Probing the unusual architecture of CRISPR machinery and its functional implications

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

by

Sunanda Chhetry

Registration no: 156106010

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

April 2024

I dedicate this thesis to

my family and friends.

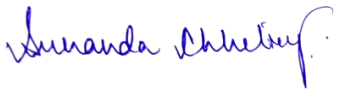
Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

Statement

I do hereby declare that the matter embodied in this thesis entitled "**Probing the unusual architecture of CRISPR machinery and its functional implications**" is the result of work carried out in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, under the supervision of **Prof. B. Anand**.

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



Sunanda Chhetry

156106010

Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

Certificate

It is certified that the work described in this thesis entitled "**Probing the unusual architecture of CRISPR machinery and its functional implications**" by **Mrs. Sunanda Chhetry** for the award of the degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati. The work embodied in this thesis has not been submitted elsewhere for a degree.

Prof. B. Anand

Thesis Advisor

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati 781039, INDIA

Acknowledgement

Having the opportunity to compose an acknowledgement is both a profound privilege and a significant responsibility. It signifies nearing the culmination of this thrilling chapter in my scientific journey while bearing the responsibility of recognizing all those whose assistance has facilitated this path of learning, growth, and self-reflection.

I extend my sincere appreciation to my thesis supervisor, Prof. B. Anand for his exemplary mentorship, invaluable guidance, and unwavering commitment throughout the journey of this research endeavour. His patience, encouragement, and scholarly insights have been pivotal in shaping the trajectory of this study and refining its scientific rigor.

I would like to express my gratitude to the esteemed members of my Doctoral committee, namely Prof. Vibin Ramakrishnan, Prof. Senthilkumar Sivaprakasam and Dr. Priyadarshi Satapati from the Department of Biosciences and Bioengineering for their invaluable suggestions. Their scholarly engagement and insightful suggestions have undoubtedly elevated the quality of this research.

I would like to thank the Heads of the Department, Prof. V. Venkata Dasu, Prof. Kannan Pakshirajan, Prof. Latha Rangan and Prof. Rakhi Chaturvedi as well as the other faculty members for lending their support and advice during my tenure.

I also extend my heartfelt appreciation to all the past and the present members of the MAB lab for fostering a warm and enriching environment throughout my time there. I would like to thank Rohan, Neha, Lopamudra, Siddhartha and Pratyusha for their constant support and insights towards the successful completion of this thesis. Special recognition is due towards Sudipta, Sandip, and Akanksha for their invaluable assistance in Chapters 4 and 5 of this thesis. I wish them all the luck for the bright future that awaits them ahead.

Furthermore, I extend my gratitude to the Department of Biosciences and Bioengineering, along with its staff, for their indispensable logistical support, which enabled the smooth execution of my research endeavors. I thank Mr. Nurul Islam, Mrs. Prarthana Swargari, Mr. Niranjan Barah, Mr Dipankar Barman and Mr Chandan Kumar Nath for their constant support. I would also like to thank Mrs. Namrata Naomi Rynjah for the immense support and comfort that she had provided me throughout this journey. Additionally, my

appreciation goes out to the Central Instrumentation Facility and the Department of Chemistry at IITG for granting access to essential scientific instruments. I also wish to acknowledge the generosity of all the researchers who shared their valuable research materials and reagents.

I am immensely grateful to my alma mater, IIT Guwahati, for providing me with the opportunity to embark on this enriching journey of learning. With its abundant facilities and the captivating natural beauty of the North-east, IITG has served as an ideal environment for academic growth and exploration. Next, I would like to thank the funding agencies such as Department of Biotechnology, Govt. of India, Department of Science and Technology, GOI, and Science and Engineering Research Board, GOI, that provided the financial support for the research grants provided to our lab. I would also like to acknowledge MHRD, GOI for the fellowship during my Ph.D. tenure.

The journey I have embarked upon wouldn't have been possible without the unwavering support and love of my dear friends, who have made every step of this voyage truly unforgettable. Foremost, I extend my heartfelt gratitude to Siddharth, Manasa, Payal, Arun, and Yoganand, who have not only been exceptional colleagues but also wonderful friends. I am deeply indebted to Himakshi and Amay for their steadfast presence during both my triumphs and tribulations. This endeavor owes its success to the incredible support of friends such as Akanksha, Paromita, Manisha, Sanjana, Monjusha, Heera, Tori, Bhagyashree, Vartika, Ganesh, Subbi, Ekramul, and Juan, who stood by me with encouragement throughout this expedition. Additionally, I am blessed to have individuals like Kiran, Sibojyoti, Pervez, Sandip, Suraj, Siddharth, and Soupal, who have showered me with love akin to that of siblings. All these people have provided me with their tireless support, encouragement and love, that have served as the driving force behind my journey. The camaraderie shared among us as colleagues and friends has left an indelible mark, and I will forever cherish the memories we have created together.

Above all, I extend my deepest gratitude to my husband, Himanshu, for accompanying me on this journey from its very inception. He has not only been the finest colleague and mentor but also the epitome of an exemplary husband. His unwavering presence, support, and guidance through every challenge and triumph over the years have been unparalleled. I feel truly blessed to have him in my life.

Finally, I wish to express my deepest appreciation to my parents, my in-laws, my brother Raj and my entire family. Their boundless love, encouragement, and steadfast belief in my aspirations have been the cornerstone of my academic pursuit. I owe a debt of gratitude beyond measure. It is their countless personal sacrifices and unwavering support that have propelled me to this significant milestone in life.

Lastly, I express my gratitude to all others that I may have missed here.

May God bless all the helping hands.

Thank you.

Sunanda Chhetry

Table of Contents

List of Figures

Abbreviations

CRISPR – Clustered Regularly Interspersed Short Palindromic Repeats

Cas – CRISPR- associated protein

MGEs – Mobile genetic elements

crRNA – CRISPR RNA

dsDNA – Double-stranded DNA

ssDNA – Single-stranded DNA

ATP – Adenosine triphosphate

6-FAM – 6-Carboxyfluorescein

CTD – C-Terminal Domain

WT – Wild type

EtBr – Ethidium bromide

PAGE – Polyacrylamide gel electrophoresis

SEC – Size exclusion chromatography

Kan^R – Kanamycin Resistance

$Spec^R$ – Spectinomycin Resistance

Amp^R –Ampicillin Resistance

Cmp^R – Chloramphenicol Resistance

Chapter 1: Introduction

1.1 Introduction

Life on Earth has developed into a vast range of forms, from prokaryotic single-celled organisms like bacteria to complex eukaryotic multicellular beings like animals and plants (Brasier, McLoughlin, Green, & Wacey, 2006). This diversity of life forms is a testament to the power of natural selection to drive the evolution of life over billions of years. Such changes have been theorized as per the Red Queen hypothesis where exposure to multiple biotic and abiotic challenges has propelled the evolution of primordial organisms to not only greater biological diversity but also to newer and higher forms of life (Paterson et al., 2010).

In this race of survival of the fittest, organisms are subjected to – amongst other challenges – constant threat from their predators. Even the most primitive life forms such as prokaryotes have been exposed to these challenges (Field, Behrenfeld, Randerson, & Falkowski, 1998) and one such threat is posed by an even more abundant life form – Phages (Edwards & Rohwer, 2005). Phages are viruses that lack cellular machinery for replication or protein synthesis for their fundamental survival. Thus, in order to thrive in nature they rely on the host cellular machinery. Based on their type of life cycle they can either infect a host or reside in them for an indefinite amount of time, lyse the host and recommence infection on newer cells.

Research in the field of bacteria-phage interactions is driven by various factors. Firstly, the curiosity to understand the role of phages in driving rapid evolution (Gregory et al., 2019). Secondly, the emergence of antibiotic resistance has prompted exploration into phage-based therapies as alternative antibacterial treatments (Dedrick et al., 2019; Sulakvelidze, Alavidze, & Morris, 2001). Understanding how bacterial pathogens may develop resistance to phages is crucial for the success of such therapies. Lastly, the need for phage-resistant strains in fermentation industries (O'Sullivan, Bolton, McAuliffe, & Coffey, 2019). This research not only supports industrial applications but also contributes to advancements in gene editing and diagnostics (Foss, Hochstrasser, & Wilson, 2019). The importance of bacterial immune systems in these areas has spurred a renewed interest in discovering and characterizing phage-resistance mechanisms. Two extensively investigated bacterial immune systems include the restriction-modification system (R-M system) and abortive infection (Abi) systems. However, contemporary research has recently redirected its attention to the CRISPR-Cas system. This exploration of bacterial immunity has led to the realization that prokaryotic immune

mechanisms are significantly more intricate than initially understood. In addition to these well-known systems, a multitude of novel defense systems continues to be unveiled regularly, revealing previously unknown facets of prokaryotic defense mechanisms (Hampton, Watson, & Fineran, 2020; Teklemariam et al., 2023).

1.1.1 Bacterial defence systems

Bacteria have evolved mechanistically complex defense methods that act at every stage of the phage life cycle to survive the continual onslaught of phage. The virus-host arms race drives the emergence of diverse defense systems. Most of these systems are guided by a few key mechanisms like the rapid evolution of genetic sequences, immense duplication of genes, and horizontal gene transfer through plasmids that carry the defense gene. There exists a potential positive relationship between the diversity of viruses in an environment and the prevalence of defense systems. The study of the genes involved in this system are typically found aggregated in genomic islands, also known as Defence Islands. The DIs (defense islands) exhibit a notable enrichment of putative operons and consist of numerous gene families that are overrepresented. In-depth sequence analysis of the proteins encoded by these overrepresented genes in DIs revealed that many of them represent divergent versions of known defense system components. Whereas, certain genes that exhibit distinct features, such as characteristic operonic organization, suggested the presence of novel defense systems (Makarova, Wolf, Snir, & Koonin, 2011). These antiviral defense mechanisms operate based on two main principles – firstly the ability to discriminate between self and non-self and secondly infection-induced dormancy, or cell suicide. Based on their ability to recall the memory of the viral encounter, these systems are broadly classified as innate and adaptive immune systems. A few such systems are described below.

1.1.2 Innate immune response

Innate immune responses are a category of immune responses that have been naturally present in organisms. They are an essential part of the immune system as they provide immediate defense against various pathogens and foreign substances. Unlike adaptive immune responses, which develop over time and have a memory of previous infections, innate immune

systems do not possess a memory of prior encounters with pathogens. These responses act as the first line of defense and help to contain and control infections until the adaptive immune system can generate a more specific and targeted response.

1.1.2.1 Surface exclusion

The initiation of a successful phage infection relies on the viral attachment to a specific receptor on the surface of the bacteria. These receptors are typically comprised of either polysaccharides or lipopolysaccharides (LPS) or porins (or other proteins), and it is equally important that these receptors are accessible to the phage particles for attachment (Bertozzi Silva, Storms, & Sauvageau, 2016; Steven et al., 1988). Thus, as a defense strategy, bacteria either may attempt to alter the molecular structure of the receptor through mutations or shield the receptor with additional barriers (Letarov & Kulikov, 2017; Seed, 2015). For instance, the outer-membrane lipoprotein TraT, which is encoded by the F plasmid, plays an important role in impeding phage infections. TraT disrupts phage attachment by interacting with outer-membrane protein A (OmpA), which acts as the entry receptor for several T-even-like *E. coli* phages. This modification prevents or reduces the attachment of phages to the host cell and inhibits their infection (Riede & Eschbach, 1986). Similarly, *Bordetella bronchiseptica* exhibits two distinct phases, namely Bvg^+ and Bvg^-, where the former is essential for pulmonary colonization as it expresses various surface proteins, including pertactin. It has been observed that phage particles attach to pertactin on the bacterial surface. In the Bvg^- strain, where pertactin is absent, phage attachment to the bacterial surface is impeded or significantly reduced (M. Liu et al., 2002).

1.1.2.2 Superinfection exclusion

Once the phage successfully attaches to the bacterial surface, superinfection exclusion (Sie) systems come into play, blocking the entry of phage DNA into the bacterial cytoplasm. Superinfection exclusion (Sie) systems are protein-based defenses that provide immunity against specific phages by hindering the entry of other phage DNA into host cells, as a result, providing bacterial immunity against superinfecting phages. These proteins are expected to be

either anchored to the cell membrane or associated with membrane components (Labrie, Samson, & Moineau, 2010). For instance, the *E. coli* phage HK97 produces a transmembrane protein called gp15, which effectively inhibits the entry of other HK97 phages, as well as a closely related phage, HK75 (Cumby, Edwards, Davidson, & Maxwell, 2012).

1.1.2.3 Restriction -Modification system

Once the viral DNA is successfully introduced into the bacteria, the host deploys various mechanisms to prevent phage replication and assembly. One of these mechanisms is the restriction-modification (R-M) system, which acts to destroy invading DNA. R-M systems exhibit two distinct activities: (a) The restriction endonuclease (REase) that recognizes and cleaves unmethylated DNA. (b) The methyltransferase (MTase) that modifies DNA bases of a specific sequence by transferring a methyl group to both strands (Tock & Dryden, 2005).

The self vs non-self discrimination happens at the level of methylation. The R-M systems typically recognize short DNA sequences of 4 to 8 base pairs. As these sequences may be present in both the host and attacking viruses, the host methylates and modifies its DNA using the MTase. This self-methylated DNA is not recognized by the host restriction endonuclease (REase). However, the unmethylated DNA of the attacking phage is cleaved upon entry into the host's cytoplasm.

R-M systems are highly diverse and are categorized into four types based on their sequence recognition and cleavage mechanisms. The Type I R-M system is a hetero-oligomeric complex system comprising three subunits: R (restriction), M (modification), and S (specificity). The R subunit includes the active site responsible for ATP hydrolysis and endonuclease activity. The M subunit functions as a methylase, modifying the host nucleic acid by adding methyl groups. Meanwhile, the S (specificity) subunit contains two target recognition domains (TRDs), which play a crucial role in conferring target sequence specificity to both the R and M subunits (Murray, 2000). The Type II R-M system has found widespread use in various biological applications as it is the simplest of all the R-M systems. It comprises of a restriction endonuclease (REase) and a methyltransferase (MTase), which are typically encoded within the same operon (Tock & Dryden, 2005). On the other hand, the Type III R-M system is similar to the Type II system as it comprises of two subunits R and M, with an ATP-

dependent helicase domain fused to the R subunit. They require two copies of their unmethylated, short asymmetric recognition site to be present in the target DNA, in an inverted repeat orientation (Meisel, Bickle, Kruger, & Schroeder, 1992). Type IV restriction enzymes differ from other types of R-M systems as they only contain a restriction enzyme (R) and lack a methylase (M). Unlike the other types, the presence of specific modifications, such as methylation, hydroxymethylation, or glucosyl-hydroxymethylation, serves as the recognition signal for Type IV restriction enzymes, making them highly selective in their DNA cleavage activity (Loenen & Raleigh, 2014). The R subunit of this system functions as both a GTPase and an endonuclease (Bourniquel & Bickle, 2002).

1.1.2.4 Abortive infection

All the above-mentioned strategies of viral defence focus on the survival of the host. Contrarily, abortive infection (Abi) is a bacterial defense mechanism by which it limits the spread of the phage particles via apoptosis. Abi systems are linked to mobile genetic elements such as prophages and plasmids. The cellular suicide can be triggered either by altering the cell membrane or by affecting internal cellular processes like transcription and translation. However, the specific mechanism behind the Abi (abortive infection) process is not well comprehended (Samson, Magadan, Sabri, & Moineau, 2013).

Lactobacillus lactis harbours almost over 20 Abi systems namely, AbiA - AbiZ. AbiP operates at an early stage in the phage replication cycle, disrupting both phage DNA replication and the transition from early to late gene expression (Domingues, Chopin, Ehrlich, & Chopin, 2004). Another example of the Abi system harboured by *Staphylococcus* species against the *Siphoviridae* which involves the Stk2 kinase gene, which is activated by a phage protein called PacK, which has been hypothesized to be involved in phage DNA packaging. Once activated, Stk2 phosphorylates multiple proteins within the cell, ultimately resulting in cell death (Depardieu et al., 2016).

1.1.2.5 Toxin-antitoxin system

Toxin-antitoxin (TA) systems are a widely spread defense mechanism in bacteria that comprise of a toxin gene and its corresponding antitoxin. In these systems, the toxins are always proteins, whereas the antitoxins can be either proteins or non-coding RNAs. The toxin gene hinders cell proliferation by interfering with essential cellular processes, whereas the antitoxin gene neutralizes the toxin's harmful effects (Unterholzner, Poppenberger, & Rozhon, 2013). TA modules are currently categorized into eight classes based on the molecular nature of the antitoxin and how it interacts with the toxin (Qiu, Zhai, Wei, Zheng, & Jiao, 2022). For example, *Erwinia carotovora* subspecies *atroseptica* harbours a ToxIN system where ToxN is a toxin protein that inhibits bacterial growth. Its cognate antitoxin, ToxI, is a non-coding RNA that inhibits ToxN's activity by forming a complex with it. When a phage attack occurs, ToxN's RNase activity leads to the destruction of both host and phage RNA. Consequently, neither the host nor the phage can survive, halting the infection (Fineran et al., 2009). *Mycobacterium tuberculosis* harbours a DarTG TA system, wherein DarT toxins have the capability to ADP-ribosylate single-stranded DNA, leading to the inhibition of phage DNA synthesis. DarG antitoxins possess ADP-glycohydrolase activity, enabling them to enzymatically reverse the modifications induced by their corresponding DarT toxins. Upon phage infection, the release of the DarT toxin is triggered, and as a consequence, a DNA damage response is initiated in cells (LeRoux et al., 2022).

1.1.2.6 Argonaute

Argonautes (Ago) were initially identified as an RNA-mediated gene silencing (RNAi) system in eukaryotes (eAgo). Similar homologs of eAgo proteins were later discovered in prokaryotes (pAgo) (Aravind, Watanabe, Lipman, & Koonin, 2000). The presence of the genes encoding the pAgo proteins on the defence island indicated their role in the defence, which was later studied to be involved in the recognition and nuclease-mediated cleavage of foreign genetic material using single-stranded RNA/DNA (Makarova, Wolf, van der Oost, & Koonin, 2009). Most pAgos, exhibit a bilobal structure comprising of four domains. The N- terminal and PAZ (PIWI-Argonaute-Zwille) domains form one lobe, while the MID (Middle) and PIWI (P-element Induced Wimpy Testis) domains constitute the other lobe (Lisitskaya, Aravin, &

Kulbachinskiy, 2018). For instance, *Thermus thermophilus* Ago (TtAgo) is an active endonuclease, that protects against foreign nucleic acids through the process of DNA-guided DNA interference (Swarts et al., 2014).

1.1.2.7 Novel bacterial defence systems

In recent times numerous previously unknown defense systems were identified from a comprehensive study examining defense islands in various bacteria and archaea. A few of them are described below.

1.1.2.7.1 The Thoeris

The thoeris defense system is a two-component based system where the first gene identified is *thsA*, which encodes a nicotinamide adenine dinucleotide (NAD) binding domain. The second component is *thsB* which comprises a Toll-interleukin receptor (TIR) domain. TIR domain plays a crucial role in certain innate immune systems found in higher organisms. Similarly, the thoeris system functions by transmitting molecular signals by ThsB, within the cell after detecting the presence of a pathogen which activates the NAD^+ degrading activity of ThsA. This finally leads to the depletion of the host NAD^+ pool and eventually to cell death by Abi (Doron et al., 2018; Ofir et al., 2021).

1.1.2.7.1 The Zorya

Based on the genetic composition, the Zorya system is classified into two types. Type I Zorya system is a four-component system that is encoded by the *zorABCD* operon, out of which ZorA and ZorB were found to be homologs of inner membrane protein that constitute the flagellar motor of bacteria, MotA and MotB. ZorC is a domain of unknown function (DUF), whereas ZorD harbours helicase domain. Whereas, the type II Zorya type is a three-component system that lacks *zorC* and *zorD* and instead harbours an additional HNH nuclease *zorE*. The defense system functions via ZorA and ZorB mediated membrane depolarisation, leading to cell suicide during phage invasion (Doron et al., 2018).

1.1.2.7.2 The Gabija

Gabija is a two-component system comprising of GajA, an ATP-dependent TOPRIM domain containing nuclease and GajB, a DNA helicase (GajB). At the time of phage attack, the process of phage DNA replication and transcription consumes cellular NTPs and dNTPs. This decline leads to the loss of activity of the ATP binding domain of GajA which in turn activates the TOPRIM domain. The activated TOPRIM domain then facilitates the cleavage of phage DNA and may also participate in the destruction of bacterial genomic DNA as part of the abortive infection response. Whereas, the role of GajB is hypothesised to be in either activating GajA or aiding in its nuclease activity (Cheng et al., 2021).

1.1.2.7.3 The Druantia

This system is characterized by a gene that encodes a remarkably large protein (~2000 amino acids), DurA, that contains a domain of unknown function (DUF) along with a helicase signature and a Walker A/B motif, which implies the potential utilization of ATP. This large gene is found to be commonly preceded by a cluster of highly variable genes that lack recognizable domains or predicted functions (Doron et al., 2018).

1.1.2.7.4 The Hachiman

The Hachiman system is a two-component system comprising of HamA, a domain of unknown function (DUF, Pfam accession: 08878) and HamB, a helicase (Doron et al., 2018).

1.1.2.7.5 The Wadjet

The Wadjet system is unique in its dedicated role as an anti-plasmid defense system. It comprises of four genes, *jetABCD*, out of which the proteins JetA, JetB, and JetC were found to show sequence similarity to subunits of the bacterial condensin complex MukBEF. Whereas, JetD is found to show distant homology to TOPRIM domain containing nucleases. The system is known to target plasmid DNA by DNA loop extrusion followed by DNA cleavage (Deep et al., 2022).

Apart from these novel defense mechanisms, numerous operons are also found on the defense islands. For instance, the Shedu is a SduA nuclease-based system. The Septu system harbours an ATP-dependent helicase (PtuA) and an HNH nuclease (PtuB). The Lamassu system consists of a DUF protein (LmuA) and an ATPase (LmuB). The Kiwa system contains genes encoding the DUF domain (KwaB) and a transmembrane protein (KwaA) (Doron et al., 2018). These newly identified genes on defense islands expand our knowledge of the diverse array of defense mechanisms bacteria possess. However, to fully comprehend their mechanisms and functions, a systematic study is required, and further research is needed.

1.1.3 The adaptive immune response

The significance of an adaptive immune system lies in its ability to establish an immunological memory of the initial encounter with a pathogen. This memory results in an amplified immune response upon subsequent invasions by the same pathogen. In prokaryotes like bacteria and archaea, the CRISPR-Cas system operates as a distinctive defense mechanism where a fragment of the foreign mobile genetic element is integrated as immunological memory which later aids the host in evading the subsequent attack by the same pathogen. Until now, the CRISPR-Cas system stands as the sole known adaptive immune system in prokaryotes. It highlights the remarkable ability of bacteria to acquire and utilize defense strategies similar to RNA interference (RNAi) machinery in eukaryotes.

1.1.3.1 The discovery of the CRISPR-Cas system

The earliest reports of documentation of an array of repeat sequences adjacent to the alkaline phosphatase (*iap*) gene in *E. coli*, were made in the year 1987 (Ishino, Shinagawa, Makino, Amemura, & Nakata, 1987). In the years 1989 and 1991, a similar series of repeats were reported to be found in Enterobacteria (*Shigella dysenteriae* and *Salmonella enterica*) (Nakata, Amemura, & Makino, 1989) and *Mycobacterium tuberculosis* (Hermans et al., 1991), respectively. Bioinformatics analysis of many bacterial and archaeal genomes showed the presence of these repeats, which were then termed as short regularly spaced repeats (SRSSs) (F. J. Mojica, Diez-Villasenor, Soria, & Juez, 2000). It was in the year 2002, when it was found that these SRSSs were variable in size (21-37 nucleotides) and were interspaced by non-repetitive unknown sequences. These arrays of repeats were then termed as CRISPR that stands

for Clustered Regularly Interspaced Short Palindromic Repeats. It was also found that certain specific genes were located in the vicinity of these arrays, which were absent in the genomes lacking CRISPR. These genes were termed as CRISPR-associated or *cas* genes and together the CRIPSR array and the *cas* genes comprised of the CRISPR-Cas loci (Jansen, Embden, Gaastra, & Schouls, 2002). Firstly, only four *cas* genes were identified, namely *cas1, cas2, cas3 and cas4* whose functions were not clear. However, upon further investigation, the *cas* operon was found to contain various components, including a superfamily II helicase (Cas3'), HD-type phosphohydrolase (Cas3''), and RecB family exonuclease (Cas4), among others. The presence of these components suggests that they might play roles in DNA repair processes (Makarova, Aravind, Grishin, Rogozin, & Koonin, 2002) or regulation of gene expression (Jansen et al., 2002). It was in the year 2005 that three groups independently showed that the non-repetitive DNA that stretch the repeat regions were derived from foreign mobile genetic elements (MGEs) such as phage genetic material or plasmids (Bolotin, Quinquis, Sorokin, & Ehrlich, 2005; F. J. Mojica, Diez-Villasenor, Garcia-Martinez, & Soria, 2005; Pourcel, Salvignol, & Vergnaud, 2005).

1.1.3.2 Comparison between the CRISPR-Cas system and RNA interference

Alongside, there were several studies made to understand this system which pointed towards the similarities it posed with the eukaryotic RNA interference (RNAi) system (Makarova, Grishin, Shabalina, Wolf, & Koonin, 2006). The fact supporting this similarity between the CRISPR system and RNA interference (RNAi) stems from the functional roles of the proteins found in CRISPR loci and the components of the RNA-induced silencing complex (RISC) involved in RNAi systems (Filipowicz, 2005; Makarova et al., 2006). The RISC is constituted of a Dicer, Argonaute, TRBP (Transactivating Response RNA-Binding Protein), and sRNA (Hammond, 2005; Iwakawa & Tomari, 2022; Sontheimer, 2005). Initially, it was hypothesized that Dicer serves as an analog of a superfamily 2 helicase fused to an HD family nuclease, which is represented by Cas3 in the CRISPR system. Similarly, the RecB family nuclease, known as Cas4 in the CRISPR loci, was believed to function similarly to the slicer in the CRISPR system.

In the RNAi pathway of eukaryotes, short interfering RNA (siRNA), microRNA (miRNA), or piwi-interfering RNA (piRNA) serves as a guide to recognise and target specific nucleic acids (Carthew & Sontheimer, 2009; Iwakawa & Tomari, 2022; Scadden, 2005). Similarly, in the

CRISPR system, a guide RNA is responsible for directing Cas proteins to recognize and cleave complementary foreign DNA, thereby acting as an analog to the RNAi pathway in eukaryotes (Brouns et al., 2008; Carte, Wang, Li, Terns, & Terns, 2008). Additionally, both systems exhibit an adaptive immune response against foreign attack. While the RNAi pathway and the CRISPR system share conceptual similarities, there are some distinct functional differences between them. The CRISPR guide RNA is matured from a single pre-CRISPR RNA transcript and is not amplified, as opposed to the miRNA or siRNAs of the RNAi system (Brouns et al., 2008; Carte et al., 2008; C. Hale, Kleppe, Terns, & Terns, 2008).

1.1.4 CRISPR-Cas system

CRISPR-Cas is an adaptive immune system found exclusively in prokaryotes (almost 36% of sequenced bacteria and 75% of sequenced archaea) (Couvin et al., 2018; Pourcel et al., 2020; Rousseau, Gonnet, Le Romancer, & Nicolas, 2009). However, it was also observed that <10% of the uncultured bacterial species harboured CRISPR-Cas defense mechanism (Burstein et al., 2016). It is hypothesised that this can be due to two factors that might guide CRISPR incidence – thermophilicity and aerobicity, which means that CRISPR incidence differs with a difference in temperature (high incidence in mesophiles than in cryophiles) and availability of oxygen (anaerobes preferred over aerobes) (Weissman, Laljani, Fagan, & Johnson, 2019).

CRISPR-Cas system is primarily composed of two components – a nucleic acid component called the CRISPR array and a protein component called Cas proteins. The CRISPR array is composed of a repeating unit of short palindromic DNA sequences, usually ranging from 25 to 40 nucleotides in length. These repeats are interspaced by similarly sized segments of foreign DNA or viral DNA fragments called spacers. Typically, spacers in CRISPR arrays are known to originate from foreign DNA, displaying homology to plasmids, viruses, or other invading genetic elements (Figure 1.1). However, it is noteworthy that in some instances, a few spacers have been observed to exhibit homology with sequences within the host genome itself (Devi, Harjai, & Chhibber, 2022; Stern, Keren, Wurtzel, Amitai, & Sorek, 2010). Whereas, the protein component of this system is termed as CRISPR-associated or Cas proteins. These Cas proteins are responsible for carrying out the molecular tasks required to acquire, identify, target, and neutralize foreign genetic material.

The functionality of the CRISPR-Cas system is achieved in three stages, namely adaptation, maturation and interference (Figure 1.1). In the first step, which referred as adaptation- the foreign DNA fragments from invading pathogens are acquired and incorporated

into the CRISPR array that allows the cell to store the memory of the encounter. In the second stage, which is referred as maturation, the CRISPR array consisting of the repeat-spacer units, is transcribed into a long precursor RNA which is then processed by Cas protein(s), to generate mature CRISPR RNAs (crRNAs). These mature crRNAs associate with the Cas protein(s) to form a ribonucleoprotein (RNP) surveillance complex. Finally, in the interference step, the crRNA guides the RNP complex to specifically recognize and bind to complementary sequences in foreign genetic material. Upon target recognition, Cas proteins with endonuclease activity, cleave the invading DNA or RNA, rendering it non-functional and neutralizing the pathogen's genetic material. The subsequent sections shall delve into detailed discussions of each stage in the CRISPR-Cas system.

