Functional Elucidation of CRISPR-Cas I-B Interference Machinery of *Leptospira interrogans*

Thesis Submitted

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

by

MD. SADDAM HUSSAIN

Under the Supervision of

Prof. Manish Kumar

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

August 2023

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
Department of Biosciences and Bioengineering
Guwahati-781039, Assam, India

Declaration

I hereby declare that the research work embodied in this thesis entitled "**Functional elucidation of CRISPR-Cas I-B interference machinery of *Leptospira interrogans***" is the result of the scientific investigation carried out by me under the supervision of Prof. Manish Kumar, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India. And, in keeping with the general practice of reporting scientific observations, due acknowledgment has been made wherever the findings of other investigators have been cited in the thesis. The work illustrated in this thesis has not been submitted elsewhere for any other degree or diploma award.

Signature: 28/08/2023

Md. Saddam Hussain
(Roll No. 166106005)
Department of Biosciences and Bioengineering,
Indian Institute of Technology Guwahati,
Guwahati-781039, Assam, India

Date: 28/08/2023

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
Department of Biosciences and Bioengineering
Guwahati-781039, Assam, India

Certificate

This is to certify the work described in this thesis entitled "**Functional elucidation of CRISPR-Cas I-B interference machinery of *Leptospira interrogans***" by Md. Saddam Hussain (Roll no. 166106005) for the award of the degree of Doctor of Philosophy is a record of an original research work carried out under my supervision at the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India. This work has not been submitted elsewhere for any other degree or diploma award.

Signature: Manish Kumar

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

Date: 28/08/2023

I dedicate this thesis to
my parents and the entire family

Acknowledgments

The accomplishment of the present research work has been made possible with the blessings of the almighty and the generous support of many individuals. Hence, I would like to express my sincere gratitude to all of them for their esteemed contributions.

First and foremost, I am extremely grateful and indebted to my Ph.D. thesis supervisor Prof. Manish Kumar for his enlightening guidance, continuous support, and patience throughout the duration of the present work. I am profoundly thankful to him for always being available for scientific discussions and eagerly addressing any challenges I encountered during this research work. His exceptionally organized way of working has been a constant inspiration, and I aspire to continue seeking his mentorships throughout my lifelong pursuits.

I would also extend my gratitude to the members of my doctoral committee, Prof. Pranab Goswami (Chairperson), Dr. Rajkumar P. Thummer, and Dr. Sunanda Chatterjee, for their timely and insightful suggestions during the course of the present work.

I sincerely acknowledge Prof. Bishwajit Kundu (National examiner), IIT Delhi, India, and Dr. Piruthivi Sukumar (Foreign examiner), Associate professor, University of Leeds, UK, for their comprehensive evaluation of my doctoral thesis.

I am also thankful to Dr. Rajeev Kumar Sharma, Dept. of Veterinary Microbiology, College of Veterinary Science, Guwahati, for his generous help with the animal house facility to generate polyclonal antibodies. I also thank Prof. Nitin Chaudhary for assisting with the fluorescence spectroscopy instrument.

I would also like to extend my acknowledgment to Prof. Kannan Pakshirajan, Prof. Latha Rangan, and Prof. Rakhi Chaturvedi, Department of Biosciences and Bioengineering (BSBE), IIT Guwahati, for lending their support as respectable Heads of the Department during my Ph. D. tenure.

I also thank the Department of BSBE and Central instrumentation facility (CIF), IIT Guwahati, for providing the infrastructural facility for the research work. My thanks are also due to the technical and non-technical staff of the Department of BSBE and CIF, IITG, for their kind cooperation during my entire Ph.D. duration.

28/06/2023

I sincerely acknowledge the Ministry of Human Resource and Development (MHRD), Government of India, for providing me with the research fellowship and the Department of Biotechnology (DBT), Department of Science and Technology (DST), and Indian Council of Medical Research (ICMR) for providing research grants to our laboratory.

I owe my heartfelt gratitude to all my senior lab colleagues, Dr. Karukriti Kaushik Ghosh, Dr. Bhuvan Dixit, Dr. Anusua Dhara, and Dr. Aman Prakash, for their generous support, constructive discussions, feedback, and the lighter moments that we shared together. I also thank my juniors in the lab, including Saswat Hota, Juhi Rajesh Rathod, Surbhi Kumari, Pratibha Koundal, Shibam Dey, and Prattusha Khan, for their active cooperation.

I extend my heartfelt appreciation to my wonderful friends Dr. Ankan Sinha, Dr. Uttariya Pal, Yashavanth P R, Dr. Rajib Shome, Dr. Manish K Gupta, Gaurav Bhatt, Riddhi Banerjee, who have been great companions throughout the highs and lows of my journey. The memories we shared, the laughter we indulged in, and the friendship we forged during our Ph.D. tenure will forever remain etched in my heart. I would also like to thank all my near & dears around the globe for encouraging me and sharing my happy and sorrowful moments at all times, with a special mention to Dr. Neelima Varshney.

Words fall short of expressing the depth of my gratitude towards my Mother, Father, Brothers, and the entire family. Their constant encouragement and unwavering support have been the driving force behind my accomplishments. Their boundless love and strength have been the pillars that have propelled me to achieve this significant milestone in my life.

28/08/2022

Md. Saddam Hussain

*****

Table of contents

List of abbreviations

Abbreviation	:	Full-form
Abi	:	Abortive infection
Agos	:	Argonautes
Amp	:	Ampicillin
BREX	:	Bacteriophage exclusion
bp	:	Base pair
BSA	:	Bovine serum albumin
CRISPR-Cas	:	Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins
crRNA	:	CRISPR RNA
Cascade	:	CRISPR-associated complex for antiviral defense
CARF	:	CRISPR-associated Rossmann fold
Csm	:	Cas subtype Mtube
Cmr	:	Cas module RAMP
cOA	:	Cyclic oligoadenylate
Chi site	:	Crossover hotspot instigator site
CRISPRi	:	CRISPR interference
CD	:	Circular dichroism
CEX	:	Cation exchange chromatography
CTD	:	C-terminal domain
Cam	:	Chloramphenicol
CFU	:	Colony-forming unit
cDNA	:	Complementary DNA
DNA	:	Deoxyribonucleic acid
DNase	:	Deoxyribonuclease
DISARM	:	Defense island associated with restriction-modification
DSB	:	Double-stranded break
DALYs	:	Disability Adjusted Life Years
ds-DNA	:	Double-stranded DNA

DTT	:	Dithiothreitol
eAgos	:	Eukaryotic Agos
ERP	:	Excision-replication-packaging
EMJH	:	Ellinghausen-McCullough-Johnson-Harris
endoDNase	:	Endodeoxyribonuclease
EDTA	:	Ethylenediaminetetraacetic acid
EtBr	:	Ethidium bromide
ELISA	:	Enzyme-linked immunosorbent assay
FAM	:	Fluorescein amidite
FETEM	:	Field emission transmission electron microscopy
FCA	:	Freund's complete adjuvant
FIA	:	Freund's incomplete adjuvant
GST	:	Glutathione S-transferase
GSH	:	Glutathione
HEPN domain	:	Higher eukaryotes and prokaryotes nucleotide-binding domain
HD domain	:	Histidine (H)-aspartate (D) domain
HNH domain	:	Histidine (H)-asparagine (N)-histidine (H) domain
HDR	:	Homology-directed repair
h	:	Hour
HRP	:	Horseradish peroxidase
IHF	:	Integration host factor
IPTG	:	Isopropyl-β-D-thiogalactopyranoside
ICMR	:	Indian Council of Medical Research
ITC	:	Isothermal titration calorimetry
Kan	:	Kanamycin
kb	:	Kilobase
kDa	:	Kilodalton
LS	:	Large subunit of Cascade
LAS	:	Leader anchoring site
LB medium	:	Luria Bertani medium
MGE	:	Mobile genetic element
MTase	:	Methyltransferase

MID domain	:	Middle domain of prokaryotic argonautes
MCS	:	Multiple cloning site
NUC	:	Nuclease lobe in Cas proteins
NHEJ	:	Non-homologous end-joining
NCBI	:	National Center for Biotechnology Information
Ni-NTA	:	Nickel-nitrilotriacetic acid
NTD	:	N-terminal domain
OD_{600}	:	Optical density at 600 nm
ORF	:	Open reading frame
PAM	:	Protospacer adjacent motif
pAgos	:	Prokaryotic Agos
PICI	:	Phage-inducible chromosomal island
PIWI-RE	:	P-element induced wimpy testis with conserved arginine (R) and Glutamate (E) residues
PAZ	:	PIWI-Argonaute-Zwille
pre-crRNA	:	Precursor crRNA
PFS	:	Protospacer flanking sequence
PDB	:	Protein data bank
PCR	:	Polymerase chain reaction
PMSF	:	Phenylmethylsulfonyl fluoride
PBS	:	Phosphate-buffered saline
qRT-PCR	:	Quantitative reverse transcription-PCR
RNA	:	Ribonucleic acid
R-M system	:	Restriction-modification system
REase	:	Restriction endonuclease
RNAi	:	RNA interference
RISC	:	RNA-induced silencing complex
RNase	:	Ribonuclease
RRM	:	RNA recognition motif
RAMP	:	Repeat-associated mysterious protein
REC	:	Recognition lobe of Cas proteins
RNAP	:	RNA polymerase

RMSD	:	Root mean square deviation
RSA	:	Relative surface accessibility
RMRC	:	Regional Medical Research Centre
RBS	:	Ribosome binding site
rRNA	:	Ribosomal RNA
SEM	:	Standard errors of the mean
SEC	:	Size exclusion chromatography
SDS-PAGE	:	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SYBR	:	Synergy brand
sgRNA	:	Single guide RNA
ss-DNase	:	Single-stranded deoxyribonuclease
ss-DNA	:	Single-stranded DNA
SS	:	Small subunit of Cascade
SaPI	:	*Staphylococcus aureus* pathogenicity island
ss-RNA	:	Single-stranded RNA
SIE	:	Superinfection exclusion
TA systems	:	Toxin-Antitoxin systems
tracrRNA	:	Transactivating CRISPR RNA
TEV	:	Tobacco etch virus
TBE	:	Tris borate EDTA
TMB	:	Tetramethyl Benzidine
TBS	:	Tris-buffered saline
UV	:	Ultraviolet light

*****

List of figures

Supplementary figures

*****

List of tables

Supplementary tables

*****

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

Abstract

Leptospira interrogans is the causative agent of leptospirosis, a zoonotic disease accounting for approximately 60,000 human deaths every year globally. The leptospires show incompetence to conventional genetic manipulation tools, and therefore, the molecular mechanism of its pathogenesis remains poorly comprehended using the reverse genetics approach. One possible reason for its incompetence in genetic manipulation is the presence of the CRISPR-Cas system in its genome. A CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins) is an RNA-directed inheritable adaptive immunity in prokaryotes against invasive mobile genetic elements (MGEs), including bacteriophages and plasmids. The system encodes an effector complex (Cascade) that utilizes small crRNA (CRISPR RNA) to sense and interfere with the invasive MGEs having the crRNA-complementary sequence next to a protospacer adjacent motif (PAM). The predominant CRISPR-Cas type I-B system, marked in several pathogenic *Leptospira* genomes, presents a promising alternative to be explored as an endogenous genome editing tool. Moreover, to exploit *Leptospira's* CRISPR-Cas type I-B (Lin_I-B) system for targeted genetic manipulation, characterization of its interference machinery and associated PAMs is a prerequisite. To this end, the present study reports the molecular and functional characterization of interference machinery of a Lin_I-B system, recently identified in *L. interrogans* serovar Copenhageni strain Fiocruz L1-130. The comprehensive biochemical analysis of the LinCas7, a major subunit of the Cascade, revealed it to be a Mg^{2+} ion-dependent non-specific endoDNase, unlike other Cas7 family proteins. Moreover, LinCas7 exhibits distinct binding to crRNA in the presence of Mg^{2+} ions, testifying its canonical role in Lin_I-B defense response. The molecular characterization of LinCas8b disclosed it as a genetic fusion of large (LinCas8b) and small subunit (LinCas11b) proteins of the Cascade (LinCascade). In this study, LinCas11b is demonstrated to co-translate from an in-frame internal translation start codon encoded within the *lincas8b* gene. LinCas11b displays structural and functional analogy with the well-studied Cas11 family proteins. In addition, the interference machinery of the Lin_I-B, when expressed in a surrogate host *E. coli*, was able to annihilate the target DNA with the predicted PAM. In sum, the current study presents the molecular and functional insight of the *Leptospira* subtype I-B interference machinery that, soon, may pave the way for scientists to harness the system as a programmable endogenous tool for genetic manipulation.

Chapter 1

Introduction to CRISPR-Cas system, *Leptospira*, and rationale of the study

1.1. Prokaryotic immune systems

Bacteria (eubacteria and archaea) are persistently under the attack of bacteriophages (phages) wherein the parasitic phages hijack the host bacterial system for their own propagation that, eventually results in the death of host bacteria (Brüssow and Hendrix, 2002; Stern and Sorek, 2011; Weinbauer, 2004). Phages are the most abundant biological entities on the globe and outnumber bacteria by a tenfold margin (Brüssow and Hendrix, 2002; Clokie et al., 2011). Bacteria encounter approximately 10^{23} phage infections every second, thus, are under immense evolutionary pressure (Clokie et al., 2011; Stern and Sorek, 2011; Suttle, 2007). Therefore, there is an evolutionary arms race between bacteria and phages to avert the peril of extinction, and this has led to the evolution of several defense mechanisms in bacteria and also the counterdefense strategies in phages (Rostøl and Marraffini, 2019; Julie E. Samson et al., 2013; Stern and Sorek, 2011). Bacteria have developed multiple lines of active defense against invasive genetic elements (phages and plasmids) (Dy et al., 2014b; Rostøl and Marraffini, 2019) that can collectively be referred to as the prokaryotic "immune system."

The immune systems active within an organism are broadly classified into two types: Innate immune systems and Adaptive immune systems. The key difference between an innate and adaptive immune system is that the former does not keep any record of past infection while the latter does. An adaptive immune system within bacteria memorizes the previous phage attacks, conferring rapid and specific defense against phages upon recurrent infections. To date, the only known adaptive immune system in bacteria is CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins), which confers immunity against invasive mobile genetic elements (MGEs), including phages and plasmids. The CRISPR-Cas system acquires a short fragment of intruding MGEs into the bacterial genome as immunological memory. Then this system uses the RNA transcripts of the acquired fragments (spacers) as guides in degrading the same intruding MGEs during the recurrent invasion, confirming adaptive immunity to the bacterial host. Detailed information on CRISPR-Cas adaptive immunity has been illustrated in the latter sections (sections 1.2 to 1.6).

To fend off phage infections, bacteria have developed a plethora of innate, multi-layered defense strategies that act at different stages of the phage life cycle (Dy et al., 2014b; Rostøl and Marraffini, 2019) (**Fig. 1.1**). These innate defense systems impede the phage infection by (1) blocking the phage adsorption, (2) jamming the phage genome injection, (3) selective nucleolytic degradation of the phage genome, and/or (4) altruistic approaches (e.g., abortive infection and toxin-antitoxin systems). These bacterial innate immune systems are briefly described below.

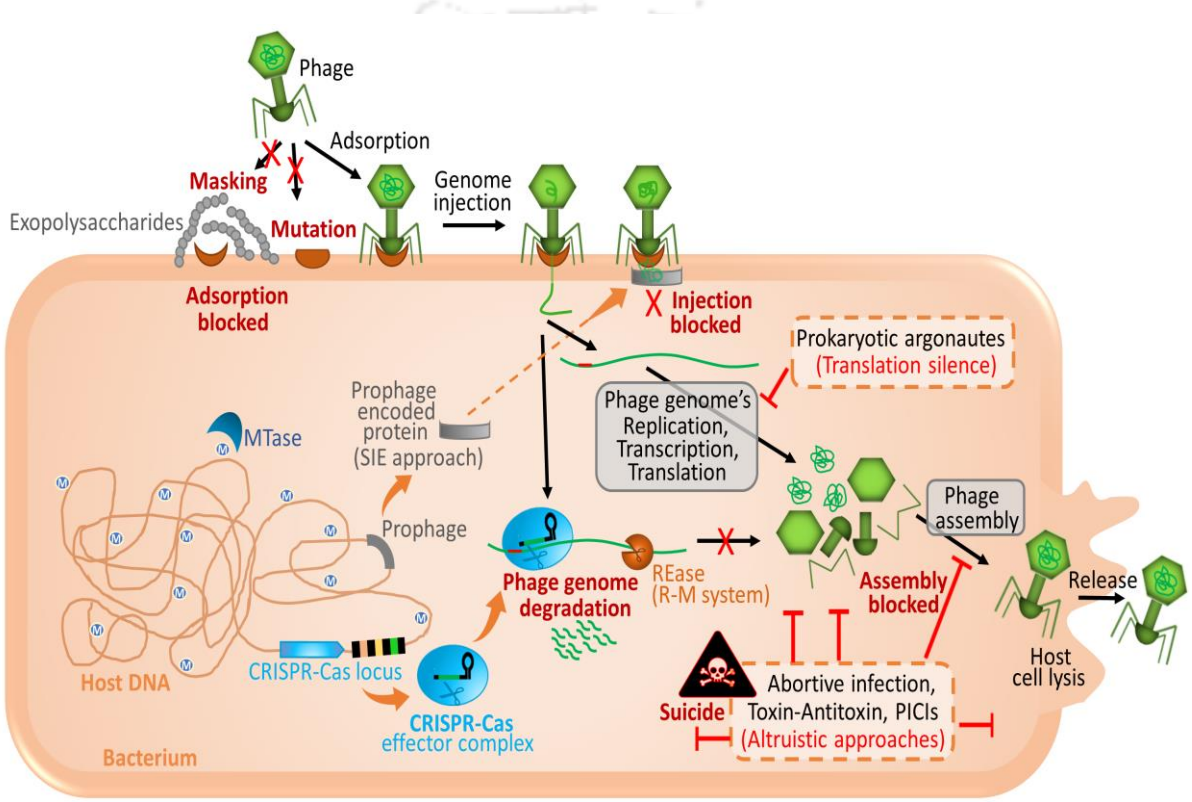

Fig. 1.1. Prokaryotic defense mechanisms against phage infection. Bacteria have developed multiple anti-phage defense strategies (innate and adaptive immunity) that act at different stages of a phage life cycle. Phage adsorption onto the host cell surface is prevented by changing the cell surface receptors (via post-translational modifications and/or mutations) or masking the receptor with extracellular coatings (exopolysaccharides). Injection of phage DNA is blocked via the SIE system encoded by an already lysogenized phage (prophage). If phage successfully enters the host, at this stage, R-M system targets the viral DNA. Wherein restriction enzyme (REase) cleaves the viral DNA (unmethylated), and methyltransferase (MTase) methylates the host's self-genome to prevent auto-targeting. CRISPR-Cas (adaptive immunity) also degrades the viral genome specifically. It memorizes the past infection by acquiring a short piece of phage DNA into the CRISPR locus on the host genome and utilizes its transcript to target the complementary viral genome during recurrent infections. Propagation of phages can also be blocked by prokaryotic argonautes that silence the translation of viral DNA, halting the generation of new phages. Bacteria also employ various altruistic approaches (abortive infection, toxin-antitoxin, and PICIs systems) as a last resort, which leads to the suicide of infected cells for the benefit of the population. The programmed cell death of the host before the maturation/assembly of viruses checks the release of multiple new phage progenies, thus, ensuring the protection to bacteria at the community level. [perceived from (Rostøl and Marraffini, 2019)]

1.1.1. Phage adsorption blockage

Phase infection begins with physical attachment (adsorption) of viruses to host bacteria via binding to specific cell surface receptor molecules, including surface proteins and extracellular glycoconjugates (Bertozzi Silva et al., 2016; Rostøl and Marraffini, 2019). To prevent phage adsorption, bacteria downregulate or modify the phage receptors by mutation or post-translational modifications (Harvey et al., 2017; Morona et al., 1984; Rostøl and Marraffini, 2019) (**Fig. 1.1**). For instance, *Bordetella bronchiseptica* during its non-virulent phase (Bvg^-) lacks the expression of the surface molecule "pertactin," unlike its virulent phase (Bvg^+). The absence of pertactin in the Bvg^- phase hinders the phase adsorption to the host cell (Liu et al., 2002). Bacteria can also block phage adsorption by developing additional extracellular coatings (e.g., biofilm and capsule) that mask the phage receptors (Dy et al., 2014b; Hammad, 1998; Vidakovic et al., 2017). For example, *E. coli* produces a K1 capsule that blocks the surface attachment of the T7 phage (Scholl et al., 2005).

1.1.2. Blockage of phage DNA injection

Upon successful adsorption onto the host cell surface, phages inject their genetic material into the bacterial cytoplasm. Bacteria check the phase infection at this stage, engaging Superinfection exclusion (SIE) systems that block the entry of the phage genome into the cytoplasm (Cwick et al., 2022; Lu and Henning, 1994) (**Fig. 1.1**). In the SIE mechanism of bacterial phage resistance, a prophage interferes with recurrent infection of the same or closely related phages (Cwick et al., 2022; Lu and Henning, 1994). For instance, the *E. coli* phage HK97, once lysogenised (prophage), encodes transmembrane protein gp15 that stops the entry of viral genome during subsequent infections of HK97 or closely related phage HK75 (Cumby et al., 2012).

1.1.3. Nucleolytic degradation of the phage genome

In case phages bypass the initial defense measures of the host and enter their genome into bacterial cytoplasm, bacteria utilize multiple strategies to sense and nucleolytically degrade the non-self-viral genome. Among various such strategies, the restriction-modification (R-M) systems, prokaryotic argonautes (pAgos), and the CRISPR-Cas adaptive immune systems are the best-comprehended defense mechanisms in bacteria (**Fig. 1.1**).

1.1.3.1. R-M systems

Restriction-modification (R-M) systems are the well-studied, most common innate immunity of prokaryotes against MGEs. Based on variations in molecular composition and DNA restriction mechanism, R-M systems are categorized into four types (I-IV). A typical R-M system consists of two subunits: a restriction endonuclease (REase) and a methyltransferase (MTase). REase recognizes a 4-8 bp long palindromic DNA sequence (recognition site) and cleaves the DNA strictly at the recognition site or at certain bases away from the recognition site. Moreover, to avoid the self-DNA cleavage by REase, an R-M system employs MTase, which specifically methylates the adenine or cytosine bases of the recognition site in the endogenous genome, preventing REase from acting on the self-genome. Thus, during phage invasion, REase cleaves the unmethylated non-self-viral DNA, restricting its replication or lysogenization while the methylated bacterial genome remains protected (Loenen et al., 2014; Murray, 2000; Sistla and Rao, 2004; Tock and Dryden, 2005; Williams, 2003) (**Fig. 1.1**).

1.1.3.2. Prokaryotic argonautes

The argonautes (Agos) were initially discovered in eukaryotes (eAgos) as the effector proteins constituting RNA interference (RNAi) system (Bohmert, 1998; Ketting, 2011; Daan C Swarts et al., 2014). Multiple argonaute proteins (eAgos) assemble with small single-stranded RNA (ssRNA) into a ribonucleoprotein complex called RISC (RNA-induced silencing complex). A RISC complex uses its ssRNA component as a guide to specifically bind and cleave the target RNA having the sequence complementarity with ssRNA; thereby, it silences the translation of targeted RNA (Hannon, 2002; Hutvagner and Simard, 2008; Daan C Swarts et al., 2014). Later, homologs of eAgos were identified in prokaryotes (pAgos) (Hegge et al., 2018; Daan C Swarts et al., 2014). Unlike eAgos, which mediates only RNA-guided RNA interference, pAgos confer either RNA-guided or DNA-guided RNA and/or DNA interference (Hegge et al., 2018; Kaya et al., 2016; Daan C. Swarts et al., 2014; Swarts et al., 2015).

On a structural basis, pAgos are grouped into three classes: long pAgos, short pAgos, and PIWI-RE proteins (P-element induced wimpy testis with conserved R (Arg) and E (Glu) residues) (Hegge et al., 2018). Among these pAgo classes, only the long pAgo class has been studied experimentally (Hegge et al., 2018; Miyoshi et al., 2016). Similar to eAgos, pAgos also exhibit a bi-lobed structure, wherein one lobe comprises the amino-terminal (N) and PIWI-

Argonaute-Zwille (PAZ) domains, while another one consists of middle (MID) and PIWI domains (Hegge et al., 2018; Miyoshi et al., 2016; Song et al., 2003; Yuan et al., 2005). The PAZ and MID domains involve in guide oligonucleotide binding, and the PIWI domain catalyzes the target-strand cleavage (Hegge et al., 2018; Miyoshi et al., 2016; Song et al., 2003; Yuan et al., 2005). These pAgos, in complex with guide RNA, silence the translation of target mRNAs without cleaving them. Moreover, the mechanism of guide RNA/DNA acquisition, target recognition, and cleavage by pAgos are yet to be fully understood.

1.1.4. Altruistic approaches to impede phage infection

Once the phages escape nucleic acid interference systems of host bacteria, the viral genome replicates rapidly to generate new phages via the lytic pathway or remains integrated into the host genome as prophage following the lysogenic path until it re-enters into the lytic cycle. In the lytic cycle, the phage genome replicates vigorously and produces viral proteins in bulk leading to the fatal depletion of the host's cellular resources. Eventually, these viral molecules (nucleic acids and proteins) assemble into new phage progenies that release themselves by lysing the host cell and further infect other surrounding bacteria. At this stage, to limit the severity of phage infection, bacteria enact some altruistic approaches which lead to the programmed death of infected cells but ensure the survivability of surrounding bacteria. Such altruistic strategies bestow protection to bacteria at the community level against phage predators. Examples of such altruistic approaches include abortive infection (Abi) (Lopatina et al., 2020), toxin-antitoxin (TA) systems (Lopatina et al., 2020; Unterholzner et al., 2013), and phage-inducible chromosomal islands (PICIs) (Ibarra-Chávez et al., 2022; Rostøl and Marraffini, 2019) (**Fig. 1.1**).

1.1.4.1. Abortive infection

Abortive infection (Abi) is a defense strategy wherein the phage-infected bacteria undergo self-cell death before the virus can complete its replication cycle. Abi precludes the spread of phage infection to nearby cells, thus protecting the bacterial colony. Abi systems can induce self-cell death of infected bacteria by damaging the cell membrane or altering the critical cellular process like DNA replication, transcription, and translation. To manifest the defense response, every Abi system contains at least two functional modules: one that senses the phage infection and another that acts as an effector killing the cell or shutting down the metabolism (Lopatina

et al., 2020). Moreover, the exact mechanism of Abi's functioning is poorly understood. The proteins of Abi systems get expressed or activated during phage infection. These proteins cause fatal membrane damage (Cohen et al., 2019; Durmaz and Klaenhammer, 2007; Lopatina et al., 2020) or indiscriminately degrade phage and host DNA/RNA (Fineran et al., 2009; Lau et al., 2020; Levitz et al., 1990), cleave proteins essential in host translation systems (Yu and Snyder, 1994). Thus, the bacterial cell undergoes an immediate growth arrest, resulting in self-cell death. A few examples of Abi systems include the Rex system in *E. coli* (Parma et al., 1992) and AbiA to AbiZ systems in *Lactococcus lactis* (Chopin et al., 2005; Durmaz and Klaenhammer, 2007).

1.1.4.2. Toxin-Antitoxin systems

Like Abi, Toxin-Antitoxin (TA) systems are another altruistic strategy evolved in bacteria to contain the spread of phage infections. The Abi and TA systems have a marginal differentiation and even have overlapped activities in a few instances (Dy et al., 2014a; J. E. Samson et al., 2013). A TA system comprises a toxin and its counter-antitoxin (Song and Wood, 2020; Unterholzner et al., 2013). In unstressed cells (free of phage infection), the antitoxin keeps the cognate toxin suppressed. Upon phage infection, the toxin becomes activated that alters the vital cellular processes, including DNA replication, translation, and cell wall synthesis, and indiscriminately degrades the phage and host nucleic acids (Aakre et al., 2013; Unterholzner et al., 2013; Winther et al., 2016). This results in host cell metabolic arrest that, in most cases, leads to self-cell death, impeding the further propagation of phages. According to the molecular nature of the toxin and antitoxin and their action mechanism, various TA systems are classified into eight types (type I-VIII) (Song and Wood, 2020). In the majority of known TA systems, toxins are proteins, while antitoxins are either proteins or non-coding RNAs (Song and Wood, 2020; Unterholzner et al., 2013). The type II TA systems are the predominant type, where both toxin and the antitoxin are protein molecules, and the antitoxin physically binds to the toxin to suppress its activity. For example, the type II TA system RnlAB in *E. coli* confers protection against certain T4 phage strains (Koga et al., 2011). The toxin RnlA exhibits endoribonuclease activity which, in unstressed cells, remains inhibited by its cognate antitoxin RnlB protein via physical interactions (Koga et al., 2011).

1.1.4.3. Phage-inducible chromosomal islands

Phage-induced chromosomal islands (PICIs) are a novel class of MGEs that remains lysogenized into bacterial genome and become active during the lytic cycle of certain phages (also termed as helper phages) (Ibarra-Chávez et al., 2022; Rostøl and Marraffini, 2019). Upon activation, PICIs parasitize the helper phages for their own benefit. They hijack the helper phage packaging machinery for the packaging of their own genetic material, thus, severely retard the production of new lytic phages. The *Staphylococcus aureus* pathogenicity islands (SaPIs) were the first documented PICIs (Lindsay et al., 1998). The excision-replication-packaging (ERP) cycle of SaPI becomes active during the lytic phase of phage 80α (Tormo-Más et al., 2010). Following ERP, the released new PICI viral particles (SaPIs) infect new cells and remain in lysogeny. Though the activation of the PICI system leads to host cell death, it severely reduces the propagation of lytic phages and protects the surrounding bacterial population.

1.1.5. Recently discovered prokaryotic immune systems

Bacterial genome analysis has disclosed that the genes comprising a defense system are frequently clustered together in a microbial genome, and such genomic regions are known as "defense islands." In recent years, comprehensive analysis of defense islands in various bacteria has led to the identification of numerous novel immune systems against MGEs, for instance, BREX (bacteriophage exclusion) (Goldfarb et al., 2015), DISARM (Defense island associated with restriction-modification) (Ofir et al., 2017), and Zorya (Doron et al., 2018), etc. Our understanding of the action mechanism of these newly discovered prokaryotic immune systems is in its infancy. In preliminary studies, these defense systems demonstrated the prevention of host bacteria against phage infection (Doron et al., 2018). However, a systematic study is warranted to decipher their exact mechanism of action.

1.2. CRISPR-Cas, the prokaryotic adaptive immunity: Overview

1.2.1. Discovery of CRISPR-Cas

The CRISPR-Cas systems are the only identified adaptive immune system in bacteria (eubacteria and archaea). They are genetically inheritable and confer RNA-directed adaptive immunity to host bacteria against invasive MGEs, including phages and plasmids. The acronym "CRISPR," stands for Clustered Regularly Interspaced Short Palindromic Repeats,

was first coined by Jansen and his colleagues for the bacterial genomic loci consisting of short, direct repeats interspaced by similarly sized hypervariable sequences (spacers) (Jansen et al., 2002). Before this, CRISPR was noticed for the first time by Ishino and coworkers in the *E. coli* genome (Ishino et al., 1987). Since then, it has been marked in almost all archaea and half of the eubacteria with sequenced genomes (Makarova et al., 2020a, 2015). The presence of CRISPR-associated (*cas*) genes in the vicinity of CRISPR loci suggested the functional relationship between the CRISPR locus and *cas* genes (Jansen et al., 2002). Further investigations revealed that spacer sequences match with foreign genetic elements (phages and plasmids), and this led to the hypothesis that CRISPR-Cas is an immune system to protect against invasive MGEs (Bolotin et al., 2005; Makarova et al., 2006; Mojica et al., 2005; Pourcel et al., 2005). Subsequently, Barrangou and coworkers experimentally established the CRISPR-Cas as an adaptive immune system of bacteria that acquires new spacers from the invading phages and confers protection against recurring phage infections utilizing Cas proteins (Barrangou et al., 2007).

1.2.2. Genetic composition of CRISPR-Cas

CRISPR-Cas loci have been identified in approximately 50 % of sequenced eubacterial and 87 % of sequenced archaeal genomes (Makarova et al., 2020a, 2015). Approximately 3 % of sequenced plasmids (eubacterial and archaeal) also encode CRISPR-Cas loci (Pinilla-Redondo et al., 2022). An active CRISPR-Cas locus entails three genetic elements: a CRISPR array, an AT-rich leader sequence preceding it, and a set of *cas* genes in the close vicinity (Karginov and Hannon, 2010) (**Fig. 1.2**). Many bacterial genomes contain multiple CRISPR loci (Jansen et al., 2002). The leader sequence is non-coding and extends several hundred base pairs in size (Karginov and Hannon, 2010). Within a multi-CRISPR loci genome, leaders share up to 80 % sequence identity, but they are pretty dissimilar among bacterial species (Karginov and Hannon, 2010). The CRISPR array element comprises a battery of short, direct repeats intervened by spacer sequences (Karginov and Hannon, 2010) (**Fig. 1.2**). The number of repeat-spacer-repeat units in a CRISPR locus can range from one to several hundred with an average of around twenty (Karginov and Hannon, 2010). The spacers' DNA sequences are selectively acquired from invasive MGEs during past attacks and serve as an immunological memory for the targeted elimination of the same genetic materials during subsequent infections (Barrangou et al., 2007; Bolotin et al., 2005; Brouns et al., 2008; Hale et al., 2009; Hille et al., 2018). Within a CRISPR locus, spacers are identical in size but highly variable in sequence

composition. Among different species, spacer size varies from 20 to 72 bp (Karginov and Hannon, 2010). Repeats of a CRISPR locus are generally identical in both size and sequence. The size of repeats among different species can vary between 21 to 47 bp (Karginov and Hannon, 2010). Closely related species can contain repeats with similar sequences, but the overall bacterial and archaeal sequence diversity of both spacers and repeats is high (Karginov and Hannon, 2010; Kunin et al., 2007). Despite high sequence diversity, most CRISPR repeats are partly palindromic, which allows them to generate RNAs with stable stem-loop (hairpin) structures- an essential feature of crRNAs (CRISPR RNAs) associated with CRISPR-Cas immunity (Karginov and Hannon, 2010; Kunin et al., 2007).

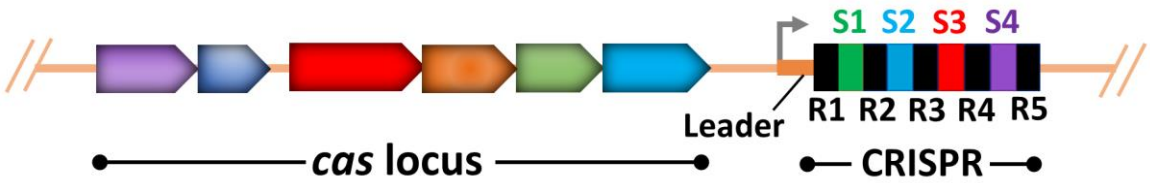

Fig. 1.2. Genomic architecture of a typical CRISPR-Cas system. CRISPR-Cas system comprises a CRISPR array containing identical direct repeats (R1-R5) interspaced by equally sized variable spacers (S1-S4), an AT-rich leader region upstream to the CRISPR array for transcription of the array, and *cas* gene cassette in close vicinity.

1.2.3. The three-stage mechanism of CRISPR-Cas immunity

Since its first validation as a prokaryotic adaptive immune system in *Streptococcus thermophilus* by Barrangou and coworkers (Barrangou et al., 2007), CRISPR-Cas systems have been characterized in a large number of bacterial genomes that led to the identification of numerous variants of *cas* genes (Makarova et al., 2020a). Despite the *cas* gene diversification, the overall mechanism by which the CRISPR-Cas systems confer immunity in various host organisms remains similar (Jiang and Marraffini, 2015; Li and Peng, 2019). Every CRISPR-Cas system develops immunity via a three-stage mechanism: 1) adaptation/spacer acquisition, 2) expression/crRNA biogenesis, and 3) interference (Jiang and Marraffini, 2015; Makarova et al., 2020a) (**Fig. 1.3**).

New immune memories are developed during the adaptation stage, wherein selected DNA fragment(s) (termed prespacer) from the intruding genomes are incorporated as unique spacers

into the CRISPR array (Yosef et al., 2012). The adaptation complex, comprised of a set of Cas proteins, recognizes the prespacer and precisely incorporates it as a new spacer into the CRISPR array at the leader-repeat (first repeat) junction via a cut-paste mechanism (Yosef et al., 2012). A concomitant duplication of the first repeat accompanies the spacer integration event (Yosef et al., 2012), maintaining the repeat-spacer-repeat arrangement of the CRISPR array. Next, in the expression stage, the CRISPR array is transcribed into a non-coding single-long precursor crRNA (pre-crRNA). Specific Cas endonuclease recognizes the structured repeat regions on pre-crRNA and processes it into multiple small, mature crRNAs, each containing a spacer sequence flanked by leftover repeat parts (Jiang and Marraffini, 2015; Li and Peng, 2019). In some CRISPR-Cas variants, the processing of pre-crRNA into crRNA involves bacterial Ribonuclease III (RNase III) and an additional RNA called tracrRNA (transactivating crRNA) (Deltcheva et al., 2011; Li and Peng, 2019). Each crRNA assembles with either one or multiple Cas proteins into a ribonucleoprotein effector complex (Li and Peng, 2019; Liu and Doudna, 2020). Then, in the interference stage, during recurrent infection, crRNA recruits the effector complex onto the intruder genome via complementary base-pairing with the protospacer sequence lying next to the protospacer adjacent motif (PAM) in the target genome. Eventually, the effector complex elicits targeted degradation of the intruder genome using either an intrinsic nuclease activity or a separate nuclease in trans (Li and Peng, 2019; Liu and Doudna, 2020), conferring the adaptive immunity to host bacteria (**Fig. 1.3**).

1.3. Classification of CRISPR-Cas systems

The arms race between bacteria and phages for survival keeps both life forms on continuous evolution leading to rapid adaptive reformations in their defense mechanisms, including bacterial CRISPR-Cas systems. Though all CRISPR-Cas systems function in a similar three-stage mechanism, they exhibit remarkable variability in gene compositions, Cas protein sequences, genomic organization, and the stoichiometry of effector complex (Koonin et al., 2017; Liu and Doudna, 2020; Makarova et al., 2020a). The ever-growing genomic and metagenomic data has led to the identification of a variety of CRISPR-Cas systems with compositional, structural, and operational diversity. Owing to the high diversity and fast evolution of the CRISPR-Cas systems, their classification is a formidable challenge. Moreover,

several classification schemes have been introduced and updated from time to time following research advancement (Koonin et al., 2017; Makarova et al., 2020a, 2018, 2015).

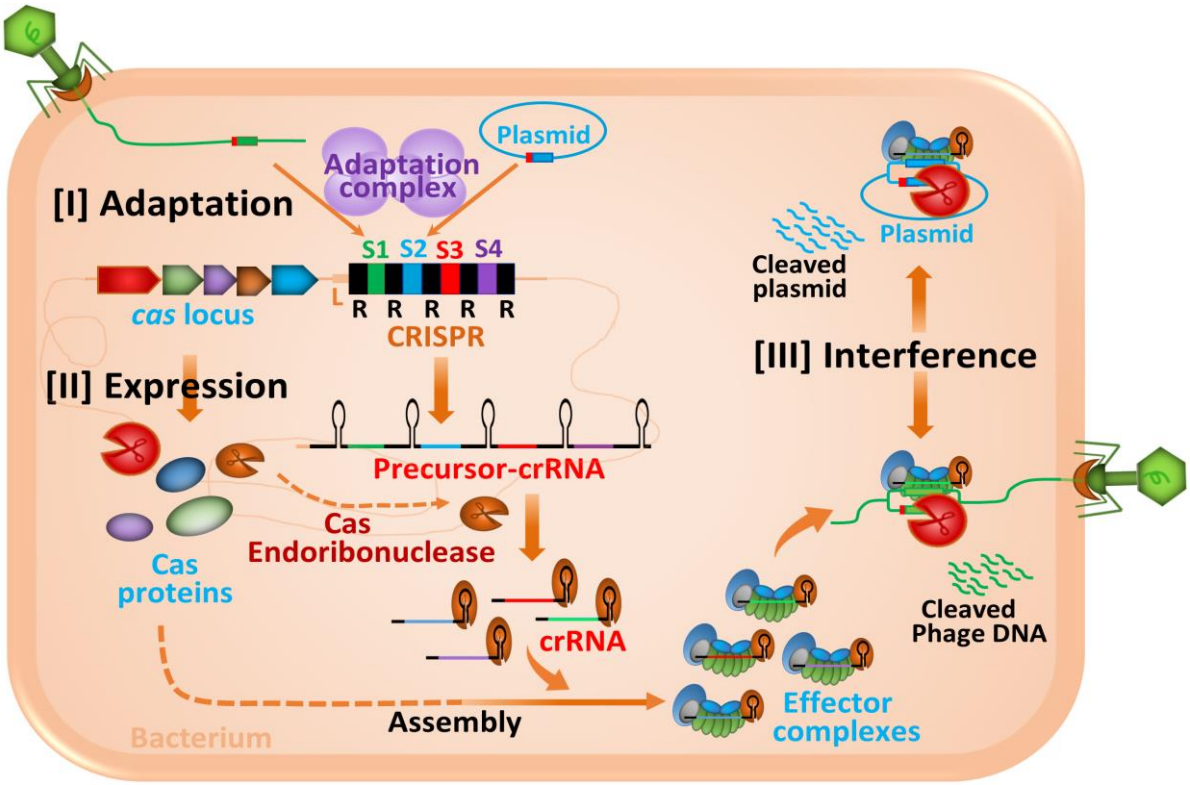

Fig. 1.3. An overview of the three-stage mechanism of CRISPR-Cas adaptive immunity. The figure presented here displays a generalized mechanism of CRISPR-Cas adaptive immunity against invasive genomes (phage and plasmid) based on the best-comprehended type I-E CRISPR-Cas system. An active CRISPR-Cas system confers immunity in three successive stages. Upon the first infection of a phage/plasmid, invasive genome is degraded by the host RecBCD complex (not shown). Subsequently, in the first stage (**adaptation**), a short viral/plasmid DNA fragment is incorporated as a new spacer (here S1 and S2) in the CRISPR array at the leader (L)-repeat (R) junction by Cas1-Cas2 integrase (adaptation complex) with the help of host integration factor (not shown), followed by duplication of the repeat. In the second stage (**expression**), crRNA biogenesis takes place. Cas proteins are expressed, and the CRISPR array is transcribed into a single precursor crRNA (pre-crRNA) transcript. Then specific Cas endoribonuclease processes the pre-crRNA into small mature crRNAs that assemble with multiple Cas proteins, forming the effector complex. In the **interference** stage during recurrent viral/plasmid infection, the effector complex recognizes the target DNA via crRNA complementary base pairing. Finally, an effector nuclease (partial circle with scissor, red) degrades the targeted DNA, conferring immunity to the host. [perceived from (Hille et al., 2018)]

The most recent classification scheme is given by Makarova and coworkers in the year 2020 (Makarova et al., 2020a). Since none of the genes is common among CRISPR-Cas systems, classification based on phylogenetic analysis of genes is implausible. Therefore, in the latest classification, a multipronged computational approach has been applied, including the identification of signature genes, comparison of gene composition and genomic organizations, sequence similarity-based clustering and phylogenetic analysis, and available experimental evidence (Makarova et al., 2020a). The latest classification categorized the various CRISPR-Cas systems into two broad classes (Class 1 and 2), six types (Type I-VI), and a total of 33 subtypes (**Fig. 1.4**).

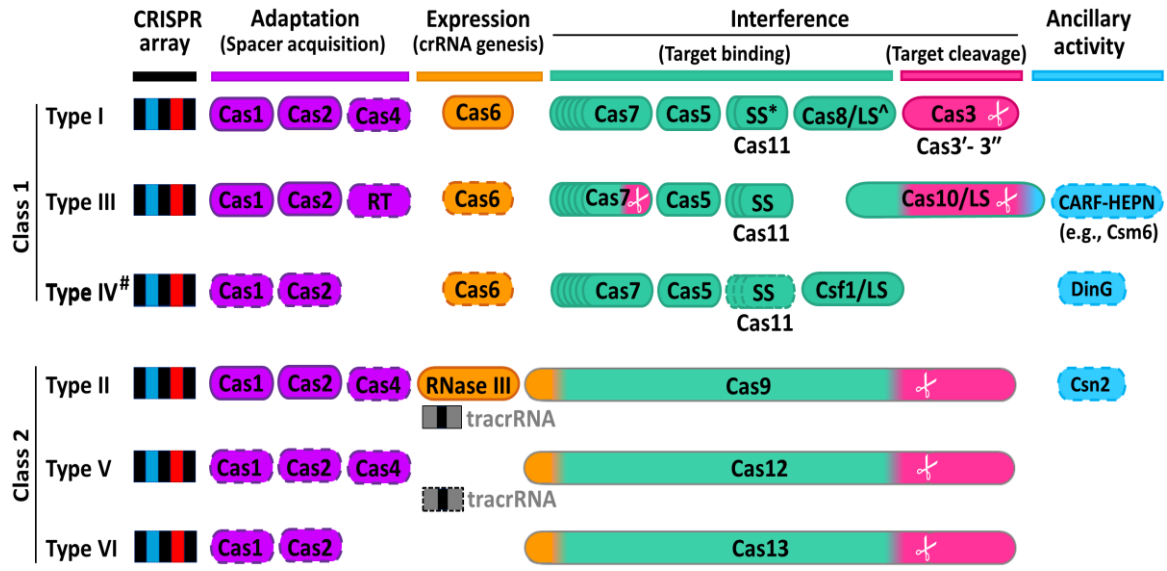

Fig. 1.4. Classification of CRISPR-Cas systems and their molecular constituents. Based on the number of proteins involved in interference (target binding and cleavage), CRISPR-Cas systems are categorized into two classes (Class 1 and 2). Class 1 systems use a multiprotein ribonucleoprotein complex for target interference. In class 2 systems, a single protein ribonucleoprotein complex performs the target interference. The figure represents the molecular constituents (proteins, crRNA encoding CRISPR array) but not their genetic organization in various CRISPR-Cas systems. Proteins are grouped (color-coded) according to their involvement in different stages of CRISPR-Cas immunity. Components depicted with dotted outlines indicate they are present in some subtypes but absent in others or encoded on other co-existing CRISPR-Cas types. CRISPR array is shown for all the types; however, it might be absent and shared with other co-existing CRISPR types/subtypes. The putative copy number is indicated as stacks of the respective protein. Proteins performing multiple functions are displayed in multiple colors accordingly. The scissor mark indicates the nuclease activity (effector DNase/RNase) of the Cas protein(s) during target interference. An asterisk (*) indicates the small subunit (SS or Cas11) that might be fused with the large subunit (LS or Cas8) in several type I subtypes. The circumflex symbol (^) shows that Cas8 might have an extension at the C-terminal representing a putative fused SS (Cas11) in several type I subtypes. The classification illustrated here is adapted from the most recent classification introduced by Makarova et al., 2020 and the putative copy number of SS (Cas11; >2) in the type IV system has been updated [hash symbol (#)] based on recent structural characterization of type IV effector complex by Zhou et al., 2021.

Based on their role, all the Cas and non-Cas proteins associated with CRISPR-Cas immunity are grouped into four distinct functional modules: adaptation, expression, interference, and ancillary (Makarova et al., 2020a) (**Fig. 1.4**). The adaptation module is largely uniform across various CRISPR-Cas systems and comprises the Cas1 and Cas2 proteins along with the CRISPR array. In some type I systems, the adaptation module also includes the restriction endonuclease superfamily enzyme Cas4 (Hooton and Connerton, 2015; Makarova et al., 2015). The expression module includes the RNA endonucleases that process the pre-crRNA into mature crRNA (typically Cas6 in Class 1 systems; RNase III and/or alternative host RNase in Class 2 systems) (Makarova et al., 2020a). The interference module encompasses the Cas subunits of the effector complex and Cas nucleases (DNase/RNase) that work together in coordination to recognize and degrade the cognate target nucleic acids. The various other proteins and domains directly or indirectly participating in the CRISPR-Cas defense mechanism are grouped as ancillary modules. Examples of ancillary proteins include CARF (CRISPR-associated Rossmann fold) and HEPN (Higher eukaryotes and prokaryotes nucleotide-binding) domain proteins (Csm6 and Csx1) in type III systems, DinG in type IV, and Csn2 in type II CRISPR-Cas systems (Makarova et al., 2020a).

1.3.1 Class 1 CRISPR-Cas systems

The categorization of CRISPR-Cas systems into two broad classes (Class 1 and 2) is based on the subunit composition of the effector complex (Makarova et al., 2020a). Class 1 CRISPR-Cas systems encompass multi-Cas protein effector complexes (**Fig. 1.5**). Various Cas proteins, namely Cas3, Cas5, Cas6, Cas7, Cas8, Cas10, and Cas11 or their functional orthologs, in different combinations (depending on the type and subtypes), assemble onto a crRNA, forming the effector complex (Makarova et al., 2020a). On the basis of signature Cas proteins that catalyze the target nucleic acid interference, the class 1 CRISPR-Cas systems are further divided into three types: type I, III, and IV (Makarova et al., 2020a).

1.3.1.1. Type I CRISPR-Cas systems

The nuclease effector Cas3 protein is the signature protein for type I CRISPR-Cas systems (Makarova et al., 2020a). Considering the variability in the sequence of Cas8 protein and the genomic architecture of CRISPR-Cas loci, type I systems are further categorized into nine subtypes: I-A, I-B, I-C, I-D, I-E, I-F1, I-F2, I-F3, and I-G (Makarova et al., 2020a).

Subtype I-B represents the most abundant CRISPR-Cas system across different bacteria (eubacteria and archaea), followed by subtypes I-C and I-E (Makarova et al., 2015). Cas1, Cas2, and Cas4 proteins constitute the adaptation module in most type I CRISPR-Cas systems; however, subtypes I-E and I-F lack Cas4 (Makarova et al., 2020a). In subtype I-G, Cas4 is fused with Cas1 (Almendros et al., 2019; Makarova et al., 2020a). In many bacteria, Cas4 is encoded outside the CRISPR-Cas locus or from the MGEs (Hudaiberdiev et al., 2017). All type I systems generally employ Cas6 as an endoribonuclease for mature crRNA genesis, but the subtype I-C locus lacks the *cas6* gene (Makarova et al., 2020a). In the subtype I-C system, the functional role of Cas6 is supplanted by Cas5c, a catalytically active variant of the typically non-catalytic Cas5 family (Hochstrasser et al., 2016; Punetha et al., 2014; Reeks et al., 2013b). In subtype I-G, Cas5 and Cas6 are expressed as a single fusion protein termed "Csb2" (Makarova et al., 2020a; Shangguan et al., 2022).

The effector complex in the CRISPR-Cas type I system is termed Cascade (a CRISPR-associated complex for antiviral defense) (Liu and Doudna, 2020). In general, a Cascade complex consists of a crRNA assembled with a Cas5, Cas6, Cas8 (large subunit), two Cas11 (small subunit), and a variable number of Cas7 subunits (Liu and Doudna, 2020). Among various CRISPR-Cas systems, the Cascade complexes, despite having enormous diversity in their stoichiometric composition of Cas proteins, exhibit a similar seahorse-like architecture with a head, foot, backbone, and belly (Liu and Doudna, 2020; O'Brien et al., 2020; Shangguan et al., 2022) (**Fig. 1.5A**). Wherein the 5′ end of crRNA is capped by a single subunit of Cas8 and Cas5, shaping the "foot" of the Cascade complex. The Cas6 subunit remains bound to the 3′ end of the crRNA forming the "head". The helical "backbone" filament is formed by the assembly of multiple Cas7 subunits along the crRNA spacer region. The "belly" filament is constituted by Cas11 subunits (Liu and Doudna, 2020; O'Brien et al., 2020) (**Fig. 1.5A**). In subtype I-D, the large subunit protein is termed Cas10d (non-catalytic), instead of Cas8, because of its close structural similarity with Cas10 (catalytic) of type III systems (Kira S Makarova et al., 2011; McBride et al., 2020). The subtype I CRISPR-Cas loci (I-B, I-C, I-D, I-F, and I-G), except I-E and I-A, lack an independent gene coding for the Cas11 subunit of the Cascade complex (Kira S Makarova et al., 2011; Makarova et al., 2015). Instead, they encode a large subunit Cas8 (or Cas10d in type I-D) with an extended C-terminus consisting of largely α-helices (Kira S Makarova et al., 2011; Makarova et al., 2015). These α-helix-rich C-terminus of Cas8 and Cas10d proteins have been postulated to be a functional substitute for Cas11

subunits in the Cascade complex of subtype I-B, I-C, I-D, I-F, and I-G systems (Kira S Makarova et al., 2011; Makarova et al., 2015). In line with this bioinformatics prediction, many of the subtype I systems (I-B, I-C, and I-D) have recently been shown to encode the Cas11 protein from within the *cas8* (or *cas10d*) gene (McBride et al., 2020). The Cas8/Cas10d and Cas11 proteins, across different type I systems (and type III systems), exhibit very low or no significant sequence similarity with their homologs (Makarova et al., 2020a; Venclovas, 2016; Vestergaard et al., 2014). Moreover, they share significant structural similarities with their homologs and therefore occupy analogous positions in various effector complexes (Jackson and Wiedenheft, 2015; Liu and Doudna, 2020; Venclovas, 2016).

Further, in a Cascade complex, the assembly of Cas7 subunits varies depending on the length of the spacer sequence in the crRNA. Typically, one Cas7 binds per 6 nucleotides (nts) of the crRNA spacer (Kuznedelov et al., 2016; Shangguan et al., 2022). The formation of the crRNA-Cas7 helical assembly is fundamental to the functioning of the CRISPR-Cas type I interference machinery (in type III systems as well) (Kuznedelov et al., 2016; Liu and Doudna, 2020). The Cas7 family proteins share a common structural element, i.e., a central non-canonical RNA recognition motif (RRM)/ferredoxin-like domain with variable insertions and C-terminal extension (Benda et al., 2014; Hrle et al., 2013; Lintner et al., 2011; Reeks et al., 2013b). Despite sequence disparity, the primary operative function of Cas7 family proteins in type I systems (also in type III) is to bind and stabilize the crRNA. The structural characterization of Cascade complex from various CRISPR-Cas type I systems, including *E. coli* type I-E (Ryan N. Jackson et al., 2014), *Desulfovibrio vulgaris* type I-C (Hochstrasser et al., 2016; O'Brien et al., 2020), *Synechocystis* spp. type I-D (McBride et al., 2020), and *Thioalkalivibrio sulfidiphilus* type I-G (Shangguan et al., 2022) via X-ray crystallography and/or cryo-electron microscopy has revealed the characteristic organization of Cas7 subunits around the crRNA in a functional Cascade complex. The Cas7 subunits lying proximal to the 5′-end of the crRNA interact with Cas8 and Cas5, while in the belly region of Cascade, Cas7 subunits bind with Cas11 (Hochstrasser et al., 2016; Liu and Doudna, 2020; O'Brien et al., 2020). Thus, the Cas7 also promotes the relative-assembly of other Cas proteins in the Cascade complex and eventually determines the size and target recognition-specificity of the effector complex (Halpin-Healy et al., 2020; Hochstrasser et al., 2016; Kuznedelov et al., 2016).

In most of the subtype I systems, the Cas3 nuclease effector does not stably bind with the Cascade complex, rather acts *in trans* during target interference post-recognition of cognate foreign DNA by the Cascade (Ryan N. Jackson et al., 2014; Liu and Doudna, 2020; McBride et al., 2020; O'Brien et al., 2020). However, in subtype I-A (Plagens et al., 2014) and I-G (Shangguan et al., 2022), Cas3 has been described as a stable component of the effector complex. A typical Cas3 protein contains an N-terminal HD (histidine-aspartate) nuclease domain fused to a helicase domain (He et al., 2020; Makarova et al., 2020a). Across subtype I systems, Cas3 exhibits variability in the domain architecture. For example, the two domains of Cas3 (nuclease and helicase) are encoded as two distinct proteins (Cas3" and Cas3', respectively) by separate genes in subtype I-A systems (He et al., 2020; Makarova et al., 2020a). In subtype I-F, Cas3 is expressed as a fusion protein with Cas2; in subtype I-D, Cas3's nuclease domain is fused to Cas10d (He et al., 2020; Makarova et al., 2020a). The domains of Cas3 in subtype I-G are interchanged: helicase domain at the N-terminal and nuclease domain at C-terminal (Makarova et al., 2020a).

1.3.1.2. Type III CRISPR-Cas systems

The CRISPR-Cas loci that encode catalytically active Cas10 protein as a large subunit of their effector complexes are classified into type III systems (Makarova et al., 2020a; Osawa et al., 2015; Staals et al., 2014). Considering the variations in the Cas10 sequence and architecture of CRISPR-Cas loci, type III systems are further categorized into six subtypes (III-A to III-F) (Makarova et al., 2020a). Among the six subtypes, III-A and III-B are the most common and best studied (Makarova et al., 2020a, 2015). All type III systems (except subtype III-A and III-E) lack adaptation modules and possibly share the CRISPR array and adaptation proteins from other CRISPR-Cas subtypes at different genomic locations (Makarova et al., 2020a). Notably, the adaptation modules in some subtype III systems encode reverse transcriptase, separately or in fusion with Cas1, for spacer acquisition from the RNA phages (González-Delgado et al., 2019; Makarova et al., 2020a; Silas et al., 2016). The enzyme responsible for crRNA genesis in type III systems is not clearly known but assumed to be Cas6 which in some cases is encoded on other CRISPR-Cas loci (Liu and Doudna, 2020; Makarova et al., 2015).

The effector complexes in type III systems are termed Csm (Cas subtype Mtube; in subtypes III-A, III-D, III-E, and III-F) or Cmr [Cas module RAMP (repeat-associated mysterious protein); in subtypes III-B and III-C]. In general, a type III effector complex (Csm and Cmr)

looks like a "seaworm" (**Fig. 1.5B**) rather than the seahorse-shaped architecture of the typical type I Cascade (Liu and Doudna, 2020; Makarova et al., 2015).

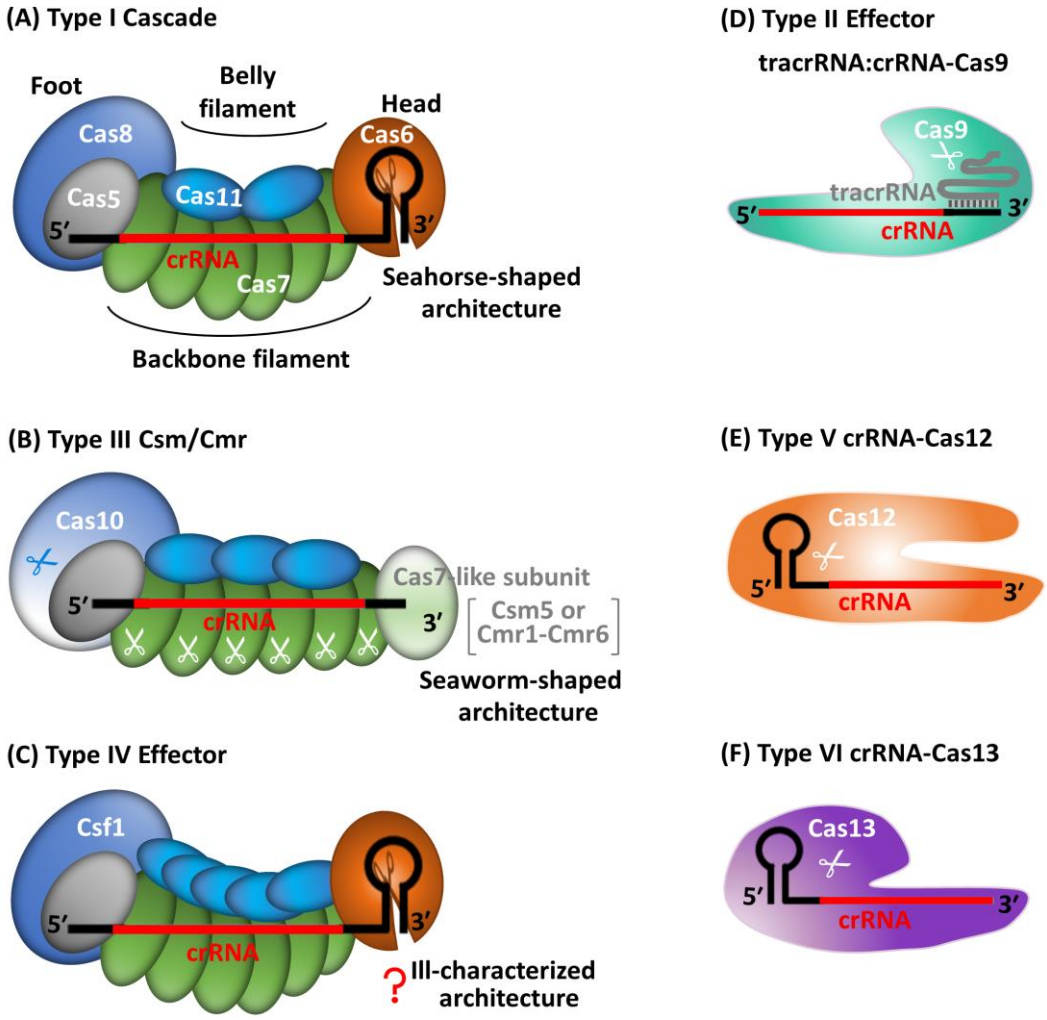

Fig. 1.5. Molecular composition and architecture of effector complex in various CRISPR-Cas types. (A) Type I effector complex (termed Cascade) typically comprises five different Cas proteins assembled onto a crRNA as a seahorse-like architecture. Six Cas7 subunits (green) oligomerize along the spacer region (red) of crRNA, forming the helical backbone filament. One Cas6 subunit (partial circle with scissor, brown) remains bound to the 3′-stem-loop (repeat-derived, black) of the crRNA forming the "head" of the Cascade complex. The repeat derived 5′-tag (black) of crRNA is capped by a single Cas8 (large subunit, sky blue) and Cas5 (gray), shaping the "foot." Two Cas11 proteins (small subunits, blue) form the "belly" filament. **(B)** The seaworm-shaped type III effector complex is called Csm (in subtypes III-A/D/E/F) or Cmr (in subtypes III-B/C). Cas subunits are color coded as in panel (A). Moreover, the large subunit in Csm/Cmr complex is called Cas10. It is larger than Cascade's Cas8 and shows DNase activity during CRISPR interference. Also, the Cas7 subunits in Csm/Cmr complex exhibit RNase activity against target RNA. The number of Cas11 subunits varies (>2) among different type III subtypes. Csm/Cmr complex does not include Cas6 but contains an additional Cas7-like subunit(s) (Csm5 in Csm; Cmr1 and Cmr6 in Cmr, greenish-white) capping the 3′-end of crRNA. **(legend continued on next page)**

(C) The effector complex of type IV systems is ill-characterized. Here, the architecture of the type IV effector complex is displayed following a previous depiction by Liu and Doudna, 2020. The number of Cas11 subunits (five) is referred from a recent structural characterization of the type IV effector complex by Zhou et al., 2021. The effector complex in **(D)** type II (tracrRNA:crRNA-Cas9), **(E)** type V (crRNA-Cas12), and **(F)** type VI (crRNA-Cas13) systems encompass a single multidomain effector nuclease (Cas9, Cas12, and Cas13, respectively) complexed with a crRNA. Moreover, the effector complex in type II systems also includes a tracrRNA base paired with the repeat tag at the 3′-end of crRNA. Similarly, several type V subtypes (V-B/E/F/G) also encompass a tracrRNA in effector complex (not shown). Unlike other types, in type V and type VI systems, crRNA has a stem-loop (repeat-derived) structure at its 5′-end. In all panels, the scissor mark shows that the indicated Cas protein (except Cas6 in type I and IV system) exhibits nuclease activity against the target genome during the interference stage of CRISPR-Cas immunity. [perceived from (Hille et al., 2018; Liu and Doudna, 2020; Zhou et al., 2021)]

Typically, a Csm/Cmr complex contains a Cas5 family protein (Csm4/Cmr3) and a Cas10 protein (Csm1/Cmr2) at 5′-end of crRNA (foot), variable number of Cas7 family protein (4-6 Csm3/Cmr4) along the crRNA spacer sequence shaping the backbone, and >2 copies of Cas11 protein (Csm2/Cmr5) as belly filament. Also, the Cas7-like protein(s), a Csm5 or Cmr1 and Cmr6, occupies the 3′-end of crRNA, forming the head of the Csm/Cmr effector complex (Liu and Doudna, 2020; Makarova et al., 2015). Cas6 is not a subunit of the type III effector complex (Csm/Cmr) (Liu and Doudna, 2020; Makarova et al., 2015) (**Fig. 1.5B**).

Unlike the type I system, the Cas7 proteins in Csm and Cmr complex (Csm3 and Cmr4, respectively) are catalytically active and cleave the target RNA complementary to crRNA (Jia et al., 2019; Liu and Doudna, 2020; Osawa et al., 2015). The notable functional divergences among proteins of the Cas7 family emphasize the significance of experimental characterization of various Cas7 orthologs in different organisms. Further, a typical type III Cas10 protein contains an active HD nuclease domain and a palm domain (like in nucleotide cyclase and polymerase) (Osawa et al., 2015; Staals et al., 2014). The binding of target ssRNA by a Csm/Cmr complex allosterically activates the HD nuclease domain in Cas10 subunit that, in turn, cleaves the ss-DNA independent of sequence (Estrella et al., 2016; T. Y. Liu et al., 2017; Liu and Doudna, 2020). The palm domain in Cas10 subunit also gets activated upon RNA binding that catalyzes the synthesis of cyclic oligoadenylate (cOA) signaling molecules. The type III systems also encode ancillary nuclease effectors, Csm6 (III-A) and Csx1 (III-B), that contain CARF and HEPN domains. Upon Cas10 activation, the synthesized cOA binds to the CARF domain of accessory nuclease effector Csm6/Csx1, activating its HEPEN domain that cleaves RNA indiscriminately (Han et al., 2018; Kazlauskiene et al., 2017; Liu and Doudna,

2020). Very recently, the characterization of type III-E effector (Cas7-11) from *Desulfonema ishimotonii* revealed it to be unique among type III systems (Goswami et al., 2022; Makarova et al., 2020a; Özcan et al., 2021). Unlike the typical type I and type III effectors, Cas7-11 is a single polypeptide effector having four Cas7- and one Cas11-like segments and processes its own crRNA similar to the type VI effector (Cas13; described later) (Goswami et al., 2022; Özcan et al., 2021).

1.3.1.3. Type IV CRISPR-Cas systems

Type IV systems are least characterized among all class 1 CRISPR-Cas loci. The Csf2 protein, a highly diverged Cas7 variant, is considered the signature protein for type IV systems (Makarova et al., 2020a; Pinilla-Redondo et al., 2020). However, the presence of Csf1 protein (a small version of Cas8) was initially used to identify the type IV CRISPR-Cas loci (Makarova et al., 2015). Type IV systems are considered highly mobile as they are primarily encoded on plasmids. They lack adaptation modules and nuclease effectors; moreover, they are believed to share Cas proteins with co-existing type I CRISPR-Cas systems (Pinilla-Redondo et al., 2020). The type IV effector complex generally includes Csf1 (Cas8), Csf2 (Cas7), Csf3 (Cas5), and crRNA (Makarova et al., 2020a; Pinilla-Redondo et al., 2020; Zhou et al., 2021) (**Fig. 1.5C**). Five subtypes (IV-A to IV-E) of the type IV system have been identified. A few of the type IV subtypes also encode Cas6-like protein (Csf5), DinG helicase (Csf4), and CysH-like signal transduction protein (Makarova et al., 2020a; Pinilla-Redondo et al., 2020).

