DNA interference by CRISPR-Cas IB system of *Leptospira interrogans* and the role of LinCas11b encoded within lincas8b. International Journal of Biological Macromolecules, 237, 124086. **(PAM prediction and protospacer construction)**

Publication from collaborative work

- Dixit, B., **Anand, V**., Hussain, M. S., & Kumar, M. (2021). The CRISPR-associated Cas4 protein from *Leptospira interrogans* demonstrate versatile nuclease activity. Current Research in Microbial Sciences, 2, 100040.

Conferences and workshop

- **Vineet Anand** and Manish Kumar "Biochemical characterization of a core Cas protein of CRISPR-Cas subtype-IC in *Leptospira interrogans* serovar Copenhageni strain Fiocruz L1-130", Challenges and Threats of Microbes to Animals and Humans. (2020) **(Poster presentation)**.
- International Seminar cum Workshop on Computer Aided Drug Design for Human Pathogens. Department of Molecular Biology & Biotechnology Tezpur University, Assam, India.
- Manish Kumar, Md Saddam Hussain, Aman Prakash, **Vineet Anand** "LinCas11, a small subunit of interference complex in *Leptospira interrogans* CRISPR-Cas I-B locus, is a pre-requisite for the targeted annihilation of DNA in a surrogate host", 12th

TH-3470_166106104

International Leptospirosis Society Conference 2022, Nov 13-16, Bangkok, Thailand. (2022) **(Oral presentation by thesis supervisor).**

****