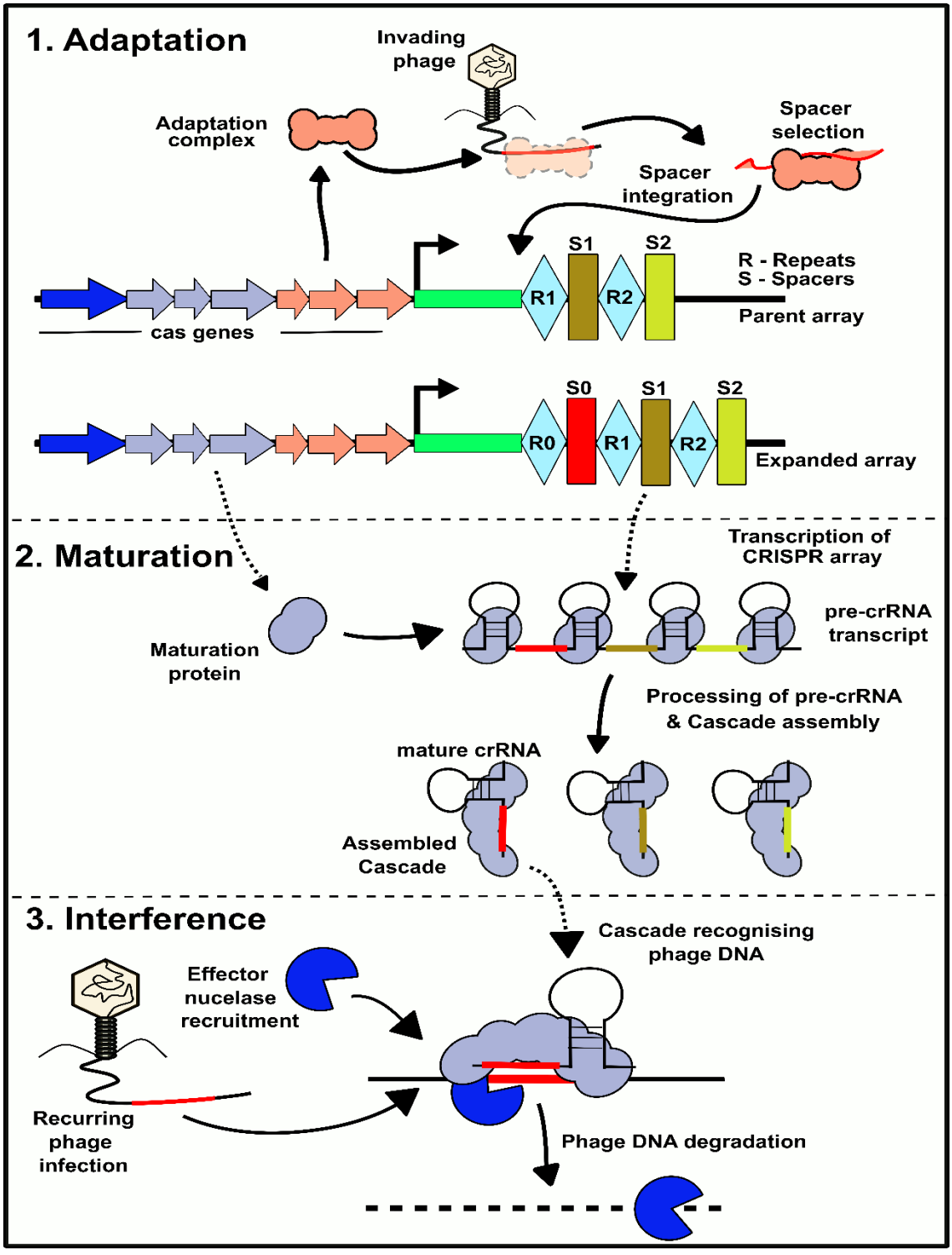


Figure 1. 1: CRISPR-Cas system- An Overview.

A schematic representation of the various stages of the CRISPR-based adaptive immunity. A typical CRISPR-Cas locus comprises of a sequence of conserved repeats separated by variable spacers. Positioned upstream of the CRISPR array is the 'leader' region (green bar), housing the promoter region responsible for CRISPR-RNA expression. Additionally, CRISPR-associated genes (*cas* genes) are found in the vicinity of the array. **(1) Adaptation:** The first stage is adaptation where the adaptation complex recognises the PAM sequence in the phage DNA/prespacer. These prespacers are integrated into the CRISPR array towards the leader proximal end followed by repeat duplication (R0) and expansion of the CRISPR array. Here, S0 denotes the newly acquired spacer, whereas, S1 and S2 denote the first and second spacers respectively, and so on. Similarly, R1 denotes

the first repeat and so on. **(2) Maturation:** This is the second step, where the CRISPR array is transcribed to produce pre-crRNA which is further processed by the Cas proteins to yield mature crRNAs. They subsequently assemble into a nucleoprotein Cascade complex. **(3) Interference:** Finally, in the third stage, the invading phage DNA is recognised by the Cascade complex that recruits an effector nuclease to cleave the target DNA.

1.1.5 CRISPR-Cas system - Classification

Comparable to other biological defense mechanisms, CRISPR-Cas systems exhibit a noteworthy diversity in their Cas protein sequences, gene compositions, and genomic locus architectures (Barrangou & Horvath, 2017; Hille & Charpentier, 2016; Ishino, Krupovic, & Forterre, 2018; Klompe & Sternberg, 2018; Koonin, Makarova, & Wolf, 2017; Mohanraju et al., 2016; Wright, Nunez, & Doudna, 2016). The ever-expanding understanding of this diversity increases with extensive genomic and metagenomic databases. To keep up with this expansion, several classification schemes of CRISPR-Cas systems based on their evolutionary relationships were proposed. Two such classification schemes from 2011 and 2015, utilized a multifaceted approach, where they integrated comparisons of CRISPR–Cas gene compositions and genomic loci architectures with sequence similarity-based clustering and phylogenetic analysis of conserved Cas proteins, to establish a comprehensive classification of the CRISPR–Cas systems (Makarova, Haft, et al., 2011; Makarova et al., 2015). In the recently updated classification, a computational approach was employed in the identification of the signature genes, comparison of genomic loci and compositions, sequence similarity-based clustering and phylogenetic analysis, neighbourhood analysis and comparison, and experimental evidence, by which researchers were able to categorize and classify the diverse CRISPR-Cas systems effectively (Makarova et al., 2020). The *cas* genes within the CRISPR-Cas systems are categorized into four distinct functional modules based on the roles played by Cas proteins (Figure 1.2). The first module is the Adaptation module, that is involved in the spacer integration process. Within this module, there is a CRISPR array and three *cas* genes: *cas1*, *cas2*, and *cas4* (Figure 1.2). The second module is the Expression module, which encompasses genes responsible for the processing of crRNA (Figure 1.2). The third one is the Interference module, where one can find genes that encode subunits of the effector complex (Figure 1.2). Lastly, the fourth module groups together all other ancillary proteins that aid in the mechanism of CRISPR-Cas mediated immunity (Figure 1.2). Based on the composition of the effector complex, the CRISPR-Cas system is broadly classified into two classes – Class 1 which

comprises of multi-subunit effector complex, and Class 2 comprising of a single subunit effector complex.

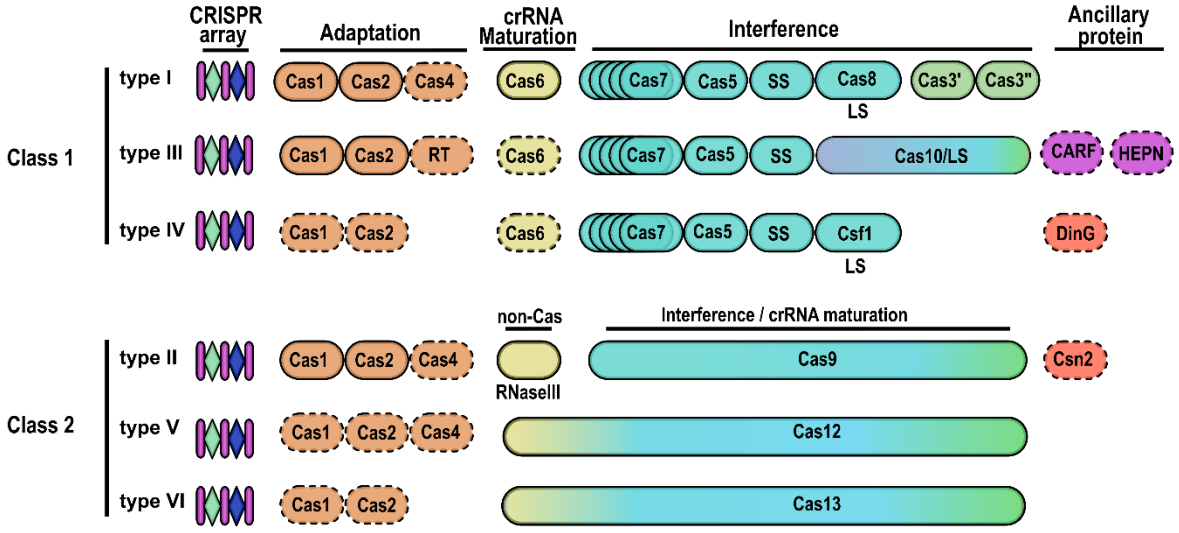


Figure 1. 2: Classification of the CRISPR-Cas system

A schema representing the classification of the CRISPR-Cas system into two different classes, based on the composition of the effector complex. The further division of the classes into various types is based on the presence of a signature protein. Class 1 CRISPR-Cas system consists of a multi-subunit effector complex that takes part in interference (colored in cyan and light green). Class 2 CRISPR-Cas system consists of a single effector protein that is Cas9, Cas12 and Cas13 in the case of type II, type V and type VI, respectively. The proteins colored in beige are important in maturation whereas those colored in peach are required for adaptation. The additional/ancillary proteins are colored in magenta and coral. SS and LS indicate small and large subunits, respectively. The dotted lines indicate dispensable or missing components.

1.1.5.1 Class 1 CRISPR-Cas system

This class of CRISPR-Cas system is categorised based on the composition of the multi-subunit effector complex. The major Cas proteins involved are Cas5, Cas6, Cas7, Cas8, Cas10 and Cas11 in varied combinations. Based on the locus architecture and repertoire of *cas* genes, Class 1 is further divided into three types, namely, type I, type III and type IV. Type I and III are the most abundant types of CRISPR in bacteria and archaea. This classification of CRISPR-Cas systems into specific types, such as Type I or Type III, is based on the presence of signature genes, such as *cas3* and *cas10*, respectively. However, *csf1* gene was proposed to be the signature gene for the type IV system (Makarova et al., 2015). But, recent studies have found that few type IV systems lack the *csf1* gene. Thus, later, *csf2, a cas7* homolog was entailed to be the signature gene of this type (Pinilla-Redondo et al., 2020; H. N. Taylor et al., 2021). Few

subtypes of type IV CRISPR-Cas system, lack the adaptation module. The multi-subunit effector complex in these systems typically consists of paralogous Cas proteins belonging to the Repeat Associated Mysterious Proteins (RAMPs) family. The RAMP family is characterized by the presence of RNA Recognition Motif (RRM) containing proteins that are commonly found in bacteria and archaea, and they are primarily associated with CRISPR loci (R. Wang & Li, 2012). The backbone of the effector complex comprises of a multimeric Cas protein, usually Cas7 (type I) or its homologs like Csf2 (type IV) and Csm3/Cmr4 (type III). The backbone stabilizes the crRNA guide which is bound by single subunit Cas proteins like Cas5, Cas6, Csm5/Cmr1, Csm4/Cmr3, or Csf5 (H. N. Taylor et al., 2021; Wright et al., 2016). Besides these, the effector complex harbours Cas proteins that interact with the guide RNA as well as other Cas proteins that bring about stability to the complex. They also comprise of Cas proteins that recognize the foreign genetic element based on the PAM sequence and interact with the effector nuclease for target recognition and degradation. Class 1 CRISPR-Cas effector complex mainly targets DNA with the exception of type III, which targets both DNA as well as RNA. A structural comparison of the available effector complex indicates a shared evolutionary origin for these proteins (Jackson & Wiedenheft, 2015; Makarova, Aravind, Wolf, & Koonin, 2011; Makarova et al., 2015).

1.1.5.1.1 Type I CRISPR-Cas system

The characteristic feature of the type I system is the presence of its signature gene *cas3* which encodes for the effector nuclease Cas3. Based on the locus architecture, type I system is further divided into seven subtypes (type I-A to I-G) (Figure 1.3). Most of these subtypes harbour a common adaptation module comprising of Cas1, Cas2 and Cas4, with the exception of types I-E and I-F which lack Cas4, that plays a crucial role in the prespacer processing and in type I-G where it is fused to Cas1 (Almendros, Nobrega, McKenzie, & Brouns, 2019; C. Hu et al., 2021; Makarova et al., 2020). The process of maturation in most of type I systems is executed by Cas6 protein, with an exception in type I-C where in the absence of Cas6, Cas5 assumes the role (Gesner, Schellenberg, Garside, George, & Macmillan, 2011; Nam et al., 2012; Punetha, Sivathanu, & Anand, 2014). In type I-G, the Cas5 and Cas6 are fused together (Makarova et al., 2015). The interference module of type I consists of Cas3, which is also the signature Cas protein of this system. It is intriguing to note that Cas3 exhibits variability in

domain architecture among different subtypes. A standard Cas3 protein consists of an N-terminal HD - nuclease followed by a DExD/H box superfamily 2 helicase domain and a C-terminal domain. However, in Type I-A systems, the two functional domains of Cas3 (nuclease and helicase) are separated, yet they work in coordination during the interference process (Majumdar & Terns, 2019; Makarova et al., 2015). The domains in type I-G Cas3 are rearranged, and it contains N-terminal helicase followed by C-terminal nuclease (Makarova et al., 2015; Makarova et al., 2020). Also, in type I-D, the HD-nuclease domain is fused with the large subunit protein Cas10, whereas, in type I-F Cas3 is fused with Cas2 protein (Makarova et al., 2015) (Figure 1.3).

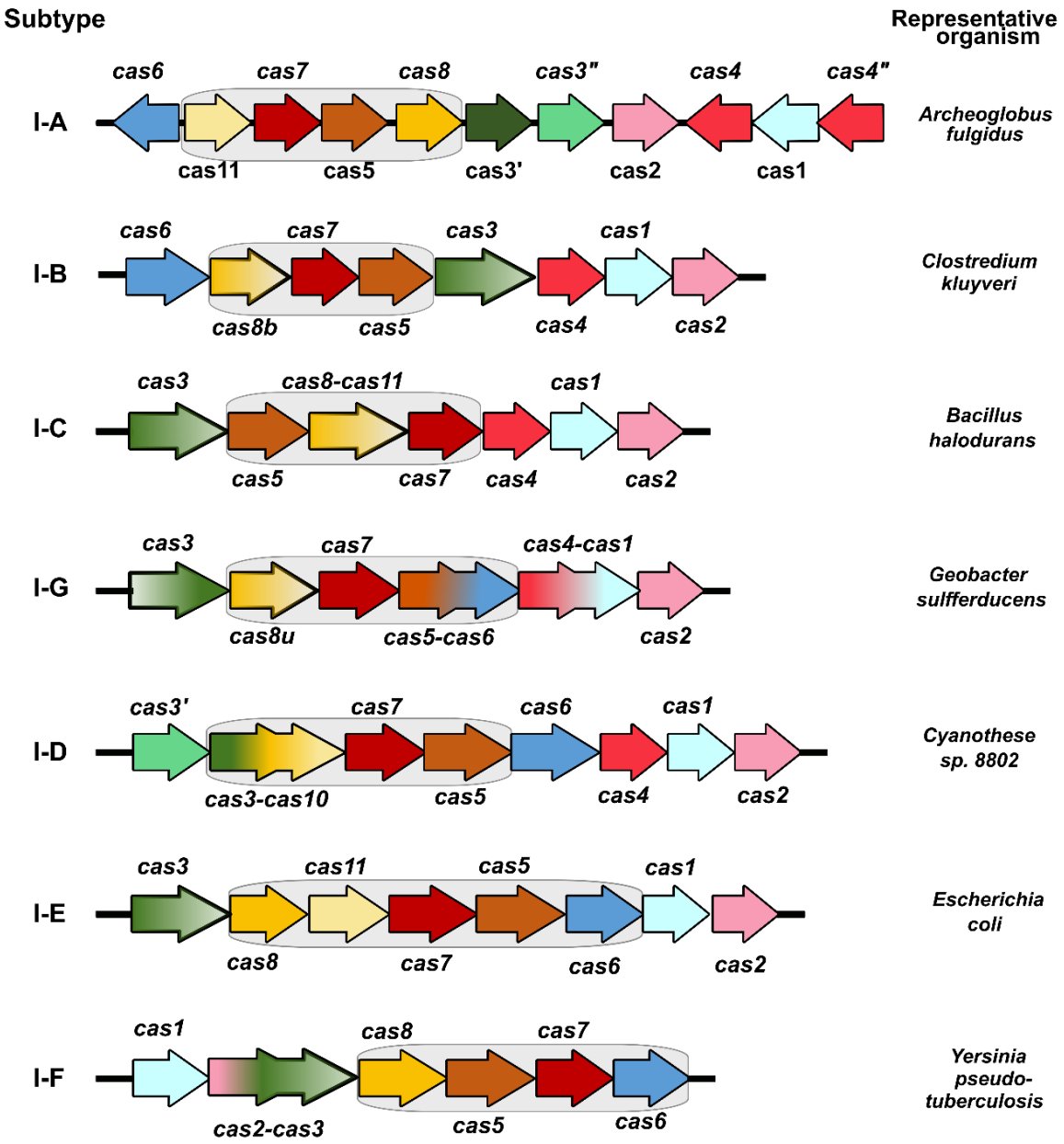


Figure 1. 3: Type I CRISPR-Cas classification

Schematic representation of the locus architecture of the seven subtypes encompassed under the type 1 system, against their representative organisms. The signature protein of the type I CRISPR-Cas system is Cas3, which comprises of an N-terminal nuclease domain (dark green) and a C-terminal helicase domain (light green). The orientation of the same is reversed in type I-G and hence depicted using a reversed colour gradient. The effector module has been highlighted in light grey boxes.

1.1.5.1.2 Type III CRISPR-Cas system

The hallmark of the Type III CRISPR-Cas system is the presence of an HD-type nuclease Cas10 protein. Cas10 serves as the largest subunit in the Type III system and contains a palm domain that includes both cyclase and polymerase activities (Osawa, Inanaga, & Numata, 2013; Osawa, Inanaga, Sato, & Numata, 2015; R. H. Staals et al., 2014; R. H. J. Staals et al., 2013). Similar to type I subtypes, type III systems also display variations in CRISPR loci architecture, and are classified into six subtypes (type III-A to III-F) (Makarova et al., 2020). Type III CRISPR-Cas systems, except for type III-A and III-E, typically lack a CRISPR array and may share the CRISPR array with other subtypes present in the same organism. Additionally, most of the type III systems also lack the adaptation module and instead acquire it from other subtypes present in the same organism (Kolesnik, Fedorova, Karneyeva, Artamonova, & Severinov, 2021; Makarova et al., 2015). Type III CRISPR-Cas system can target RNA, whose binding to the effector complex also activates the DNase activity of Cas10 that recognises and cleaves the DNA component associated within the transcriptosome (Elmore et al., 2016; Estrella, Kuo, & Bailey, 2016; Kazlauskiene, Tamulaitis, Kostiuk, Venclovas, & Siksnys, 2016). In addition to the *cas* genes, the type III CRISPR-Cas system is linked to accessory proteins that are implicated in membrane transport or signal transduction processes. These accessory proteins possess two distinctive domains: the CRISPR -Associated Rossmann Fold (CARF) domain and the Higher Eukaryotes Prokaryote Nucleotide-binding (HEPN) domain. Notably, this system can produce a signaling molecule known as cyclic oligoadenylates (cOA). When cOAs bind to the CARF domain, it activates the HEPN domain, leading to the indiscriminate cleavage of RNA (Jia, Jones, Yang, Ouerfelli, & Patel, 2019; Molina et al., 2019). Additionally, the recently identified Type III-E system encodes a potential caspase-like protease, TPR-CHAT (also known as Csx29), along with accessory proteins Csx30, Csx31, and RpoE. This gRAMP Cas7-11 protein associates with crRNA and TPR-CHAT, forming a stable complex termed Craspase (CRISPR-guided caspase) that initiates an RNA-guided and protease-based immune response as an alternative to the conventional system (C. Hu et al., 2022; G. Huo, Shepherd, & Pan, 2023; Strecker et al., 2022; X. Wang et al., 2022; H. Yang & Patel, 2022).

1.1.5.1.3 Type IV CRISPR-Cas system

Type IV CRISPR-Cas systems possess certain discernible characteristics from other CRISPR-Cas systems and are still not fully characterized. The hallmark of this system is the Cas7 homolog, Csf2. Similar to its homolog, Csf2 also binds the guide RNA within a helical backbone (H. N. Taylor et al., 2021). Type IV CRISPR-Cas system is further classified into five subtypes: type IV-A-E (Pinilla-Redondo et al., 2020). Most of the organisms harbouring this system also lack the adaptation module as well as an effector nuclease. Additionally, they are primarily encoded by extra-chromosomal genetic elements like plasmids (Faure et al., 2019; H. N. Taylor et al., 2021). They mostly target dsDNA damage but studies also indicate its role in RNA targeting (Zhou et al., 2021). The effector complex of type IV systems comprises of Csf2 (Cas7), Csf3 (Cas5), and Csf1 (large subunit) (Makarova et al., 2020; Ozcan et al., 2019; Pinilla-Redondo et al., 2020). Type IV-A and IV-E have an ancillary DinG helicase – Csf4 that helps in unwinding the target DNA, as well as a Cas6 homolog – Csf5 (Faure et al., 2019; Makarova et al., 2020; S. A. Shmakov, Makarova, Wolf, Severinov, & Koonin, 2018). The type IV-B encodes a putative signal transduction protein CysH-like that is a homolog of the CysH alongside a putative Cas11 (small subunit) (Faure et al., 2019; S. A. Shmakov et al., 2018). Type IV-D systems carry a helicase of the RecD family instead of the archetypal DinG (Pinilla-Redondo et al., 2020). Interestingly, the structure of Csf2 from type IV-C has similarities to Cmr4/Csm3, the Cas7-like helical backbone subunit in type III-A/B system (Pinilla-Redondo et al., 2020).

1.1.5.2 Class 2 CRISPR-Cas system

All the CRISPR-Cas systems that harbour a single subunit effector complex are categorized under the Class2 system. This property of Class 2 further aids in establishing several tools for genetic manipulation and gene silencing. The understanding of this class is ever expanding, since its discovery. Similar to Class 1, Class 2 is also further subdivided into three types, namely type II, V and VI, based on their difference in locus architecture (Makarova et al., 2020). The variants of class 2 differ majorly with respect to the mechanisms of pre-crRNA processing (Chylinski, Makarova, Charpentier, & Koonin, 2014; Chyou & Brown, 2019).

1.1.5.2.1 Type II CRISPR-Cas system

Type II system harbours Cas9, which is its signature gene. Cas9 effector is a bilobed structure made of a Recognition lobe (REC) and a Nuclease lobe (NUC). The NUC lobe contains two juxtaposed nuclease domains, responsible for target DNA cleavage; RuvC endonuclease and HNH nuclease domain (Chylinski et al., 2014; Jinek et al., 2014). Apart from the Cas9 effector nuclease, type II adaptation module comprises of Cas1 and Cas2. A few of its variants also harbour the Cas4 endonuclease required for the adaptation process, while in the rest, Csn2 replaces Cas4 (Heler et al., 2015; Nussenzweig, McGinn, & Marraffini, 2019; Wei, Terns, & Terns, 2015). The guide RNA here is also accompanied by an additional RNA component called as the tracrRNA that has a complementarity sequence to the repeats in the CRISPR array, which upon binding to the former induces an RNaseIII-mediated cleavage event, which results in the processing and maturation of the crRNA (Deltcheva et al., 2011). Finally, in the interference stage, the effector complex that comprises of Cas9, tracrRNA and crRNA, recognises the target DNA, and the nuclease domains of Cas9 cleave both strands of DNA (Garneau et al., 2010; Gasiunas, Barrangou, Horvath, & Siksnys, 2012; Jinek et al., 2012). Type II is further divided into three subtypes: Type II-A, II-B and II-C (Makarova et al., 2020).

1.1.5.2.2 Type V CRISPR-Cas system

One of the most diverse CRISPR types, type V comprises of the signature protein Cas12, which similar to Cas9, is also bilobed harbouring both REC and NUC lobes. The NUC lobe lacks the HNH domain but contains a RuvC endonuclease domain that cleaves target DNA. The diversity in the locus architecture further divides this system into eleven subtypes (V-A to I, V-K and CRISPR-CasΦ) (Tong et al., 2020). The adaptation module of type V system consists of Cas1, Cas2 and Cas4 with an exception in types V-C, D, G, H, I, K and CRISPR-CasΦ, suggesting a shared adaptation module with other existing CRISPR loci (Makarova et al., 2020; Tong et al., 2020; Zetsche et al., 2017). Additionally, certain types (V-F, V-G, V-H, V-K and Φ) contain a relatively smaller version of Cas12 protein (Tong et al., 2020). The Cas12-crRNA effector complex typically targets double-stranded DNA (dsDNA) but recent studies have revealed that it can also target only ssDNA (type V-F), both dsDNA and ssDNA (type V-A, B, C, D, H, I and Φ) as well as RNA targets too (type V-G). Except for Cas12k (type V-K), Cas12 shows trans cleavage activity leading to collateral damage (Chen et

al., 2019; Harrington et al., 2018; Harrington et al., 2020; J. J. Liu et al., 2019; Pausch et al., 2020; S. Shmakov et al., 2015; Strecker et al., 2019; Yan et al., 2019; Zetsche et al., 2017). In contrast, Cas12k lacks CRISPR-based DNA cleavage activity (both *cis* and *trans*) and it is associated with transposons that facilitate crRNA-dependent transposition (Park et al., 2023; Strecker et al., 2019; Yan et al., 2019).

1.1.5.2.3 Type VI CRISPR-Cas system

Type VI CRISPR-Cas system is the most minimal system amongst the other types and harbours the signature protein Cas13. Cas13 also harbours a REC and NUC lobe, where the NUC lobe is comprised of two HEPN domains (L. Liu, Li, Wang, et al., 2017). Based on the difference in locus architecture, this system is further subdivided into four subtypes, type VI-A to VI-D. The adaptation module is lacking in types VI-B and VI-C, whereas types VI-A and VI-D contain Cas1 and Cas2, but, lack Cas4 (Makarova et al., 2020). This system is primarily an RNA-dependent RNA targeting system, which also shows collateral damage due to indiscriminate host targeting (Abudayyeh et al., 2016; Perculija, Lin, Zhang, & Ouyang, 2021). Type VI-B harbours additional ancillary proteins like Csx27 and Csx28 that regulate the interference activity of Cas13 (Smargon et al., 2017). Similarly, type VI-D consists of a WYL-domain containing ancillary protein that enhances the activity of Cas13 (Yan et al., 2018).

1.1.6 CRISPR adaptation

Adaptation is a process in which the variable DNA sequence coming from the foreign mobile genetic elements, is acquired into the CRISPR array. These variable DNA sequences called spacers, are the basis of the CRISPR immunological memory. It has been also observed that few spacers are acquired within the host genome itself (Devi et al., 2022; Levy et al., 2015). There can be 2 to more than 200 variable spacers in a CRISPR array, and thus, these variable spacers can provide immunity against a wide range of MGEs (S. A. Shmakov et al., 2017).

The process of adaptation within the CRISPR-Cas system is intricate and involves several stages, leading to the selection and incorporation of a segment from a Mobile Genetic Element (MGE) into the CRISPR array. Initially, Cas proteins discern a brief sequence pattern, known as the protospacer-adjacent motif (PAM), consisting of 2 to 7 nucleotides on the invasive MGE. Subsequently, this specific DNA segment is transformed into a sequence spanning 20 to 40 base pairs. This alteration is accomplished through collaboration between

Cas proteins and, in certain prokaryotes, host exonucleases (Barrangou et al., 2007; Garneau et al., 2010; Marraffini & Sontheimer, 2010b; Yoganand, Muralidharan, Nimkar, & Anand, 2019; Yosef, Goren, & Qimron, 2012). The abbreviated segment of the MGE is referred to as "prespacers," which are precursors to the actual spacers. In the final phase of this process, the altered DNA fragment is incorporated into the CRISPR array with the assistance of Cas integrase. The DNA that gets integrated into the array are thereafter termed as spacers. The entire arrangement of repeats and spacers is termed as CRISPR array. The array is preceded by a leader region that harbours the promoter element required for the transcription of the array. Typically, the process of adaptation is polarised in nature, in which the new spacers are incorporated towards the leader proximal region of the CRISPR array (Barrangou et al., 2007; Diez-Villasenor, Guzman, Almendros, Garcia-Martinez, & Mojica, 2013; Erdmann & Garrett, 2012; Yosef et al., 2012). Sometimes, multiple spacers can be acquired from a phage genome, consequently heightening the potential for interference in scenarios where specific segments of phage DNA undergo mutations, which in turn helps to pace up with constantly evolving phages (Paez-Espino et al., 2013; R. H. Staals et al., 2016; van Houte et al., 2016). The detailed process of adaptation is discussed in the upcoming sections.