1.3.2. Class 2 CRISPR-Cas systems

Class 2 CRISPR-Cas systems exhibit a minimal architecture with a single multidomain effector nuclease that, alone in complex with crRNA, executes the interference of target nucleic acid. Based on the type of effector nuclease (i.e., Cas9, Cas12, and Cas13), class 2 systems are grouped into three types: type II, V, and VI, respectively (Makarova et al., 2020a).

1.3.2.1. Type II CRISPR-Cas systems

Cas9 is the signature protein for type II systems. Cas9 exhibits a bilobed structure having a recognition (REC) and a nuclease (NUC) lobe. During the interference stage, the REC lobe recognizes the target ds-DNA, and two nuclease domains of the NUC lobe [RuvC and HNH (histidine-asparagine-histidine)] cleave the target ds-DNA (Jinek et al., 2014). Based on

variations in loci architecture, three subtypes of type II systems have been identified: II-A to II-C (Makarova et al., 2020a). The adaptation proteins Cas1 and Cas2 are common in all type II systems; some variants also encode Cas4 (Makarova et al., 2020a; Nussenzweig et al., 2019). Interestingly, for processing of pre-crRNA into crRNA, type II systems employ an additional variant of small RNA called transactivating CRISPR RNA (tracrRNA) encoded within the same loci. For crRNA genesis, tracrRNAs base pair with repeats in pre-crRNA and duplex is stabilized by Cas9. Thereafter, site-specific cleavage in pre-crRNA by RNase III generates mature crRNA that remains base-paired with tracrRNA and complexed with Cas9, forming the effector complex tracrRNA:crRNA-Cas9 (Deltcheva et al., 2011; Jinek et al., 2012) (**Fig. 1.5D**).

1.3.2.2. Type V CRISPR-Cas systems

Type V systems encode the signature protein Cas12, a bilobed (REC and NUC) protein similar to Cas9 (Makarova et al., 2020a; Yan et al., 2019; Zetsche et al., 2015) (**Fig. 1.5E**). However, unlike Cas9, Cas12's NUC lobe lacks the HNH nuclease domain; instead, it has a RuvC endonuclease domain that cleaves both the strands of target ds-DNA (Swarts and Jinek, 2019; Yamano et al., 2017; Zetsche et al., 2015). The functional divergence of Cas12 proteins and variations in CRISPR-Cas loci led to the classification of type V systems into ten subtypes (V-A to V-I and V-U or V-K) (Makarova et al., 2020a; Yan et al., 2019). All the adaptation proteins (Cas1, Cas2, and Cas4*)* are present in subtypes V-A, V-B, V-F, and V-E. Subtypes V-C and V-D encode only the Cas1, while subtypes V-G, V-I, V-H, and V-U completely lack the adaptation Cas proteins; presumably share the adaptation module with other existing CRISPR-Cas loci (Makarova et al., 2020a; Yan et al., 2019). Cas12 directly processes the pre-crRNA or may also involve tracrRNA (Fonfara et al., 2016; Zetsche et al., 2015). Type V effector complex (crRNA-Cas12) typically cleaves target ds-DNA. In addition, the effector complex of several subtypes has been demonstrated to show collateral non-specific ss-DNase activity (Chen et al., 2018; Yan et al., 2019; Zetsche et al., 2015).

1.3.2.3. Type VI CRISPR-Cas systems

Cas13 is the signature protein for type VI systems which also exhibits a bilobed (REC and NUC) architecture like the Cas9 and Cas12 but functionally diverged (Abudayyeh et al., 2016; L. Liu et al., 2017b; Makarova et al., 2020a) (**Fig. 1.5F**). Unlike Cas9 and Cas12, the NUC

lobe of Cas13 contains two HEPEN domains that catalyze the RNA-guided cleavage of target RNA (Abudayyeh et al., 2016; L. Liu et al., 2017b). The Cas13-crRNA effector complex also shows the collateral non-specific ssRNase activity (Abudayyeh et al., 2016; L. Liu et al., 2017b). Cas13 alone can process its own crRNA (East-Seletsky et al., 2016; Zhang et al., 2019). Type VI systems are further characterized into four subtypes: VI-A to VI-D (Makarova et al., 2020a). The adaptation module Cas proteins are almost absent or not well distinguished in type VI systems; therefore, they are believed to be shared by other CRISPR-Cas loci (Hoikkala et al., 2021; Makarova et al., 2020a).

1.4. CRISPR adaptation

CRISPR adaptation is a complex multistep process wherein a selected portion (prespacer; precursor of spacer) of MGEs (primarily) is precisely incorporated into the CRISPR array as a new spacer. The acquired spacers serve as heritable infection memory data banks. Such adaptation ability to memorize the past infection of MGEs, certifies the CRISPR-Cas to be a legitimate adaptive immune system in prokaryotes. Naïve and primed adaptation are the two spacer acquisition mechanisms identified in various prokaryotes. When spacers are acquired from a MGE that a bacteria encounters for the first time, it is called Naïve adaptation. However, to fend off the CRISPR-Cas interference, phages mutate their protospacer regions, escaping the recognition by the CRISPR effector complex in the already immunized host. To counter this escape mutant phage infections, many bacterial hosts further acquire spacers from the mutated intruder genomes upon recognizing mismatched protospacers. This process of spacer acquisition is defined as primed adaptation, wherein the DNA fragments resulting from nucleolytic degradation of escape mutant intruder genomes by the CRISPR effector complex serve as the primary source of prespacer (Datsenko et al., 2012; Jackson et al., 2019; Semenova et al., 2016).

The established molecular understanding of the CRISPR adaptation mechanism is primarily gained from well-characterized type I-E and II-A systems (Hille et al., 2018; Sasnauskas and Siksnys, 2020). The adaptation process involves three subsequent events: selection/capture of the prespacer, its processing, and finally, integration into the CRISPR array (**Fig. 1.6**).

1.4.1. Selection of prespacer with self versus non-self discrimination

Detection and differentiation of intruder genomes for spacer acquisition during the adaptation stage is the key to avoiding autoimmunity by CRISPR effector complex. In a MGE, the DNA segment (s) bordering the PAM sequence (2-5 nt) qualifies to become a prespacer for acquisition into type I and II CRISPR array (Mojica et al., 2009; Wang et al., 2015; Yosef et al., 2012). Prespacers are processed and integrated into a CRISPR array without PAM sequence, thus averting the chance of self-targetting at the CRISPR locus by effector complex during interference. The level of PAM distribution in MGEs and the bacterial genome is estimated to be almost the same (Levy et al., 2015). Therefore, the PAM-directed preferential spacer acquisition mechanism alone is insufficient to differentiate between self and non-self-genomes. Research (based on data mining) suggests that around 90 % of the acquired spacer are processed from MGEs (Shmakov et al., 2017). Nonetheless, the differentiation of self and non-self-genome and preferential uptake of spacer from a foreign genome by CRISPR-Cas system is benefited by additional host mechanisms (e.g., RecBCD DNA repair) and discriminating features of MGEs (no/very less Chi sites and high rate of transcription/replication).

Prespacers are mostly sourced from the genomic regions prone to double-stranded breaks (DSBs) like replication origin and termination sites (Levy et al., 2015). In prokaryotes, such DNA damages are sensed by the RecBCD DNA repair system (in gram-negative bacteria) or its functional homolog AddAB in gram-positive bacteria (Dillingham and Kowalczykowski, 2008; Wigley, 2013). Activated RecBCD complex unwinds and cleaves the DNA until it encounters a Chi (Crossover hotspot instigator) site, an 8-nucleotide sequence that decelerates the RecBCD mediated cleavage (Dillingham and Kowalczykowski, 2008) (**Fig. 1.6A**). The produced DNA fragments by the RecBCD complex are suitable in size to capture as prespacers for CRISPR adaptation (Levy et al., 2015). Since Chi sites are abundantly present in the bacterial genome (per 4.6 kb), it limits the prespacer generation by RecBCD complex from self-DNA. In comparison, MGEs generally lack the Chi site; therefore, their unhindered degradation by RecBCD generates ample non-self prespacers (Levy et al., 2015; Modell et al., 2017). In addition, many MGEs enter the host cells in a linear form (exposed ds-DNA ends mimicking the DSBs), and usually, they replicate and transcribe rapidly (more prone to DNA damage). These features of MGEs make them an easy target of RecBCD nucleolytic cleavage

leading to an excess supply of non-self prespacers (over self-sequences) to CRISPR adaptation machinery for spacer acquisition (Levy et al., 2015; Modell et al., 2017).

The Cas1 and Cas2 proteins are the fundamental molecular players for the CRISPR adaptation across all the CRISPR-Cas systems (Yoganand et al., 2019; Yosef et al., 2012). The Cas1 and Cas2 together form a heterohexameric complex (2×Cas1-homodimer complexed with 1×Cas2-homodimer) called Cas1-2 integrase (**Fig. 1.6B**). The Cas1-Cas2 integrase senses the PAM sequence in a DNA fragment (dual-forked DNA pieces generated by RecBCD or by CRISPR effector complex during interference) and captures/binds it (Yoganand et al., 2019; Yosef et al., 2012). The binding by Cas1-Cas2 integrase authenticates a DNA fragment as a prespacer to be further processed and integrated into the CRISPR array. Apart from Cas1-Cas2 integrase, several CRISPR-Cas systems (I-A, I-B, I-C, I-G, II-B, V-A, V-B, V-E, and V-F) also employ Cas4 for identification and processing of prespacers with a cognate PAM sequence (Almendros et al., 2019; Lee et al., 2019, 2018; Makarova et al., 2020a) (**Fig. 1.6C**). In type II-A, the adaptation process essentially involves all the Cas proteins (Cas1, Cas2, and Cas9), Csn2 (auxiliary protein), and tracrRNA, wherein Cas9 is required for PAM identification (Heler et al., 2015; Wei et al., 2015) (**Fig. 1.6D**).

1.4.2. Prespacer processing and integration

Post-capturing of authentic prespacer by Cas1-Cas2 integrase (and additional Cas proteins), the bound prespacer is further processed into 20-40 bp sequence with the help of additional proteins that vary across different CRISPR-Cas systems (Ramachandran et al., 2020; Yoganand et al., 2019; Yosef et al., 2012). In type I-E, the prespacer is trimmed at both ends (PAM-proximal and distal) by host exonucleases such as ExoIII, ExoT, and DnaQ (Ramachandran et al., 2020; Yoganand et al., 2019) (**Fig. 1.6B**). In type II-A, the Csn2 protein trims the prespacer (Heler et al., 2015; Wei et al., 2015) (**Fig. 1.6D**). In other type I and type V systems, Cas4 proteins are identified to trim the PAM-proximal end of prespacers (**Fig. 1.6C**); moreover, which protein(s) trims the PAM-distal end is not sure yet (Almendros et al., 2019; Lee et al., 2019, 2018; Makarova et al., 2020a).

The mechanism of prespacer integration has mainly been studied in type I-E systems. Once the authentic prespacers are captured and processed, Cas1-Cas2 integrase catalyzes the prespacer integration into the CRISPR array. The prespacer integration by the Cas1-Cas2 complex is

directed to the leader-proximal repeat (1st repeat) by a host factor termed IHF (Integration host factor) (**Fig. 1.6**). IHF binds to the CRISPR leader sequence (~60 bp AT reach motif in *E. coli*), inducing a bend in DNA locally that directs the Cas1-Cas2 complex to integrate the prespacer at leader-repeat junction (Nuñez et al., 2016; Yoganand et al., 2017).

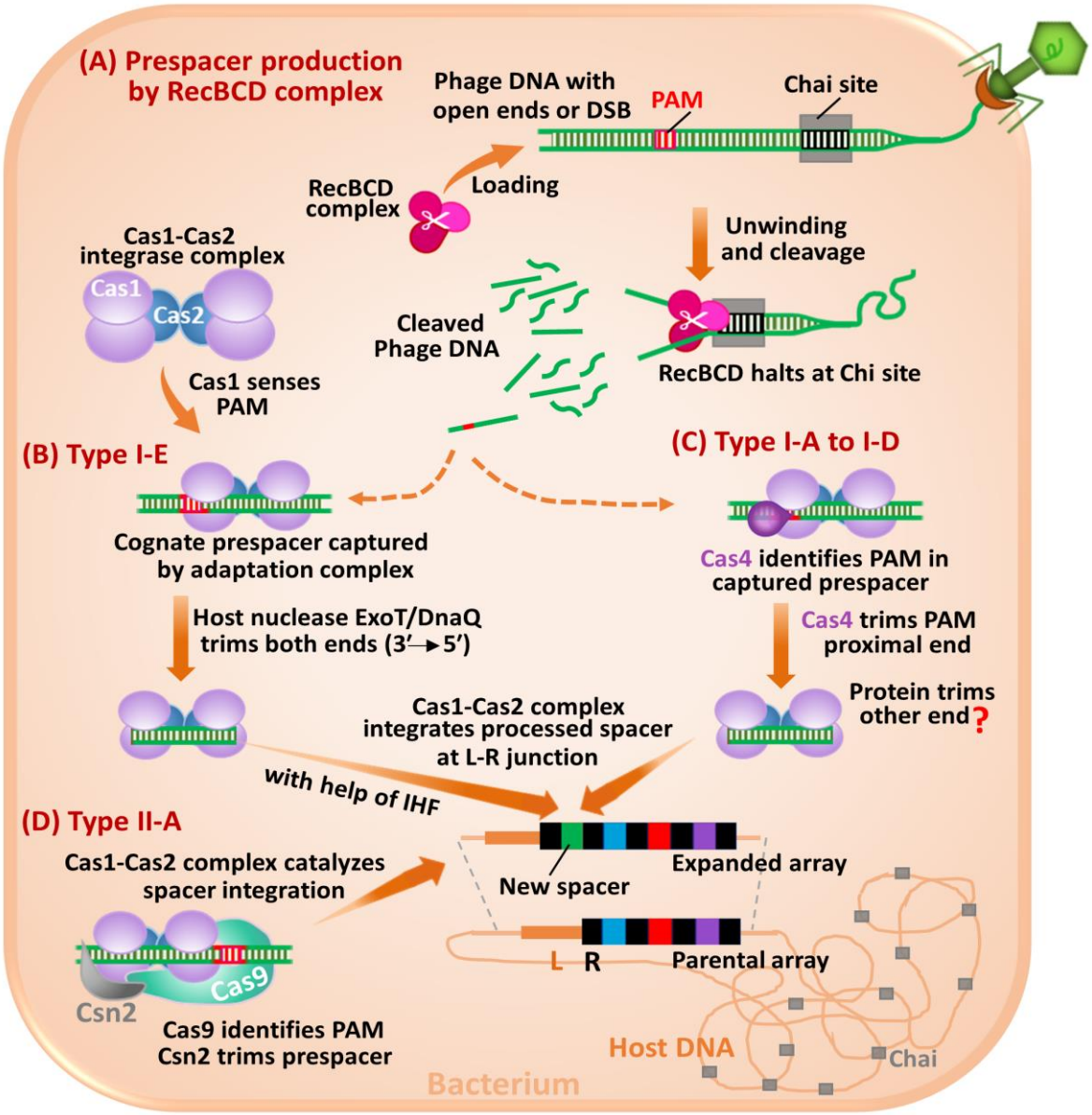

Fig. 1.6. Mechanism of CRISPR adaptation. CRISPR adaptation mechanism is primarily described based on well-comprehended type I-E and II-A systems. **(A)** During the first encounter with an intruder genome (phage DNA), the host RecBCD complex senses the open double-stranded end or double-stranded break (DSB) in actively replicating/transcribing phage DNA and cleaves it. This supplies the CRISPR-Cas system with ample non-self-DNA fragment(s) suitable for spacer acquisition (naïve adaptation). The rare occurrence of the Chi site in a phage genome permits its active degradation by RecBCD compared to cleavage of damaged host genome rich in Chi site. The Cas1 and Cas2 proteins are the fundamental molecular players for the CRISPR adaptation across all the CRISPR-Cas systems. **(legend continued on next page)**

(B) RecBCD degraded phage DNA fragment(s) with cognate PAM and appropriate size is captured by Cas1-Cas2 integrase complex (or adaptation complex). In the type I-E system, Cas1 recognizes the PAM sequence in a DNA fragment, authenticating it as a prespacer. Subsequently, host nucleases (ExoT/DnaQ) trim the captured prespacer at both ends in 3′ to 5′ direction to make it of suitable size for integration. **(C)** In other type I systems (I-A to I-D), Cas4 recognizes the PAM and trims the PAM proximal end of the prespacer. The protein(s) that are responsible for trimming the PAM distal end is still largely unknown. Once a prespacer is processed, the Cas1-Cas2 complex integrates it as a new spacer into the CRISPR array via a cut-paste mechanism, followed by a concomitant duplication of repeat sequence involving DNA polymerases and ligases. A CRISPR-unrelated protein, integration host factor (IHF), directs the new spacer integration by Cas1-Cas2 integrase precisely at the leader (L)-repeat (R) junction of the CRISPR array. **(D)** Spacer acquisition in type II-A system, besides Cas1-Cas2 integrase, requires Cas9 for PAM recognition and ancillary Csn2 protein for trimming of prespacer. [perceived from (Hille et al., 2018; Xiao et al., 2017b)]

The integration process starts with a nucleophilic attack of the 3'-OH group of the prespacer followed by ligation to 5'-phosphate of the leader-repeat junction, resulting in half-site integration. Successively, a second nucleophilic attack at the existing repeat-spacer junction causes full-site integration of the prespacer into the CRISPR array (Nuñez et al., 2015; Rollie et al., 2015). The integrated prespacer is called a spacer (**Fig. 1.6**). The correct orientation of integration is guided by the presence of partial/full PAM in the prespacer. However, after integration, only one nucleotide from PAM is retained in spacers (Shipman et al., 2016; Wang et al., 2015).

Differing from the type I systems, the adaptation process in type II systems does not require IHF. Instead, integration is dependent on a short (~5 nt) motif on the leader region termed as LAS (Leader anchoring site) (Wright and Doudna, 2016; Xiao et al., 2017b). LAS is directly recognized by the Cas1-Cas2 complex and is sufficient to integrate the prespacer in the correct orientation. Like the type I-E system, spacer integration starts with a 3'-OH nucleophilic attack at the LAS-repeat junction leading to half-site integration products. Subsequently, a second nucleophilic attack on the first repeat-spacer junction leads to full-site integration (Wright and Doudna, 2016; Xiao et al., 2017b). Despite multiple repeat-spacer units in a CRISPR array, a new prespacer is always integrated at the leader-repeat junction, followed by a concomitant duplication of the first repeat (McGinn and Marraffini, 2016; Yosef et al., 2013, 2012) (**Fig. 1.6**). Such a polarized integration mechanism maintains the chronology of MGE infections and ensures efficient interference response upon recurrent infection of most recently encountered MGEs (McGinn and Marraffini, 2016).

1.5. Biogenesis of crRNA and effector complex

crRNA biogenesis is elementary to the CRISPR-Cas defense response in a prokaryotic host for targeted cleavage of MGEs during the recurrent invasion. The process starts with transcription of the CRISPR locus into a single long precursor CRISPR RNA (pre-crRNA), which in turn enzymatically processed into multiple short mature crRNAs, each containing a complete spacer sequence flanked by leftover repeat parts (Brouns et al., 2008; Pougach et al., 2010).

The endoribonuclease that processes the pre-crRNA varies across the different CRISPR-Cas types. In most class 1 CRISPR-Cas types (I, III, and IV), Cas6 endoribonuclease catalyzes the crRNA genesis (**Fig. 1.7A-B**). Moreover, the mechanism of pre-crRNA processing by Cas6 varies depending on the structured/unstructured conformations of repeats in a pre-crRNA (Sefcikova et al., 2017). More than 50 % of the known repeat sequences exhibit palindromic motif that folds into a stable stem-loop structure in RNA. Such structured CRISPR repeats are categorized as canonical repeats. Rest CRISPR repeats display weak or no palindromic features, therefore, lack the characteristic stem-loop structure and are classified as non-canonical repeats (unstructured) (Kunin et al., 2007; Sefcikova et al., 2017).

In a pre-crRNA with canonical repeats, Cas6 typically cleaves the structured repeats at the 3′-end of the stem-loop, generating mature crRNAs, each comprising one complete spacer flanked by adjacent repeats derived 8 nt sequence at 5′-end (called 5′-handle or -tag) and stem-loop at 3′-end (Sefcikova et al., 2017). Cas6 proteins (except in type I-A, I-B, and III systems) exhibit single-turnover enzyme kinetics and remain bound with mature crRNA at 3′-stem-loop (**Fig. 1.7A**), shaping the head of the Cascade complex (Behler and Hess, 2020; Liu and Doudna, 2020; Sefcikova et al., 2017). In the type I-C system, which lacks the Cas6, catalytically active Cas5c protein supplants the functional role of Cas6 in crRNA genesis (Liu and Doudna, 2020; Nam et al., 2012a; Punetha et al., 2014). Moreover, Cas5c remains bound at 5′ handle of processed crRNA. Pre-crRNAs with non-canonical repeats (in type I-A, I-B, and III systems) are also processed by Cas6 protein but slightly differently. The repeats in type I-A, I-B, and III systems are mostly non-palindromic and lack the canonical stem-loop structure (Kunin et al., 2007; Sefcikova et al., 2017) (**Fig. 1.7B**). In these systems, Cas6 dimers induce the formation of a transient stem-loop structure in repeats required for their cleavage (**Fig. 1.7B**). Here, Cas6 acts as a multi-turnover enzyme and dissociates from the cleaved mature crRNA, and does not

form part of the effector complex (Richter et al., 2013; Sefcikova et al., 2017; Shao et al., 2016).

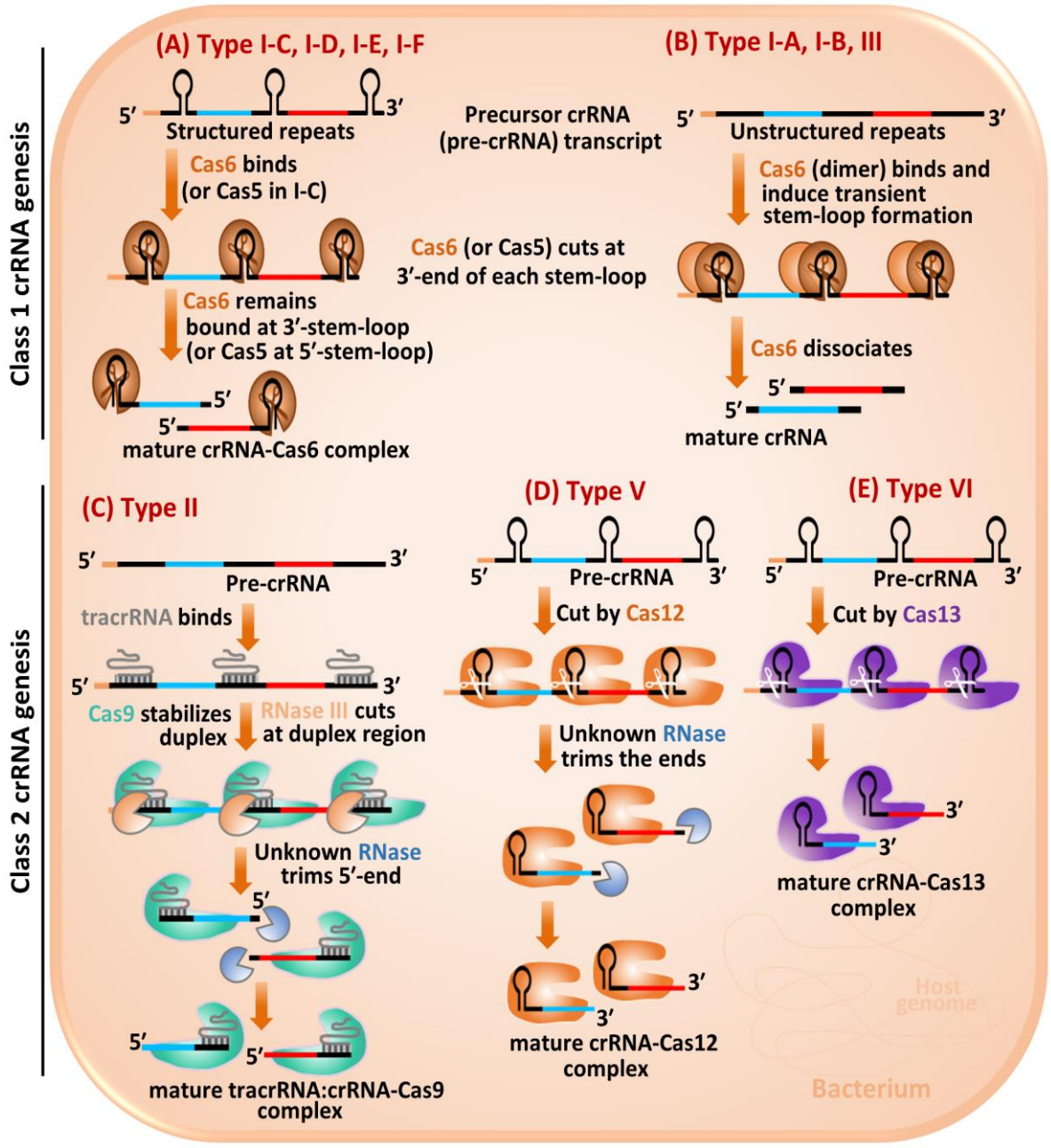

Fig. 1.7. Biogenesis of mature crRNA in Class 1 and Class 2 CRISPR-Cas systems. (A) crRNA maturation in type I-C, I-D, I-E, and I-F systems. In these systems, pre-crRNA contains palindromic repeat sequences (black), which form stem-loop structures. Cas6 (or Cas5 in I-C) binds to the structured repeats and cleaves at 3′-end of each stem-loop, generating mature crRNAs each containing a spacer flanked by left-over repeat sequences. Post-cleavage, Cas6 (or Cas5) remains bound to the mature crRNA. **(B)** crRNA genesis in type I-A, I-B, and type III systems. The associated pre-crRNA encompasses unstructured repeats. Here, the Cas6 dimer induces a transient stem-loop-like fold in repeats and cleaves them at 3′-end. Cas6 dissociates from the processed mature crRNA. **(legend continued on next page)**

(C) Pre-crRNA processing in type II systems. The unstructured repeats in pre-crRNA of type II systems hybridize with tracrRNAs, generating duplex regions. Cas9 stabilizes the crRNA:tracrRNA complex, and RNase III cleaves in the duplex regions towards the 3′-end. An unknown nuclease further trims off the repeat remnant from the 5′-end of crRNA, generating mature crRNA complexed with tracrRNA and Cas9. The resulting stable tracrRNA:crRNA-Cas9 complex acts as an effector complex during CRISPR interference. **(D)** crRNA maturation in type V systems. The depicted mechanism is of a type V-A system that encodes pre-crRNA with structured repeats. Other type V subtypes (V-B to V-G) engage an additional RNA (tracrRNA or scoutRNA) for pre-crRNA maturation. Cas12 cleaves the repeats at a few bases upstream from the 5′-end of the stem-loop. An unknown nuclease trims the 5′-end, generating mature crRNA having a 5′-stem-loop. Post-processing, Cas12 remains bound with Cas12 and acts as an effector complex in the CRISPR interference stage. **(E)** In type VI systems, Cas13 alone processes the pre-crRNA (structured repeats) into mature crRNA having a 5′-stem-loop. The resulting mature crRNA-Cas13 complex catalyzes the CRISPR defense reaction during the interference stage. [perceived from (Hille et al., 2018)]

On the contrary, the Cas6 of type I-B system in *Heloferax volcanii* (Brendel et al., 2014; Hochstrasser and Doudna, 2015) and *Leptospira interrogans* (Prakash and Kumar, 2021) shows stable binding to processed crRNA. In type IV systems, the Cas6-like protein, Csf5, catalyzes the pre-crRNA processing into mature crRNA (Özcan et al., 2018). The processed mature crRNA, in all the class 1 systems, allows relative assembly of Cas subunits to shape into an effector complex.

Unlike class 1, class 2 CRISPR-Cas systems (type II, V, and VI) employ the multidomain effector nuclease (Cas9, Cas12, and Cas13, respectively) for both crRNA biogenesis (**Fig. 1.7C-D**) and subsequent target interference (Koonin et al., 2017; Makarova et al., 2020a; Shmakov et al., 2015). Additionally, pre-crRNA processing in type II systems also requires RNase III and a trans-encoded small RNA termed trans-activating crRNA (tracrRNA; ~75 nt) that contains an anti-repeat motif (~25 nt) having partly complementary to the CRISPR repeats (Chyou and Brown, 2019). The tracrRNA complex makes complementary base pairing with the unstructured repeat sequence of pre-crRNA, forming a duplex region stabilized by Cas9 binding (**Fig. 1.7C**). This duplex is recognized by RNase III, that co-processes both CRISPR repeat and tracrRNA (Deltcheva et al., 2011). Resulting tracrRNA:crRNA-Cas9 complex (**Fig. 1.7C**) acts as an effector complex that catalyzes the targeted cleavage of intruder DNAs during the interference stage (Deltcheva et al., 2011; Jinek et al., 2012).

Like the type II system, the pre-crRNA repeats in most type V systems (V-B, V-E, V-F, and V-G) are unstructured, and crRNA maturation involves Cas12 effector nuclease and tracrRNA (Koonin et al., 2017; Makarova et al., 2020a; Shmakov et al., 2015). On the other hand, in the

type V-A system, repeats exhibit a stem-loop structure (Fonfara et al., 2016). Cas12a of type V-A recognizes the structured repeats in pre-crRNA and cleaves the repeats at a few bases upstream of the stem-loop (**Fig. 1.7D**), unlike at the base of the stem-loop by Cas6 in the case of class 1 systems (Dong et al., 2016; Fonfara et al., 2016). Subsequently, the intermediate crRNA is further trimmed at both ends by unknown host nuclease(s), forming the effector complex containing a Cas12a bound with a mature crRNA having a 5′-repeat derived stem-loop (Dong et al., 2016; Fonfara et al., 2016) (**Fig. 1.7D**). Similar to type V-A, type VI systems process the pre-crRNA (contain structured repeats) employing a multidomain Cas13 effector nuclease independent of tracrRNA (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Knott et al., 2017; Makarova et al., 2020a) (**Fig. 1.7E**).

1.6. CRISPR interference

The targeted degradation of intruder genomes is the primary purpose of every CRISPR-Cas system encoded within bacteria. During the interference stage, the infection records stored in a CRISPR locus are utilized as crRNA in complex with Cas protein(s) (effector complex) for specific recognition and degradation of invader genomes engaging Cas effector nuclease. Owing to the diverse composition of effector complexes in various CRISPR-Cas systems, the mechanism of CRISPR interference also varies across the different CRISPR-Cas types (Liu and Doudna, 2020; Makarova et al., 2020a).

1.6.1. Class 1 CRISPR-Cas interference

Class 1 CRISPR-Cas systems (I, III, and IV) employ a multiprotein-crRNA effector complex to interfere with intruder genomes (**Fig. 1.8A-B**). In type I CRISPR-Cas systems, the multiprotein-crRNA effector complex "Cascade" identifies the non-self-DNA for interference via a two-factor authentication practice. The first factor is the strong complementarity between crRNA and the target sequence (protospacer) in invading genome. The second is the presence of a cognate PAM sequence immediately adjacent to the protospacer (Gleditzsch et al., 2019; Leenay and Beisel, 2017; Marraffini and Sontheimer, 2010). Conventionally, the trinucleotide sequence present at the immediate 5′-end of the protospacer serves as a functional PAM for type I systems, and it is highly conserved for a particular Cascade (Fischer et al., 2012; Gleditzsch et al., 2019; Leenay and Beisel, 2017; Rollins et al., 2015). PAM is thus explicitly identified by a Cascade complex in a foreign genome, thereby deterring the auto-target of the

host's DNA at CRISPR loci that lack the PAM sequence (Gleditzsch et al., 2019; Hayes et al., 2016; Leenay and Beisel, 2017; Marraffini and Sontheimer, 2010). The Cas8 (or Cas10 in type I-D) subunit (LS) of Cascade identifies the cognate PAM on MGEs. The high sequence diversity among Cas8 orthologs across the CRISPR-Cas subtypes ensures the recognition of specific PAM by a Cascade complex for efficient interference (Cass et al., 2015a; Leenay and Beisel, 2017; Rollins et al., 2015; Xiao et al., 2017a). The recognition of a cognate PAM on MGEs by Cascade complex via its Cas8 subunit induces a local unwinding of target DNA, facilitating the complementary base pairing of crRNA with adjacent protospacer (Hayes et al., 2016; Xiao et al., 2017a) (**Fig. 1.8A**). A tight base paring of crRNA with perfectly matching protospacer in MGEs forms a three-stranded DNA-RNA hybrid structure termed R-loop. In an R-loop, the protospacer's target strand base paired with crRNA is firmly supported by the Cas7 backbone, and the displaced non-target strand is stabilized by Cas11 subunits (SS), preventing the collapse of the R-loop (Hayes et al., 2016; Xiao et al., 2017a). The formation of a stable R-loop confirms the recognition of bonafide target DNA by the Cascade complex for interference (**Fig. 1.8A**).

Once the Cascade complex binds a bonafide target DNA, certain structural rearrangements in the complex enable the recruitment of Cas3 helicase/nuclease effector protein onto the single-stranded non-target strand of R-loop (Hayes et al., 2016; Nimkar and Anand, 2020; Xiao et al., 2017a). Upon recruitment, Cas3's HD (histidine-aspartate) nuclease domain makes an initial nick onto the non-target strand in the R-loop that helps properly position the Cas3 helicase domain (Nimkar and Anand, 2020; Redding et al., 2015; Xiao et al., 2017a). Subsequently, the Cas3's helicase domain continuously unwinds the target ds-DNA in a 3′ to 5′ direction, and the HD nuclease domain cleaves the free non-target strand (Liu and Doudna, 2020; Loeff et al., 2018; Nimkar and Anand, 2020; Redding et al., 2015) (**Fig. 1.8A**). Moreover, how a Cas3 cleaves the target strand of DNA has not been well understood yet. It is speculated that post-degradation of the non-target strand, the exposed single-stranded target strand is degraded by a second Cas3 molecule (Liu and Doudna, 2020).

Like type I systems, type III systems also employ a multiprotein-crRNA effector complex (Csm or Cmr complex) for targeted interference of MGEs; however, the mechanism of CRISPR interference differs greatly from type I systems (Liu and Doudna, 2020) (**Fig. 1.8B**). Unlike type I systems, type III systems can target both RNA and DNA and do not need a PAM

sequence on non-self-DNAs to recognize and cleave them (Makarova et al., 2020a; Osawa et al., 2015; Samai et al., 2015; Staals et al., 2014). The effector complex employs a different set of molecular information to avert auto-targeting at the CRISPR locus of the host genome.

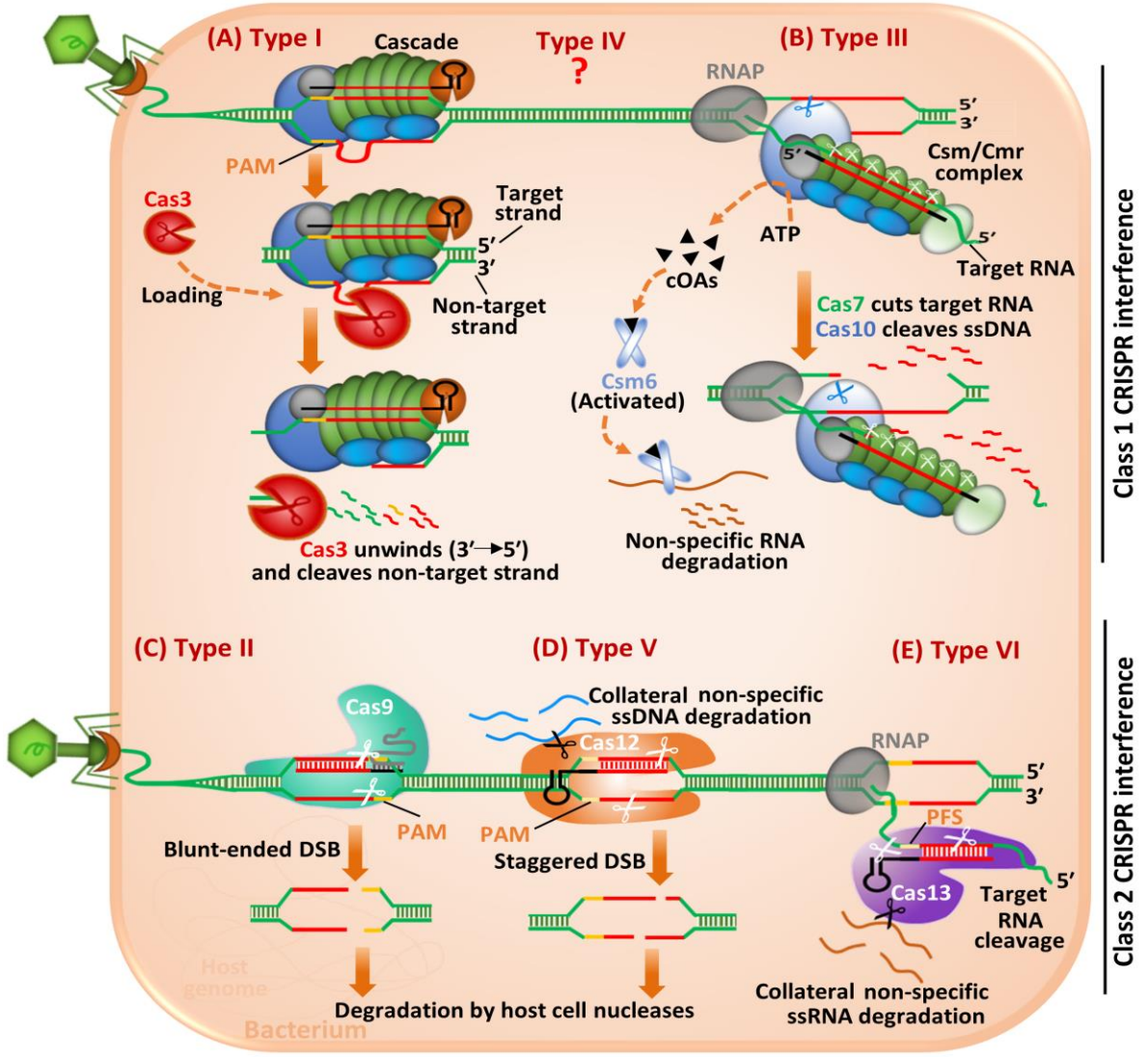

Fig. 1.8. The interference mechanism of CRISPR-Cas systems. (A) Type I CRISPR-Cas interference. The Cascade complex (via Cas8 subunit) recognizes the PAM on the intruder phage DNA. Thereafter, the base pairing of crRNA with the target strand of the protospacer induces R-loop formation. Cas3 binds to the displaced single-stranded non-target strand of the R-loop and starts cleaving it while moving in 3′ to 5′ direction via helicase/nuclease activity. The mechanism of interference by CRISPR-Cas type IV systems is not well-comprehended yet. **(B)** Type III CRISPR-Cas interference. The effector complex (Csm or Cmr), via crRNA complementary base-pairing, binds to the protospacer on the target RNA during active transcription of phage DNA. Cas7 subunits (Csm3 or Cmr4) cut the RNA transcript. The target binding activates the Cas10 to cleave the exposed ss-DNA at the transcribing R-loop region. Target binding also triggers Cas10 to catalyze the conversion of ATP into cyclic oligoadenylates (cOAs), which activate ancillary Csm6 nuclease to degrade RNA indiscriminately. **(legend continued on next page)**

(C) Type II CRISPR-Cas interference. A single protein effector complex, tracrRNA:crRNA-Cas9, catalyzes the target DNA interference. Correct PAM recognition by Cas9 and perfect base-pairing of crRNA with protospacer on phage DNA induces R-loop formation. Cas9 nicks both strands, resulting in a blunt end double-stranded break (DSB) on target phage DNA which is further degraded by host nucleases. **(D)** Type V CRISPR-Cas interference. The depicted mechanism is based on the well-studied type V-A system. Post recognition of a cognate PAM on the phage DNA and perfect base-pairing of crRNA with protospacer, Cas12 cleaves the target DNA. Cleavage by Cas12 produces a staggered DSB on phage DNA, making it prone to degradation by host nucleases. Target binding also triggers a collateral ss-DNase activity in Cas12, leading to ss-DNAs degradation non-specifically. **(E)** Type VI CRISPR-Cas interference. The type VI effector complex (crRNA-Cas13) catalyzes an RNA-guided RNA interference. Cas13 binds to the crRNA matching-target RNA region next to a protospacer flanking sequence (PFS) on the phage RNA transcript and degrades it. Target RNA binding also unleashes the collateral ss-RNase activity of Cas13, resulting in the ss-RNA degradation promiscuously. RNAP: RNA polymerase. [perceived from (Hille et al., 2018; Liu and Doudna, 2020)]

During the maturation of crRNA, a leftover repeat sequence remains on either end of the spacer sequence. The cognate protospacer on intruder genomes does not contain any adjacent sequences that resemble repeat. Therefore, only the spacer-derived region in a crRNA shows complementarity with protospacer on intruder genomes. Whereas, at the spacer sites on the CRISPR array, the complementarity is extended towards remnant repeat residues on mature crRNA. Such base pairing at the 5′-end of crRNA unlicensed the effector complex to proceed for interference, protecting the host from an autoimmune response (Marraffini and Sontheimer, 2010).

The molecular understanding of type III CRISPR interference is referred mainly to the most common and best-studied type III-A and III-B systems. Type III-A and III-B utilize a Cascade-like effector complex termed Csm and Cmr, respectively, for target identification and interference (Liu and Doudna, 2020) (**Fig. 1.8B**). Unlike nucleolytically inactive Cascade complex, Csm and Cmr complex recognize and degrade the target genomes via their intrinsic nuclease activity. Once a cognate target RNA is bound by the Csm/Cmr complex, the catalytically active Cas7 subunits (Csm3/Cmr4) cleave the target RNA at every sixth base (Liu and Doudna, 2020; Osawa et al., 2015; Samai et al., 2015) (**Fig. 1.8B**). However, the degradation of DNA by a Csm/Cmr complex is conditional and occurs only when it is in the transcribing state. The binding of the Csm/Cmr complex to target RNA during transcription activates the HD nuclease domain in the Cas10 subunit (LS) that cleaves the template ss-DNA (Elmore et al., 2016; Estrella et al., 2016; Kazlauskiene et al., 2016; Liu and Doudna, 2020; Osawa et al., 2015; Samai et al., 2015) (**Fig. 1.8B**). Additionally, upon RNA binding, the palm domain in Cas10 subunit also gets activated and catalyzes the synthesis of cyclic oligoadenylate

(cOA) signaling molecules. Subsequently, cOAs activate the ancillary nuclease effectors, Csm6 (III-A) and Csx1 (III-B), that indiscriminately cleave host and viral RNA, leading to cell growth arrest (Han et al., 2018; Kazlauskiene et al., 2017; Liu and Doudna, 2020) (**Fig. 1.8B**). In the type IV system, an effector complex generally includes Csf1 (Cas8), Csf2 (Cas7), Csf3 (Cas5), and crRNA (Makarova et al., 2020a; Pinilla-Redondo et al., 2020), but effector nuclease protein has not been identified. For type IV systems, the mechanism of CRISPR interference is yet to be understood.

1.6.2. Class 2 CRISPR-Cas interference

Class 2 CRISPR-Cas systems (II, V, and VI) develop a single multidomain protein-crRNA effector complex to interfere with intruder genomes (Koonin et al., 2017; Makarova et al., 2020a). The model for interference mechanism in type II systems is referred from the extensively studied type II-A system of *Streptococcus pyogenes* (Jinek et al., 2014, 2012; Nishimasu et al., 2015, 2014). In type II systems, the effector complex comprises Cas9 nuclease, crRNA, and a tracrRNA (**Fig. 1.5D**). Cas9 recognizes the cognate PAM sequence at the 3′-end of the non-target strand (Leenay and Beisel, 2017) and facilitates the base-pairing of crRNA with adjacent protospacer. A perfect complementary base pairing of crRNA forms a stable R-loop structure, leading to the activation of Cas9 nuclease domains (Jinek et al., 2014, 2012; Nishimasu et al., 2015, 2014) (**Fig. 1.8C**). Subsequently, well-poised HNH and RuvC nuclease domains of Cas9 cut the target and non-target strands of protospacer, respectively. This results in a blunt double-stranded DNA break lethal for DNA stability, and thus type II effector complex (Cas9-crRNA-tracrRNA) confers interference of the intruder genome (Jinek et al., 2014, 2012; Nishimasu et al., 2015, 2014) (**Fig. 1.8C**).

Interference in type V systems is catalyzed by the crRNA-Cas12 effector complex (Koonin et al., 2017; Makarova et al., 2020a) (**Fig. 1.8D**). A few type V effector complexes (V-B, V-C, V-G) also contain tracrRNA (Koonin et al., 2017; Makarova et al., 2020a; Yan et al., 2019). The single RuvC nuclease domain in the Cas12's NUC lobe nicks both the target and non-target strands of protospacer DNA having cognate 5′-PAM sequence (Dong et al., 2016; L. Liu et al., 2017a; Shmakov et al., 2015; Yamano et al., 2017; Yan et al., 2019). Unlike blunt-ended cleavage by Cas9 in type II systems, Cas12 nuclease activity induces sticky DSB (5-7 nt overhangs) in the crRNA-targeted alien ds-DNA (Dong et al., 2016; L. Liu et al., 2017a; Shmakov et al., 2015; Yamano et al., 2017) (**Fig. 1.8D**). In addition, upon binding to target ds-

DNA, effector complex of several type V systems (Cas12a/V-A, Cas12b/V-B, and Cas12c/V-C) also display non-specific collateral ss-DNase activity (Chen et al., 2018; L. Li et al., 2019; Yan et al., 2019) (**Fig. 1.8D**).

Type VI systems employ Cas13 effector nuclease that exclusively targets the RNA, unlike the Cas9 (type II) and Cas12 (type V) that interfere with DNA molecules (Koonin et al., 2017; Makarova et al., 2020a; Shmakov et al., 2015). The crRNA-Cas13 effector complex senses a PAM-like motif termed PFS (protospacer flanking sequence) in alien RNA to discriminate from self-transcripts (Leenay and Beisel, 2017). Once a crRNA matching target-ssRNA with cognate 3′-PFS is identified, the Cas13's HEPEN domains cleave the targeted RNA (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Koonin et al., 2017) (**Fig. 1.8E**). Similar to Cas12, once bound with crRNA-targeted RNA, Cas13 exhibits promiscuous collateral ssRNase activity that causes degradation of host transcripts, leading to cell dormancy or suicide (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Koonin et al., 2017) (**Fig. 1.8E**).

1.7. Applications of CRISPR-Cas systems

Owing to crRNA navigated specificity and the requirement of a single endonucleolytic Cas protein to interfere with a target nucleic acid, the class 2 CRISPR-Cas systems (type II/Cas9, V/Cas12, and VI/Cas13) are being exploited as targeted genetic manipulation tool (Aquino-Jarquin, 2019; Cox et al., 2017; Moon et al., 2019; Murugan et al., 2017). Ten years ago, for the first time, a type II CRISPR/Cas9 system from *Streptococcus pyogenes* was illustrated as a programmable DNA cleavage tool (Jinek et al., 2012). Heretofore, the number of class 2 CRISPR/Cas systems from different bacteria has been extensively harnessed as genetic-manipulation tools exploiting its ability to bind and cleave specific DNA or RNA sequences (Aquino-Jarquin, 2019; Cox et al., 2017; Moon et al., 2019; Murugan et al., 2017). Moreover, Cas9 from *S. pyogenes* is the most extensively exploited Cas endonuclease for targeted genetic manipulation (Jinek et al., 2012). Typically, Cas9-based genome editing functions by creating a double-stranded break (DSB) on crRNA-specified target DNA that consequently induces the host cell's endogenous DNA repair systems (Jiang and Marraffini, 2015). DNA damage (DSB) is repaired by either of two mechanisms: non-homologous end-joining (NHEJ) and homology-directed repair (HDR) (Jiang and Marraffini, 2015). Repair by the NHEJ pathway is error-prone, resulting in unpredictable small indel mutation (insertion or deletion) at the cleavage site, producing functional gene knockouts. On the other hand, the HDR path relies on

homologous recombination, thus generating precise alteration at the DSB site upon furnishing the homologous DNA template (knockin, knockout, or base editing) (Jiang and Marraffini, 2015).

Since its advent, the CRISPR-Cas9 system has been rapidly optimized into a handy, fast, and easily customizable genome editing tool. To simplify the CRISPR-Cas9-based sequence targeted tools, the crRNA and tracrRNA have been fused to a single chimeric RNA called single guide RNA (sgRNA) (Jinek et al., 2012). Also, Cas9 has been engineered into different variants for various genetic manipulation purposes (Xu and Qi, 2019). For example, the catalytically dead Cas9 (dCas9) has been generated that can bind but not cleave the target DNA, thus, causing a steric hindrance to RNA polymerase, resulting in gene silencing (Xu and Qi, 2019). This dCas9-based gene silencing strategy is termed as CRISPRi (CRISPR interference). More examples include: dCas9 fused with deaminase enzyme (e.g., dCas9-cytidine deaminase) for gene-targeted base editing (Komor et al., 2016) and dCas9 functionalized with transcriptional or epigenetic effector proteins (e.g., dCas9-VP64 and dCas9-HAT) for gene regulation studies (Xu and Qi, 2019). Thus, owing to the easily programmable crRNA-mediated specificity and need for a single effector protein to achieve DNA targeting, class 2 CRISPR-Cas systems-based toolkits have become a preferred choice for genetic manipulation in a broad range of organisms, especially eukaryotes (Knott and Doudna, 2018; Zhang, 2019). Besides being evolved as a method of choice for genetic manipulation, the various CRISPR-Cas systems have now been creatively exploited to develop advanced biomaterials and diagnostics for disease management, food surveillance, and environmental monitoring (English et al., 2019; Kaminski et al., 2021; Y. Li et al., 2019).

However, the applications of class 2 CRISPR-Cas systems in prokaryotes are limited, largely because of the exceptional diversity of DNA homeostasis in microorganisms. Many prokaryotes lack NHEJ machinery for DNA repair; hence, effector protein (Cas9) induced DSB becomes cytotoxic (Shuman and Glickman, 2007; Xu et al., 2021). Notably, the class 2 CRISPR-Cas systems represent only ~10 % of all CRISPR-Cas systems identified so far, and the remaining 90 % belongs to the class 1 system (type I, III, and IV) (Makarova et al., 2017, 2015). Among class 1 systems, type I is the predominant type in both eubacteria and archaea, comprising 50 % of all CRISPR-Cas systems identified up to recently (Makarova et al., 2017, 2015; Xu et al., 2021). Type I systems with such a wide presence in prokaryotes hold immense

potential to be harnessed for endogenous genetic manipulation, especially in genetically recalcitrant bacteria. On the same line, in recent years, the repurposing of type I systems as an efficient endogenous genome editing tool has been demonstrated in a few bacteria (Xu et al., 2021). For example, Li and coworkers have demonstrated the gene deletion in *Sulfolobus islandicus* via repurposing the endogenous type I-A CRISPR-Cas system (Li et al., 2016). They transformed the *S. islandicus* with an editing plasmid that codes for target-specified pre-crRNA and donor DNAs with homology with sequence flanking the genomic target site. Thereby, successful homologous recombination disrupts the target sequence on the chromosome, leading to no CRISPR interference of self-genome, resulting in the survival of only mutant cells (target sequence deletion). While in another case (no recombination), wild-type cells could not survive due to self-targeting, allowing markerless selection of mutant cells as well (Li et al., 2016).

1.8. *Leptospira*: a bacterial genus of human and animal health importance

Leptospira is a 6-12 µm long, thin (~0.1 µm diameter), highly motile spirochete bacteria with a helical cell body (~0.7 µm helix pitch; ~0.2 µm helix diameter) and two periplasmic flagella (~3 µm in length) (Nakamura, 2020; Picardeau, 2020). Unlike other spirochetes, leptospires have a characteristic hook or question-mark-shaped ends (**Fig. 1.9**). Leptospires do not take up the Gram stain and, in the absence of any stain, they are best visualized under a dark-field microscope (Nakamura, 2020; Picardeau, 2020). Leptospires are slow-growing (fastidious) obligate aerobes in both liquid and solid media with a 6-20 h generation time at an optimal growth temperature of 28-29 °C (Adler and de la Peña Moctezuma, 2010; Murray et al., 2009). On solid agar medium, leptospires grow as subsurface colonies. The most widely used medium is based on oleic acid, 1 % bovine serum albumin (BSA), and polysorbate (Tween 80- a source of long-chain fatty acids) medium EMJH (Ellinghausen-McCullough-Johnson-Harris) (Ellinghausen and McCullough, 1965; Johnson and Harris, 1967).

The genus *Leptospira* is highly diverse. Since the first isolation (*L. interrogans*) in 1915 (Inada et al., 1916), a total of 64 different species have been identified by 2019 (Vincent et al., 2019). In a recent study, phylogenetic analysis based on whole-genome sequences (previously based on 16S rRNA gene sequencing) clustered the genus into two major clades: saprophytes (clade S) and pathogens (clade P) (Vincent et al., 2019). Clade "saprophytes" groups the *Leptospira* species, which are isolated from the natural environment and do not cause infections. The clade

"pathogens" comprises all the species responsible for infections in humans and/or animals and environmental species with undefined virulence status. Based on serological classification (agglutinating lipopolysaccharide antigens), *Leptospira* species are classified into 20 serogroups and over 300 serovars (Picardeau, 2013). The seven main pathogenic *Leptospira* species are *L. interrogans, L. borgpetersenii, L. santarosai, L. noguchii, L. weilii, L. kirschneri,* and *L. alexanderi* (Ahmed et al., 2006).

Pathogenic leptospires are the etiological agent of the zoonotic disease “Leptospirosis” in humans and animals. Rodents are the primary reservoir of pathogenic leptospires as the bacteria asymptomatically colonize the proximal renal tubules of the infected animals. Leptospires are disseminated in the environment through the urine of infected animals, and thus disease transmission to humans and other animals occurs mostly by direct contact with urine or indirectly via contaminated soil, water, and food (Bharti et al., 2003; Rojas et al., 2010)(Ko et al., 2009) (**Fig. 1.9**). Humans get infected with *Leptospira* accidentally (Bharti et al., 2003; Ko et al., 2009; Rojas et al., 2010). Leptospires penetrate the host body through mucous membranes, conjunctiva, or abraded skins, enter the bloodstream, and rapidly disseminate to multiple tissues and organs, including kindly, liver, and lungs (Ko et al., 2009; Matsunaga et al., 2007; Plank and Dean, 2000; Rojas et al., 2010).

The leptospirosis infection is mostly caused by serovars of *L. interrogans* and *L. borgpetersenii*, and the clinical manifestation in both cases are similar, though severity may vary. However, the modes of transmission of the two species differ: *L. interrogans* is usually acquired from contaminated water, while *L. borgpetersenii* is transmitted directly from host to host (Bulach et al., 2006). The clinical manifestation in humans is variable, the incubation period can range from 2-30 days with an average of 5-14 days post-exposure, and illness can last for months (Ko et al., 2009; Plank and Dean, 2000). The disease ranges from a mild, flu-like illness (headache, fever, diarrhea, and/or myalgia) to a severe and fatal state resulting in lung hemorrhage, jaundice, multiple organ failure (including kidney and liver), pulmonary distress, and death (the classical Weil's disease) (Adler and de la Peña Moctezuma, 2010; Ko et al., 2009; Levett, 2004).

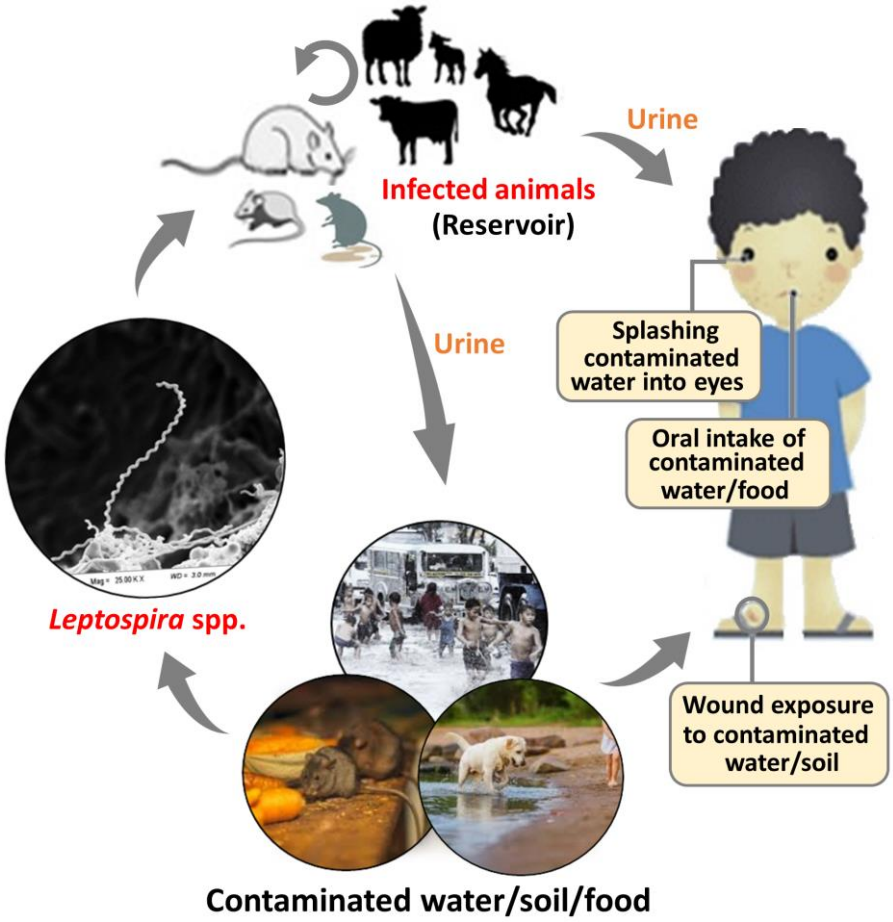

Fig. 1.9. The infection cycle of pathogenic *Leptospira* species. Leptospires are disseminated in the environment through the urine of infected mammalian species. Thus, the disease (leptospirosis) transmission to humans and other animals occurs mostly by direct contact with urine or by accidental exposure to water, soil, or food contaminated with the urine of reservoir animals (primarily rodents). [perceived from (Ko et al., 2009); Digital images from google image; *Leptospira*'s scanning electron micrograph from stock images available in Prof. Manish Kumar's laboratory, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India]

More than one million severe human cases of leptospirosis are reported every year (Adler and de la Peña Moctezuma, 2010). The disease is often misdiagnosed in humans since its symptoms mimic other illnesses (e.g., dengue fever, influenza, hepatic disease, Hantavirus infections), which contributes to underestimating the incidence of leptospirosis (Hartskeerl et al., 2011; Victoriano et al., 2009). For treatment, antibiotics doxycycline, ampicillin, azithromycin, or amoxicillin are prescribed in case of mild leptospirosis (Griffith et al., 2006). For severe leptospirosis, intravenous penicillin G has long been the drug of choice. Porcine, bovine, and canine vaccines are available against *Leptospira* infections. Human vaccines are available in some countries containing killed spirochetes of that region's most prevalent serovar (Barazzone et al., 2022; Bashiru and Bahaman, 2018; Levett, 2001; Wang et al., 2007).

Leptospirosis has a worldwide distribution, though the disease burden is largely on the improvised population in developing countries under tropical areas and is primarily associated with poor sanitary conditions (Costa et al., 2015; Torgerson et al., 2015). Nonetheless, there is a rise in reported leptospirosis cases in developed countries with temperate climates (Pijnacker et al., 2016; Traxler et al., 2014). It is estimated that there are 1.03 million leptospirosis cases and approximately 60,000 human deaths every year globally, based on a systematic review of 80 published morbidity and mortality studies and databases from 34 countries with annual morbidity of 14.8 per 100,000 population (Costa et al., 2015; Torgerson et al., 2015). Also, leptospirosis-induced reproductive ailments (abortion, stillbirths, or weak offspring birth) in livestock cause substantial economic loss each year (Ellis, 2015).

Though leptospirosis has been known for decades and is considered a major public health issue, the paucity of literature on the disease burden and erratic clinical manifestations has made it a highly underreported disease. Multiple epidemics have been reported due to natural disasters and poor sanitary conditions in developing countries (Holla et al., 2018; Victoriano et al., 2009). The disease has been endemic in East and Southeast Asia, Australia, Oceania, and South America (Miraglia et al., 2013; SMITH et al., 2013; Zhang et al., 2012). Nonetheless, in recent years, occasional cases and outbreaks of human leptospirosis have often been reported from many European countries, North America, and Africa, where leptospirosis is now being considered an emerging or re-emerging infectious disease (Biggs et al., 2013; Goris et al., 2013; Traxler et al., 2014). Estimates of Disability Adjusted Life Years (DALYs) suggest that globally around 2.90 million DALYs are lost per annum from the approximately annual 1.03 million cases (Costa et al., 2015; Torgerson et al., 2015). Tropical regions of South and Southeast Asia, the Western Pacific, Central and South America, and Africa had the highest estimated leptospirosis disease burden (Torgerson et al., 2015).

In India, leptospirosis is a major health problem related to the monsoons and poor sanitary conditions, with multiple epidemics reported in recent years (Chaudhry et al., 2002; Holla et al., 2018; Kembhavi et al., 2021; Pappas et al., 2008; Sethi et al., 2010; Shukla et al., 2022). Most outbreaks of leptospirosis in India are reported from the coastal regions of Gujarat, Maharashtra, West Bengal, Orissa, Kerala, Tamil Nadu, Karnataka, and the Andaman Islands (Holla et al., 2018; Kembhavi et al., 2021; Sethi et al., 2010). Annually, >500 human leptospirosis cases are reported in Kerala, Tamil Nadu, Maharashtra, and Gujarat (Gautam et

al., 2021). Leptospirosis is the leading cause of mortality among infectious diseases prevalent in Kerala (James et al., 2018).

Therefore, the *Leptospira* species are of medical and veterinary importance. Yet, the biology and pathogenesis mechanism of *Leptospira* is not well understood because the genus, in general, suffers from the lack of efficient, targeted genetic manipulation tools (Fernandes et al., 2021; Picardeau, 2015). The development of more efficient genetic tools is required for functional genomic analysis of *Leptospira* to understand its biology in order to treat and prevent the infection (Fernandes et al., 2021; Picardeau, 2015).

1.9. Rationale of the present study

Leptospirosis presents a severe problem accounting for over one million human cases, with approximately 60,000 deaths every year globally (Costa et al., 2015; Torgerson et al., 2015) and a substantial economic loss due to leptospirosis-induced reproductive disorders in livestock (Ellis, 2015). Despite the significant global burden of disease and its economic importance, *Leptospira*'s biology and pathogenesis mechanism have not been well comprehended yet. Reportedly, pathogenic species of *Leptospira* are recalcitrant for genetic manipulation, making their virulence study difficult with a reverse genetic approach (Fernandes et al., 2021; Picardeau, 2015). Pathogenic leptospires, including the *L. interrogans* serovar Copenhageni strain Fiocruz L1-130, show incompetence to most of the available genetic manipulation techniques, primarily due to poor transformation/conjugation compatibility of the genus (Fernandes et al., 2019; Picardeau, 2015). The presence of the CRISPR-Cas system is inferred to be one of the conceivable justifications behind obstacles being confronted in the genetic manipulation of the pathogenic *Leptospira* (Fouts et al., 2016; Xiao et al., 2019a). Analysis of sequenced *Leptospira* genomes has revealed the presence of CRISPR-Cas systems (type I-B, I-C, and I-E) in species of pathogenic clade but not in saprophytes, suggesting their role in virulence as well (Fouts et al., 2016; Kira S. Makarova et al., 2011; Xiao et al., 2019a). However, recently, a metagenomics study has identified the type V CRISPR-Cas system in saprophytic *L. biflexa* as well (Martínez Arbas et al., 2020).

Up until now, the techniques that have primarily been used for functional genomic studies of pathogenic leptospires include random transposon insertion or homologous recombination-based targeted gene silencing (Fernandes et al., 2019; Murray et al., 2009; Pětrošová and

Picardeau, 2014; Slamti and Picardeau, 2012), although both show very low efficiency in virulent *Leptospira*. Further, the Cas9-induced DSB in the bacterial genome has been reported to be lethal in *Leptospira* due to the lack of NHEJ DNA repair machinery (Fernandes et al., 2019; Nascimento et al., 2004). Moreover, very recently, the targeted genetic manipulation in saprophyte *L. biflexa* and pathogenic *Leptospira* species have been achieved using the dCas9-based CRISPRi approach (Fernandes et al., 2021). Nonetheless, the CRISPRi strategy is limited to being used for only gene silencing via transcriptional blockage (Xu and Qi, 2019).

Therefore, the predominant CRISPR-Cas type I-B system, marked in many pathogenic *Leptospira* genomes (Dixit et al., 2016; Xiao et al., 2019a), presents a promising alternative to be exploited for endogenous genome editing. The repurposing of endogenous CRISPR-Cas type I system for genome editing has recently been demonstrated in some bacteria, including *Sulfolobus islandicus* (type I-A) (Li et al., 2016), *Clostridium pasteurianum* (type I-B) (Pyne et al., 2016), *Haloarcula hispanica* (type I-B) (Cheng et al., 2017), *Lactobacillus crispatus* (type I-E) (Hidalgo-Cantabrana et al., 2019), and *Zymomonas mobilis* (type I-F) (Zheng et al., 2019). CRISPR-Cas systems exhibit high diversity in composition and operational mechanism across different bacterial genera and species (Gleditzsch et al., 2019; Liu and Doudna, 2020; Makarova et al., 2020a), necessitating their experimental validation in each bacterium instead of directly reckoning from the studies reported in related species. Therefore, to harness *Leptospira's* CRISPR-Cas type I-B (Lin_I-B) system for genetic manipulation, the characterization of its interference machinery is a prerequisite. To this end, the present study aims at the functional characterization of interference machinery of a transcriptionally active CRISPR-Cas type I-B system (Lin_I-B), recently identified in the *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 (Dixit et al., 2016). We investigated the molecular and/or biochemical characteristics of interference-associated Cas7 and Cas8b proteins to interpret their mechanistic role in the defense response of the Lin_I-B system. Underlining the absence of a distinct gene coding for the small subunit (Cas11) in the Lin_I-B locus (Dixit et al., 2016), the molecular composition of Lin_I-B's Cascade complex and its functionality has been assessed.