There are two types of adaptation - naïve adaptation, where spacers are acquired from completely new MGEs. This process is gradual, and the spacers are procured without any specific pattern. Given that MGEs undergo constant mutations, the process of naïve adaptation is inadequate for countering mutated MGEs effectively. To address this challenge, spacers are swiftly acquired from mutated MGEs via a mechanism known as primed adaptation. Both the mechanism of naïve and primed adaptation are explained briefly in the subsequent sections.

1.1.6.1 Self and non-self-discrimination during adaptation

The functioning of the CRISPR-Cas system as an adaptive immune mechanism relies on its ability to discern between self and non-self elements. However, self-targeting in case of CRISPR-Cas immunity is a potential threat as the process of adaptation might lead to the acquisition of host genome sequence too (Levy et al., 2015; Wei et al., 2015; Yosef et al., 2012). Additionally, the crRNA that is complementary to the CRISPR array, might also lead to autoimmunity (Bikard, Hatoum-Aslan, Mucida, & Marraffini, 2012; Edgar & Qimron, 2010; W. Jiang, Bikard, Cox, Zhang, & Marraffini, 2013). To prevent such an autoimmune response, the CRISPR system employs multiple mechanisms to tilt its specificity in favour of recognizing

and targeting foreign MGEs. Indeed, non-self recognition in CRISPR immunity has been observed both during spacer acquisition and target degradation.

In the case of CRISPR adaptation, the Cas1-2 complex acquires DNA substrates from the breakdown products resulting from RecBCD (in gram -ve bacteria) and AddAB (in gram +ve bacteria) activity during double-stranded DNA (dsDNA) breaks repair (Ivancic-Bace, Cass, Wearne, & Bolt, 2015; Levy et al., 2015; Modell, Jiang, & Marraffini, 2017). However, the activity of RecBCD-mediated ssDNA cleavage is decelerated on encountering a crossover host instigator (Chi) site. The bacterial chromosome is less prone to extensive spacer acquisition due to the high density of Chi sites, whereas the phage DNA is more susceptible. Furthermore, given that most phages enter the cell as linear DNA, and RecBCD binds and processes exposed linear DNA until it reaches the nearest Chi site, unprotected phage DNA becomes an immediate target for spacer acquisition upon cell entry (Dillingham & Kowalczykowski, 2008; Levy et al., 2015; Modell et al., 2017; Weissman, Stoltzfus, Westra, & Johnson, 2020).

1.1.6.2 Naïve adaptation

Naïve adaptation in CRISPR-Cas system includes the acquisition of immunity *de novo*. This acquisition is independent of the already acquired spacers. The initiative step of this process is the selection and processing of prespacers. This is followed by the integration of this prespacer into the existing repeat-spacer array. The CRISPR array functions as a molecular memory system, storing information about past interactions with foreign genetic elements. In contrast to higher-order eukaryotes, the immunological memories produced by CRISPRs are inheritable and transmitted to offspring cells, facilitating the swift propagation of immunity throughout the microbial community.

1.1.6.2.1 Selection and processing of prespacer

CRISPR adaptation is limited in its ability to capture a spacer from any arbitrary location on MGEs. Within the type I, II, and V CRISPR-Cas systems, the choice of prespacers is directed by a DNA sequence found in the PAM (protospacer adjacent motif) region of MGEs (F. J. M. Mojica, Diez-Villasenor, Garcia-Martinez, & Almendros, 2009; J. Wang et al., 2015). However, in the course of adaptation, the PAM sequence is excluded and not incorporated into

the CRISPR array. As the CRISPR array lacks the PAM, it remains immune to targeting by the CRISPR interference machinery, thereby safeguarding the 'self' genome (F. J. M. Mojica et al., 2009). Nonetheless, the MGE containing the corresponding PAM sequence remains susceptible to CRISPR interference. Furthermore, research indicates that the frequency of spacer uptake is elevated from DNA segments that are actively transcribing and replicating (Goldberg, Jiang, Bikard, & Marraffini, 2014; Ivancic-Bace et al., 2015; Levy et al., 2015; Samai et al., 2015). The distinguishing feature of phage and plasmid DNA is their elevated rates of transcription and replication, in contrast to the relatively lower activity of the host genomic DNA. Consequently, spacers are selectively obtained from MGEs, preserving the host genome from such acquisitions.

While CRISPR-Cas systems display significant diversity, the essential Cas protein components of the adaptation machinery, i.e., Cas1 and Cas2 proteins, are universally conserved across most CRISPR-Cas types. Typically, the adaptation complex is hetero-hexameric with two dimers of Cas1 sandwiched between a dimer of Cas2 ($Cas1_4$-$Cas2_2$) (J. Wang et al., 2015). The Cas1-Cas2 complex recognizes the PAM region of the DNA substrates with dual forks, that are generated by the RecBCD complex, featuring a 23 bp duplex region and 3’ extended overhangs. In *E. coli*, the Cas1-Cas2 complex is structurally adapted to handle DNA substrates approximately 33 nucleotides in length. The initial phase of DNA substrate processing involves the binding of the Cas1-Cas2 complex to the forked DNA substrate, followed by trimming facilitated by host exonucleases like ExoIII, ExoT, and DnaQ (Ramachandran, Summerville, Learn, DeBell, & Bailey, 2020; Yoganand et al., 2019) (Figure 1.4).

Exceptionally, in a few type I, II, and V systems Cas4 cleaves the 3' overhangs of prespacer DNA until it encounters a bound Cas1-Cas2 complex, resulting in the production of spacers with a specific length and a correct PAM (Almendros et al., 2019; Kieper et al., 2018; H. Lee, Dhingra, & Sashital, 2019; H. Lee, Zhou, Taylor, & Sashital, 2018; Rollie, Graham, Rouillon, & White, 2018; Shiimori, Garrett, Graveley, & Terns, 2018).

Prespacer recognition in type II-A CRISPR-Cas systems involves additional Cas9 protein and ancillary protein Csn2. It is proposed that the $Cas1_8$–$Cas2_4$–$Csn2_8$ complex binds to the free ends of prespacers and moves along the DNA until it reaches Cas9, which is bound to a PAM. At this point, the spacer may be excised from the DNA (Ellinger et al., 2012; K. H. Lee et al., 2012; Wilkinson et al., 2019). Interestingly, the adaptation process in the type III-B system involves a unique reverse

transcriptase (RT)-fused-Cas1 protein. In contrast to RT-free systems previously reported, which can only adapt DNAs as CRISPR spacers, the type III-B system exhibits the capability to utilize both RNAs and DNAs as substrates. The adaptation against RNAs is specifically reliant on the reverse transcriptase (RT). This distinctive ability to adapt to RNAs allows the system to preferentially acquire new spacers from transcribing regions. This preference proves advantageous for the system's functionality, as target interference by type III systems necessitates the transcription of the targets (Gonzalez-Delgado, Mestre, Martinez-Abarca, & Toro, 2019; Silas et al., 2016).

1.1.6.2.2 Integration of prespacer into CRISPR array

After the prespacers are captured and processed, they are ready to be integrated into the CRISPR array and be further served as an immunological memory to the host. The Cas1-Cas2 complex that is bound to the prespacer helps further in its integration. One of the four subunits of Cas1 latches the PAM region to its active site, which eventually helps in maintaining the directionality of the spacers that get integrated into the array (Nunez, Harrington, Kranzusch, Engelman, & Doudna, 2015; Shipman, Nivala, Macklis, & Church, 2016; S. Shmakov et al., 2014; Swarts et al., 2014; J. Wang et al., 2015) (Figure 1.4). Another prerequisite of this process is the presence of the leader region upstream to the first repeat of the array. Prespacer integration exhibits polarity, with newly acquired spacers being integrated at the leader proximal end of the CRISPR array (Figure 1.4). This polarized integration preserves the chronological order of encounters with MGEs, thereby enhancing efficient interference against the most recent infections (Barrangou et al., 2007; McGinn & Marraffini, 2016; Pourcel et al., 2005).

Typically, in type I-E system, the 3’-OH groups on both strands of the prespacer sequentially initiate nucleophilic attacks on the junctions between the leader and the first repeat, and between the first repeat and the initial pre-existing spacer, resulting in half-site integration. The recognition of the leader-repeat boundary is guided by the interaction of the leader sequence with the Integration Host Factor (IHF) (Nunez, Bai, Harrington, Hinder, & Doudna, 2016; Nunez, Harrington, et al., 2015; Yoganand et al., 2019) (Figure 1.4). Through transesterification reactions, the Cas1–Cas2 complex integrates the double-stranded prespacer into the CRISPR array (full-site integration), causing a separation between the plus and minus strands of the first repeat and leaving two gaps. To complete the process, DNA polymerase(s) and ligase(s) are believed to participate in filling the gaps (Ivancic-Bace et al., 2015; Nunez,

Lee, Engelman, & Doudna, 2015; Rollie, Schneider, Brinkmann, Bolt, & White, 2015) (Figure 1.4).

In contrast, the adaptation mechanism in type II systems does not necessitate IHF and instead, relies on a brief motif (~5 nt) located in the leader region known as the Leader Anchoring Site (LAS). The Cas1-Cas2 complex directly recognizes LAS, and this interaction is adequate for generating correctly oriented spacers (McGinn & Marraffini, 2016; Wright et al., 2016; Xiao, Ng, Nam, & Ke, 2017).

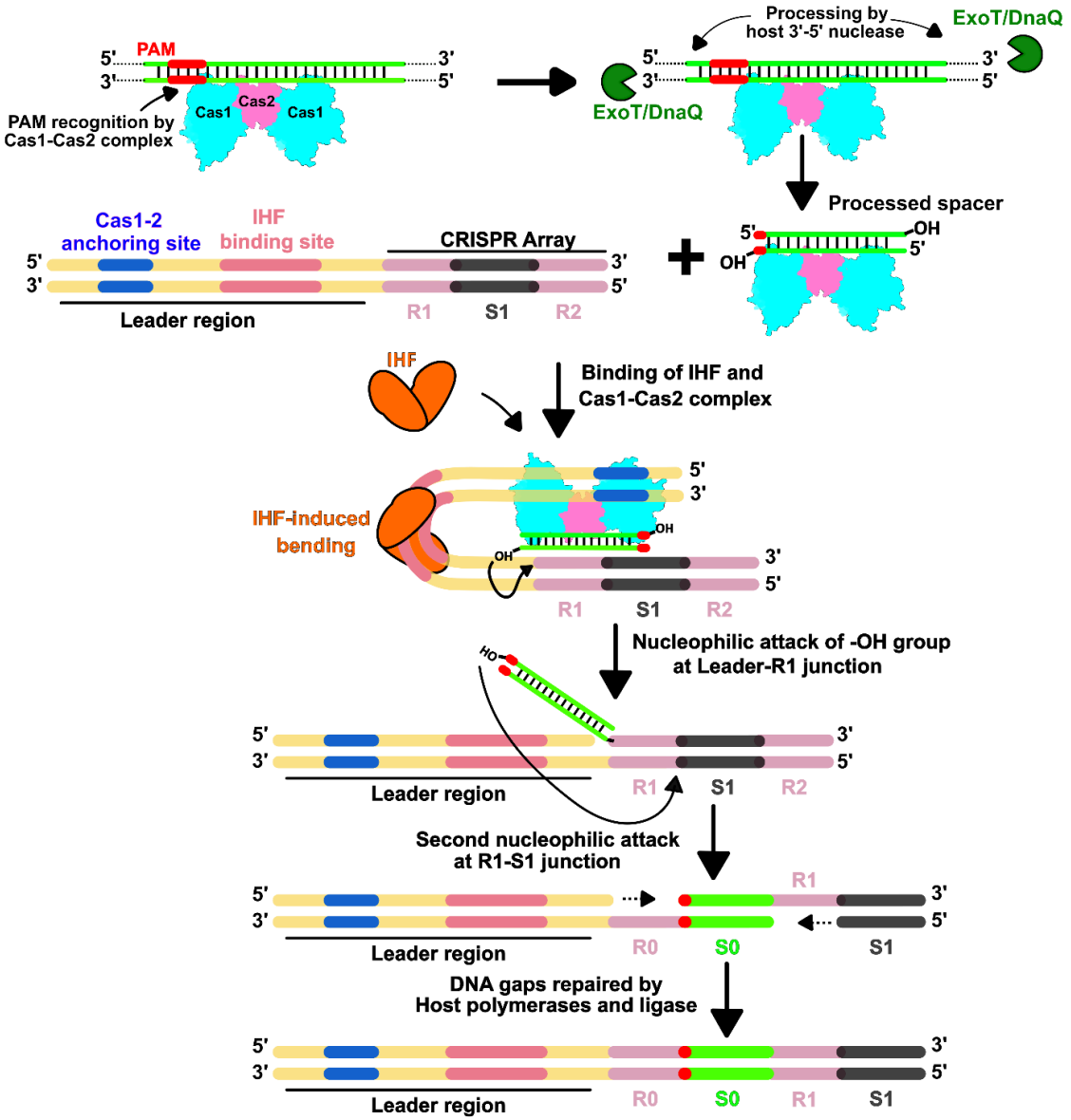


Figure 1. 4: CRISPR adaptation

Schematic representation of the adaptation process in a typical type I-E system. Cas1 binds to the prespacer recognising the PAM sequence. Thereafter, certain host 3'-5' nucleases like ExoIII, ExoT and DnaQ (dark

green) process the prespacer, sized by the adaptation complex. In the subsequent steps, the host recruits a protein called integration host factor (IHF), that binds to the leader region (IHF binding site in pink) of the CRISPR array and bends the DNA such that the Cas1-Cas2 anchoring site (in blue) binds the adaptation complex and localizes it to the leader-repeat junction. The 3' -OH group of the prespacer makes a nucleophilic attack at the leader-R1 junction leading to a half-site integration intermediate product. A full-site integration product is subsequently generated via a second nucleophilic attack by free 3'-OH at the R1-S1 junction. Finally, the host polymerases and ligases repair the breaks and gaps in the array. R1 – first repeat in parent array, S1 – First spacer in parent array, R0 – newly generated repeat after integration, S0 – newly acquired spacer after integration.

1.1.6.3 Primed adaptation

The MGEs have high mutation rates which can easily alter the PAM or the seed sequence. This can in turn result in compromised CRISPR interference. To counteract potential 'escape' by MGEs with such mutations, the CRISPR system has evolved a mechanism to enhance spacer acquisition frequency by utilizing existing spacers as bait (Datsenko et al., 2012; Swarts, Mosterd, van Passel, & Brouns, 2012). In this process, multiple spacers are acquired from the priming region of the interference complex during the interference stage, leading to a robust immune response from the newly acquired spacers. This process is initiated by the recognition of the target sequence by the Cascade complex. The conformation of the Cascade complex, after binding to cognate or mutated target DNA, determines the prevalence of interference or adaptation, respectively. When there is a mutation in the PAM or 'seed' region, the Cascade complex assumes a conformation that is unfavourable for interference. Instead, Cas1-Cas2 and Cas3 are recruited, forming a Primed Adaptation Complex (PAC), leading to the acquisition of new spacers (Blosser et al., 2015; Hayes et al., 2016; Nussenzweig et al., 2019; Redding et al., 2015; Savitskaya, Semenova, Dedkov, Metlitskaya, & Severinov, 2013; Swarts et al., 2012; Xue, Whitis, & Sashital, 2016). Moreover, the natural fusion of Cas2 and Cas3 (Cas2-3) in type I-F elucidates the evolutionary advantage of having an overlap between the adaptation and interference stages (Fagerlund et al., 2017; Makarova et al., 2020; R. H. Staals et al., 2016).

1.1.7 CRISPR maturation

The second stage of CRISPR-Cas immunity is termed as CRISPR maturation. In this stage, the immunological memory that was acquired during adaptation is prepared to be utilised in selectively binding and targeting specific sequences within invading genetic elements.

CRISPR maturation is a two-step process that firstly involves the transcription of the CRISPR array yielding unprocessed transcripts referred as precursor-crRNAs (pre-crRNAs). The pre-crRNA harbours all the acquired spacers and needs to be processed further in the subsequent step with the help of specific endoribonucleases. In some CRISPR-Cas systems, an additional trimming or secondary cleavage step is necessary to produce active, mature crRNAs (Charpentier, Richter, van der Oost, & White, 2015; Hille et al., 2018; Hochstrasser & Doudna, 2015). Both of these steps are described in detail in the later sections.

Mature crRNAs consist of a constant repeat segment recognized by various Cas proteins and a variable spacer region complementary to the invading MGEs. Essentially, every CRISPR system needs to differentiate between crRNAs and other transcripts to effectively carry out immunity. To address this challenge, the constant region serves as a distinctive marker, allowing the system to differentiate it from other cellular RNAs and indicating its selection for RNA processing (C. Hale et al., 2008; Haurwitz, Jinek, Wiedenheft, Zhou, & Doudna, 2010). Moreover, the acknowledgment of the constant region in the crRNA by the CRISPR immunosurveillance complex is frequently linked to the activation of the complex. Therefore, the maturation of crRNA potentially serves as a pivotal regulatory step before the onset of the targeting phase in the immune response (Nussenzweig & Marraffini, 2020).

1.1.7.1 Transcription of the CRISPR locus

The first step of CRISPR maturation is the transcription of the CRISPR array of repeat and spacers. The initial observation of transcribing a CRISPR locus into a primary transcript or pre-crRNA emerged from extensive analyses of non-coding RNAs in *Archaeoglobus fulgidus* and *Sulfolobus solfataricus* P2 (Tang et al., 2002; Tang et al., 2005). The identified sequences were associated with various positions within the CRISPR array, indicating that the entire locus undergoes transcription as a lengthy transcript, subsequently undergoing processing into smaller repeat-spacer units. It was later observed in many different bacterial and archeal species (Agari et al., 2010; Brouns et al., 2008; C. R. Hale et al., 2009; Marraffini & Sontheimer, 2010b; Pougach et al., 2010; Pul et al., 2010). Additionally, the studies also indicated the unidirectional transcription of the pre-crRNAs from a leader proximal end of the locus that harbours putative BRE and TATA box sites. However, in a few *Sulfolobus* species, transcription of the CRISPR array was observed from the complementary strand indicating the presence of these promoter elements downstream of the array (Lillestol, Redder, Garrett, &

Brugger, 2006; Lillestol et al., 2009). However, further processing of these pre-crRNAs has not been well detected and thus remains undetermined.

In general, Cas proteins and pre-crRNA are expressed continuously, yet under specific conditions, their levels can be subject to regulation, indicating the existence of a feedback mechanism to monitor the presence of invasive nucleic acids. The regulatory process exhibits notable variations among different CRISPR-Cas systems, underscoring the remarkable adaptability and evolutionary capacity of CRISPR systems in response to environmental pressures. Various types of regulation have been observed in the transcription of CRISPR arrays and *cas* genes.

- Bacterial quorum sensing (QS) plays an important role in regulating the expression of the CRISPR-Cas system. QS entails the production, release, and detection of extracellular signal molecules known as autoinducers (AI). This communication system regulates behaviors that necessitate synchronized actions among cells to achieve effective outcomes. Studies have shown that the expression of the *cas* genes as well as the CRISPR array of type I-E, I-F and III-A in *Serratia* sp. are found to be regulated by SmaI (autoinducer) and SmaR (repressor) based system of the host (Patterson et al., 2016). Similar observations have been made about the interplay between quorum sensing (QS) and CRISPR-Cas systems in *Pseudomonas aeruginosa* and *Aliivibrio wodanis*, highlighting how the presence of autoinducers influences the regulation and functioning of the CRISPR defense mechanism (Hoyland-Kroghsbo et al., 2017; Maharajan, Hjerde, Hansen, & Willassen, 2022).
- Nucleoid-associated proteins (NAPs) that possess the capacity to bind double-stranded DNA, also serve as global regulators of gene expression. Histone-like nucleoid structuring (H-NS) protein, belonging to the NAPs family, non-specifically binds to DNA at adenine and thymine rich regions, resulting in significant DNA condensation that might hinder gene transcription (Navarre, McClelland, Libby, & Fang, 2007; Stoebel, Free, & Dorman, 2008). Studies indicate that H-NSs not only play a role in the general repression of gene expression but also downregulate the transcription of CRISPR-Cas genes. In *E. coli*, H-NS activity almost entirely suppresses *cas* gene expression under laboratory conditions (Pul et al., 2010). Moreover, certain phages have been discovered to encode their own *hns* genes, suggesting that it might counteract the defense response of the CRISPR-Cas system (Skennerton et al., 2011). The transcription of the CRISPR array is also influenced, albeit to a lesser degree, by the regulatory characteristics of these proteins. This dependence is attributed to the reduced binding affinity of H-NS to the leader portion of the CRISPR array, which allows

for the generation of pre-crRNA at lower levels, as the expression is not entirely suppressed (Pul et al., 2010).

- However, LeuO, another DNA binding protein, binds to the regulatory region of the *cas* gene as a tetramer and interferes with the cooperative binding of H-NS proteins (Guadarrama, Medrano-Lopez, Oropeza, Hernandez-Lucas, & Calva, 2014). Studies on *E. coli* mutants have revealed that LeuO not only inhibits the binding of repressors but also stimulates the transcription of the *cas* genes (Westra et al., 2010). The regulation of *leuO* gene expression indirectly impacts the regulation of the CRISPR-Cas system. Activation of *leuO* expression occurs through the RcsB-BglJ heterodimer, belonging to the LuxR family of transcription regulators (Venkatesh et al., 2010). The transcription of *leuO* is also repressed by H-NS and/or StpA (H-NS homolog). Additionally, a negative feedback loop is evident at high LeuO concentrations. The LeuO protein can bind to its operator on DNA, leading to the inhibition of transcription, indirectly leading to repression of the CRISPR-Cas expression (Stratmann, Pul, Wurm, Wagner, & Schnetz, 2012).
- StpA protein is a homologue of H-NS protein and belongs to the NAPs family. It has been observed that StpA and H-NS paralogs share a common DNA binding site but have opposite roles in the regulation of transcription. In *E. coli* type I-E system, StpA was found to repress the expression of *cas* genes (Mitic, Radovcic, Markulin, & Ivancic-Bace, 2020). Leucine-responsive regulatory proteins (LRPs), like H-NS proteins, are members of the NAPs family and share a similar function of suppressing the CRISPR operon that is regulated by LeuO. In *S. typhi*, both LRP and H-NS bind upstream and downstream of the transcriptional start site of *cas* genes suggesting their collaborative role in controlling gene expression within the CRISPR-Cas system (Medina-Aparicio et al., 2011).
- The complex of cAMP and cAMP receptor proteins, also known as catabolite activator proteins (CAP), is generally a global transcription modulator. However, it was observed to have influenced the CRISPR locus expression in various organisms. In the case of *Pectobacterium atrosepticum* harbouring I-F CRISPR-Cas subtype, a deficiency in glucose activates the expression of *cas* genes, and its abundance has the opposite regulatory effect (Patterson, Chang, Taylor, & Fineran, 2015). Whereas, in *E. coli* with subtype I-E, elevated glucose levels result in the indirect activation of transcription, whereas inhibition of transcription occurs at lower glucose levels (C. D. Yang, Chen, Huang, Huang, & Tseng, 2014). This difference arises due to the difference in CAP binding sites. In case of *E. coli* the CAP binding site is located upstream of the *cse1* transcriptional start site and overlaps the binding

site of the LeuO activator, which directly effects the expression of *cas* genes and CRISPR array. On the other hand, *P. atrosepticum* CAP-box is situated optimally which enhances RNAP binding without effecting the *leuO* binding site and thus upregulating the expression of the CRISPR locus.

- Various two-component systems are also found to control the expression of the CRISPR locus. For example, the VicR/K system regulates the expression of subtypes II-A and I-C in *Streptococcus mutans*. This two-component signaling system favoured the expression of type II-A CRISPR locus over type I-C (Serbanescu et al., 2015). However, the mechanism of action is not clear yet. Similarly, in *P. aeruginosa*, the I-F CRISPR-Cas system is regulated by a two-component system – KinB/AlgB, that takes part in alginate biosynthesis regulation (Borges et al., 2020).

1.1.7.2 Processing of the pre-crRNA

The second step of maturation is the processing of pre-crRNAs (consisting of constant repeat units interspaced by variable spacer units) into mature crRNAs by specific endoRNases. The processing takes place within the repeat sequences as a result of which the mature crRNAs consist of a repeat segment recognized by Cas proteins in a manner dependent on both structure and sequence and a target-specific spacer region. In a few cases, an additional processing step is required to generate the mature crRNAs. These endoRNases serve two distinct functions: first, they identify and process the precursor transcript, leading to the generation of mature crRNAs, and second, they typically retain the mature crRNA for subsequent loading onto the respective effector proteins or complexes responsible for mediating interference. The diversification of CRISPR-Cas into different classes and types and the presence of a wide array of distinct Cas proteins reflect the evolution of distinct mechanisms for crRNA processing.

1.1.7.2.1 Class 1 crRNA maturation

Class 1 CRISPR-Cas systems are subdivided into types I, III and IV, wherein the process of crRNA maturation shows similarity in types I and III. The majority of type I CRISPR-Cas systems, excluding subtype I-C systems, utilize Cas6 endoribonucleases to cleave the repeat sequence within the pre-crRNA sequence (Charpentier et al., 2015; H. Li, 2015).

Cas6 belongs to the repeat-associated mysterious proteins (RAMP) superfamily, which includes a diverse range of protein families characterized by tandem or single ferredoxin-like folds. These folds are also known as RNA recognition motifs (RRM) and are associated with RNA binding (Haurwitz et al., 2010; R. Wang, Preamplume, Terns, Terns, & Li, 2011). The RRM folds are characterized by two sequentially connected units of βαβ, resulting in an overall fold shaped like βαββαβ. In general, RAMP proteins set themselves apart from other ferredoxin-fold proteins by incorporating additional secondary elements onto the fundamental βαββαβ unit and organizing two such domains in tandem. The N-terminal domain in all RAMP proteins typically closely resembles the classic ferredoxin fold, while the C-terminal domain is often interrupted by insertions. The distinctive G-loop links the final α-helix and β-strand (α4 and β8) and is positioned at the junction between the two ferredoxin-like fold/RRM domains. The G-loop has been identified as crucial for both the folding of Cas6 and RNA binding (Haurwitz et al., 2010; Sashital, Jinek, & Doudna, 2011; R. Wang & Li, 2012; R. Wang et al., 2011). The Cas6-mediated processing step is metal-independent and proposed to be catalyzed through a potential general acid-base mechanism, typically resulting in crRNAs with 5′-OH and 2′,3′-cyclic phosphate termini (Carte, Pfister, Compton, Terns, & Terns, 2010; Carte et al., 2008; Gesner et al., 2011; Jore, Brouns, & van der Oost, 2012).

Cas6 recognises and binds to the repeat regions of the pre-crRNA initiating the processing of the precursor substrate. The repeats found in different CRISPR-Cas types and subtypes exhibit variations in both sequence and structure. These repeats can adopt either a structured architecture, such as a stem-loop, or remain unstructured. The ability of each repeat sequence to form stable secondary structures, often in the form of a stem-loop, is influenced by the palindromic nature of the repeat sequence (F. J. Mojica et al., 2000). Sequence similarities and alignment of the repeats allow the organization of CRISPR repeats into 33 repeat clusters. (Kunin, Sorek, & Hugenholtz, 2007). Among these, 12 clusters included 10 or more members, and the most extensive cluster (cluster 1) encompassed 94 repeat sequences. Consequently, various CRISPR-Cas systems can be linked to distinct clusters, showcasing differences in the sequence and structure of repeats. Therefore, the unique recognition of a repeat RNA necessitates a specific mechanistic solution for substrate discrimination by the endoRNases, implying variations in the mechanism of substrate recognition and processing among different CRISPR variants.

CRISPR repeats in subtype I-A and I-B lack palindromic sequences, resulting in an unstructured conformation. Cas6 proteins in these systems form dimers and play a crucial role

in remodeling the repeats to create the necessary stem-loop structure and reposition the cleavage site (Figure 1.5). This is achieved by either two Cas6 proteins forming dimers to process the unstructured repeat, or two non-interacting Cas6 proteins binding to separate sequence motifs within the same repeat, facilitating processing activity (Reeks et al., 2013; Richter, Lange, Backofen, & Randau, 2013; Shao & Li, 2013; Shao et al., 2016). Cas6a (type I-A) and Cas6b (type I-B) have a metal-independent ribonuclease activity, that specifically generates an 8-nt 5′ repeat tag followed by a complete spacer sequence. All known homologs of Cas6a (type I-A) and some homologs of Cas6b (type I-B) release the crRNA after the processing event and do not become integrated into the effector complex (Carte et al., 2008; Charpentier et al., 2015) (Figure 1.5).

Cas6 in type I-D, I-E, and I-F CRISPR-Cas systems identifies stable hairpins that develop between palindromic sequence stretches within their repeats (Kunin et al., 2007) (Figure 1.5). The conserved nucleotide sequence of the repeats within pre-crRNA is crucial for its recognition (Brouns et al., 2008; Jore et al., 2011). Structural studies show that the phosphate backbone in the 3' region of the RNA establishes electrostatic contacts with positively charged residues in both ferredoxin-like domains, while the 5' region remains exposed to the solvent. Specifically, the N-terminal ferredoxin domain (N-ferredoxin) of Cas6/I-E is involved in RNA hydrolysis. It interacts with the lower portion of the RNA stem containing the scissile phosphate and with the two unpaired nucleotides at the 3'-end of the RNA. On the other hand, the C-terminal ferredoxin domain (C-ferredoxin), responsible for recognizing the RNA substrate, plays crucial roles in maintaining (i) the groove-binding element (GBE) situated between β1′ and α1′, where the prime symbol refers to the second, C-terminal RRM, probing the major groove of the pre-crRNA stem and often providing sequence specificity, (ii) the glycine-rich loop (G-loop) motif GhGxxxxGhG (where h represents a hydrophobic residue and xxxxx contains at least one arginine or lysine) at the C-terminus positioned between α2′ and β4′ that stabilizes the pre-crRNA interaction with the Cas6 nuclease, and (iii) the β hairpin formed by β2′ and β3′ that assists in positioning the scissile phosphate of the pre-crRNA transcript in the active site (Carte et al., 2008; Gesner et al., 2011; Haurwitz et al., 2010; Makarova, Aravind, et al., 2011; Sashital et al., 2011). The catalytic residues Tyr23, His26, Arg27, and Arg158 of Cas6e play a role in RNA cleavage (Gesner et al., 2011; Sashital et al., 2011).

Cas6d from type I-D binds to the pre-crRNA in a manner dependent on both sequence and secondary structure. It cleaves the repeat sequence after the stem-loop, resulting in an

approximately 72-nucleotide crRNA with an 8-nucleotide 5′ and a 29-nucleotide 3′ repeat-derived handles. However, in the natural host *Synechocystis*, further maturation to 45 and 39 nucleotides is observed (Jesser, Behler, Benda, Reimann, & Hess, 2019; Scholz, Lange, Hein, Hess, & Backofen, 2013). A similar maturation pattern is observed in type I-D from *S. islandicus*, suggesting step-wise processing by host nuclease(s), similar to type III systems, as well as explains processing in complexes lacking Cas6 and a Cas7 subunit(s) (Lin et al., 2020; Scholz et al., 2013).

Cas6f from type I-F tightly binds to the hairpin region of the repeat RNA. However, In *P. aeruginosa*, Cas6f features an N-terminal ferredoxin-like fold, whereas, its C-terminal region adopts an extended conformation, despite maintaining a basic secondary structure connectivity reminiscent of a ferredoxin-like fold (Haurwitz et al., 2010). Cas6f utilizes a serine and a histidine residue to facilitate cleavage of the pre-crRNA within the repeat at the 3′ side of a stable RNA stem-loop structure. Notably, in contrast to crRNA processing by type I-E Cas6e, crRNAs produced by Cas6f have a non-cyclic phosphate at the 3′ end (Wiedenheft et al., 2011). In the I-F system, Cas6 remains bound to the cleavage product through base-specific interactions with the RNA allowing for the subsequent utilization of the mature crRNA by the Cascade complex (Rollins, Schuman, Paulus, Bukhari, & Wiedenheft, 2015; Sternberg, Haurwitz, & Doudna, 2012).

The type I-C CRISPR-Cas system lacks Cas6 maturase and Cas5d, identified within the locus, serves as the endoribonuclease responsible for cleaving pre-crRNA within the repeats (Figure 1.5). Cas5d recognizes both the base of the pre-crRNA stem-loop and the 3′ single-stranded overhang in the pre-crRNA repeat and subsequently cleaves the substrate in a metal-independent manner (Garside et al., 2012; Nam et al., 2012). Unlike Cas6 enzymes, Cas5d generates an 11-nucleotide 5′ tag instead of the canonical 8 nucleotides. Cas5d cleavage produces crRNA products with a 5′ OH and a 2′,3′-cyclic phosphate. The crystal structure of Cas5d reveals a ferredoxin-based architecture and a catalytic triad consisting of histidine, tyrosine, and lysine, indicative of a general acid-base mechanism. Cas5 proteins feature a single N-terminal RAMP (Repeat-Associated Mysterious Proteins) domain and a smaller C-terminal β-sheet domain, where the RAMP (RRM) domain is responsible for substrate binding and cleavage. Cas5 from certain subgroups belonging to type I-C harbour an insertion (~23 residues) in between the β3 and β4 strands of the RRM fold (Garside et al., 2012; Koo, Ka, Kim, Suh, & Bae, 2013; Nam et al., 2012). Further biochemical and structural analysis demonstrates that after pre-crRNA cleavage, Cas5d assembles into a 400-kDa complex with

the mature crRNA, Cas8c (Csd1), and Cas7 (Csd2), the other two Cas proteins specific to type I-C. Few Cas5d variants have also been shown to possess an additional metal-dependent DNase activity (Punetha et al., 2014).

Unlike the swaths of structural and biochemical insights available for most of the abovementioned type I systems, an unexplored but fascinating system is the type I-G CRISPR-Cas system. Interestingly, type I-G, a newly identified subtype of type I CRISPR-Cas system, does not harbour either Cas5 or Cas6 as a standalone maturase. Instead, it comprises of Csb2 which is a fusion of both Cas5 and Cas6 proteins and was thus predicted to be involved in crRNA maturation (Makarova et al., 2015). Recent biochemical insights into the role of Csb2 from *Thioalkalivibrio sulfidiphilus*, have shown its role in the processing of pre-crRNA (Shangguan, Graham, Sundaramoorthy, & White, 2022). However, its interplay with other components of the CRISPR system is poorly understood.

Type III CRISPR-Cas system is further subdivided into six subtypes (A-F) (Makarova et al., 2020). Amongst them only type III-A and III-B retain *cas6* in their locus. The repeat regions in these systems are unstructured and the Cas6 proteins exhibit high sequence similarity to Cas6 homologs of type I-A and I-B. Additionally, the overall processing mechanism in type III closely resembles that of type I-A and I-B, where Cas6-mediated crRNA maturation occurs through a two-step process. Firstly, processing includes the cleavage of pre-crRNA by Cas6 within the repeats, and then they undergo further processing by unknown host nucleases, at the 3′ ends of the crRNA to generate the active, mature crRNAs (Carte et al., 2010; Carte et al., 2008; Kunin et al., 2007; Marraffini & Sontheimer, 2010a). Moreover, like the Cas6 homologs of type I-A and I-B, type III Cas6 proteins are not part of the interference complex. The catalytic and binding sites in Cas6/III-B are distantly located and connected by the substrate, which interacts weakly or transiently with the signature Glycine-rich loop. Cleavage of the pre-crRNA transcript takes place 8 nucleotides upstream of each spacer, generating the conserved 5' handle and a variable 22-nucleotide repeat-derived sequence at the 3'-end. This processing is facilitated by a catalytic triad consisting of Tyrosine, Histidine, and Lysine. The resulting product remains bound to Cas6 until it is transferred to the respective effector complex (R. Wang et al., 2011). Type III-C, D and F systems do not have genes that express the Cas6 maturase and thus are hypothesised to either utilize Cas5 similar to type I-C or Cas6 from neighbouring CRISPR locus. Interestingly, type III-E harbours a giant RAMP family protein (Cas7-11), which is found to take part in the maturation of the pre-crRNA to mature crRNA.

The *cas7-11* gene consists of four subunits of *cas7*, whose last subunit is interrupted by an insertion element and the first subunit is succeeded by *cas11*. The Cas7.1 and Cas7.2 subunits are responsible for processing the pre-crRNA (Goswami, Rai, Das, & Li, 2022; Ozcan et al., 2021; van Beljouw et al., 2021; Yu et al., 2022).

Type IV CRISPR-Cas system is further divided into five subtypes (A-E), out of which type IV-B lacks Cas6 (Pinilla-Redondo et al., 2020). Type IV systems encompass various Cas6 variants, incorporating sequences resembling Cas6e and Cas6f and an additional, Cas6 homolog known as Csf5 (Makarova et al., 2020). The structure of Cas6/IV-A reveals the presence of two RRM folds wherein the C-terminal RRM fold harbors the GBE, the G-loop and the β hairpin. The metal-independent cleavage by Cas6 is catalysed by tyrosine and histidine (H. N. Taylor et al., 2019). However, sequence alignments comparing biochemically characterized and putative type IV Csf5/Cas6 enzymes have indicated that Csf5 enzymes cleave RNA using arginine active site residues (Ozcan et al., 2019; H. N. Taylor et al., 2021). Cas6/IV from *Mahella australiensis* is found to remain bound to the 5′ side of its cleaved crRNA product, indicating that it might function as a component within a multi-subunit complex that assembles on the processed crRNA (H. N. Taylor et al., 2019). The mechanism of crRNA processing in subtypes IV-C, D and E awaits characterization.

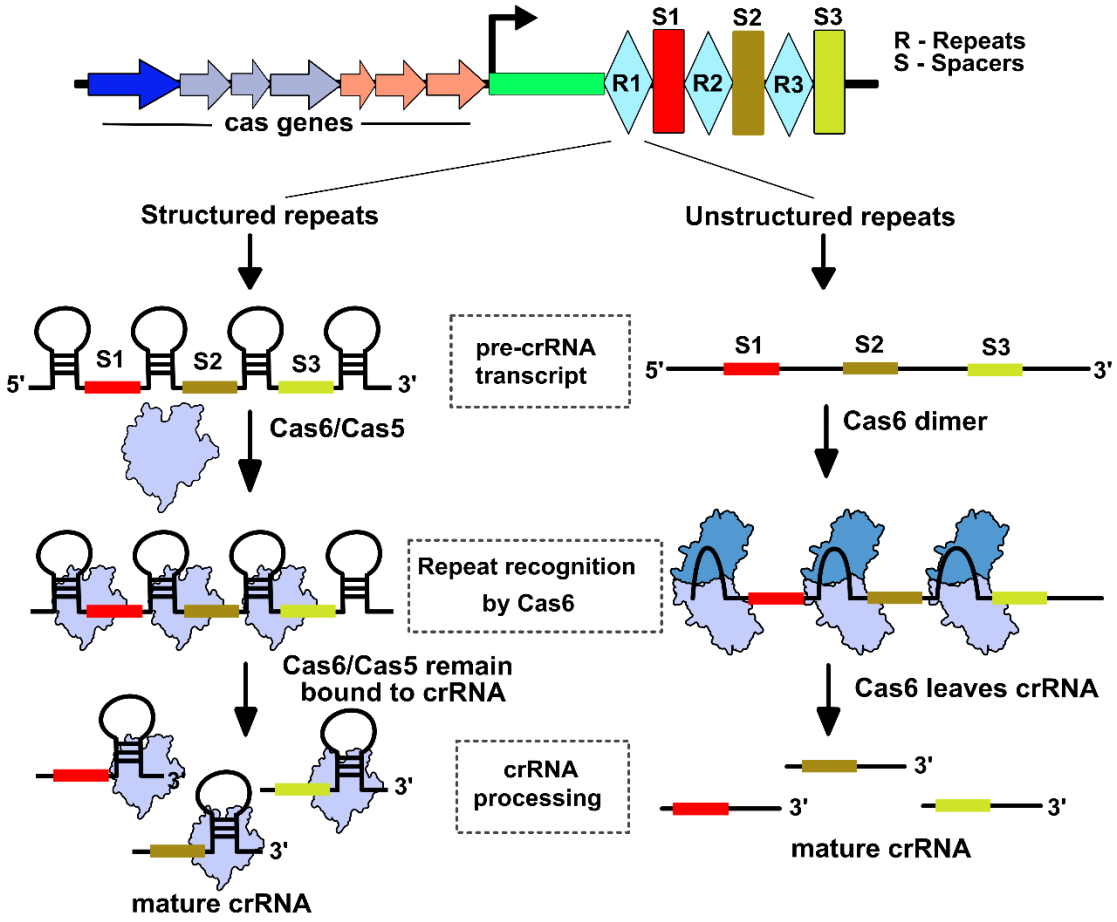


Figure 1. 5: Type I CRISPR maturation

Type I systems harbour both structured (left) and unstructured (right) repeat regions.
Left panel: The expression of the CRISPR array yields a long pre-crRNA transcript which is processed further by either Cas6 or Cas5. Both of them recognise the stem-loop of the repeat RNA and cleave at the base of it. Cas6/Cas5 generally remains bound to the mature crRNA facilitating the assembly of the Cascade complex.
Right panel: The unstructured repeats fail to form a stem-loop structure. In such cases, the Cas6 dimers recognise the repeat sequence and bind to it forming a transient stem-loop like structure that facilitates its further processing. Unlike the previous case, Cas6 here is a multi-turnover enzyme and does not form a part of the Cascade complex.

1.1.7.2.2 Class 2 crRNA maturation

In contrast to Class 1 systems, Class 2 systems employ a single-subunit multidomain effector nuclease and, in some instances, a non-CRISPR host protein for crRNA maturation. Class 2 system is further catagorized into three types- type II, V and VI systems. Amongst these type II and V requires an additional trans-acting small RNA (tracrRNA) for successful

processing of the pre-crRNA (Deltcheva et al., 2011; S. Shmakov et al., 2015; Zhang et al., 2013). The tracrRNA is a small RNA encoded in trans, often transcribed several nucleotides upstream of the CRISPR locus from the opposite strand.

Type II CRISPR system is further subdivided into three subtypes, type II A-C. The maturation of crRNA in type II involves a distinct Cas protein known as Cas9, which collaborates with RNase III, a host factor, to facilitate the processing of CRISPR RNA (Figure 1.6). The apo state structures of Cas9 feature an alpha-helical recognition (REC) lobe and a nuclease (NUC) lobe, which contains the conserved HNH and split RuvC nuclease domains, along with the more variable C-terminal domain (CTD). These two lobes are connected by two linking segments, one formed by the arginine-rich bridge helix and the other by a disordered linker (residues 712–717). The REC lobe consists of three alpha-helical domains (Hel-I, Hel-II, and Hel-III) and lacks structural similarity with other known proteins. The elongated CTD exhibits a Cas9-specific fold and includes PAM-interacting sites essential for PAM interrogation (Jinek et al., 2014). Type II repeat RNAs lack the potential to form hairpin structures. The tracrRNA, which is complementary to the crRNA, plays a crucial role by providing an interacting RNA structure and consequently contributing towards overcoming the deficiency of repeat secondary structures (Deltcheva et al., 2011) (Figure 1.6). Moreover, the tracrRNA is indispensable for crRNA-mediated DNA recognition and Cas9-mediated targeting, even for crRNAs that skip the processing (Jinek et al., 2012; Zhang et al., 2013). The initiation of the maturation process in the type II-A system begins with the binding of tracrRNA to the complementary sequence on the repeat region of pre-crRNA, forming a tracrRNA:crRNA duplex. The effector protein Cas9 (II-A) stabilizes the duplex and recruits the host RNase III protein for the initial processing of crRNA. Following this, another unknown host nuclease trims the 5' repeat-derived overhang (Deltcheva et al., 2011; Le Rhun, Escalera-Maurer, Bratovic, & Charpentier, 2019) (Figure 1.6). The mature crRNA, bound to tracrRNA with approximately 20 nucleotides in the 5' spacer and 19-22 nucleotides in the 3' repeat-derived region, along with Cas9, forms the interference complex necessary for target cleavage (Deltcheva et al., 2011; Gasiunas et al., 2012; Jinek et al., 2012). In contrast, the type II-C system features individual promoters towards the terminal 9 nucleotides of each repeat, leading to the transcription of individual crRNAs. This results in bypassing the need for further processing and yielding mature crRNA (Zhang et al., 2013). Although trimming at the 3’-end may occur, involving RNase III and tracrRNA, it is dispensable for interference. The dispensability of pre-crRNA processing represents a notable mechanistic variation in crRNA generation among type II variants (Zhang et al., 2013).

Type V CRISPR-Cas system is one of the highly diversed systems which is further subdivided into twelve subtypes (A-H, I, K, L and CRISPR-CasΦ) (Tong et al., 2020; Yan et al., 2019). Similar to the type II system, some variants of type V also requires a tracrRNA element and the host RNaseIII for processing the pre-crRNA. Apart from these, Cas12 – a major Cas protein – is involved in the crRNA maturation. Cas12 is bilobed in structure, consisting of an α-helical recognition (REC) lobe and a nuclease (NUC) lobe. The two lobes are joined by a bridge helix (BH) motif. The REC lobe features two REC domains (REC1 and REC2), primarily responsible for accommodating and stabilizing the crRNA-target DNA hybrid upon the formation of the "R-loop". Within the NUC lobe, Cas12a features a RuvC domain, a PAM identical (PI) domain, an oligonucleotide-binding (OB) domain, and a Nuc domain (Gao, Yang, Rajashankar, Huang, & Patel, 2016). The CRISPR repeat RNAs of the type V system are typically structured with a stem-loop region (Figure 1.6). Cas12a/Cpf1 mediated crRNA processing is metal independent and the catalytic residues responsible for crRNA processing are one histidine and two lysines, that are located in the WED domain of the NUC lobe (Fonfara, Richter, Bratovic, Le Rhun, & Charpentier, 2016; Swarts, van der Oost, & Jinek, 2017). Cas12b/C2c1 is the type V-B effector nuclease that demonstrates cleavage properties similar to Cpf1/V-A. However, Cas12b mediated processing requires both crRNA and tracrRNA for targeting, as opposed to Cpf1/V-A (H. Yang, Gao, Rajashankar, & Patel, 2016). The structure of Cas12b/C2c1 is a bi-lobed architecture with REC and NUC lobes, but it lacks the PI domain (L. Liu, Chen, et al., 2017).

Studies on type V-C, G, H and I show involvement of the respective Cas12 (Cas12c, g, h and i) proteins, but the mechanism of their action is poorly understood. Amongst these, type V-G reveals the involvement of tracrRNA in the processing of pre-crRNA, and is also hypothesised to utilise additional host nucleases for this process (Yan et al., 2019). On the other hand, Cas12c from type V-C requires a 71nt long scout RNA for pre-crRNA processing. Cas12d also recruits a scout RNA, but its involvement in pre-crRNA processing is poorly understood (Harrington et al., 2020). Both Cas12c and Cas12d harbour a single RuvC nuclease domain, which is known to help in RNA-guided DNA cleavage, in other type V family enzymes (Haurwitz et al., 2010). The process of maturation in types V-E, F and K is not yet studied. Type V-J/ CRISPR-CasΦ system features CasΦ which is a minimal effector protein consisting of a C-terminal RuvC domain which is responsible for metal-dependent processing of the pre-crRNA (Pausch et al., 2020).

The last member of the Class 2 CRISPR-Cas system is type VI which is further subdivided into four subtypes (A-D) (Makarova et al., 2020). Type VI CRISPR-Cas systems

feature Cas13 as the only effector protein. Unlike Cas9 and Cas12 from type II and V systems respectively, Cas13 does not require tracrRNA to aid in the process of maturation (L. Liu, Li, Ma, et al., 2017; L. Liu, Li, Wang, et al., 2017). Cas13 contains two nuclease domains belonging to the higher eukaryotes and prokaryotes nucleotide-binding (HEPN) superfamily, connected by a Helical-2 domain. The NUC (nuclease) lobe is preceded by an N-terminal REC (recognition) lobe that is composed of an N-terminal domain (NTD) and a Helical-1 domain. The repeat regions of the crRNA form a stem-loop structure (Figure 1.6). In the case of Cas13a/C2c2 from type VI-A the Helical-1 domain of the REC lobe harbours the catalytic sites (arginine and lysine) responsible for the processing of the crRNA (East-Seletsky, O'Connell, Burstein, Knott, & Doudna, 2017; East-Seletsky et al., 2016; Knott et al., 2017; L. Liu, Li, Wang, et al., 2017; O'Connell, 2019). Cas13a cleaves at the pre-crRNA at the base of the stem-loop region towards the 5' end (Figure 1.6). The crRNA maturation process follows a single-turnover mechanism and is not dependent on divalent cations (East-Seletsky et al., 2016; L. Liu, Li, Wang, et al., 2017). Cas13b (type VI-B), pre-crRNA processing active site recognizes the base of the stem-loop region and cleaves at the 3' end of it. Interestingly, it has been observed that in type VI-B, the CRISPR array exhibits variable repeat lengths, while the spacer sequence maintains a consistent length of approximately 30 nucleotides. A mature crRNA originating from this system may possess either 36-nucleotide or 88-nucleotide long repeats, with both variants demonstrating equal efficiency in target cleavage (Smargon et al., 2017). Cas13d (type VI-D) pre-crRNA processing produces a mature crRNA with a repeat-spacer structure similar to the Cas13a family, specifically cleaving at the 5' end of the stem-loop region. However, unlike Cas13a, which is a metal-independent RNase, Cas13d pre-crRNA processing requires divalent ions (Konermann et al., 2018; L. Liu, Li, Wang, et al., 2017; Yan et al., 2018). Given the conservation of pre-crRNA processing in Type VI-A, B, and D systems, it is likely that Cas13c will also maintain the ability to cleave pre-crRNAs, but it requires further investigation.

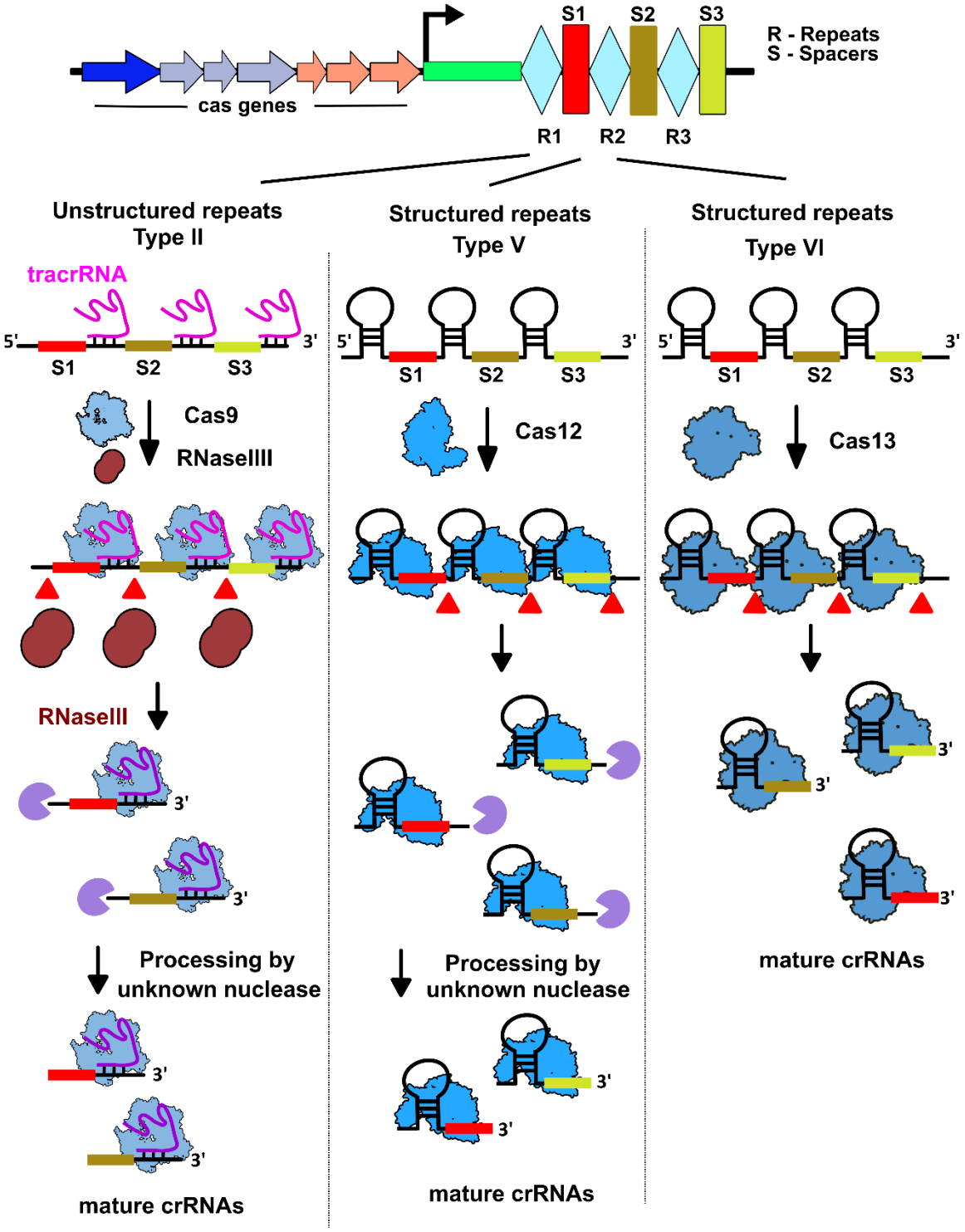


Figure 1. 6: CRISPR maturation in class 2

Class 2 CRISPR-Cas systems feature a single subunit effector protein that takes part in the maturation of pre-crRNA.
Left: Type II systems utilise a tracrRNA component (magenta) in the recognition of the pre-crRNA. This complementary recognition of the repeat region by tracrRNA facilitates Cas9 binding, which further deploys the host RNaseIII (in brown) for cleavage. Post RNaseIII mediated nicking, the crRNA bound to Cas9 is further subjected to processing by certain unknown host nucleases (in lavender).

Centre: Type V CRISPR system harbours Cas12 that recognises the repeat region and makes a nick at the 3' end of the spacer. The individual crRNAs are then further processed by an unknown nuclease (lavender).
Right: Cas13 is responsible for crRNA maturation in the type VI system. It recognises the repeat regions and cleaves at the 5' end of the base of the stem-loop region, producing mature crRNAs.
Red triangles denote nicking positions.

1.1.8 CRISPR interference

The conclusive step of the CRISPR-Cas system is interference, and involves the targeted destruction of invasive elements or mobile genetic elements (MGEs). This is achieved through the complementarity of the specific target DNA/RNA sequence with its corresponding spacer region on the crRNA within the effector complex. Given that crRNA is derived from the CRISPR array, there is a potential risk of self-targeting, leading to the cleavage of its own genome. However, an evolutionary mechanism has evolved to prevent such an autoimmune response. In most CRISPR-Cas systems, the protospacer adjacent motif (PAM) sequence, which is the major criterion for the recognition of the foreign DNA by the adaptation complex, is either not acquired or only partially integrated into the CRISPR array during the adaptation process (Goren, Yosef, Auster, & Qimron, 2012; Heler et al., 2015). Consequently, the PAM sequence is exclusively present in invading MGEs and absent in the CRISPR locus and since the PAM sequence is essential for effective CRISPR interference, the "cleavage of the self-DNA" is precluded (Abudayyeh et al., 2016; Anders, Niewoehner, Duerst, & Jinek, 2014; Deveau et al., 2008; Hayes et al., 2016; F. J. M. Mojica et al., 2009; Westra et al., 2013; Yosef et al., 2012; Zetsche et al., 2017). In contrast, type III systems do not rely on the PAM sequence for such discrimination and instead, the 5' and 3' portions of repeats on the matured crRNA serve as markers to identify invading MGEs. The complementary base pairing between crRNA and the host CRISPR DNA makes interference machinery less effective, thereby safeguarding it from unintended cleavage (Marraffini & Sontheimer, 2010b).

1.1.8.1 Class 1 CRISPR-Cas Interference

Class 1 CRISPR-Cas systems are further categorized into types I, III, and IV, each characterized by a multi-subunit interference complex responsible for the interference process. While our understanding of interference in Class 1 primarily stems from types I and III, the

knowledge regarding the functionality of recently discovered type IV systems is in its early stages and necessitates comprehensive investigation.

Type I system is the most prevalent among CRISPR-Cas systems. It utilizes a multiprotein complex, known as the CRISPR-associated complex for antiviral defense (Cascade), bound to crRNA, that recognizes the target sequences. Thereafter it employs Cas3, a helicase-nuclease protein for the cleavage of targeted DNA sequences. While the general structure of Cascade is maintained across different subtypes, there can be variations in its composition (Makarova et al., 2015; Makarova et al., 2020). The major protein components of the most widely studied Cascade complex from type I-E, are $Cas5_1$, $Cas6_1$, $Cas7_6$, $Cas8_1$ and $Cas11_2$, where the suffix indicates the number of subunits (Figure 1.7A). Following pre-crRNA processing by Cas6e, it remains associated with the 3' repeat portion of the crRNA and facilitates Cascade assembly (Gesner et al., 2011; Sashital et al., 2011). The crRNA becomes an integral component of Cascade, binding along the backbone of the complex and capped by Cas5e at the 5' end (Figure 1.7A). Six subunits of Cas7e proteins constitute the helical backbone of the Cascade, that adopt hand-like shapes, with the thumb domains inducing a kink in the crRNA at every sixth nucleotide. Cascade consists of small (Cas11e) and large (Cas8e) subunits. Two Cas11e subunits (small subunit) interact directly with the Cas7e backbone, forming the 'belly' of the complex, while Cas8e (large subunit) interacts with Cas5e, Cas7e, and Cas11e, constituting the 'tail' (Figure 1.7A). The Cascade effector complex finally assembles into a seahorse-like structure (Jackson et al., 2014; Mulepati, Heroux, & Bailey, 2014; Zhao et al., 2014) (Figure 1.7A). Cas8 recognises the PAM of the target DNA and initiates the local unwinding. Following this, the crRNA binds to the complementary seed sequence, comprising the first eight PAM-proximal nucleotides of the crRNA (Hayes et al., 2016; Xiao, Luo, et al., 2017) (Figure 1.7B). The non-target strand of the DNA is bound by two Cas11e subunits, leading to the formation and stabilization of the R-loop structure. The R-loop structure triggers the recruitment of the endonuclease Cas3, which initiates the degradation of the non-target strand (Y. Huo et al., 2014; Redding et al., 2015; Westra et al., 2012; Zhao et al., 2014) (Figure 1.7B). Cas3 comprises of N-terminal HD domain followed by a superfamily 2 DEAD-box helicase domain followed by a C-terminal domain (CTD). The metal-dependent HD domain of Cas3 nicks the displaced non-target DNA strand, inducing structural changes in the protein that activates its ATP-dependent helicase activity. Consequently, Cas3 translocates and progressively degrades the non-target DNA strand in the 3' to 5' direction (Beloglazova et al., 2011; Gong et al., 2014; D. Han & Krauss, 2009; Y. Huo et al., 2014; Jackson et al., 2014; Sinkunas et al., 2011; Sinkunas, Gasiunas, & Siksnys, 2015)

(Figure 1.7B). It was also shown that the CTD is crucial for the interference step in type I-C Cas3 (Nimkar & Anand, 2020). Nevertheless, in certain variants of Type I systems, such as in Type I-A, the nuclease and helicase domains are expressed separately and yet function synergistically. Additionally, there are instances where these domains are fused to other Cas proteins, such as the Cas2-Cas3 fusion observed in Type I-F. Another example is Type I-G, where Cas3 domains are rearranged, resulting in the presence of N-terminal helicase and C-terminal nuclease domains (Makarova et al., 2015; Makarova et al., 2020; Tuminauskaite et al., 2020). The full degradation of the target DNA may be mediated by host nucleases, whose exact mechanism remains the subject of further investigation.