*****

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

Chapter 2

Assembly of Cas7 subunits of *Leptospira* on the mature crRNA of CRISPR-Cas I-B is modulated by divalent ions [1]

2.1. Summary

The spirochete *Leptospira interrogans* serovar Copenhageni harbors the genetic elements of the CRISPR-Cas type I-B system in its genome. CRISPR-Cas is a crRNA-mediated adaptive immune system in most prokaryotes against mobile genetic elements (MGEs). To eliminate the intruding MGEs, CRISPR-Cas type I systems utilize a Cascade (CRISPR-associated complex for antiviral defense) complex composed of Cas5, Cas6, Cas7, Cas8, and Cas11 bound with a crRNA. The Cas7 is essentially known to constitute the major component of the Cascade complex. The present study reports the biochemical characterization of the Cas7 (LinCas7) from the CRISPR-Cas type I-B system of *L. interrogans* serovar Copenhageni. The pure recombinant LinCas7 (rLinCas7) exists as a monomer in the solution revealed by size exclusion chromatography. The rLinCas7 demonstrates an endoDNase activity dependent upon divalent Mg^{2+} ions, monovalent ions, pH, temperature, and substrate size. Analysis of ribonucleoprotein composite (rLinCas7-crRNA) by electron microscopy and native-PAGE demonstrated that rLinCas7 could oligomerize on the mature crRNA framework in the presence of Mg^{2+} ions. The ribonucleoprotein composite attains a helical shape similar to the backbone of the Cascade complex. However, in the absence of Mg^{2+} ions, rLinCas7 acts as an RNase. The fluorescence spectroscopy disclosed a weak interaction (K_d = 26.81 mM) between rLinCas7 and Mg^{2+} ions, leading to an overall conformational change in rLinCas7 that modulates the rLinCas7's activity on DNA and RNA substrates. The nuclease activity of LinCas7 characterized in this study aids the functional divergences among proteins of the Cas7 family from different CRISPR-Cas systems in various organisms.

[1] This chapter is partly adapted from the published article and reprinted with the authors' permission. **Hussain, M.S.**, Kumar, M., 2022. Assembly of Cas7 subunits of *Leptospira* on the mature crRNA of CRISPR-Cas I-B is modulated by divalent ions. Gene 818, 146244.

2.2. Results

2.2.1. LinCas7 copurifies with host DNA

In the NCBI (National Center for Biotechnology Information) database, the gene *LIC10936* (840 bp) has been annotated to encode for Cas7 protein (LinCas7; 279 residues; 31 kDa) in the CRISPR-Cas type I-B system of *Leptospira*. Therefore, the gene *LIC10936* (*lincas7*) was cloned in the pET-28a vector **(Appendix Fig. A1a-b)** to overexpress with an N-terminal 6×His tag. The IPTG (Isopropyl-β-D-thiogalactopyranoside)-induced overexpression of the *LIC10936* (*lincas7*) in *E. coli* BL21 (DE3) cells resulted in the production of recombinant LinCas7 (rLinCas7; 33.6 kDa) that was purified by Ni-NTA (Nickel-nitrilotriacetic acid) affinity chromatography **(Fig. 2.1A)**. An attempt to remove the purification tag (N-terminal 6×His) from rLinCas7 using thrombin was conducted to avoid any nuclease artifact with the fused 6×His tag (Schoonen et al., 2017). However, efficient cleavage with thrombin could not be accomplished **(Appendix Fig. A2a)**. It is possible that the thrombin cleavage site might not be accessible within rLinCas7. Alternatively, *lincas7* (*LIC10936*) was cloned in a pGST-TEV expression vector **(Appendix Fig. A1a-c)** and overexpressed with an N-terminal GST (Glutathione S-transferase)-tag (57 kDa). The GST-tag was cleaved off efficiently with the TEV (Tobacco etch virus) protease, producing LinCas7 of the native size (31 kDa) **(Fig. 2.1B & Appendix Fig. A2b)**.

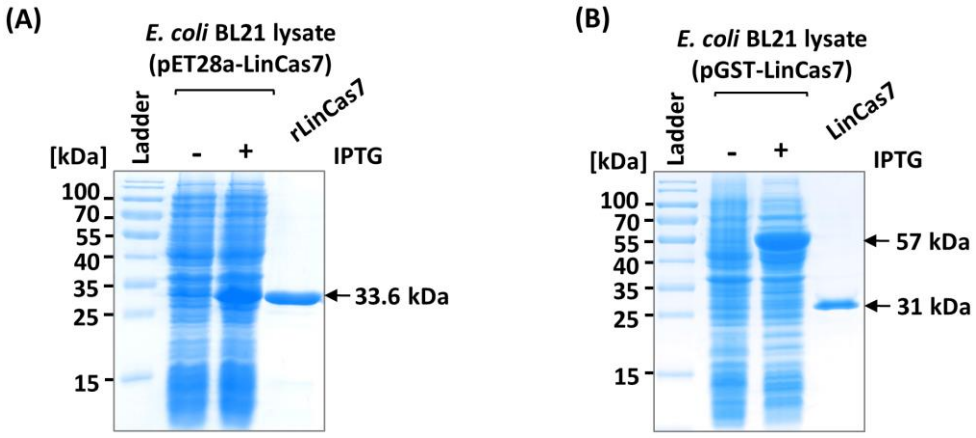

Fig. 2.1. Purification of recombinant Cas7 protein of *Leptospira*. (A) Overexpression and purification of recombinant Cas7 protein of *Leptospira* with 6×His tag (rLinCas7, 33.6 kDa). The SDS-PAGE (12 %; Coomassie-stained) analysis of whole-cell lysates of *E. coli* BL21 cells (pET28a-LinCas7) induced with (+) and without (-) IPTG (1 mM) and purified pure rLinCas7. **(B)** Overexpression and purification of recombinant LinCas7 (untagged LinCas7) of the native size (31 kDa). The SDS-PAGE (12 %; Coomassie-stained) analysis of whole-cell lysates of *E. coli* BL21 cells (pGST-LinCas7) overexpressing GST-LinCas7 (57 kDa) post-induction with (+) and without (-) IPTG (1 mM) and the purified untagged LinCas7 (31 kDa).

The TEV protease (28 kDa) having a fused 6×His tag was eliminated by passing the purified LinCas7 (untagged) through the Ni-NTA affinity column. Hereafter, in this study, we demarcate the two variants of recombinant LinCas7 with and without the tag as rLinCas7 and LinCas7, respectively. During the size exclusion chromatography (SEC), rLinCas7 eluted at a single peak (~36 kDa), closer to the theoretical monomeric unit (33.6 kDa) **(Fig. 2.2A)**. However, on SEC, the native LinCas7 eluted with two peaks: a minor peak (P-I) at the size of ~55 kDa and the other major peak (P-II) at ~36 kDa **(Fig. 2.2B)**.

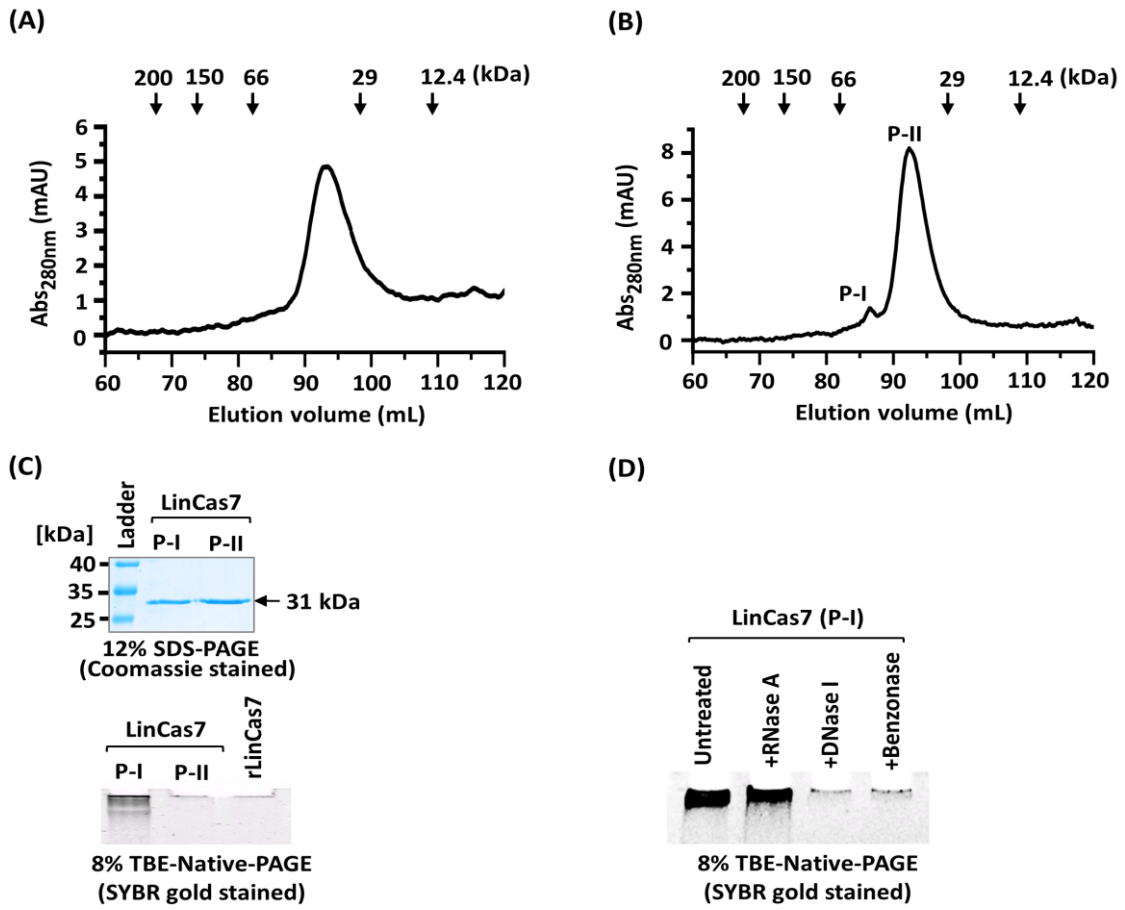

Fig. 2.2. LinCas7 copurifies with host DNA. The size exclusion chromatography (SEC) of affinity purified **(A)** rLinCas7 (6×His tagged, 33.6 kDa) and **(B)** LinCas7 (untagged, 31 kDa). The rLinCas7 eluted at the size of ~36 kDa, whereas LinCas7 eluted with two peaks (P-I of ~55 kDa size and P-II of ~36 kDa). The elution profiles of standard molecular weight protein markers are indicated with inverted arrows on top of the chromatogram (β-Amylase, 200 kDa; Alcohol dehydrogenase, 15 kDa; Albumin, 66 kDa; Carbonic anhydrase, 29 kDa; Cytochrome c, 12.4 kDa). **(C)** The SDS-PAGE analysis of P-I and P-II fractions of LinCas7 (upper panel). In the lower panel, fractions P-I and P-II of LinCas7′s along with rLinCas7 (1 µg of each) were resolved on TBE-Native-PAGE (8 %) and visualized with SYBR (Synergy brand) Gold stain. DNA was exclusively detected in the P-I fraction of LinCas7. **(D)** The P-I fraction of LinCas7 (2 µg) was treated with either RNase A, DNase I, or benzonase. The reaction product was resolved on TBE-Native-PAGE (8 %) and stained with SYBR gold.

The minor peak fraction (P-I) of LinCas7 was concentrated and resolved along with the P-II fraction on SDS-PAGE (Sodium dodecyl sulfate-polyacrylamide gel electrophoresis) gel to discern any presence of contaminant **(Fig. 2.2C, upper panel)**. The SDS-PAGE analysis of the P-I fraction showed the presence of LinCas7, similar to the P-II fraction. Interestingly, analysis of P-I fraction on TBE (Tris borate EDTA)-Native-PAGE-coupled with SYBR (Synergy brand) gold staining revealed the copurification of nucleic acid with LinCas7 **(Fig. 2.2C, lower panel)**.

The nucleic acid copurified with LinCas7 was cleaved by DNase I and benzonase but not with RNase A, suggesting it to be DNA **(Fig. 2.2D)**. Moreover, the copurification of DNA bound with LinCas7 turned out to be due to the use of low-salt buffers (100-150 mM NaCl) in comparison to the purification of pure rLinCas7 done at high-salt buffers (300 mM NaCl). When rLinCas7 was purified at low-salt buffers, DNA was bound with rLinCas7, similar to LinCas7 **(Appendix Fig. A2d)**. The high salt concentration (NaCl at ≥300 mM) is reported to destabilize the protein-nucleic acid interaction and dissociate the DNA or RNA from a nucleic acid-binding protein (Kalwani et al., 2020; Xiao et al., 2010). Moreover, the secondary structures of Cas7 protein variants (LinCas7 and rLinCas7) were also evaluated by Circular dichroism (CD) spectroscopy to understand the disparity for DNA copurification. CD profile of both the Cas7 variants (LinCas7 and rLinCas7) was nearly similar, with 35.6/34.4 % α-helix and 37.9/33.9 % β-sheets in LinCas7/rLinCas7 **(Appendix Fig. A3a-b)**, respectively and close to the theoretical values predicted by PSIPRED (Buchan and Jones, 2019). The proteins of the Cas7 family are conventionally comprehended to fasten at the variable region (spacer sequence) of crRNA and shape the backbone of the effector complex in CRISPR-Cas type I & III systems. Thus, the detection of LinCas7 bound with DNA during purification motivated us to examine its activity on various DNA substrates.

2.2.2. LinCas7 exhibits divalent metal ion-dependent endoDNase activity

In our preliminary nuclease assay, LinCas7 variants (LinCas7 and rLinCas7) exhibited Mg^{2+} dependent endodeoxyribonuclease (endoDNase) activity on ds-DNA substrate (circular double-stranded plasmid) at a related grade **(Appendix Fig. A4a)**. Consequently, the DNA-free rLinCas7 variant was selected for downstream biochemical characterization. The rLinCas7 activity on DNA substrates was explored under different reaction conditions. In a time-bound nuclease assay set for an hour, a gradual increase in the DNA degradation was evident with the

increasing concentrations (0.5-3 μM) of rLinCas7 **(Fig. 2.3A)**. The optimum degradation of DNA was attained at a concentration of 3 μM of rLinCas7. The rLinCas7 action on DNA substrate illustrated a defined preference for plasmid conformation. The different conformations of plasmid DNA noticed on agarose gel were supercoiled covalently closed circular (SC), nicked open circular (OC), and linear (L) DNA. The nuclease activity of rLinCas7 on plasmid occurred sequentially on SC, OC, and L forms before it gets entirely digested **(Fig. 2.3A and B)**. Such refined digestion of supercoiled plasmid is often perceived with restriction enzymes (Keatch, 2004). To rule out the DNase activity detected is not due to any other artifacts during protein (LinCas7 or rLinCas7) purification, nuclease activity was validated with another Cas7 ortholog from *D. vulgaris* (DvuCas7c) (Hochstrasser et al., 2016) similarly cloned and purified **(Appendix Fig. A1a-c & 2c)**. Nuclease assay was conducted utilizing the two Cas7 orthologs (rLinCas7 and DvuCas7c) independently on a circular ds-DNA plasmid **(Appendix Fig. A4b)**. Under similar reaction conditions, rLinCas7 could entirely digest the circular plasmid within an hour, but the same was not evident with the DvuCas7c. Nevertheless, DvuCas7c was competent in cleaving the ds-DNA upon prolonged incubation (24 h) of the nuclease reaction, inferring DvuCas7c has weaker DNase activity than rLinCas7 **(Appendix Fig. A4b)**.

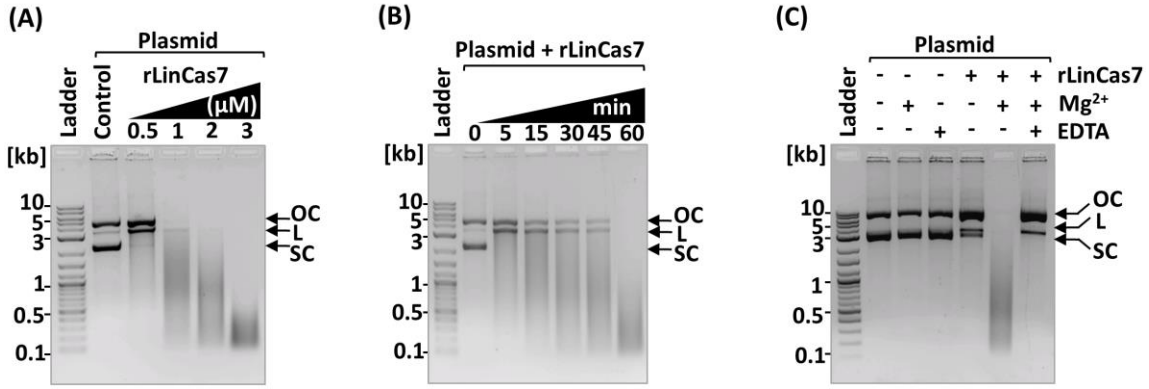

Fig. 2.3. Recombinant LinCas7 has divalent metal ion-dependent endoDNase activity. All the assays were performed at 37 °C for 60 min (if not specified) using rLinCas7 (3 μM or as indicated) and ~300 ng of plasmid DNA as a circular ds-DNA substrate (Plasmid) in 1× nuclease buffer with and without Mg^{2+} ions (10 mM). EDTA (10 mM) was used as a divalent metal ion chelator. The different conformations of plasmid DNA have been indicated as supercoiled covalently closed circular (SC), nicked open circular (OC), and linear (L) forms. Reactions were analyzed on ethidium bromide (EtBr) stained-agarose gel (1 %). The leftmost lane, in all the gels, contains 2 log DNA ladders. For clarity, the gel images have been presented after inversion. **(A)** The DNase activity assay of rLinCas7 against plasmid DNA (plasmid) at varying concentrations (0.5-3 μM). **(B)** Time chase nuclease assay of rLinCas7. Nuclease reactions were stopped by adding 1× DNA loading dye at the indicated time interval. **(C)** Divalent metal ion dependence assessment of rLinCas7 for DNase activity. At least two independent experiments confirmed the results represented.

Prompted by the pronounced DNase activity of rLinCas7, a set of nuclease assays was conducted using the circular ds-DNA plasmid as a substrate to determine the optimum mandated conditions (time, temperature, pH, co-factors, and monovalent ions). The rLinCas7 at 3 μM concentration exhibited complete degradation of ds-DNA (300 ng) in one hour **(Fig. 2.3B)**. The DNase activity of rLinCas7 was divalent metal ion-dependent as the absence of Mg^{2+} ions or chelation of metal with EDTA (ethylenediaminetetraacetic) developed no activity **(Fig. 2.3C)**. However, rLinCas7 can nick the supercoiled ds-DNA into an open circular conformation independent of divalent metal ions **(Fig. 2.3C)**.

After that, the nuclease assay was executed in the presence of different divalent cations to evaluate the rLinCas7′ s predisposition toward a particular type of divalent metal ions **(Fig. 2.4A)**. The divalent cations (Mg^{2+}, Mn^{2+}, Ni^{2+}, Ca^{2+}, Cu^{2+}, and Zn^{2+}) are typically associated with the activity of nucleic acid-binding proteins (Belkebir and Azeddoug, 2013; Miroshnikova et al., 2016). Distinguished nuclease activity was perceived in the presence of Mg^{2+} ions. On the other hand, in the presence of Mn^{2+} ions, reasonable degradation of DNA substrate was noticed. The Ni^{2+} and Ca^{2+} could not promote the rLinCas7 for DNase activity. No DNase activity was evident in the presence of two other metal ions (Cu^{2+} and Zn^{2+}) compared with the controls containing only DNA substrate in nuclease activity buffer with or without rLinCas7 **(Fig. 2.4A)**. Furthermore, the rLinCas7 nuclease assay demonstrated the Mg^{2+} ions concentration-dependent DNase activity **(Fig. 2.4B)**. The DNase activity augmented with the increasing concentrations of Mg^{2+} ions and attained saturation at 10-14 mM. Next, the influence of electrostatic interaction on the DNase activity of rLinCas7 was also examined by utilizing various monovalent salts (viz. NaCl, KCl, and NH_4Cl) at increasing concentrations (50, 100, and 150 mM) as illustrated in **Fig. 2.4C**. Among these three salts, the DNase activity of rLinCas7 was prominently witnessed in the presence of KCl than in NH_4Cl and NaCl. However, a low grade of DNase activity was also noticed with 50 mM NaCl which reduced about to the control grade at a higher NaCl concentration. The KCl at 100 mM concentration was estimated as the optimum salt concentration for rLinCas7′ s DNase activity. After that, the pH dependency of rLinCas7 for DNase activity was deduced by arranging the cleavage reactions at varying pH 3.0-11.0. The DNase activity of rLinCas7 was evident over a wide range of pH 6.0-10.0 **(Fig. 2.4D)**. The maximal DNase activity was witnessed between pH 7.0-9.0, whereas activity got drastically reduced at pH below 6.0 and above 9.0 **(Fig. 2.4D)**.

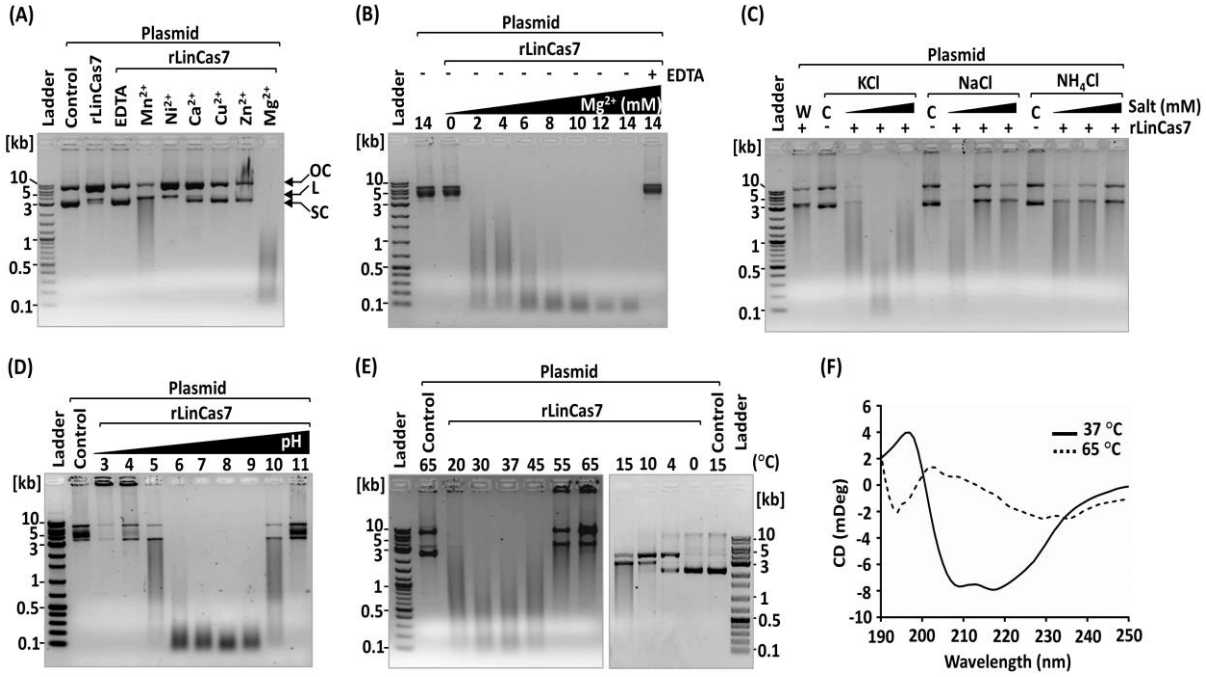

Fig. 2.4. Different parameters for DNase activity in rLinCas7. (A) DNase activity of rLinCas7 in the presence of different divalent metal ions. **(B)** DNase activity of rLinCas7 at varying concentrations (0-14 mM) of Mg^{2+}ions. **(C)** The influence of various monovalent salts (KCl, NaCl, and NH_4Cl) at varying concentrations (50, 150, and 200 mM) on the DNase activity of rLinCas7. "W" represents the control reaction without rLinCas7. "C" is the control reaction containing rLinCas7 without supplementation of indicated monovalent salts. **(D)** The DNase activity of rLinCas7 at varying pH. **(E)** The DNase activity of rLinCas7 at different temperatures. **(F)** Change in CD spectra of rLinCas7 due to the temperature. The CD spectra of the rLinCas7 were acquired after a 10 min incubation at 37 °C or 65 °C and compared. Nuclease reactions were analyzed on EtBr stained-agarose gel (1 %). For clarity, the gel images have been illustrated after inversion. In all the gels, either 2 log or 1 kb DNA ladders were loaded. At least two independent experiments confirmed the results represented.

The stacking of DNA in wells at pH 3.0-4.0 proposes the strong binding character of rLinCas7 to DNA at a highly acidic pH. The range of 30-37 °C was estimated to be the optimum temperature for DNase activity where complete DNA degradation was evident **(Fig. 2.4E)**. At temperature >45 °C, the protein was found to be aggregated, driving it to get affixed in the wells of electrophoretic agarose gel along with temporarily bound DNA molecules. The failure of nuclease activity because of aggregation was also ascertained by CD spectroscopy, where deformation of rLinCas7′ s secondary configurations was recorded when incubated at 65 °C for 10 min compared to the protein pre-incubated at 37 °C **(Fig. 2.4F)**.

2.2.3. LinCas7 is a non-specific endoDNase

A set of nuclease assays were accomplished using DNA substrates of varying nucleotide sequences, sizes, and conformations to determine the explicitness of rLinCas7 for DNase activity towards the type of DNA. The nuclease assay was executed against linear ds-DNA substrates of varied sizes (840 and 105 bp) and sequences **(Fig. 2.5A)**. Incision of both the DNA substrates was pronounced conditionally if the reaction contained Mg^{2+} ions, signifying the sequence-independent rLinCas7′ s endoDNase activity against ds-DNA molecules. However, against duplex DNA oligo substrate (40 bp), rLinCas7 was inert for nuclease activity but exhibited binding property **(Fig. 2.5B)**.

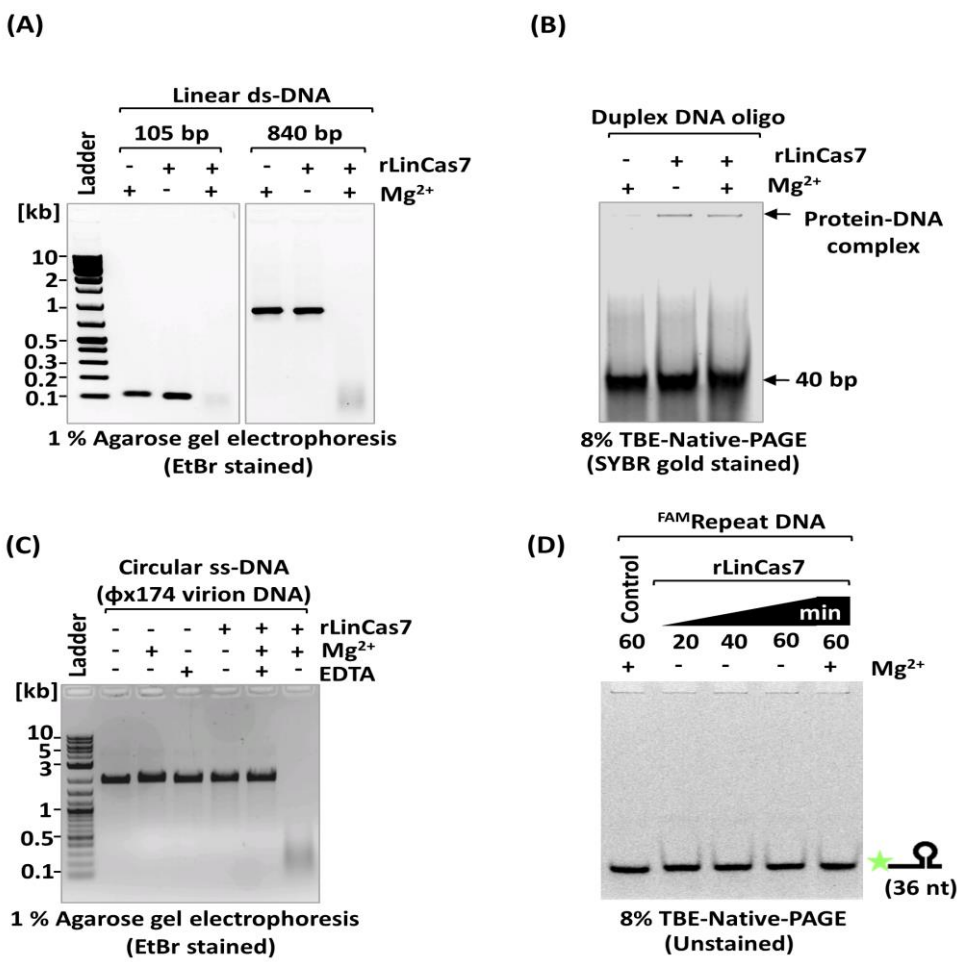

Fig. 2.5. Specificity of rLinCas7 endoDNase activity. The specificity of rLinCas7′ s DNase activity towards varied DNA sequences, sizes, and conformations was assayed (vide methods). **(A)** The rLinCas7′ s DNase activity on linear ds-DNA substrates of varied sizes (105 bp, 840 bp) and nucleotide sequences. **(B)** Nuclease assay of rLinCas7 against 40 bp duplex DNA oligos (1 μM). The reactants were resolved on 8 % TBE-Native- PAGE and visualized with SYBR gold staining. **(legend continued on next page)**

(C) Shows the rLinCas7′ s DNase activity on circular ss-DNA (ϕx174 virion DNA). Post-completion, nuclease reactions in A & C were analyzed on ethidium bromide (EtBr) stained-agarose gel (1 %). Gel images are presented after inversion. The leftmost lane in all the agarose gels contains 2 log DNA ladders. **(D)** Electrophoretic mobility shift assay (EMSA) of repeat DNA in the presence of rLinCas7 at variable time intervals. For EMSA, reactions of rLinCas7 (3 μM) on short ss-DNA oligos (0.5 μM) were set in 1× binding buffer with and without Mg^{2+} ions ($MgCl_2$; 10 mM) for the indicated time (20-60 min) at 37 °C and resolved on TBE-Native-PAGE (8 %) at 4 °C. The short ss-DNA oligos used is a 36 nucleotide 5′ fluorescent-labeled (marked with a green star) repeat DNA (^{FAM}Repeat DNA). At least two independent experiments confirmed the results presented.

The binding of rLinCas7 to the duplex DNA oligo (40 bp) without cleavage activity is in consensus with the observation that LinCas7 copurifies with DNA of size apparently ~30 bases long. The DNase assay indicates that the sensitivity of rLinCas7's activity relies on the size of the DNA substrates. The nuclease activity assay set on circular single-stranded DNA (ss-DNA) substrate (ϕx174 virion DNA) **(Fig. 2.5C)** also confirmed the non-specific character of LinCas7′ s endoDNase activity. However, rLinCas7 demonstrated neither cleavage nor stable binding with single-stranded short DNA oligos [synthetic Fluorescein amidite-labeled repeat DNA (^{FAM}Repeat DNA) of CRISPR2 in type I-B system of *L. interrogans* Copenhageni] under the specified experiment conditions, i.e., with/without Mg^{2+} ions at 37 °C **(Fig. 2.5D)**. The inert nature of rLinCas7 on short ss-DNA and duplex DNA oligo substrates demonstrated that the discerned endoDNase activity is specific to rLinCas7 and not because of any protein purification artifact.

2.2.4. LinCas7 displays divalent metal ion tunable RNA binding as well as cleavage

The various Cas7 orthologs characterized so far from different CRISPR-Cas systems of bacteria have been reported as RNA binding proteins that fasten at the variable region of mature crRNA and play a role in sculpting the helical backbone of the Cascade complex (R. N. Jackson et al., 2014; Lintner et al., 2011; Plagens et al., 2014). Cas7 family proteins can attach unstructured RNAs in a non-specific sequence mode. Nevertheless, there is stability in the binding of Cas7 to ss-RNA with an increase in ss-RNA length as reported for the Cas7 of *Methanopyrus kandleri* (MkCsm3) from the CRISPR-Cas type III-A system (Hrle et al., 2013). Furthermore, Cas7-like proteins that constitute the effector complex (termed Cmr or Csm complex) in several CRISPR-Cas type III systems are recognized to cleave the specific RNA target complementary to the crRNA of the effector complex (Liu and Doudna, 2020). In contrast, the BhaCas7 (also known as Csd2) encoded in the CRISPR-Cas type I-C system of

Bacillus halodurans has been reported as inert against DNA and RNA substrates (Punetha et al., 2014).

Here, the unusual DNase activity of LinCas7 motivated us to investigate its activity on RNA substrates. A time-bound activity assay was set up using an *in vitro* synthesized non-specific RNA substrate with varying concentrations of Mg^{2+} ions **(Fig. 2.6A)**. To our astonishment, instead of conventional RNA binding activity, rLinCas7, in the absence of Mg^{2+} ions, exhibited RNase activity. Notably, the RNase activity of rLinCas7 was inhibited in the presence of increasing concentrations of Mg^{2+} ions. The Mg^{2+} ions at an apparent molar ratio of 1:10000 (rLinCas7: Mg^{2+}) impeded the RNase activity of rLinCas7. The possibility of any RNase contamination in the protein preparation was ruled out by scrutinizing the rLinCas7 heat sensitivity towards RNase activity. The omnipresent RNases are highly stable to heat and can stay active even upon heating at 120 °C (Green and Sambrook, 2019; Miyamoto et al., 2009). Therefore, the pure rLinCas7 was heat-denatured at 65 °C for 10 min, and then the RNase activity was analyzed **(Fig. 2.6B)**. Heat denaturation of rLinCas7 compromised its RNase activity, ensuring that the observed RNase activity was specific.

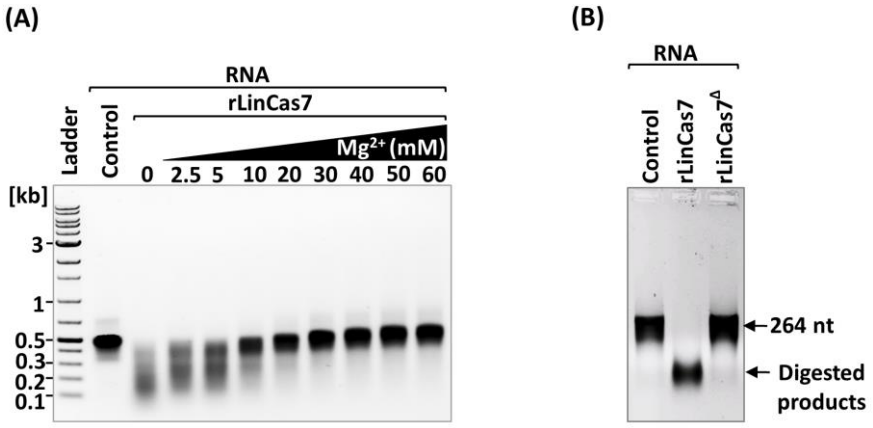

Fig. 2.6. rLinCas7 demonstrates metal ions tuneable RNase activity. (A) The time-bound nuclease activity of rLinCas7 (3 µM) against non-specific RNA substrate (264 nt). Nuclease reaction was performed in 1× nuclease activity buffer containing increasing concentrations (0-60 mM) of ions for 1 h at 37 °C and resolved on EtBr stained-agarose gel (1.5 %). The leftmost lane contains DNA ladders. **(B)** The effect of heat on pure rLinCas7′s RNase activity. The rLinCas7 was heated for 10 min at 65 °C (rLinCas7$^{\Delta}$) before setting the nuclease reaction against RNA substrate for 1 h at 37 °C. Post-completion, reactions were resolved on EtBr stained-agarose gel (1.5 %). The results presented were confirmed by at least two independent experiments.

Next, the rLinCas7 activity on its cognate mature crRNA (^{FAM}crRNA, 70 nt) and repeat RNA (^{FAM}Repeat RNA, 36 nt) was assessed **(Fig. 2.7A)**. The rLinCas7 displayed distinct binding to crRNA only in the presence of Mg^{2+} ions, and a shift in RNA migration was evident on Native-PAGE. Interestingly, in the absence of Mg^{2+} ions, rLinCas7 showed cleavage activity against both crRNA and repeat RNA substrates **(Fig. 2.7B),** consistent with its perceived action on a non-specific RNA **(Fig. 2.6A)**. The binding of rLinCas7 to crRNA was also in agreement through field emission transmission electron microscopy (FETEM) **(Fig. 2.7B and C)**, where the oligomerization of rLinCas7 was evident when incubated with crRNA. A ribonucleoprotein composite structure was seen **(Fig. 2.7C)**, aligning similar in shape to the reported helical backbone of the CRISPR-Cas type I Cascade complex (Hochstrasser et al., 2016; Liu and Doudna, 2020).

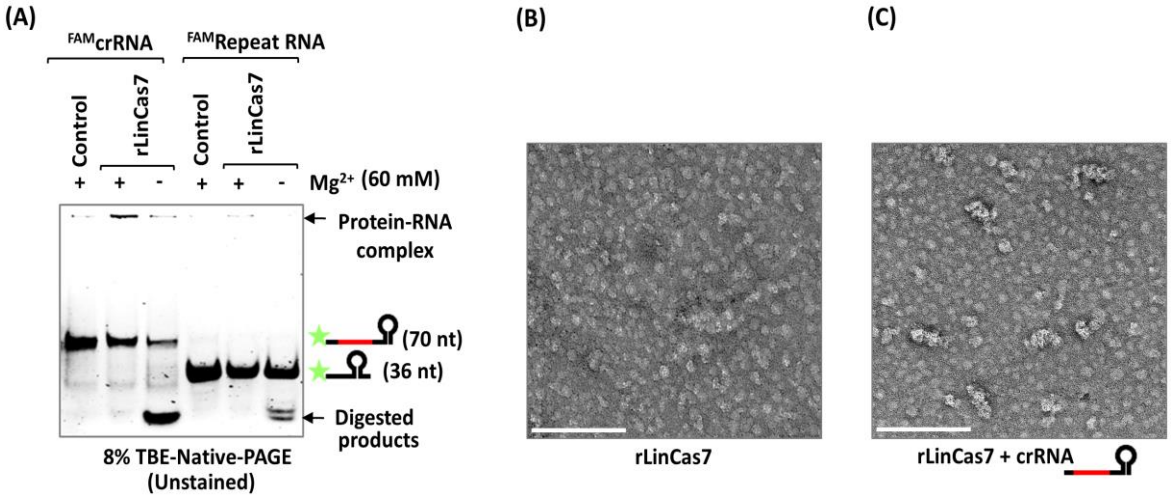

Fig. 2.7. Pure rLinCas7 binds its cognate crRNA in presence of Mg^{2+} ions. (A) The rLinCas7 demonstrates binding to mature crRNA in an electrophoretic mobility shift assay (EMSA). The EMSA assay was performed by incubating 5′ FAM-labeled (marked with a green star) mature crRNA (^{FAM}crRNA, 70 nt; 0.5 µM) and repeat RNA (^{FAM}Repeat RNA, 36 nt; 0.5 µM) substrates independently with rLinCas7 (6 µM) in 1× binding buffer in the presence and absence of Mg^{2+} (60 mM) for 1 h at 37 °C. Post-completion, reactions were resolved on TBE-Native-PAGE (8 %) at 4 °C. **(B)** Field emission transmission electron microscopy (FETEM) micrograph of pure rLinCas7. **(C)** FETEM micrograph of rLinCas7 after incubation with crRNA. The micrograph scale bars are 200 nm. The results presented were confirmed by at least two independent experiments.

Recently, it has been reported that the 6×His tag of the recombinant protein may contribute to RNase activity as an artifact (Schoonen et al., 2017). In the RNA-binding assay, each Cas7 variant (LinCas7 and rLinCas7) of *Leptospira* exhibited RNase activity without Mg^{2+} ions **(Appendix Fig. A5a)**. Thus, an artifact RNase activity within Cas7 variants of *Leptospira* was

overruled. To further substantiate any artifact's role in RNase activity within the LinCas7 variant, the RNA binding assay was executed with DvuCas7c (untagged) under similar experimental conditions. The DvuCas7c has been demonstrated to have an affinity for the crRNA without any RNA cleavage elsewhere (Hochstrasser et al., 2016). In agreement with the reported finding, the DvuCas7c in this study ascertained the RNA substrate binding with no cleavage activity **(Appendix Fig. A5b)**. In the DvuCas7c RNA binding assay, a shift in RNA substrate migration on Native-PAGE was detected **(Appendix Fig. A5b)**. Interestingly, the binding efficiency of DvuCas7c is reasonably compromised in the presence of Mg^{2+} ions. The detectable RNA binding affinity without the cleavage activity of DvuCas7c ruled out the possible RNase artifact and advocated the LinCas7 variants to possess an intrinsic RNase characteristic.

2.2.5. Mg^{2+} ions modulate the overall structural conformation of LinCas7

The structure prediction for LinCas7 by the I-TASSER program (Yang and Zhang, 2015) projected a 3D model with a confidence score of -0.28 that showed the closest architectural analogy with SsoCas7 (PDB: 3ps0A) of CRISPR-Cas type I-A system in *Sulfolobus solfataricus* (Lintner et al., 2011), scoring a root mean square deviation (RMSD) of 1.86 Å for all C_α atoms **(Appendix Fig. A6)**. Likewise, the LinCas7 shares the best sequence match with SsoCas7 among all functionally and/or structurally characterized Cas7 orthologs as revealed by BlastP alignment (24 % sequence identity and 61 % query coverages). The computational analysis of modeled LinCas7 revealed it to contain five tryptophan (W) and seven tyrosine (Y) residues, which are largely buried. The relative surface accessibility (RSA) of tryptophan residues in LinCas7 ranges from 0 to 33.7 % [W35 (33.7 %), W56 (0 %), W75 (13 %), W119 (15.1 %), and W166 (6.32 %)] while the RSA of tyrosines ranges between 0-65.8 % [Y50 (24.7 %), Y86 (0 %), Y142 (65.8 %), Y151 (0 %), Y153 (0.38 %), Y189 (21 %), and Y210 (9.9 %)]. The largely buried positioning of tryptophan and tyrosine residues allowed the probing the nature of Mg^{2+} ions interaction with LinCas7 by measuring the LinCas7′ s intrinsic fluorescence in the presence and absence of Mg^{2+} ions. The intrinsic fluorescence measurement of a protein provides information on the nature of ligand binding and conformational changes in protein (Bougie et al., 2003; Punetha et al., 2014; Suryawanshi et al., 2016).

The intrinsic fluorescence measurement of pure rLinCas7 in the absence of Mg^{2+} ions displayed the maximum emission at 342 nm (λ_{max}) upon excitation at 290 nm **(Fig. 2.8A)**. A steady decrease in pure rLinCas7′ s intrinsic fluorescence intensity (i.e., λ_{max}) was recorded with the increase in Mg^{2+} ions concentrations (0-60 mM) and reached saturation at 100-300 mM of Mg^{2+} ions **(Fig. 2.8A-C)**. The reduction in intrinsic fluorescence suggests the quenching of tryptophan and tyrosine residues, which can be possible if they are getting exposed to solvent or become closer to each other due to conformational changes in rLinCas7 by metal ion binding. The binding of the metal ions at the site(s) closer to the emission traits of tryptophan and tyrosine residues in rLinCas7 may also result in the quenching of rLinCas7's intrinsic fluorescence.

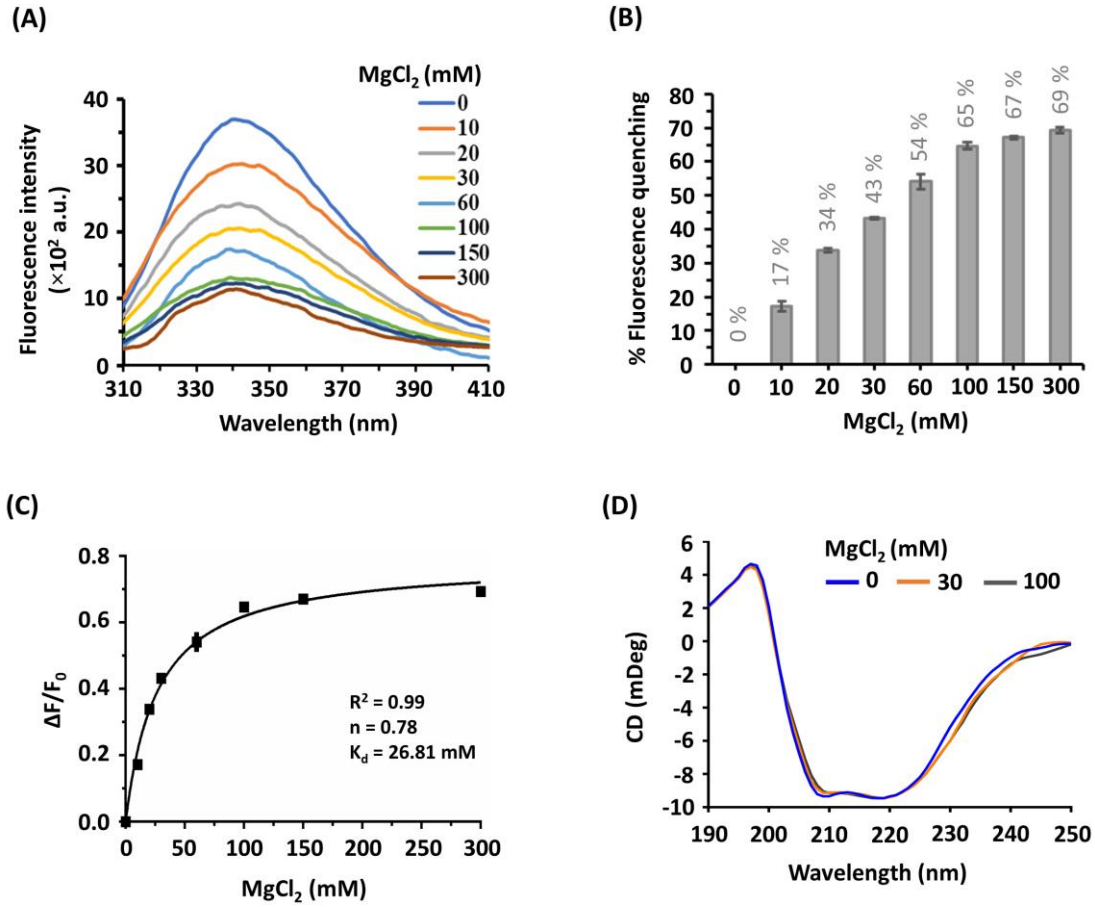

Fig. 2.8. Interaction between Mg^{2+} ions and rLinCas7. (A) The representative intrinsic fluorescence spectra of rLinCas7 (1 µM) in the presence of Mg^{2+} ions of variable concentrations (0-300 mM). The fluorescence emission spectra of rLinCas7 ranging from 310 to 410 nm at an excitation wavelength of 290 nm have been shown. **(B)** Percentage quenching of intrinsic fluorescence of rLinCas7 (1 µM) at maximum emission wavelength (342 nm, λ_{max}) as a function of added Mg^{2+} ions (0-300 mM). The percent quenching of fluorescence was calculated considering rLinCas7's intrinsic fluorescence in the absence of Mg^{2+} (0 mM) as 100 %. **(legend continued on next page)**

(C) The saturation binding isotherm generated by plotting the change in rLinCas7′s intrinsic fluorescence intensity ($\Delta F/F_0$) at 342 nm (λ_{max}) as a function of variable $MgCl_2$ concentrations (0-300 mM). Here ΔF is the magnitude of difference between the fluorescence intensity of rLinCas7 with (F) and without (F_0) the Mg^{2+} ions. R^2 indicates the square of the goodness of fit, n represents the number of binding sites, and K_d denotes the apparent dissociation constant, respectively. **(D)** The CD spectra of rLinCas7 (10 μM) in the presence of varying concentrations (0, 30 & 100 mM) of Mg^{2+} ions. The presented CD spectra are the average of 5 scans after baseline correction. The error bars in (B) & (C) show the standard deviations (SDs) from the two independent experiments.

A saturation binding isotherm was plotted using the fluorescence data to get insight into the nature of the interaction of Mg^{2+} ions with rLinCas7 **(Fig. 2.8C)**. The non-linear curve of the saturation binding isotherm of rLinCas7 projected the number of binding sites for Mg^{2+} ions to be close to one (n = 0.78) with a dissociation constant (K_d) of 26.81 mM. Furthermore, the CD profile of rLinCas7 was also recorded to reveal any alteration in secondary structures due to the presence of Mg^{2+} ions **(Fig. 2.8D)**. The CD analysis of rLinCas7 in the absence and presence of Mg^{2+} ions showed similar secondary structure content (36.0/34.8/33.5 % α-helix and 37.6/37.2/ 38.6 % β-sheets in rLinCas7 at 0/30/100 mM Mg^{2+} ions, respectively).

2.3. Discussion

CRISPR-Cas adaptive immunity in prokaryotes utilizes the effector complex to fend off the intruding genetic elements based on memory acquired during past infections with similar genetic elements (Barrangou and Horvath, 2017). In CRISPR-Cas type I and type III systems, the effector complex is a large ribonucleoprotein composite comprised of a set of Cas proteins assembled on crRNA in a specific structural order (Liu and Doudna, 2020). Cas7 is the major subunit protein of the effector complex. Multiple units of Cas7 polymerize helically on the unstructured variable spacer region of crRNA, creating the central backbone of the effector complex (Hochstrasser et al., 2016; Liu and Doudna, 2020). In agreement with its high affinity for hypervariable spacer region of crRNA in effector complex, the Cas7 orthologs from various CRISPR-Cas systems have been characterized as non-specific RNA binding proteins (Hrle et al., 2013; Kalwani et al., 2020; Plagens et al., 2014). However, the RNA binding strength of Cas7 proteins is weakened by the smaller size and structured conformations of RNA, as demonstrated for the Cas7 ortholog- MkCsm3 (Hrle et al., 2013). The MkCsm3 showed a strong binding affinity towards the unstructured ss-RNAs of ≥15 nt size independent of sequence. While, for the structured RNAs (CRISPR repeat RNAs) and ss-RNA of smaller size (8 nt), MkCsm3 did not exhibit significant binding activity. On the same line, the Cas7

orthologs from *Thermoproteus tenax* (TtxCas7) and cyanobacterium *Anabaena* sp. PCC 7120 (Alr1562) has been documented to copurify with non-specific RNA (Kalwani et al., 2020; Plagens et al., 2014). In addition, the copurification of DNA as an impurity with Cas7-bound RNA has also been reported in the *S. solfataricus* CRISPR-Cas type I-A system (Lintner et al., 2011). On the other hand, in this study, LinCas7 copurified exclusively with DNA. The apparent size of DNA copurified with LinCas7 was calculated to be around 30 bases long based on the difference in the molecular size of DNA-bound LinCas7 (~55 kDa) and the pure LinCas7 (~36 kDa) on SEC. The purification of recombinant LinCas7 bound with DNA intimated its affinity for DNA and drove us to scrutinize its activity on DNA substrates. The DNase activity of Cas7 orthologs has been underreported to date, possibly due to milder nuclease action on DNA as noticed for DvuCas7c in this study **(Appendix Fig. A4b)**. Similarly, the BhaCas7 from *B. halodurans* has been demonstrated to be inert on DNA and RNA substrate (Punetha et al., 2014).

Unlike BhaCas7 and DvuCas7c, the rLinCas7 displayed a pronounced non-specific Mg^{2+} ion-dependent endoDNase activity. The endoDNase activity in the rLinCas7 is non-canonical for its proposed roles in the CRISPR-Cas defense system. Nevertheless, such unconventional DNase/ RNase activity has also been recorded for other Cas proteins, including BhaCas5c and BhaCas8c of *B. halodurans* (Punetha et al., 2014). Similarly, multiple Cas proteins of *L. interrogans*, including LinCas1 (Dixit et al., 2021b), LinCas2 (Dixit et al., 2016; Xiao et al., 2019b), and LinCas4 (Dixit et al., 2021a), demonstrated the non-specific DNase activity. Similar to the LinCas2 and LinCas4, the rLinCas7 showed prominent endoDNase activity in the presence of divalent Mg^{2+} ions and, to a minor extent, with Mn^{2+} ions. The other physiological conditions mandated for rLinCas7 DNase activity viz. pH (7.0-9.0), temperature (30-37 °C), and preference for specific monovalent ions (KCl) were also similar to the ones reported for LinCas2 and LinCas4 (Dixit et al., 2021a, 2016; Xiao et al., 2019b). The different nuclease assays together suggest that *Leptospira* may have a similar mechanism to regulate the nuclease activity of Cas proteins. Further, the structure disorderliness of rLinCas7 at 65 °C was evident by CD spectroscopy data. CD data of rLinCas7 suggests its thermal instability beyond 45 °C results in a gradual decrease in DNase activity.

At this juncture, the functional applicability of LinCas7 DNase activity towards CRISPR-Cas immunity in the *Leptospira* genome remains unclear. However, the promiscuous DNase

activity in LinCas7 can be considered an evolutionary adaptation that may potentially benefit the bacteria to confront the intruding phage genome with fewer restriction sites or CRISPR escape mutations (Vasu et al., 2012). The additional DNase activity in a Cascade complex furnished by Cas7, along with Cas3 helicase/nuclease, may accelerate the alien DNA destruction by CRISPR interference. Furthermore, the non-specific DNase activity in the subunit proteins of a Cascade can also be advantageous for CRISPR adaptation. Although the mechanistic details are still elusive, there are accumulating pieces of evidence suggesting the involvement of Cascade complex and Cas3 along with Cas1/Cas2 in the adaptation phage of CRISPR defense (Datsenko et al., 2012; Garrett et al., 2020; Musharova et al., 2021). The primed adaptation essentially involves the CRISPR interference complex wherein the Cascade recognizes the invasive DNA with perfectly matching protospacer and/or altered PAM/protospacer sequence (escape mutations) and recruits the Cas3 helicase/nuclease for subsequent degradation of target DNA (Musharova et al., 2021). The target DNA degradation by Cas3 generates short ds-DNA fragments (prespacers) that are processed and integrated into the CRISPR array as new spacers by Cas1-Cas2 heterocomplex (Musharova et al., 2021). Recently, the Cascade-, Cas3- and Cas1/Cas2-dependent primed as well as naïve adaptation has been demonstrated in the CRISPR-Cas type I-B system of *Pyrococcus furiosus* (Garrett et al., 2020). The deletion of Cas3 or the type I-B Cascade complex individually led to a drastic reduction in new spacer acquisition compared to wild type *P. furiosus* strain (Garrett et al., 2020). Therefore, a Cascade complex with promiscuous DNase activity may promote the efficient acquisition of multiple new spacers from a phage genome which hoax its destruction by the Cas3 nuclease due to escape mutations in the PAM/protospacer region.

The rLinCas7 showed size sensitivity for DNase activity as it could not degrade the short DNA duplex (40 bp), and in DNase assay, accumulation of degraded products of <100 bp was always evident. The observed sensitivity of rLinCas7 DNase activity towards the size of the ds-DNA substrates is in line with our speculation that the non-specific DNase activity of LinCas7 may foster the generation of short DNA fragments for incorporation into the CRISPR array, hence may endorse the adaptation process. The divalent metal ion-independent DNA nicking activity of LinCas7 is also in conformity with the requirement of the CRISPR defense as it may enable the Cascade complex to settle on the target DNA by relaxing the supercoiled DNA and facilitating the construction of an open platform for the binding of Cas3 nuclease. Also, the non-specific nucleases are implicated in a comprehensive range of cellular tasks, including

regulating endogenous gene expression, replication, recombination, and repair (Kang et al., 2010; Yang, 2011). Hence, it is conceivable that besides the canonical functions in CRISPR-Cas immunity, the non-specific nuclease activity of Cas proteins may be utilized by the host bacterium for other physiological tasks. For instance, the nuclease activity of Cas2 in *Legionella pneumophila* promotes the intracellular infection of amoebal host cells (Gunderson et al., 2015). Also, the Cas9 of *Francisella novicida* has been reported to foster bacterial virulence by degrading the endogenous transcript encoding a bacterial lipoprotein that otherwise triggers the host's innate immune response (Sampson et al., 2013). Moreover, the low affinity of LinCas7 for Mg^{2+} ions (K_d = 26.81 mM) can be one of the regulatory features that suppress the non-specific DNase activity of LinCas7 inside the bacterial cell where the total Mg^{2+} ions concentration ranges up to 100 mM (Maguire and Cowan, 2002; Moncany and Kellenberger, 1981).

Similarly, rLinCas7 also exhibits divalent metal ion tuneable RNA binding and cleavage activity. In consensus with its canonical role in CRISPR-Cas immunity, i.e., to attach the crRNA and shape the helical backbone of the CRISPR effector complex, rLinCas7 showed distinct binding to its authentic mature crRNA in the presence of Mg^{2+} ions. However, the absence of Mg^{2+} ions rendered the rLinCas7 to cleave the RNA molecules non-specifically. Under similar conditions, rLinCas7 did not show detectable binding with the repeat RNA, unlike its affinity for mature crRNA. The shorter and mostly structured conformation of repeat RNA compared to crRNA could be the possible reason that obstructed the rLinCas7 binding as previously documented for its orthologs MkCsm3 (Hrle et al., 2013) and Alr1562 (Kalwani et al., 2020). MkCsm3 and Alr1562 have been demonstrated to have a higher affinity for RNA with a more extended linear unstructured segment (mature crRNA) in comparison to short-structured RNA (repeat RNA) (Hrle et al., 2013; Kalwani et al., 2020). Although the Cas7 family proteins share structural analogy, sequence-wise, they show extremely poor homology. The disparity in the amino acid sequence among Cas7 orthologs can lead to uneven oligomerization properties (Gu et al., 2019; Kalwani et al., 2020). For instance, Alr1562 (38 kDa), a Cas7 protein of Cyanobacterium Anabaena PCC 7120, exists as a dimer in solution (Kalwani et al., 2020). However, Alr1562, along with RNA, can form higher oligomeric complexes of >600 kDa. Similarly, pure ZmCsy3 (38 kDa) (Cas7 of *Zymomonas mobilis*) requires RNA for oligomerization in solution (Gu et al., 2019). Moreover, ZmCsy3 can self-

assemble into a distinct oligomeric structure (7-mer) when it is in the crystalline state and resembles the backbone of the Cascade complex (Gu et al., 2019).

The transmission electron microscopy revealed the ability of rLinCas7 to oligomerize onto the crRNA framework into distinct structures. The assembled structure roughly resembles the helical backbone of the Cascade complex. Likewise, the various Cas7 orthologs from different type I and type III CRISPR-Cas systems have been demonstrated to form helical arrangement on RNA, including the DvuCas7c (Hochstrasser et al., 2016), SsoCas7 (Lintner et al., 2011), and Cas7e of *E. coli* (Wright et al., 2016). The RNA-induced helical assembly of rLinCas7 agrees with its mandate function in the CRISPR-Cas system of *Leptospira*. Furthermore, the oligomerization of rLinCas7 is RNA dependent since the pure rLinCas7 alone, in SEC, showed an elution profile equivalent to the monomeric size and appeared as a homogenous single unit entity under the electron microscope. Similarly, the SsoCas7, ZmCsy3 and MkCsm3 were reported to exist as a monomer in solution and requires crRNA for helical assembly (Gu et al., 2019; Hrle et al., 2013; Lintner et al., 2011).

The Mg^{2+} ions at an apparent molar ratio of 1:10000 (rLinCas7: Mg^{2+}) deterred the RNase activity of rLinCas7. At this molar ratio, Mg^{2+} ions reduced the intrinsic fluorescence of rLinCas7 by approximately 17 % without altering its secondary structures. The free Mg^{2+} ions concentration in *Leptospira* has been estimated to be 0.4 mM (Szatmári et al., 2020). However, the total Mg^{2+} ions concentration, including free and bound forms in *Leptospira* cells, is unknown. In general, the total Mg^{2+} ions content in a bacterial cell (*E. coli*) can range up to 100 mM (Maguire and Cowan, 2002; Moncany and Kellenberger, 1981). The expression level of LinCas7 (17 copies/cell = ~80 nM) in *Leptospira* detected by mass spectrometry (Malmström et al., 2009) is very low and thus may allow Mg^{2+} ions to modulate the LinCas7 nuclease activity *in vivo*. Such divalent metal ion modulation can reasonably impede the LinCas7 RNase activity, permit LinCas7 to bind the crRNA, and facilitate the degradation of intruder DNA by the CRISPR-Cas interference complex. Furthermore, in a CRISPR-Cas system, a team of Cas proteins forms the functional effector complex characterized by collective interactions between different Cas proteins and with crRNA as exemplified in the CRISPR-Cas type I-C system of *D. vulgaris* (Hochstrasser et al., 2016). In *D. vulgaris* CRISPR-Cas I-C, multiple Cas proteins and the crRNA cooperatively shape a functional Cascade complex (Hochstrasser et al., 2016). Hence, it is plausible that other interacting Cas

proteins partners in the Cascade complex may regulate the LinCas7 non-canonical activities (DNase and RNase) and direct it to execute its mandate role in CRISPR immunity. A similar mechanism to seize the unexpected nuclease activities has been described for BhaCas8c (or Csd1) (Punetha et al., 2014). The RNase action of BhaCas8c was inhibited by pure BhaCas7 (or Csd2) when supplemented into the nuclease reaction (Punetha et al., 2014). In general, for Cas8 of CRISPR-Cas type I systems, the uncontested function is to bind and cap the 5′ end of crRNA with the aid of Cas7 (Liu and Doudna, 2020). Overall, the present study suggests that the LinCas7 is an evolved Cas7 variant with additional nuclease benefits that may enable the bacterium to cope-up more effectively against the impersonator phage genome. Moreover, the exact role of the promiscuous nuclease activity of LinCas7 in the CRISPR-Cas type I-B defense system of *Leptospira* needs to be further interpreted.

2.4. Conclusion

The pure rLinCas7 displays Mg^{2+} ion tunable activity against DNA and RNA substrates. In agreement with its canonical role in CRISPR-Cas immunity, rLinCas7 displays distinct binding to mature crRNA exclusively in the presence of Mg^{2+} ions. In the absence of divalent metal ions, rLinCas7 behaves as a non-specific RNase. LinCas7 seems unique among proteins classified under the Cas7 family as it demonstrates Mg^{2+} ion-dependent endoDNase activity. The LinCas7's non-canonical nuclease activity, concurrently with the previously reported dual nuclease nature of LinCas2 (Dixit et al., 2016) and the non-specific DNase action of LinCas1 (Dixit et al., 2021b) and LinCas4 (Dixit et al., 2021a) implies a distinctly evolved CRISPR-Cas I-B immune system in *Leptospira* compared to other prokaryotes.

2.5 Materials and methods

2.5.1. Bacterial strains, media, and growth conditions

L. interrogans serovar Copenhageni strain Fiocruz L1-130 was procured from the Indian Council of Medical Research (ICMR), Regional Medical Research Centre (RMRC), Port Blair, Andaman, and Nicobar Island, India. The spirochete culture was maintained in EMJH (Ellinghausen-McCullough-Johnson-Harris) medium (Difco, Becton Dickinson India Pvt. Ltd., New Delhi, India) at 29 °C by regular subculturing at an interval of 6–7 days. The *E. coli* strains DH5α (NEB #C2987H) and BL21 (DE3) (Novagen #69450-M) were used to propagate plasmid constructs and the over-expression of recombinant proteins, respectively. Both *E. coli*

strains were grown in Miller Luria Bertani (LB)-broth (Himedia #M1245) or Miller LB-agar (Himedia #M1151) medium supplemented with respective antibiotics at 37 °C.

2.5.2. Cloning, overexpression, and purification of Cas7 proteins

The complete *lincas7* ORF (Open reading frame) (*LIC10936*; 840 bp) was PCR (Polymerase chain reaction) amplified from the genomic DNA of *L. interrogans* Copenhageni strain Fiocruz L1-130 using gene-specific primer pair. The gene-specific primers and plasmids used in this study are listed in (Appendix Table A1). The amplified *lincas7* gene was cloned into a pET-28a expression vector between *Nhe*I and *Xho*I restriction sites to overexpress the LinCas7 with an N-terminal 6×His tag. For overexpression with an N-terminal GST-tag, the PCR amplified *lincas7* gene was cloned into a pGST-TEV expression vector between *Eco*RI and *Xho*I restriction sites. Similarly, for comparative analysis, the *cas7c* gene (*dvucas7c*, 873 bp) of *D. vulgaris* was PCR amplified from a commercially available plasmid construct Cas7/pHMGWA (Addgene plasmid # 81184) (Hochstrasser et al., 2016) and cloned into a pGST-TEV expression vector between *Eco*RI and *Xho*I restriction sites to overexpress the recombinant DvuCas7c protein with an N-terminal GST tag (GST-DvuCas7c; 58 kDa). The recombinant plasmid constructs (pET28a-LinCas7, pGST-LinCas7, and pGST-DvuCas7c) were confirmed by double restriction digestion and Sanger sequencing (Eurofins, India). The recombinant plasmid constructs were independently transformed into *E. coli* DH5α and BL21 (DE3) cells for *in-vivo* plasmid propagation and overexpression of Cas7 proteins, respectively.

The Cas7 proteins were overexpressed in *E. coli* BL21 (DE3) cells by inducing the cultures at OD600 = 0.5-0.6 with 1 mM IPTG for 4 h at 180 rpm and 37 °C. Induced cells were harvested and washed with 1× phosphate-buffered saline (1× PBS composition: 10 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl, and 2.7 mM KCl; pH 7.4) by centrifugation at 3000 × g for 10 min. The 6×His-LinCas7 (rLinCas7; 33.6 kDa) was purified by affinity chromatography using nickel-nitrilotriacetic acid (Ni-NTA) resins (GCC Biotech, India #GPC-01D) under native conditions as described before in our laboratory (Dhara et al., 2021). Briefly, the induced culture pellets were lysed in ice-cold lysis buffer-1 (50 mM Tris- HCl, 300 mM NaCl, 1 % Triton X-100, and 10 % glycerol; pH 8.0) for 30 min on ice and sonicated for 5 min at 25 % amplitude with 5 sec on/off cycles. The resulting homogenate was clarified by centrifugation at 13,000 × *g* for 20 min at 4 °C. The supernatant was allowed to bind the Ni-NTA resins in a 10 mL gravity-flow column. The column was washed with 6-7 column volumes of ice-cold

wash buffer-1 (50 mM Tris-HCl and 300 mM NaCl; pH 8.0) supplemented with 30 mM imidazole. Post-washing recombinant protein was eluted in elution buffer (50 mM Tris-HCl, 300 mM NaCl, 10 % Glycerol, and 250 mM imidazole; pH 8.0). The induced *E. coli* cultures overexpressing GST-LinCas7 (57 kDa) proteins were lysed in ice-cold lysis buffer-2 (50 mM Tris-HCl, 150 mM NaCl, 1 % Triton X-100, and 10 % glycerol; pH 8.0) to prepare the untagged LinCas7. After that, the clarified lysates were allowed to bind the glutathione agarose beads (BioBharati LifeScience, India #BB-GA004E) overnight at 4 °C. Post-binding, protein-bound beads were washed with a 10-column volume of ice-cold wash buffer-2 (50 mM Tris-HCl and 100 mM NaCl; pH 8.0). Afterward, the untagged LinCas7 (LinCas7; 31 kDa) was eluted by cleaving the GST-tag through in-bead TEV protease (6×His-TEV) (NEB #P8112S) cleavage in 1× TEV protease cleavage buffer at 4 °C for over-night. After elution, 6×His-TEV was removed by an additional Ni-NTA purification step.

The affinity-purified rLinCas7 and LinCas7 were individually subjected to the second purification step by size exclusion chromatography (SEC). The SEC fractions containing rLinCas7 or LinCas7 were concentrated using the 10 kDa centrifugal filter units (Amicon #UFC901024) and stored in a storage buffer (25 mM Tris-HCl, 100 mM NaCl, 5 % Glycerol; pH 8.0) at -20 °C until further experimental use. The untagged DvuCas7c was purified following a similar procedure as performed for LinCas7 (untagged) purification, except lysis buffer-1 and wash buffer-1 were used. The purified proteins were analyzed on 12 % sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) coupled with Coomassie brilliant blue R250 staining.

2.5.3. Size exclusion chromatography

The size exclusion chromatography (SEC) was carried out using HiLoad™ 16/600 Superdex™ 200 pg column (GE Healthcare #28-9893-35) connected to an AKTAprime plus chromatography system (GE Healthcare). The column was equilibrated with storage buffer-1 and calibrated with molecular weight markers (Sigma-Aldrich #MWGF200) of gel filtration at a 1 mL/min flow rate. After that, the SEC of affinity-purified rLinCas7, LinCas7, and DvuCas7c was executed under identical operating conditions.

2.5.4. Nuclease activity assay

The nuclease activity of LinCas7 was explored on various DNA and RNA substrates listed in Table 2.1. To generate a non-specific RNA substrate, the full-length *LIC10940* gene (213 bp) of *L. interrogans* was cloned into pTZ57R/T transcription vector directionally and *in vitro* transcribed using Hiscribe™ T7 high yield RNA synthesis kit (NEB #E2040S) following manufacturer's instructions. The synthesized RNA contains a vector-derived additional 51 nt sequence at its 5′ end. All the nuclease assays (DNase/RNase) were performed using 3 μM of pure Cas7 proteins (rLinCas7, LinCas7, or DvuCas7c) in a 10 μL reaction volume independently in 1× nuclease activity buffer (20 mM Tris-HCl and 100 mM KCl; pH 8.0) and 10 mM $MgCl_2$ at 37 °C for 1 h unless specified. The buffer composition was formulated as described elsewhere (Punetha et al., 2014). In a time-bound nuclease activity assay, samples were pulled out at the specified time intervals, and the reaction was terminated using either 50 mM ethylenediaminetetraacetic acid (EDTA; pH 8.0) or 1× DNA loading dye. Divalent metal-ion dependence for nuclease activity was explored by substituting the Mg^{2+} with numerous other divalent cations (Mn^{2+}, Ni^{2+}, Ca^{2+}, Cu^{2+}, and Zn^{2+}) in the reactions. EDTA (10 mM) was utilized to chelate the metals in the nuclease reactions.

The influence of electrostatic interaction on rLinCas7 nuclease activity was scrutinized utilizing various salts (NaCl, KCl, and NH_4Cl) at varying concentrations (50, 100, and 150 mM). The optimal temperature mandated for nuclease activity was determined by incubating the nuclease reactions at varying temperatures (0-65 °C) for 1 h. The optimal pH was discerned by establishing the reactions in a buffer containing 100 mM KCl, 10 mM $MgCl_2$ and 20 mM of either sodium citrate (pH 3.0-5.0), MES (pH 6.0), Tris-HCl (pH 7.0-8.0) or CAPS (pH 9.0-11.0). All the nuclease activity reactions, after completion, were electrophoresed either on agarose gel (1-1.5 %) stained with ethidium bromide (EtBr) or tris borate EDTA (TBE)-Native-PAGE (8 %) coupled with SYBR Gold staining and imaged using a gel documentation instrument (Bio-rad). The images acquired are depicted after inversion for clarity.

2.5.5. TBE-Native-PAGE and EMSA

The SEC fractions of the LinCas7 and the rLinCas7 were concentrated, and 1 μg of each protein was resolved on TBE (Tris borate EDTA)- Native-PAGE (Polyacrylamide gel electrophoresis) (8 %). SYBR Gold (Invitrogen #S11494) stain was used for visualizing the resolved product.

The nucleic acid (DNA or RNA) copurified with LinCas7 was identified by treating the SEC fraction of LinCas7 protein with either DNase I (1 U), RNase A (2 µg), or benzonase (10 U) independently for 30 min at 37 °C. The reactions were resolved on 8 % TBE- Native-PAGE and visualized after staining with SYBR Gold (Invitrogen #S11494). Electrophoretic mobility shift assays (EMSA) were executed to determine the binding activity of Cas7 proteins on fluorescein-labeled DNA/RNA oligos. The 5′ Fluorescein amidite-labeled (FAM) synthetic RNA or DNA oligos (0.5 µM) of specified sequences (^{FAM}Repeat RNA, ^{FAM}crRNA, or ^{FAM}Repeat DNA) (Table 2.1) were incubated with and without Cas7 proteins (rLinCas7, LinCas7, or DvuCas7c; 3-6 µM) in 1× binding buffer (20 mM Tris-HCl, 100 mM KCl, 5 % glycerol, 1 mM dithiothreitol (DTT), 10 µg/mL heparin, and 0.01 % Igepal; pH 8.0) in presence and absence of Mg^{2+} ions ($MgCl_2$; 10-60 mM) at 37 °C for 1 h. The reactions were resolved on 8 % TBE-Native-PAGE at 4 °C and were visualized directly (unstained) using a gel documentation instrument.

Table 2.1. DNA and RNA substrates used in this study

DNA/RNA substrate	Sequence/Specificity	Size	Source
^{FAM}Repeat RNA	**(5′FAM)**CUGAAUAUAACUUUGAUGCCGUUAGGCGUUGAGCAC	36 nt	Prakash and Kumar, 2021
^{FAM}crRNA	**(5′FAM)**UUGAGCACAAGGGGAAAACAUUCGUCACCCCGUGAAAAACUUCUGAAUAUAACUUUGAUGCCGAUAGGCG	70 nt	Prakash and Kumar, 2021
ϕx174 DNA	Circular ss-DNA	5.3 kb	NEB #N3023S
RNA	Transcript of *LIC10940* gene of *L. interrogans* Copenhageni	264 nt	This study
^{FAM}Repeat DNA	**(5′FAM)**GTGCTCAACGCCTAACGGCATCAAAGTTATATTCAG	36 nt	
Plasmid	pTZ57R/T plasmid with an insert; circular ds-DNA	3.7 kb	
PCR amplicons	Linear ds-DNA	840 bp, 105 bp	
Duplex DNA Oligo	*In vitro* annealed duplex DNA oligo	40 bp	

nt: nucleotide; kb: kilobase pair; bp: base pair

2.5.6. Circular dichroism (CD) spectroscopy

The purified LinCas7 protein variants (LinCas7 and rLinCas7) were brought to a 10 µM final concentration by diluting in a buffer (20 mM Tris-HCl and 20 mM NaCl; pH 8.0). CD spectra in different conditions were recorded on a CD spectrometer (Jasco, model J-1500-150, Jasco Inc., Maryland) using a 1 mm quartz cuvette (Starna Scientific Ltd). The obtained CD spectra

were assessed to compute secondary structure configurations using the program beta structure selection (BeStSel) (Micsonai et al., 2015).

2.5.7. Fluorescence spectroscopy

The intrinsic fluorescence measurement was carried out at 26 °C using the FP-8500 spectrofluorometer (Jasco, Japan)**.** The rLinCas7 (1 μM) in storage buffer-1 was added with Mg^{2+} ions ($MgCl_2$) at varying concentrations (0-300 mM). At an excitation wavelength of 290 nm, fluorescence emission data of rLinCas7 were acquired in the range of 300 to 500 nm at 0.5 nm intervals. At 290 nm wavelength, both tryptophan and tyrosine absorb (Bougie et al., 2003). Emission spectra were recorded with the excitation and emission bandwidth of 5 nm and with a default parameter of high sensitivity. The experiment was performed twice independently. The spectrum was generated with the average value of three scans after background correction by subtracting the signal from the control having the storage buffer-1 mixed with Mg^{2+} ions of variable concentrations (0-300 mM). A saturation binding isotherm was generated by plotting the change in fluorescence intensity at λ_{max} = 342 nm of rLinCas7 ($\Delta F/F_0$) vs. the concentration of Mg^{2+} ions (0-300 mM) as described before (Bougie et al., 2003). In this assay, ΔF represents the magnitude of the difference between the fluorescence intensity at a given concentration of Mg^{2+} ions (F) and the fluorescence intensity in the absence of Mg^{2+} ions (F_0). The Prism v.9 software was used for plotting the graph and curve fitting.

2.5.8. Field emission transmission electron microscopy

The pure rLinCas7 (6 μM) was incubated with crRNA (0.5 μM) in 1× binding buffer containing Mg^{2+} ions ($MgCl_2$; 60 mM) in a total 10 μL reaction volume for 1 h at 37 °C. An identical reaction without crRNA was included as a control. Then, each reaction sample (5 μL) after 2× dilution in nuclease-free water was independently dropped-cast on UV-activated 400-mesh carbon-coated copper grids (Agar Scientific Ltd, UK #AGS147-4H). After adsorption for 5 min, the excess buffer was blotted off with filter paper gently, and the samples were stained with 2 % (w/v) uranyl acetate solution for 2 min. After that, the residual stain was wicked off and air-dried. Electron micrographs were acquired using a JEOL-JEM-2100F field emission transmission electron microscope operating at an accelerating voltage of 200 keV.

*****

Chapter 3

The belly filament subunit, LinCas11b, co-translates from within the *lincas8b* gene of CRISPR-Cas I-B in *Leptospira* [2]

3.1. Summary

Type I CRISPR-Cas systems are highly diverse and constitute the most abundant CRISPR-Cas type in prokaryotes. They develop a ribonucleoprotein complex called Cascade to sense and interfere with foreign DNA. Typically, a functional Cascade complex encloses a belly filament comprised of two Cas11 subunits together with an α-helical C-terminal domain of the large subunit (Cas8 homologs). Nevertheless, most type I CRISPR-Cas subtypes (I-B, I-C, I-D, I-F, and I-G), except I-E and I-A, are devoid of an autonomous gene coding for the Cas11. Thus, the source and engagement of Cas11 subunits in various type I CRISPR-Cas lacking distinct *cas11* genes remain elusive until recently. The present study reports the molecular and biochemical characterization of a Cas11 (LinCas11b) protein from CRISPR-Cas subtype I-B (Lin_I-B) of *Leptospira interrogans*. Within the *lincas8b* gene of Lin_I-B, we have identified a small open reading frame that independently co-translates into LinCas11b (22 kDa). The co-translating LinCas11b exhibits an all α-helix structure that shares the structural analogy with Cas11 family proteins. Pure recombinant LinCas11b (rLinCas11b) forms a homodimer in solution and strongly interacts with LinCas8b. rLinCas11b demonstrates non-specific DNA binding activity independent of divalent metal ions. In sum, LinCas11b, with a non-canonical origin from *lincas8b*, is identified to be a structural and functional analog of Cas11 family proteins. Moreover, further investigation of LinCas11b is warranted to apprehend its role in CRISPR defense response within *Leptospira*.

[2] This chapter is partly adapted from a publication and reprinted with the authors' permission. **Hussain, M.S.**, Anand, V., and 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*, p.124086.

3.2. Results

3.2.1. Overexpression and purification of rLinCas8b protein of *Leptospira*

The open reading frame *LIC10937* (*lincas8b*, 1887 bp) encoding full-length Cas8b protein (LinCas8b) of CRISPR-Cas subtype I-B system in *L. interrogans* was cloned into pET28a plasmid to overexpress recombinant LinCas8 (rLinCas8b, 75 kDa) in *E. coli* **(Appendix Fig. B1a-b)**. The soluble rLinCas8b overexpression was achieved by IPTG induction of *E. coli* culture grown under osmotic stress with sorbitol (0.5 M) **(Appendix Fig. B2a)**. Initially, we attempted to purify the rLinCas8b from soluble cellular fraction by nickel-affinity chromatography. However, rLinCas8b could not be purified adequately because of its inappreciable elution post-binding to affinity chromatography resins **(Appendix Fig. B2b)**. Several elution conditions, including different concentrations of imidazole (0.2-1 M), low pH buffer (pH 4-6), 8 M urea, or guanidine-HCl (6M), were tested for elution of rLinCas8b bound to affinity chromatography resins. However, appreciable elution of rLinCas8b could not be achieved with any of the tested elution conditions. To check the intrinsic binding behavior of LinCas8b to affinity chromatography resins, LinCas8b was overexpressed without any fusion tag by cloning into the pET23a vector **(Appendix Fig. B1c & B2c)**. The overexpressed LinCas8b without any tag (73 kDa) bound to Ni-NTA-agarose beads, demonstrating LinCas8b's intrinsic affinity for purification resins **(Appendix Fig. B2c)**. Ni-NTA resins can also bind the Tryptophan and Cysteine residues in addition to Histidine (Hage et al., 2013). Thus, the high content of Histidine, Tryptophan, and Cysteine (16, 17, and 6 residues, respectively) in LinCas8b could be the reason for its strong binding to Ni-NTA resins.

Next, the purification of rLinCas8b was attempted with cation exchange chromatography (CEX) **(Appendix Fig. B3a-b)**. The elutes obtained from the CEX contained rLinCas8b as well as multiple impurities of larger and smaller sizes. Therefore, the CEX elutes containing rLinCas8b were subjected to the second purification step by SEC **(Appendix Fig. B3c-d)**. To our disappointment, on SEC, rLinCas8b was eluted as a single peak with multiple impurities of larger size. Thus, pure rLinCas8b could not be obtained via CEX coupled with SEC. Alternatively, the full-length *lincas8b* gene was cloned and overexpressed with an N-terminal GST-tag (GST-LinCas8b, 99 kDa) and subjected to glutathione (GSH) affinity chromatography for GST-LinCas8b purification **(Appendix Fig. B1d-e & B4a-c)**. We reasoned that post-binding to GSH-agarose resins, LinCas8b could be released by cleaving the

GST tag through in-bead TEV protease cleavage if in case GST-LinCas8b could not be eluted by reduced glutathione. During purification, the GST-LinCs8b bound to GSH-agarose beads was eluted using reduced glutathione (10-40 mM), although elution was inefficient. Moreover, pure GST-LinCas8b tended to precipitate during the dialysis set to remove glutathione. Therefore, we attempted to elute the LinCas8b post-binding to GSH-agarose beads via in-bead TEV protease cleavage **(Appendix Fig. B4d)**. Alas, after TEV protease cleavage, the untagged LinCas8b could not be eluted appreciably, most likely due to its intrinsic strong binding nature to affinity chromatography resins, as noticed during the purification of 6×His-LinCas8b also. Thus, plagued with the strong binding nature of LinCas8b to affinity chromatography resins, LinCas8b could not be purified adequately to perform any *in vitro* assays toward its biochemical characterization.

Moreover, interestingly, one additional protein of size 22 kDa was found to be overproduced exclusively in BL21 (DE3) cells overexpressing rLinCas8b **(Fig. 3.1A, top panel; Appendix Fig. B1-B4)**. Hereafter, we refer to this additional protein as recombinant LinCas11b (rLinCas11b) due to its similarity to other Cas11 subunits of various Cascade complexes, as identified later in this study.

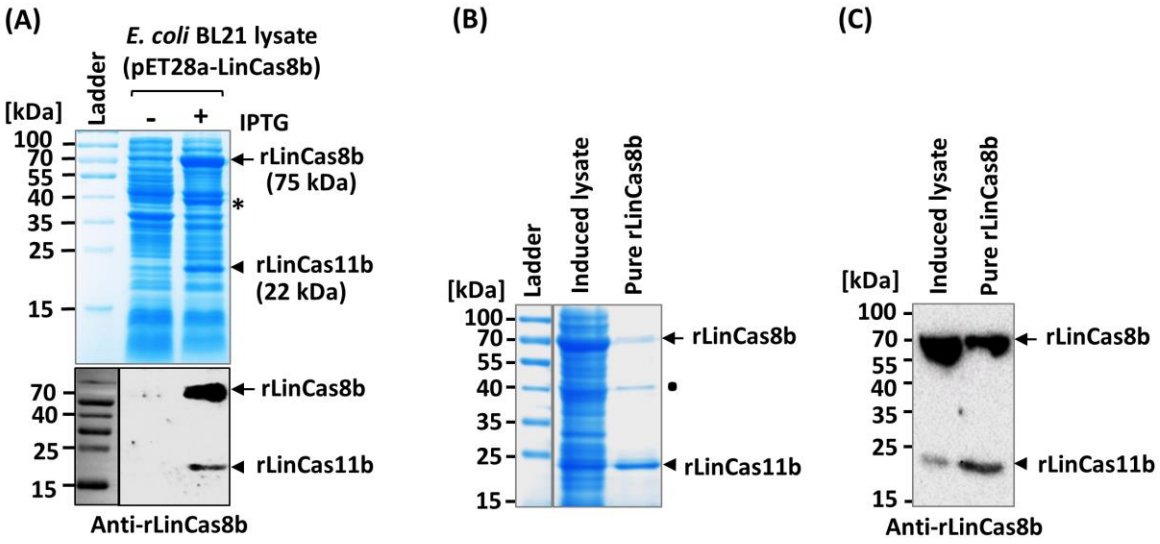

Fig. 3.1. Overexpression and purification of recombinant LinCas8b protein of *Leptospira*. (A) 10 % SDS-PAGE analysis (Coomassie stained) of the whole-cell lysates of uninduced (-) and IPTG-induced (+) *E. coli* BL21 (DE3) cells housing the plasmid pET28a-LinCas8b **(top panel).** Immunoblot with anti-rLinCas8b antibody of *E. coli* BL21 (DE3) lysates overexpressing rLinCas8b **(bottom panel)**. **(legend continued on next page)**

(B) 10 % SDS-PAGE analysis (Coomassie stained) and **(C)** immunoblot of whole-cell lysate of IPTG-induced *E. coli* BL21 cells overexpressing rLinCas8b and the nickel-affinity purified rLinCas8b. The induction of heterologous overexpression of rLinCas8b results in enhanced expression of a protein (marked by an asterisk) that is most likely of host *E. coli* origin. No detection of it (asterisk marked) by anti-rLinCas8b polyclonal antibodies infers its origin from the host. The purified rLinCas8b contains a non-specific protein (marked by a black dot) of *E. coli* origin judged by immunoblot with anti-rLinCas8b polyclonal antibodies.

On immunoblotting, the anti-rLinCas8b antibody could detect the rLinCas8b (75 kDa) as well as the additional protein rLinCas11b (22 kDa) exclusively in the induced whole cell lysate of *E. coli* BL21 (DE3) cells **(Fig. 3.1A, bottom panel)**. Furthermore, the purification of rLinCas8b through nickel-affinity chromatography resulted in copurification of the rLinCas11b (**Fig. 3.1B)** that cross-reacted with anti-LinCas8b on immunoblotting **(Fig. 3.1C)**.

3.2.2. LinCas11b is encoded within *lincas8b* gene

The detection of additional protein rLinCas11b (22 kDa) by anti-rLinCas8b suggested its sequence identity with rLinCas8b and instigated us to probe the source of rLinCas11b from LinCas8b. The analysis of the LinCas8b sequence (628 residues; UniProt ID. Q72TT0) revealed that it is longer than its orthologs in type I-E systems, including EcoCas8e (502 residues; UniProt ID. Q46901) and TthCas8e (502 residues; UniProt ID. Q53VY1) of *E. coli* and *Thermus thermophilus,* respectively. The subtype I systems (I-B, I-C, I-D, I-F, and I-G), except I-E and I-A, lack an independent gene coding for the Cas11 subunit of the Cascade complex (Kira S Makarova et al., 2011; Makarova et al., 2015). Instead, they encode a large subunit Cas8 (or Cas10d in type I-D) with an extended C-terminus consisting of largely α-helices (Kira S Makarova et al., 2011; Makarova et al., 2015). These α-helix-rich C-terminus of Cas8 and Cas10d proteins have been postulated to be a functional substitute for Cas11 subunits in the Cascade complex of subtype I-B, I-C, I-D, I-F, and I-G systems (Kira S Makarova et al., 2011; Makarova et al., 2015). In agreement, recently, CthCas8b (613 residues; UniProt ID. A3DKB7) and MmaCas8b (650 residues; UniProt ID. A4FXZ7) of type I-B system in *Clostridium thermocellum* and *Methanococcus maripaludis,* respectively, have been described to contain α-helix rich extended C-terminus (Richter et al., 2017). Moreover, the CthCas8b and MmaCas8b were reported to undergo auto-fragmentation resulting in the generation of a Cas11-like C-terminal fragment of ~17 kDa (Richter et al., 2017).

Similarly, in this study, the Phyre2 program (Kelley et al., 2015) predicted an α-helix rich (84 %) (0 % β-strand) secondary structure composition for the extended C-termini of LinCas8b (200 residues; Met429-Ala628) with a high confidence level **(Fig. 3.2)**. Also, the size of the extended C-termini (23 kDa) of LinCas8b aligns with the co-expressing LinCas11b's size (22 kDa) on SDS-PAGE.

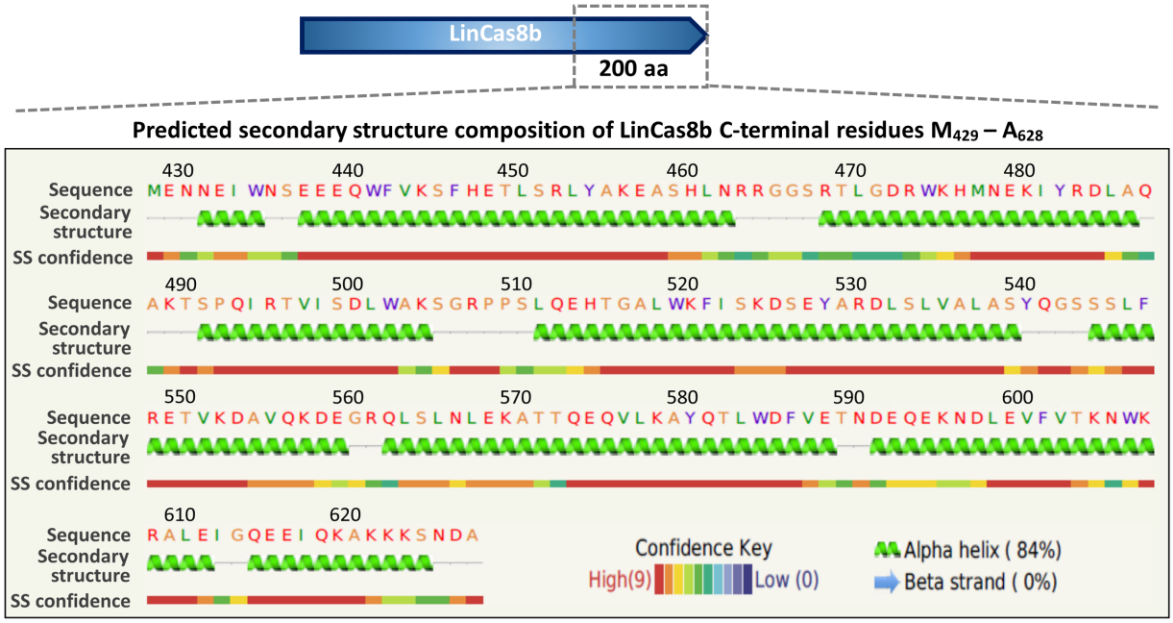

Fig. 3.2. The α-helix rich C-terminal of LinCas8b. The secondary structure elements of the LinCas8b C-terminus (200 residues; Met429-Ala628), predicted by Phyre2 program, demonstrates an α-helix rich composition. SS stands for secondary structure.

Therefore, overexpression of the truncated variants of LinCas8b (rLinCas8b$^{\Delta N}$, 53 kDa; rLinCas8b$^{\Delta C}$, 53 kDa) in *E. coli* was explored **(Fig. 3.3A & Appendix Fig. B1f-g)**. On immunoblotting with anti-rLinCas8b, rLinCas11b (22 kDa) was evident in cells overexpressing either full-length rLinCas8b or rLinCas8b$^{\Delta N}$ but not in the cells producing rLinCas8b$^{\Delta C}$ **(Fig. 3.3B)**. This confirmed LinCas8b's C-terminus as the source of LinCas11b. Moreover, results of immunoblotting with anti-rLinCas8b antibody negate the auto-fragmentation hypothesis for rLinCas8b as its cleavage may have resulted in the detection of a third intermediate-size protein (53 kDa) in addition to full-length rLinCas8b and rLinCas11b. Thus, we speculated that the LinCas11b is a co-translational product from the C-termini of LinCas8b rather than a result of LinCas8b's auto-fragmentation.

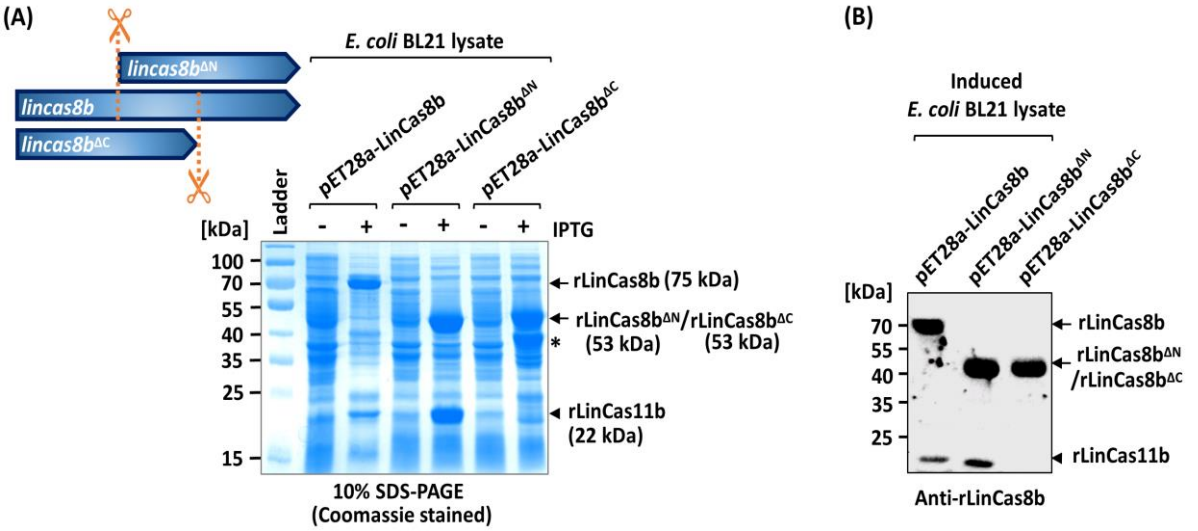

Fig. 3.3. LinCas11b originates from the C-terminal of LinCas8b. (A) The schematics displayed on top-left shows the partial ORFs (*lincas8b*$^{\Delta N}$ and *lincas8b*$^{\Delta C}$) of full-length *lincas8b* individually cloned into pET28a plasmid to overexpress N- and C-terminal truncated variants of LinCas8b, rLinCas8b$^{\Delta N}$ and rLinCas8b$^{\Delta C}$, respectively. SDS-PAGE analysis of the whole-cell lysates of uninduced (-) and IPTG-induced (+) *E. coli* BL21 (DE3) cells overexpressing the full-length rLinCas8b or its truncated variants (rLinCas8b$^{\Delta N}$ or rLinCas8b$^{\Delta C}$). **(B)** Immunoblot of whole-cell lysates of *E. coli* BL21 cells overexpressing full-length rLinCas8b or its truncated variants (rLinCas8b$^{\Delta N}$ or rLinCas8b$^{\Delta C}$). The induction of heterologous overexpression of LinCas8b or its variants also results enhanced expression of a protein (marked by asterisk) that most likely of host *E. coli* origin, inferred by no detection of it by anti-rLinCas8b polyclonal antibodies.

Next, we searched for the alternative translation start codon within the *lincas8b* gene using the ribosome binding site (RBS) calculator program (Salis et al., 2009). The *in-silico* analysis of the *lincas8b* gene revealed a highly active (ΔG_{Total} = -2.76 kcal/mol) in-frame translation start codon (ATG encoding Met429) near downstream of an RBS-like GA-rich stretch **(Fig. 3.4A)**. The in-frame translation start codon has been previously correlated to the total Gibbs free energy change (ΔG_{Total}) required for the base pairing of the ribosome to mRNA sequence (Salis et al., 2009). To validate the *in silico* analysis, the predicted internal start codon ATG (Met429) was mutated to GCG (Ala429) in *lincas8b* within the pET28a-LinCas8b plasmid. Immunoblot of induced *E. coli* lysate overexpressing rLinCas8b mutant variant (rLinCas8b^{M429A}) demonstrated complete abolition in co-translation of rLinCas11b **(Fig. 3.4B-C)**. Our finding of LinCas11b protein encoded within the *lincas8b* also agrees with the recent reports of Cas11 co-translation in other organisms harboring CRISPR-Cas type I-C (*Neisseria lactamica* and *Desulfovibrio vulgaris*) (O'Brien et al., 2020; Tan et al., 2022) and I-D systems (*Synechocystis* spp.) (McBride et al., 2020).

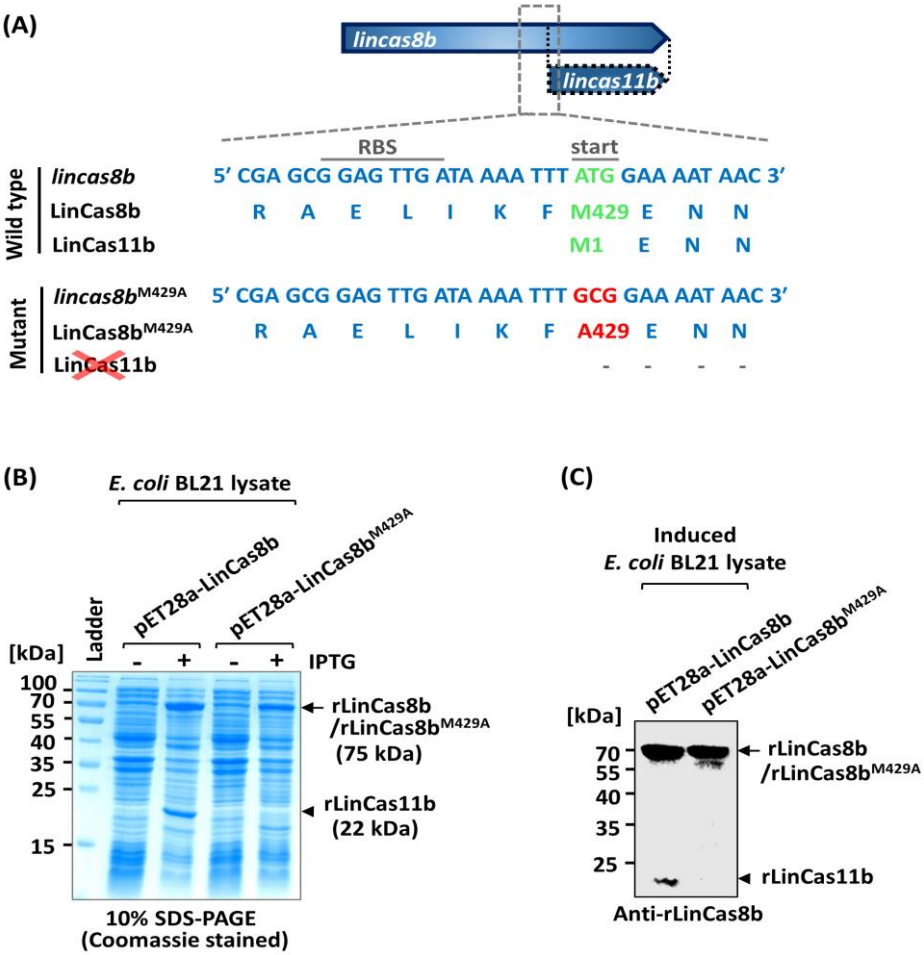

Fig. 3.4. LinCas11b co-translates from an in-frame internal translation start codon encoded within the *lincas8b* gene. **(A)** Schematics of *lincas8b* and *lincas11b* ORF. The codons within *lincas8b* [ATG (M429) to TAG (stop codon)] represent a small fused *lincas11b* ORF. The alternative RBS and internal translation start codon (ATG in green) within the *lincas8b* are indicated and labeled. The amino acid residues in LinCas8b and LinCas11b are indicated below the corresponding codons in the *lincas8b*. The codon marked in red fonts indicates the mutation introduced in the *lincas8b* to overexpress LinCas8b^{M429A} variant without LinCas11b co-translation. **(B)** SDS-PAGE analysis of the whole-cell lysates of uninduced (-) and IPTG-induced (+) *E. coli* BL21 (DE3) cells overexpressing the full-length rLinCas8b or its mutant variant rLinCas8b^{M429A}. **(C)** Immunoblot, using anti-rLinCas8b antibody, of induced whole-cell lysates of *E. coli* BL21 cells overexpressing rLinCas8b or rLinCas8b^{M429A}. Mutation of the internal start codon within *lincas8b* results in the abolition of the rLinCas11b co-translation.

3.2.3. LinCas11b shares structural similarities with its orthologs

In different type I and III CRISPR-Cas systems, the large (Cas8 and Cas10) and small subunit (Cas11) proteins exhibit very low or insignificant sequence similarity with their homologs (Makarova et al., 2020a; Venclovas, 2016; Vestergaard et al., 2014). Nevertheless, these Cas proteins (Cas8, Cas10, and Cas11) exhibit structural homology with their homologs and thus occupy analogous positions in the effector complex of type I and III systems (Jackson and

Wiedenheft, 2015; Liu and Doudna, 2020; Venclovas, 2016). Similarly, no detectable sequence similarity was noted on the BLASTP (Johnson et al., 2008) alignment of LinCas8b or the unconventionally encoded LinCas11b to their respective homologs from various bacteria. The structure of pure Cas8b, Cas11b, or the Cascade complex (type I-B) has not yet been determined experimentally. In this study, the tertiary structure prediction of LinCas8b and LinCas11b via Swiss-Model (Waterhouse et al., 2018) or I-TASSER server (Roy et al., 2010) could not produce quality models. Moreover, the LinCas8b (AF-Q72TT0) model structure, available from the AlphaFold database (Jumper et al., 2021; Varadi et al., 2022), allowed the analysis of LinCas11b's structural features. Since, LinCas11b (200 residues, Met1-Ala200) and C-terminal of full-length LinCas8b (200 residues, Met429-Ala628) have an identical amino acid sequence, they are anticipated to exhibit highly similar tertiary structure folding. Therefore, the retrieved tertiary structure of LinCas8b's C-terminal (Met429-Ala628) was considered for LinCas11b structural analysis. In agreement with this study, previously, SynCas11b, CthCas11b, and MmaCas11b were reported to be encoded within SynCas8b (Tan et al., 2022), CthCas8b (Richter et al., 2017), and MmaCas8b (Richter et al., 2017), respectively. Therefore, to compare with LinCas11b, the tertiary structure of SynCas11b, CthCas11b, and MmaCas11b were also retrieved from the respective full-length Cas8b proteins available in the AlphaFold database.

The structural similarity between LinCas11b and its characterized orthologs across various type I and III systems was assessed by structural alignment using the Dali web server (Holm and Laakso, 2016). For structural alignment, SynCas11b, CthCas11b, and MmaCas11b were included to represent subtype I-B Cas11 proteins. In addition, one structural representative each from I-A (SsoCsa5 (Reeks et al., 2013a)), I-E (EcoCas11e (Zhao et al., 2014)), I-C (DvuCas11c (O'Brien et al., 2020)), I-D (SynCas11d (McBride et al., 2020)), type III-A (SepCsm2 (Takeshita et al., 2019)), and type III-B (AfuCmr5 (Osawa et al., 2015)) were compared. The results of structural comparison based on the Dali Z-score have been represented as a structural similarity matrix (heat map) **(Fig. 3.5)**. The similarity between two protein structures is proportional to the Z-score achieved. However, a Z-score of less than two is considered spurious (Holm and Laakso, 2016). The closest structural match for LinCas11b noted in this study was SynCas11b (Z-score 9.6), followed by CthCas11b (5.6) and SsoCsa5 (5.5). The lowest similarity of LinCas11b was detected with SepCsm2 (2.5) and SynCas11d (2.1). Within

the given set of Cas11 orthologs, the highest similarity in tertiary structure was between CthCas11b and MmaCas11b (Z-score 14).

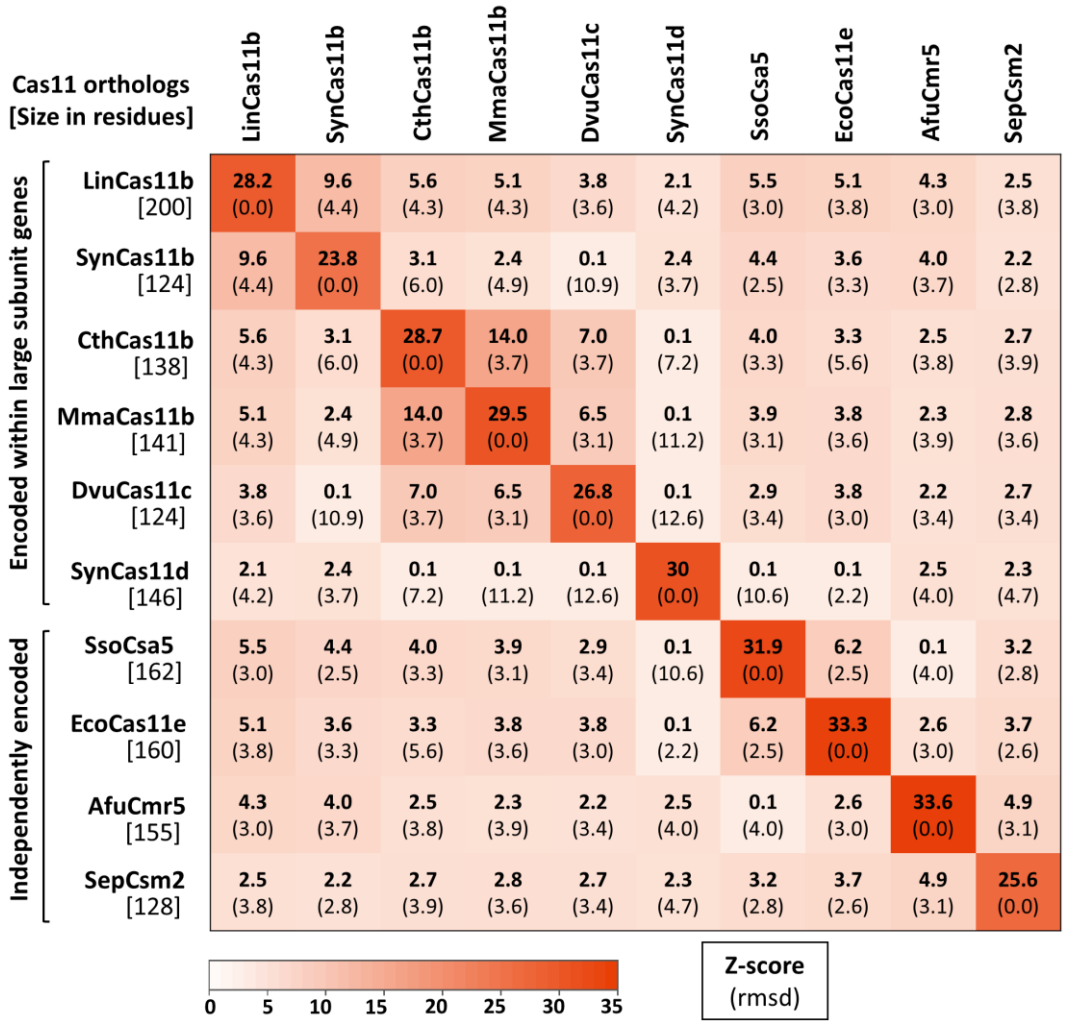

Fig. 3.5. The structural similarity between selected Cas11 orthologs. Structural alignment of selected Cas11 orthologs was performed against each other using the Dali web server. The resulting similarity matrix based on Z-score is presented as a heat map. For each structural alignment, the root mean square deviation (rmsd in Å) over all C_α atoms is indicated in parenthesis. The Cas11 orthologs used for comparison are LinCas11b [associated with subtype I-B of *Leptospira interrogans*, encoded within LinCas8b (AlfaFold ID AF-Q72TT0), residues Met429-Ala628, ΔG_{Total} = -2.76 kcal/mol], SynCas11b [I-B, *Synechocystis* spp., SynCas8b (AF-A0A068N831), Met396-Glu519, ΔG_{Total} = -6.48 kcal/mol], CthCas11b [I-B, *Clostridium thermocellum*, CthCas8b (AF-A3DKB7), Met476-Glu613, ΔG_{Total} = -3.49 kcal/mol], MmaCas11b [I-B, *Methanococcus maripaludis*, MmaCas8b (AF-A4FXZ7), Met510-Glu650, ΔG_{Total} = -4.65 kcal/mol], DvuCas11c [I-C, *Desulfovibrio vulgaris*, DvuCas8c (PDB ID 7KHA chain K), Val489-Asn612, ΔG_{Total} = -0.99 kcal/mol], SynCas11d [I-D, *Synechocystis* spp., SynCas10d (7SBA chain K), Met830-Asn975, ΔG_{Total} = -5.47 kcal/mol], SsoCsa5 [I-A, *Saccharolobus solfataricus* (3ZC4 chain A)], EcoCas11e [I-E, *Escherichia coli*, (4U7U chain B)], AfuCmr5 [III-B, *Archaeoglobus fulgidus*, (3X1L chain G)], and SepCsm2 [III-A, *Staphylococcus epidermidis*, (6AE1 chain A)].

The LinCas11b structure analysis revealed a helical bundle topology arranged as two domains- an N-terminal domain (NTD) with five α-helices (α1- α5) and a C-terminal domain (CTD) with four α-helices (α6- α9) **(Fig. 3.6A)**. Analysis of the other large subunit-encoded Cas11 orthologs (SynCas11b, CthCas11b, MaCas11b, DvuCas11c, and SynCas11d) revealed that they share a similar topology with LinCas11b **(Fig. 3.6A-5F)**. The given set of Cas11 orthologs, including LinCas11b, shares a common α-helical fold of five helices, organized as two intercalating helix pairs (α1- α2 and α4-α5), each resembling the letter "V" and connected by the α3-helix **(Fig. 3.6G)**. The α-helical fold common in the large subunit-encoded Cas11 orthologs **(Fig. 3.6G)** closely matches with the consensus fold **(Fig. 3.6H)** identified in the independently-encoded Cas11 proteins, including EcoCas11e, SsoCsa5, AfuCmr5, and SepCsm2 (Osawa et al., 2015; Takeshita et al., 2019; Venclovas, 2016). Nevertheless, among all Cas11 orthologs, the corresponding helices' size and orientation vary (Osawa et al., 2015; Takeshita et al., 2019; Venclovas, 2016).

LinCas11b has an additional α-helical domain (Ser117- Ala200; CTD) consisting of four helices (α6-α9) **(Fig. 3.6A)**, though the model prediction confidence score for CTD is very low [predicted local distance difference test (pLDDT) < 50]. Moreover, such additional elements are also evident in CthCas11b (α1′) **(Fig. 3.6C)** and MmaCa11b (α1′) **(Fig. 3.6D)**. Additional elements are also present in SsoCsa5 (Reeks et al., 2013a) and Cas11e (also termed as Cse2 or CasB) of *E. coli* (EcoCas11e) (Zhao et al., 2014), *T. thermophilus* (TthCse2) (Nam et al., 2012b), and *Thermobifida fusca* (TfuCse2) (Nam et al., 2012b). Cas11e proteins exhibit a two-domain all α-helical topology (NTD with four α-helices and CTD with five α-helices) (Nam et al., 2012b; Venclovas, 2016). While SsoCsa5 contains the consensus α-helical domain with the insertion of a weakly conserved β-strand domain (Reeks et al., 2013a; Venclovas, 2016). The importance of these additional elements (CTD in LinCas11b, α1′ in CthLinCas11b, and MmaCas11b) in the functioning of Cas11b proteins requires further investigation. Even though LinCas11b demonstrates no significant sequence similarity, a substantial structural homology with its orthologs from various type I and III systems is evident and thus may have the same function.

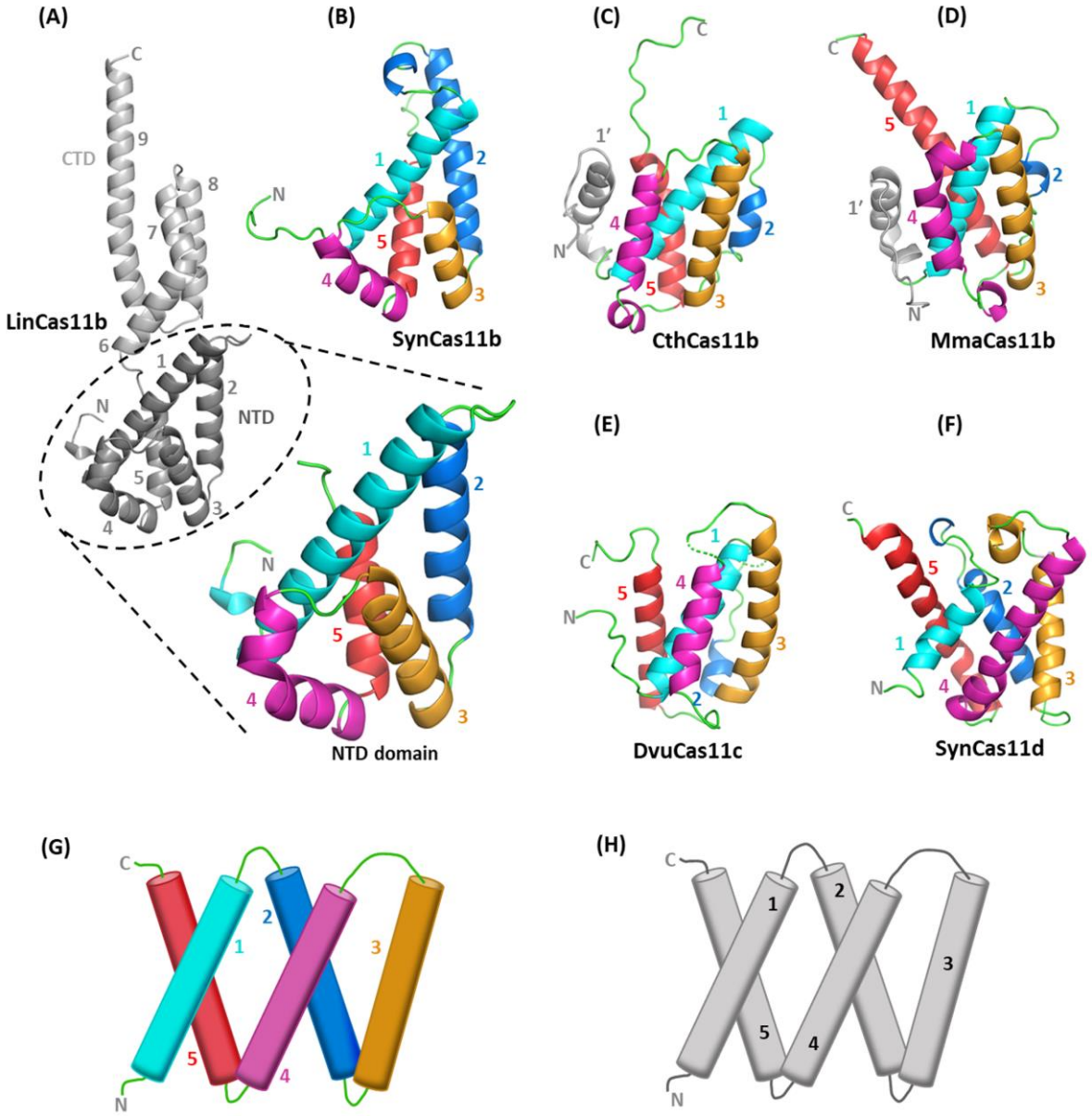

Fig. 3.6. Structural comparison of LinCas11b with large-subunit encoded Cas11 proteins. Structures compared are **(A)** LinCas11b (LinCas8b's C-terminus, residues Met429-Ala628; AF-Q72TT0), **(B)** SynCas11b (SynCas8b's C-terminus, residues Met396-Glu519; AF-A0A068N831), **(C)** CthCas11b (CthCas8b's C-terminus, residues Met476-Glu613; AF-A3DKB7), **(D)** MmaCas11b (MmaCas8b's C-terminus, residues Met510-Glu650; AF-A4FXZ7), **(E)** DvuCas11c (DvuCas8c's C-terminus, residues Val489-Asn612; PDB ID 7KHA chain K), **(F)** SynCas11d (SynCas10d's C-terminus, residues Met830-Asn975; 7SBA chain K). LinCas11b has a helical bundle topology arranged as two domains- NTD (α1-α5; dark gray) and CTD (α6-α9; light gray). Structures are presented in a similar orientation, and corresponding α-helices (α1-α5) are indicated in the same color. The additional elements (α1′) in CthCas11b and MmaCas11b are indicated in light gray. **(G)** The architectural layout of the common α-helical fold shared by Cas11 orthologs (encoded within large subunit proteins). **(H)** Schematic representation of the consensus α-helical fold previously deduced in the Cas11 proteins family (encoded independently) (Venclovas, 2016).

3.2.4. LinCas11b forms a homodimer and exhibits non-specific affinity for nucleic acids

The role of LinCas11b is anticipated to constitute a functional LinCascade, similar to its structural homologs in other organisms harboring CRISPR-Cas type I and type III systems (McBride et al., 2020; O'Brien et al., 2020; Venclovas, 2016). To better understand the molecular and biochemical properties of LinCas11b, the determined ORF (Open reading frame) was cloned in an expression plasmid (pCDF1b-LinCas11b), overexpressed, and purified (6×His-LinCas11b, 26 kDa) from *E. coli* **(Fig. 3.7A & Appendix Fig. B1h-i)**. During the second purification step by size exclusion chromatography, rLinCas11b resolved at ~48 kDa, close to its dimeric size **(Fig. 3.7B)**.

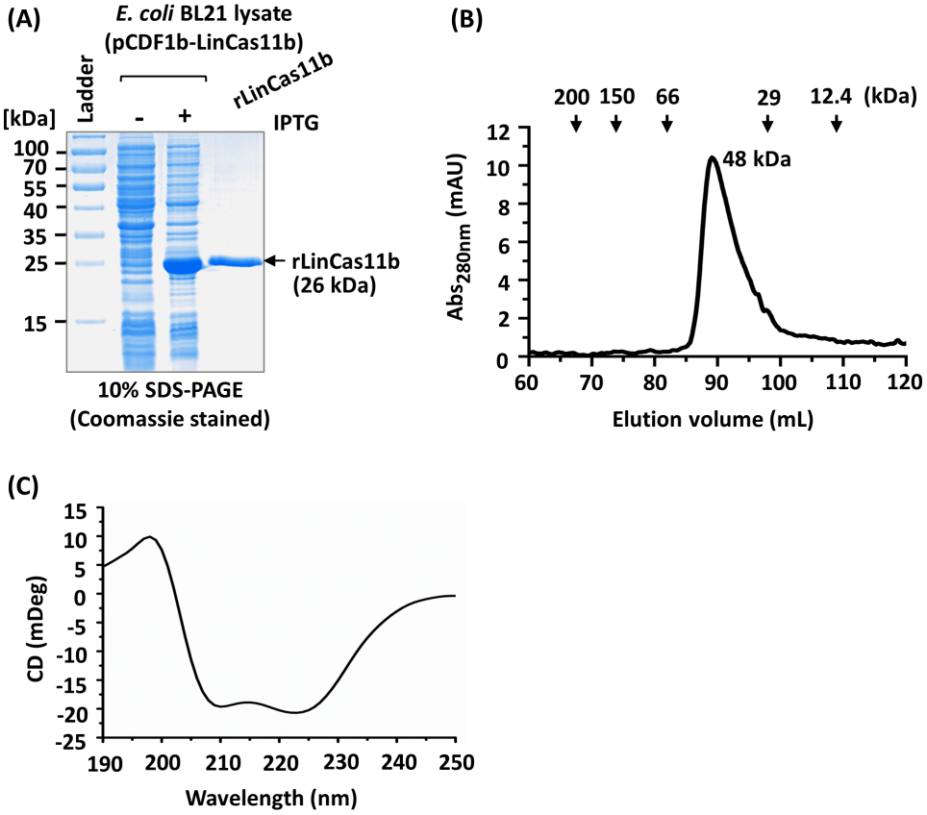

Fig. 3.7. Molecular characterization of pure rLinCas11b. (A) SDS-PAGE analysis of the whole-cell lysates of uninduced (-) & IPTG-induced (+) *E. coli* BL21 (DE3) cells housing pCDF1b-LinCas11b plasmid and purified 6×His-tagged recombinant LinCas11b (rLinCas11b, 26 kDa). **(B)** The size exclusion chromatography (SEC) of nickel-affinity purified rLinCas11b. The rLinCas11b resolved at size of ~48 kDa close to its dimeric size. The elution profiles of standard molecular weight protein markers are shown with inverted arrows on top of the chromatogram (β-Amylase, 200 kDa; Alcohol dehydrogenase, 15 kDa; Albumin, 66 kDa; Carbonic anhydrase, 29 kDa; Cytochrome c, 12.4 kDa). **(C)** Circular dichroism (CD) spectra of rLinCas11b. The estimated secondary structure compositions of rLinCas11b by CD spectroscopy are 80.51 % α-helix, 0.68 % β-strand and 18.81 % others.

Next, CD spectroscopy of pure rLinCas11b revealed an α-helix rich (80.51 % α-helix, 0.68 % β-strand, 18.81 % others) secondary structure composition **(Fig. 3.7C)** that matches with the *in-silico* predictions **(Fig. 3.2 & 3.6A)**, suggesting the integrity of protein is sustained post-purification.

The pure Cas11e (CasB) ortholog has been biochemically demonstrated as a non-specific DNA-binding protein from various bacteria harboring the CRISPR-Cas subtype I-E system (Nam et al., 2012b). Therefore, the DNA binding activity of pure rLinCas11b was assayed. In the DNA binding assay, the rLinCas11b (at increasing concentrations; 2-8 µM) demonstrated an affinity for linear ds-DNA (PCR amplicons, non-methylated) without any nuclease activity and independent of divalent metal ions **(Fig. 3.8A)**. The supplementation of divalent metal ions (Mg^{2+}, 10 mM) or chelation of possible metal traces (if any) using EDTA (10 mM) in the reaction did not influence the rLinCas11b's DNA binding activity. Likewise, rLinCas11b demonstrated an indiscriminate affinity for circular ds-DNA (Plasmid; methylated) and circular ss-DNA (ϕx174 virion DNA) substrates **(Fig. 3.8B)**. In the presence of rLinCas11b, the detection of DNA substrates in the wells on gel electrophoresis was most likely due to the non-specific binding of rLinCas11b to DNA that restricted the DNA migration **(Fig. 3.8A-B)**.

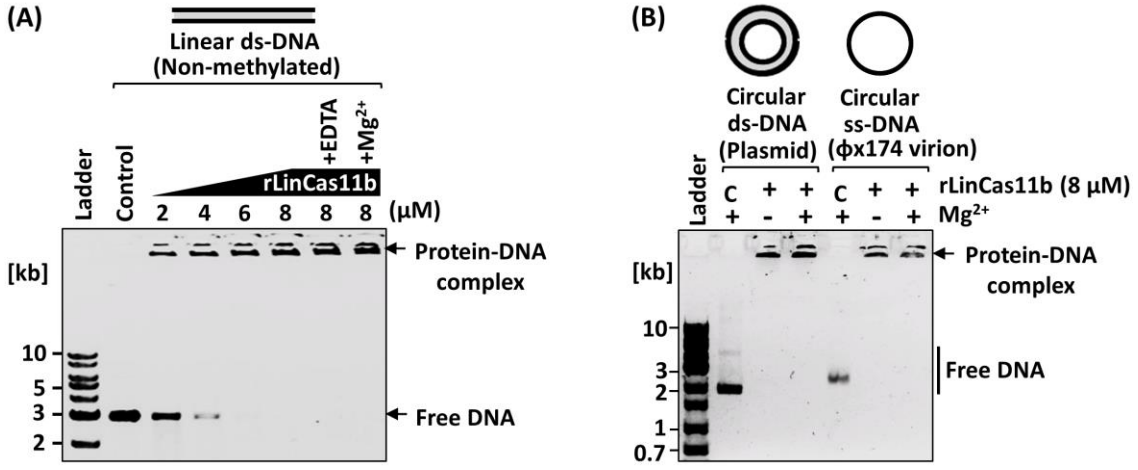

Fig. 3.8. Pure rLinCas11b binds DNA non-specifically. (A) DNA binding assay of rLinCas11b (2-8 µM) on a linear ds-DNA substrate (PCR amplicon, nonmethylated; 3.5 nM). EDTA (10 mM) served as a divalent metal ion chelator. **(B)** DNA binding assay of rLinCas11b (8 µM) on a circular ds-DNA (plasmid, methylated; 3.5 nM) and a circular ss-DNA (ϕx174 virion DNA; 3.5 nM) substrates. Assays were performed in the presence and absence of Mg^{2+} ions ($MgCl_2$; 10 mM) at 37 °C for 1 h and resolved on EtBr-stained 1 % agarose gels (native). Gel images are presented after inversion for clarity. At least two independent experiments confirmed the results represented.

Thus, the biochemical characteristics demonstrated by rLinCas11b are consistent with previously characterized Cas11e (CasB) proteins encoded by independent genes in *E. coli*, *T. thermophilus*, and *Thermobifida fusca* (Nam et al., 2012b).

3.3. Discussion

The type I CRISPR-Cas systems (I-B, I-C, I-D, I-F, and I-G), except I-E and I-A, are devoid of an autonomous gene coding for the Cas11 (Kira S Makarova et al., 2011; Makarova et al., 2015). Nevertheless, a functional Cascade complex contains a belly filament of two Cas11 subunits concurrently with an α-helical C-terminal domain of the large subunit (Cas8 homologs) (Liu and Doudna, 2020; Venclovas, 2016). Recently, in subtype I-C and I-D systems, the Cas11 was documented to be expressed from an internal translation start site within the large subunit's ORF (McBride et al., 2020; Tan et al., 2022). Also, the internal translation of Cas11 from *cas8* was hypothesized to be an overall common means for Cas11 production in type I systems that lack distinct *cas11* genes (McBride et al., 2020). Therefore, LinCas11b expression from the *lincas8b* agrees with the previously hypothesized mechanism for Cas11 production in various type I systems. Moreover, deviation from the suggested general mechanism for Cas11 expression in type I systems lacking distinct *cas11* genes has also been reported. The generation of a Cas11-like polypeptide (~17 kDa) in *M. maripaludis* and *C. thermocellum* was proposed to result from auto-cleavage of full-length Cas8b (MmaCas8b and CthCas8b) at C-terminal instead of co-translation from *cas8b* genes (Richter et al., 2017). Met510 and Met476 in MmaCas8b and CthCas8b, respectively, were identified as the 1st residue of Cas11-like C-terminal fragments (~17 kDa) (Richter et al., 2017). Also, the codons encoding Met510 and Met476 in full-length MmaCas8b and CthCas8b, respectively, represent the potential internal translation start codon with high confidence (ΔG_{Total} = -4.65 and -3.49 at codons encoding Met510 and Met476 in MmaCas8b and CthCas8b, respectively). However, an alanine-coding substitution mutation of the predicted potential internal translation start codon failed to abolish the production of the C-terminal fragment (Richter et al., 2017). This presents the apparent deviation from the suggested general mechanism for Cas11 production in type I systems, thus necessitating experimental validation for each CRISPR-Cas type I system.

Further, the structural homology of LinCas11b to various Cas11 proteins in different type I and III systems suggests its identity and function as a small subunit protein in the LinCascade

complex. The extended α-helical C-terminals of the large subunits in various type I systems (I-B, I-C, and I-D) represent a fused Cas11 protein. Such extended α-helical C-terminals presumably act as nucleation points for the assembly of Cas11 belly filament for the Cascade (McBride et al., 2020; O'Brien et al., 2020). Similarly, in this study, the rLinCas11b co-purification with rLinCas8b suggests a sturdy inter-molecular interaction between the two proteins. The rLinCas11b dimerization property witnessed on SEC also implies a plausible binding of LinCas11b to the C-terminal domain (fused LinCas11) of full-length LinCas8b in the LinCascade complex. Such strong self-assembling protein-protein interactions using Cas8's extended C-terminus as a nucleation point are sufficient to form the belly filament of a Cascade complex (McBride et al., 2020). Agreeing with this, pure rLinCas11b did not demonstrate perceptible binding to rLinCas7 on SEC **(Appendix Fig. B5)** and isothermal titration calorimetry (K_d = 152 ±10.5 μM) **(Appendix Fig. B6)**. Nevertheless, in type III and I-E systems, the formation of the belly filament in the effector complex usually requires the interaction of Cas11 with additional subunits, including Cas5 and Cas7 (Liu and Doudna, 2020; Reeks et al., 2013b; Venclovas, 2016). In the type III system of *Pyrococcus furiosus*, an unfused (separately encoded) Cas11 (pfuCmr5) was depicted to bind the Cas7 subunit (pfuCmr4) (Park et al., 2013). However, in the type I-A system of *Sulfolobus solfataricus*, the unfused Cas11 (SsoCsa5) does not exhibit binding with SsoCas7 (Reeks et al., 2013a). Furthermore, the self-oligomerization is evident for many unfused Cas11 orthologs, including TfuCse2 (dimer; type I-E, *T. fusca*) (Nam et al., 2012b), TmaCsm2 (dimer; type III-A, *Thermotoga maritima*) (Venclovas, 2016), TthCmr5 (trimer; type III-B, *T. thermophilus*) (Sakamoto et al., 2009), and SsoCsa5 (multimer; type I-A, *S. solfataricus*) (Reeks et al., 2013a). While, TthCse2 (type I-E, *T. thermophilus*) (Nam et al., 2012b) and PfuCmr5 (type III-B, *P. furiosus*) (Park et al., 2013) are the Cas11 orthologs that exhibit monomeric states. Although the prevailing oligomerization of Cas11 is common, it is not a conserved trait of Cas11 family proteins and thus mandates separate assessment.

In addition, the rLinCas11b revealed non-specific DNA binding activity similar to that of previously characterized unfused Cas11 proteins, including EcoCas11e, TthCse2, and TfuCse2 (Nam et al., 2012b). The non-specific DNA binding activity in Cas11 family proteins conforms with its role in Cascade-mediated DNA interference: (1) facilitate the initial binding of Cascade to target DNA for PAM recognition; (2) bind to the displaced non-target strand to prevent the collapse of R-loop (Liu and Doudna, 2020; McBride et al., 2020). Moreover, to better interpret

the role of Cas11b proteins, including LinCas11b, in subtype I-B adaptive immunity, their structure determination alone or in complex with Cas8b or subtype I-B Cascade is warranted.

3.4. Conclusion

The LinCas8b is a genetic fusion of large (LinCas8b) and belly filament (LinCas11b) subunit proteins of the CRISPR-Cas I-B Cascade complex in *L. interrogans*. The *lincas8b* gene encloses a small ORF (*lincas11b*) that codes for the C-terminal of LinCas8b as well as independently co-translates into LinCas11b protein. LinCas11b's self-dimerization and its stable interaction with LinCas8b present molecular insights into the assembly of belly filament in a Cascade complex of various CRISPR-Cas type I systems that lack distinct *cas11* genes. Taken together, the present study demonstrates the co-translating LinCas11b to be a structural and functional analog of Cas11 family proteins (Venclovas, 2016) that constitutes the belly filament in the type I and III CRISPR-Cas effector complex.

3.5 Materials and methods

3.5.1. Bacterial strains, media, and growth conditions

The spirochete *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 culture was maintained in EMJH medium (Difco, Becton Dickinson India Pvt. Ltd., New Delhi, India) as described in 2.5.1. The *E. coli* strains DH5α (NEB #C2987H) and BL21 (DE3) (Novagen #69450-M) were used for cloning and recombinant proteins' overexpression purposes, respectively. Both *E. coli* strains were grown in LB (broth or semi-solid) medium vide section 2.5.1.

3.5.2. Molecular cloning

The DNA sequences of interest were PCR amplified individually from the genomic DNA of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 using sequence-specific primer pairs and cloned into different plasmids, as detailed below. The target DNA sequence-specific primers and plasmids used in this study are listed in (Appendix Table B1). The full-length *lincas8b* gene (*LIC10937*, 1887 bp) was cloned into the pET28a expression vector between *Nhe*I and *Sal*I restriction sites. The resulting plasmid (pET28a-LinCas8b) encodes full-length recombinant LinCas8b with an N-terminal 6×His tag (rLinCas8b, 75 kDa). Similarly, two truncated variants of *lincas8b* (*lincas8b*$^{\Delta N}$ and *lincas8b*$^{\Delta C}$) were PCR amplified and cloned into

pET28a to overexpress rLinCas8b$^{\Delta N}$ (59 kDa) and rLinCas8b$^{\Delta C}$ (59 kDa) with a deletion of 200 residues from the N-terminus and C-terminus, respectively. To overexpress the full-length LinCas8b with an N-terminal GST-tag, *lincas8b* was cloned into the pGST-TEV expression vector between *Bam*HI and *Sal*I restriction sites. The fused *lincas11b* ORF identified within the *lincas8b* in this study (603 bp from 3′ end) was PCR amplified and ligated into the pCDF-1b plasmid between *BamH*I and *Sal*I sites to overexpress the recombinant protein with an N-terminal 6×His tag (rLinCas11b; 26 kDa). A single amino acid substitution mutation (Met429 to Ala) was introduced in *lincas8b* encoded within the plasmid pET28a-LinCas8b using Q5 site-directed mutagenesis kit (NEB #E0554S) to overexpress rLinCas8b^{M429A} to intervene co-translation of rLinCas11b. All the plasmid constructs were sequenced before downstream uses.

3.5.3. Protein overexpression and purification

For heterologous overexpression of recombinant proteins, respective plasmid constructs were individually transformed into *E. coli* BL21 (DE3) cells. To check the overexpression of the full-length rLinCas8b or its variants (rLinCas8b$^{\Delta N}$, rLinCas8b$^{\Delta C,}$ or rLinCas8b^{M429A}), *E. coli* BL21 (DE3) cells housing the respective plasmids (pET28a-LinCas8b, pET28a-LinCas8b$^{\Delta N}$, pET28a-LinCas8b$^{\Delta C}$ or pET28a-LinCas8b^{M429A}) were induced with IPTG (1 mM) at culture OD_{600}= 0.6 for next 4 h at 37 °C. For purification purposes, the soluble rLinCas8b overexpression in *E. coli* was also examined by inducing with IPTG (1 mM) at 18 °C for 20 h. Moreover, the soluble rLinCas8b overexpression could be achieved by inducing the BL21 cells grown under osmotic stress, as described before (Oganesyan et al., 2007). Briefly, the *E. coli* BL21 cells (pET28a-LinCas8b) were grown to culture OD_{600} = 0.4-0.5 in 0.5 M Sorbitol containing LB media at 37 °C. Then, the culture was transferred to a shaking water bath (42 °C) and incubated till culture OD_{600}= 0.8-0.9, followed by induction with 1 mM IPTG for the next 20 h at 18 °C. Post-induction, cells were harvested and washed once with PBS (pH 7.4) by centrifugation at 3000 × g for 10 min. Then cells were processed for rLinCas8b purification from soluble cellular fraction by Ni-NTA affinity chromatography under native conditions following the similar procedure described in section 2.5.2. Briefly, the induced cell pellet (two liter culture) was resuspended in 30 mL of chilled lysis buffer-1 (50 mM Tris-HCl, 300 mM NaCl, 1 % Triton X-100, 10 % glycerol, and 10 mM imidazole; pH 8.0) supplemented with 1 mM PMSF (Phenylmethylsulfonyl fluoride). Following a 30 min incubation on ice, the cell suspension was subjected to sonication for 5 min at 30 % amplitude with 5 sec on and 10 sec off cycles. The resulting lysate was clarified by centrifugation at 13000 × g for 20 min at 4 °C.