The interference complex in Type III exhibits a structure similar to Cascade, referred to as Csm for III-A and Cmr for III-B which has structural similarity to the type I interference complex (Hochstrasser et al., 2014; Jackson et al., 2014; Makarova, Haft, et al., 2011; Mulepati et al., 2014; Osawa et al., 2015; R. H. Staals et al., 2014; D. W. Taylor et al., 2015; Zhao et al., 2014) (Figure 1.7C). However, the major difference lies in its ability to target both invading RNA and DNA of mobile genetic elements. Following crRNA maturation, Cas5 binds to a 5' repeat-derived handle, serving as the scaffold for the assembly of other Cas proteins. The backbone of the complex is formed by Cas7 family proteins (Csm3 and Csm5 in III-A and Cmr1, Cmr4, and Cmr6 in III-B), whereas, Cas10 (Csm1 in III-A and Cmr2 in III-B) and Cas11 (Csm2 in III-A and Cmr5 in III-B) constitute the large and small subunits, respectively (Makarova et al., 2020; Osawa et al., 2015; R. H. Staals et al., 2014; D. W. Taylor et al., 2015) (Figure 1.7C). In Type III, the degradation of target DNA is contingent upon its transcriptional activity. Upon binding to a nascent target RNA through crRNA-mediated complementary base pairing, the Cas7 family subunits of the interference complex cleave the target RNA after every sixth base (Figure 1.7C). Simultaneously, the transcriptionally active DNA is cleaved by the HD-nuclease domain of the Cas10 subunit (Osawa et al., 2015; Samai et al., 2015; R. H. Staals et al., 2014; Tamulaitis et al., 2014; D. W. Taylor et al., 2015) (Figure 1.7C). The Type III complex frequently associates with Csm6 and Cxs1 family RNases, playing an auxiliary role in interference (Koonin et al., 2017; Makarova et al., 2020). Besides target cleavage, the target-bound Cas10 converts ATP to cyclic adenylate, activating RNase Csm6, leading to the degradation of non-specific proximate RNA transcripts, including both host and invader RNA (Kazlauskiene, Kostiuk, Venclovas, Tamulaitis, & Siksnys, 2017; Kazlauskiene et al., 2016; Niewoehner et al., 2017) (Figure 1.7C). Furthermore, the newly discovered Type III-E system includes a putative caspase-like protease called TPR-CHAT (also referred to as Csx29). This system also involves accessory proteins Csx30, Csx31, and RpoE. The Cas7-11 protein

associated with crRNA and TPR-CHAT forms a stable complex known as Craspase (CRISPR-guided caspase) initiating an RNA-mediated protease activity in addition to the Target RNA degradation (C. Hu et al., 2022; G. Huo et al., 2023; Strecker et al., 2022; X. Wang et al., 2022; H. Yang & Patel, 2022).

In type IV systems, Csf3 (Cas5) and Csf2(Cas7) are consistently present across all subtypes (IV-A to IV-D), and in some subtypes, Csf1(Cas8) and Cas11 are also observed, that constitute the effector complex (Makarova et al., 2020). Type IV-A recruits a class 2 superfamily helicase called DinG that exhibits ssDNA-stimulated and ATP-dependent 5'-3' DNA helicase activity that can unwind both the non-target strand (NTS) and the target strand (TS) (Cui et al., 2023; H. N. Taylor et al., 2021). Further research is crucial to unravel the intricacies of type IV systems and their interplay with other CRISPR-Cas systems for a more comprehensive understanding of their interference mechanisms (H. N. Taylor et al., 2021).

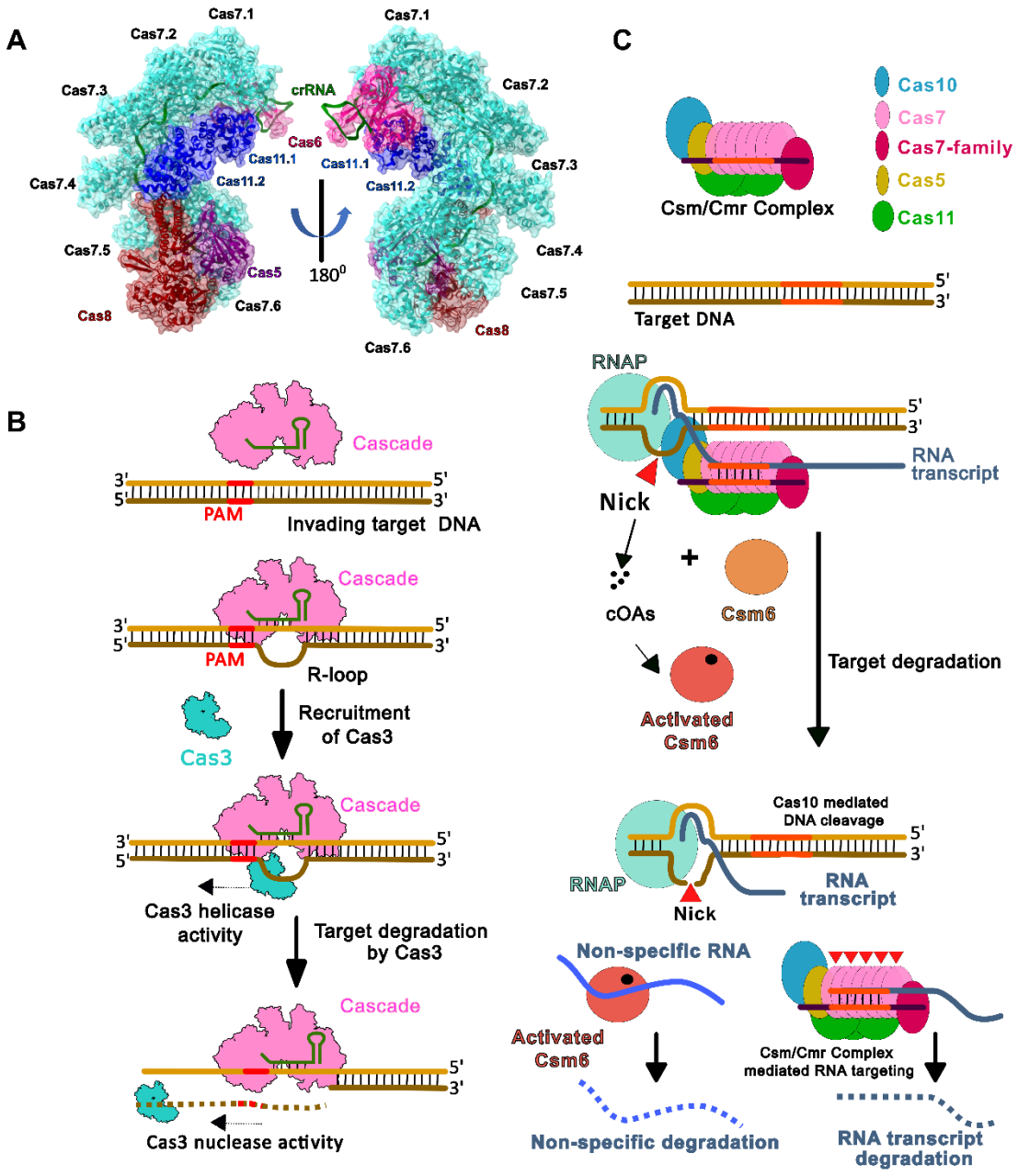


Figure 1. 7: CRISPR interference in Class 1 CRISPR-Cas system

A. Seahorse-shaped structure of the classical type I-E Cascade complex bound to crRNA (dark green) from *E. coli* is shown above. Type I-E Cascade complex consists of six units of Cas7 (cyan) that form the helical backbone. Cas6 (pink) caps this backbone at the top, whereas, Cas5 (purple) is located towards the tail of the structure along with Cas8 (dark red). Two subunits of Cas11 (blue) form the belly region of the Cascade complex. PDB ID: 5CD4.

B. A schematic representation of a typical type I interference mechanism is shown. The Cascade complex (pink) recognises the PAM region (red) and binds to the target DNA, after which it undergoes a conformational change and recruits Cas3 (teal). Cas3 loads onto the non-target strand (dark brown) where its helicase domain unwinds the DNA and the nuclease domain facilitates degradation in a 3' to 5' direction.

C. Type III interference mechanism is multifaceted. It encompasses three major events. First, the target RNA is recognised by the Csm/Cmr effector complex, whose catalytic Cas7 subunits (pink) cleave the RNA. Second, the Cas10 subunit (light blue) of the complex cleaves the DNA region that is bound to the target

RNA transcript. Parallelly, Cas10 is also involved in generating cyclic oligoadenylates (cOA) that act as secondary messengers. These cOAs activate another protein called Csm6 (light red) which subsequently degrades non-specific RNA.

1.1.8.2 Class 2 CRISPR-Cas Interference

The characteristic feature of class 2 CRISPR-Cas is the single multi-domain Cas protein that is responsible for target degradation. Based on the domain architecture of the effector protein, the class is further divided into three types – type II, V and VI. The presence of a single effector complex in Class 2 CRISPR-Cas systems renders them particularly well-suited for genome engineering applications.

The interference complex in type II comprises Cas9, crRNA, and a tracrRNA (Barrangou et al., 2007; Deltcheva et al., 2011; Gasiunas et al., 2012; Jinek et al., 2012; Jinek et al., 2014). The HNH and RuvC domains of the NUC lobe are responsible for cleaving the target DNA during interference. Guided by crRNA, Cas9 locates and binds to the target DNA that induces a conformational change promoting PAM recognition by the C-terminal domain, leading to unwinding of the target DNA (Anders et al., 2014; F. Jiang & Doudna, 2015; Jinek et al., 2014; Nishimasu et al., 2015; Nishimasu et al., 2014) (Figure 1.8A). Similar to type I, an incorrect PAM or 'seed' sequence does not support target degradation. However, perfectly matched target DNA results in the conformational activation of the HNH and RuvC nuclease domains, leading to the stabilization of the R-loop. The HNH domain cleaves the target strand, whereas the RuvC domain nicks the non-target strand, resulting in a blunt double-stranded DNA break (Jinek et al., 2012; Sternberg, LaFrance, Kaplan, & Doudna, 2015) (Figure 1.8A).

Type V CRISPR-Cas system features Cas12 effector nuclease that harbours REC and NUC lobes similar to Cas9. However, few of its subtypes operate without the tracrRNA (Fonfara et al., 2016; Makarova et al., 2020; S. Shmakov et al., 2015). The PI (PAM-interacting) domain in the NUC lobe, identifies the protospacer adjacent motif (PAM) and initiates the unwinding of the targeted double-stranded DNA (dsDNA), the seed sequence promptly engages with the target strands. Thereafter, the seed sequence immediately interacts with the target strands, providing the groundwork for the extension of the R-loop structure (Stella, Alcon, & Montoya, 2017) (Figure 1.8B). In the case of perfect target binding, the RuvC domain is properly positioned and activated, leading to the cleavage of both the target and non-

target strands (H. Yang et al., 2016) (Figure 1.8B). In the Type V-K CRISPR system, Cas12k contains a catalytically inactive RuvC domain, preventing it from participating in target degradation. Instead, the crRNA-Cas12k-target dsDNA complex interacts with a Tn7-like transposase to form a transposome, which facilitates the integration of cargo DNA elements into the target DNA (Faure et al., 2019; Park et al., 2023).

The effector nuclease of type VI CRISPR-Cas system is Cas13, which harbours HEPN domains specialized in cleaving single-stranded RNA (ssRNA). Similar to Cas9 and Cas12, Cas13 exhibits a bilobed structure, featuring two HEPN domains in the nuclease (NUC) lobe (Anantharaman, Makarova, Burroughs, Koonin, & Aravind, 2013; Makarova et al., 2020; S. Shmakov et al., 2015). The binding of the Cas13:crRNA complex to the target ssRNA, activates the HEPN domains, leading to the cleavage of both the target ssRNA and collateral RNA (Abudayyeh et al., 2016; East-Seletsky et al., 2016; S. Shmakov et al., 2015; Smargon et al., 2017) (Figure 1.8C). Cas13 activation is permissive to mismatches at the edges of crRNA:target-ssRNA base pairing, although the central region is crucial for its function (Abudayyeh et al., 2016; L. Liu, Li, Ma, et al., 2017). The interaction between Cas13 and the protospacer flanking sequence (PFS) is also essential during interference (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Smargon et al., 2017) (Figure 1.8C). Once all the necessary conditions are met by the target-ssRNA, the two HEPN domains cleave both the target and proximally present ssRNAs. In the interference process, host RNAs are also cleaved, leading to cellular dormancy or programmed cell death (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Smargon et al., 2017)(123, 220, 221) (Figure 1.8C). The Cas13 mediated targeting is modulated by certain accessory proteins like Csx27, Csx28 and WYL-domain containing proteins (Smargon et al., 2017; Yan et al., 2018).

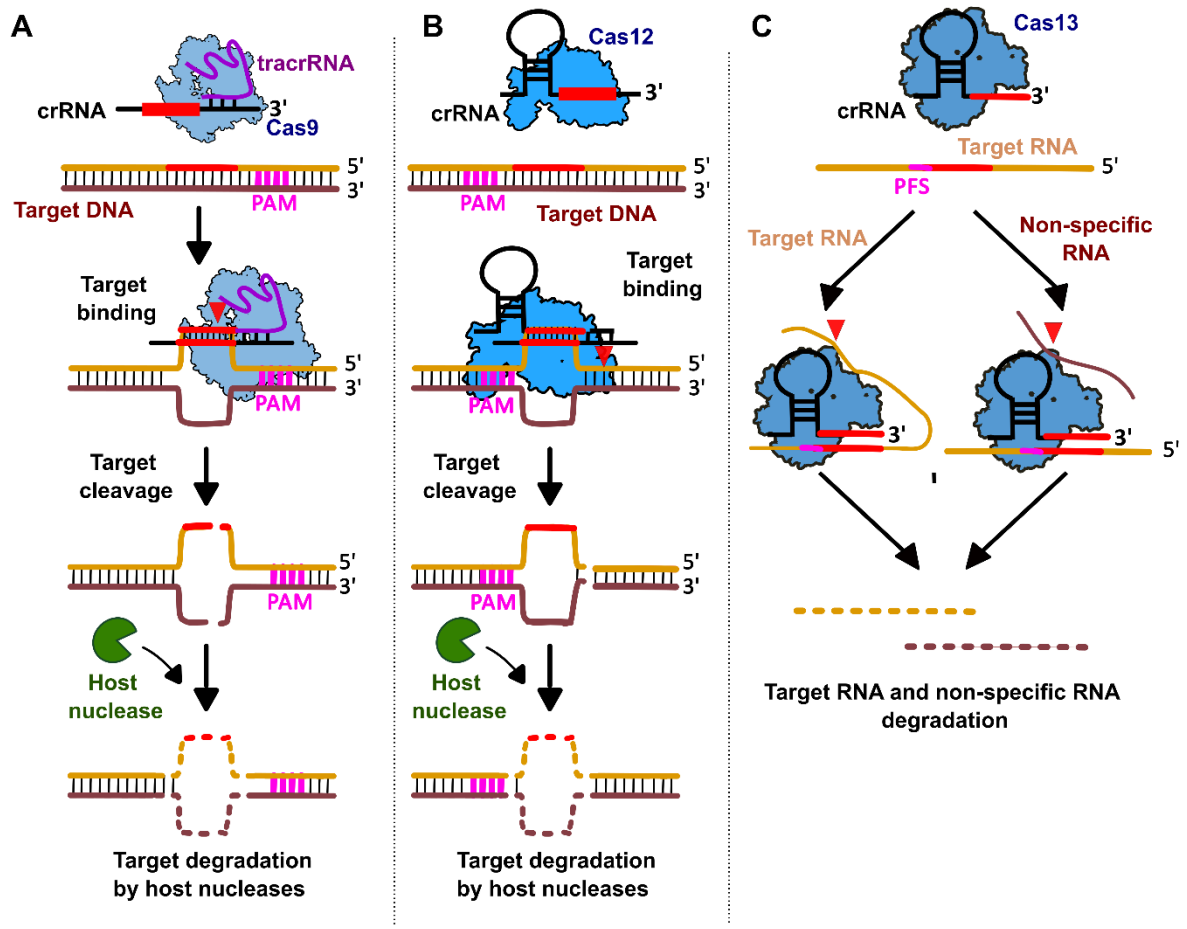


Figure 1. 8: CRISPR interference in Class 2 CRISPR-Cas system.

A. In type II systems the tracrRNA, crRNA and Cas9 facilitate the process of interference. Cas9 recognises the PAM and the seed region of the target DNA sequence and induces a double-stranded break on the target DNA. This is followed by complete degradation by host nucleases (green).

B. Type V CRISPR system employs a Cas12-crRNA complex that recognises and binds the target DNA and cleaves it generating staggered ends. Similar to the type II system, the type V system also employs host nucleases (green) for further degradation of the target.

C. In case of type VI systems, the Cas13-crRNA complex binds to target DNA flanked by protospacer flanking sequence (PFS). This results in two situations, firstly target RNA degradation and secondly, non-specific RNA degradation.

1.1.9 CRISPR inhibition by phage

The mechanisms of phage resistance displayed by bacteria exhibit a remarkable diversity, and in response, phages have evolved a wide range of counteracting strategies. Amongst these, the CRISPR-Cas systems alone can be evaded by phages in multiple ways. Phages can acquire point mutations or deletions in the PAM sequences or the seed region of

the protospacer (Malone, Birkholz, & Fineran, 2021). Alternatively, some phages use anti-CRISPR (Acr) proteins, first described in phages of *P. aeruginosa* (Bondy-Denomy, Pawluk, Maxwell, & Davidson, 2013). In general, Acr proteins work by either preventing recruitment of the crRNA-Cas complex to the target DNA by binding the complex or occluding the PAM sequence, or by inhibiting the endonuclease domain so that cleavage cannot take place (Y. Li & Bondy-Denomy, 2021; Stanley & Maxwell, 2018). Recent studies have also discovered a novel CRISPR–Cas inhibition strategy involving small non-coding RNA anti-CRISPRs (Racrs), that mimic the repeats present in CRISPR arrays (Camara-Wilpert et al., 2023). Glucosylation of phage genetic material has also been shown to protect phages against some CRISPR-Cas systems (Vlot et al., 2018). A different strategy is employed by the jumbo *Serratia* phage PCH45 and *Pseudomonas* phage ϕKZ, which form a protein shell that encloses phage DNA in a nucleus-like compartment, physically shielding it from the CRISPR-Cas complexes (Malone et al., 2020; Mendoza et al., 2020). Of note, this mechanism does not protect the phage from RNA-targeting CRISPR-Cas systems, as the transcribed mRNA is not contained within the nucleus-like compartment during translation.

1.2 Applications of CRISPR-Cas system

Genetic engineering encompasses the usage of several DNA-binding proteins that alter the DNA locus. For example, several prokaryotic proteins, such as restriction enzymes, TALENS, zinc fingers, and recombinases are utilized for gene editing (Danna & Nathans, 1971; Kim, Cha, & Chandrasegaran, 1996; Kim & Chandrasegaran, 1994; Kim, Shi, Berg, & Chandrasegaran, 1997; Ramalingam, Annaluru, Kandavelou, & Chandrasegaran, 2014). However, the challenge lies in specificity. In contrast, CRISPR-Cas systems rely on the spacer sequence in the crRNA, and PAM or PFS on target, for guidance and sequence specificity. This allows for sequence-specific cleavage at any location near PAM or PFS by designing a unique crRNA or engineered guide-RNA (gRNA).

The single effector complex of the class 2 CRISPR-Cas system makes it particularly suitable for genome engineering applications. Among Class 2 CRISPR variants, *Streptococcus pyogenes* Cas9 (SpCas9) is the first and most widely used genome editing tool (Jinek et al., 2012; Jinek et al., 2013; Wiedenheft, Sternberg, & Doudna, 2012). Cas9-mediated genome editing involves generating a double-stranded break (DSB), initiating a DNA damage response,

and inducing an endogenous DNA repair mechanism. Two repair mechanisms, non-homologous end-joining (NHEJ) and homology-directed repair (HDR), can occur. NHEJ is error-prone, often resulting in insertions or deletions (indels) that create knockouts. HDR relies on homologous recombination and can be directed to a precise location using a template DNA sequence, leading to exact genome alterations (Lieber, 2010; Patsali, Kleanthous, & Lederer, 2019; Richardson, Ray, DeWitt, Curie, & Corn, 2016). The fusion of Cas9 and reverse transcriptase is also developed as a prime editing tool, to create insertion and deletion mutants (Anzalone et al., 2019). Besides mutation induction, a nuclease mutant variant of Cas is being widely used for gene regulation. Studies have also demonstrated the use of Cas9 fused to cytidine deaminase or uracil glycosylase to edit a single base (Gaudelli et al., 2017). While, Cas12 (type V) has been employed in genome editing, similar to Cas9, Cas13 (type VI) on the other hand is utilized for RNA editing and regulation (Freije et al., 2019; Gootenberg et al., 2017; Mahas, Aman, & Mahfouz, 2019). However, recent research has shown the effectiveness of employing type I systems for genome editing in prokaryotes (X. Han et al., 2024; Hidalgo-Cantabrana, Goh, & Barrangou, 2019; C. Hu et al., 2024; Z. Xu, Li, Li, Xiang, & Yan, 2021). Besides these, CRISPR-Cas systems are also exploited for various other applications such as live imaging of nucleic acids, chromatin immunoprecipitation, DNA/RNA detection, genome-wide high throughput screening, etc (Anton, Karg, & Bultmann, 2018; Fujita & Fujii, 2013; Y. Hu & Li, 2022; Y. Xu & Li, 2020).

1.3 Definition of the problem

Based on the composition of the effector complex, the CRISPR-Cas system is broadly classified as Class 1, comprising of multi-subunit effector complex and Class 2, which utilises a single subunit effector complex. Class 1 is further classified into type I, III and IV, whereas Class 2 is categorised further into types II, V and VI; each type is subdivided into various subtypes. Amongst all, type I belonging to the Class 1 CRISPR-Cas system, is the most abundant system in bacteria. Type I system is further subdivided into 7 subtypes (A-G). The genomic loci of all the subtypes harbour a varied CRISPR architecture. The Cas proteins that form the effector complex in these subtypes display unique structural and compositional features, suggesting that they have evolved with distinct functional mechanisms. In line with this, the newly discovered subtype, type I-G, displays atypical architecture that may

differentiate its functional mechanism from other type I systems. With respect to the maturation stage, type I CRISPR-Cas systems follow a nearly conserved theme with the involvement of Cas6 as the processing enzyme, except for type I-C, where Cas6 is absent, and Cas5 assumes its role. Interestingly, the type I-G system harbours a fusion protein of Cas5 and Cas6. Intriguingly, it was not known whether Cas5 or Cas6 or both are required for processing the pre-crRNA and how this fusion of Cas5 and Cas6 influences the assembly and function of the effector complex. When the author began her studies, the type I-G system was just identified and these aforementioned questions were not addressed. Therefore, this thesis investigates the mechanistic basis of CRISPR RNA maturation in the type I-G system and its functional implications in CRISPR interference.

1.3.1 Objectives of the study

- To characterise the functional mechanism of the type I-G CRISPR RNA processing enzyme
- To probe the role of type I-G effector complex in crRNA maturation
- To assess the impact of type I-G CRISPR RNA maturation on the interference mechanism.

Chapter 2: Investigating the nuclease activity of Csb2 in crRNA maturation

2 Chapter 2

2.1 Introduction

The fundamental feature of CRISPR-Cas defense is the use of crRNAs to specifically recognize and target invading genetic elements. The process of maturation ensures the proper generation of these crRNA molecules, which makes it a hallmarking step in the CRISPR-based targeting system. The mature crRNA harbours a segment of the repeat region, followed by the complete spacer sequence (Nussenzweig & Marraffini, 2020). The pivotal protein responsible for the maturation of crRNA in type I systems belongs to the Cas6 family, which features two tandemly arranged ferredoxin-like domains, also known as RNA recognition motif (RRM) domains, and a glycine-rich region at the C-terminus (Haurwitz et al., 2010; R. Wang et al., 2011) (Figure 2.1 A-B). Exceptionally, in the type I-C system, where Cas6 is absent, a catalytically active Cas5 drives maturation. Unlike Cas6, Cas5 comprises of a single RRM domain and a C-terminal β-sheet domain (Figure 2.1 A, C) (Garside et al., 2012; Koo et al., 2013; Nam et al., 2012). Though Cas5 and Cas6 are two distinct Cas proteins, their functionality as a maturase is somewhat similar. Cas6 and Cas5 (in case of type I-C), are metal-independent endoRNases. They recognise the stem-loop structure of the repeat RNA and cleave at the base of the stem-loop structure, generating products with a 5′ OH and a 2′,3′-cyclic phosphate groups (Carte et al., 2010; Carte et al., 2008; Garside et al., 2012; Gesner et al., 2011; Jore et al., 2012; Nam et al., 2012). Additionally, Cas5/I-C also displays a tunable metal-dependent DNase activity (Punetha et al., 2014). Remarkably, the type I-G system features a fusion protein encompassing Cas5 and Cas6-like domains (Figure 2.1A) (Makarova et al., 2015; Makarova et al., 2020). This fusion poses an intriguing question: whether Cas5 and/or Cas6-like domain(s) is/are essential for processing the pre-crRNA?

When the author began her studies, the type I-G system was just identified and the mechanism of crRNA maturation in the type I-G was not understood. Therefore, we embarked on a comprehensive study of the biochemical characteristics of Csb2, which harbours the Cas5-Cas6 fusion. Additionally, we tested the RNase activity of Csb2 on various substrates and probed its specificity determinant. Since Cas6 and Cas5 (in type I-C) are metal-independent nucleases, we also checked the metal ions requirement of Csb2 nuclease activity. Subsequently, we also carried out experiments to assess which domain of Csb2 (Cas5-like or

Cas6-like) was involved in the processing of pre-crRNA. We also tested for the nuclease activity of Csb2 on various dsDNA substrates. Altogether, in Chapter 2, our efforts were directed towards *in vitro* characterization of Csb2 as a stand-alone protein.

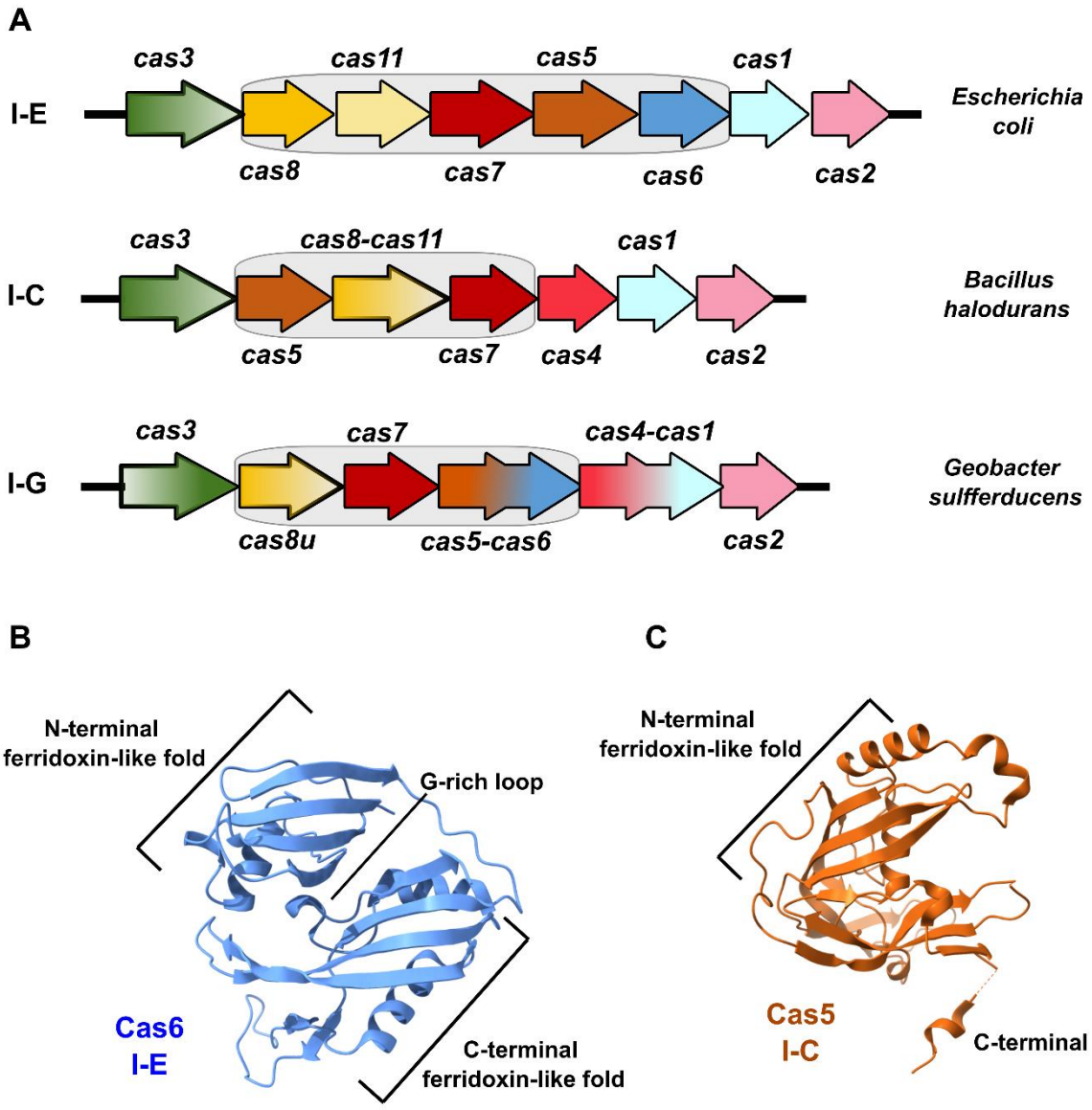


Figure 2. 1: Type I-G has unique locus architecture

A. A schematic representation of the CRISPR-Cas locus of the type I-E, I-C and I-G from *E. coli*, *B. halodurans* and *B. animalis* respectively. Genes are drawn to represent the locus architecture of the respective subtype. The effector module has been highlighted in light grey boxes.