The clarified lysate was allowed to bind to the Ni-NTA-agarose resins (1 mL) (GCC Biotech, India #GPC-01D) pre-equilibrated with lysis buffer-1 in a 10 mL gravity-flow column. Post-binding, the resins were washed with 50 mL of chilled wash buffer-1 (50 mM Tris-HCl, 300 mM NaCl, and 10 % glycerol; pH 8.0) supplemented with 1 mM PMSF and 30 mM imidazole. Then, the recombinant protein was eluted with a step gradient of imidazole (200-1000 mM) in chilled elution buffer (50 mM Tris-HCl, 300 mM NaCl, 10 % Glycerol; pH 8.0) containing 1 mM PMSF. Owing to inappreciable elution of rLinCas8b with imidazole (200-1000 mM), low pH phosphate-citrate buffer (pH 4.0-6.0), 8 M urea, or guanidine-HCl (6M) containing buffer was applied to elute the rLinCas8b from Ni-NTA-agarose resins. Moreover, appreciable elution of rLinCas8b could not be achieved with any of the tested elution conditions.

Next, to purify the rLinCas8b, cation exchange chromatography (CEX) was also performed using HiPrep^{TM} SP HP 16/10 column (GE Healthcare #29018183) operated by NGC^{TM} chromatography system (BioRad). For CEX, the lysate of induced *E. coli* cells was prepared in chilled binding buffer (50 mM Tris-HCl, 100 mM NaCl, 1 % Triton X-100, 10 % glycerol; pH 8.0) supplemented with 1 mM PMSF. Post-binding to pre-equilibrated CEX column, washing with 5-column volume (5 × 20 mL) binding buffer lacking the Triton X-100 was given. Elution was performed with a linear gradient of NaCl up to 1 M in a total of 80 mL of chilled elution buffer (50 mM Tris-HCl, 10 % glycerol; pH 8.0) added with 1 mM PMSF. The obtained elutes containing the rLinCas8b were subjected to the second purification step by size exclusion chromatography (SEC). Alternatively, LinCas8b was overexpressed with an N-terminal glutathione S-transferase (GST) tag (GST-LinCas8b, 98 kDa) to purify it by glutathione (GSH) affinity chromatography. The induced *E. coli* cells were lysed in chilled lysis buffer-1 (pH 8.0) containing 1 mM PMSF without imidazole, as described above, and clarified lysate was applied to GSH-agarose beads (1 mL) (BioBharati LifeScience #BB-GA004E). Post-binding resins were washed with 50 mL chilled wash buffer-1 (pH 8.0). Elution was performed using reduced glutathione (10-40 mM) in chilled wash buffer-1 (pH 8.0) with incubation of 4-24 h at 4 °C. Alternatively, to elute the bound GST-LinCas8, in-bead TEV protease (6×His-TEV) (NEB #P8112S) cleavage in 1× TEV protease cleavage buffer was performed at 4 °C for over-night.

For rLinCas11b (25.7 kDa) purification, the *E. coli* BL21 (DE3) cells transformed with pCDF1b-LinCas11b were induced with 1 mM IPTG at culture OD_{600} = 0.6 in LB broth at 37 °C for 4 h. The rLinCas11b was purified by Ni-NTA affinity chromatography under the native condition following the similar procedure as described above for rLinCas8b (6×His tagged) purification. rLinCas11b was eluted with a linear gradient of imidazole up to 400 mM in a total of 50 mL of chilled wash buffer-1 containing 1 mM PMSF. The obtained elutes were subjected to the second purification step by size exclusion chromatography (SEC). The SEC fractions containing rLinCas11b were concentrated using a centrifugal concentrator (Amicon #UFC901024) and stored at -20 °C until further experimental use.

3.5.4. Size exclusion chromatography

The size exclusion chromatography (SEC) was carried out using a HiLoad™ 16/600 Superdex™ 200 pg column (GE Healthcare #28-9893-35) operated through the NGC™ system (BioRad) vide section 2.5.3. To check the interaction between pure rLinCas11b and rLinCas7, 500 µg of each protein alone or together in a total volume of 1 mL (in storage buffer) was incubated at 37 °C for 1 h, followed by SEC.

3.5.5. Generation of polyclonal antibody against pure rLinCas8b protein

Polyclonal antibody against pure rLinCas8b (75 kDa) was generated in mice as described elsewhere (Dhara et al., 2019; Ghosh et al., 2018). For primary immunization, approximately 15 µg of gel-eluted rLinCas8b protein in PBS was emulsified with Freund's complete adjuvant (FCA; Santa Cruz Biotechnology #SC-3727) and was subcutaneously injected into each mouse of a group of the 4-6 weeks-old BALB/c mice (n = 4-5). A negative control group was injected with an equal volume of PBS along with the FCA. Thereafter, immunized mice were given two subsequent booster injections of rLinCas8b (15 µg/ mouse) with Freund's incomplete adjuvant (FIA; Santa Cruz Biotechnology #SC-3726) at 14 and 21 days of primary immunization. After seven days of the second booster, blood was collected from each mouse by retro-orbital bleeding, and then the mouse was sacrificed using the atlantooccipital dislocation method as described before (Ghosh et al., 2018). Sera obtained were pooled and stored at -80 °C until further use. The polyclonal antibody (anti-rLinCas8b) titer was analyzed by enzyme-linked immunosorbent assay (ELISA) **(Appendix Fig. B7)**. The mice immunization experiments were performed at the Department of Veterinary Microbiology, College of Veterinary Science,

Assam Agriculture University, Guwahati, India, after approval by the institutional animal ethics committee (approval no. 770/ac/CPCSEA/FVSC/AAU/IAEC/13-14).

3.5.6. Enzyme-linked immunosorbent assay

To determine the titer of the anti-rLinCas8b polyclonal antibody in the mice serum, ELISA was performed as described elsewhere (Ghosh et al., 2018). A disposable 96-well polystyrene plate was coated with 50 µL of pure rLinCas8b (200 ng/well) and incubated overnight at 4 °C. Thereafter, protein-coated wells were blocked for 2 h at 37 °C with 100 µL of 3 % bovine serum albumin (BSA). Post-blocking plates were washed three times with 100 µL/well of PBS containing 0.05 % Tween 20 (PBS-T), followed by incubation with 50 µL of mice serum at various dilutions (1:100 to 1:10000) at 37 °C for 2 h. As a negative control, pooled pre-immune serum of mice was used. After three times washing with PBS-T, wells were probed with 50 µL of horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG (1:5000; Sigma #A4416) for 1 hour at 37 °C. Plates were washed as described above, and then the binding was detected by adding Tetramethyl Benzidine (TMB) peroxidase substrate (Thermo Scientific #34022) for 10 min at 37 °C. The reaction was terminated by adding 50 µL/well of 1 M H_2SO_4, and final absorbance at 450 nm wavelength (OD_{450}) was recorded using an ELISA plate reader (Infinite 200 Pro, Tecan). The endpoint titer was determined visually, being the highest serum dilution giving a positive color development. All serum/antibody dilutions were prepared in 0.05 % PBS-T.

3.5.7. Immunoblotting

The whole cell lysates of *E. coli* and the pure rLinCas8b were resolved on 10 % SDS-polyacrylamide and electroblotting onto the nitrocellulose membrane (Santa Cruz Biotechnology #SC-3718). Membranes were blocked with 5 % non-fat dried milk prepared in Tris-buffered saline (TBS, pH 8.0) at 4 °C overnight, followed by probing with mice anti-rLinCas8b antibodies (1:500 dilution) for 2 h at room temperature. After five washes with 0.1% TBS-T (TBS containing 0.1 % Tween 20), the membranes were incubated with secondary antibody HRP-conjugated goat anti-mouse IgG (1:5000; Sigma #A4416) for 1 h at room temperature. After five washes with 0.1 % TBS-T, immunoblots were developed by adding a chemiluminescence substrate (Thermo Scientific #32209). All antibody dilutions were prepared in 2 % non-fat dried milk dissolved in 0.1 % TBS-T.

3.5.8. Bioinformatics analysis

The DNA sequences of interested Cas8 orthologs were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov). The potential internal in-frame translation start codons within the genes of interest (*cas8* orthologs) were predicted by the ribosome binding site (RBS) calculator program (https://salislab.net/software/). The amino acid sequences of identified Cas11 orthologs encoded within large subunit proteins were retrieved from respective full-length Cas8 or Cas10d proteins available from UniProt (https://www.uniprot.org/). The sequences of Cas11 orthologs (encoded independently) were retrieved from UniProt. Pairwise sequence comparison of Cas8 or Cas11 orthologs was made using Protein BLAST (BLASTP) web server (https://blast.ncbi.nlm.nih.gov/) with default parameters. For structural analysis, the structures of interest were retrieved from either the AlphaFold protein structure database (https://alphafold.ebi.ac.uk/) or the Protein data bank (PDB; https://www.rcsb.org) based on their availability. The structural comparison of proteins by pairwise alignment was performed using the Dali web server (http://ekhidna2.biocenter.helsinki.fi/dali/). For structure visualization and representation, the program PyMOL v2.4.1 (PyMOL molecular graphic system, Schrodinger, LLC, New York, USA) was used.

The Cas8 and Cas11 orthologs used in this study for sequence/structural analysis include LinCas8b (associated with subtype I-B system of *Leptospira interrogans*; UnipProt ID: Q72TT0; AlphaFold ID: AF-Q72TT0), SynCas8b (I-B; *Synechocystis* spp.*;* A0A068N831; AF-A0A068N831), CthCas8b (I-B; *Clostridium thermocellum*; A3DKB7; AF-A3DKB7), MmaCas8b (I-B, *Methanococcus maripaludis*; A4FXZ7, AF-A4FXZ7), DvuCas8c (I-C; *Desulfovibrio vulgaris*; Q72WF8), SynCas10d (a Cas8 ortholog in type I-D of *Synechocystis* spp.; Q6ZEI7). DvuCas11c (encoded within DvuCas8c; PDB ID: 7KHA chain K), SynCas11d (encoded within SynCas10d; 7SBA chain K), EcoCas11e (I-E; *E. coli*; P76632; 4U7U chain B), SsoCsa5 (a Cas11 ortholog in subtype I-A of *Saccharolobus solfataricus*; Q97YC8; 3ZC4 chain A); SepCsm2 (Cas11 in type III-A of *Staphylococcus epidermidis*; Q5HK90; 6AE1 chain A) and AfuCmr5 (Cas11 in type III-B of *Archaeoglobus fulgidus*; O28417, 3X1L chain G).

3.5.9. Circular dichroism (CD) spectroscopy

The purified rLinCas11b (10 µM) in a dilution buffer (20 mM Tris-HCl and 20 mM NaCl; pH 8.0) was subjected to CD spectroscopy as described in section 2.5.6. The CD spectra were

assessed to compute secondary structure configurations using the K2D3 webserver (Louis-Jeune et al., 2012).

3.5.10. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were performed on a MicroCal iTC200 system (GE Healthcare) at 30 °C with constant stirring at 300 rpm. Pure protein samples (rLinCas7 & rLinCas11b) were prepared in storage buffer (25 mM Tris-HCl, 100 mM NaCl, 5 % Glycerol; pH 8.0). The sample cell and the syringe were washed twice with storage buffer before loading the samples. The pure rLinCas7 (280 µL of 8 µM) and the rLinCas11b (40 µL of 120 µM) were loaded into the ITC sample cell and the syringe, respectively. rLinCas11b (1.5 µL) was titrated into the sample cell at an interval of 120 sec for a total of 25 injections over a period of 50 min. Power was recorded at a "high" gain setting with a reference power of 6 µcal sec^{-1} and a 5 sec filter period. A control experiment to measure the heat of dilution was also performed by titrating the rLinCas11b into the storage buffer. The obtained heat of dilution in the control experiment was subtracted from the protein-protein titration to obtain the real binding affinity. The parameters of the ITC experiment were determined by reaction-heat integration using nonlinear regression and a one-site binding model. Data analysis, including baseline correction and evaluation, was carried out using OriginPro 8.5 ITC.

3.5.11. DNA binding Assay

The rLinCas11b at the specified concentration(s) was incubated with various DNA substrates (3.5 nM) in a binding buffer (20 mM Tris-HCl and 100 mM KCl; pH 8.0) in the presence and absence of Mg^{2+} ions ($MgCl_2$; 10 mM) at 37 °C for 1h. The types of DNA substrates used were linear ds-DNA (PCR amplicon, nonmethylated, 2.8 kb), circular ds-DNA (pT7Blue, methylated, 2.8 kb; Sigma-Aldrich #69967), and circular ss-DNA (ϕx174 virion DNA, 5.3 kb; NEB #3023S). EDTA (10 mM) served as a divalent metal ion chelator. The reactions were resolved on 1 % agarose gel stained with EtBr and visualized using a gel documentation instrument.

*****

Chapter 4

Insights to the assembly of functional CRISPR-Cas type I-B Cascade complex of *Leptospira interrogans* and characterization of interference-efficient PAM sequence [2]

4.1. Summary

Pathogenic species of *Leptospira* are recalcitrant for genetic manipulation using conventional tools, and therefore there is a need to explore techniques of higher efficiency. Application of endogenous CRISPR-Cas tool is emerging and efficient; nevertheless, it is limited by a poor understanding of interference machinery in the bacterial genome and its associated protospacer adjacent motif (PAM). In this study, the functionality of the interference machinery of CRISPR-Cas subtype I-B (Lin_I-B) from *L. interrogans* was assessed in surrogate host *E. coli*. The overexpression of the Lin_I-B interference machinery in *E. coli* demonstrated that LinCas5, LinCas6, LinCas7, LinCas8b, and LinCas11b can self-assemble on cognate crRNA to form an effector complex (LinCascade), conferring the operativity of the system for maturation stage of CRISPR immunity. Further, LinCascade demonstrated a functional phenotype resulting in a robust interference of target plasmids containing a protospacer with a predicted 5′-PAM (TGA, ATG, and ATA). Moreover, a mutant variant LinCascade$^{-Cas11b}$ that lacks LinCas11b co-expression erred to mount target plasmid interference. At the same time, LinCas11b complementation in LinCascade$^{-Cas11b}$ rescued target plasmid interference suggesting the prerequisite of LinCas11b for a functional LinCascade. Thus, the present study establishes *Leptospir*a's subtype I-B interference machinery to be functional and, soon, may pave the way for scientists to harness it as a programmable endogenous genetic manipulation tool.

[2] This chapter is partly adapted from a publication and reprinted with the authors' permission. **Hussain, M.S.**, Anand, V., and 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*, p.124086.

4.2. Results

4.2.1. Overexpression of *Leptospira* CRISPR-Cas type I-B interference machinery in heterologous host *E. coli*

The CRISPR-Cas subtype I-B (Lin_I-B) system in *L. interrogans* contains a CRISPR array (LIC_Cr^2) flanked by two independent *cas*-operons (operon-I and -II) **(Fig. 4.1A)** (Dixit et al., 2016). Operon-I encloses *cas* genes (*lincas4*, *lincas1*, and *lincas2*) essential for the adaptation phase and thus serves as an adaptation module. On the other hand, Operon-II encloses interference-phase specific *cas* genes (*lincas6*, *lincas3*, *lincas8b*, *lincas7*, and *lincas5*) and is referred to as the interference module. Therefore, to understand the interference mechanism of the Lin_I-B system, the complete interference Cas module (*lincas6*, *lincas3*, *lincas8b*, *lincas7*, and *lincas5*) and the associated CRISPR array (LIC_Cr^2) were co-expressed in BL21 (DE3) cells via plasmid constructs pCas and pCr^2, respectively. The induction of BL21 strain (BL-pCas+pCr^2) housing pCas and pCr^2 plasmid with IPTG resulted in the overexpression of interference-associated LinCas proteins (rLinCas6, 23 kDa; rLinCas3, 85 kDa; rLinCas8b, 75 kDa; rLinCas11b, 22 kDa rLinCas7, 31 kDa and rLinCas5, 23 kDa) at their expected size **(Fig. 4.1B)**. However, because of the similar size of overexpressed Cas proteins (rLinCas8b; 75 kDa & rLinCas3; 85 kDa) and (rLinCas6; 23 kDa & rLinCas5; 23 kDa), these could not be differentiated on SDS-PAGE (10 %). Therefore, the whole-cell lysates of the BL-pCas+pCr^2 strain (induced and uninduced) were immunoblotted using specific antibodies against each LinCas protein **(Fig. 4.1C)**. Immunoblotting confirmed the Cas protein overexpression (rLinCas6, rLinCas3, rLinCas8b, rLinCas11b, rLinCas7, and rLinCas5) in *E. coli* at their expected size.

4.2.2. *Leptospira* CRISPR-Cas type I-B interference machinery overexpressed in *E. coli* self-assembles into Cascade complex

We investigated if the interference-associated Cas proteins of the Lin_I-B system, overexpressed in *E. coli*, can self-assemble with the co-transcribed pre-crRNA into a Cascade complex (rLinCascade). Previously our research group has demonstrated that the pre-crRNA transcript of LIC_Cr^2 could be processed into mature crRNAs by rLinCas6 under *in vitro* conditions (Prakash and Kumar, 2021). Here, we explored the LIC_Cr^2 pre-crRNA transcript processing by LinCas6 within *E. coli*. The Ni-NTA affinity pull-down assay, employing rLinCas8b as bait, resulted in co-purification of all the Cas subunits (rLinCas8b, rLinCas7,

rLinCas6, and rLinCas5), including the co-translating rLinCas11b from induced BL-pCas+pCr^2 cells evident by SDS-PAGE analysis **(Fig. 4.2A)** and immunoblotting **(Fig. 4.2B)** of pull-down elutes.

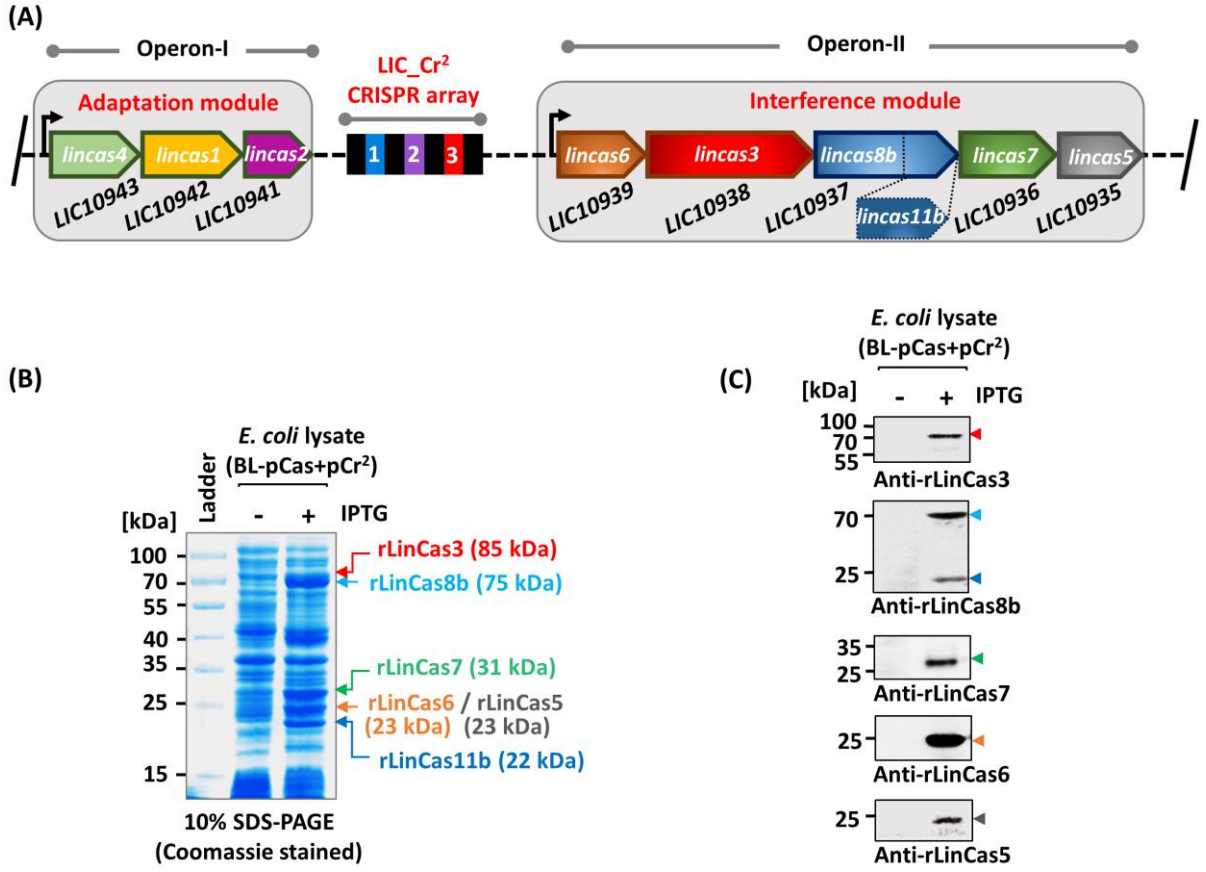

Fig. 4.1. Overexpression of the Lin_I-B interference machinery in *E. coli*. (A) CRISPR-Cas subtype I-B (Lin_I-B) loci in *L. interrogans* genome (adapted from Dixit et. al., 2016). The CRISPR array (LIC_Cr^2) comprised of four repeats (black rectangles) and three spacers (colored rectangles) is flanked by two *cas*-operons: I and II. Operon-I (*lincas4-lincas1-lincas2*) encodes adaptation module Cas proteins, while Operon-II (*lincas6-lincas3-lincas8b-lincas7-lincas5*) encodes the interference module Cas proteins. The length of each *cas* gene is drawn approximately to scale, whereas CRISPR array elements are drawn much larger for clarity. The corresponding gene locus tag is indicated below each *cas* gene. The *lincas11b* ORF encoded within *lincas8b* at 3′-end has been shown in zoom view. **(B)** The SDS-PAGE (Coomassie stained) analysis of whole-cell lysates of uninduced (-) and IPTG-induced (+) *E. coli* BL21 (DE3) cells (BL-pCas+pCr^2). **(C)** Immunoblotting of whole-cell lysates of *E. coli* BL-pCas+Cr^2. The uninduced (-) and IPTG-induced (+) lysates were resolved on SDS-PAGE (10 %), electroblotted on nitrocellulose membrane, and then probed with antibodies specific to rLinCas proteins. Detection of specific rLinCas proteins in induced (+) lysates is demonstrated and directed with an arrowhead.

On the other hand, in the *E. coli* BL-pCas cells lacking the array component, the nickel affinity pull-down resulted in the purification of only the rLinCas8b (bait) and rLinCas11b **(Fig. 4.2A**

& B). Thus, one may suggest that the pre-crRNA is essential for the assembly of rLinCascade in *E. coli* cells (BL-pCas+pCr^2). Moreover, denaturing urea-PAGE analysis of pull-down eluates containing the rLinCascade exhibited a predominant nucleic acid band similar to mature crRNA (70 nt; custom synthesized) of Lin_I-B locus **(Fig. 4.2C)**. The RNase A treatment to the elutes completely cleaved the nucleic acid content suggesting it to be crRNA **(Fig. 4.2C)**. Thus, rLinCas6 catalyzed efficient processing of pre-crRNA into crRNA within *E. coli* served as a framework for the assembly of rLinCascade.

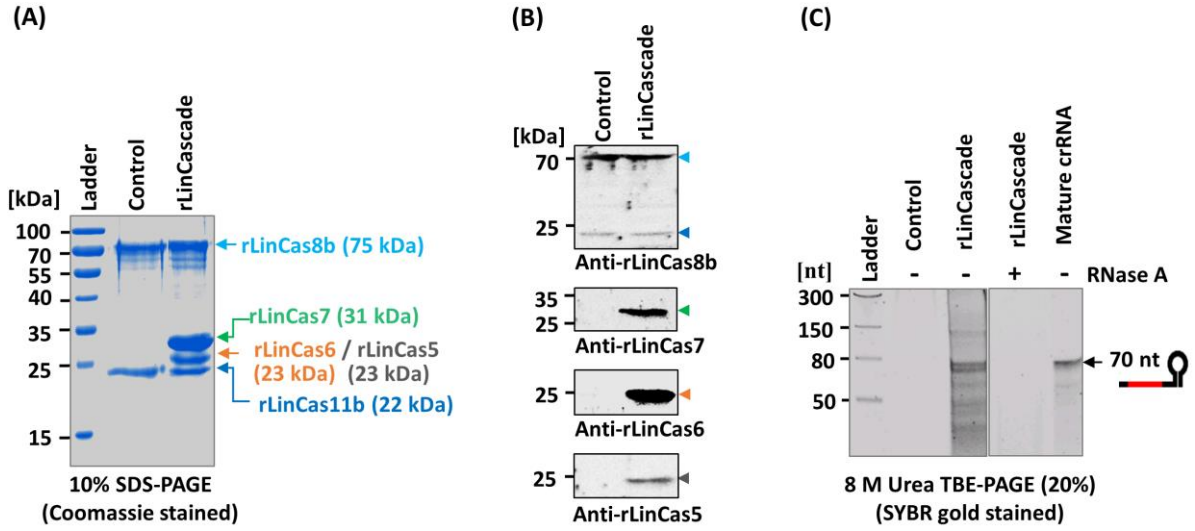

Fig. 4.2. Self-assembly of Lin_I-B effector complex (LinCascade) overexpressed in *E. coli*. (A) SDS-PAGE (10 %; Coomassie stained) analysis of nickel affinity pull-down elutes from *E. coli* BL21 (DE3) cells overexpressing Operon-II of Lin_I-B loci with (rLinCascade) or without (control) LIC_Cr^2. The 6×His tagged rLinCas8b served as bait in the pull-down assay. **(B)** Immunoblot of pull-down elutes. The elutes analyzed in panel-A were immunoblotted with rLinCas protein specific antibodies. Specific detection of rLinCas proteins is marked with an arrowhead. **(C)** Denaturing PAGE (20 %; SYBR gold stained) of pull-down elutes (control and rLinCascade). The elute containing the nucleoprotein complex was treated with RNase A (+) to know if the bound nucleic acids were DNA or RNA. The commercially synthesized mature crRNA (70 nt) of LIC_Cr^2 was run as a marker. The distant wells of the same gel have been cropped and re-aligned for clarity. The left-most lane in denaturing PAGE contains an RNA ladder, and the image is presented after inversion for clarity.

4.2.3. CRISPR-Cas type I-B interference machinery of *Leptospira* confers PAM-dependent target plasmid interference in *E. coli*

In order to validate the functionality of the Lin_I-B system, a target plasmid interference assay was arranged in *E. coli* against the plasmids incorporated with protospacers containing specific 5′-PAM. The 5′-PAM sequence(s) was identified from predicted protospacers matching the three spacers (S1, S2, and S3) of the LIC_Cr^2 array. For all three spacers of the LIC_Cr^2, the

CRISPRTarget program (Biswas et al., 2013) presented a total of seven protospacer hits (Table 4.1) from various plasmid and phage genomes available in the database. Analysis of 5′-eight nucleotides overhangs of the predicted protospacers revealed 5′-TGA and 5′-ATA as the consensus 5′-PAM sequences **(Fig. 4.3A)**. The predicted PAM sequences in this study differed from earlier predicted PAM sequences (5′-ATG and 5′-TAC) for the CRISPR-Cas I-B locus of the genus *Leptospira* (Mendoza and Trinh, 2018; Prakash and Kumar, 2022; Xiao et al., 2019a). Therefore, we included the previously predicted PAM (5′-ATG) along with 5′-TGA and 5′-ATA for our experimental validation via target plasmid interference assay.

Table 4.1. Putative protospacer identified through *in-silico* analysis of the LIC_Cr^2 spacers

LIC_Cr^2 spacers	*Trinucleotide protospacer match (5′-3′)	Invader plasmid/phage strain	#Coordinates
Spacer-1	TCTCA**ATA**AAGTTTTTCACGGG GTGACGAATGTTTTCCCCCT	Ga0123519_10003459	22167-22136
	TTTGA**TGA**AGGGGGAAAACAT TCGTCACCCCGTGAAAAACTT	IMGVR_UViG_2531839053_000001	22136-22167
	TTTAA**TGA**AGGGCGAGAACATT CGTCATCCTGTGAAAAACTT	IMGVR_UViG_2541047030_000003	12676-50382
	TTTAA**TGA**AGGGCGAGAACATT CGTCATCCTGTGAAAAACTT	IMGVR_UViG_2541047030_000002	19810-19779
	TTTAA**TGA**AGGGCGAGAACATT CGTCATCCTGTGAAAAACTT	IMGVR_UViG_2541047029_000002	19822-19791
	TTTAA**TGA**AGGGCGAGAACATT CGTCATCCTGTGAAAAACTT	IMGVR_UViG_2534681715_000001	16951-16920
Spacer-2	No hits	-	-
Spacer-3	TTTAA**TGA**AGGGCGAGAACATT CGTCATCCTGTGAAAAACTT	IMGVR_UViG_3300025691_000035	12213-12239

*The putative PAM sequences are indicated in bold and underlined.
#The coordinates of the protospacer sequence in the invader plasmid and phage genome are provided.

Various 5′-PAM (TGA, ATA, ATG) containing oligos with identical spacer (spacer-1 of LIC_Cr^2) were synthesized by outsourcing (Table 4.2) and cloned into pT7Blue plasmid. It resulted in the generation of three different target plasmids [p(TGA)Spacer, p(ATA)Spacer, and p(ATG)Spacer]. Similarly, a non-PAM sequence (5′-CAC) containing protospacer was cloned into pT7Blue (non-PAM-target plasmid). At the same time, the blank wild-type pT7Blue was used as a reference plasmid (pBlank) in the target plasmid interference assay. Afterward, the individual target plasmid, along with the pCas and pCr^2, was co-transformed into *E. coli* BL21-AI cells. As a control, in place of the target plasmid, the non-PAM-target

plasmid or reference plasmid was used. The generated BL21-AI strains housing all three plasmids were induced in LB broth and then plated on LB-agar plates supplemented with three antibiotics (Cam+Kan+Amp). For clarity, the procedure has been schematically represented in **Fig. 4.3B**. In the designed plasmid interference assay, the target plasmid having a functional 5′-PAM confers ampicillin resistance to its carrier bacteria. Therefore, the target plasmid degradation by the Lin_I-B interference machinery would result in a loss of antibiotic resistance rendering the *E. coli* unable to grow on an ampicillin-containing medium. Thus, the depletion in bacterial growth [colony-forming unit (CFU) count] on the test plate would be due to the degradation of target plasmids (Amp^R) by Lin_I-B interference machinery representing a CRISPR interference phenotype. The magnitude of DNA interference was measured as the CFU depletion ratio (the ratio of CFU on the control and test plate).

Table 4.2. Protospacers composed from spacer-1 of LIC_Cr^2 and the putative 5′-PAM

Protospacer		***Sequence (5′-3′)**
Duplex name	**DNA strand**	
(TGA)Spacer	Positive	*AGCTT***TGA**AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTT*CA*
	Negative	*TATG*AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT**TCA***A*
(ATA)Spacer	Positive	*AGCTT***ATA**AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTT*CA*
	Negative	*TATG*AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT**TAT***A*
(ATG)Spacer	Positive	*AGCTT***ATG**AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTT*CA*
	Negative	*TATG*AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT**CAT***A*
#(CAC)Spacer (Non-PAM-protospacer; control)	Positive	*AGCTT***CAC**AAAGGATCCTTTGATCAAAAGAATTCGTCCTTGATTT*CA*
	Negative	*TATG*AAATCAAGGACGAATTCTTTTGATCAAAGGATCCTTT**GTG***A*

*Cohesive restriction overhangs (*Hind*III at 5′- & *Nde*I at 3′-end) are indicated in italics; PAM sequences are shown in bold and underlined.
#Non-PAM-protospacer consists of a spacer-1 and a consensus trinucleotide (5′-CAC) derived from 3′-end of repeats in array LIC_Cr^2.

The survival of the induced strains containing the target plasmids [p(TGA)Spacer, p(ATA)Spacer or p(ATG)Spacer] was drastically reduced (up to ~149-fold) compared to the controls **(Fig. 4.3C)**. The uninduced culture of each BL21-AI strain used in the target plasmid interference assay demarcated the existing basal level of depletion ratio.

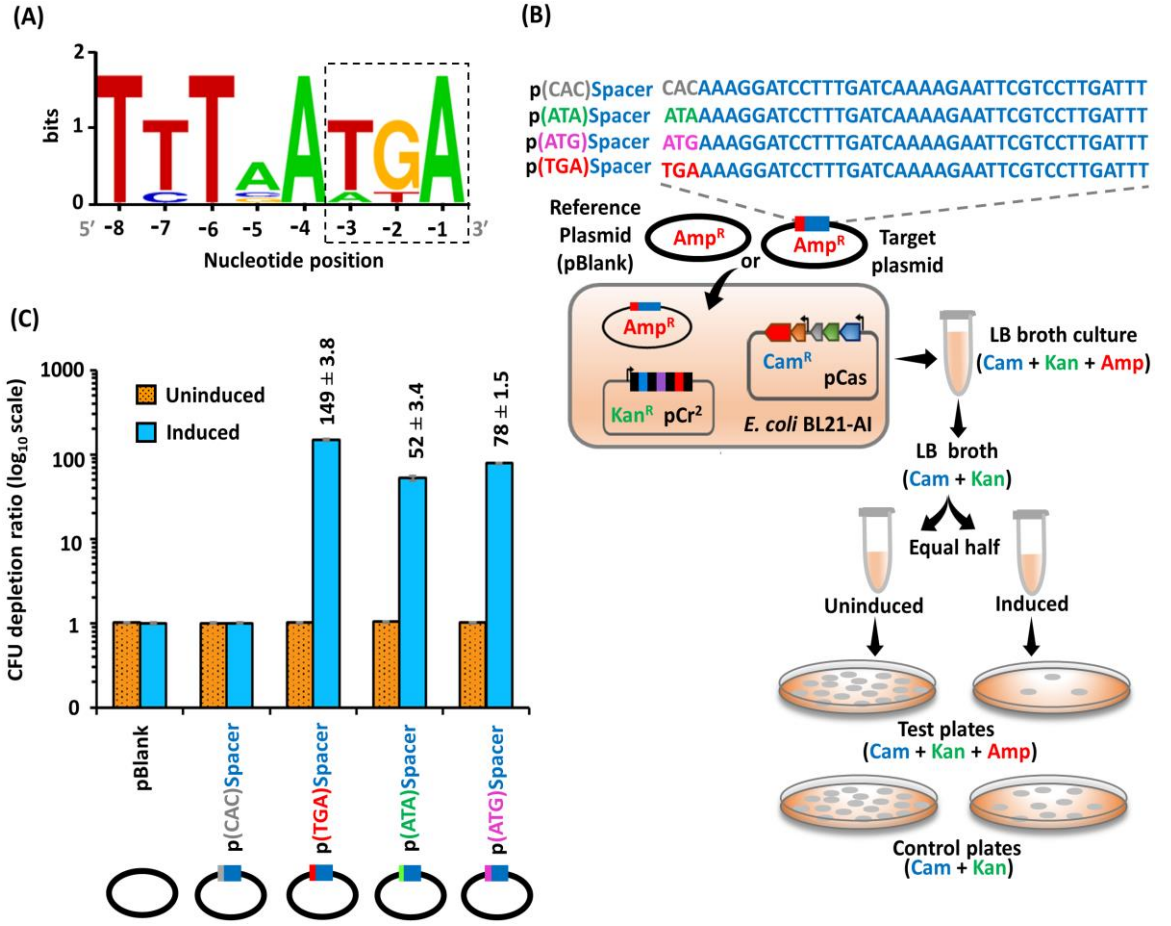

Fig. 4.3. Targeted plasmid interference by Lin_I-B interference machinery in *E. coli*. (A) The sequence logo of eight nucleotide 5′-flanking sequence of predicted protospacers matching the spacers of Lin_I-B locus (LIC_Cr^2) of *L. interrogans* serovar Copenhageni. The residue position is indicated on the X-axis. The height of residue (in bits) represents the extent of conservation at each position. The consensus trinucleotide PAM (5′-TGA and 5′-ATA) at the 5′-end of protospacers is shown (discontinued black square). **(B)** Schematic of plasmid interference assay in *E. coli*. The *E. coli* BL21-AI cells were co-transformed with pCas (Cam^R), pCr^2 (Kan^R), and one of the given target plasmids [p(TGA)Spacer, p(ATA)Spacer, or p(ATG)Spacer; Amp^R]. For experimental control against target plasmid, pBlank or p(CAC)Spacer was used as reference or non-PAM target plasmid, respectively. The obtained *E. coli* BL21-AI strains housing three plasmids were cultured to mid-log phase in LB broth containing three antibiotics (Cam+Kan+Amp). Cultures were pelleted and resuspended in fresh LB broth containing two antibiotics (Cam+Kan). Equal half of the cultures were grown as uninduced and induced (IPTG + L-arabinose) for 4 h at 37 °C followed by spread plating (24 h at 37 °C) on LB agar plates containing three (test plates) and two antibiotics (control plates). The reduced CFU count on the test plate for the induced culture represents a CRISPR interference phenotype. **(legend continued on next page)**

(C) PAM-elicited target plasmid interference by Lin_I-B interference machinery in *E. coli*. Expression of Lin_I-B interference machinery resulted in efficient interference of cohabiting target plasmids containing protospacer with either of the three predicted 5′-PAM sequence [p(TGA)Spacer, p(ATA)Spacer, and p(TGA)Spacer]. Interference was not evident for plasmids [pBlank and p(CAC)Spacer]. The magnitude of plasmid DNA interference was measured in terms of CFU depletion ratio: the ratio of CFU count on the control plate and test plate. Data are represented as mean ± standard errors of the mean (SEM) from four-independent experiments (n = 4).

Also, as expected, bacteria housing non-PAM-target plasmid or reference plasmid exhibited a basal CFU depletion ratio. Among the three 5′-PAM sequences (TGA, ATA, ATG), maximum plasmid interference was recorded with 5′-TGA leading to ~149-fold depletion in *E. coli* growth (CFU) followed by 5′-ATG (~78-fold) and 5′-ATA (~52-fold). The assay thus signifies the necessity of a functional PAM for the Lin_I-B system to confer CRISPR immunity effectively. The PAM-dependent target plasmid DNA degradation also demonstrates that the Lin_I-B system is functionally active for the interference phase of CRISPR immunity in surrogate host *E. coli*.

4.2.4. LinCas11b is indispensable for the LinCascade-mediated target DNA interference

We have identified a small in-frame ORF within the *lincas8b* gene that co-translates into LinCas11b protein. To better understand if the co-translating LinCas11b is critical for the function of LinCascade, plasmid DNA interference was assessed in *E. coli* with omission and complementation of LinCas11b. Wherein, by site-directed mutagenesis in pCas, we inactivated the co-expression of the LinCas11b subunit of the rLinCascade (rLinCascade$^{-Cas11b}$) **(Appendix Fig. C2.)** and plasmid interference assay was repeated with one of the validated PAM (5′-TGA). The interference of target plasmid [p(TGA)Spacer] was assessed in BL21-AI strain and was plotted in terms of depletion ratio **(Fig. 4.4)**. Overexpression of the rLinCascade lacking LinCas11b (rLinCascade$^{-Cas11b}$) erred to cause the target plasmid interference effectively. A drastically reduced CFU depletion ratio (~2-fold) was observed compared to ~141-fold recorded with wild-type rLinCascade (WT). With the rLinCascade$^{-Cas11b}$ variant of rLinCascade, the depletion ratio of *E. coli* containing target plasmid [p(TGA)Spacer] was close to the strains housing reference plasmid (pBlank). Moreover, the interference ability of the rLinCascade$^{-Cas11b}$ was rescued to the WT level when rLinCas11b was trans-complemented (rLinCascade$^{-Cas11b\ com}$) in the *E. coli* strain. The assays thus suggest LinCas11b is essential for LinCascade to mediate the target plasmid interference.

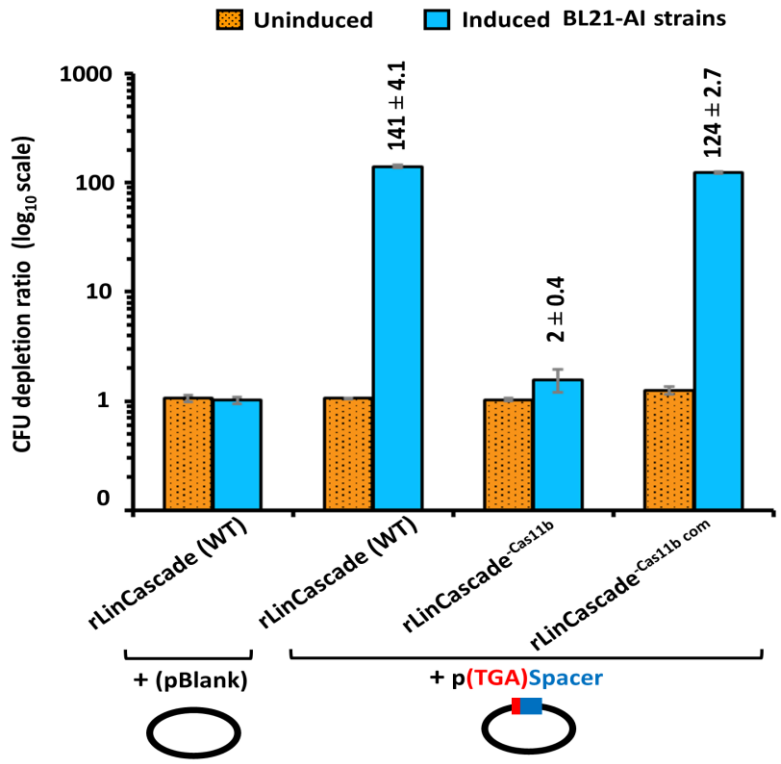

Fig. 4.4. LinCas11b is a prerequisite for LinCascade-mediated target plasmid interference. The target plasmid [p(TGA)spacer] interference was assessed in *E. coli* BL21-AI strain (rLinCascade$^{-Cas11b}$) expressing all Cas proteins except LinCas11b. Another *E. coli* BL21-AI strain (rLinCascade$^{-Cas11b\ com}$), trans-complemented with rLinCas11b, was used to check the specificity of plasmid interference and was compared with the strain expressing wild-type rLinCascade (WT). The magnitude of plasmid interference was measured in the CFU depletion ratio (vide methods). The mutant rLinCascade (rLinCascade$^{-Cas11b}$) failed to cause the interference of the target plasmid within *E. coli*. On the other hand, the trans-complementation of LinCas11b to rLinCascade$^{-Cas11b}$ (rLinCascade$^{-Cas11b\ com}$) within *E. coli* rescued the plasmid interference. Data are represented as mean ± standard errors of the mean (SEM) from three independent experiments (n = 3).

4.3. Discussion

CRISPR-Cas systems are highly diverse (Makarova et al., 2020b, 2015), and the molecular details of type-I CRISPR-Cas operation are referred to date primarily from subtypes I-E (Hayes et al., 2016; Kuznedelov et al., 2016; Yosef et al., 2012) and I-C systems (Hochstrasser et al., 2016; O'Brien et al., 2020). The operational details of CRISPR-Cas I-B, representing the most abundant CRISPR-Cas system in prokaryotes (Makarova et al., 2015), are sparse, including *Leptospira*. Our previous investigations on pure recombinant LinCas proteins (LinCas1 (Dixit et al., 2021b), LinCas2 (Dixit et al., 2016), LinCas4 (Dixit et al., 2021a), LinCas6 (Prakash and Kumar, 2021), and LinCas7 (Hussain and Kumar, 2022)) of the CRISPR-Cas I-B (Lin_I-B) of

L. interrogans lead us to anticipate an operational defense system within *Leptospira* genome. Also, the cognate pre-crRNA transcript was processed into mature crRNA by the rLinCas6 (Prakash and Kumar, 2021) and served as a platform for the rLinCas7 to assemble (Hussain and Kumar, 2022). In a follow-up to these studies, we addressed two queries related to the interference phase of CRISPR-Cas immunity in *Leptospira* spp. We explored whether LinCas proteins (Cas3, Cas5, Cas6, Cas7, Cas8b, and Cas11b) associated with the interference phase can self-assemble to form a Cascade and recognize the PAM-labeled protospacers in target plasmid within an *E. coli* host. The functional status of the self-assembled Cascade and its PAM recognition was assessed by target plasmid interference assay in an *E. coli* host.

In the nickel-affinity pull-down assay of rLinCas8b, other interference-associated LinCas proteins (rLinCas7, rLinCas6, and rLinCas5) demonstrated co-purification exclusively when there were cognate pre-crRNA transcripts in *E. coli* (BL-pCas+pCr^2). As identified previously **(Fig. 3.1B-C)**, the direct interaction of rLinCas11b with rLinCas8b led to co-purification of rLinCas11b under both conditions, in the presence and absence (control) of pre-crRNA transcripts. The size of crRNA co-purified with LinCas proteins agrees with that of processed mature crRNA variants (70, 73, and 75 nt) from pre-crRNA by LinCas6, described before (Prakash and Kumar, 2021). Thus, the co-purification of LinCas proteins (rLinCas8b, rLinCas11b, rLinCas7, rLinCas6, and rLinCas5) with mature crRNA suggests crRNA-nucleated self-assembly of LinCascade. Further, the co-purified protein complex demonstrated rLinCas7 overrepresentation. Such overrepresentation is likely due to Cascade backbone genesis with multiple rLinCas7 subunits on a crRNA- a characteristic feature of the Cascade complex (Liu and Doudna, 2020). The formation of the rLinCas7 subunits backbone agrees with our previous study, where rLinCas7 oligomerizes onto the cognate mature crRNA *in vitro* (Hussain and Kumar, 2022). These findings infer that the Lin_I-B interference machinery is active for the expression phase of CRISPR immunity, leading to the formation of a stable rLinCascade when overexpressed in *E. coli.*

Motivated by the perceptible assembly of rLinCascade in *E. coli*, the proficiency of the complex to mount CRISPR defense against targeted plasmid was assessed. A type I CRISPR-Cas system mounts the defense reaction when it recognizes a cognate PAM next to a spacer-matching protospacer sequence in a foreign DNA/plasmid (Gleditzsch et al., 2019; Liu and

Doudna, 2020). Conventionally, the trinucleotide sequence located at the immediate 5′-end of the protospacer serves as a functional PAM for type I CRISPR-Cas systems, including subtype I-B (Fischer et al., 2012; Gleditzsch et al., 2019; Leenay and Beisel, 2017). Similarly, in this study, the putative PAM sequence (5′-TGA and 5′-ATA) was identified for the Lin_I-B system. In addition, there have been reports of other computationally predicted PAM sequences (5′-ATG and 5′-TAC) for type I-B locus in *Leptospira* spp. (Mendoza and Trinh, 2018; Prakash and Kumar, 2022; Xiao et al., 2019b). Such discrepancy in the predicted PAM sequence may be due to the *in-silico* analysis of a varying number of spacers. The predicted 5′-PAMs (TGA, ATA, and ATG) from this or elsewhere studies (Mendoza and Trinh, 2018; Prakash and Kumar, 2022) were analyzed to probe their efficiency to trigger the Lin_I-B's DNA interference response by plasmid interference assay. The bacterial growth depletion recorded in the plasmid interference assay demonstrates the degradation of the target plasmids by Lin_I-B interference machinery, affirming that the Lin_I-B system is functionally active for the interference phase of CRISPR immunity. Maximum plasmid interference (~149-fold) was recorded against the target plasmid containing 5′-TGA PAM followed by 5′-ATG (~78-fold) and 5′-ATA (~52-fold).

The defensive CRISPR response against all three 5′-PAMs (TGA, ATA, and ATG) indicates the promiscuity in PAM recognition by the rLinCascade. Such promiscuity in PAM recognition by Cascade complex is inherent in all type I CRISPR-Cas systems across the various bacteria (Gleditzsch et al., 2019; Shah et al., 2013). For instance, in *C. thermocellum*, all the trinucleotides with a consensus 5′-TNA and 5′-TTN sequence (where N= A/T/G/C), including 5′-TGA, support the interference by the type I-B system (Walker et al., 2020). Similarly, the type I-B Cascade in *Haloferax volcani* recognizes six different 5′-PAMs, including CAC, TAG, TAA, ACT, TTC, and TAT (Fischer et al., 2012). The type I-B of *Synchocyctis* spp. recognizes 5′-ATG as PAM for interference (Tan et al., 2022). In *C. pasteurianum*, three 5′-PAMs (TCT, TTG, and TCA) have been reported functional for interference by type I-B system (Pyne et al., 2016).

Thus, the determined 5′-PAMs (TGA and ATG) for Lin_I-B interference find a match with the previously identified PAMs for subtype I-B in *C. thermocellum* (5′-TGA) and *Synchocyctis* spp. (5′-ATG). Also, in this study, the control protospacer with a non-PAM sequence (5′-CAC, derived from the repeats of Lin_I-B locus) failed to trigger DNA interference. This suggests

the spirochete may use strategies to avert self-attack on its native CRISPR locus. However, the same PAM (5′-CAC) was previously demonstrated functional for type I-B CRISPR interference in *H. volcani*. Overall, the identification of multiple functional 5′-PAMs (TGA, ATG, and ATA) for the Lin_I-B system agrees with the previously witnessed promiscuity in PAM recognition by the type I-B systems in various bacteria. Also, the observed variable response of an effector complex towards a PAM in different bacteria sustains the necessity of determining the PAM sequence experimentally in each bacterium instead of directly adopting it from other bacteria. Further, the promiscuity in PAM recognition is often noticed because of poor or no significant sequence similarity between the large subunit proteins (Cas8 orthologs)-the subunits in different effector complex that recognizes the cognate PAM (Cass et al., 2015b; Gleditzsch et al., 2019). Nevertheless, the inherent flexibility of PAM recognition during interference is in line with the fitness need of bacteria to cope with the mutational evasion of viral intruders (Gleditzsch et al., 2019). Moreover, in this study, the experimentally validated 5′-PAMs (TGA, ATA, and ATG) portray only a small subset of interference-efficient PAMs. Therefore, a systematic review of all possible trinucleotides, as illustrated for other subtypes of CRISPR-Cas I systems (Fischer et al., 2012; Musharova et al., 2019), may deduce a plausible number of interference-efficient PAMs for the Lin_I-B system.

Further, in the target plasmid interference assay, the rLinCascade$^{-Cas11b}$ that lacks the LinCas11b subunits failed to bring about the interference of the target plasmid. On the other hand, the complementation of rLinCas11b to rLinCascade$^{-Cas11b}$ (rLinCascade$^{-Cas11b\ com}$) rescued the interference response. This endorses the essentiality of LinCas11b for LinCascade-mediated target plasmid interference. Our result agrees with recent studies wherein the fused Cas11 proteins (encoded within the *cas8* or *cas10d* genes) were reported essential for target plasmid interference by subtype I-C, I-B, and I-D systems of various bacteria, including *D. vulgaris*, *B. halodurans*, and *Synechocystis* spp.(McBride et al., 2020; Tan et al., 2022).

4.4. Conclusion

The present study of CRISPR-Cas mediated interference in the heterologous host leads us to conclude that the CRISPR-Cas subtype I-B (Lin_I-B) of *L. interrogans* may be functionally active. We determined the interference-efficient PAMs for the Lin_I-B system. However, a systematic analysis of trinucleotide combinations is warranted to deduce all possible interference-efficient PAM motifs for the Lin_I-B system. Importantly, LinCas11b encoded

within the *lincas8b* ORF is indispensable for the defense response of the Lin_I-B system, as in its absence, Lin_I-B interference machinery could not accomplish efficient target DNA interference. Further insight into the molecular mechanism of interference by the Lin_I-B system awaits experimental structure determination of a type I-B effector complex, including LinCascade. Taken together, it is presumed that the characterization of the Lin_I-B system would pave the way for scientists to harness it as a targeted endogenous genome editing tool and augment the toolbox for precise genetic manipulation of *Leptospira* spp.

4.5. Materials and methods

4.5.1. Bacterial strains, media, and growth conditions

L. interrogans serovar Copenhageni strain Fiocruz L1-130 was maintained in EMJH medium (Difco, Becton Dickinson India Pvt. Ltd., India) as described in section 2.5.1. *E. coli* strain DH5α (NEB #C2987H) was used for cloning and plasmid maintenance. *E. coli* strains BL21 (DE3) (Invitrogen #C6000-03) and BL21-AI (Invitrogen #C6070-03) were used for heterologous overexpression of the proteins. LB-broth (Himedia #M1245) or LB-agar media (Himedia #M1151) supplemented with appropriate antibiotics at 37 °C were used for growing *E. coli* strains.

4.5.2. Construction of plasmids encoding type I-B interference machinery of *Leptospira*

The genomic sequence of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130, available at the NCBI, was used for primer designing and cloning. The DNA sequences of interest were PCR amplified individually from the genomic DNA of *L. interrogans* using sequence-specific primers and cloned into different plasmids, as detailed below. The primers and plasmids used in this study are listed in (Appendix Table C1). The DNA sequence encompassing three consecutive *cas* genes (*lincas8b*/*LIC10937*, 1887 bp; *lincas7*/*LIC10936*, 840 bp; *lincas5*/*LIC10935*, 624 bp) was ligated into the pACYCDuet-1 plasmid at multiple cloning site-I (MCS-I) between *Sal*I and *Not*I. At the MCS-II site of the same plasmid, the DNA sequence enclosing two *cas* genes (*lincas6*/*LIC10939*; 633 bp, and *lincas3*/*LIC10938*; 2253 bp) was ligated between *Kpn*I and *Pac*I **(Appendix Fig. C1a-b)**. The obtained recombinant plasmid 'pCas' can overexpress a set of Cas proteins (rLinCas8b, rLinCas11b, rLinCas7, rLinCas5, rLinCas6, and rLinCas3) required for interference, wherein rLinCas8b possesses an

N-terminal 6×His-tag. The CRISPR array (LIC_Cr^2) associated with the Lin_I-B locus (Dixit et al., 2016; Prakash and Kumar, 2021) was ligated into pRSF-1b between *Kpn*I and *Hind*III sites, generating the recombinant plasmid construct 'pCr^2' **(Appendix Fig. C1c-d)** that can transcribe full-length array (pre-crRNA).

The plasmid constructs available in our laboratory [pET28a-LinCas7 (Hussain and Kumar, 2022), pCDF1b-LinCas11b, pET-SUMO-LinCas5, and pET-SUMO-LinCas3] were used for individual production of rLinCas7, rLinCas11b, rLinCas5, and rLinCas3, respectively. Wherein genes *lincas5* and *lincas3* were individually cloned into pET SUMO plasmid following the manufacturer's instructions (Invitrogen #K300-01). A single amino acid substitution mutation (Met429 to Ala) was introduced in *lincas8b* encoded within the plasmid pCas using Q5 site-directed mutagenesis kit (NEB # E0554S). The resulting plasmid (pCas$^{-Cas11b}$) overexpresses the interference-associated Cas proteins (rLinCas8b^{M429A}, rLinCas7, rLinCas5, rLinCas6, and rLinCas3) without rLinCas11b. All the plasmid constructs were sequenced before downstream uses.

4.5.3. Protein overexpression and purification

The *E. coli* BL21 (DE3) were co-transformed with two plasmids (pCas and pCr^2) for overexpression of Cas proteins (rLinCas8b, rLinCas11b, rLinCas7, rLinCas5, rLinCas6, and rLinCas3) and generation of LIC_Cr^2 pre-crRNA transcripts. Such co-transformation resulted in strain BL-pCas+pCr^2 of *E. coli* BL21 (DE3) and is anticipated to allow Cascade complex (rLinCascade) formation within *E. coli*. Induction of LinCas protein(s) overexpression in BL-pCas+pCr^2 strain was done by adding IPTG (1 mM) for 4 h at 37 °C. The ribonucleoprotein complex (rLinCascade) was purified by nickel affinity pull-down as described elsewhere (Beloglazova et al., 2015; Tan et al., 2022) with rLinCas8b as bait. The bound complex (rLinCascade) with the Ni-NTA beads was separated using the elution buffer [250 mM imidazole, 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 % glycerol and 1 mM PMSF (phenylmethylsulfonyl fluoride)]. The obtained elutes (rLinCascade) were concentrated using a centrifugal concentrator (Amicon #UFC901024) and stored at -20 °C. Similarly, as a control, purification was performed from *E. coli* cells (BL-pCas strain) overexpressing only the interference-associated Cas proteins. The 6×His-SUMO tagged rLinCas3 (98 kDa) was individually purified by Ni-NTA affinity chromatography under the native condition vide

section 2.5.2. For the purification of N-terminal 6×His-SUMO tagged rLinCas5 (37 kDa), since the protein was primarily in the insoluble fraction, Ni-NTA affinity chromatography was pursued under denaturing conditions followed by renaturation of protein via dialysis as described before (Ghosh et al., 2018).

4.5.4. Generation of polyclonal antibodies against the pure recombinant protein

Polyclonal antibodies specific to pure recombinant Cas proteins (rLinCas3, rLinCas5, and rLinCas7) of the Lin_I-B system were separately generated in mice, as described in section 3.5.5. The anti-rLinCas protein-specific polyclonal antibody titer in the mice serum was analyzed by enzyme-linked immunosorbent assay (ELISA) **(Appendix Fig. C3)**, as vide section 3.5.6. The immunization experiments were performed with the approval of the institutional animal ethics committee (approval no. 770/ac/CPCSEA/FVSC/AAU/IAEC/13-14) at the Department of Veterinary Microbiology, College of Veterinary Sciences, Assam Agricultural University, Guwahati, India.

4.5.5. Immunoblotting

Various experimental samples, including *E. coli* whole cell lysates and elutes of the pull-down assay, were resolved on 10 % SDS-polyacrylamide gel(s) and electroblotted onto nitrocellulose membrane(s) (Santa Cruz Biotechnology #SC-3718). Immunoblotting was performed (vide section 3.5.7) using rLinCas protein-specific polyclonal primary antibodies (mice antisera). The specific primary antibodies anti-rLinCas6 (Prakash and Kumar, 2021), anti-rLinCas3, and anti-rLinCas5 were used at a dilution of 1:1000, while anti-rLinCas7 and anti-rLinCas8b were used at 1:500.

4.5.6. Denaturing urea PAGE

The pull-down elutes of ribonucleoprotein complex from Ni-NTA chromatography, and its control was either treated with RNase A (2µg) or left untreated for 30 min at 37 °C. Post-treatment, the elutes were resolved on TBE-PAGE (20 %) with 8 M urea. As a marker, the commercially synthesized mature crRNA (70 nt) of LIC_Cr^2 was included. The nucleic acid component of the rLinCascade complex was visualized by SYBR Gold staining.

4.5.7. In silico PAM prediction for Lin_I-B system and the generation of plasmids containing the protospacer

The spacer matching-protospacer sequences were predicted by submitting the sense strand (+) DNA sequence of each spacer of LIC_Cr^2 (Prakash and Kumar, 2021) individually to the CRISPRTarget webserver (Biswas et al., 2013) under default parameters. The eight nucleotide flanking sequence at the 5′-end of each predicted protospacers (Table 4.1) was analyzed by the WebLogo webserver (Crooks, 2004) to deduce the consensus 5′-PAM. The protospacer DNA oligos containing 5′-PAM with spacer-1 of LIC_Cr^2 (Table 4.2) were custom synthesized (Eurofins, India). The custom-synthesized protospacer oligos (positive and negative strands) were annealed to have cohesive ends (*Hind*III and *Nde*I) and ligated into similarly restricted pT7Blue plasmid, as described previously (Strotksaya et al., 2015). Prepared PAM-protospacer plasmids were confirmed by sequencing.

4.5.8. Plasmid interference assay in *E. coli*

The target plasmid interference by Lin_I-B interference machinery in *E. coli* was assessed as described previously (Tan et al., 2022). Briefly, the plasmid constructs (pCas, pCr^2, and a target plasmid) were co-transformed into BL21-AI cells using the TransformAid bacterial transformation kit (Thermo Scientific #K2711). The target plasmids constructed in a backbone of pT7Blue [p(TGA)Spacer, p(ATA)Spacer, or p(ATG)Spacer] contained the identified 5′-PAM and the spacer-1 sequence of array LIC_Cr^2. As a control, pT7Blue plasmid with spacer-1 containing a non-PAM sequence [p(CAC)Spacer, non-PAM target plasmid] or without any protospacer (pBlank, reference plasmid) was utilized. For target plasmid interference assay, BL21-AI cells housing three plasmids (pCas, pCr^2, a target plasmid) were grown overnight at 37 °C in LB broth containing three antibiotics (Cam; 33 µg/mL, Kan; 100 µg/mL, and Amp; 100 µg/mL). Overnight grown cultures of each strain were subcultured in LB broth (2 mL) containing three antibiotics (Cam+Kan+Amp) till-mid log phase (OD_{600} = 0.5-0.6). After that, cultures were pelleted by centrifugation and gently resuspended in fresh LB broth (2 mL) containing two antibiotics (Cam+Kan). One-half (1 mL) of each culture was induced with IPTG (1 mM) and L-arabinose (0.2 %) at 37 °C for the next 4 h. The remaining half (1 mL) of the culture was grown uninduced for 4 h. Next, an equal volume of 100-fold dilution of uninduced and induced cultures of each strain was spread on LB-agar test plates with three antibiotics (Cam+Kan+Amp) and LB-agar control plates with two antibiotics (Cam+Kan).

After 24 h incubation at 37 °C, the culture plates were scored for bacterial growth by counting colony-forming units (CFU). The CFU depletion ratio in uninduced and induced cells of each BL21-AI strain was estimated by calculating the ratio of CFU scored on control and test plates.

Similarly, the CFU depletion ratio was estimated in the BL21-AI strain (pCas$^{-Cas11b}$, pCr^{2}, and target plasmid), lacking rLinCas11b. The same strain was also complemented with a fourth plasmid, pCDF1b-LinCas11b bearing a spectinomycin resistance cassette. The interference of potential target plasmid in rLinCas11b complemented BL21-AI cells was estimated following a similar procedure as detailed above, except that an additional antibiotic spectinomycin (Spe; 50 µg/mL) was used. The experiment was repeated at least three times, and the statistical analysis was performed using Microsoft office excel 2019.

*****

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

Chapter 5

Differential expression of *cas* genes of CRISPR-Cas type I-B locus in *L. interrogans* under stress conditions

5.1. Summary

The pathogenic *Leptospira interrogans* serovar Copenhageni strain Fiocruz L1-130 carries a transcriptionally active CRISPR-Cas type I-B (Lin_I-B) system. However, at the protein level, the expression of *cas* genes (*lincas3*, *lincas5*, *lincas6*, *lincas7*, and *lincas8b*) associated with the interference module of the Lin_I-B system is murky. In this study, on immunoblotting using specific polyclonal antibodies, none of the LinCas proteins associated with Lin_I-B's interference module could be detected in the whole-cell lysate of *L. interrogans* cultivated under optimal *in vitro* growth conditions. Therefore, we assessed the transcription level of Lin_I-B's *cas* genes within *L. interrogans* under various abiotic stress conditions (high osmolarity, oxidative stress, and UV-C irradiation) that leptospires frequently experience while living in soil, water, or infected animal tissues. The treatment of mid-log phase grown *Leptospira* with hydrogen peroxide (H_2O_2, 1 mM; oxidative stress) for 1 h resulted in significant down-regulation (fold change (FC) = 0.2<FC<0.6) of interference-associated *cas* genes of Lin_I-B locus. Moreover, oxidative stress did not trigger any significant change in transcription of the adaptation module-associated *cas* genes (*lincas1*, *lincas2*, and *lincas4*). Similarly, a 24 h exposure of high physiologic osmolarity (300 mOsM; 120 mM NaCl) to log-phase grown *Leptospira* significantly down-regulated (0.3<FC<0.6) the transcription of *cas* genes of Lin_I-B's interference module except for *lincas6*, which remained unchanged. Also, high osmolarity stress induced a significant up-regulation in the transcription of adaptation module-associated *cas* genes (*lincas1*: FC = ~5, $P < 0.05$; *lincas2*: FC = ~2.4, $P < 0.01$; *lincas4*: FC = ~3, $P < 0.01$). Further, the UV-C exposure (8 J/m^2) to exponentially growing *L. interrogans* had no effect on the transcription of the Lin_I-B associated *cas* genes. Overall, the present study demonstrates the influence of abiotic stress on the transcription of *cas* genes of the Lin_I-B system in *L. interrogans* serovar Copenhageni.

5.2. Results

5.2.1. Undetectable expression of Cas proteins of Lin_I-B's interference module in *in-vitro* grown *L. interrogans* by immunoblotting

Earlier, the genes coding for Cas proteins of the Lin_I-B system were reported to be transcriptionally active in *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 (Dixit et al., 2016). In this study, we examined the expression of Cas proteins associated with the interference module of the Lin_I-B system in *Leptospira* by immunoblotting using protein-specific mice antisera. On immunoblotting of whole-cell lysates of *in vitro* grown *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 (~5×10^9 spirochetes), none of the Cas proteins of Lin_I-B's interference module could be detected **(Appendix Fig. D1)**. However, in the same lysate, we could detect the LinCas2 of Lin_I-B's adaptation module **(Appendix Fig. D1)**, as reported previously (Dixit et al., 2021b). No detection on immunoblot may infer a very low expression level of Cas proteins of Lin_I-B's interference module in *Leptospira* when cultivated under optimal growth conditions- in EMJH media at 29 °C. The expression of *cas* genes in prokaryotes is differentially regulated in response to various environmental stimuli, including oxidative stress, high osmolarity, and UV irradiation (Götz et al., 2007; Nickel et al., 2013; Plagens et al., 2012; Strand et al., 2010; Zavala-Alvarado et al., 2020). Pathogenic leptospires encounter various stress conditions, including oxidative stress, high osmolarity, and UV irradiation during their life in the soil, water, or infected animal host. Therefore, next, we analyzed the expression of Lin_I-B associated *cas* genes in *L. interrogans* under various abiotic stress conditions (i.e., high physiologic osmolarity, oxidative stress, and UV irradiation) via qRT-PCR to check if it can be up-regulated to a level significant enough to detect by immunoblotting.

5.2.2. Influence of oxidative stress on transcription of Lin_I-B's *cas* genes

Previously, in *Pyrococcus furiosus*, a significant up-regulation (up to ~6-fold) in the transcription of the *cas7* gene associated with CRISPR-Cas type I-A system (Strand et al., 2010) was recorded in response to oxidative stress induced upon a 30 min treatment with 0.5 mM H_2O_2. However, in *L. interrogans* serovar Manaliae, a 1 h treatment with 1 mM H_2O_2 (oxidative stress) resulted in a significant down-regulation in the transcription of *cas* genes (*cas3*, *cas5*, *cas7*, and *cas8*) associated with type I-C system (Zavala-Alvarado et al., 2020). Moreover, the transcriptional response to different environmental stress conditions can vary

among prokaryotes (Götz et al., 2007; Nickel et al., 2013; Plagens et al., 2012; Strand et al., 2010; Zavala-Alvarado et al., 2020). In this study, we assessed the transcription level of *cas* genes of the Lin_I-B system within *L. interrogans* serovar Copenhageni in response to oxidative stress employing real-time quantitative reverse transcription-PCR (qRT-PCR). Details of *cas* gene-specific primers used for qRT-PCR analysis are enlisted in Table 5.1.