B. Molecular structure of type I-E Cas6 from *E. coli* (PDB -5CD4) is shown in the figure. Cas6 consists of an N-terminal and a C-terminal ferredoxin-like fold followed by a Glycine-rich loop.

C. Molecular structure of type I-C Cas5 from *H. halodurans* (PDB -4F3M) is shown in the figure. Cas5 consists of an N-terminal ferredoxin-like fold followed by a C-terminal β-sheet region.

2.2 Materials and Methods

2.2.1 Molecular cloning

The genomic DNA of *Bifidobacterium animalis subsp lactis* (DSM-10140*), Geobacter sulfurreducens* PCA (DSM-12127) and *Kyrpidia tusciae* (DSM-2912) were procured from DSMZ. The genes encoding Csb2 from all these organisms were cloned in pET Strep II TEV LIC cloning vector (1R, Addgene #29664) at the SspI restriction site by using the Gibson Assembly method. The point mutants of Csb2 from *B. animalis,* namely, Y14A, E24A, R31A, E65A, Y90A, Y236A, E410A, H520A and D530A were created using PCR-based mutagenesis. These constructs, as well as ΔN (1-250 aa) and ΔC (251-545 aa) truncation domains, were PCR amplified and cloned into pET Strep II TEV LIC cloning vector (1R, Addgene #29664).

2.2.2 Purification of proteins

To purify Csb2 from various organisms or its mutant variants, *E. coli* BL21 (DE3) cells were transformed with plasmids carrying the respective genes. The cells were grown in freshly prepared LB broth supplemented with 50 µg/mL of kanamycin at 37 °C until OD_{600} reached 0.6. The cells were then induced with 0.2 mM IPTG and were expressed at 25 °C for 12 hours. The cells were then harvested by centrifugation at 25 °C, 10,000 g for 5 min and the pellet was washed and resuspended in binding buffer (20 mM Tris-Cl (pH 7.6), 500 mM NaCl and 6 mM β-mercaptoethanol) containing 1 mM phenylmethylsulfonyl fluoride (PMSF) and Protease inhibitor cocktail (Sigma-Aldrich) (20 mg/g of cell pellet). Cells were lysed by using a sonicator (Pulse - 2 sec ON and 20 sec OFF) and clarified by centrifugation at 4 °C and 20,000 g for 30 min. The supernatant was then loaded onto a pre-equilibrated 5 mL StrepTrap HP (GE Healthcare) column. After binding, the column the column was washed with 10 column volumes of binding buffer. Subsequently, the protein was eluted in the binding buffer enriched with 2.5 mM d-Desthiobiotin (Sigma-Aldrich). The eluted fractions were subjected to additional purification through a HiLoad 16/600 Superdex 75 prep grade column (GE Healthcare). The proteins were concentrated using ultrafiltration through a 50 kDa cutoff filter (Sartorius), and subsequently, they were snap-frozen and stored at -80 °C until further use.

2.2.3 Promoter prediction

To predict the directionality of the repeat-spacer array, previously well studied parameters were used (Biswas, Fineran, & Brown, 2014). 200 nts flanking both sides of the CRISPR repeat-spacer array was chosen for the analysis. The methods used for prediction involved several approaches. Firstly, a sequence pattern similar to ATTGAAA was searched for within the repeat regions. Secondly, variations in nucleotide composition, particularly a higher proportion of AT bases, in the leader and trailing regions of the arrays were examined. Thirdly, mutations within the repeats leading to the degeneracy of the array was analyzed. Fourthly, the secondary structure of the repeat was assessed. Fifthly, the distance between the repeats and the nearest coding gene at both the 5' and 3' ends was determined. Here, the longer distances indicate the potential functional significance. Lastly, a higher occurrence of the nucleotide A compared to T within the repeat sequences was looked for. This analysis revealed that the sense strand RNA conforms with the first, third and fifth criteria while the anti-sense RNA agreed with the rest three criteria, suggesting an equal probability of repeat RNA to be synthesized from either ends. To confirm the second criteria, we employed Promotech (Chevez-Guardado & Pena-Castillo, 2021), which is a machine learning-based tool for the detection of promoter region in diverse bacterial species. Of the two strands, Promotech predicted promoter sequence upstream of Cas operon in one of the strands only. Whereas for the repeat-spacer array, it predicted that the upstream region of the CRISPR array harbours promoter sequence in both orientations, suggesting that crRNA is synthesised from both the strands. It is worth mentioning that due to the presence of promoter elements in both strands, some of the standard prediction tools couldn't ascertain the transcriptional direction of CRISPR array unambiguously (Briner et al., 2015; Couvin et al., 2018).

2.2.4 Preparation of substrates

Repeat RNA substrate of 36 nt length expressed from both the Cas sense orientation (CSO) and Cas anti-sense orientation (CAO) labelled with either 5' or 3' 6-FAM were chemically synthesized from IDT. To create various mutants of the RNA substrates, their DNA sequence along with the minus strand of the T7 promoter was synthesized (IDT). These were annealed with the plus strand of T7 promoter and were further used as substrates for RNA synthesis via *in vitro* transcription using T7 RNA polymerase. To remove the remains of the

DNA template, they were subjected to DNase I treatment followed by phenol-chloroform extraction. They were further purified by gel extraction followed by ethanol precipitation.

2.2.5 *In vitro* RNase assay

To test the nuclease activity of Csb2, 0.5 µM of the CAO repeat RNA (5' 6-FAM labelled) substrate was incubated with 1 µM of Csb2/I-G or their variants in a reaction buffer containing 20 mM Tris-Cl (pH 7.6), 100 mM KCl and 6 mM β-Mercaptoethanol (β-ME) and incubated at 37 °C for 30 min. Cleavage products were analysed on a 20% (w/v) denaturing polyacrylamide gel and viewed using Bio-Rad gel documentation system.

Time-dependent nuclease assays were performed using 5' 6-FAM labelled CAO repeat RNA where 0.25 µM and 50 nM RNA in reaction buffer 1 were incubated with 1 µM of either wild type (WT) or H520A Csb2, at 37 °C from 0-120 min. The samples were analyzed on a 20% (w/v) denaturing polyacrylamide gel and viewed using Bio-Rad gel documentation system.

2.2.6 Cleavage site mapping assay

In order to map the cleavage sites on the RNA substrates, two types of ladders were generated – RNase T1 ladder and an alkaline hydrolysis ladder (Nilsen, 2013). For RNase T1 digestion reaction, CRISPR repeat RNA (5'/3' 6-FAM labelled in CSO and CAO) was treated with 0.1 units of RNase T1 in 10 mM Tris-Cl (pH 7), 100 mM KCl and 10 mM $MgCl_2$. An alkaline hydrolysis ladder was generated by heating the CRISPR repeat RNA (5'/3' 6-FAM labelled in CSO and CAO) with 1X alkaline hydrolysis buffer (50 mM Sodium carbonate (pH 9.2) and 1 mM EDTA) at 95 °C for 5 min and snap cooled in ice. These digests along with the test sample where the CRISPR repeat RNA (5'/3' 6-FAM labelled in CSO and CAO) was treated with Csb2, were subsequently resolved on a 20% (w/v) denaturing polyacrylamide gel and visualized using a Bio-Rad gel documentation system.

2.2.7 Assay for characterisation of the nature of the ends of the fragment

To assess the nature of the fragment released by Csb2 activity on the repeat RNA in CAO, the cleaved fragment was isolated and treated with T4 polynucleotide kinase (T4-PNK) in 100 mM Tris-acetate (pH 6), 10 mM $MgCl_2$ and 2 mM EDTA at 37 °C for 30 min (Das & Shuman, 2013). The end-repaired RNA was purified by ethanol precipitation and treated with 1 unit of T4 RNA ligase 1 in 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP at 37 °C for 30 min, to ligate the 3' end of the dephosphorylated 5' cleavage fragment to the phosphorylated 5' end of the 3' cleavage fragment (Kaufmann, Klein, & Littauer, 1974). It was subsequently resolved in 20% (w/v) denaturing polyacrylamide gel and visualized using a Bio-Rad gel documentation system.

2.2.8 Fluorescence anisotropy assay

Both 5' and 3' 6-FAM labelled CSO and CAO RNA were used for the study. 10 nM of the RNA substrates were pre-incubated with varied concentrations of either Csb2 in binding buffer containing 10 mM Tris-Cl (pH 8), 50 mM KCl and 6 mM β-Mercaptoethanol (β-ME) for 30 min. The anisotropy value of the reactions was measured using FluoroMax®-4 (Horiba Scientific) by exciting the fluorescein-labelled RNA at 495 nm and measuring the emitted fluorescence at 518 nm. The anisotropy values for the respective protein concentrations were plotted and fitted to the following equation using in GraphPad Prism.

$$A = Amin + \left[(D + E + K_D) - \{(D + E + K_D)^2 - (4DE)\}^{\frac{1}{2}}\right](Amax - Amin)/(2D)$$

where *A* is the anisotropy of the RNA at a protein concentration E, *D* is the total RNA concentration, *A*max and *A*min are maximum and minimum anisotropy (Riede & Eschbach, 1986).

2.2.9 *In vitro* DNase assay

To assess the DNase activity of Csb2, circular or linear pUC19 (dsDNA) was used as substrates. The substrate was incubated with varied concentrations of Csb2 in reaction buffer 1. The samples were visualized in 0.8% agarose gel using EtBr stain. The reaction buffer 1 was supplemented with Mg^{2+} ions, wherever necessary.

2.3 Results

2.3.1 Identification of a suitable type I-G system

Based on the sequence similarity and availability of the genomes, we selected three microorganisms for our study. The microorganisms were *Bifidobacterium animalis*, *Geobacter sulfurreducens* and *Kyrpidia tusciae*. The respective gene encoding Cas5-Cas6 fusion protein, Csb2, was amplified using gene-specific primers. The positive clones were selected for expression studies. The overexpressed proteins were evaluated for their solubility, wherein, Csb2 from *B. animalis* was found to be most amenable to purification under native conditions (Figure 2.2A). Thus, we chose *B. animalis* as our system for further studies (Figure 2.2B).

A time-dependent stability check of Csb2 revealed that Csb2 is unstable and gets degraded within two weeks (Figure 2.2C, upper panel). Western blot analysis of the same samples showed that the degradation was indeed happening from the C-terminal end of the protein (Figure 2.2 C, lower panel). Therefore, all the assays were performed either immediately or within a couple of days from the purification.

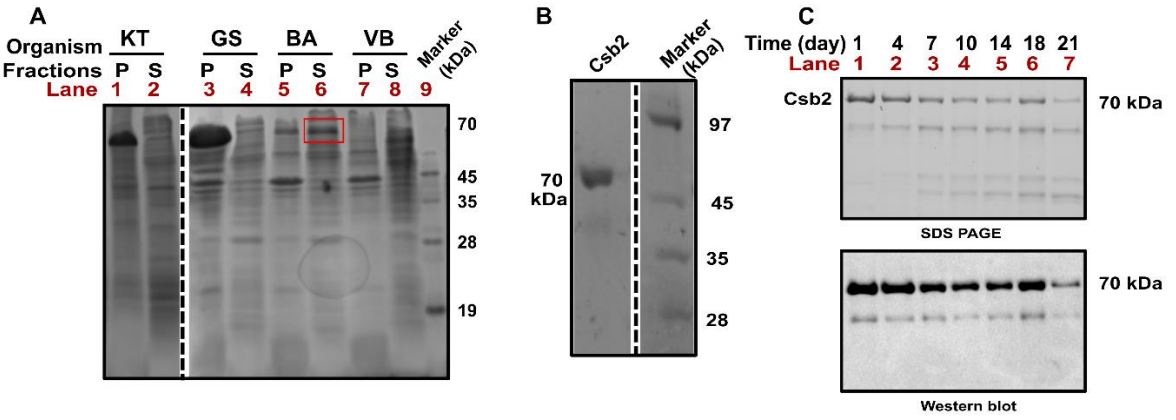


Figure 2. 2: Expression and purification of Csb2

A. A 15% SDS PAGE showing the expression profile of Csb2 (70 kDa) from various organisms; KT - *Kyrpidia tusciae*, GS - *Geobacter sulfurreducens*, BA - *Bifidobacterium animalis*, and VB – empty vector backbone (1R). A protein marker is loaded on the right and the respective molecular weights are indicated. Lanes marked as P and S indicate pellet and supernatant fractions, respectively. The vertical broken line indicates discontinuity between the lanes, which is introduced for the sake of clarity. The expression of Csb2 from *B. animalis* is indicated by a red box.

B. A 15% SDS PAGE showing the purified 70 kDa Csb2 protein (marked as Csb2). A protein marker is loaded on right and the respective molecular weights are indicated. To improve visual clarity, a vertical broken line has been introduced to emphasize the separation between lanes.

C. A 15% SDS PAGE showing the time-dependent degradation of 1 µM Csb2 over a period of 21 days at 4 °C. The corresponding immunoblot analysis is presented below, employing an antibody specifically targeting the N-terminal Strep-tag.

2.3.2 Assessing the orientation of the *B. animalis* CRISPR array

Before embarking upon the understanding of the functionality of Csb2, we tried to understand the features of its substrate RNA. While assessing the orientation of the CRISPR array in *B. animalis*, an intriguing observation was made where it was found that there was an equal probability of the presence of promoter-like elements in both strands, which might result in the transcription of the repeat-spacer array from both the directions (Chevez-Guardado & Pena-Castillo, 2021). To distinguish between the two orientations, we hereafter refer to one as Cas sense orientation (CSO) where the transcription direction of CRISPR array aligns with that of *cas* genes and the other as Cas anti-sense orientation (CAO) (Figure 2.3 A-C).

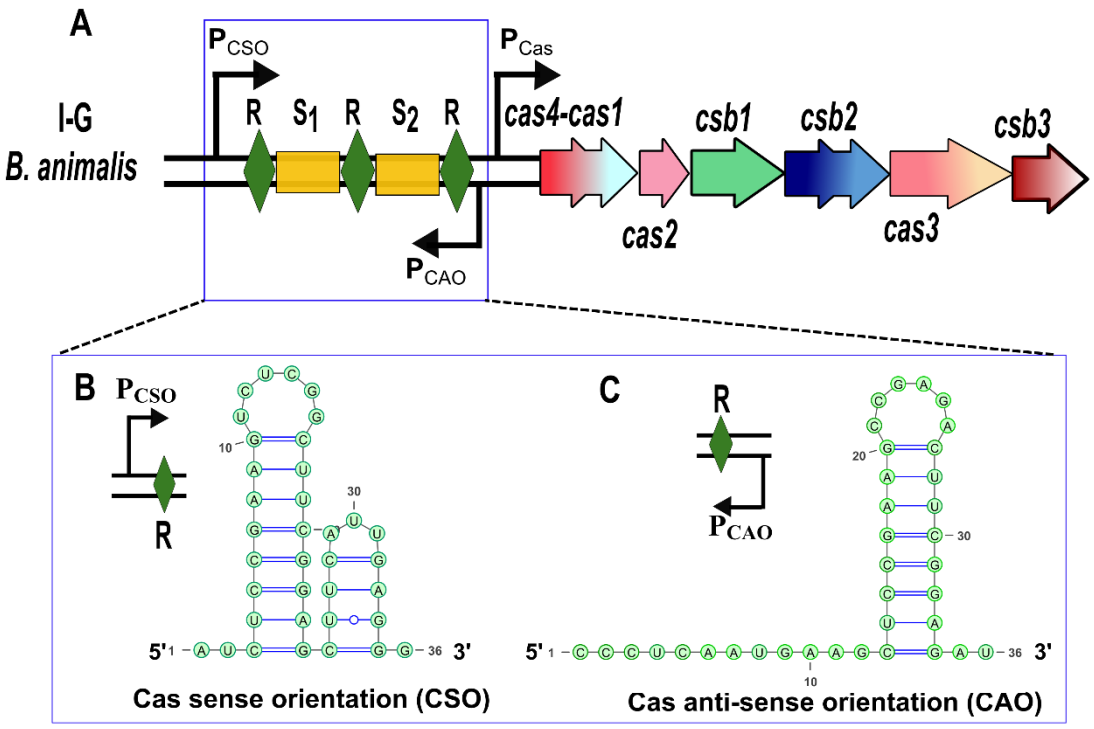


Figure 2. 3: Dual orientation of *B. animalis* CRISPR array

A. A schematic representation of the CRISPR-Cas locus of the type I-G from *B. animalis subsp lactis*. Genes are drawn to represent the locus architecture of the respective subtype. The repeats regions (R) are indicated

as green diamonds, whereas, the spacers (S) are shown as rectangles in yellow. The predicted promoter elements (P_{CSO}, P_{CAO} and P_{Cas}) are indicated by the arrows in black.

B. A predicted secondary structure representation of the *B. animalis* CRISPR repeat RNA in Cas sense orientation (CSO) is shown. The secondary structure prediction was made using the MFOLD Web Server and illustrated using the Varna tool (Darty, Denise, & Ponty, 2009; Zuker, 2003).

C. A predicted secondary structure representation of the *B. animalis* CRISPR repeat RNA in Cas anti-sense orientation (CAO). The secondary structure prediction was made using the MFOLD Web Server and illustrated using the Varna tool (Darty et al., 2009; Zuker, 2003).

2.3.3 Characterization of the nuclease activity of Csb2

To test Csb2 substrate specificity, we incubated both CSO and CAO RNA with Csb2 and noted a prominent small RNA fragment, suggesting that Csb2 is an RNase that could process repeat RNA from both orientations (Figure 2.4 A & B). However, even after prolonged incubation of the CAO RNA, no further processing was observed until 60 mins, suggesting that Csb2 is an endonuclease that cleaves at a specific region within the repeat RNA (Figure 2.4D). To check the dependency of metal ions for the RNase activity, we incubated the reaction with metal and EDTA, separately and found that Csb2 is a metal-independent RNase, similar to its counterpart Cas6 (Figure 2.4C).

To identify the site of cleavage within RNA substrates, we performed alkaline hydrolysis and RNase T1 based structural mapping. This revealed that Csb2 cleaves the CSO repeat RNA at 8 nt from the 3' end, that is between C_{28} and A_{29} (Figure 2.5 A & B). Similarly, it also cleaves the CAO repeat RNA 5 nt from the 5' end, between C_5 and A_6 (Figure 2.5 C & D). On further investigation, it was found that the repeat RNAs in both the orientations harbour a unique conserved sequence at the site of cleavage, that is 5' U**CA**[AU]U 3', where **C** and **A** are the site of cleavage and AU within the bracket indicates that either A or U may occur at this position. Since it was observed that Csb2 could cleave the repeat RNA substrate from both the orientations in a similar manner, we have used the Cas anti-sense orientation (CAO) RNA substrate for our further studies, unless mentioned otherwise.

Studies suggest that both Cas6 and in a few instances Cas5, cleave the substrate RNA via nucleophilic reaction generating fragments that comprise of 5' hydroxyl (OH) and 2',3'-cyclic phosphate RNA ends. To understand the nature of the RNA ends generated by Csb2, we treated the CAO 5' 6-FAM labelled cleavage product with T4 Polynucleotide Kinase (PNK) in an acidic condition such that it favours the removal of phosphates (Das & Shuman, 2013)

(Figure 2.6A). On analyzing the reactions on a denaturing PAGE, it was found that the T4 PNK treated fragments migrated slower than that of the untreated ones indicating the conversion of the phosphate group to a less negatively charged hydroxyl group. To further confirm this, we attempted to ligate the 3' end of the dephosphorylated 5' cleavage fragment to the phosphorylated 5' end of the 3' cleavage fragment using T4 RNA ligase. However, we could not succeed as the size of the generated RNA fragment (5 nt) is too small for T4 RNA ligase to act upon (Kaufmann et al., 1974) (Figure 2.6B). Thus, based on the difference in migration, it appears that RNA cleavage by Csb2 yields fragments with 2', 3'- cyclic phosphate ends.

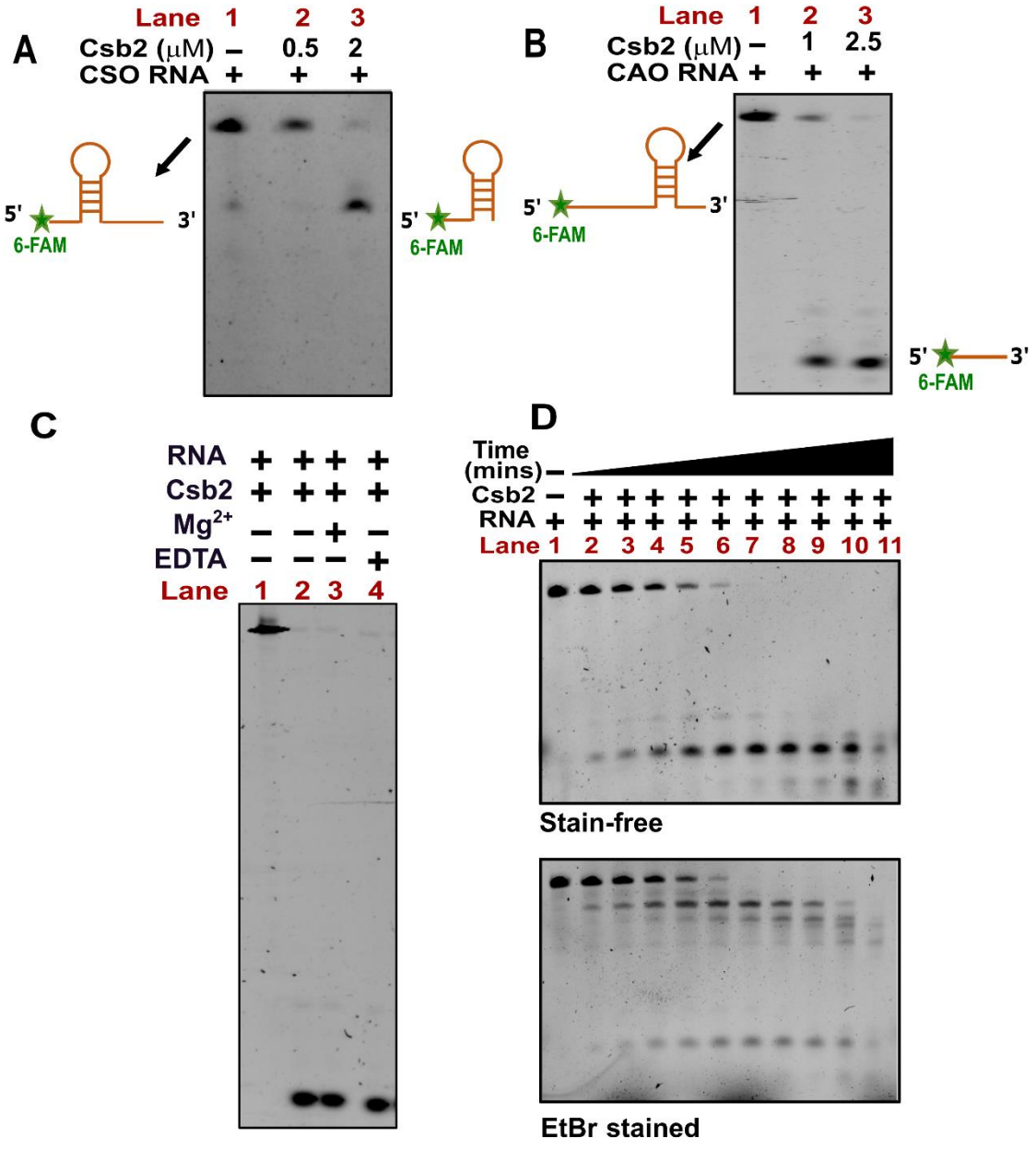


Figure 2. 4: Characterising the RNase activity of Csb2

A. A 20% denaturing PAGE showing the RNase activity of Csb2 (0.5 µM and 2 µM) on 5' 6-FAM labelled CSO repeat RNA substrate. The schema shows both the substrate and the cleaved fragment.

B. A 20% denaturing PAGE showing the RNase activity of Csb2 (1 µM and 2.5 µM) on 5' 6-FAM labelled CAO repeat RNA substrate. The schema shows both the substrate and the cleaved fragment.

C. A 20% denaturing PAGE shows RNase activity of Csb2 (1 µM) on 5' 6-FAM labelled WT CAO repeat RNA substrate (0.25 µM), in presence of metal and EDTA.

D. A 20% denaturing PAGE shows the time-dependent (0, 5, 10, 15, 30, 60, 90, 120 min) RNase activity of Csb2 (1 µM) on 5' 6-FAM labelled WT CAO repeat RNA substrate (0.25 µM). Both stain-free and EtBr stained gels are shown.

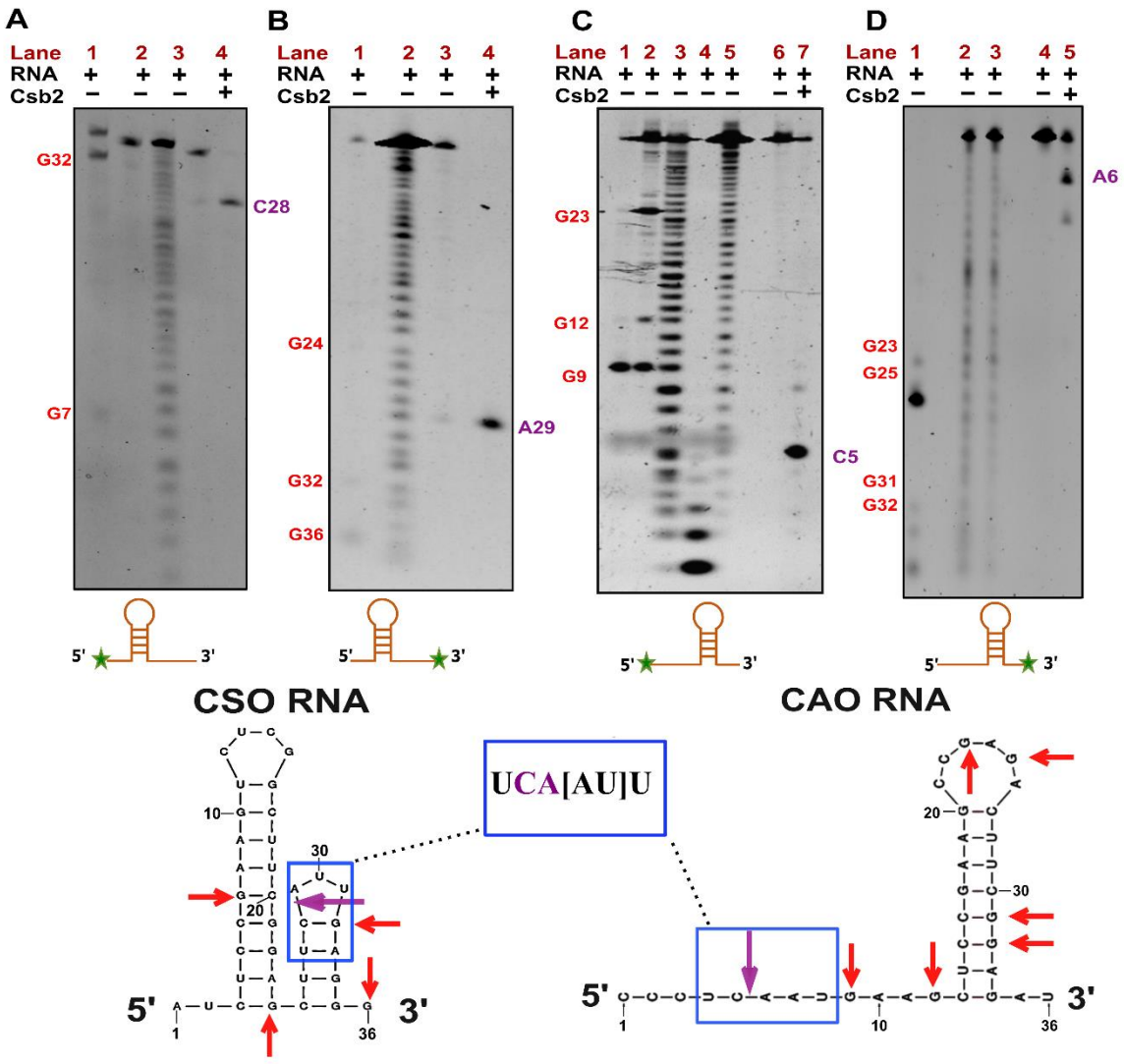


Figure 2. 5: Mapping the site of cleavage by Csb2

A 20% denaturing PAGE is shown for mapping the cleavage site in the 6-FAM labelled CSO (**A, B**) and CAO (**C, D**) repeat RNA by Csb2, using an RNase T1 digestion ladder and an alkaline hydrolysis ladder.