Table 5.1. *cas* gene specific primers used for qRT-PCR in this study.

Gene / locus tag	Gene size (bp)	Primer sequence (Forward / Reverse), 5′- 3′
lincas1 / LIC10942	855	CTAGCTAGCATGATTAAATTTTTAGTTTCCGAAGAG / ACTCACATTTTGCGGATACTAGTTT
lincas2 / LIC10941	285	CATGCCATGGACATGAAACATTGGAGATTAGTATCTT / ACGCGTCGACTTATAGAATTTTTAATGTTTTTGGAGC
lincas3 / LIC10938	2253	CTAGCTAGCGTGATCCTTCTCGCAAAAT / ATTGTATTCGGATTCTGAAAGTCGG
lincas4 / LIC10943	630	GGAATTCCATATGATGGACCAATACGGAAACA / GAAAGGATTCGATTCTTCCCGA
lincas5 / LIC10935	624	CTAGCTAGCATGGATCCATTGATTCTTTACT / AACACATTTCGCAGCTGGAG
lincas6 / LIC10939	633	CTAGCTAGCATGTCCATTCCGAACGTCA / CTCAGTCGTGAATTGGCTGTTAGTCGAAGG
lincas7 / LIC10936	840	CTAGCTAGCATGAGTCTGCATATTTTCGGAAC / CTGCCAGTCGAATCCTTCGCTT
lincas8 / LIC10937	1887	CTAGCTAGCATGAGTTCCCGTCCAAGC / GAATAGAATCCTGAGTCAGTTCCCAGGACAAAGTC
linrrs1 / LIC11010	1513	GCGTAGGCGGACATGTAAGT / AATCCCGTTCACTACCCACG

For expression analysis of Lin_I-B's *cas* genes in response to the harmful oxidative stress encountered during infection, mid-log phase grown *L. interrogans* were treated with or without (control) 1 mM H_2O_2 for 1 h in EMJH medium at 29 °C. The transcripts of *cas* genes measured in *Leptospira* post-exposure to hydrogen peroxide (H_2O_2) were normalized with internal standard 16S rRNA (*rrs1*) transcripts using the $2^{-\Delta\Delta CT}$ method and represented as the number of gene transcripts per 10^{10} copies of the 16S rRNA. The fold-change in the transcript level of the target *cas* genes upon exposure of spirochetes to oxidative stress were calculated with respect to basal-level expression under control condition. As reported in *L. interrogans* serovar Manaliae (Zavala-Alvarado et al., 2020), oxidative stress caused significant down-regulation [expression relative to control or fold change (FC) = 0.2<FC<0.6] in interference-associated *cas* genes of Lin_I-B locus **(Fig. 5.1)**. Maximum transcriptional repression (FC = ~0.3, P < 0.01) was recorded for *lincas5*, *lincas7*, and *linas8b*, followed by *lincas3* (FC = ~0.5, P < 0.05),

and *lincas6* (FC = ~0.6, $P < 0.05$). No significant change in the transcripts level of the Lin_I-B's adaptation module (*lincas4-lincas1-lincas2*) was evident under the tested oxidative stress condition **(Fig. 5.1)**.

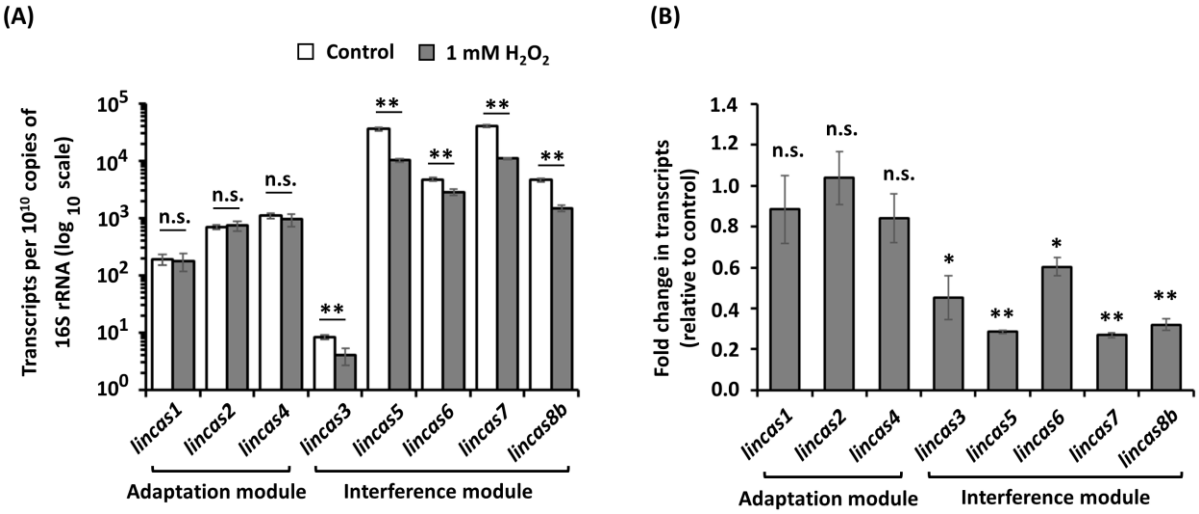

Fig. 5.1. **Effect of oxidative stress on the transcription of *cas* genes of Lin_I-B locus.** Mid-log phase grown *L. interrogans* cultures in EMJH media were treated with or without (control) 1 mM H_2O_2 for 1 h at 29 °C. From each culture sample (treated and control), total RNA was isolated followed by cDNA synthesis using random hexamers. qRT-PCR was performed using gene specific primers. Transcripts of *cas* genes were quantified based on threshold cycle (C_T) values using the $2^{-\Delta\Delta CT}$ method and normalized against internal standard 16S rRNA (*rrs1*). **(A)** Transcript level of Lin_I-B's *cas* genes in *L. interrogans* under optimal (control) and oxidative stress (1 mM H_2O_2) conditions. **(B)** Fold-change in transcript level of *cas* genes of Lin_I-B locus within *Leptospira* in response to applied oxidative stress relative to control. Data are presented as mean ± SEM, n = 3. (n.s. not significant; *$P < 0.05$; **$P < 0.01$)

5.2.3. Influence of high osmolarity on transcription of Lin_I-B's *cas* genes

Upon entry into the animal host, leptospires encounter high physiological osmolarity stress (~300 mOsM) (Kratz et al., 2004; Suckow et al., 2005). Previously, exposure of physiological osmolarity to *in vitro* cultivated pathogenic leptospires via treatment with 120 mM NaCl in EMJH media has been reported to up-regulate the expression of many virulence-associated genes, including *ligA* (*LIC10465*), *ligB* (*LIC10464*), *sph2* (*LIC12631*) (Matsunaga et al., 2007), and *lsa66* (*LIC10258*) (Oliveira et al., 2011). Also, in *Thermoproteus tenax*, the expression of the *cas3* gene of CRISPR-Cas type I-A locus was induced up to 10-fold in response to treatment with 100-150 mM NaCl (Plagens et al., 2012). However, treating *T. tenax* with a lower concentration of NaCl (50 mM) caused a 7-fold decrease in *cas3* expression level (Plagens et

al., 2012). This prompted us to analyze the expression of *cas* genes of the Lin_I-B locus in *L. interrogans* grown *in vitro* under physiological osmolarity conditions (~300 mOsM). Therefore, the exponentially growing *Leptospira* culture was treated with 120 mM NaCl in EMJH medium for 24 h, followed by gene expression analysis by qRT-PCR. The addition of 120 mM NaCl to the EMJH medium results in an osmolarity of ~300 mOsM (Matsunaga et al., 2007) that mimics the physiologic osmolar environment experienced by leptospires during mammalian host infection. The applied high osmolarity stress significantly down-regulated (0.3<FC<0.6) the expression of *cas* genes associated with Lin_I-B interference module except the *lincas6* gene, which remained unchanged **(Fig. 5.2)**. Interestingly, under high osmolarity stress, a significant up-regulation in the transcription of adaptation module-associated *cas* genes (*lincas1*: FC = ~5, $P < 0.05$; *lincas2*: FC = ~2.4, $P < 0.01$; *lincas4*: FC = ~3, $P < 0.01$) was evident **(Fig. 5.2)**.

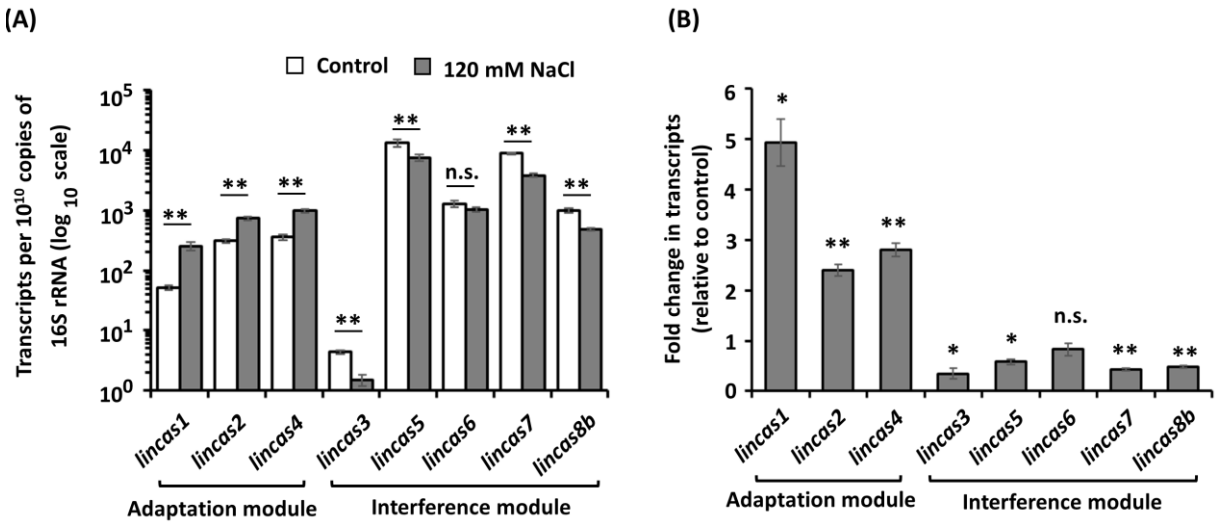

Fig. 5.2. Differential expression of Lin_I-B's *cas* genes on physiologic osmolarity stress. Mid-log phase grown *L. interrogans* cultures in EMJH media were treated with or without (control) 120 mM NaCl for 24 h at 29 °C. Transcripts of *cas* genes were quantified by qRT-PCR (vide methods). **(A)** Transcript level of Lin_I-B's *cas* genes in *L. interrogans* under optimal (control) and physiologic osmolarity stress (120 mM NaCl) conditions. **(B)** Fold-change in transcript level of *cas* genes of Lin_I-B locus within *Leptospira* in response to applied osmolarity stress relative to control. Data are presented as mean ± SEM, n = 3. (n.s. not significant; *$P < 0.05$; **$P < 0.01$)

5.2.4. Influence of UV irradiation on transcription of Lin_I-B's *cas* genes

Earlier, UV-C (254 nm) irradiation has been shown to up-regulate the expression of CRISPR associated *cas* genes in some prokaryotes. In *T. tenax*, a UV-C exposure of 20 J/m^2 resulted in ~3-fold and ~2-fold increased transcription of *cas3* and *cas4* genes of CRISPR-Cas type I-A locus, respectively (Plagens et al., 2012). Similarly, in *Sulfolobus solfataricus*, UV-C exposure of 200 J/m^2 caused ~2-fold up-regulation in the transcription of *cas8* of type I-A locus (Götz et al., 2007). Pathogenic leptospires also encounter UV exposure during their free-living state in soil and water. The sensitivity of pathogenic spirochetes to UV irradiation has previously been determined (Stamm and Charon, 1988). For *L. interrogans* serovar Pomona, cell survival was estimated to be 1 % and 10 % after UV-C exposure of 12.5 J/m^2 and 7.5 J/m^2, respectively (Stamm and Charon, 1988). In another study, UV-C exposure of 100 J/m^2 to *L. interrogans* serovar Manilae has been shown to reduce bacterial survivability by 99.9 % (three logs reduction) (Richard et al., 2021). Therefore, we chose a UV-C dose of 8 J/m^2 to analyze the effect of UV irradiation on the expression of *cas* genes associated with the Lin_I-B locus in *L. interrogans* serovar Copenhageni. The exponential phase culture of *L. interrogans* was exposed to UV-C at a dose of 8 J/m^2. After UV-C exposure, the culture was incubated at 29 °C for the next 3 h, followed by cas gene transcripts analysis by qRT-PCR. The UV-C exposure (8 J/m^2) did not result in any significant change in the transcription of *cas* genes of the Lin_I-B locus **(Fig. 5.3)**.

5.3. Discussion

The CRISPR-Cas systems are highly diverse in molecular organization and operational mechanism. In addition to its primary role as adaptive immunity, CRISPR-Cas systems appear to be involved in various other cellular functions, including DNA repair, virulence, and programmed cell death (Faure et al., 2019). Multiple lines of evidence indicate the involvement of Cas proteins in bacterial virulence, although the detailed mechanism remains elusive. For instance, the Cas9 of *Francisella novicida* has been reported to foster bacterial virulence by degrading the endogenous transcript encoding a bacterial lipoprotein that otherwise triggers the host's innate immune response (Sampson et al., 2013). Similarly, in *Streptococcus agalactiae*, Cas9 enhances bacterial virulence by downregulating the expression of the transcription regulator *reg*R, the repressor of the gene encoding the hyaluronidase enzyme (Ma et al., 2018).

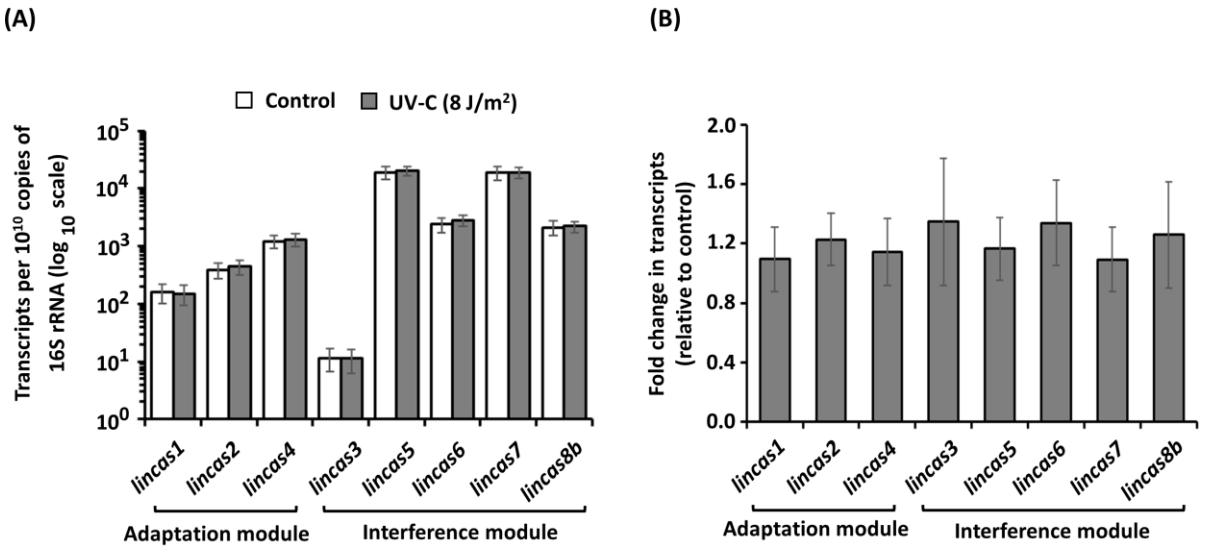

Fig. 5.3. Transcript analysis of *cas* genes in *Leptospira* exposed to UV-C irradiation. Mid-log phase grown *L. interrogans* cultures in EMJH media were exposed to UV-C light at a dose of 8 J/m^2 followed by incubation in dark for next 3 h at 29 °C. Similarly grown *Leptospira* culture without UV irradiation served as control. Transcripts of *cas* genes were quantified by qRT-PCR (vide methods). **(A)** Transcript level of Lin_I-B's *cas* genes in *L. interrogans* grown without (control) or with UV-C irradiation (8 J/m^2). **(B)** Fold-change in transcript level of *cas* genes of Lin_I-B locus within *Leptospira* in response to UV-C exposure relative to control. Data are presented as mean ± SEM, n = 3. No significant change in transcript level of Lin_I-B's *cas* genes was evident.

The nuclease activity of the Cas2 protein of CRISPR-Cas type II-B locus in *Legionella pneumophila* has been demonstrated to promote bacterial infectivity for amoebae (Gunderson et al., 2015). Adding more, in *Myxococcus xanthus*, the starvation condition positively regulates the expression of Cas proteins (Cas7, Cas5, and Cas8: previously described as Dev-R, -S, and -T, respectively) of CRISPR-Cas type I-B locus leading to sporulation of bacteria, a bacterial virulence-associated strategy to survive the stress condition (Rajagopalan and Kroos, 2017). Also, in pathogenic leptospires, the CRISPR-Cas systems have been described as a virulence factor (Fouts et al., 2016) similar to some other bacteria, including *Listeria monocytogenes* (Mraheil et al., 2011) and *Campylobacter jejuni* (Louwen et al., 2014, 2013).

Further, pathogenic bacteria regulate the expression of virulence factors as per environmental cues. For instance, the expression of virulence-associated LigA protein could not be seen in *L. interrogans* cultivated *in vitro* but was detected in leptospires colonized in the kidneys of infected hamsters (Palaniappan et al., 2002). Similarly, the expression of Sph2 and Lsa66 was found to be triggered in leptospires when cultured under host-like high osmolarity stress, which

otherwise failed to be detected under standard *in vitro* culture conditions (Matsunaga et al., 2007; Oliveira et al., 2011). Thus, in this study, on immunoblotting, no detection of Cas proteins of Lin_I-B's interference module led us to presume that the expression of Cas proteins is too low to be detected by immunoblot when spirochetes were cultivated in EMJH media at 29 °C, an optimal growth condition with no stress. This agrees with the previous report on mass-spectrometry-based absolute quantification of protein copies per cell in *L. interrogans* by Malmstrom and his colleagues (Malmström et al., 2009). Wherein none of the Cas proteins of Lin_I-B locus except LinCas7 could be detected in *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 grown in EMJH media at 30 °C. By mass-spectrometry, the amount of LinCas7 protein was determined to be 17 copies per cell (~80 nM) (Malmström et al., 2009), which is too low (in the range of fmol) to be detected by the immunoblotting technique that has sensitivity down to pmol (Prasad and Unadkat, 2014). On the other hand, the detection of the LinCas2 protein of Lin_I-B locus in *L. interrogans* serovar Copenhageni by immunoblotting rather suggests a knotty setting for the expression of LinCas proteins in *Leptospira*. The existing literature suggests that the expression of *cas* genes in prokaryotes is subject to differential regulation under various abiotic stress conditions, such as high osmolarity, oxidative stress, and irradiation (Götz et al., 2007; Nickel et al., 2013; Plagens et al., 2012; Strand et al., 2010; Zavala-Alvarado et al., 2020). Therefore, we wonder if the expression of Cas proteins of the Lin_I-B locus can be manipulated by mimicking the abiotic stress conditions (i.e., high osmolarity, oxidative stress, and UV irradiation) that leptospires experience while living in the soil, water, or infected animal hosts.

Pathogens, including *Leptospira*, during infection, encounter DNA-damaging stress from the host tissue environment and innate immune response. Inside the animal body, bacteria experience a high physiologic osmolar environment (300 mOsM) (Kratz et al., 2004; Suckow et al., 2005) compared to what they face outside the host. The host's innate immune response generates oxidative stress by producing various reactive oxygen species (ROS), including hydrogen peroxide (H_2O_2), hydroxyl radicals, and superoxide anions (Araújo et al., 2014; Erdogan et al., 2008; Eshghi et al., 2012; Marangoni, 2006). The oxidative stress, induced by an increase in ROS concentrations to the supraphysiological level (H_2O_2 >1 mM), causes extensive damage to bacterial DNA (Eshghi et al., 2012; Fan et al., 2010). Also, outside the host, bacteria experience DNA damage by UV irradiation (Götz et al., 2007). DNA damage induced by such stress conditions, to some extent, mocks the situation a bacterium experiences

during phage infection; therefore, it plausibly stimulates the expression of proteins involved in DNA repair and maintenance, including Cas proteins (Babu et al., 2011; Plagens et al., 2012). Agreeing with this, previously, noticeable up-regulation in the transcription of CRISPR-Cas-associated genes in different bacteria has been reported in response to abiotic stress conditions. Examples include, in *P. furiosus*, an up to ~6-fold increase in the transcription of the *cas7* gene (CRISPR-Cas type I-A) was recorded upon treating the bacteria with 0.5 mM H_2O_2 for 30 min (Strand et al., 2010). The UV-C exposure of 20 J/m^2 to *T. tenax* resulted in ~3-fold and ~2-fold up-regulation in the transcription of *cas3* and *cas4* genes (CRISPR-Cas type I-A), respectively (Plagens et al., 2012). Similarly, in *S. solfataricus*, UV-C stress of 200 J/m^2 caused ~2-fold up-regulation in the transcription of *cas8* of type I-A locus (Götz et al., 2007). Further, treating the *T. tenax* with 100-150 mM NaCl (osmolarity stress) resulted in up to a 10-fold increased expression of the *cas3* gene (CRISPR-Cas type I-A) (Plagens et al., 2012). However, treating *T. tenax* with a lower concentration of NaCl (50 mM) caused a 7-fold down-regulation in *cas3* expression level (Plagens et al., 2012). Similarly, the treatment of *Methanosarcina mazei* with 500 mM NaCl did not affect the transcription level of *cas* genes of the type I-B system, although the expression and maturation of crRNA were increased by ~5-10-fold (Nickel et al., 2013). In *L. interrogans* serovar Manaliae, a significant down-regulation in the transcription of *cas* genes (*cas3, cas5*, cas7, and *cas8*) of type I-C locus was recorded post-treatment with 1 mM H_2O_2 (Zavala-Alvarado et al., 2020).

Similar to *L. interrogans* serovar Manaliae, in this study, the oxidative stress (1 mM H_2O_2) to *L. interrogans* serovar Copenhageni significantly down-regulated ($0.2<FC<0.6$) the interference-associated *cas* genes of Lin_I-B locus. Such diverse transcriptional responses of different bacteria may be correlated with their different sensitivity towards a similar environmental stress condition. For *L. interrogans* serovar Copenhageni and Manaliae, 1 mM of H_2O_2 is sublethal (Eshghi et al., 2012). While for *P. furiosus*, H_2O_2 at 0.5 mM concentration is sublethal (Strand et al., 2010). Significant down-regulation ($0.3<FC<0.6$) in the transcription of *cas* genes (except *lincas6*) of Lin_I-B's interference module was also detected in the case of high osmolarity stress (~300 mOsM), while the adaptation module's *cas* genes got up-regulated. The up-regulation of adaptation *cas* genes of the Lin_I-B locus in response to the physiologic osmolar condition may infer their possible role in counter strategies of *Leptospira* to survive the stress during animal host infection, hence suggesting their role in *Leptospira* virulence. Previously, Cas1 and Cas2 proteins' role as virulence factors has been illustrated in

different bacteria, including *Legionella pneumophila* (Campbell and Cianciotto, 2022) and *E. coli* (Sampson and Weiss, 2013). Earlier, CRISPR adaptation-associated Cas proteins are suggested to be actively involved in DNA repair (Babu et al., 2011; Faure et al., 2019; Wörtz et al., 2022). High osmolarity (NaCl) stress can lead to cellular DNA damage (Dmitrieva et al., 2004); thus, it may cause stimulation in the expression of proteins that serve in DNA repair. Therefore, in this study, under high osmolar conditions, the observed up-regulation in the Lin_I-B's adaptation *cas* genes (*lincas1*, *lincas2*, and *lincas4*) indicates their active role in DNA repair. On the other hand, the down-regulation in the interference *cas* genes due to oxidative and high osmolarity stress or no change in expression upon UV-C exposure suggests their no active involvement, at least, in the stress management strategies of *Leptospira*. Overall, the present study demonstrates the effect of abiotic stress on the transcription of *cas* genes of the Lin_I-B system in pathogenic *L. interrogans* serovar Copenhageni. Our results complement previous studies reporting the variable transcriptional response of different microorganisms to the same abiotic stress.

5.4. Conclusion

Although it is hard to come to a more specific conclusion, the interference associated *cas* genes of the Lin_I-B locus appear to be more responsive towards the tested abiotic stress conditions, including high osmolarity and oxidative stress. Moreover, the down-regulation in the interference *cas* genes due to oxidative and high osmolarity stress or no change in expression upon UV-C exposure suggests their no active involvement, at least, in stress management strategies of *Leptospira*. At the same time, an up-regulation in the *cas* genes of Lin_I-B's adaptation module in response to high physiologic osmolar stress infers their active role in *Leptospira* virulence. Overall, the present study demonstrates the influence of abiotic stress on the transcription of *cas* genes of the Lin_I-B system in pathogenic *L. interrogans* serovar Copenhageni.

5.5. Materials and methods

5.5.1. Bacterial strains, media, and growth conditions

L. interrogans serovar Copenhageni strain Fiocruz L1-130 (hereafter referred to as *L. interrogans* for simplicity) was procured from the ICMR, RMRC, Port Blair, Andaman, and Nicobar Island, India. The spirochete culture was maintained in EMJH medium (Difco, Becton

Dickinson India Pvt. Ltd., New Delhi, India) vide section 2.5.1. To avoid other bacterial contamination, EMJH culture media was supplemented with antibiotic 5′-fluorouracil (100 µg/mL).

5.5.2. Immunoblotting

The whole-cell lysates of late-log phase grown *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 (~5 × 10^9 spirochetes) were prepared in 1× SDS-PAGE loading dye and heat denatured at 95 °C for 20 min. Lysates were clarified by centrifugation at 13000 × g for 5 min. Supernatant fractions were resolved on 12 % SDS-polyacrylamide gel(s) and electroblotted onto nitrocellulose membrane(s) (Santa Cruz Biotechnology #SC-3718). As a positive control, pure rLinCas proteins were also included. Immunoblotting was performed using rLinCas protein-specific polyclonal primary antibodies following the procedure described in section 3.5.7. The specific primary antibodies (mice antisera), anti-rLinCas6 (Prakash and Kumar, 2021), anti-rLinCas3, and anti-rLinCas5 were used at a dilution of 1:1000, while anti-rLinCas7 and anti-rLinCas8b were used at 1:500 dilution. The rabbit antisera anti-rLinCas2 primary antibody (Dixit et al., 2021b) was used at 1:1000 dilution. The respective secondary antibodies horseradish peroxidase (HRP)-conjugated goat anti-mouse or anti-rabbit IgG (1:5000; Sigma) were applied for 1 h at room temperature.

5.5.3. Stress treatment to *Leptospira*

For assessing the transcriptional response of *L. interrogans* under abiotic stress conditions, spirochetes were grown in EMJH media at 29 °C till the mid-log phase. Spirochetes were enumerated using a Patroff-Hausser counting chamber under a dark-field microscope (Zeiss Axio Lab.A1). To induce oxidative stress, mid-log phase grown *L. interrogans* culture (3 × 10^8 cells/mL; 10 mL) in EMJH medium was treated with or without (control) 1 mM H_2O_2 for 1 h at 29 °C, as described elsewhere (Eshghi et al., 2012; Zavala-Alvarado et al., 2020). To induce high osmolarity stress, the *L. interrogans* culture, in the mid-log phase, was added with or without (control) 120 mM of NaCl (filter sterilized) followed by 24 h incubation at 29 °C, as illustrated previously (Matsunaga et al., 2007). The addition of 120 mM of NaCl raises the osmolarity of the EMJH culture medium from ~67 mOsM to ~300 mOsM, mimicking the physiologic osmolar environment in animal tissue (Matsunaga et al., 2007). For UV irradiation, mid-log phase grown *L. interrogans* culture (3 × 10^8 cells/mL; 10 mL), under sterile condition,

was transferred to transparent petri dishes (Tarsons #T460050) and treated with UV-C (254 nm) at a dose of 8 J/m^2 using UV crosslinker (Analytik Jena CL-1000). Post UV-C exposure, culture samples (control and treated) were incubated in the dark at 29 °C for the next 3 h.

5.5.4. RNA isolation and real-time qRT-PCR

Post-treatment with specified abiotic stress conditions, the *L. interrogans* culture samples (control and treated) were harvested by centrifugation at 4000 × g for 20 min at room temperature. Total RNA was isolated using the TRIzol reagent (Life technologies #15596018) as described elsewhere (Dixit et al., 2016). Possible DNA contamination in isolated total RNA was removed by DNase I (NEB #M0303S) treatment. An equal amount (1 – 1.5 µg) of total RNA samples (control and treated) was converted to cDNA (Complementary DNA) using the Verso cDNA synthesis kit (Thermo Scientific #AB1453B). To rule out the genomic DNA contamination, as a negative control, the cDNA synthesis was set without the reverse transcriptase. The primers for qRT-PCR were designed by the OligoPerfect primer designer tool (Thermo Fisher Scientific) using the target gene sequence of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 available from the NCBI database. The primer pairs (Table 5.1) for qRT-PCR were designed with similar annealing temperatures (~60 °C) and similar amplicon sizes (~200 bp). Each primer pair was verified for single target-specific PCR amplification. The qRT-PCR reaction was performed using a DyNAmo ColorFlash SYBR Green qPCR kit (Thermo Scientific #F-416) and a CFX96^{TM} Real-Time machine (Bio-rad) programmed as 50 °C for 2 min, 95 °C for 10 min, and 40 cycles of 95 °C for 15 sec and 60 °C for 1 min, followed by melt curve analyses of the PCR products. As described previously in our laboratory (Ghosh et al., 2018), the transcripts of target genes were quantified based on threshold cycle (C_T) values using the $2^{-\Delta\Delta CT}$ method (Schmittgen and Livak, 2008) and normalized with the internal standard 16S rRNA (*rrs1*) gene of *Leptospira*. The transcription level of target genes in response to applied abiotic stress conditions relative to the control was determined in three independent experiments to obtain statistically significant results.

*****

Chapter 6

Conclusion and future prospects

Leptospira infection poses a significant public health challenge and causes a substantial economic loss, as well, due to *Leptospira*-induced ailments in livestock (Costa et al., 2015; Ellis, 2015; Torgerson et al., 2015). Even so, the pathogenesis mechanism of *Leptospira* remains poorly understood, largely because the genus exhibits incompetence to most of the available genetic manipulation tools (Fernandes et al., 2021). Thus, there is a need to explore alternative techniques of higher efficiency for investigating pathogenic leptospires. The predominant CRISPR-Cas type I-B system, identified in several pathogenic *Leptospira* genomes (Dixit et al., 2016; Xiao et al., 2019a), presents a promising alternative to be explored as an endogenous genome editing tool, as recently demonstrated in some other bacteria (Xu et al., 2021). Moreover, to exploit *Leptospira's* CRISPR-Cas type I-B (Lin_I-B) system for genetic manipulation, characterization of its interference machinery is a prerequisite, owing to the high variability in composition and target interference mechanism of CRISPR-Cas systems in different bacteria. To this end, the present study aims to the functional characterization of interference machinery of a transcriptionally active CRISPR-Cas type I-B system (Lin_I-B), recently identified in the *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 (Dixit et al., 2016).

We investigated the molecular and/or biochemical characteristics of LinCas7 and LinCas8b proteins to interpret their mechanistic role in the defense response of the Lin_I-B system. In line with the mandated role of Cas7 family proteins to shape the Cascade's backbone, pure LinCas7 exhibits distinct binding to mature crRNA in the presence of Mg^{2+} ions. The RNA-induced assembly of LinCas7, evident by transmission electron microscopy and SEC, confirms its canonical function in Lin_I-B defense response. Moreover, without divalent metal ions (Mg^{2+}), LinCas7 behaves as a non-specific RNase. Interestingly, the comprehensive biochemical analysis of the pure rLinCas7 revealed it to be a Mg^{2+} ion-dependent non-specific endoDNase, unlike other Cas7 family proteins. Perhaps, catalytically active Cas7 variants may also exist across other type I systems but remain underreported due to milder nuclease action on DNA, as noticed for DvuCas7c in this study.

At this juncture, the importance of the Cas7 subunit's non-specific nuclease activity for CRISPR-Cas immunity remains elusive. Moreover, the additional DNase activity in a Cascade complex furnished by Cas7, along with Cas3 helicase/nuclease, may accelerate the target DNA destruction and thus may benefit the bacteria to confront the invasive genome more effectively. Nonetheless, the observed divalent metal ions- based tuneability of LinCas7's nuclease activity may represent one of the regulatory mechanisms that operate inside bacterial cells to seize LinCas7's nuclease activity. We also reasonably hypothesized that, in a Cascade complex, the other interacting Cas subunits may regulate the LinCas7 non-canonical nuclease activity and direct it to execute its mandate role in CRISPR immunity. In line with this hypothesis, the expression of LinCas7 with cognate pre-crRNA and other subunits of the effector complex (LinCascade) resulted in the formation of LinCascade containing mature crRNA. The formation of LinCascade infers that other co-working Cas proteins regulate LinCas7 non-canonical nuclease activities, letting LinCas7 assemble onto crRNA without cleaving it. Further scrutiny of which Cas subunit(s) of the LinCascade and additional factor(s) (if any) regulates the functioning of LinCas7 would be an insightful study. Very recently, a Cas7 variant (Csb1) in the type I-G system of *Bifidobacterium animalis* has also been demonstrated to exhibit RNase activity independent of divalent metal ions (Chhetry and Anand, 2022). On the contrary, Cas7 in type I-G of *Thioalkalivibrio sulfidiphilus* was reported inert for RNase activity (Shangguan et al., 2022). These examples, along with LinCas7 (characterized in this study), convey the existing high diversity in the operational mechanism of CRISPR-Cas immunity and emphasize the importance of studying the same CRISPR-Cas type in different bacteria.

The molecular characterization of LinCas8b- the predicted large subunit of the effector complex in the Lin_I-B system- revealed it to be a genetic fusion of large (Cas8) and small subunit (Cas11) proteins found in a typical Cascade complex. We identified that the *lincas8b* gene encloses a small ORF (*lincas11b*) that codes for the C-terminal of LinCas8b as well as independently co-translates into LinCas11b (small subunit) protein. Further analysis revealed that the co-translating LinCas11b is a structural and functional analog of Cas11 family proteins constituting the belly filament of effector complex in type I and III CRISPR-Cas systems (McBride et al., 2020; Venclovas, 2016). LinCas11b's self-dimerization and its sturdy interaction with LinCas8b present molecular insights into the assembly of belly filament in the Cascade of various CRISPR-Cas type I systems that lack distinct *cas11* genes. The self-

dimerization property of LinCas11b suggests that the LinCas8b's C-terminal (representing a fused LinCas11b) may act as nucleation points for the assembly of LinCas11b belly filament in the LinCascade.

Pure LinCas11b demonstrated metal ion-independent non-specific DNA binding activity, leading us to interpret its functional role in Cascade-mediated DNA interference. The Cas11 subunit's non-specific DNA binding activity may benefit the Cascade in initial binding to target DNA for PAM detection. Also, such non-specific affinity allows Cas11 to hold the displaced DNA strand leading to stable R-loop formation during target interference. The structural analysis of LinCas11b disclosed a two-domain architecture comprising NTD and CTD. The NTD (α1- α5) exhibits a close structural match with the various known Cas11 family proteins (McBride et al., 2020; Venclovas, 2016). However, the CTD (α6- α9; 84 residues) is identified as unique to LinCas11b and requires further investigation to probe its role in LinCas11b's functioning. To better interpret the contribution of LinCas11b in the CRISPR-Cas adaptive immunity of *Leptospira*, the experimental structure determination of LinCas11b alone or in complex with LinCas8b or LinCascade is warranted.

In the type I-B system, typically, Cas6 processes the pre-crRNA into mature crRNA following multi-turnover kinetics. Cas6 gets dissociated from crRNA post-processing and does not form a subunit of the type I-B Cascade (Liu and Doudna, 2020; Sefcikova et al., 2017). In contrast, Cas6 (LinCas6) from a type I-B (Lin_I-B) system of *L. interrogans* has recently been demonstrated to process the cognate pre-crRNA in a single-turnover mode and remains bound with mature crRNA in vitro (Prakash and Kumar, 2021). This contradictory observation raised an immediate query to us- whether LinCas6 forms a subunit in the Cascade (LinCascade) complex of type I-B system in *L. interrogans*, unlike other bacteria. Therefore, in the present study, we examined the crRNA generation by LinCas6 in the presence of the other Cas proteins of Lin_I-B interference machinery within surrogate host *E. coli* and checked for LinCascade assembly. The co-expression of Lin_I-B interference machinery resulted in the formation of LinCascade comprising mature crRNA complexed with LinCas5, LinCas6, LinCas7, LinCas8b, and LinCas11b subunits. This infers that the Lin_I-B interference machinery is active for the expression phase of CRISPR immunity, leading to the crRNA biogenesis and assembly of a stable LinCascade complex which also encloses LinCas6 as its subunit.

Motivated by the perceptible assembly of LinCascade, the proficiency of the complex to recognize PAM and mount CRISPR defense against target plasmid was assessed within surrogate host *E. coli*. Wherein, LinCascade demonstrated a functional phenotype resulting in a robust interference of target plasmids containing a protospacer with a predicted 5′-PAM (TGA, ATG, and ATA). The active defense response of LinCascade against all the protospacers (containing 5′-TGA, -ATA, or -ATG PAM) suggests promiscuity in PAM recognition by LinCascade for target interference. Such promiscuity in PAM recognition for target interference is often noticed in all type I CRISPR-Cas systems across different bacteria and is presumed to be in line with the fitness need of the host to counter the mutational viral evasions (Cass et al., 2015a; Gleditzsch et al., 2019). Thus, in this study, the experimentally determined 5′-PAMs (TGA, ATA, and ATG) may portray only a small subset of interference-efficient PAMs for the Lin_I-B system. Therefore, a systematic analysis of all the trinucleotide combinations is warranted to deduce a plausible number of interference-efficient PAM for the Lin_I-B system. Furthermore, previous attempts to record the naïve acquisition in Lin_I-B CRISPR locus via establishing the CRISPR adaptation machinery of *Leptospira* in a heterologous host *E. coli* failed (Dixit, 2021), hinting the alternative mechanism, primed adaptation, may be functional. Therefore, the functional characterization of Lin_I-B interference machinery and its PAM in this study may allow the researcher to assay the primed adaptation in the Lin_I-B system.

We validated that the co-translating LinCas11b is indispensable for the defense response of the Lin_I-B system, as in its absence, Lin_I-B interference machinery could not accomplish target DNA interference. Moreover, further insight into the molecular mechanism of interference by the Lin_I-B system awaits experimental structure determination of the LinCascade. To date, the experimentally determined structure for type I-B effector Cascade or its subunits (except Cas6) from any bacteria is not available.

In surrogate *E. coli*, Lin_I-B interference machinery displayed a functional phenotype, thus anticipated to be active in its native host, *L. interrogans*, as well. As reported earlier (Dixit et al., 2021b), on immunoblotting, we could detect the expression of the LinCas2 in the whole-cell lysate of the native host, *L. interrogans* serovar Copenhageni strain Fiocruz L1-130. To our surprise, none of the Cas proteins of Lin_I-B's interference module could be detected. Previously, during mass-spectrometry-based proteome profiling of the same strain, Malmstrom

and co-workers reported the detection of only LinCas7 but at a very low amount (17 copies/cell) (Malmström et al., 2009). This led us to presume that the expression of interference-associated Cas proteins is too low to be detected by immunoblot when spirochetes were cultivated under optimal growth conditions- in EMJH media at 29 °C. Therefore, as described elsewhere (Plagens et al., 2012; Zavala-Alvarado et al., 2020), we analyzed the expression of Lin_I-B associated *cas* genes in *L. interrogans* under various abiotic stress conditions (i.e., high physiologic osmolarity, oxidative stress, and UV irradiation) via qRT-PCR to check if it can be up-regulated to a level significant enough to detect by immunoblotting. Under tested stress conditions, none of the interference-associated *cas* genes showed up-regulation; instead, a significant downregulation was recorded in response to oxidative and high osmolarity stress. Moreover, high osmolarity stress resulted in a significant up-regulation of the adaptation-associated *cas* genes, suggesting their active role in stress management, hence, in *Leptospira* virulence. Immunoblotting of the fresh *L. interrogans* isolates from an infected host may confirm the expression of Lin_I-B's Cas proteins. In sum, the current study presents the molecular and functional insight of the Lin_I-B interference machinery that may soon pave the way for scientists to harness it for endogenous genome editing and expand the toolbox for targeted genetic manipulation in recalcitrant *Leptospira* spp.

*****

Appendix-A: Supplementary data to chapter 2

Table A1. Plasmids and primers used in this study

Plasmid	Specificity	Source
pET28a (5369 bp)	Bacterial expression vector with T7lac promoter, N-terminal 6×His tag, thrombin cleavage site, C-terminal His tag; kanamycin resistance	Novagen #69864-3
pGST-TEV (5720 bp)	Bacterial expression vector with T7lac promoter, N-terminal GST tag, TEV cleavage site; ampicillin resistance	BioBharati LifeScience #BB-V0021
pTZ57R/T (2886 bp)	Linearized TA vector with 3′-ddT overhangs for TA cloning of PCR products; ampicillin resistance	Part of a kit, Thermo Scientific #K1214
Primers	***Sequence (5′-3′)**	**Purpose**
LinCas7-F (*Nhe*I)	CTA**GCTAGC**ATGAGTCTGCATATTTTCGGAAC	Cloning of *lincas7* (*LIC10936*) into pET28a
LinCas7-R (*Xho*I)	CCG**CTCGAG**TTATGCGCTCTTTCCAATCC	
LinCas7-F (*Eco*RI)	CG**GAATTC**ATGAGTCTGCATATTTTCGGAACG	Used with primer LinCas7-R (*Xho*I) for cloning of *lincas7* into pGST-TEV
DvuCas7c-F (*Eco*RI)	CG**GAATTC**ATGACCGCCATTGCCAACAGAT	Cloning of *dvucas7c* into pGST-TEV
DvuCas7c-R (*Xho*I)	CCG**CTCGAG**CTACAGCATCTCTTTGACCTCT	
LIC10940-F (*Nhe*I)	CTA**GCTAGC**GTGGATATGAGTGTTTTCAAAG	Cloning of *LIC10940* gene into pTZ57R/T
LIC10940-R (*Xho*I)	CCG**CTCGAG**CTATCTAATGATTGGCCAAC	

*Restriction sites are indicated in bold and underlined; F: Forward; R: Reverse

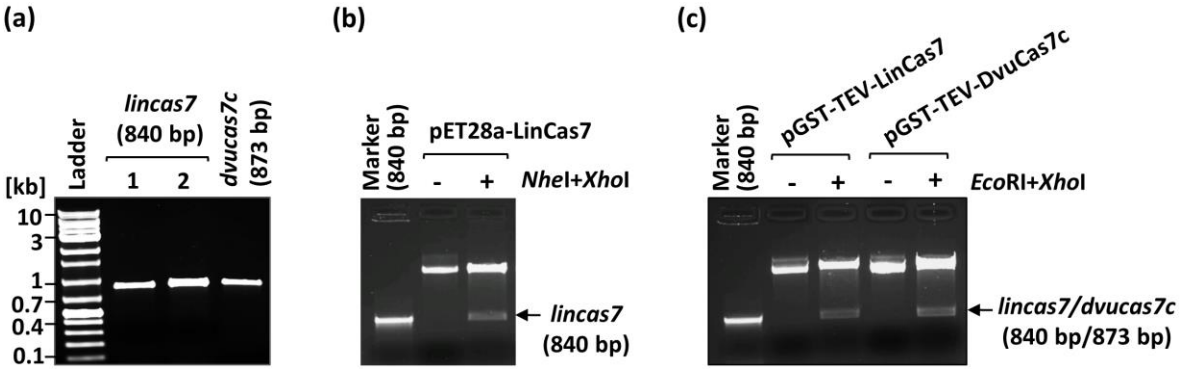

Fig. A1. Cloning of *lincas7* and *dvucas7c* into expression vectors. (a) PCR amplicons of *lincas7* (840 bp) and *dvucas7c* (873 bp) gene. The *lincas7* gene was PCR amplified from genomic DNA of *L. interrogans* serovar Copenhageni using gene-specific primers designed for cloning into (1) pET28a and (2) pGST-TEV vectors. The *dvucas7c* gene of *D. vulgaris* was PCR amplified from a commercially available plasmid construct *D. vulgaris* Cas7/pHMGWA. **(b)** Double restriction digestion (*Nhe*I and *Xho*I) of plasmid pET28a-LinCas7 to confirm the cloning of *lincas7*. **(c)** Double restriction digestion (*Eco*RI and *Xho*I) of plasmid pET28a-LinCas7 and pGST-DvuCas7c to confirm the cloning of *lincas7* and *dvucas7c*, respectively. In all panels, DNA samples have been resolved on 1 % agarose gel stained with EtBr.

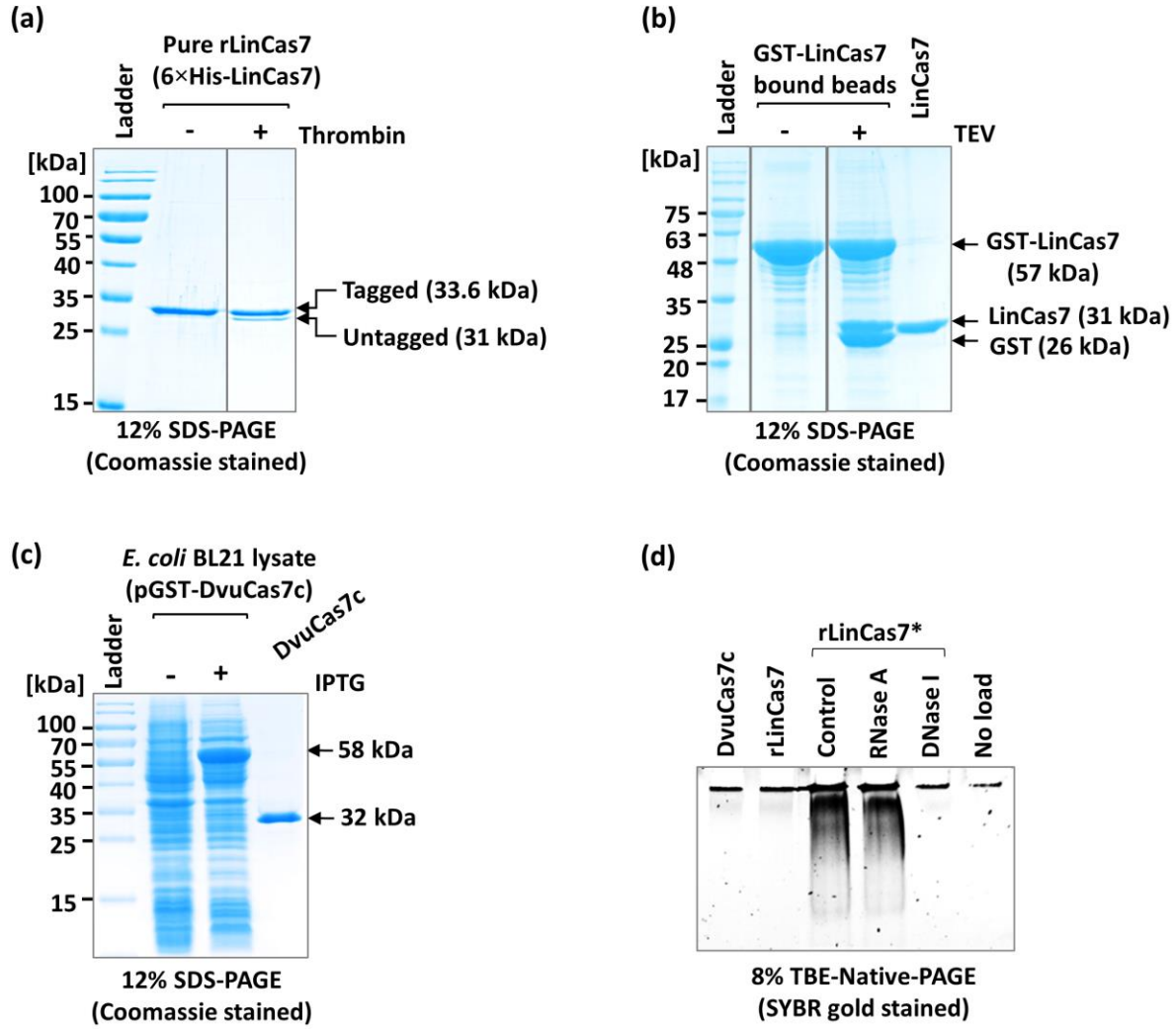

Fig. A2. Purification of untagged recombinant Cas7 proteins. (a) Untagged LinCas7 (31 kDa) preparation from 6×His-LinCas7 (rLinCas7, 33.6 kDa). SDS-PAGE analysis of the Ni-NTA affinity purified rLinCas7 (1 µg) post-treatment with (+) and without (-) thrombin enzyme (0.01 U) for 24 h at 4 ºC. Efficient cleavage of 6×His tag could not be achieved. **(b)** Untagged LinCas7 preparation from GST-LinCas7 (57 kDa). The GST-LinCas7 protein bound to Glutathione-agarose beads was treated with TEV protease (1 U TEV protease per 15 µg of GST-LinCas7) for 16 h at 4 ºC followed by collection of untagged LinCas7 in the elute. Post treatment, the TEV treated (+) and untreated (-) samples along with the collected elute containing untagged LinCas7 (31 kDa) were analysed by SDS-PAGE. **(c)** Overexpression and purification of recombinant Cas7c of *D. vulgaris* (DvuCas7c). The SDS-PAGE analysis of whole-cell lysates of *E. coli* BL21 cells (pGST-DvuCas7c) overexpressing GST-DvuCas7c (58 kDa) post-induction with (+) and without (-) IPTG (1 mM) and the purified untagged DvuCas7c (32 kDa). **(d)** TBE-Native-PAGE (SYBR gold stained) analysis of purified Cas7 proteins. The purified rLinCas7 lot marked with an asterisk (rLinCas7*) indicates the purification was performed using low-salt buffers containing 100-150 mM NaCl. Whereas, rLinCas7 and DvuCas7c were purified using high-salt buffers (300 mM NaCl). On SYBR gold staining, detection near the wells depicts the presence of nucleic acid in rLinCas7* lot. DNase I treatment vanished the nucleic acid content in rLinCas7* suggesting copurification of bound DNA. The well without sample loading (No load) have been shown to demonstrate the artefact detection in wells, if any.

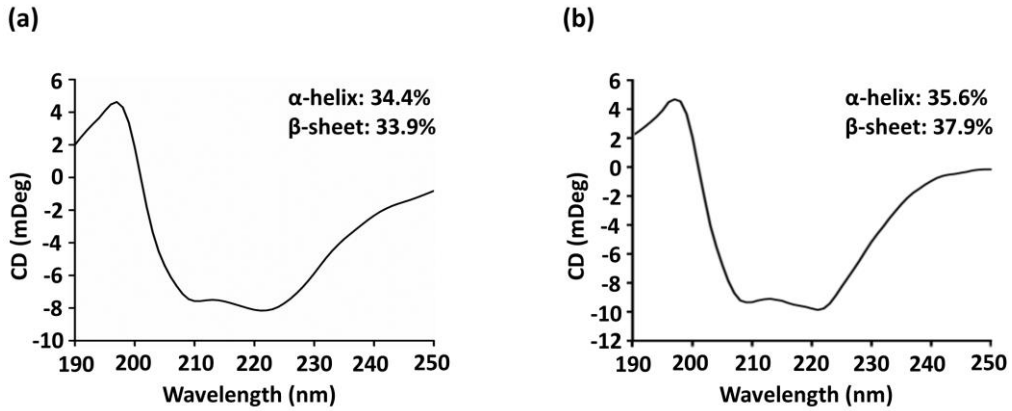

Fig. A3. Circular dichroism (CD) spectroscopy of recombinant LinCas7 protein. CD profile of purified **(a)** 6×His-LinCas7 (rLinCas7, 33.6 kDa) and **(b)** untagged LinCas7 (31 kDa). CD profile was recorded at 190-250 nm wavelength at room temperature. The presented CD spectra are the average of 5 scans after baseline correction. The computed secondary structure composition based on CD spectra have been displayed on top right corner of each spectrogram.

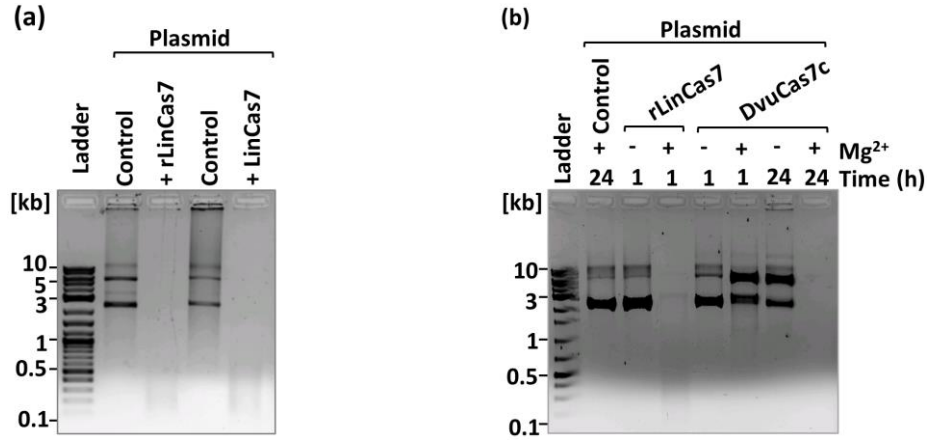

Fig. A4. Comparative assessment of nuclease activity in recombinant Cas7 variants. (a) Nuclease assay of LinCas7 variants (6×His-LinCas7 denoted as rLinCas7 and untagged as LinCas7) on the circular ds-DNA substrate (Plasmid). **(b)** Assessment of DNase activity in rLinCas7 and DvuCas7c. All the DNase assays were performed employing a 3 µM of indicated variants of Cas7 protein (rLinCas7, LinCas7, or DvuCas7c) against ~300 ng of circular ds-DNA substrate (Plasmid) in 1× nuclease activity buffer (20 mM Tris-HCl, 100 mM KCl; pH 8.0) with 10 mM Mg^{2+} ions (if not specified) at 37 ºC for 1 h or as indicated. Post-completion reactions were resolved on 1 % agarose gel stained with EtBr. For clarity, the gel images have been illustrated after inversion. The leftmost lane in all the agarose gels contains DNA ladders. DNase assays illustrated were confirmed by at least two independent experiments.

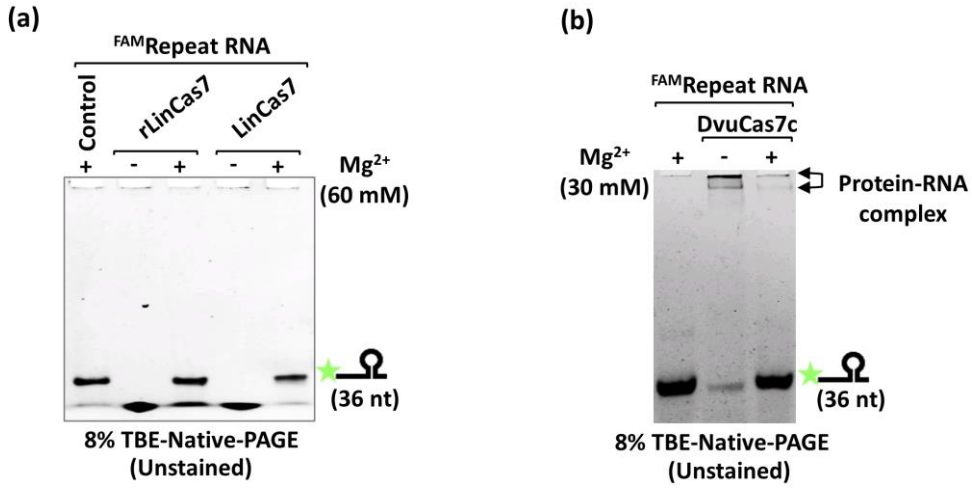

Fig. A5. Activity of Cas7 variants on RNA substrate. (a) EMSA assay of ^{FAM}Repeat RNA (36 nt) incubated with LinCas7 variants (rLinCas7 or LinCas7; 6 μM). **(b)** EMSA assay of ^{FAM}Repeat RNA (36 nt) incubated with DvuCas7c (3 μM). For EMSA assays, 0.5 μM of the RNA substrate (^{FAM}Repeat RNA; 36 nt) was incubated with Cas7 protein variants (rLinCas7, LinCas7, or DvuCas7c) in 1× binding buffer in the presence (+) and absence (-) of Mg^{2+} ions (as specified) for 1 h at 37 °C and resolved on TBE-Native-PAGE (8 %) at 4 °C. The results presented were confirmed by at least two independent experiments. The green star indicates the fluorescent label (FAM) at 5′ end of the used RNA substrate.

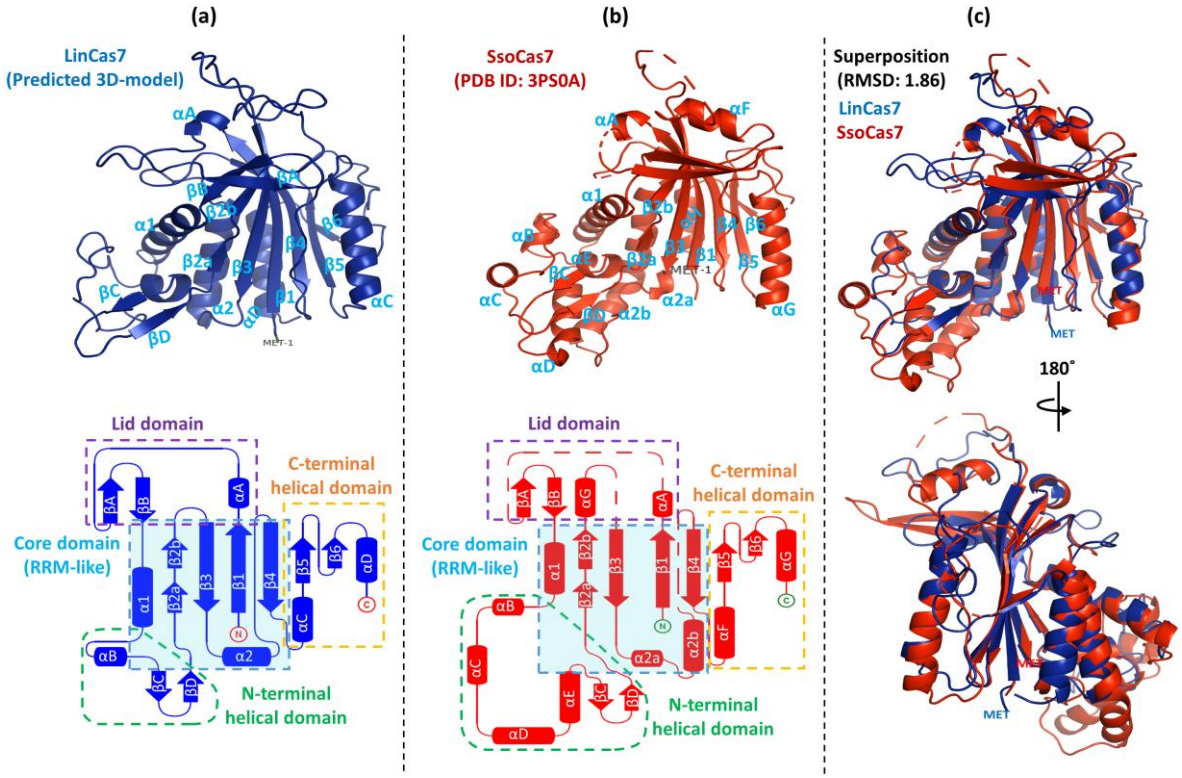

Fig. A6. Structural features of *Leptospira* Cas7 protein. (a) The predicted 3D model of LinCas7 **(top)** and its topology diagram showing the arrangement of auxiliary domains relative to core RRM (RNA recognition motif)-like domain **(bottom)**. **(b)** The 3D model of SsoCas7 (PDB: 3ps0A) **(top)** and its topology diagram **(bottom)** to compare with LinCas7's predicted structure. The topological arrangement of the insertions (lid domain and N-terminal domain) in core RRM-like domain (shaded in cyan) and extension at C-terminal are similar in both proteins. **(c)** Superposition of the predicted LinCas7's structure (blue) on to SsoCas7 (red; PDB: 3ps0A). The two views are related by a 180° rotation as indicated. The LinCas7 shares the structural homology with SsoCas7 with an RMSD 1.86 Å. The highest similarity is evident in RRM-like domain (β-α-β-β-α-β). The major structural difference lies in the N-terminal helical domain. Overall, the relative structural arrangement in LinCas7 is almost the same as in SsoCas7.

****

Appendix-B: Supplementary data to chapter 3

Table B1. Plasmids and primers used in this study

Plasmids	Specificity	Source
pET28a (5369 bp)	Bacterial expression vector with T7lac promoter, N- and C-terminal 6×His tag, thrombin cleavage site; kanamycin resistance	Novagen #69864-3
pGST-TEV (5720 bp)	Bacterial expression vector with T7lac promoter, N-terminal GST tag, TEV cleavage site; ampicillin resistance	BioBharati LifeScience #BB-V0021
pET23a (3666 bp)	Bacterial expression vector with T7 promoter, N-terminal T7 tag, C-terminal 6×His tag; ampicillin resistance	Novagen #69745-3
pCDF-1b (3621 bp)	Bacterial expression vector with T7lac promoter, N-terminal 6×His tag, enterokinase cleavage site, C-terminal S tag; streptomycin/spectinomycin resistance	Novagen #71330-3
Primers	***Sequence (5′-3′)**	**Purpose**
LIC10937-F (*Nhe*I)	CTA**GCTAGC**ATGAGTTCCCGTCCAAGC	Cloning of *LIC10937 (lincas8b)* into pET28a and pET23a
LIC10937-R (*Sal*I)	ACGC**GTCGAC**CTAGGCGTCCTTCGACTTCT	
LIC10937(ΔN)-F (*Nhe*I)	CTA**GCTAGC**ACCATTGCGGATGCGTTTC	Used with LIC10937-R (*Sal*I) for cloning of *lincas8b* ORF partially into pET28a to overexpress rLinCas8b$^{\Delta N}$
LIC10937(ΔC)-R (*Sal*I)	GC**GTCGAC**TTAAAATTTTATCAACTCCGCT CGT	Used with LIC10937-F (*Nhe*I) for cloning of *lincas8b* ORF partially into pET28a to overexpress rLinCas8b$^{\Delta C}$
LIC10937-F (*Bam*HI)	CGC**GGATCC**ATGAGTTCCCGTCCAAGCG	Used with LIC10937-R (*Sal*I) for cloning of *lincas8b* ORF into pGST-TEV
LinCas11b-F (*BamH*I)	CGC**GGATCC**CATGGAAAATAACGAAATCT GGA	Used with LIC10937-R (*Sal*I) for cloning of identified *lincas11b* ORF into pCDF-1b
LIC10937^{M429A} -F	GATAAAATTTGCGGAAAATAACGAAATCTG GAATTC	Mutagenesis of *lincas8b* in pET28a-LinCas8b plasmid construct
LIC10937^{M429A} -R	AACTCCGCTCGTTCGAAT	

*Restriction sites are indicated in bold and underlined; F: Forward; R: Reverse

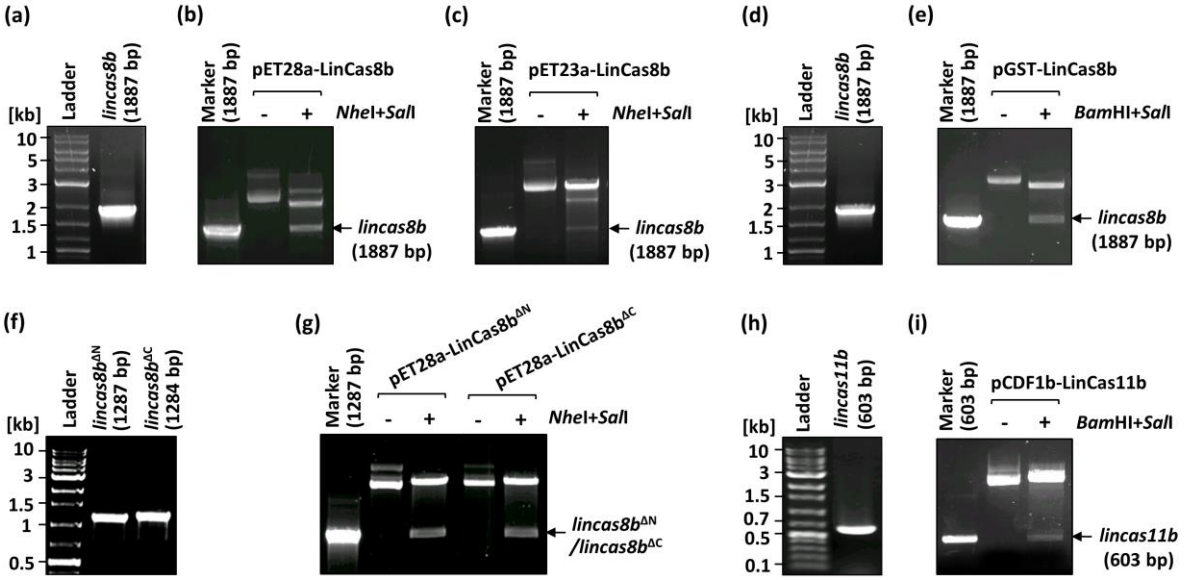

Fig. B1. Cloning of full-length *lincas8b* or its variants into expression vectors. Using sequence-specific primers, target DNA sequences were PCR amplified from the genomic DNA of *L. interrogans* serovar Copenhageni and individually restriction cloned into indicated expression vectors. **(a)** PCR amplification of *lincas8b* for cloning into pET28a and pET23a. **(b) & (c)** Double restriction digestion (*Nhe*I + *Sal*I) of plasmid pET28a-LinCas8b & pET23a-LinCas8b, respectively, to confirm the cloning of *lincas8b*. **(d)** PCR amplification of *lincas8b* for cloning into pGST-TEV. **(e)** Double restriction digestion (*Bam*HI + *Sal*I) of plasmid pGST-LinCas8b to confirm the cloning of *lincas8b*. **(f)** PCR amplification of *lincas8b*'s 1287 bp from 3′-end (*lincas8b*$^{\Delta N}$ encodes LinCas8b$^{\Delta N}$) and 1284 bp from 5′-end (*lincas8b*$^{\Delta C}$ encodes LinCas8b$^{\Delta C}$) for cloning into pET28a. **(g)** Double restriction digestion (*Nhe*I + *Sal*I) of plasmids pET28a-LinCas8b$^{\Delta N}$ and pET28a-LinCas8b$^{\Delta C}$ to confirm the cloning of *lincas8b*$^{\Delta N}$ and *lincas8b*$^{\Delta C}$, respectively. **(h)** PCR amplification of *lincas11b* ORF identified within *lincas8b*'s (603 bp from 3′-end) for cloning into pCDF-1b. **(i)** Double restriction digestion (*Bam*HI + *Sal*I) of plasmids pCDF1b-LinCas11b to confirm the cloning of *lincas11b*. In all panels, DNA samples have been resolved on 1 % agarose gel stained with EtBr.