A. The RNase T1 digestion ladder in lanes 1 and 2 was generated by treating the 5' 6-FAM labelled CSO repeat RNA with 1 and 0.1 units of RNase T1, respectively. The alkaline hydrolysis ladder in lane 3 was generated by incubating it for 5 min at 95 °C. Positions that are specific to cleavage by RNase T1 are indicated as G32 and G7.

B. The RNase T1 digestion ladder in lane 1 was generated by treating the 3' 6-FAM labelled CSO repeat RNA with 0.1 units of RNase T1. The alkaline hydrolysis ladder in lane 2 was generated by incubating it for 5 min at 95 °C. Positions that are specific to cleavage by RNase T1 are indicated as G24 through G36 in red.

C. The RNase T1 digestion ladder in lane 1 and 2 was generated by treating the 5' 6-FAM labelled CAO repeat RNA with 1 and 0.1 units of RNase T1, respectively. The alkaline hydrolysis ladder in lanes 3, 4 and 5 was generated by varying the hydrolysis incubation times, 5 and 10 min at 95 °C and 20 min at 60 °C, respectively. Positions that are specific to cleavage by RNase T1 are indicated as G9 through G23 in red.

D. The RNase T1 digestion ladder in lane 1 was generated by treating the 3' 6-FAM labelled CAO repeat RNA with 1 unit of RNase T1. The alkaline hydrolysis ladder in lanes 2 and 3 was generated by varying the hydrolysis incubation times (5 and 10 min at 95 °C), respectively. Positions that are specific to cleavage by RNase T1 are indicated as G23 through G32 in red.

The site of cleavage by Csb2 on both the RNA substrates, along with the sites of cleavage by RNase T1 are indicated on the predicted repeat RNA structures, using purple and red arrows, respectively. The conserved sequence motif 'U**CA**[AU]U' is also highlighted.

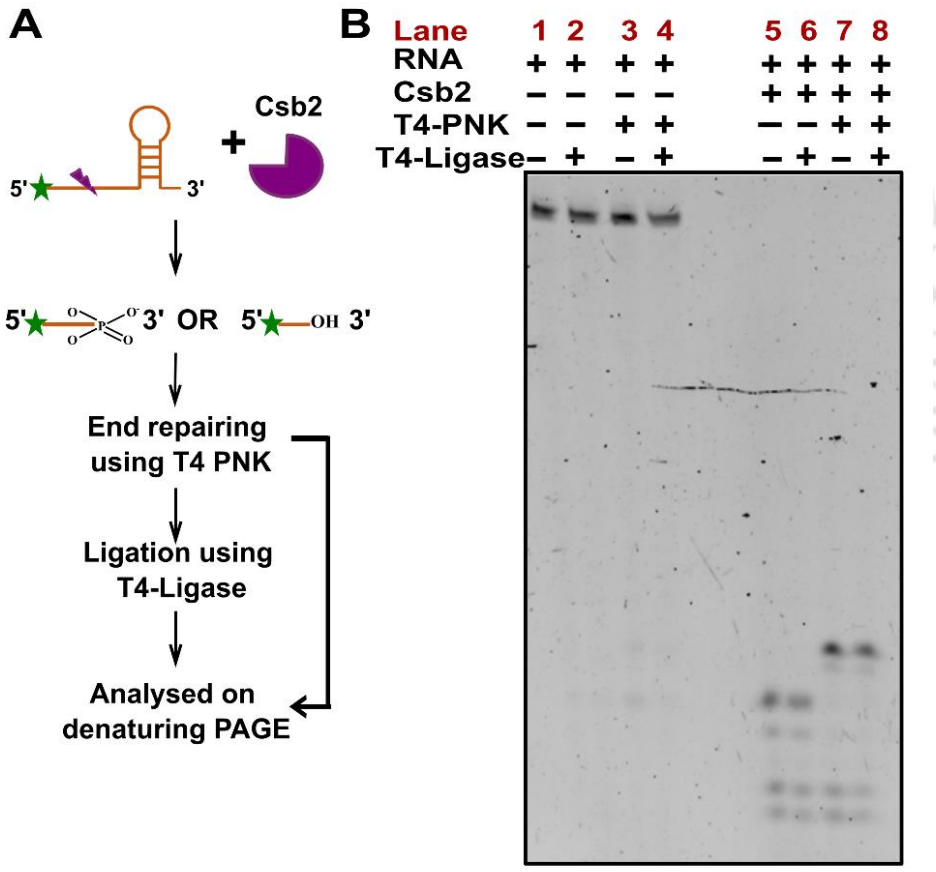


Figure 2. 6: Understanding the nature of the cleaved fragments by Csb2

A. A schematic representation of the experiment designed to understand the nature of the cleaved fragments, whose experimental results are shown in panel D. The cleaved RNA product was treated with T4 PNK at pH 5.2 in the absence of ATP to replace terminal phosphates with hydroxyl groups.

B. A 20% denaturing PAGE is displayed for deciphering the nature of cleaved RNA fragment by Csb2. The difference in migration between T4 PNK treated and untreated samples is attributed to the charge difference between these groups.

2.3.4 Substrate specificity of Csb2

To investigate the sequence and structural preference of Csb2 for the RNA substrate, we designed several RNA constructs to generate structural and sequence variations in the substrate RNA (Figure 2.7A). On incubating Csb2 with each variant of RNA that has altered stem-loop structure, we observed reduced cleavage, suggesting the crucial role of the stem-loop in the nuclease activity. However, to evaluate sequence conservation, we introduced alterations upstream of the cleavage site, along with deletion of the 5' region. This resulted in a noticeable reduction in RNA processing thereby confirming the significance of the 5' region for the nuclease activity (Figure 2.7B).

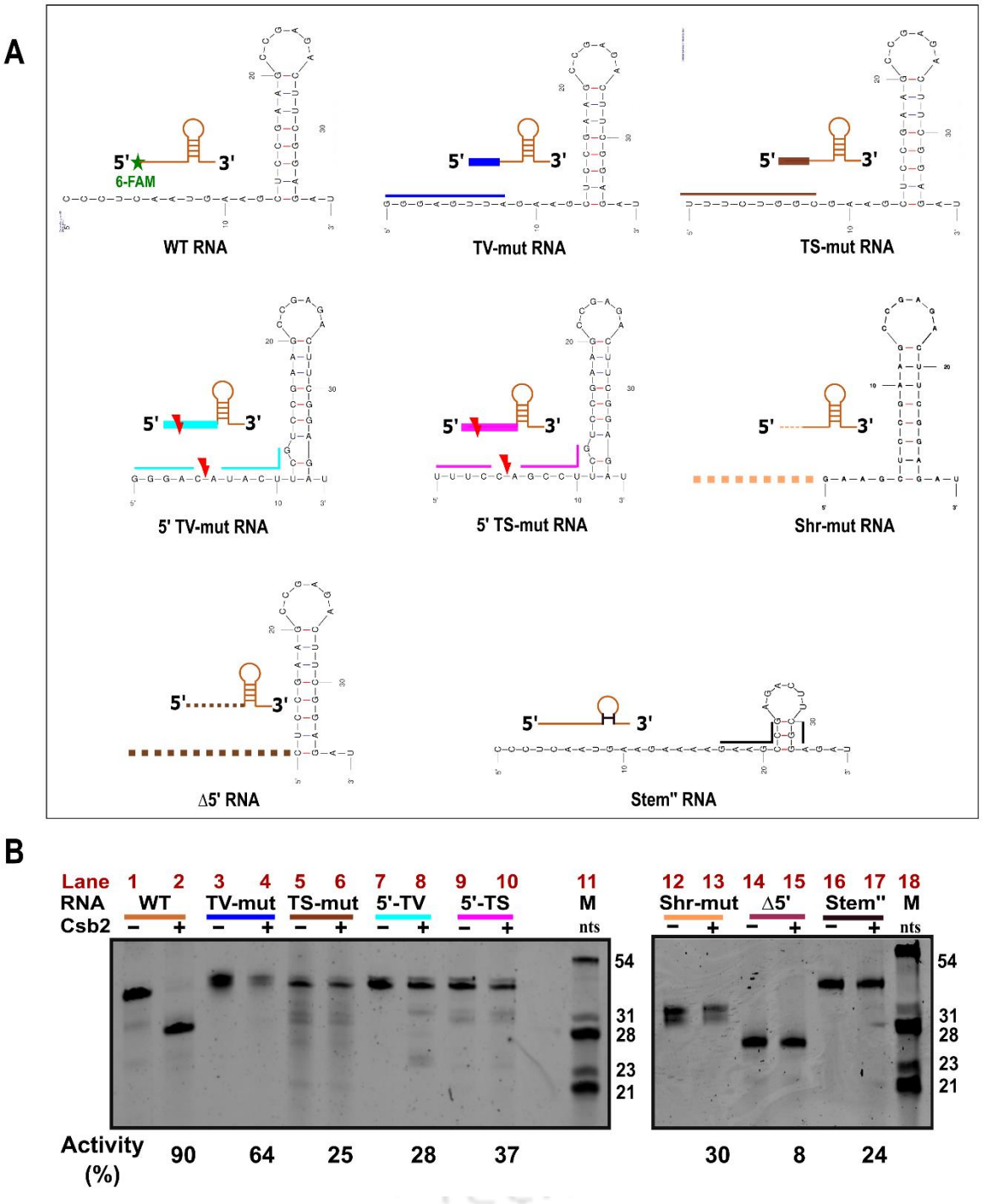


Figure 2. 7: Various RNA mutant substrates used to understand the substrate specificity of Csb2

A. Secondary structure representation of the varied repeat CAO RNA substrates that were used to study the influence of sequence and structure of repeat RNA on the nuclease activity of Csb2. The secondary structure prediction was made using the MFOLD Web Server (Zuker, 2003). TV-mut (blue) and TS-mut (brown) RNA indicate transversion and transition mutations incorporated in the first 8 nt from the 5' end, respectively. 5' TV-mut (cyan) and 5' TS-mut (magenta) RNA indicate transversion and transition mutations incorporated in the first 12 nt from the 5' end, respectively, but keeping the site of cleavage (C_5 and A_6) intact in both cases. Shr-mut RNA (orange dashed lines) indicates a deletion of the first 8 nt from the 5' end, whereas, Δ5' (dark brown dashes) indicates deletion of 5' overhang (12 nt from the 5' end).

Stem" RNA (black) variant indicates destabilisation of the stem region, yet, the total length with respect to the WT remains unaltered (36 nt).

B. An EtBr stained 20% denaturing PAGE is shown for assessing RNA substrate specificity of Csb2 (1 µM) on structural and sequence variants of WT RNA (0.5 µM) as described in (A).

Upon identifying the RNA structural elements preferred for cleavage by Csb2, we next wanted to understand the repeat RNA regions where the protein binds for processing. Towards this, we designed a fluorescence anisotropy-based assay with 6-FAM end-labelled CSO and CAO repeat RNA. We hypothesised that Csb2 bound RNA would display increased fluorescence anisotropy in comparison to the unbound RNA. Further, by using a substrate labelled on either ends, we could also track the region of the repeat RNA that remains bound after cleavage. (Figure 2.7 C-F). Using this experimental design, we observed a steady increase in the fluorescence anisotropy in the CSO repeat RNA construct labelled at the 5' end that retained the stem-loop region, suggesting that CSb2 stably binds to this region of the repeat RNA (Figure 2.8A). Contrastingly, no such trend was observed on using the 3'6-FAM labelled substrate (Figure 2.8B). Interestingly, we previously observed that Csb2 cleaves towards the 3' end of the CSO repeat RNA and hence may not remain bound to the cleaved 3' fragment. Similar observations were made for CAO RNA, where Csb2 preferably binds to the stem-loop region, with lower affinity as compared to the CSO RNA substrate (Figure 2.8 C & D). Overall our study indicated the importance of the stem-loop region for Csb2 binding and the 5' overhang region for its nuclease activity, suggesting an unconventional mode of RNA processing by Csb2 (Figure 2.8 A & D).

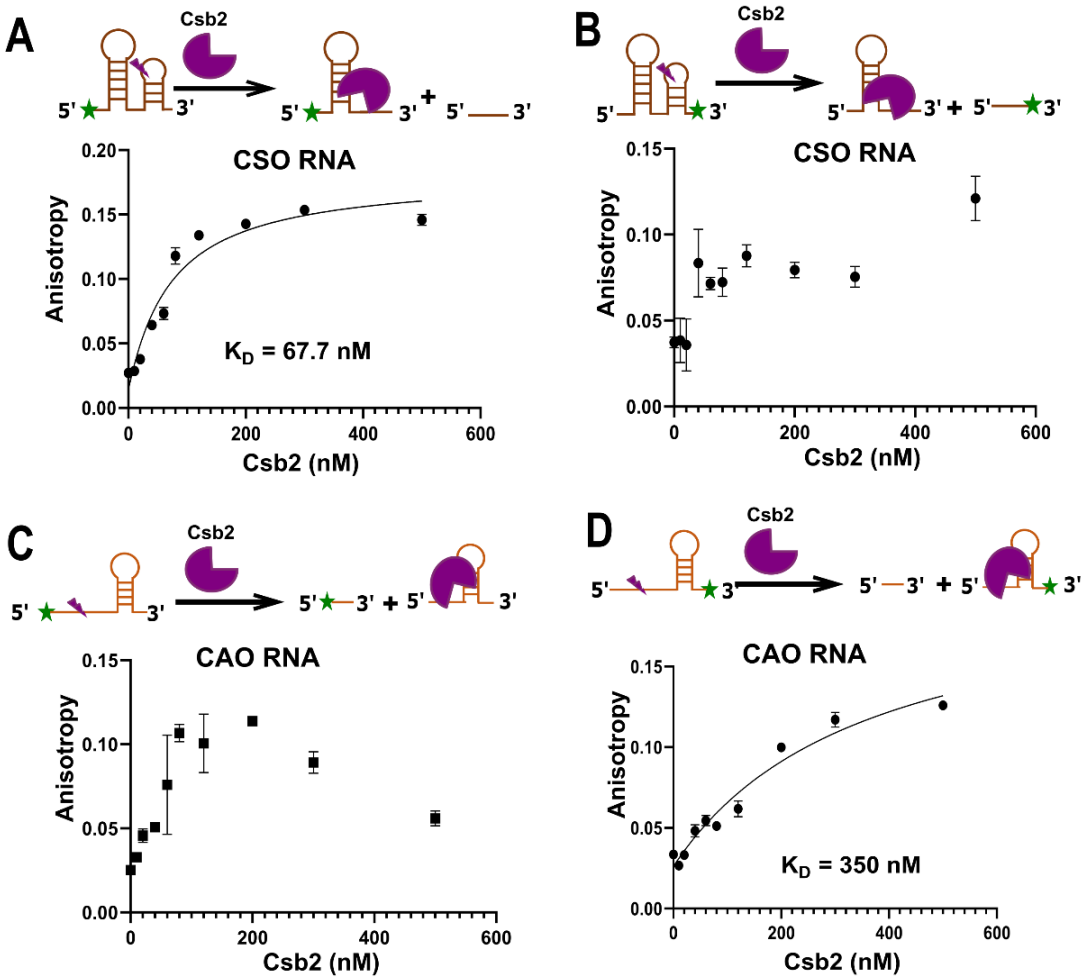


Figure 2. 8: Characterising the substrate specificity of Csb2

A-D. Fluorescence anisotropy analysis of the binding of Csb2 to the CSO (A, B) and CAO (C, D) repeat RNA. The schema above each plot indicates the probable RNA binding region of Csb2 (purple). The 6-FAM label is indicated by a green star. The K_D values were shown if the R^2 value for the fit was greater than 0.8.

2.3.5 Understanding the role of the different domains of Csb2

After verifying the RNase activity of Csb2, our objective was to investigate which of the individual domains within Csb2 houses the nuclease activity. To address this, we conducted domain assignments using the FFAS server (Jaroszewski, Li, Cai, Weber, & Godzik, 2011), which indicated that the N-terminal region (1-250 aa) shares homology with Cas5, while the C-terminal region (251-545 aa) is homologous to Cas6 (Figure 2.9 A&D). To evaluate whether the catalytic activity resides in the Cas5-like or Cas6-like domain (or both), we generated N-terminal (ΔN) and C-terminal (ΔC) deletions. *In vitro* nuclease activity of these truncated proteins demonstrated that ΔN retained the nuclease activity, whereas ΔC exhibited no activity

(Figure 2.9B). This implied that the catalytic activity lies in the C-terminal Cas6-like domain, while the N-terminal Cas5-like domain remains inactive in the type I-G system.

Following the identification of the catalytic domain in Csb2, our next step was to identify the active site residues. We conducted a comprehensive analysis of Csb2 orthologs across 57 organisms through a multiple sequence alignment using Clustal Omega (Sievers et al., 2011). From the conserved residues, we specifically targeted charged residues for site-directed mutagenesis (SDM). The chosen residues (Y14A, E24A, R31A, E65A, Y90A, Y236A, E410A, H520A, and D530A) were substituted with alanine and subsequently, cloned, and purified for the assessment of nuclease activity (Figure 2.10). Interestingly, all mutants, except H520A and D530A, retained RNase activity comparable to the wild type (Figure 2.9 C). This observation suggests that H520 and D530 serve as the catalytic residues in Csb2. Notably, both H520 and D530 are located in the C-terminal region of Csb2, providing additional evidence supporting the catalytic function of the C-terminal Cas6-like domain (Figure 2.9 D).

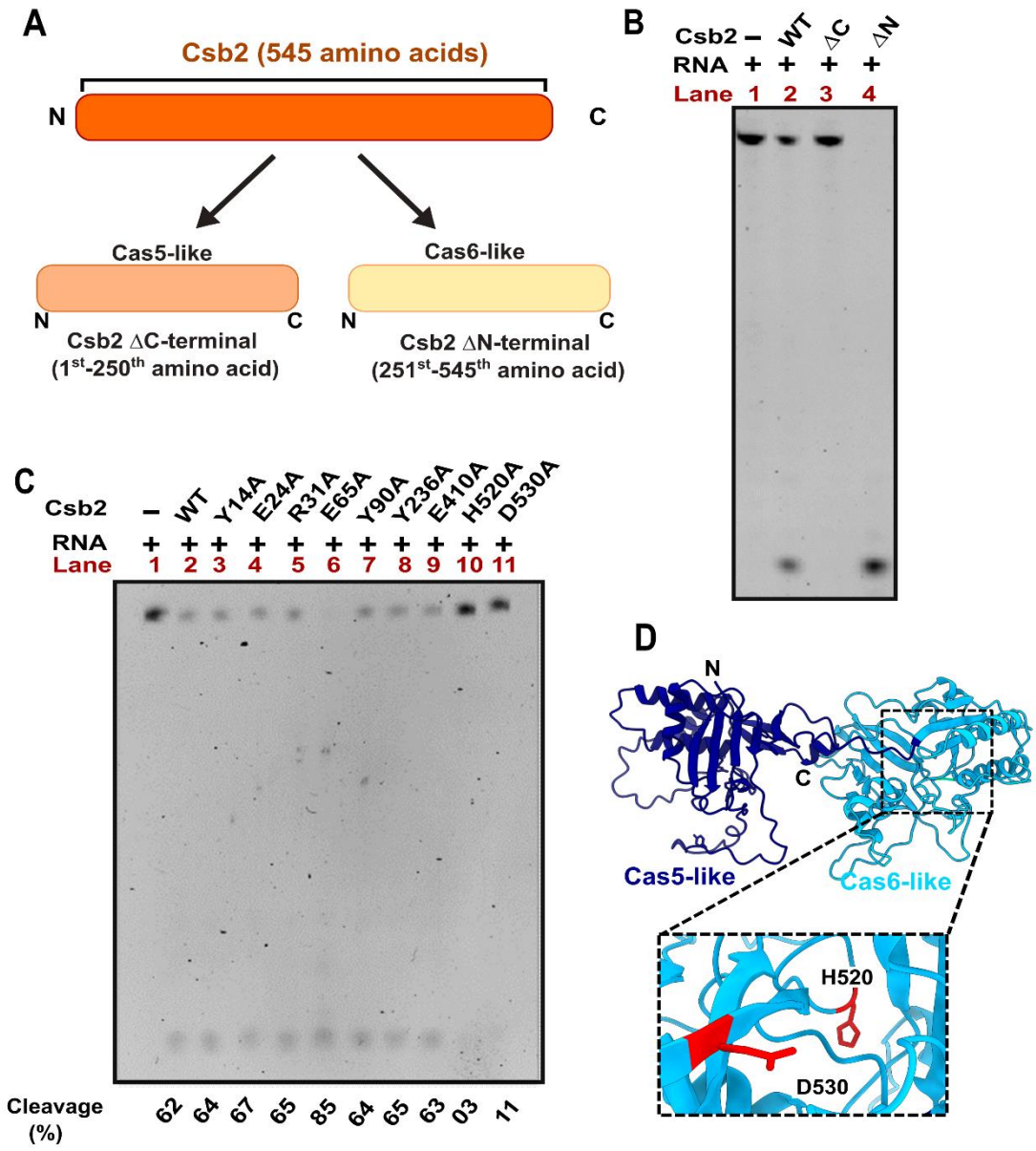


Figure 2. 9: Investigating the catalytic domain of Csb2

A. A schematic representation of the domain assignment of Csb2 (orange) that shows the presence of Cas5-like and Cas6-like domains. FFAS analysis (Jaroszewski et al., 2011) revealed that the N-terminal region of Csb2 from 1st -250th amino acid (peach) is homologous to Cas5, whereas the rest from 251st – 545th amino acid (light yellow) is homologous to Cas6.

B. A 20% denaturing PAGE shows the RNase activity of 1 µM of truncated version of Csb2 such as the ΔC (lane 3) and ΔN (lane 4) on 0.5 µM 5' 6-FAM labelled repeat RNA substrate in comparison to the WT (lane 2).

C. A 20% denaturing PAGE shows the RNase activity of 1 µM of the Csb2 point mutants Y14A, E24A, R31A, E65A, Y90A, Y236A, E410A, H520A, and D530A (lanes 3-11, respectively) on 0.5 µM of 5' 6-FAM labelled repeat RNA substrate in comparison to WT (lane 2). The percentage of cleaved RNA fragment is indicated below the gel.

D. An Alphafold2 predicted 3D structural model of Csb2 indicating the predicted N-terminal Cas5-like domain (in dark blue) and C-terminal Cas6-like domain (in light blue) is shown. The active site residues, namely, H520 and D530 (in red) are shown in the inset.

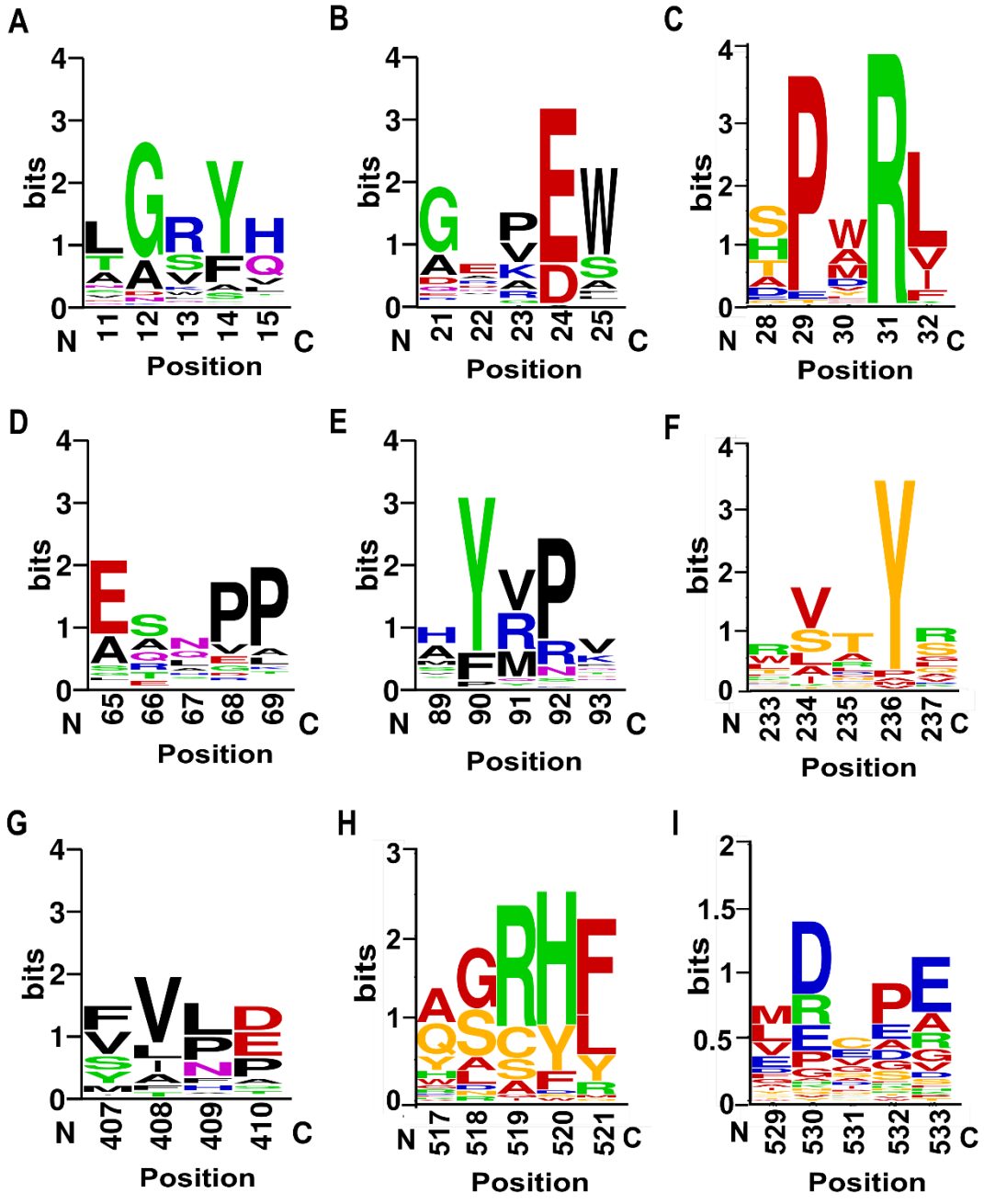


Figure 2. 10: The sequence logo depicts the conservation of amino acids Y14 (A), E24 (B), R31 (C) E65 (D), Y90 (E), Y236 (F) and E410 (G), H520 (H) and D530 (I) across Csb2 proteins (57 organisms) in the type I-G CRISPR-Cas system. The X-axis indicates the amino acid residue number, whereas, the Y-axis indicates the extent of its conservation (in bits).

2.3.6 Investigating the DNase activity of Csb2

Studies have shown that Cas5/I-C is an RNase but also has a tunable DNase activity (Punetha et al., 2014). We wanted to examine whether Csb2 which harbours an N-terminal Cas5 domain, exhibits any DNase activity similar to Cas5/I-C. To explore this, we incubated Csb2 with both circular and linear DNA substrates, in the presence or absence of metal. While we detected nickase activity on circular DNA (dsDNA) at higher concentrations, no significant nuclease activity was observed on linear dsDNA (Figure 2.11 A-D).

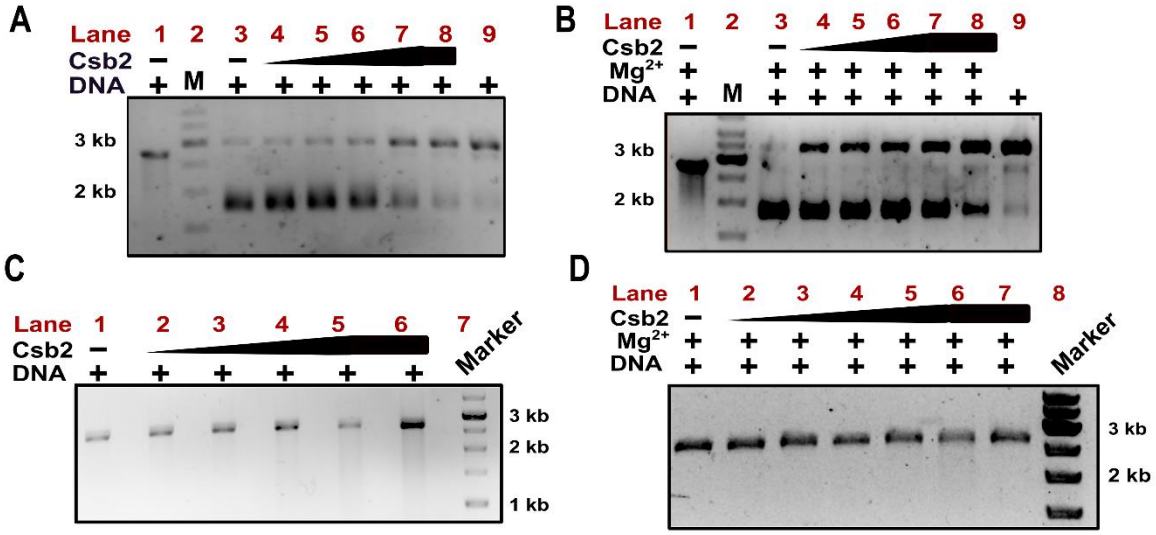


Figure 2. 11: DNase activity of Csb2

A. A 1% agarose gel showing the DNase activity of increasing concentration of Csb2 (0.5, 1, 2.5, 5 µM) on circular dsDNA substrate in the absence of divalent metal ion (Mg^{2+}), (lanes 4-8). Lane 8 consists of DNA treated with 5 µM of Csb2, followed by Proteinase K treatment. A DNA marker is loaded in lane 2, and the respective bands are marked for reference. Lane 1 and lane 9 show linearized substrate DNA and DNA treated with Topoisomerase I, respectively.