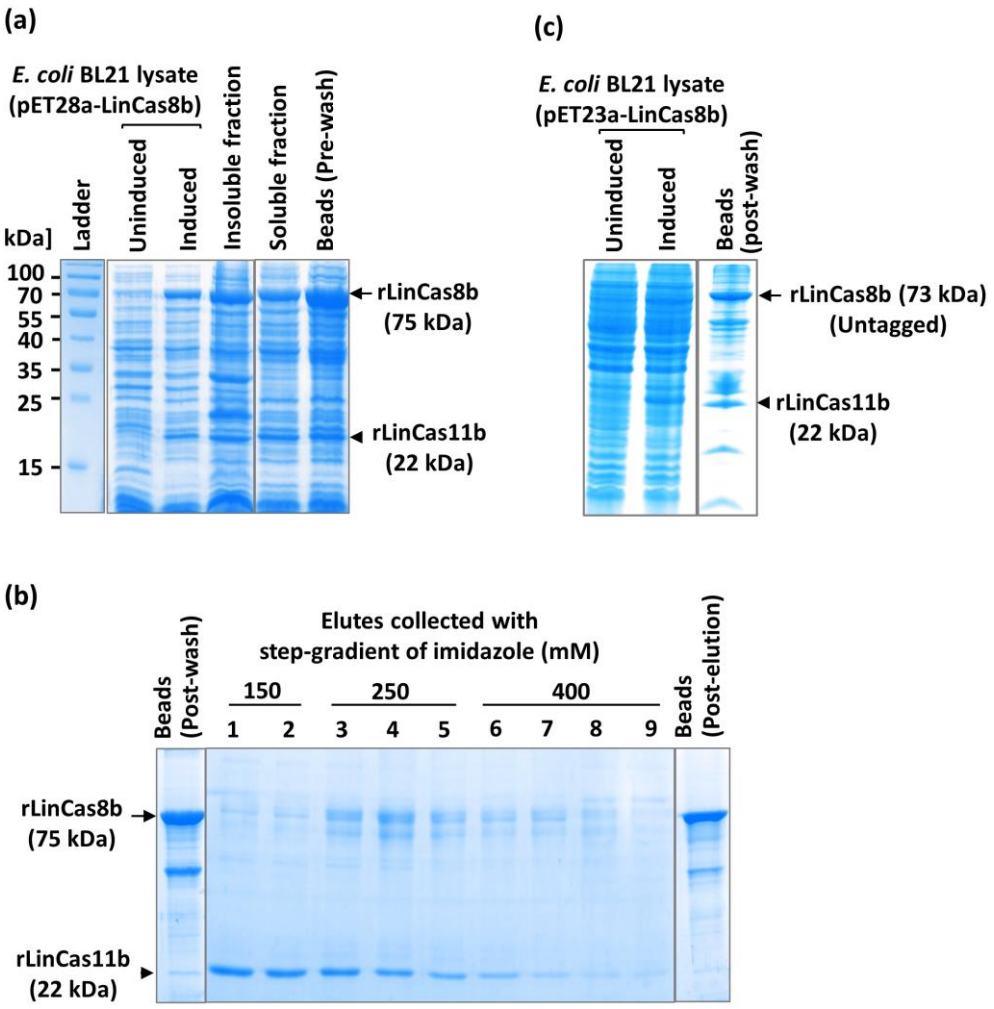

Fig. B2. Overexpression and purification of recombinant LinCas8b via nickel-affinity chromatography. (a) SDS-PAGE analysis of rLinCas8b's purification samples using Ni-NTA-agarose beads. The whole-cell lysates of uninduced and IPTG-induced *E. coli* BL21 (DE3) cells (pET28a-LinCas8b), soluble & insoluble fractions post-lysis of the induced *E. coli* cells and protein-bound beads (pre-washing) were resolved on 10 % SDS-PAGE and visualized by Coomassie staining. The IPTG-induction resulted in the overexpression of rLinCas8b (6×His-LinCas8b,75 kDa) and an additional protein of size 22 kDa that was identified as rLinCas11b in this study later on. **(b)** SDS-PAGE (10 %; Coomassie-stained) analysis of post-washing rLinCas8b-bound beads, elutes collected with step-gradient of imidazole (150-400 mM), and post-elution beads. **(c)** Binding check of rLinCas8b (untagged, 73 kDa) to the Ni-NTA-agarose beads. The rLinCas8 was overexpressed without 6×His tag via pET23a plasmid and incubated with the Ni-NTA-agarose beads. Post-binding and washing, protein-bound beads and whole-cell lysates of uninduced and IPTG-induced *E. coli* BL21 (DE3) cells overexpressing rLinCas8b (untagged) were analyzed by 10 % SDS-PAGE (Coomassie stained).

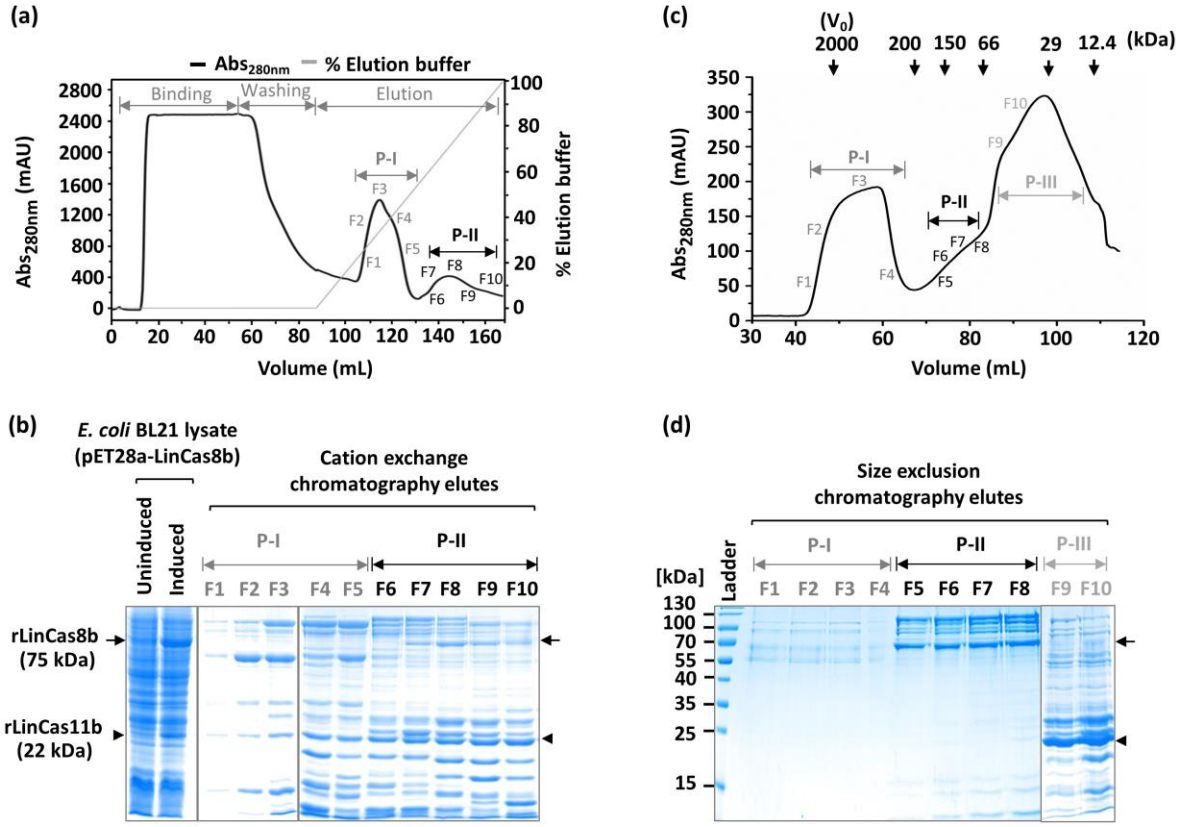

Fig. B3. Purification of rLinCas8b via cation exchange chromatography. (a) The representative chromatogram obtained during cation exchange chromatography (CEX) for rLinCas8b (6×His-LinCas8b,75 kDa) purification from the whole-cell lysate of induced *E. coli* BL21 cells (pET28a-LinCas8b). Purification events are indicated, including lysate binding to the chromatography column (binding), column washing (washing), and elution. The bound-protein(s) eluted in two major peaks (P-I & P-II). The collected fractions from each peak are marked and indicated as F1-F10. **(b)** SDS-PAGE (10 %; Coomassie-stained) analysis of fractions (F1-F10) collected during CEX. As a marker, the whole-cell lysates of uninduced and IPTG-induced *E. coli* BL21 cells (pET28a-LinCas8b) overexpressing rLinCas8b were resolved. **(c)** Size exclusion chromatography (SEC) of CEX fractions (pulled F7-F10) containing rLinCas8b. Proteins resolved as three peaks (P-I, P-II & P-III). The collected fractions from each peak are marked and indicated as F1-F10. The elution profiles of standard molecular weight protein markers are shown with inverted arrows on top of the chromatogram (Blue dextran eluting in the void volume, V_o, ~2000 kDa; β-Amylase, 200 kDa; Alcohol dehydrogenase, 15 kDa; Albumin, 66 kDa; Carbonic anhydrase, 29 kDa; Cytochrome c, 12.4 kDa). **(d)** SDS-PAGE (10 %; Coomassie-stained) analysis of fractions (F1-F10) collected during SEC. P-II contains the rLinCas8b as well as multiple impurities of larger size. In panels (b) & (d), the bands of rLinCas8b and rLinCas11b are marked and indicated with an arrow and arrowhead, respectively.

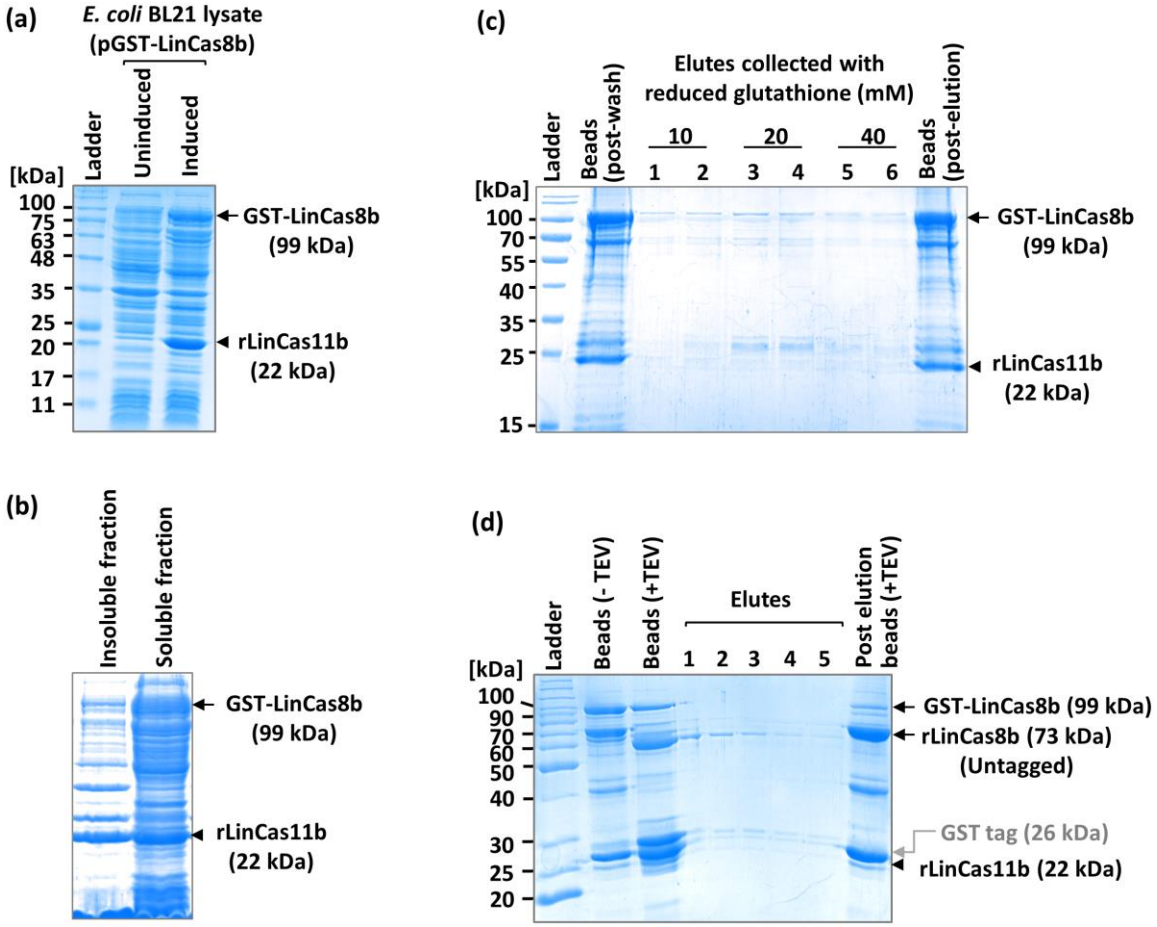

Fig. B4. Overexpression and purification of recombinant LinCas8b via glutathione affinity chromatography. (a) SDS-PAGE (10 %; Coomassie-stained) analysis whole-cell lysates of uninduced and induced (1 mM IPTG) *E. coli* BL21 (DE3) cells housing pGST-LinCas8b plasmid. The overexpression of GST-LinCas8b (99 kDa) and rLinCas11b (22 kDa) are marked. **(b)** The soluble & insoluble fractions of whole-cell lysate of the induced *E. coli* cells overexpressing GST-LinCas8b are analyzed by 10 % SDS-PAGE (Coomassie-stained). **(c)** SDS-PAGE (10 %; Coomassie-stained) analysis of post-washing GST-LinCas8b bound glutathione (GSH)-agarose beads, elutes collected with step-gradient of reduced glutathione (10-40 mM), and post-elution beads. **(d)** In-bead TEV protease cleavage of GST-LinCas8b. The GST-LinCas8b protein bound to glutathione-agarose beads was treated with TEV protease (1 U TEV protease per 15 µg of GST-LinCas7) for 16 h at 4 ºC followed by the collection of untagged rLinCas8b (73 kDa) in the elute. Post-treatment, the TEV treated (+) and untreated (-) samples, along with the post-elution beads and collected elutes containing untagged rLinCas8b (73 kDa), were analyzed by SDS-PAGE (10 %; Coomassie-stained). Post-cleavage, the untagged rLinCas8b remains bound to beads and does not elute efficiently.

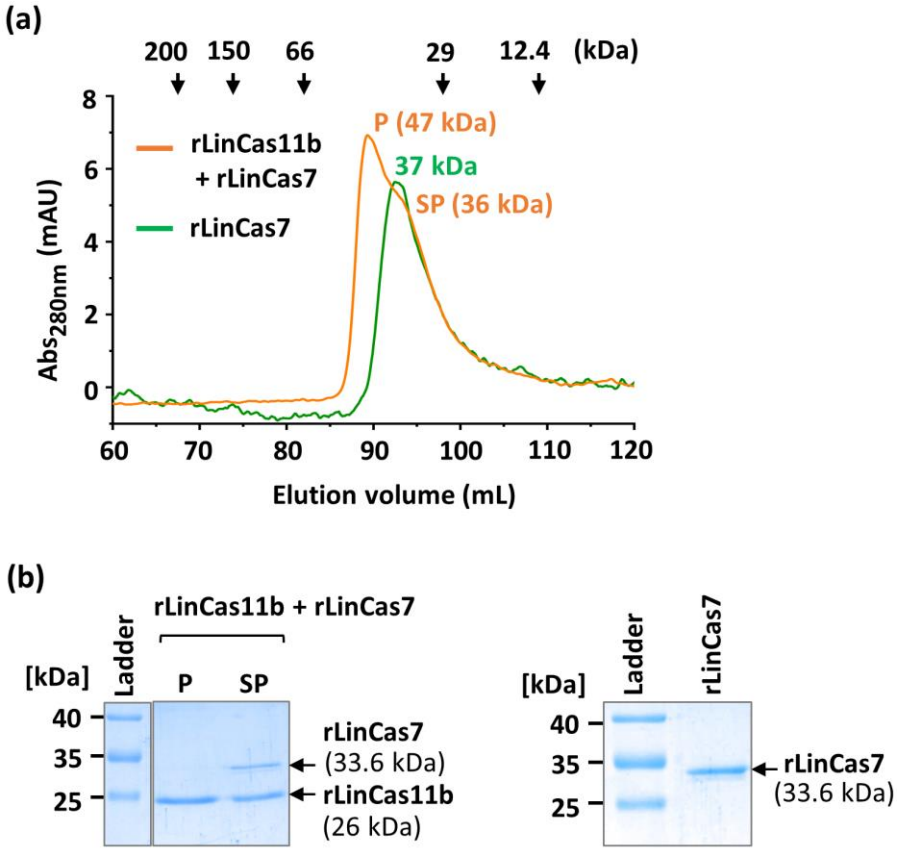

Fig. B5. Interaction check between rLinCas11b and rLinCas7 by SEC. (a) Size exclusion chromatography (SEC). Pure rLinCas7 (500 µg) alone (green) or in combination with rLinCas11b (500 µg each; orange) in a total volume of 1 mL (in storage buffer) was incubated at 37 °C for 1 h, followed by SEC. **(b)** SDS-PAGE (10 %; Coomassie-stained) of fractions (**left:** P & SP; **right:** rLinCas7) collected during SEC. Analysis revealed no stable interaction between rLinCass11b and rLinCas7. In both conditions, alone (green) or with rLinCas11b (orange, as shoulder peak SP), rLinCas7 eluted at a size of ~37 kDa close to its momomeric size (33.6 kDa). While, rLinCas11b resolved near its dimeric size ~47 kDa (orange, P) and as a shoulder peak (SP) also due to overlapping of elution profiles. On top of the chromatogram, the elution profiles of standard molecular weight protein markers are shown with inverted arrows (β-Amylase, 200 kDa; Alcohol dehydrogenase, 15 kDa; Albumin, 66 kDa; Carbonic anhydrase, 29 kDa; Cytochrome c, 12.4 kDa).

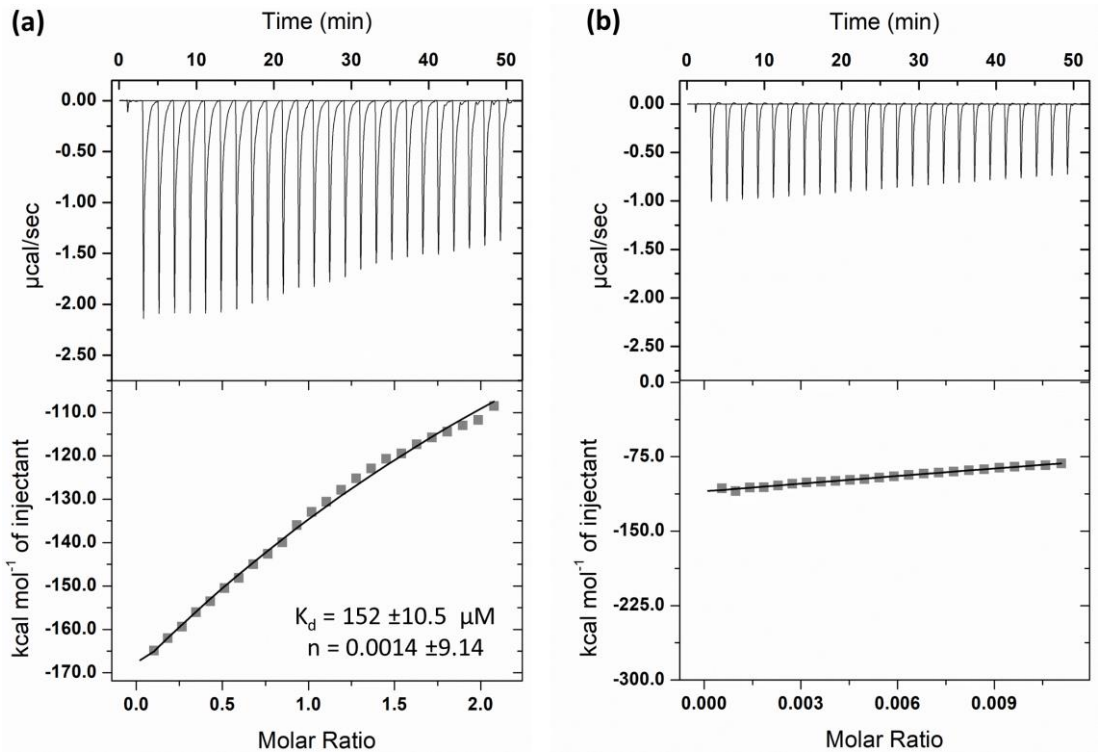

Fig. B6. Interaction check between rLinCas11b and rLinCas7 by ITC. Representative isothermal titration calorimetry (ITC) thermogram obtained from **(a)** titration of rLinCas7 with rLinCas11b (control subtracted) and **(b)** titration of storage buffer with rLinCas11b (control). Though saturation of binding could not be attained in multiple ITC trials. With the tested conditions (vide methods), the determined apparent dissociation constant is (K_d) = 152 ±10.5 µM, demonstrating a subtle binding affinity between rLinCas11b and rLinCas7. Other parameters determined from ITC experiments includes the stoichiometry factor (n) = 0.0014 ±9.14; enthalpy (ΔH) = $-2.28 \times 10^8 \pm 1.5 \times 10^{11}$ cal/mol; entropy (ΔS) = -7.53×10^6 cal/mol/deg.

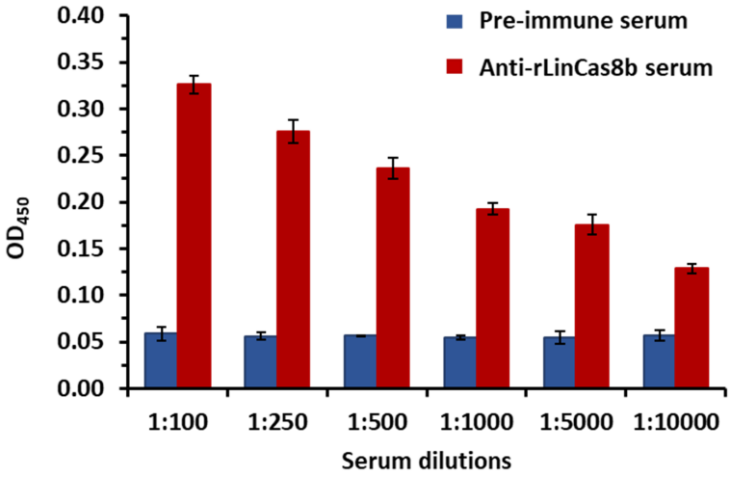

Fig. B7. Titer determination of anti-rLinCas8b polyclonal antibody in mice serum by ELISA. The mice immune sera obtained after seven days of second booster dose were used at different dilutions to determine the titer of anti-rLinCas8b polyclonal antibody through ELISA. Serum obtained from unimmunized mice (pre-immune serum) was used as a control for evaluation of antibody titer, and data is represented as mean ± standard error of two independent experiments.

*****

Appendix-C: Supplementary data to chapter 4

Table C1. Plasmids and primers used in this study

Plasmids	Specificity	Source
pACYCDuet-1 (4008 bp)	A bacterial expression vector with two multiple cloning sites (MCS), each preceded by a T7 lac promoter for coexpression of two genes, adds an N-terminal 6×His tag to insert cloned at MCS-I, chloramphenicol resistance	Novagen #71147-3
pRSF-1b (3669 bp)	Bacterial expression vector with T7 lac promoter, N-terminal 6×His tag, C-terminal S tag, enterokinase cleavage site; Kanamycin resistance	Novagen #71363-3
pT7Blue (2887 bp)	Bacterial cloning vector with T7 lac promoter; ampicillin resistance as a selection marker; allow blue/white screening	Sigma-Aldrich #69967
pET28a-LinCas7 (6142 bp)	pET28a plasmid inserted with *lincas7* gene to overexpress N-terminal 6×His tagged LinCas7 (rLinCas7) protein; kanamycin resistance	Laboratory stock (Hussain and Kumar, 2022)
pET28a-LinCas8b (7210 bp)	pET28a plasmid inserted with *lincas8b* gene to overexpress N-terminal 6×His tagged LinCas8b (rLinCas7b) protein; kanamycin resistance	Laboratory stock (Hussain et al., 2023)
pCDF1b-LinCas11b (4191 bp)	pCDF-1b plasmid inserted with *lincas11b ORF* to overexpress N-terminal 6×His tagged LinCas11b (rLinCas11b) protein; streptomycin/spectinomycin resistance	Laboratory stock (Hussain et al., 2023)
pET-SUMO-LinCas5 (6267 bp)	pET SUMO plasmid inserted with *lincas5* gene to overexpress N-terminal 6×His-SUMO tagged LinCas5 (rLinCas5) protein; kanamycin resistance	Laboratory stock
pET-SUMO-LinCas3 (7896 bp)	pET SUMO plasmid inserted with *lincas3* gene to overexpress N-terminal 6×His-SUMO tagged LinCas3 (rLinCas3) protein; kanamycin resistance	Laboratory stock

Primers	*Sequence (5′-3′)	Purpose
LIC10937-F (*Sal*I)	GC**GTCGAC**ATGAGTTCCCGTCCAAGC	Cloning of *lincas8b*, *lincas7*, and *lincas5* in pACYCDuet-1
LIC10935-R (*Not*I)	ATAAGAAT**GCGGCCGC**TTAAGAGGGTCG GAGTTTAA	
LIC10939-F (*Kpn*I)	GG**GGTACC**ATGTCCATTCCGAACGTCA	Cloning of *lincas6* and *lincas3* in pACYCDuet-1
LIC10938-R (*Pac*I)	G**TTAATTAA**TCAATTCCATTCTCCGCC	
LIC10937^{M429A} -F	GATAAAATTTGCGGAAAATAACGAAATCT GGAATTC	Mutagenesis of *LIC10937* (*lincas8b*) in pCas plasmid constructs
LIC10937^{M429A} -R	AACTCCGCTCGTTCGAAT	
LIC_Cr^{2} R1-F (*Kpn*I)	CGG**GGTACC**CTGAATATAACTTTGATGCC GTTAGG	Cloning of LIC_Cr^{2} array into pRSF-1b
LIC_Cr^{2} R4-R (*Hind*III)	CCC**AAGCTT**TTCTAAACCGCCTATCGGC	

*Restriction sites are indicated in bold and underlined; F: Forward; R: Reverse

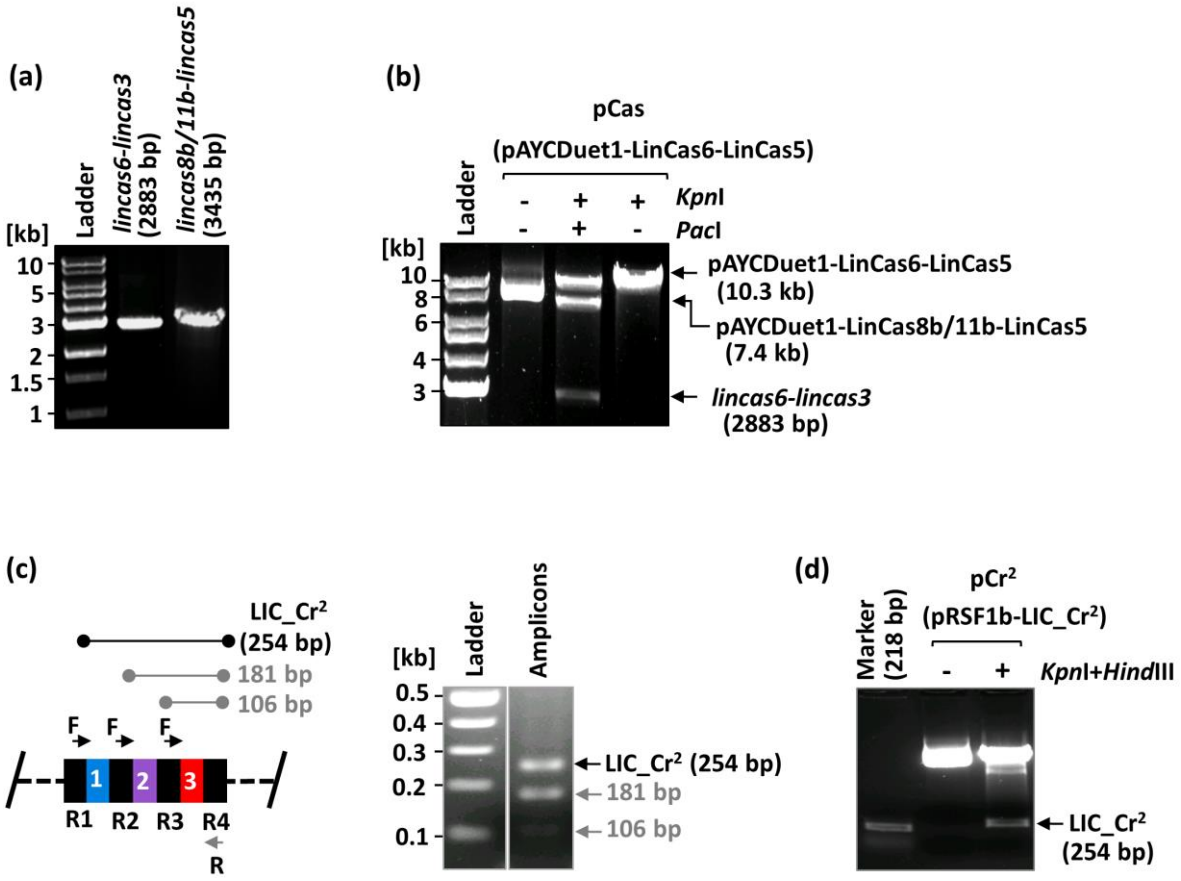

Fig. C1. Cloning of Lin_I-B interference machinery of *L. interrogans*. In *L. interrogans* serovar Copenhageni, the interference-associated *cas* genes of CRISPR-Cas subtype I-B (Lin_I-B) system are organized into a single operon (Operon-II) in order of *lincas6-lincas3-lincas8b/11b-lincas7-lincas5*. Using sequence-specific primers, DNA sequences of interest were PCR amplified from the genomic DNA of *L. interrogans* serovar Copenhageni and individually cloned into indicated expression vectors. **(a)** PCR amplification of *lincas6-lincas3* (2883 bp) and *lincas8b/11b-lincas5* (3435 bp) DNA segments. **(b)** The PCR amplified *lincas8b-lincas5* (3435 bp), and *lincas6-lincas3* (2883 bp*)* were restriction cloned into pACYDuet-1 expression plasmid at MCS-I and MCS-II, respectively. The prepared plasmid construct (pAYCDuet1-LinCas6-LinCas5, 10.3 kb; pCas) was double (*Kpn*I + *Pac*I) or single (*Kpn*I) restriction digested to confirm the cloning. Double restriction digestion released the *lincas6-lincas3* insert cloned at MCS-II. **(c)** The CRISPR array (LIC_Cr^2) associated with Lin_I-B loci consists of four repeats (black rectangles: R1, R2, R3, and R4) and three spacers (colored diamonds: 1, 2, and 3). Repeats R1, R2, and R3 are identical in sequence. Thus, as indicated in the left panel, PCR for full-length LIC_Cr^2 (254 bp) amplification gives two additional amplicons (181 bp and 106 bp) (right panel). The distant wells of the same gel have been cropped and re-aligned for clarity. **(d)** Full-length amplicon (LIC_Cr^2; 254 bp) was restriction cloned into pRSF-1b vector. The resulting plasmid construct (pRSF1b-LIC_Cr^2; pCr^2) was double restriction digested (*Kpn*I + *Hind*III) to confirm the cloning. In all panels, DNA samples are analyzed on 1.5 % agarose gel stained with EtBr.

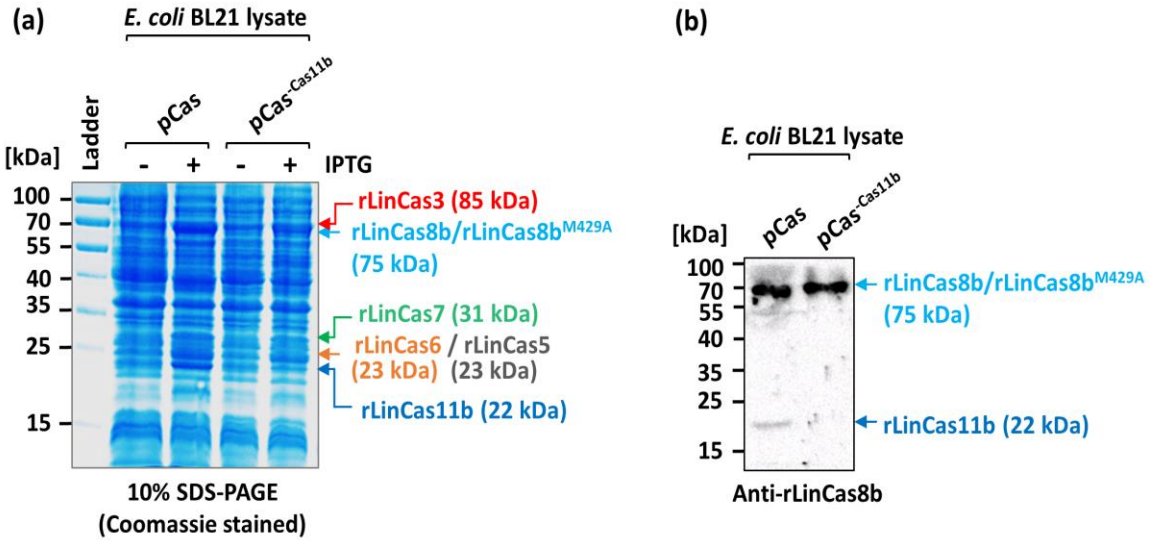

Fig. C2. Site-directed mutation inactivated the co-expression of rLinCas11b encoded within pCas plasmid. (a) SDS-PAGE analysis of the whole-cell lysates of uninduced (-) and IPTG-induced (+) *E. coli* BL21 (DE3) cells transformed with pCas or mutant pCas$^{-Cas11b}$ plasmid. Overexpression of rLinCas proteins at their expected size are indicated. **(b)** Immunoblot, using anti-rLinCas8b antibody, of induced whole-cell lysates of *E. coli* BL21 cells overexpressing rLinCas8b or rLinCas8b^{M429A} via plasmid pCas or pCas$^{-Cas11b}$, respectively. A single amino acid substitution mutation (Met429 to Ala; rLinCas8b^{M429A}) in *lincas8b* encoded within the plasmid pCas (pCas$^{-Cas11b}$) resulted in the abolition of the rLinCas11b co-expression.

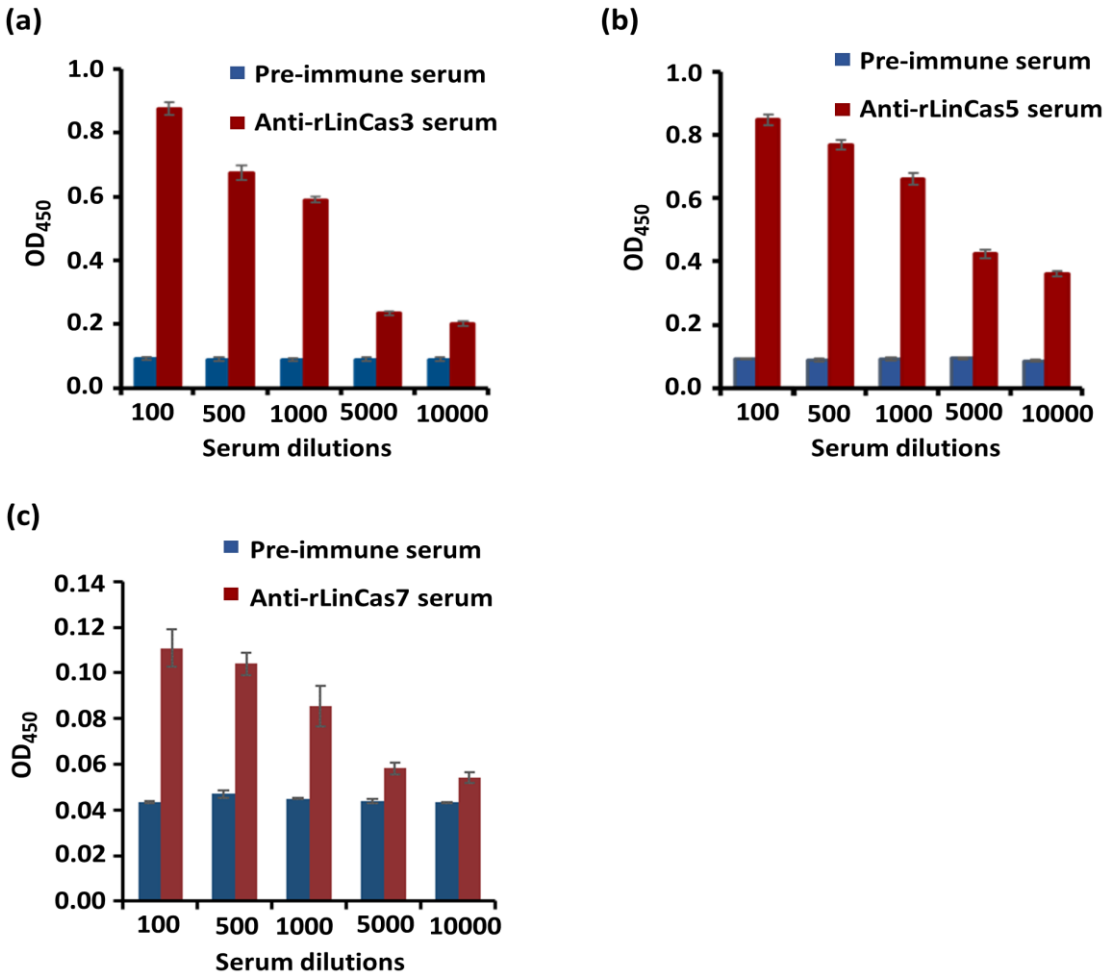

Fig. C3. Titer determination of anti-rLinCas protein polyclonal antibody in mice serum by ELISA. The mice immune sera obtained after seven days of second booster dose were used at different dilutions to determine the titer of **(a)** anti-rLinCas3, **(b)** anti-rLinCas5, and **(c)** anti-rLinCas7 polyclonal antibodies through ELISA. Serum obtained from unimmunized mice (pre-immune serum) was used as a control for evaluation of respective antibody titer, and data is represented as mean ± standard error of two independent experiments.

*****

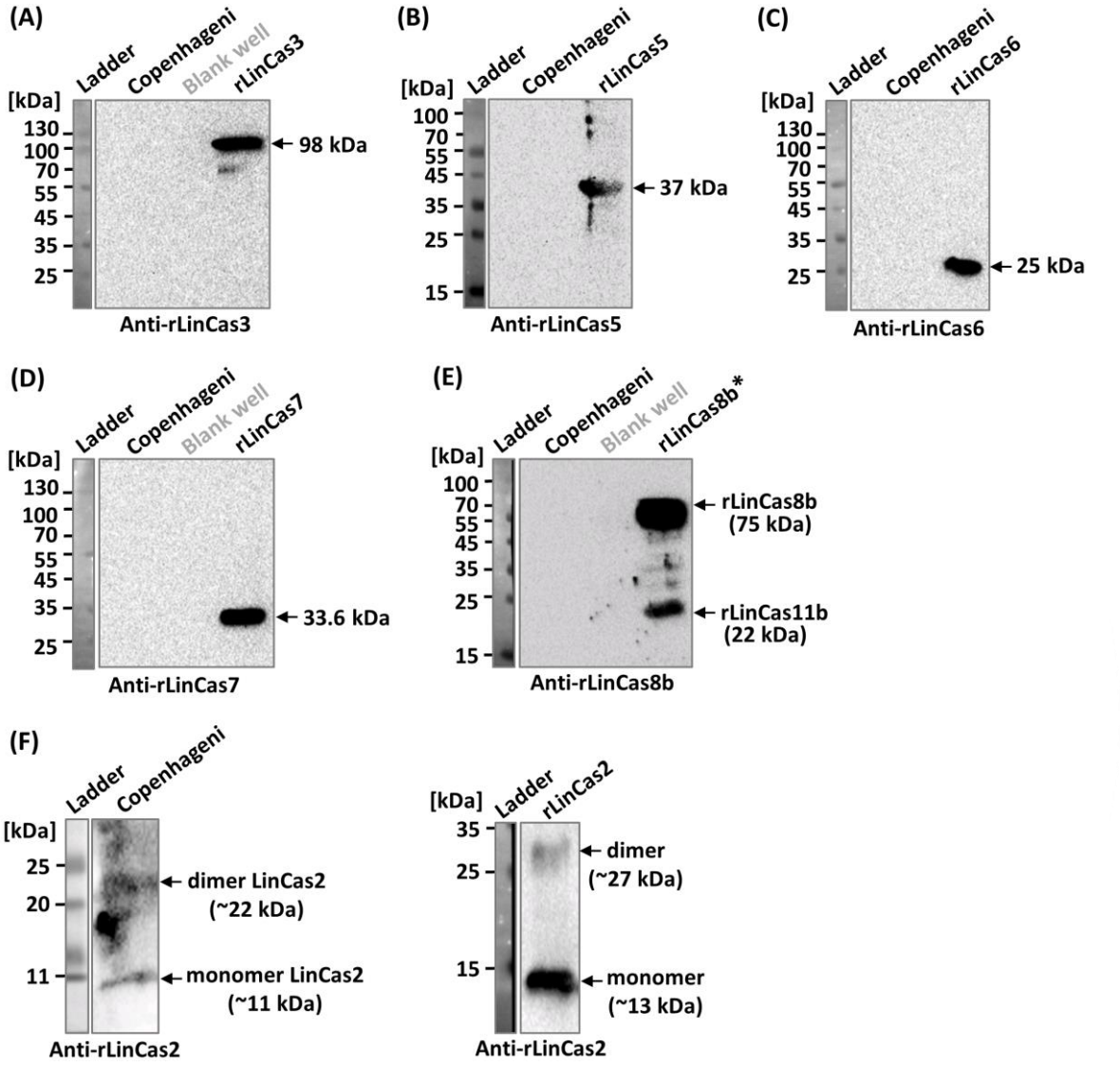

Fig. D1. Immunoblot of whole-cell lysate of *L. interrogans*. To detect the Cas proteins of Lin_I-B locus in its native host *L. interrogans* serovar Copenhageni strain Fiocruz L1-130, immunoblotting was performed (vide methods). The whole-cell lysate of *Leptospira* was probed with rLinCas protein-specific polyclonal primary antibodies (indicated at the bottom of each blot). As a positive control, pure rLinCas proteins were also included. Wells without sample load (in blot A, D, and E) have been marked as blank wells. Asterisk mark (*; in blot E) indicates that in place of pure rLinCas8b, whole-cell lysate of IPTG-induced *E. coli* BL21 (DE3) strain overexpressing rLinCas8b has been used. Panel (F) shows two blots developed independently for *Leptospira* lysate (left) and pure rLinCas2 (right).

*****

References

Aakre, C.D., Phung, T.N., Huang, D., Laub, M.T., 2013. A Bacterial Toxin Inhibits DNA Replication Elongation through a Direct Interaction with the β Sliding Clamp. Mol. Cell 52, 617–628. https://doi.org/10.1016/j.molcel.2013.10.014

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E.S., Koonin, E. V., Zhang, F., 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science (80-.). 353, aaf5573. https://doi.org/10.1126/science.aaf5573

Adler, B., de la Peña Moctezuma, A., 2010. Leptospira and leptospirosis. Vet. Microbiol. 140, 287–296. https://doi.org/10.1016/j.vetmic.2009.03.012

Ahmed, N., Devi, S.M., Valverde, M. de los A., Vijayachari, P., Machang'u, R.S., Ellis, W.A., Hartskeerl, R.A., 2006. Multilocus sequence typing method for identification and genotypic classification of pathogenic Leptospira species. Ann. Clin. Microbiol. Antimicrob. 5, 28. https://doi.org/10.1186/1476-0711-5-28

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 Res. 47, 5223–5230. https://doi.org/10.1093/nar/gkz217

Aquino-Jarquin, G., 2019. CRISPR-Cas14 is now part of the artillery for gene editing and molecular diagnostic. Nanomedicine Nanotechnology, Biol. Med. 18, 428–431. https://doi.org/10.1016/j.nano.2019.03.006

Araújo, A.M., Noronha-Dutra, A., Reis, E.A.G., Couto, N.S., Reis, M.G., Ko, A.I., Araujo, G.C., Ribeiro, G.S., Damião, A.O., Athanazio, D.A., Hagan, J.E., 2014. Oxidative Stress Markers Correlate with Renal Dysfunction and Thrombocytopenia in Severe Leptospirosis. Am. J. Trop. Med. Hyg. 90, 719–723. https://doi.org/10.4269/ajtmh.13-0667

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., Phanse, S., Joachimiak, A., Koonin, E. V., Savchenko, A., Emili, A., Greenblatt, J., Edwards, A.M., Yakunin, A.F., 2011. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol. Microbiol. 79, 484–502. https://doi.org/10.1111/j.1365-2958.2010.07465.x

Barazzone, G.C., Teixeira, A.F., Azevedo, B.O.P., Damiano, D.K., Oliveira, M.P., Nascimento, A.L.T.O., Lopes, A.P.Y., 2022. Revisiting the Development of Vaccines Against Pathogenic Leptospira: Innovative Approaches, Present Challenges, and Future Perspectives. Front. Immunol. 12. https://doi.org/10.3389/fimmu.2021.760291

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 (80-.). 315, 1709–1712. https://doi.org/10.1126/science.1138140

Barrangou, R., Horvath, P., 2017. A decade of discovery: CRISPR functions and applications. Nat. Microbiol. 2, 17092. https://doi.org/10.1038/nmicrobiol.2017.92

Bashiru, G., Bahaman, A., 2018. Advances & challenges in leptospiral vaccine development. Indian J. Med. Res. 147, 15. https://doi.org/10.4103/ijmr.IJMR_1022_16

Behler, J., Hess, W.R., 2020. Approaches to study CRISPR RNA biogenesis and the key players involved. Methods 172, 12–26. https://doi.org/10.1016/j.ymeth.2019.07.015

Belkebir, A., Azeddoug, H., 2013. Metal ion dependence of DNA cleavage by SepMI and EhoI restriction endonucleases. Microbiol. Res. 168, 99–105. https://doi.org/10.1016/j.micres.2012.08.003

Beloglazova, N., Kuznedelov, K., Flick, R., Datsenko, K.A., Brown, G., Popovic, A., Lemak, S., Semenova, E., Severinov, K., Yakunin, A.F., 2015. CRISPR RNA binding and DNA target recognition by purified Cascade complexes from Escherichia coli. Nucleic Acids Res. 43, 530–543. https://doi.org/10.1093/nar/gku1285

Benda, C., Ebert, J., Scheltema, R.A., Schiller, H.B., Baumgärtner, M., Bonneau, F., Mann, M., Conti,

E., 2014. Structural Model of a CRISPR RNA-Silencing Complex Reveals the RNA-Target Cleavage Activity in Cmr4. Mol. Cell 56, 43–54. https://doi.org/10.1016/j.molcel.2014.09.002

Bertozzi Silva, J., Storms, Z., Sauvageau, D., 2016. Host receptors for bacteriophage adsorption. FEMS Microbiol. Lett. 363, fnw002. https://doi.org/10.1093/femsle/fnw002

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., Vinetz, J.M., 2003. Leptospirosis: a zoonotic disease of global importance. Lancet Infect. Dis. 3, 757–771. https://doi.org/10.1016/S1473-3099(03)00830-2

Biggs, H.M., Hertz, J.T., Munishi, O.M., Galloway, R.L., Marks, F., Saganda, W., Maro, V.P., Crump, J.A., 2013. Estimating Leptospirosis Incidence Using Hospital-Based Surveillance and a Population-Based Health Care Utilization Survey in Tanzania. PLoS Negl. Trop. Dis. 7, e2589. https://doi.org/10.1371/journal.pntd.0002589

Biswas, A., Gagnon, J.N., Brouns, S.J.J., Fineran, P.C., Brown, C.M., 2013. CRISPRTarget. RNA Biol. 10, 817–827. https://doi.org/10.4161/rna.24046

Bohmert, K., 1998. AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J. 17, 170–180. https://doi.org/10.1093/emboj/17.1.170

Bolotin, A., Quinquis, B., Sorokin, A., Ehrlich, S.D., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551–2561. https://doi.org/10.1099/mic.0.28048-0

Bougie, I., Charpentier, S., Bisaillon, M., 2003. Characterization of the Metal Ion Binding Properties of the Hepatitis C Virus RNA Polymerase. J. Biol. Chem. 278, 3868–3875. https://doi.org/10.1074/jbc.M209785200

Brendel, J., Stoll, B., Lange, S.J., Sharma, K., Lenz, C., Stachler, A.-E., Maier, L.-K., Richter, H., Nickel, L., Schmitz, R.A., Randau, L., Allers, T., Urlaub, H., Backofen, R., Marchfelder, 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. J. Biol. Chem. 289, 7164–7177. https://doi.org/10.1074/jbc.M113.508184

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E. V., van der Oost, J., 2008. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. Science (80-.). 321, 960–964. https://doi.org/10.1126/science.1159689

Brüssow, H., Hendrix, R.W., 2002. Phage Genomics. Cell 108, 13–16. https://doi.org/10.1016/S0092-8674(01)00637-7

Buchan, D.W.A., Jones, D.T., 2019. The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res. 47, W402–W407. https://doi.org/10.1093/nar/gkz297

Bulach, D.M., Zuerner, R.L., Wilson, P., Seemann, T., McGrath, A., Cullen, P.A., Davis, J., Johnson, M., Kuczek, E., Alt, D.P., Peterson-Burch, B., Coppel, R.L., Rood, J.I., Davies, J.K., Adler, B., 2006. Genome reduction in Leptospira borgpetersenii reflects limited transmission potential. Proc. Natl. Acad. Sci. 103, 14560–14565. https://doi.org/10.1073/pnas.0603979103

Campbell, J.A., Cianciotto, N.P., 2022. Legionella pneumophila Cas2 Promotes the Expression of Small Heat Shock Protein C2 That Is Required for Thermal Tolerance and Optimal Intracellular Infection. Infect. Immun. 90. https://doi.org/10.1128/iai.00369-22

Cass, S.D.B., Haas, K.A., Stoll, B., Alkhnbashi, O.S., Sharma, K., Urlaub, H., Backofen, R., Marchfelder, A., Bolt, E.L., 2015a. The role of Cas8 in type I CRISPR interference. Biosci. Rep. 35. https://doi.org/10.1042/BSR20150043

Cass, S.D.B., Haas, K.A., Stoll, B., Alkhnbashi, O.S., Sharma, K., Urlaub, H., Backofen, R., Marchfelder, A., Bolt, E.L., 2015b. The role of Cas8 in type I CRISPR interference. Biosci. Rep. 35. https://doi.org/10.1042/BSR20150043

Chaudhry, R., Premlatha, M.M., Mohanty, S., Dhawan, B., Singh, K.K., Dey, A.., 2002. Emerging Leptospirosis, North India. Emerg. Infect. Dis. 8, 1526–1527. https://doi.org/10.3201/eid0812.020052

Chen, J.S., Ma, E., Harrington, L.B., Da Costa, M., Tian, X., Palefsky, J.M., Doudna, J.A., 2018. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science (80-.). 360, 436–439. https://doi.org/10.1126/science.aar6245

Cheng, F., Gong, L., Zhao, D., Yang, H., Zhou, J., Li, M., Xiang, H., 2017. Harnessing the native type

I-B CRISPR-Cas for genome editing in a polyploid archaeon. J. Genet. Genomics 44, 541–548. https://doi.org/10.1016/j.jgg.2017.09.010

Chhetry, S., Anand, B., 2022. Csb1 moonlighting gives rise to functional redundancy with Csb2 in processing the pre-CRISPR transcript in type I-G CRISPR-Cas system. bioRxiv 2022.08.19.504415.

Chopin, M.-C., Chopin, A., Bidnenko, E., 2005. Phage abortive infection in lactococci: variations on a theme. Curr. Opin. Microbiol. 8, 473–479. https://doi.org/10.1016/j.mib.2005.06.006

Chyou, T., Brown, C.M., 2019. Prediction and diversity of tracrRNAs from type II CRISPR-Cas systems. RNA Biol. 16, 423–434. https://doi.org/10.1080/15476286.2018.1498281

Clokie, M.R.J., Millard, A.D., Letarov, A. V., Heaphy, S., 2011. Phages in nature. Bacteriophage 1, 31–45. https://doi.org/10.4161/bact.1.1.14942

Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G., Sorek, R., 2019. Cyclic GMP–AMP signalling protects bacteria against viral infection. Nature 574, 691–695. https://doi.org/10.1038/s41586-019-1605-5

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 Negl. Trop. Dis. 9, e0003898. https://doi.org/10.1371/journal.pntd.0003898

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., Zhang, F., 2017. RNA editing with CRISPR-Cas13. Science (80-.). 358, 1019–1027. https://doi.org/10.1126/science.aaq0180

Crooks, G.E., 2004. WebLogo: A Sequence Logo Generator. Genome Res. 14, 1188–1190. https://doi.org/10.1101/gr.849004

Cumby, N., Edwards, A.M., Davidson, A.R., Maxwell, K.L., 2012. The Bacteriophage HK97 gp15 Moron Element Encodes a Novel Superinfection Exclusion Protein. J. Bacteriol. 194, 5012–5019. https://doi.org/10.1128/JB.00843-12

Cwick, J.P., Owen, J.E., Kochetkova, I., Hain, K.S., Horssen, N. Van, Taylor, M.P., 2022. Superinfection Exclusion of Alphaherpesviruses Interferes with Virion Trafficking. Microbiol. Spectr. 10. https://doi.org/10.1128/spectrum.00684-22

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., Semenova, E., 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. 3, 945. https://doi.org/10.1038/ncomms1937

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607. https://doi.org/10.1038/nature09886

Dhara, A., Hussain, M.S., Datta, D., Kumar, M., 2019. Insights to the Assembly of a Functionally Active Leptospiral ClpP1P2 Protease Complex along with Its ATPase Chaperone ClpX. ACS Omega 4, 12880–12895. https://doi.org/10.1021/acsomega.9b00399

Dhara, A., Hussain, M.S., Kanaujia, S.P., Kumar, M., 2021. Acyldepsipeptide activated ClpP1P2 macromolecule of Leptospira, an ideal Achilles' heel to hamper the cell survival and deregulate ClpP proteolytic activity. Res. Microbiol. 172, 103797. https://doi.org/10.1016/j.resmic.2021.103797

Dillingham, M.S., Kowalczykowski, S.C., 2008. RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. Microbiol. Mol. Biol. Rev. 72, 642–671. https://doi.org/10.1128/MMBR.00020-08

Dixit, B., 2021. Elucidation of the adaptive stage of CRISPR-Cas subtype I-B defense system in Leptospira interrogans Copenhageni strain Fiocruz L1-130.

Dixit, B., Anand, V., Hussain, M.S., Kumar, M., 2021a. The CRISPR-associated Cas4 protein from Leptospira interrogans demonstrate versatile nuclease activity. Curr. Res. Microb. Sci. 2, 100040. https://doi.org/10.1016/j.crmicr.2021.100040

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 Lett. 590, 1002–1016. https://doi.org/10.1002/1873-3468.12124

Dixit, B., Prakash, A., Kumar, P., Gogoi, P., Kumar, M., 2021b. The core Cas1 protein of CRISPR-Cas I-B in Leptospira shows metal-tunable nuclease activity. Curr. Res. Microb. Sci. 100059.

https://doi.org/10.1016/j.crmicr.2021.100059
Dmitrieva, N.I., Cai, Q., Burg, M.B., 2004. Cells adapted to high NaCl have many DNA breaks and impaired DNA repair both in cell culture and in vivo. Proc. Natl. Acad. Sci. 101, 2317–2322. https://doi.org/10.1073/pnas.0308463100
Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., Ma, C., Wang, S., Wu, D., Ma, Y., Fan, S., Wang, J., Gao, N., Huang, Z., 2016. The crystal structure of Cpf1 in complex with CRISPR RNA. Nature 532, 522–526. https://doi.org/10.1038/nature17944
Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., Sorek, R., 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science (80-.). 359. https://doi.org/10.1126/science.aar4120
Durmaz, E., Klaenhammer, T.R., 2007. Abortive Phage Resistance Mechanism AbiZ Speeds the Lysis Clock To Cause Premature Lysis of Phage-Infected Lactococcus lactis. J. Bacteriol. 189, 1417–1425. https://doi.org/10.1128/JB.00904-06
Dy, R.L., Przybilski, R., Semeijn, K., Salmond, G.P.C., Fineran, P.C., 2014a. A widespread bacteriophage abortive infection system functions through a Type IV toxin–antitoxin mechanism. Nucleic Acids Res. 42, 4590–4605. https://doi.org/10.1093/nar/gkt1419
Dy, R.L., Richter, C., Salmond, G.P.C., Fineran, P.C., 2014b. Remarkable Mechanisms in Microbes to Resist Phage Infections. Annu. Rev. Virol. 1, 307–331. https://doi.org/10.1146/annurev-virology-031413-085500
East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H.D., Tjian, R., Doudna, J.A., 2016. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature 538, 270–273. https://doi.org/10.1038/nature19802
ELLINGHAUSEN, H.C., MCCULLOUGH, W.G., 1965. NUTRITION OF LEPTOSPIRA POMONA AND GROWTH OF 13 OTHER SEROTYPES: FRACTIONATION OF OLEIC ALBUMIN COMPLEX AND A MEDIUM OF BOVINE ALBUMIN AND POLYSORBATE 80. Am. J. Vet. Res. 26, 45–51.
Ellis, W.A., 2015. Animal Leptospirosis. pp. 99–137. https://doi.org/10.1007/978-3-662-45059-8_6
Elmore, J.R., Sheppard, N.F., Ramia, N., Deighan, T., Li, H., Terns, R.M., Terns, M.P., 2016. Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. Genes Dev. 30, 447–459. https://doi.org/10.1101/gad.272153.115
English, M.A., Soenksen, L.R., Gayet, R. V., de Puig, H., Angenent-Mari, N.M., Mao, A.S., Nguyen, P.Q., Collins, J.J., 2019. Programmable CRISPR-responsive smart materials. Science (80-.). 365, 780–785. https://doi.org/10.1126/science.aaw5122
Erdogan, H.M., Karapehlivan, M., Citil, M., Atakisi, O., Uzlu, E., Unver, A., 2008. Serum sialic acid and oxidative stress parameters changes in cattle with leptospirosis. Vet. Res. Commun. 32, 333–339. https://doi.org/10.1007/s11259-008-9036-z
Eshghi, A., Lourdault, K., Murray, G.L., Bartpho, T., Sermswan, R.W., Picardeau, M., Adler, B., Snarr, B., Zuerner, R.L., Cameron, C.E., 2012. Leptospira interrogans Catalase Is Required for Resistance to H 2 O 2 and for Virulence. Infect. Immun. 80, 3892–3899. https://doi.org/10.1128/IAI.00466-12
Estrella, M.A., Kuo, F.-T., Bailey, S., 2016. RNA-activated DNA cleavage by the Type III-B CRISPR–Cas effector complex. Genes Dev. 30, 460–470. https://doi.org/10.1101/gad.273722.115
Fan, X., Hussien, R., Brooks, G.A., 2010. H2O2-induced mitochondrial fragmentation in C2C12 myocytes. Free Radic. Biol. Med. 49, 1646–1654. https://doi.org/10.1016/j.freeradbiomed.2010.08.024
Faure, G., Makarova, K.S., Koonin, E. V., 2019. CRISPR–Cas: Complex Functional Networks and Multiple Roles beyond Adaptive Immunity. J. Mol. Biol. 431, 3–20. https://doi.org/10.1016/j.jmb.2018.08.030
Fernandes, L.G. V., Guaman, L.P., Vasconcellos, S.A., Heinemann, M.B., Picardeau, M., Nascimento, A.L.T.O., 2019. Gene silencing based on RNA-guided catalytically inactive Cas9 (dCas9): a new tool for genetic engineering in Leptospira. Sci. Rep. 9, 1839. https://doi.org/10.1038/s41598-018-37949-x
Fernandes, L.G. V., Hornsby, R.L., Nascimento, A.L.T.O., Nally, J.E., 2021. Genetic manipulation of

pathogenic Leptospira: CRISPR interference (CRISPRi)-mediated gene silencing and rapid mutant recovery at 37 °C. Sci. Rep. 11, 1768. https://doi.org/10.1038/s41598-021-81400-7

Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S., Salmond, G.P.C., 2009. The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair. Proc. Natl. Acad. Sci. 106, 894–899. https://doi.org/10.1073/pnas.0808832106

Fischer, S., Maier, L.-K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., Dyall-Smith, M., Marchfelder, A., 2012. An Archaeal Immune System Can Detect Multiple Protospacer Adjacent Motifs (PAMs) to Target Invader DNA. J. Biol. Chem. 287, 33351–33363. https://doi.org/10.1074/jbc.M112.377002

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., Charpentier, E., 2016. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature 532, 517–521. https://doi.org/10.1038/nature17945

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., Haake, D.A., Haft, D.H., Hartskeerl, R., Ko, A.I., Levett, P.N., Matsunaga, J., Mechaly, A.E., Monk, J.M., Nascimento, A.L.T., Nelson, K.E., Palsson, B., Peacock, S.J., Picardeau, M., Ricaldi, J.N., Thaipandungpanit, J., Wunder, E.A., Yang, X.F., Zhang, J.-J., Vinetz, J.M., 2016. What Makes a Bacterial Species Pathogenic?:Comparative Genomic Analysis of the Genus Leptospira. PLoS Negl. Trop. Dis. 10, e0004403. https://doi.org/10.1371/journal.pntd.0004403

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 Res. 48, 6120–6135. https://doi.org/10.1093/nar/gkaa381

Gautam, G., Jais, M.B., Kaur, R., Dhakkad, M.S., 2021. Human leptospirosis in patients with undifferentiated acute febrile illness in a tertiary care hospital in New Delhi. Int. J. Res. Med. Sci. 9, 485. https://doi.org/10.18203/2320-6012.ijrms20210428

Ghosh, K.K., Prakash, A., Balamurugan, V., Kumar, M., 2018. Catecholamine-Modulated Novel Surface-Exposed Adhesin LIC20035 of Leptospira spp. Binds Host Extracellular Matrix Components and Is Recognized by the Host during Infection. Appl. Environ. Microbiol. 84. https://doi.org/10.1128/AEM.02360-17

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 Biol. 16, 504–517. https://doi.org/10.1080/15476286.2018.1504546

Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G., Sorek, R., 2015. <scp>BREX</scp> is a novel phage resistance system widespread in microbial genomes. EMBO J. 34, 169–183. https://doi.org/10.15252/embj.201489455

González-Delgado, A., Mestre, M.R., Martínez-Abarca, F., Toro, N., 2019. Spacer acquisition from RNA mediated by a natural reverse transcriptase-Cas1 fusion protein associated with a type III-D CRISPR–Cas system in Vibrio vulnificus. Nucleic Acids Res. 47, 10202–10211. https://doi.org/10.1093/nar/gkz746

Goris, M.G.A., Boer, K.R., Duarte, T.A.T.E., Kliffen, S.J., Hartskeerl, R.A., 2013. Human Leptospirosis Trends, the Netherlands, 1925–2008. Emerg. Infect. Dis. 19, 371–378. https://doi.org/10.3201/eid1903.111260

Goswami, H.N., Rai, J., Das, A., Li, H., 2022. Molecular mechanism of active Cas7-11 in processing CRISPR RNA and interfering target RNA. Elife 11. https://doi.org/10.7554/eLife.81678

Götz, D., Paytubi, S., Munro, S., Lundgren, M., Bernander, R., White, M.F., 2007. Responses of hyperthermophilic crenarchaea to UV irradiation. Genome Biol. 8, R220. https://doi.org/10.1186/gb-2007-8-10-r220

Green, M.R., Sambrook, J., 2019. How to Win the Battle with RNase. Cold Spring Harb. Protoc. 2019, pdb.top101857. https://doi.org/10.1101/pdb.top101857

Griffith, M.E., Hospenthal, D.R., Murray, C.K., 2006. Antimicrobial therapy of leptospirosis. Curr. Opin. Infect. Dis. 19, 533–537. https://doi.org/10.1097/QCO.0b013e3280106818

Gu, D.-H., Ha, S.C., Kim, J.-S., 2019. A CRISPR RNA Is Closely Related With the Size of the Cascade Nucleoprotein Complex. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02458

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. Infect. Immun. 83, 1008–1018. https://doi.org/10.1128/IAI.03102-14

Hage, D.S., Anguizola, J.A., Li, R., Matsuda, R., Papastavros, E., Pfaunmiller, E., Sobansky, M., Zheng, X., 2013. Affinity Chromatography, in: Liquid Chromatography. Elsevier, pp. 1–23. https://doi.org/10.1016/B978-0-12-415806-1.00001-2

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., Terns, M.P., 2009. RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. Cell 139, 945–956. https://doi.org/10.1016/j.cell.2009.07.040

Halpin-Healy, T.S., Klompe, S.E., Sternberg, S.H., Fernández, I.S., 2020. Structural basis of DNA targeting by a transposon-encoded CRISPR–Cas system. Nature 577, 271–274. https://doi.org/10.1038/s41586-019-1849-0

Hammad, A.M.M., 1998. Evaluation of alginate-encapsulatedAzotobacter chroococcum as a phage-resistant and an effective inoculum. J. Basic Microbiol. 38, 9–16. https://doi.org/10.1002/(SICI)1521-4028(199803)38:1<9::AID-JOBM9>3.0.CO;2-4

Han, W., Stella, S., Zhang, Y., Guo, T., Sulek, K., Peng-Lundgren, L., Montoya, G., She, Q., 2018. A Type III-B Cmr effector complex catalyzes the synthesis of cyclic oligoadenylate second messengers by cooperative substrate binding. Nucleic Acids Res. https://doi.org/10.1093/nar/gky844

Hannon, G.J., 2002. RNA interference. Nature 418, 244–251. https://doi.org/10.1038/418244a

Hartskeerl, R.A., Collares-Pereira, M., Ellis, W.A., 2011. Emergence, control and re-emerging leptospirosis: dynamics of infection in the changing world. Clin. Microbiol. Infect. 17, 494–501. https://doi.org/10.1111/j.1469-0691.2011.03474.x

Harvey, H., Bondy-Denomy, J., Marquis, H., Sztanko, K.M., Davidson, A.R., Burrows, L.L., 2017. Pseudomonas aeruginosa defends against phages through type IV pilus glycosylation. Nat. Microbiol. 3, 47–52. https://doi.org/10.1038/s41564-017-0061-y

Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B.G., 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, 499–503. https://doi.org/10.1038/nature16995

He, L., St. John James, M., Radovcic, M., Ivancic-Bace, I., Bolt, E.L., 2020. Cas3 Protein—A Review of a Multi-Tasking Machine. Genes (Basel). 11, 208. https://doi.org/10.3390/genes11020208

Hegge, J.W., Swarts, D.C., van der Oost, J., 2018. Prokaryotic Argonaute proteins: novel genome-editing tools? Nat. Rev. Microbiol. 16, 5–11. https://doi.org/10.1038/nrmicro.2017.73

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., Marraffini, L.A., 2015. Cas9 specifies functional viral targets during CRISPR–Cas adaptation. Nature 519, 199–202. https://doi.org/10.1038/nature14245

Hidalgo-Cantabrana, C., Goh, Y.J., Barrangou, R., 2019. Characterization and Repurposing of Type I and Type II CRISPR–Cas Systems in Bacteria. J. Mol. Biol. 431, 21–33. https://doi.org/10.1016/j.jmb.2018.09.013

Hille, F., Richter, H., Wong, S.P., Bratovič, M., Ressel, S., Charpentier, E., 2018. The Biology of CRISPR-Cas: Backward and Forward. Cell 172, 1239–1259. https://doi.org/10.1016/j.cell.2017.11.032

Hochstrasser, M.L., Doudna, J.A., 2015. Cutting it close: CRISPR-associated endoribonuclease structure and function. Trends Biochem. Sci. 40, 58–66. https://doi.org/10.1016/j.tibs.2014.10.007

Hochstrasser, M.L., Taylor, D.W., Kornfeld, J.E., Nogales, E., Doudna, J.A., 2016. DNA Targeting by a Minimal CRISPR RNA-Guided Cascade. Mol. Cell 63, 840–851. https://doi.org/10.1016/j.molcel.2016.07.027

Hoikkala, V., Ravantti, J., Díez-Villaseñor, C., Tiirola, M., Conrad, R.A., McBride, M.J., Moineau, S., Sundberg, L.-R., 2021. Cooperation between Different CRISPR-Cas Types Enables Adaptation in an RNA-Targeting System. MBio 12. https://doi.org/10.1128/mBio.03338-20

Holla, R., Darshan, B., Pandey, L., Unnikrishnan, B., Kumar, N., Thapar, R., Mithra, P., Kulkarni, V., 2018. Leptospirosis in Coastal South India: A Facility Based Study. Biomed Res. Int. 2018, 1–5.

https://doi.org/10.1155/2018/1759125
Holm, L., Laakso, L.M., 2016. Dali server update. Nucleic Acids Res. 44, W351–W355. https://doi.org/10.1093/nar/gkw357
Hooton, S.P.T., Connerton, I.F., 2015. Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. Front. Microbiol. 5. https://doi.org/10.3389/fmicb.2014.00744
Hrle, A., Su, A.A., Ebert, J., Benda, C., Randau, L., Conti, E., 2013. Structure and RNA-binding properties of the Type III-A CRISPR-associated protein Csm3. RNA Biol. 10, 1670–1678. https://doi.org/10.4161/rna.26500
Hudaiberdiev, S., Shmakov, S., Wolf, Y.I., Terns, M.P., Makarova, K.S., Koonin, E. V., 2017. Phylogenomics of Cas4 family nucleases. BMC Evol. Biol. 17, 232. https://doi.org/10.1186/s12862-017-1081-1
Hussain, M.S., Kumar, M., 2022. Assembly of Cas7 subunits of Leptospira on the mature crRNA of CRISPR-Cas I-B is modulated by divalent ions. Gene 818, 146244. https://doi.org/10.1016/j.gene.2022.146244
Hutvagner, G., Simard, M.J., 2008. Argonaute proteins: key players in RNA silencing. Nat. Rev. Mol. Cell Biol. 9, 22–32. https://doi.org/10.1038/nrm2321
Ibarra-Chávez, R., Brady, A., Chen, J., Penadés, J.R., Haag, A.F., 2022. Phage-inducible chromosomal islands promote genetic variability by blocking phage reproduction and protecting transductants from phage lysis. PLOS Genet. 18, e1010146. https://doi.org/10.1371/journal.pgen.1010146
Inada, R., Ido, Y., Hoki, R., Kaneko, R., Ito, H., 1916. THE ETIOLOGY, MODE OF INFECTION, AND SPECIFIC THERAPY OF WEIL'S DISEASE (SPIROCHÆTOSIS ICTEROHÆMORRHAGICA). J. Exp. Med. 23, 377–402. https://doi.org/10.1084/jem.23.3.377
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. J. Bacteriol. 169, 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987
Jackson, Ryan N., Golden, S.M., van Erp, P.B.G., Carter, J., Westra, E.R., Brouns, S.J.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 (80-.). 345, 1473–1479. https://doi.org/10.1126/science.1256328
Jackson, R. N., Golden, S.M., van Erp, P.B.G., Carter, J., Westra, E.R., Brouns, S.J.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 (80-.). 345, 1473–1479. https://doi.org/10.1126/science.1256328
Jackson, R.N., Wiedenheft, B., 2015. A Conserved Structural Chassis for Mounting Versatile CRISPR RNA-Guided Immune Responses. Mol. Cell 58, 722–728. https://doi.org/10.1016/j.molcel.2015.05.023
Jackson, S.A., Birkholz, N., Malone, L.M., Fineran, P.C., 2019. Imprecise Spacer Acquisition Generates CRISPR-Cas Immune Diversity through Primed Adaptation. Cell Host Microbe 25, 250-260.e4. https://doi.org/10.1016/j.chom.2018.12.014
James, S., Sathian, B., Teijlingen, E. Van, Asim, M., 2018. Outbreak of Leptospirosis in Kerala. Nepal J. Epidemiol. 8, 745–747. https://doi.org/10.3126/nje.v8i4.23876
Jansen, R., Embden, J.D.A. van, Gaastra, W., Schouls, L.M., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 43, 1565–1575. https://doi.org/10.1046/j.1365-2958.2002.02839.x
Jia, N., Mo, C.Y., Wang, C., Eng, E.T., Marraffini, L.A., Patel, D.J., 2019. Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity. Mol. Cell 73, 264-277.e5. https://doi.org/10.1016/j.molcel.2018.11.007
Jiang, W., Marraffini, L.A., 2015. CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. Annu. Rev. Microbiol. 69, 209–228. https://doi.org/10.1146/annurev-micro-091014-104441
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A Programmable

Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science (80-.). 337, 816–821. https://doi.org/10.1126/science.1225829

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A.T., Charpentier, E., Nogales, E., Doudna, J.A., 2014. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. Science (80-.). 343. https://doi.org/10.1126/science.1247997

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T.L., 2008. NCBI BLAST: a better web interface. Nucleic Acids Res. 36, W5–W9. https://doi.org/10.1093/nar/gkn201

Johnson, R.C., Harris, V.G., 1967. Differentiation of Pathogenic and Saprophytic Leptospires I. Growth at Low Temperatures. J. Bacteriol. 94, 27–31. https://doi.org/10.1128/jb.94.1.27-31.1967

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kalwani, P., Rath, D., Ballal, A., 2020. Novel molecular aspects of the CRISPR backbone protein 'Cas7' from cyanobacteria. Biochem. J. 477, 971–983. https://doi.org/10.1042/BCJ20200026

Kaminski, M.M., Abudayyeh, O.O., Gootenberg, J.S., Zhang, F., Collins, J.J., 2021. CRISPR-based diagnostics. Nat. Biomed. Eng. 5, 643–656. https://doi.org/10.1038/s41551-021-00760-7

Kang, S.O., Caparon, M.G., Cho, K.H., 2010. Virulence Gene Regulation by CvfA, a Putative RNase: the CvfA-Enolase Complex in Streptococcus pyogenes Links Nutritional Stress, Growth-Phase Control, and Virulence Gene Expression. Infect. Immun. 78, 2754–2767. https://doi.org/10.1128/IAI.01370-09

Karginov, F. V., Hannon, G.J., 2010. The CRISPR System: Small RNA-Guided Defense in Bacteria and Archaea. Mol. Cell 37, 7–19. https://doi.org/10.1016/j.molcel.2009.12.033

Kaya, E., Doxzen, K.W., Knoll, K.R., Wilson, R.C., Strutt, S.C., Kranzusch, P.J., Doudna, J.A., 2016. A bacterial Argonaute with noncanonical guide RNA specificity. Proc. Natl. Acad. Sci. 113, 4057–4062. https://doi.org/10.1073/pnas.1524385113

Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., Siksnys, V., 2017. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science (80-.). 357, 605–609. https://doi.org/10.1126/science.aao0100

Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č., Siksnys, V., 2016. Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. Mol. Cell 62, 295–306. https://doi.org/10.1016/j.molcel.2016.03.024

Keatch, S.A., 2004. Alleviation of restriction by DNA condensation and non-specific DNA binding ligands. Nucleic Acids Res. 32, 5841–5850. https://doi.org/10.1093/nar/gkh918

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858. https://doi.org/10.1038/nprot.2015.053

Kembhavi, R., Velhal, G., Shah, A., 2021. Epidemiological determinants of leptospirosis in rural and urban districts of Maharashtra, India. J. Fam. Med. Prim. Care 10, 3361. https://doi.org/10.4103/jfmpc.jfmpc_674_21

Ketting, R.F., 2011. The Many Faces of RNAi. Dev. Cell 20, 148–161. https://doi.org/10.1016/j.devcel.2011.01.012

Knott, G.J., Doudna, J.A., 2018. CRISPR-Cas guides the future of genetic engineering. Science (80-.). 361, 866–869. https://doi.org/10.1126/science.aat5011

Knott, G.J., East-Seletsky, A., Cofsky, J.C., Holton, J.M., Charles, E., O'Connell, M.R., Doudna, J.A., 2017. Guide-bound structures of an RNA-targeting A-cleaving CRISPR–Cas13a enzyme. Nat. Struct. Mol. Biol. 24, 825–833. https://doi.org/10.1038/nsmb.3466

Ko, A.I., Goarant, C., Picardeau, M., 2009. Leptospira: the dawn of the molecular genetics era for an

emerging zoonotic pathogen. Nat. Rev. Microbiol. 7, 736–747. https://doi.org/10.1038/nrmicro2208

Koga, M., Otsuka, Y., Lemire, S., Yonesaki, T., 2011. Escherichia coli rnlA and rnlB Compose a Novel Toxin–Antitoxin System. Genetics 187, 123–130. https://doi.org/10.1534/genetics.110.121798

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., Liu, D.R., 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424. https://doi.org/10.1038/nature17946

Koonin, E. V, Makarova, K.S., Zhang, F., 2017. Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. 37, 67–78. https://doi.org/10.1016/j.mib.2017.05.008

Kratz, A., Ferraro, M., Sluss, P.M., Lewandrowski, K.B., 2004. Normal Reference Laboratory Values. N. Engl. J. Med. 351, 1548–1563. https://doi.org/10.1056/NEJMcpc049016

Kunin, V., Sorek, R., Hugenholtz, P., 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol. 8, R61. https://doi.org/10.1186/gb-2007-8-4-r61

Kuznedelov, K., Mekler, V., Lemak, S., Tokmina-Lukaszewska, M., Datsenko, K.A., Jain, I., Savitskaya, E., Mallon, J., Shmakov, S., Bothner, B., Bailey, S., Yakunin, A.F., Severinov, K., Semenova, E., 2016. Altered stoichiometry Escherichia coli Cascade complexes with shortened CRISPR RNA spacers are capable of interference and primed adaptation. Nucleic Acids Res. 44, 10849–10861. https://doi.org/10.1093/nar/gkw914

Lau, R.K., Ye, Q., Birkholz, E.A., Berg, K.R., Patel, L., Mathews, I.T., Watrous, J.D., Ego, K., Whiteley, A.T., Lowey, B., Mekalanos, J.J., Kranzusch, P.J., Jain, M., Pogliano, J., Corbett, K.D., 2020. Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. Mol. Cell 77, 723-733.e6. https://doi.org/10.1016/j.molcel.2019.12.010

Lee, H., Dhingra, Y., Sashital, D.G., 2019. The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. Elife 8. https://doi.org/10.7554/eLife.44248

Lee, H., Zhou, Y., Taylor, D.W., Sashital, D.G., 2018. Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. Mol. Cell 70, 48-59.e5. https://doi.org/10.1016/j.molcel.2018.03.003

Leenay, R.T., Beisel, C.L., 2017. Deciphering, Communicating, and Engineering the CRISPR PAM. J. Mol. Biol. 429, 177–191. https://doi.org/10.1016/j.jmb.2016.11.024

Levett, P.N., 2004. Leptospirosis: A forgotten zoonosis? Clin. Appl. Immunol. Rev. 4, 435–448. https://doi.org/10.1016/j.cair.2004.08.001

Levett, P.N., 2001. Leptospirosis. Clin. Microbiol. Rev. 14, 296–326. https://doi.org/10.1128/CMR.14.2.296-326.2001

Levitz, R., Chapman, D., Amitsur, M., Green, R., Snyder, L., Kaufmann, G., 1990. The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease. EMBO J. 9, 1383–1389. https://doi.org/10.1002/j.1460-2075.1990.tb08253.x

Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., Sorek, R., 2015. CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature 520, 505–510. https://doi.org/10.1038/nature14302

Li, L., Li, S., Wu, N., Wu, J., Wang, G., Zhao, G., Wang, J., 2019. HOLMESv2: A CRISPR-Cas12b-Assisted Platform for Nucleic Acid Detection and DNA Methylation Quantitation. ACS Synth. Biol. 8, 2228–2237. https://doi.org/10.1021/acssynbio.9b00209

Li, Y., Li, S., Wang, J., Liu, G., 2019. CRISPR/Cas Systems towards Next-Generation Biosensing. Trends Biotechnol. 37, 730–743. https://doi.org/10.1016/j.tibtech.2018.12.005

Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., Chen, L., Liang, Y.X., She, Q., 2016. Harnessing Type I and Type III CRISPR-Cas systems for genome editing. Nucleic Acids Res. 44, e34–e34. https://doi.org/10.1093/nar/gkv1044

Li, Y., Peng, N., 2019. Endogenous CRISPR-Cas System-Based Genome Editing and Antimicrobials: Review and Prospects. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02471

Lindsay, J.A., Ruzin, A., Ross, H.F., Kurepina, N., Novick, R.P., 1998. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in Staphylococcus aureus. Mol. Microbiol. 29, 527–543. https://doi.org/10.1046/j.1365-2958.1998.00947.x

Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q., Copié, V., Young, M.J., White, M.F., Lawrence, C.M., 2011. Structural and Functional Characterization of an Archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Complex for Antiviral Defense (CASCADE). J. Biol. Chem. 286, 21643–21656. https://doi.org/10.1074/jbc.M111.238485

Liu, L., Chen, P., Wang, M., Li, X., Wang, J., Yin, M., Wang, Y., 2017a. C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism. Mol. Cell 65, 310–322. https://doi.org/10.1016/j.molcel.2016.11.040

Liu, L., Li, X., Ma, J., Li, Z., You, L., Wang, J., Wang, M., Zhang, X., Wang, Y., 2017b. The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. Cell 170, 714-726.e10. https://doi.org/10.1016/j.cell.2017.06.050

Liu, M., Deora, R., Doulatov, S.R., Gingery, M., Eiserling, F.A., Preston, A., Maskell, D.J., Simons, R.W., Cotter, P.A., Parkhill, J., Miller, J.F., 2002. Reverse Transcriptase-Mediated Tropism Switching in Bordetella Bacteriophage. Science (80-.). 295, 2091–2094. https://doi.org/10.1126/science.1067467

Liu, T.Y., Doudna, J.A., 2020. Chemistry of Class 1 CRISPR-Cas effectors: Binding, editing, and regulation. J. Biol. Chem. 295, 14473–14487. https://doi.org/10.1074/jbc.REV120.007034

Liu, T.Y., Iavarone, A.T., Doudna, J.A., 2017. RNA and DNA Targeting by a Reconstituted Thermus thermophilus Type III-A CRISPR-Cas System. PLoS One 12, e0170552. https://doi.org/10.1371/journal.pone.0170552

Loeff, L., Brouns, S.J.J., Joo, C., 2018. Repetitive DNA Reeling by the Cascade-Cas3 Complex in Nucleotide Unwinding Steps. Mol. Cell 70, 385-394.e3. https://doi.org/10.1016/j.molcel.2018.03.031

Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A., Wilson, G.G., Murray, N.E., 2014. Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic Acids Res. 42, 3–19. https://doi.org/10.1093/nar/gkt990

Lopatina, A., Tal, N., Sorek, R., 2020. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. Annu. Rev. Virol. 7, 371–384. https://doi.org/10.1146/annurev-virology-011620-040628

Louis-Jeune, C., Andrade-Navarro, M.A., Perez-Iratxeta, C., 2012. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. Proteins Struct. Funct. Bioinforma. 80, 374–381. https://doi.org/10.1002/prot.23188

Louwen, R., Horst-Kreft, D., Boer, A.G., Graaf, L., Knegt, G., Hamersma, M., Heikema, A.P., Timms, A.R., Jacobs, B.C., Wagenaar, J.A., Endtz, H.P., Oost, J., Wells, J.M., Nieuwenhuis, E.E.S., Vliet, A.H.M., Willemsen, P.T.J., Baarlen, P., Belkum, A., 2013. A novel link between Campylobacter jejuni bacteriophage defence, virulence and Guillain–Barré syndrome. Eur. J. Clin. Microbiol. Infect. Dis. 32, 207–226. https://doi.org/10.1007/s10096-012-1733-4

Louwen, R., Staals, R.H.J., Endtz, H.P., van Baarlen, P., van der Oost, J., 2014. The Role of CRISPR-Cas Systems in Virulence of Pathogenic Bacteria. Microbiol. Mol. Biol. Rev. 78, 74–88. https://doi.org/10.1128/MMBR.00039-13

Lu, M.-J., Henning, U., 1994. Superinfection exclusion by T-even-type coliphages. Trends Microbiol. 2, 137–139. https://doi.org/10.1016/0966-842X(94)90601-7

Ma, K., Cao, Q., Luo, S., Wang, Z., Liu, G., Lu, C., Liu, Y., 2018. cas9 Enhances Bacterial Virulence by Repressing the regR Transcriptional Regulator in Streptococcus agalactiae. Infect. Immun. 86. https://doi.org/10.1128/IAI.00552-17

Maguire, M.E., Cowan, J.A., 2002. Magnesium chemistry and biochemistry. BioMetals 15, 203–210. https://doi.org/10.1023/A:1016058229972

Makarova, Kira S, Aravind, L., Wolf, Y.I., Koonin, E. V, 2011. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol. Direct 6, 38. https://doi.org/10.1186/1745-6150-6-38

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. Biol. Direct 1, 7. https://doi.org/10.1186/1745-6150-1-7

Makarova, Kira S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., van der Oost, J., Koonin, E. V., 2011. Evolution and classification of the CRISPR–Cas systems. Nat. Rev. Microbiol. 9, 467–477. https://doi.org/10.1038/nrmicro2577

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., van der Oost, J., Backofen, R., Koonin, E. V., 2015. An updated evolutionary classification of CRISPR–Cas systems. Nat. Rev. Microbiol. 13, 722–736. https://doi.org/10.1038/nrmicro3569

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Scott, D., Shah, S.A., Siksnys, V., Terns, M.P., Venclovas, Č., White, M.F., Yakunin, A.F., Yan, W., Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J., Barrangou, R., Koonin, E. V., 2020a. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. 18, 67–83. https://doi.org/10.1038/s41579-019-0299-x

Makarova, K.S., Wolf, Y.I., Koonin, E. V., 2018. Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? Cris. J. 1, 325–336. https://doi.org/10.1089/crispr.2018.0033

Makarova, K.S., Wolf, Y.I., Shmakov, S.A., Liu, Y., Li, M., Koonin, E. V., 2020b. Unprecedented Diversity of Unique CRISPR-Cas-Related Systems and Cas1 Homologs in Asgard Archaea. Cris. J. 3, 156–163. https://doi.org/10.1089/crispr.2020.0012

Makarova, K.S., Zhang, F., Koonin, E. V., 2017. SnapShot: Class 1 CRISPR-Cas Systems. Cell 168, 946-946.e1. https://doi.org/10.1016/j.cell.2017.02.018

Malmström, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E.W., Aebersold, R., 2009. Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. Nature 460, 762–765. https://doi.org/10.1038/nature08184

Marangoni, A., 2006. Production of reactive oxygen species and expression of inducible nitric oxide synthase in rat isolated Kupffer cells stimulated by Leptospira interrogans and Borrelia burgdorferi. World J. Gastroenterol. 12, 3077. https://doi.org/10.3748/wjg.v12.i19.3077

Mark A. Suckow, Steven H. Weisbroth, C.L.F., 2005. The laboratory rat, 2nd ed. Elsevier, Burlington, MA.

Marraffini, L.A., Sontheimer, E.J., 2010. Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 463, 568–571. https://doi.org/10.1038/nature08703

Martínez Arbas, S., Narayanasamy, S., Herold, M., Lebrun, L.A., Hoopmann, M.R., Li, S., Lam, T.J., Kunath, B.J., Hicks, N.D., Liu, C.M., Price, L.B., Laczny, C.C., Gillece, J.D., Schupp, J.M., Keim, P.S., Moritz, R.L., Faust, K., Tang, H., Ye, Y., Skupin, A., May, P., Muller, E.E.L., Wilmes, P., 2020. Roles of bacteriophages, plasmids and CRISPR immunity in microbial community dynamics revealed using time-series integrated meta-omics. Nat. Microbiol. 6, 123–135. https://doi.org/10.1038/s41564-020-00794-8

Matsunaga, J., Lo, M., Bulach, D.M., Zuerner, R.L., Adler, B., Haake, D.A., 2007. Response of Leptospira interrogans to Physiologic Osmolarity: Relevance in Signaling the Environment-to-Host Transition. Infect. Immun. 75, 2864–2874. https://doi.org/10.1128/IAI.01619-06

McBride, T.M., Schwartz, E.A., Kumar, A., Taylor, D.W., Fineran, P.C., Fagerlund, R.D., 2020. Diverse CRISPR-Cas Complexes Require Independent Translation of Small and Large Subunits from a Single Gene. Mol. Cell 80, 971-979.e7. https://doi.org/10.1016/j.molcel.2020.11.003

McGinn, J., Marraffini, L.A., 2016. CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. Mol. Cell 64, 616–623. https://doi.org/10.1016/j.molcel.2016.08.038

Mendoza, B.J., Trinh, C.T., 2018. In Silico Processing of the Complete CRISPR-Cas Spacer Space for Identification of PAM Sequences. Biotechnol. J. 13, 1700595. https://doi.org/10.1002/biot.201700595

Micsonai, A., Wien, F., Kernya, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., Kardos, J., 2015. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proc.

Natl. Acad. Sci. 112, E3095–E3103. https://doi.org/10.1073/pnas.1500851112

Miraglia, F., Matsuo, M., Morais, Z.M., Dellagostin, O.A., Seixas, F.K., Freitas, J.C., Hartskeerl, R., Moreno, L.Z., Costa, B.L., Souza, G.O., Vasconcellos, S.A., Moreno, A.M., 2013. Molecular characterization, serotyping, and antibiotic susceptibility profile of Leptospira interrogans serovar Copenhageni isolates from Brazil. Diagn. Microbiol. Infect. Dis. 77, 195–199. https://doi.org/10.1016/j.diagmicrobio.2013.08.003

Miroshnikova, A.D., Kuznetsova, A.A., Vorobjev, Y.N., Kuznetsov, N.A., Fedorova, O.S., 2016. Effects of mono- and divalent metal ions on DNA binding and catalysis of human apurinic/apyrimidinic endonuclease 1. Mol. Biosyst. 12, 1527–1539. https://doi.org/10.1039/C6MB00128A

Miyamoto, T., Okano, S., Kasai, N., 2009. Irreversible thermoinactivation of ribonuclease-A by soft-hydrothermal processing. Biotechnol. Prog. NA-NA. https://doi.org/10.1002/btpr.267

Miyoshi, T., Ito, K., Murakami, R., Uchiumi, T., 2016. Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. Nat. Commun. 7, 11846. https://doi.org/10.1038/ncomms11846

Modell, J.W., Jiang, W., Marraffini, L.A., 2017. CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. Nature 544, 101–104. https://doi.org/10.1038/nature21719

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., Almendros, C., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733–740. https://doi.org/10.1099/mic.0.023960-0

Mojica, F.J.M., Diez-Villasenor, C., Garcia-Martinez, J., Soria, E., 2005. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. J. Mol. Evol. 60, 174–182. https://doi.org/10.1007/s00239-004-0046-3

Moncany, M.L.J., Kellenberger, E., 1981. High magnesium content ofEscherichia coli B. Experientia 37, 846–847. https://doi.org/10.1007/BF01985672

Moon, S. Bin, Kim, D.Y., Ko, J.-H., Kim, Y.-S., 2019. Recent advances in the CRISPR genome editing tool set. Exp. Mol. Med. 51, 1–11. https://doi.org/10.1038/s12276-019-0339-7

Morona, R., Klose, M., Henning, U., 1984. Escherichia coli K-12 outer membrane protein (OmpA) as a bacteriophage receptor: analysis of mutant genes expressing altered proteins. J. Bacteriol. 159, 570–578. https://doi.org/10.1128/jb.159.2.570-578.1984

Mraheil, M.A., Billion, A., Mohamed, W., Mukherjee, K., Kuenne, C., Pischimarov, J., Krawitz, C., Retey, J., Hartsch, T., Chakraborty, T., Hain, T., 2011. The intracellular sRNA transcriptome of Listeria monocytogenes during growth in macrophages. Nucleic Acids Res. 39, 4235–4248. https://doi.org/10.1093/nar/gkr033

Murray, G.L., Morel, V., Cerqueira, G.M., Croda, J., Srikram, A., Henry, R., Ko, A.I., Dellagostin, O.A., Bulach, D.M., Sermswan, R.W., Adler, B., Picardeau, M., 2009. Genome-Wide Transposon Mutagenesis in Pathogenic Leptospira Species. Infect. Immun. 77, 810–816. https://doi.org/10.1128/IAI.01293-08

Murray, N.E., 2000. Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). Microbiol. Mol. Biol. Rev. 64, 412–434. https://doi.org/10.1128/MMBR.64.2.412-434.2000

Murugan, K., Babu, K., Sundaresan, R., Rajan, R., Sashital, D.G., 2017. The Revolution Continues: Newly Discovered Systems Expand the CRISPR-Cas Toolkit. Mol. Cell 68, 15–25. https://doi.org/10.1016/j.molcel.2017.09.007

Musharova, O., Medvedeva, S., Klimuk, E., Guzman, N.M., Titova, D., Zgoda, V., Shiriaeva, A., Semenova, E., Severinov, K., Savitskaya, E., 2021. Prespacers formed during primed adaptation associate with the Cas1–Cas2 adaptation complex and the Cas3 interference nuclease–helicase. Proc. Natl. Acad. Sci. 118, e2021291118. https://doi.org/10.1073/pnas.2021291118

Musharova, O., Sitnik, V., Vlot, M., Savitskaya, E., Datsenko, K.A., Krivoy, A., Fedorov, I., Semenova, E., Brouns, S.J.J., Severinov, K., 2019. Systematic analysis of Type I-E Escherichia coli CRISPR-Cas PAM sequences ability to promote interference and primed adaptation. Mol. Microbiol. 111, 1558–1570. https://doi.org/10.1111/mmi.14237

Nakamura, S., 2020. Measurement of the Cell-Body Rotation of Leptospira. pp. 139–148. https://doi.org/10.1007/978-1-0716-0459-5_13

Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M.P., Ke, A., 2012a. Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System. Structure 20, 1574–1584. https://doi.org/10.1016/j.str.2012.06.016

Nam, K.H., Huang, Q., Ke, A., 2012b. Nucleic acid binding surface and dimer interface revealed by CRISPR-associated CasB protein structures. FEBS Lett. 586, 3956–3961. https://doi.org/10.1016/j.febslet.2012.09.041

Nascimento, A.L.T.O., Verjovski-Almeida, S., Van Sluys, M.A., Monteiro-Vitorello, C.B., Camargo, L.E.A., Digiampietri, L.A., Harstkeerl, R.A., Ho, P.L., Marques, M.V., Oliveira, M.C., Setubal, J.C., Haake, D.A., Martins, E.A.L., 2004. Genome features of Leptospira interrogans serovar Copenhageni. Brazilian J. Med. Biol. Res. 37, 459–477. https://doi.org/10.1590/S0100-879X2004000400003

Nickel, L., Weidenbach, K., Jäger, D., Backofen, R., Lange, S.J., Heidrich, N., Schmitz, R.A., 2013. Two CRISPR-Cas systems in Methanosarcina mazei strain Gö1 display common processing features despite belonging to different types I and III. RNA Biol. 10, 779–791. https://doi.org/10.4161/rna.23928

Nimkar, S., Anand, B., 2020. Cas3/I-C mediated target DNA recognition and cleavage during CRISPR interference are independent of the composition and architecture of Cascade surveillance complex. Nucleic Acids Res. 48, 2486–2501. https://doi.org/10.1093/nar/gkz1218

Nishimasu, H., Cong, L., Yan, W.X., Ran, F.A., Zetsche, B., Li, Y., Kurabayashi, A., Ishitani, R., Zhang, F., Nureki, O., 2015. Crystal Structure of Staphylococcus aureus Cas9. Cell 162, 1113–1126. https://doi.org/10.1016/j.cell.2015.08.007

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O., 2014. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. Cell 156, 935–949. https://doi.org/10.1016/j.cell.2014.02.001

Nuñez, J.K., Bai, L., Harrington, L.B., Hinder, T.L., Doudna, J.A., 2016. CRISPR Immunological Memory Requires a Host Factor for Specificity. Mol. Cell 62, 824–833. https://doi.org/10.1016/j.molcel.2016.04.027

Nuñez, J.K., Lee, A.S.Y., Engelman, A., Doudna, J.A., 2015. Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. Nature 519, 193–198. https://doi.org/10.1038/nature14237

Nussenzweig, P.M., McGinn, J., Marraffini, L.A., 2019. Cas9 Cleavage of Viral Genomes Primes the Acquisition of New Immunological Memories. Cell Host Microbe 26, 515-526.e6. https://doi.org/10.1016/j.chom.2019.09.002

O'Brien, R.E., Santos, I.C., Wrapp, D., Bravo, J.P.K., Schwartz, E.A., Brodbelt, J.S., Taylor, D.W., 2020. Structural basis for assembly of non-canonical small subunits into type I-C Cascade. Nat. Commun. 11, 5931. https://doi.org/10.1038/s41467-020-19785-8

Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., Sorek, R., 2017. DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat. Microbiol. 3, 90–98. https://doi.org/10.1038/s41564-017-0051-0

Oganesyan, N., Ankoudinova, I., Kim, S.-H., Kim, R., 2007. Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization. Protein Expr. Purif. 52, 280–285. https://doi.org/10.1016/j.pep.2006.09.015

Oliveira, R., de Morais, Z.M., Gonçales, A.P., Romero, E.C., Vasconcellos, S.A., Nascimento, A.L.T.O., 2011. Characterization of Novel OmpA-Like Protein of Leptospira interrogans That Binds Extracellular Matrix Molecules and Plasminogen. PLoS One 6, e21962. https://doi.org/10.1371/journal.pone.0021962

Osawa, T., Inanaga, H., Sato, C., Numata, T., 2015. Crystal Structure of the CRISPR-Cas RNA Silencing Cmr Complex Bound to a Target Analog. Mol. Cell 58, 418–430. https://doi.org/10.1016/j.molcel.2015.03.018

Özcan, A., Krajeski, R., Ioannidi, E., Lee, B., Gardner, A., Makarova, K.S., Koonin, E. V., Abudayyeh,

O.O., Gootenberg, J.S., 2021. Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. Nature 597, 720–725. https://doi.org/10.1038/s41586-021-03886-5

Özcan, A., Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G., Randau, L., 2018. Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. Nat. Microbiol. 4, 89–96. https://doi.org/10.1038/s41564-018-0274-8

Palaniappan, R.U.M., Chang, Y.-F., Jusuf, S.S.D., Artiushin, S., Timoney, J.F., McDonough, S.P., Barr, S.C., Divers, T.J., Simpson, K.W., McDonough, P.L., Mohammed, H.O., 2002. Cloning and Molecular Characterization of an Immunogenic LigA Protein of Leptospira interrogans. Infect. Immun. 70, 5924–5930. https://doi.org/10.1128/IAI.70.11.5924-5930.2002

Pappas, G., Papadimitriou, P., Siozopoulou, V., Christou, L., Akritidis, N., 2008. The globalization of leptospirosis: worldwide incidence trends. Int. J. Infect. Dis. 12, 351–357. https://doi.org/10.1016/j.ijid.2007.09.011

Park, J.-H., Sun, J., Park, S.-Y., Hwang, H.-J., Park, M.-Y., Shin, M., Kim, J.-S., 2013. Crystal structure of Cmr5 from Pyrococcus furiosus and its functional implications. FEBS Lett. 587, 562–568. https://doi.org/10.1016/j.febslet.2013.01.029

Parma, D.H., Snyder, M., Sobolevski, S., Nawroz, M., Brody, E., Gold, L., 1992. The Rex system of bacteriophage lambda: tolerance and altruistic cell death. Genes Dev. 6, 497–510. https://doi.org/10.1101/gad.6.3.497

Pětrošová, H., Picardeau, M., 2014. Screening of a Leptospira biflexa Mutant Library To Identify Genes Involved in Ethidium Bromide Tolerance. Appl. Environ. Microbiol. 80, 6091–6103. https://doi.org/10.1128/AEM.01619-14

Picardeau, M., 2020. Leptospira and Leptospirosis. pp. 271–275. https://doi.org/10.1007/978-1-0716-0459-5_24

Picardeau, M., 2015. Genomics, Proteomics, and Genetics of Leptospira. pp. 43–63. https://doi.org/10.1007/978-3-662-45059-8_4

Picardeau, M., 2013. Diagnosis and epidemiology of leptospirosis. Médecine Mal. Infect. 43, 1–9. https://doi.org/10.1016/j.medmal.2012.11.005

Pijnacker, R., Goris, M.G.A., te Wierik, M.J.M., Broens, E.M., van der Giessen, J.W.B., de Rosa, M., Wagenaar, J.A., Hartskeerl, R.A., Notermans, D.W., Maassen, K., Schimmer, B., 2016. Marked increase in leptospirosis infections in humans and dogs in the Netherlands, 2014. Eurosurveillance 21. https://doi.org/10.2807/1560-7917.ES.2016.21.17.30211

Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R.A., Randau, L., Sørensen, S.J., Shah, S.A., 2020. Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. Nucleic Acids Res. 48, 2000–2012. https://doi.org/10.1093/nar/gkz1197

Pinilla-Redondo, R., Russel, J., Mayo-Muñoz, D., Shah, S.A., Garrett, R.A., Nesme, J., Madsen, J.S., Fineran, P.C., Sørensen, S.J., 2022. CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids. Nucleic Acids Res. 50, 4315–4328. https://doi.org/10.1093/nar/gkab859

Plagens, A., Tjaden, B., Hagemann, A., Randau, L., Hensel, R., 2012. Characterization of the CRISPR/Cas Subtype I-A System of the Hyperthermophilic Crenarchaeon Thermoproteus tenax. J. Bacteriol. 194, 2491–2500. https://doi.org/10.1128/JB.00206-12

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 I-A Cascade interference complex. Nucleic Acids Res. 42, 5125–5138. https://doi.org/10.1093/nar/gku120

Plank, R., Dean, D., 2000. Overview of the epidemiology, microbiology, and pathogenesis of Leptospira spp. in humans. Microbes Infect. 2, 1265–1276. https://doi.org/10.1016/S1286-4579(00)01280-6

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. Mol. Microbiol. 77, 1367–1379. https://doi.org/10.1111/j.1365-2958.2010.07265.x

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, 653–663. https://doi.org/10.1099/mic.0.27437-0

Prakash, A., Kumar, M., 2022. Transcriptional analysis of CRISPR I-B arrays of Leptospira interrogans serovar Lai and its processing by Cas6. Front. Microbiol. 13. https://doi.org/10.3389/fmicb.2022.960559

Prakash, A., Kumar, M., 2021. Characterizing the transcripts of Leptospira CRISPR I-B array and its processing with endoribonuclease LinCas6. Int. J. Biol. Macromol. 182, 785–795. https://doi.org/10.1016/j.ijbiomac.2021.04.066

Prasad, B., Unadkat, J.D., 2014. Optimized Approaches for Quantification of Drug Transporters in Tissues and Cells by MRM Proteomics. AAPS J. 16, 634–648. https://doi.org/10.1208/s12248-014-9602-y

Punetha, A., Sivathanu, R., Anand, B., 2014. Active site plasticity enables metal-dependent tuning of Cas5d nuclease activity in CRISPR-Cas type I-C system. Nucleic Acids Res. 42, 3846–3856. https://doi.org/10.1093/nar/gkt1335

Pyne, M.E., Bruder, M.R., Moo-Young, M., Chung, D.A., Chou, C.P., 2016. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. Sci. Rep. 6, 25666. https://doi.org/10.1038/srep25666

Rajagopalan, R., Kroos, L., 2017. The dev Operon Regulates the Timing of Sporulation during Myxococcus xanthus Development. J. Bacteriol. 199. https://doi.org/10.1128/JB.00788-16

Ramachandran, A., Summerville, L., Learn, B.A., DeBell, L., Bailey, S., 2020. Processing and integration of functionally oriented prespacers in the Escherichia coli CRISPR system depends on bacterial host exonucleases. J. Biol. Chem. 295, 3403–3414. https://doi.org/10.1074/jbc.RA119.012196

Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B., Doudna, J.A., Greene, E.C., 2015. Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. Cell 163, 854–865. https://doi.org/10.1016/j.cell.2015.10.003

Reeks, J., Graham, S., Anderson, L., Liu, H., White, M.F., Naismith, J.H., 2013a. Structure of the archaeal Cascade subunit Csa5. RNA Biol. 10, 762–769. https://doi.org/10.4161/rna.23854

Reeks, J., Naismith, J.H., White, M.F., 2013b. CRISPR interference: a structural perspective. Biochem. J. 453, 155–166. https://doi.org/10.1042/BJ20130316

Richard, E., Bourhy, P., Picardeau, M., Moulin, L., Wurtzer, S., 2021. Effect of disinfection agents and quantification of potentially viable Leptospira in fresh water samples using a highly sensitive integrity-qPCR assay. PLoS One 16, e0251901. https://doi.org/10.1371/journal.pone.0251901

Richter, H., Lange, S.J., Backofen, R., Randau, L., 2013. Comparative analysis of Cas6b processing and CRISPR RNA stability. RNA Biol. 10, 700–707. https://doi.org/10.4161/rna.23715

Richter, H., Rompf, J., Wiegel, J., Rau, K., Randau, L., 2017. Fragmentation of the CRISPR-Cas Type I-B signature protein Cas8b. Biochim. Biophys. Acta - Gen. Subj. 1861, 2993–3000. https://doi.org/10.1016/j.bbagen.2017.02.026

Rojas, P., Monahan, A.M., Schuller, S., Miller, I.S., Markey, B.K., Nally, J.E., 2010. Detection and quantification of leptospires in urine of dogs: a maintenance host for the zoonotic disease leptospirosis. Eur. J. Clin. Microbiol. Infect. Dis. 29, 1305–1309. https://doi.org/10.1007/s10096-010-0991-2

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. https://doi.org/10.7554/eLife.08716

Rollins, M.F., Schuman, J.T., Paulus, K., Bukhari, H.S.T., Wiedenheft, B., 2015. Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. Nucleic Acids Res. 43, 2216–2222. https://doi.org/10.1093/nar/gkv094

Rostøl, J.T., Marraffini, L., 2019. (Ph)ighting Phages: How Bacteria Resist Their Parasites. Cell Host Microbe 25, 184–194. https://doi.org/10.1016/j.chom.2019.01.009

Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738. https://doi.org/10.1038/nprot.2010.5

Sakamoto, K., Agari, Y., Agari, K., Yokoyama, S., Kuramitsu, S., Shinkai, A., 2009. X-ray crystal structure of a CRISPR-associated RAMP superfamily protein, Cmr5, from Thermus thermophilus HB8. Proteins Struct. Funct. Bioinforma. 75, 528–532. https://doi.org/10.1002/prot.22358

Salis, H.M., Mirsky, E.A., Voigt, C.A., 2009. Automated design of synthetic ribosome binding sites to

control protein expression. Nat. Biotechnol. 27, 946–950. https://doi.org/10.1038/nbt.1568
Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A., Marraffini, L.A., 2015. Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. Cell 161, 1164–1174. https://doi.org/10.1016/j.cell.2015.04.027
Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.-L., Weiss, D.S., 2013. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature 497, 254–257. https://doi.org/10.1038/nature12048
Sampson, T.R., Weiss, D.S., 2013. Alternative Roles for CRISPR/Cas Systems in Bacterial Pathogenesis. PLoS Pathog. 9, e1003621. https://doi.org/10.1371/journal.ppat.1003621
Samson, J. E., Belanger, M., Moineau, S., 2013. Effect of the Abortive Infection Mechanism and Type III Toxin/Antitoxin System AbiQ on the Lytic Cycle of Lactococcus lactis Phages. J. Bacteriol. 195, 3947–3956. https://doi.org/10.1128/JB.00296-13
Samson, Julie E., Magadán, A.H., Sabri, M., Moineau, S., 2013. Revenge of the phages: defeating bacterial defences. Nat. Rev. Microbiol. 11, 675–687. https://doi.org/10.1038/nrmicro3096
Sasnauskas, G., Siksnys, V., 2020. CRISPR adaptation from a structural perspective. Curr. Opin. Struct. Biol. 65, 17–25. https://doi.org/10.1016/j.sbi.2020.05.015
Schmittgen, T.D., Livak, K.J., 2008. Analyzing real-time PCR data by the comparative CT method. Nat. Protoc. 3, 1101–1108. https://doi.org/10.1038/nprot.2008.73
Scholl, D., Adhya, S., Merril, C., 2005. Escherichia coli K1's Capsule Is a Barrier to Bacteriophage T7. Appl. Environ. Microbiol. 71, 4872–4874. https://doi.org/10.1128/AEM.71.8.4872-4874.2005
Schoonen, L., van Esterik, K.S., Zhang, C., Ulijn, R. V., Nolte, R.J.M., Hest, J.C.M. van, 2017. Alternative application of an affinity purification tag: hexahistidines in ester hydrolysis. Sci. Rep. 7, 14772. https://doi.org/10.1038/s41598-017-15310-y
Sefcikova, J., Roth, M., Yu, G., Li, H., 2017. Cas6 processes tight and relaxed repeat RNA via multiple mechanisms: A hypothesis. BioEssays 39, 1700019. https://doi.org/10.1002/bies.201700019
Semenova, E., Savitskaya, E., Musharova, O., Strotskaya, A., Vorontsova, D., Datsenko, K.A., Logacheva, M.D., Severinov, K., 2016. Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex. Proc. Natl. Acad. Sci. 113, 7626–7631. https://doi.org/10.1073/pnas.1602639113
Sethi, S., Sharma, N., Kakkar, N., Taneja, J., Chatterjee, S.S., Banga, S.S., Sharma, M., 2010. Increasing Trends of Leptospirosis in Northern India: A Clinico-Epidemiological Study. PLoS Negl. Trop. Dis. 4, e579. https://doi.org/10.1371/journal.pntd.0000579
Shah, S.A., Erdmann, S., Mojica, F.J.M., Garrett, R.A., 2013. Protospacer recognition motifs. RNA Biol. 10, 891–899. https://doi.org/10.4161/rna.23764
Shangguan, Q., Graham, S., Sundaramoorthy, R., White, M.F., 2022. Structure and mechanism of the type I-G CRISPR effector. Nucleic Acids Res. 50, 11214–11228. https://doi.org/10.1093/nar/gkac925
Shao, Y., Richter, H., Sun, S., Sharma, K., Urlaub, H., Randau, L., Li, H., 2016. A Non-Stem-Loop CRISPR RNA Is Processed by Dual Binding Cas6. Structure 24, 547–554. https://doi.org/10.1016/j.str.2016.02.009
Shipman, S.L., Nivala, J., Macklis, J.D., Church, G.M., 2016. Molecular recordings by directed CRISPR spacer acquisition. Science (80-.). 353. https://doi.org/10.1126/science.aaf1175
Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., Zhang, F., Koonin, E. V., 2015. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol. Cell 60, 385–397. https://doi.org/10.1016/j.molcel.2015.10.008
Shmakov, S.A., Sitnik, V., Makarova, K.S., Wolf, Y.I., Severinov, K. V., Koonin, E. V., 2017. The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes. MBio 8. https://doi.org/10.1128/mBio.01397-17
Shukla, S., Mittal, V., Karoli, R., Singh, P., Singh, A., 2022. Leptospirosis in central & eastern Uttar Pradesh, an underreported disease: A prospective cross-sectional study. Indian J. Med. Res. 155, 66. https://doi.org/10.4103/ijmr.IJMR_1811_19
Shuman, S., Glickman, M.S., 2007. Bacterial DNA repair by non-homologous end joining. Nat. Rev.

Microbiol. 5, 852–861. https://doi.org/10.1038/nrmicro1768
Silas, S., Mohr, G., Sidote, D.J., Markham, L.M., Sanchez-Amat, A., Bhaya, D., Lambowitz, A.M., Fire, A.Z., 2016. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase–Cas1 fusion protein. Science (80-.). 351. https://doi.org/10.1126/science.aad4234
Sistla, S., Rao, D.N., 2004. S-Adenosyl-L-methionine–Dependent Restriction Enzymes. Crit. Rev. Biochem. Mol. Biol. 39, 1–19. https://doi.org/10.1080/10409230490440532
Slamti, L., Picardeau, M., 2012. Construction of a Library of Random Mutants in the Spirochete Leptospira Biflexa Using a Mariner Transposon. pp. 169–176. https://doi.org/10.1007/978-1-61779-603-6_9
SMITH, J.K.G., YOUNG, M.M., WILSON, K.L., CRAIG, S.B., 2013. Leptospirosis following a major flood in Central Queensland, Australia. Epidemiol. Infect. 141, 585–590. https://doi.org/10.1017/S0950268812001021
Song, J.-J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R.A., Hannon, G.J., Joshua-Tor, L., 2003. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat. Struct. Mol. Biol. 10, 1026–1032. https://doi.org/10.1038/nsb1016
Song, S., Wood, T.K., 2020. A Primary Physiological Role of Toxin/Antitoxin Systems Is Phage Inhibition. Front. Microbiol. 11. https://doi.org/10.3389/fmicb.2020.01895
Staals, R.H.J., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K., Sakamoto, K., Suzuki, T., Dohmae, N., Yokoyama, S., Schaap, P.J., Urlaub, H., Heck, A.J.R., Nogales, E., Doudna, J.A., Shinkai, A., van der Oost, J., 2014. RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of Thermus thermophilus. Mol. Cell 56, 518–530. https://doi.org/10.1016/j.molcel.2014.10.005
Stamm, L. V, Charon, N.W., 1988. Sensitivity of pathogenic and free-living Leptospira spp. to UV radiation and mitomycin C. Appl. Environ. Microbiol. 54, 728–733. https://doi.org/10.1128/aem.54.3.728-733.1988
Stern, A., Sorek, R., 2011. The phage-host arms race: Shaping the evolution of microbes. BioEssays 33, 43–51. https://doi.org/10.1002/bies.201000071
Strand, K.R., Sun, C., Li, T., Jenney, F.E., Schut, G.J., Adams, M.W.W., 2010. Oxidative stress protection and the repair response to hydrogen peroxide in the hyperthermophilic archaeon Pyrococcus furiosus and in related species. Arch. Microbiol. 192, 447–459. https://doi.org/10.1007/s00203-010-0570-z
Strotksaya, A., Semenova, E., Savitskaya, E., Severinov, K., 2015. Rapid Multiplex Creation of Escherichia coli Strains Capable of Interfering with Phage Infection Through CRISPR. pp. 147–159. https://doi.org/10.1007/978-1-4939-2687-9_9
Suryawanshi, V.D., Walekar, L.S., Gore, A.H., Anbhule, P. V., Kolekar, G.B., 2016. Spectroscopic analysis on the binding interaction of biologically active pyrimidine derivative with bovine serum albumin. J. Pharm. Anal. 6, 56–63. https://doi.org/10.1016/j.jpha.2015.07.001
Suttle, C.A., 2007. Marine viruses — major players in the global ecosystem. Nat. Rev. Microbiol. 5, 801–812. https://doi.org/10.1038/nrmicro1750
Swarts, D.C., Hegge, J.W., Hinojo, I., Shiimori, M., Ellis, M.A., Dumrongkulraksa, J., Terns, R.M., Terns, M.P., van der Oost, J., 2015. Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA. Nucleic Acids Res. 43, 5120–5129. https://doi.org/10.1093/nar/gkv415
Swarts, D.C., Jinek, M., 2019. Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. Mol. Cell 73, 589-600.e4. https://doi.org/10.1016/j.molcel.2018.11.021
Swarts, Daan C., Jore, M.M., Westra, E.R., Zhu, Y., Janssen, J.H., Snijders, A.P., Wang, Y., Patel, D.J., Berenguer, J., Brouns, S.J.J., van der Oost, J., 2014. DNA-guided DNA interference by a prokaryotic Argonaute. Nature 507, 258–261. https://doi.org/10.1038/nature12971
Swarts, Daan 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. Nat. Struct. Mol. Biol. 21, 743–753. https://doi.org/10.1038/nsmb.2879
Szatmári, D., Sárkány, P., Kocsis, B., Nagy, T., Miseta, A., Barkó, S., Longauer, B., Robinson, R.C.,

Nyitrai, M., 2020. Intracellular ion concentrations and cation-dependent remodelling of bacterial MreB assemblies. Sci. Rep. 10, 12002. https://doi.org/10.1038/s41598-020-68960-w

Takeshita, D., Sato, M., Inanaga, H., Numata, T., 2019. Crystal Structures of Csm2 and Csm3 in the Type III-A CRISPR–Cas Effector Complex. J. Mol. Biol. 431, 748–763. https://doi.org/10.1016/j.jmb.2019.01.009

Tan, R., Krueger, R.K., Gramelspacher, M.J., Zhou, X., Xiao, Y., Ke, A., Hou, Z., Zhang, Y., 2022. Cas11 enables genome engineering in human cells with compact CRISPR-Cas3 systems. Mol. Cell 82, 852-867.e5. https://doi.org/10.1016/j.molcel.2021.12.032

Tock, M.R., Dryden, D.T., 2005. The biology of restriction and anti-restriction. Curr. Opin. Microbiol. 8, 466–472. https://doi.org/10.1016/j.mib.2005.06.003

Torgerson, P.R., Hagan, J.E., Costa, F., Calcagno, J., Kane, M., Martinez-Silveira, M.S., Goris, M.G.A., Stein, C., Ko, A.I., Abela-Ridder, B., 2015. Global Burden of Leptospirosis: Estimated in Terms of Disability Adjusted Life Years. PLoS Negl. Trop. Dis. 9, e0004122. https://doi.org/10.1371/journal.pntd.0004122

Tormo-Más, M.Á., Mir, I., Shrestha, A., Tallent, S.M., Campoy, S., Lasa, Í., Barbé, J., Novick, R.P., Christie, G.E., Penadés, J.R., 2010. Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. Nature 465, 779–782. https://doi.org/10.1038/nature09065

Traxler, R.M., Callinan, L.S., Holman, R.C., Steiner, C., Guerra, M.A., 2014. Leptospirosis-Associated Hospitalizations, United States, 1998–2009. Emerg. Infect. Dis. 20. https://doi.org/10.3201/eid2008.130450

Unterholzner, S.J., Poppenberger, B., Rozhon, W., 2013. Toxin–antitoxin systems. Mob. Genet. Elements 3, e26219. https://doi.org/10.4161/mge.26219

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., Velankar, S., 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50, D439–D444. https://doi.org/10.1093/nar/gkab1061

Vasu, K., Nagamalleswari, E., Nagaraja, V., 2012. Promiscuous restriction is a cellular defense strategy that confers fitness advantage to bacteria. Proc. Natl. Acad. Sci. 109, E1287–E1293. https://doi.org/10.1073/pnas.1119226109

Venclovas, Č., 2016. Structure of Csm2 elucidates the relationship between small subunits of CRISPR-Cas effector complexes. FEBS Lett. 590, 1521–1529. https://doi.org/10.1002/1873-3468.12179

Vestergaard, G., Garrett, R.A., Shah, S.A., 2014. CRISPR adaptive immune systems of Archaea. RNA Biol. 11, 156–167. https://doi.org/10.4161/rna.27990

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., Yanagihara, Y., Yoshida, S., Adler, B., 2009. Leptospirosis in the Asia Pacific region. BMC Infect. Dis. 9, 147. https://doi.org/10.1186/1471-2334-9-147

Vidakovic, L., Singh, P.K., Hartmann, R., Nadell, C.D., Drescher, K., 2017. Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. Nat. Microbiol. 3, 26–31. https://doi.org/10.1038/s41564-017-0050-1

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., Nakao, R., Amara Korba, A., Bourhy, P., Veyrier, F.J., Picardeau, M., 2019. Revisiting the taxonomy and evolution of pathogenicity of the genus Leptospira through the prism of genomics. PLoS Negl. Trop. Dis. 13, e0007270. https://doi.org/10.1371/journal.pntd.0007270

Walker, J.E., Lanahan, A.A., Zheng, T., Toruno, C., Lynd, L.R., Cameron, J.C., Olson, D.G., Eckert, C.A., 2020. Development of both type I–B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium Clostridium thermocellum. Metab. Eng. Commun. 10, e00116. https://doi.org/10.1016/j.mec.2019.e00116

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, 840–853.

https://doi.org/10.1016/j.cell.2015.10.008
Wang, Z., Jin, L., Węgrzyn, A., 2007. Leptospirosis vaccines. Microb. Cell Fact. 6, 39. https://doi.org/10.1186/1475-2859-6-39
Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303. https://doi.org/10.1093/nar/gky427
Wei, Y., Terns, R.M., Terns, M.P., 2015. Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation. Genes Dev. 29, 356–361. https://doi.org/10.1101/gad.257550.114
Weinbauer, M.G., 2004. Ecology of prokaryotic viruses. FEMS Microbiol. Rev. 28, 127–181. https://doi.org/10.1016/j.femsre.2003.08.001
Wigley, D.B., 2013. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. Nat. Rev. Microbiol. 11, 9–13. https://doi.org/10.1038/nrmicro2917
Williams, R.J., 2003. Restriction Endonucleases: Classification, Properties, and Applications. Mol. Biotechnol. 23, 225–244. https://doi.org/10.1385/MB:23:3:225
Winther, K., Tree, J.J., Tollervey, D., Gerdes, K., 2016. VapCs of Mycobacterium tuberculosis cleave RNAs essential for translation. Nucleic Acids Res. 44, 9860–9871. https://doi.org/10.1093/nar/gkw781
Wörtz, J., Smith, V., Fallmann, J., König, S., Thuraisingam, T., Walther, P., Urlaub, H., Stadler, P.F., Allers, T., Hille, F., Marchfelder, A., 2022. Cas1 and Fen1 Display Equivalent Functions During Archaeal DNA Repair. Front. Microbiol. 13. https://doi.org/10.3389/fmicb.2022.822304
Wright, A. V., Nuñez, J.K., Doudna, J.A., 2016. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell 164, 29–44. https://doi.org/10.1016/j.cell.2015.12.035
Wright, A. V, Doudna, J.A., 2016. Protecting genome integrity during CRISPR immune adaptation. Nat. Struct. Mol. Biol. 23, 876–883. https://doi.org/10.1038/nsmb.3289
Xiao, B., Johnson, R.C., Marko, J.F., 2010. Modulation of HU–DNA interactions by salt concentration and applied force. Nucleic Acids Res. 38, 6176–6185. https://doi.org/10.1093/nar/gkq435
Xiao, G., Yi, Y., Che, R., Zhang, Q., Imran, M., Khan, A., Yan, J., Lin, X., 2019a. Characterization of CRISPR-Cas systems in Leptospira reveals potential application of CRISPR in genotyping of Leptospira interrogans. APMIS 127, 202–216. https://doi.org/10.1111/apm.12935
Xiao, G., Yi, Y., Che, R., Zhang, Q., Imran, M., Khan, A., Yan, J., Lin, X., 2019b. Characterization of CRISPR-Cas systems in Leptospira reveals potential application of CRISPR in genotyping of Leptospira interrogans. APMIS 127, 202–216. https://doi.org/10.1111/apm.12935
Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M., Ke, A., 2017a. Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. Cell 170, 48-60.e11. https://doi.org/10.1016/j.cell.2017.06.012
Xiao, Y., Ng, S., Nam, K.H., Ke, A., 2017b. How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. Nature 550, 137–141. https://doi.org/10.1038/nature24020
Xu, X., Qi, L.S., 2019. A CRISPR–dCas Toolbox for Genetic Engineering and Synthetic Biology. J. Mol. Biol. 431, 34–47. https://doi.org/10.1016/j.jmb.2018.06.037
Xu, Z., Li, Y., Li, M., Xiang, H., Yan, A., 2021. Harnessing the type I <scp>CRISPR-Cas</scp> systems for genome editing in prokaryotes. Environ. Microbiol. 23, 542–558. https://doi.org/10.1111/1462-2920.15116
Yamano, T., Zetsche, B., Ishitani, R., Zhang, F., Nishimasu, H., Nureki, O., 2017. Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1. Mol. Cell 67, 633-645.e3. https://doi.org/10.1016/j.molcel.2017.06.035
Yan, W.X., Hunnewell, P., Alfonse, L.E., Carte, J.M., Keston-Smith, E., Sothiselvam, S., Garrity, A.J., Chong, S., Makarova, K.S., Koonin, E. V., Cheng, D.R., Scott, D.A., 2019. Functionally diverse type V CRISPR-Cas systems. Science (80-.). 363, 88–91. https://doi.org/10.1126/science.aav7271
Yang, J., Zhang, Y., 2015. I-TASSER server: new development for protein structure and function

predictions. Nucleic Acids Res. 43, W174–W181. https://doi.org/10.1093/nar/gkv342

Yang, W., 2011. Nucleases: diversity of structure, function and mechanism. Q. Rev. Biophys. 44, 1–93. https://doi.org/10.1017/S0033583510000181

Yoganand, K.N., Muralidharan, M., Nimkar, S., Anand, B., 2019. Fidelity of prespacer capture and processing is governed by the PAM-mediated interactions of Cas1-2 adaptation complex in CRISPR-Cas type I-E system. J. Biol. Chem. 294, 20039–20053. https://doi.org/10.1074/jbc.RA119.009438

Yoganand, K.N.R., Sivathanu, R., Nimkar, S., Anand, B., 2017. Asymmetric positioning of Cas1–2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. Nucleic Acids Res. 45, 367–381. https://doi.org/10.1093/nar/gkw1151

Yosef, I., Goren, M.G., Qimron, U., 2012. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res. 40, 5569–5576. https://doi.org/10.1093/nar/gks216

Yosef, I., Shitrit, D., Goren, M.G., Burstein, D., Pupko, T., Qimron, U., 2013. DNA motifs determining the efficiency of adaptation into the Escherichia coli CRISPR array. Proc. Natl. Acad. Sci. 110, 14396–14401. https://doi.org/10.1073/pnas.1300108110

Yu, Y.T., Snyder, L., 1994. Translation elongation factor Tu cleaved by a phage-exclusion system. Proc. Natl. Acad. Sci. 91, 802–806. https://doi.org/10.1073/pnas.91.2.802

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. Mol. Cell 19, 405–419. https://doi.org/10.1016/j.molcel.2005.07.011

Zavala-Alvarado, C., Sismeiro, O., Legendre, R., Varet, H., Bussotti, G., Bayram, J., G. Huete, S., Rey, G., Coppée, J.-Y., Picardeau, M., Benaroudj, N., 2020. The transcriptional response of pathogenic Leptospira to peroxide reveals new defenses against infection-related oxidative stress. PLOS Pathog. 16, e1008904. https://doi.org/10.1371/journal.ppat.1008904

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., Koonin, E. V., Zhang, F., 2015. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell 163, 759–771. https://doi.org/10.1016/j.cell.2015.09.038

Zhang, B., Ye, Y., Ye, W., Perčulija, V., Jiang, H., Chen, Y., Li, Y., Chen, J., Lin, J., Wang, S., Chen, Q., Han, Y.-S., Ouyang, S., 2019. Two HEPN domains dictate CRISPR RNA maturation and target cleavage in Cas13d. Nat. Commun. 10, 2544. https://doi.org/10.1038/s41467-019-10507-3

Zhang, C., Wang, H., Yan, J., 2012. Leptospirosis prevalence in Chinese populations in the last two decades. Microbes Infect. 14, 317–323. https://doi.org/10.1016/j.micinf.2011.11.007

Zhang, F., 2019. Development of CRISPR-Cas systems for genome editing and beyond. Q. Rev. Biophys. 52, e6. https://doi.org/10.1017/S0033583519000052

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, 147–150. https://doi.org/10.1038/nature13733

Zheng, Y., Han, J., Wang, B., Hu, X., Li, R., Shen, W., Ma, X., Ma, L., Yi, L., Yang, S., Peng, W., 2019. Characterization and repurposing of the endogenous Type I-F CRISPR–Cas system of Zymomonas mobilis for genome engineering. Nucleic Acids Res. 47, 11461–11475. https://doi.org/10.1093/nar/gkz940

Zhou, Y., Bravo, J.P.K., Taylor, H.N., Steens, J.A., Jackson, R.N., Staals, R.H.J., Taylor, D.W., 2021. Structure of a type IV CRISPR-Cas ribonucleoprotein complex. iScience 24, 102201. https://doi.org/10.1016/j.isci.2021.102201

*****

List of publications

Publication from thesis research work

1. **Hussain, M.S.** and Kumar, M., 2022. Assembly of Cas7 subunits of *Leptospira* on the mature crRNA of CRISPR-Cas I-B is modulated by divalent ions. *Gene*, *818*, p.146244.

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

Publications from other collaborative research work

1. Dhara, A., **Hussain, M.S.**, Datta, D. and Kumar, M., 2019. Insights to the assembly of a functionally active leptospiral ClpP1P2 protease complex along with its ATPase chaperone ClpX. *ACS omega*, *4*(7), pp.12880-12895.

2. Ghosh, K.K., Prakash, A., Dhara, A., **Hussain, M.S.**, Shrivastav, P., Kumar, P., Balamurugan, V. and Kumar, M., 2019. Role of supramolecule ErpY-like lipoprotein of *Leptospira* in thrombin-catalyzed fibrin clot inhibition and binding to complement factors H and I, and its diagnostic potential. *Infection and Immunity*, *87*(12), pp.e00536-19.

3. Dhara, A., **Hussain, M.S.**, Kanaujia, S.P. and Kumar, M., 2021. Acyldepsipeptide activated ClpP1P2 macromolecule of *Leptospira*, an ideal Achilles' heel to hamper the cell survival and deregulate ClpP proteolytic activity. *Research in microbiology*, *172*(2), p.103797.

4. Dixit, B., Anand, V., **Hussain, M.S.** and Kumar, M., 2021. The CRISPR-associated Cas4 protein from *Leptospira interrogans* demonstrates versatile nuclease activity. *Current Research in Microbial Sciences*, *2*, p.100040.

5. Hota, S., **Hussain, M.S.** and Kumar, M., 2022. ErpY-like Lipoprotein of *Leptospira* Outsmarts Host Complement Regulation by Acquiring Complement Regulators, Activating Alternative Pathways, and Intervening in the Membrane Attack Complex. *ACS Infectious Diseases*, *8*(5), pp.982-997.

6. Anand, V., Prabhakaran, H.S., Prakash, A., **Hussain**, **M.S.**, Kumar, M. Differential processing of CRISPR RNA by LinCas5c and LinCas6 of *Leptospira*. **(Under revision)**

7. Banerjee, R., Mukherjee, A., Adhikary, A., **Hussain, M.S.**, Sharma, S., Ali, M.E. and Nagotu, S. Insights into the role of the conserved GTPase domain residues T62 and S277 in the structure-function relationship of yeast Dnm1. **(Under revision)**

Book chapter

1. **Hussain, M.S.** and Kumar, M., 2022. CRISPR/Cas System and Stem Cell Editing: Prospects and Possibilities in Veterinary Sciences. In *Stem Cells in Veterinary Science* (pp. 323-354). Singapore: Springer Nature Singapore.

*****

List of conferences/workshops participated/attended

Conferences

1. **M.S. Hussain**, M. Kumar "Biochemical characterization of CRISPR-Cas Cascade complex subunit protein-Cas7 in *Leptospira interrogans* Copenhageni Strain Fiocruz L1- 130", Research conclave'19 (2019). [Organized by IIT Guwahati, Guwahati, India (March 14-17, 2019)] **(Poster presentation)**

2. **M.S. Hussain**, M. Kumar "Biochemical characterization of Cas7 subunit of CRISPR-Cas I-B Cascade complex subunit in *Leptospira interrogans* Copenhageni Strain Fiocruz L1-130", 2nd National biomedical research competition, 2019 (NBRCom-2019). [PGIMER, Chandigarh, India (November 17, 2019)] **(Oral presentation)**

3. **M.S. Hussain**, M. Kumar "*Leptospira* Cas8b protein: a genetic fusion of large and small subunit of CRISPR-Cas type I-B Cascade complex", Emerging areas in Biosciences and Biomedical Technologies-2 (eBBT-2), P.P 46-46, (2020). [IIT Indore, Indore, India (February 7-9, 2020)] **(Poster presentation)**

4. Virtual symposium on "COVID-19 Disease Control - Opportunities and Challenges for Vaccines, Bio-therapeutics and Diagnostics". [Centre for Bio-Separation Technology (CBST), Vellore Institute of Technology (VIT), Vellore, India (July 9-10, 2020] **(Attended)**

5. Virtual conference on "A Paradigm Shift in Infectious Diseases: Lessons from Pandemic." [Merck KGaA, Germany (September 1-2, 2020)]. **(Attended)**

6. **M.S. Hussain**, M. Kumar "The burgeoning CRISPR-Cas applications," Biotechnological Approaches in Animal Research and Disease Diagnosis (2021). [GADVASU, Ludhiana, India (February 1-12, 2021), Virtual] **(Invited lecture)**

7. M. Kumar**, M.S. Hussain**, A. Prakash, V. 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 International Leptospirosis Society Conference. P.P 37, (2022). [Bangkok, Thailand (November 13-16, 2022)] **(Oral presentation by thesis supervisor)**

8. **M.S. Hussain**, M. Kumar "Molecular and functional insight of CRISPR-Cas type I-B interference complex in pathogenic *Leptospira interrogans*," International symposium on zoonotic and transboundary diseases: breaking the chain through multidisciplinary approach" and XVIIIth annual conference of IAVPHS. P.P 168-169, (2022). [ICAR research complex for NEH region, Meghalaya, India (December 1-2, 2022)] **(Poster presentation)**

Workshops

1. International seminar cum workshop on "Computer aided drug design for human pathogens (CADDHP)". [Tezpur University, Tezpur, India (February 12-17, 2018)]

2. International training course on "Biotechnological approaches in animal research and disease diagnosis." [GADVASU, Ludhiana, India (February 1-12, 2021), Virtual]

*****