B. A 1% agarose gel showing the DNase activity of increasing concentration of Csb2 (0.5, 1, 2.5, 5 µM) on circular dsDNA substrate in the presence of divalent metal ion (Mg^{2+}), (lanes 4-8). Lane 8 consists of DNA treated with 5 µM of Csb2, followed by Proteinase K treatment. A DNA marker is loaded in lane 2, and the respective bands are marked for reference. Lane 1 and lane 9 show linearized substrate DNA and DNA treated with Topoisomerase I, respectively.

C. A 1% agarose gel depicts the DNase activity of increasing concentration of Csb2 (0.5, 1, 2.5, 5 µM) on linear dsDNA substrate in the absence of divalent metal ion (Mg^{2+}), (lanes 2-5). Lane 6 consists of DNA treated with 5 µM of Csb2, followed by Proteinase K treatment. A DNA marker is loaded in lane 7, and the respective bands are marked for reference.

D. A 1% agarose gel shows the DNase activity of increasing concentration of Csb2 (0.5, 1, 2, 3, 5 µM) on linear dsDNA substrate (lanes 2-6) in the presence of divalent metal ion (Mg^{2+}). Lane 7 consists of DNA treated with 5 µM of Csb2, followed by Proteinase K treatment. A DNA marker is loaded in lane 8, and the respective bands are marked for reference.

2.4 Discussion

In most studied type I systems Cas6 (Cas5 in case of type I-C system), is typically responsible for generating the guide crRNA by cleaving within the repeat region of the pre-CRISPR transcript (Carte et al., 2008; Garside et al., 2012; Gesner et al., 2011; Nam et al., 2012; Sashital et al., 2011; R. Wang et al., 2011). Both Cas6 and Cas5 possess ferredoxin-like domains, also known as RNA recognition motif (RRM) domains (Haurwitz et al., 2010; Koo et al., 2013; R. Wang et al., 2011). Therefore, the fusion of Cas5-like and Cas6-like domains in Csb2 raises several questions about its role in type I-G crRNA maturation and the mechanism by which this enzyme recognises and processes its substrate. To address this, we biochemically characterised the type1-G Csb2 protein from *B. animalis*.

We attempted to express Csb2 homologs from three different organisms including the representative organism *G. sulferreducens*. Except for Csb2 from *B. animalis*, our efforts didn't succed for others due to protein insolubility. We were successful in purifying Csb2 from *B. animalis*, but, the protein displayed significant instability *in vivo* (Figure 2.2C), suggesting that the instability might arise from its ectopic expression or its inherent need for a stabilizing partner. Next, we also observed a novel feature of the CRISPR array of type I-G system from *B. animalis*, which harbours bidirectional promoter-like elements for generating pre-crRNA transcripts. We also observed that these RNA transcripts had distinct structures and sequences from both orientations (Figure 2.3 A-C). Interestingly, such bidirectional expression of the CRISPR array had only been observed in type III CRISPR-Cas systems in *Sulfolobus acidocaldarius*, where it was suggested that the complementary base-pairing between the two RNA transcripts may serve to inactivate the processing of pre-crRNA in the absence of an invading foreign genetic element (Lillestol et al., 2006; Lillestol et al., 2009). Further, we demonstrate that both these RNA transcripts can be processed by Csb2, which is unique across CRISPR-Cas systems reported till date (Figure 2.4 A, B). Even more fascinating is the fact that our mapping studies reveal that the sites of Csb2-mediated cleavage are conserved among both the orientations (Figure 2.5). The nature of the RNA fragments produced by Csb2 (generating 5'-OH and 2',3' cyclic phosphate) is characteristic of both Cas5 and Cas6 nuclease activities, which employ metal independent RNA hydrolysis (Carte et al., 2010; Carte et al., 2008; Garside et al., 2012; Gesner et al., 2011; Jore et al., 2012; Nam et al., 2012) (Figure 2.6 A, B). Furthermore, unlike most type I maturases that prefer the base of the stem-loop structure for binding and processing of CRISPR repeats, Csb2 exhibits specificity towards the stem-loop

region for binding while cleaving at the overhang region (Figure 2.7). Here, Csb2 binds to the 5'- proximal stem-loop region of the CSO repeat RNA and cleaves at the 3' overhang region (Figure 2.8 A, B). Whereas, it also binds the 3'- proximal stem-loop region of the CAO repeat RNA while cleaving the 5' overhang region (Figure 2.8 C, D). Collectively, these observations suggest that Csb2 displays an unconventional mode of specificity towards the repeat RNA. Csb2 harbours an N-terminal Cas5-like domain and a C-terminal Cas6-like domain. Here, the nuclease activity resides in the C-terminal Cas6-like domain, while the N-terminal Cas5-like domain is catalytically inert (Figure 2.9B). Previous studies have shown that in type I-C, in the absence of Cas6, Cas5 assumes the catalytic role. Interestingly, Cas5/I-C also possesses a tunable DNase activity (Punetha et al., 2014); however, Csb2/I-G exhibited no substantial DNase activity, suggesting that Cas5-like domain is not active against both DNA and RNA (Figure 2.11).

Csb2 is an endoRNase that can process the repeat regions of the crRNA that are transcribed from both orientations. *In vitro* nuclease assays have shown that Csb2 can act as a standalone maturase, but, in a few other type I systems, like type I-E, Cas6-mediated maturation occurs only after the formation of the multi-subunit Cascade complex. However, Csb2-mediated maturation as a part of the Cascade complex has not been studied yet. We have tried to understand the Cascade formation of the type I-G system and its role in crRNA maturation, in chapter 3.

2.5 Summary

In this chapter, we have characterized a Cas nuclease that is involved in crRNA maturation from a newly identified type I-G CRISPR-Cas system. We assessed the RNase activity on repeat RNA substrates from both the orientations. We showed that Csb2 is a metal-independent RNase that cleaves repeat RNA from both sense and anti-sense orientations, recognising a common motif. We also analysed the type of fragments generated during this cleavage and found that the processing of pre-crRNA generates fragments with 2',3'cyclic phosphates and 5'-OH respectively, that is characteristic of a metal-independent RNA hydrolysis. Additionally, we observed that the stem-loop region of the repeat RNA is important for Csb2 binding, whereas, the 5' overhang of the repeat region is important for the nuclease activity. Csb2 harbours an N-terminal Cas5-like and a C-terminal Cas6-like domains. Amongst

these, the C-terminal Cas6-like domain harbours the catalytic center. We also found two of the catalytically active site residues for Csb2 RNase activity. With these findings, we characterized Csb2 and proceeded further to understand the maturation mechanism in Chapter 3.

Chapter 3: Identification of dual nucleases in type I-G Cascade complex

3 Chapter 3

3.1 Introduction

Having characterised the Csb2 as a site-specific endo-RNase, we next set out to understand the functional role of Csb2 as a subunit of the effector complex. Our curiosity towards the type I-G effector complex stems from the differential mode of Cas6-mediated crRNA processing in other type 1 systems. For example, Cas6a (from type I-A) processes pre-crRNA and delivers the mature crRNA to a pre-formed Cascade complex during a transient association with the type I-A Cascade (Charpentier et al., 2015; Lintner et al., 2011; Plagens, Tjaden, Hagemann, Randau, & Hensel, 2012), suggesting it to be a multi-turnover enzyme complex. On the other hand, Cas6e (type I-E) remains bound to the mature crRNA and facilitates the assembly of the Cascade complex around it (Brouns et al., 2008; Pougach et al., 2010). Similarly, Cas5 also is a single turnover RNase that remains bound to the crRNA and leads to the assembly of the effector complex (Nam et al., 2012; O'Brien et al., 2020). Our biochemical characterization of Csb2 in chapter 2, clearly establishes its ability to process pre-crRNA as a standalone protein *in vitro*. Therefore, we next wanted to investigate whether Csb2 can also process crRNA, as a part of the Cascade complex. Additionally, in I-E Cascade complex, Cas6 caps the backbone of the Cascade complex at the 3' end and Cas5 constitutes the tail region of the complex (Figure 1.7A) (Gesner et al., 2011; Jackson et al., 2014; Mulepati et al., 2014; Sashital et al., 2011; Zhao et al., 2014). Such an arrangement of Cas proteins in other type I Cascade complexes sparked our curiosity regarding the potential association of a Cas5-Cas6 fusion protein (Csb2) with the remaining components of the Cascade complex. Additionally, the type I-G effector complex is anticipated to be more minimalistic (three-component system) compared to other type I effector complexes, enhancing its intrigue for in-depth study of its assembly process (Makarova et al., 2015; Makarova et al., 2020).

Thus, in this chapter we primarily focus on the assembly of the minimal type I-G effector complex. As it was already predicted that Csb1 and Csb3 is homologous to Cas7 and Cas8, respectively, we hypothesised that Csb1, Csb2, and Csb3 could collectively form a Cascade complex analogous to those observed in other type I systems. We then aimed at investigating the pre-crRNA processing abilities of Csb2 as a part of the Cascade complex. Intriguingly, we also observed the involvement of additional Cas proteins besides Csb2, in pre-

crRNA processing. Overall, in Chapter 3, we shed light on distinct features exhibited by the type I-G effector complex, setting it apart from other type I systems.

3.2 Materials and methods

3.2.1 Molecular cloning

To reconstitute the type I-G Cascade complex *in vivo*, a polycistronic construct consisting of *B. animalis* derived genes encoding Csb2, Csb1 and Csb3 (referred as pCascade) was designed via overlap extension PCR and cloned into pQE2 (Qiagen) using Gibson assembly. Cascade complex with Csb2 point mutant (Csb2 H520A) was similarly cloned into pQE2. A type I-G CRISPR array derived from *B. animalis*, containing six identical repeat-spacer units with the Cas anti-sense orientation (CAO), was incorporated under the regulation of a T7 promoter. This was obtained through commercial synthesis (GeneArt) and subsequently cloned into 13S-R (Addgene # 48328) to generate pCR_Array. The gene encoding Csb1 and Csb3 from *B. animalis* were PCR amplified and cloned into pET Strep II co-transformation cloning vector (13S-R Addgene) and pQE2 (Qiagen) vector, respectively, using Gibson assembly.

3.2.2 *In vivo* Cascade complex reconstitution

For *in vivo* reconstitution of the Cascade complex, *E. coli* BL21 (DE3) was co-transformed with pCascade and pCR_Array. The cells were cultivated in LB broth supplemented with 50 µg/mL kanamycin and 100 µg/mL spectinomycin. Upon reaching 0.6 OD_{600}, the cells were induced by adding 0.2 mM IPTG, and the culture was further incubated at 16 °C for an additional 12 hours. Subsequently, cells were harvested, resuspended in binding buffer A (20 mM Tris-Cl, pH 7.6, 30 mM NaCl, 10% glycerol, and 6 mM β-mercaptoethanol), and supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF), Protease inhibitor cocktail (Sigma Aldrich) (20 mg/g of cell pellet), and RNase-free DNase I. Cell lysis was performed using a cell disruptor operating at 20 kpsi. The lysate was clarified through centrifugation at 4 °C and 20,000 g for 30 minutes, followed by an additional

ultracentrifugation step at 4 °C and 120,000 g for 2 hours. The clarified lysate was then loaded to a pre-equilibrated 5 mL HiTrap SP HP (GE Healthcare) column connected in series with 5 mL HiTrap Q HP (GE Healthcare) column. The column was then washed individually with 10 CV of the binding buffer A. The complex was eluted from the HiTrap Q HP column using a linear gradient of 30 mM to 1 M NaCl in the binding buffer A. The eluted fractions were pooled and loaded onto a pre-equilibrated 5 mL HisTrap HP (GE Healthcare) column. After washing with 10 CV of buffer B (20 mM Tris-Cl (pH 7.6), 300 mM NaCl, 30 mM imidazole, 10% glycerol and 6 mM β-mercaptoethanol) the complex was eluted using a linear gradient of 30 mM to 1 M imidazole in buffer B. For the final step of purification, the fractions were loaded onto a pre-equilibrated HiLoad 16/600 Superdex 200 prep grade column (GE Healthcare) and eluted in the storage buffer (20 mM Tris-Cl (pH 7.6), 300 mM NaCl, 10% glycerol and 6 mM β-mercaptoethanol). The complex was concentrated using ultrafiltration through a 100 kDa cutoff filter (Merck) and analyzed on a 10% SDS PAGE for the presence of Cas subunits. Similar to Csb2, the Cascade complex was also observed to undergo spontaneous disintegration within few days even at -80 °C. Therefore, after the complex was purified, all assays were performed immediately or within a couple of days. To ascertain that no dissociated subunits were used while performing the experiments, each time the purified complex was re-passed through a 100 kDa cutoff filter and only the retentate was used for the *in vitro* experiments. The Cascade complex harbouring the Csb2 point mutant, namely pC2H_Cascade was reconstituted similarly.

3.2.3 crRNA extraction from Cascade complex

To isolate the crRNA from the intact Cascade complex, 500 µL of the 20 µM sample was heated at 95 °C for 15 min. It was then subjected to phenol-chloroform extraction, followed by ethanol precipitation. The sample thus obtained was treated with RNase-free DNase I (50 µg/mL) at 37 °C for 1 hour to remove any traces of DNA. The RNA that remained was visualized using an EtBr stained 20% denaturing polyacrylamide gel.

3.2.4 Purification of proteins

For the purification of Csb1, *E. coli* BL21 (DE3) cells carrying pCsb1 were cultured in LB broth supplemented with 50 µg/mL spectinomycin at 37 °C until the optical density at 600 nm (OD600) reached 0.6. The cells were then induced with 0.2 mM IPTG and were expressed at 25 °C for 12 hours. Following expression, cells were collected by centrifugation at 25 °C, 10,000 g for 5 minutes. The harvested cell pellets underwent washing and resuspension in binding buffer C (20 mM Tris-Cl, pH 7.6, 500 mM NaCl, and 6 mM β-mercaptoethanol), supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) and Protease inhibitor cocktail (Sigma-Aldrich) (20 mg/g of cell pellet). Cells were lysed through sonication (Pulse - 2 sec ON and 20 sec OFF), and the resulting lysate was clarified by centrifugation at 4 °C and 20,000 g for 30 minutes. The supernatant was loaded to a pre-equilibrated 5 mL StrepTrap HP (GE Healthcare) column and subsequently washed with 10 CV of binding buffer C, and eluted in binding buffer containing 2.5 mM d-Desthiobiotin (Sigma-Aldrich). The eluted fractions were further refined using a HiLoad 16/600 Superdex 75 prep-grade column (GE Healthcare). The purified protein was concentrated via ultrafiltration through a 50 kDa cutoff filter (Sartorius), snap-frozen, and stored at -80 °C until needed for subsequent use.

To purify Csb3, *E. coli* BL21 (DE3) harboring pCsb3 was grown in LB broth supplemented with 100 µg/mL ampicillin at 37 °C until OD_{600} reached 0.6 and induced with 0.2 mM IPTG. The expression was continued at 25 °C for 12 hours and then the cells were harvested by centrifugation at 25 °C, 10,000 g for 5 min. The pellet was then resuspended in binding buffer D (20 mM Tris-Cl (pH 7.6), 300 mM NaCl, 30 mM imidazole and 6 mM β-mercaptoethanol) containing 1mM phenylmethylsulfonyl fluoride (PMSF) and Protease inhibitor cocktail (20 mg/g of cell pellet). After sonication (Pulse - 2 sec ON and 20 sec OFF), the lysate was centrifuged at 4 °C and 20,000 g for 30 min. The supernatant was loaded onto a pre-equilibrated 5mL HisTrap HP (GE Healthcare) column. After the sample binding, the column was washed with 10 CV of binding buffer B and eluted on a linear gradient of imidazole (30 mM to 1 M) in buffer D. The eluted fractions were loaded onto a pre-equilibrated HiLoad 16/600 Superdex 75 prep grade column (GE Healthcare) and purified further. The protein was concentrated using ultrafiltration through a 30 kDa cutoff filter (Sartorius), snap-frozen and stored at -80 °C until further used.

3.2.5 *In vitro* RNase assay

To test whether Csb2 could process repeat RNA as a part of the Cascade complex, 0.5 µM of the CAO repeat RNA (5' 6-FAM labelled) substrate was incubated with 1 µM of WT or its mutant variant (Csb2 H520A) Cascade complex in a reaction buffer containing 20 mM Tris-Cl (pH 7.6), 100 mM KCl and 6 mM β-Mercaptoethanol (β-ME) and incubated at 37 °C for 30 min. Cleavage products were analysed on a 20% (w/v) denaturing polyacrylamide gel and viewed in Bio-Rad gel documentation system. Similar reactions were set up to assess the RNase activity of Csb1 and Csb3.

3.3 Results

3.3.1 *In vivo* reconstitution of the Cascade complex

Following the characterization of the nuclease activity of Csb2, our objective was to examine whether Csb2 could process the pre-CRISPR transcript too as part of the effector complex. To achieve this, the initial step involved the successful reconstitution of the Cascade complex. In other type I systems, core components such as Cas5, Cas6, Cas7, and Cas8 typically constitute the Cascade complex in conjunction with crRNA. Based on our predictions, Csb1 shares homology with Cas7, Csb2 contains N-terminal Cas5-like and C-terminal Cas6-like domains, and Csb3 is homologous to Cas8. Consequently, these proteins were identified as potential constituents of the Cascade complex in the type I-G system. To test the assembly of these proteins into the Cascade complex, we cloned *csb1*, *csb2*, and *csb3* from *B. animalis* as a polycistronic construct via overlap extension PCR (pCascade) (Figure 3.1 A-D) and co-expressed them with the CRISPR repeat-spacer array (pCR_Array) (Figure 3.1E and 3.2A). We purified the intact complex through series of purification steps followed by final size-exclusion chromatography step (Figure 3.2B) (refer to methods for details, section 3.2.2). Notably, during the analysis of the purified complex on SDS-PAGE, we successfully identified these proteins, indicating that Csb1, Csb2, and Csb3 indeed form a complex with crRNA. This analysis also suggests that Csb1 is the most enriched whilst Csb2 and Csb3 might be present in lower stoichiometries (Figure 3.2 C). Further, the nucleic acid isolated from the purified

complex was treated with DNase, following which, a distinct band was observed on a denaturing PAGE, with its size corresponding to the predicted length (approximately 60-70 nt) of crRNA (Figure 3.2D). This additional evidence strongly suggests that Csb1, Csb2, and Csb3 co-assemble to form a complex with crRNA.

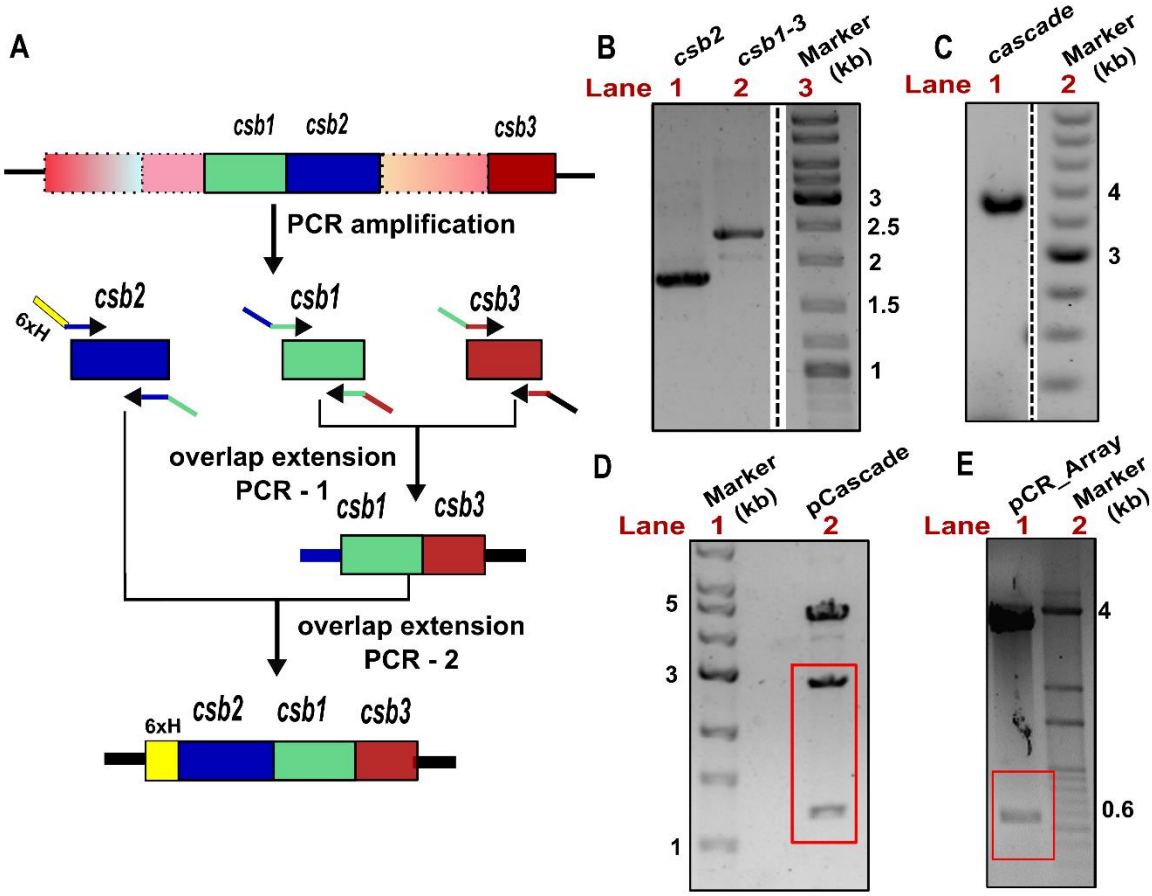


Figure 3. 1: Cloning of the type I-G Cascade operon

A. A schematic representation of the experimental outline of constructing the type I-G Cascade operon through overlap-extension PCR. The genes encoding Csb1, Csb2 and Csb3 were amplified separately using specific primers. *csb1* and *csb3* were amplified via overlap extension PCR to give *csb1-csb3* amplicon which was later combined with *csb2* via a second overlap-extension PCR reaction to generate *csb2-csb1-csb3* (*cascade*) operon.

B. A 0.8% agarose gel showing PCR amplicons *csb2* (1.6 kb, lane 1) and *csb1-csb3* (2.2 kb, lane 2) overlap. A DNA marker is loaded in lane 3, and the respective bands are marked for reference. The dashed line denotes discontinuity in the gel for the purpose of clarity.

C. A 0.8% agarose gel showing PCR amplicon of the cascade operon, *csb2- csb1-csb3* (3.8 kb, lane 1). A DNA marker is loaded in lane 2, and the respective bands are marked for reference. The dashed line denotes discontinuity in the gel for the purpose of clarity.

D. A 0.8% agarose gel showing the clone confirmation of pCascade construct using restriction digestion. The gene of interest released after digestion in shown in red box. A DNA marker is loaded in lane 1, and the respective bands are marked for reference.

E. A 0.8% agarose gel showing the clone confirmation of pCR_Array construct using restriction digestion. The gene of interest released after digestion in shown in red box. A DNA marker is loaded in lane 2, and the respective bands are marked for reference.

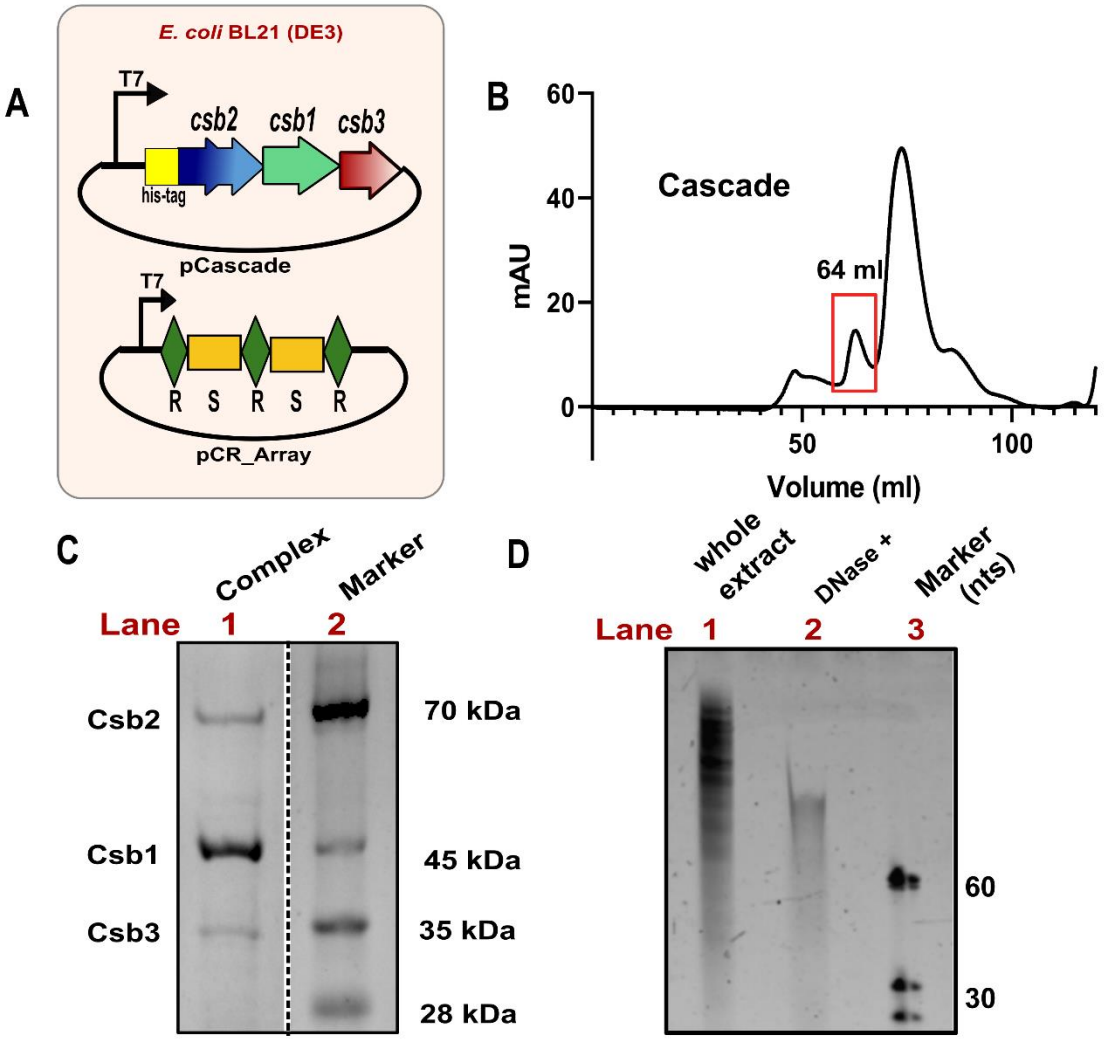


Figure 3. 2: *In vivo* Cascade reconstitution

A. A schematic representation of the constructs used for overexpression and *in vivo* reconstitution of the Cascade complex. The plasmids expressing the Cascade complex (pCascade) and crRNA (pCR_Array) were co-expressed upon IPTG induction in *E. coli* BL21 (DE3).

B. A size exclusion chromatogram of the HiLoad 16/600 Superdex 200 prep grade column indicating the peak corresponding to the Cascade complex (red box), eluted at 64.5 mL is shown.

C. A 10% SDS PAGE shows the individual subunits of the purified Cascade complex (lane 1) extracted from the peak fraction in (B). A protein marker is loaded on lane 2 to validate the molecular sizes. The respective subunits, namely, Csb1, Csb2 and Csb3 are also indicated on the right. The dashed line indicates the discontinuity in the gel for clarity.

D. An EtBr stained 20% denaturing PAGE shows the RNA isolated from the purified Cascade complex. Lane 1 shows the nucleic acid extract from the Cascade complex, while lane 2 displays the same extract after treatment with DNase, confirming the presence of RNA. An ssDNA marker is loaded onto lane 3 for reference.

3.3.2 Pre-crRNA processing by the Cascade complex *in vitro*

Inspired by these findings, we investigated whether the effector complex possesses the capability to process CRISPR repeat RNA. Remarkably, we observed that the effector complex exhibited endonuclease activity when incubated with CRISPR repeat RNA, and this activity was more efficient compared to the Csb2 nuclease activity (Figure 3.3 B-C). This suggests that the pre-CRISPR transcript undergoes processing by Csb2 both as an independent protein and as a subunit within the RNP effector complex.

To confirm that the nuclease activity of the effector complex is attributed to Csb2, we reconstituted the complex *in vivo* with a Csb2 nuclease-dead mutant (H520A) (Figure 3.3A). Intriguingly, this mutant Cascade complex exhibited nuclease activity similar to the WT Cascade complex, suggesting the presence of an additional nuclease within the complex (Figure 3.3 B-C).

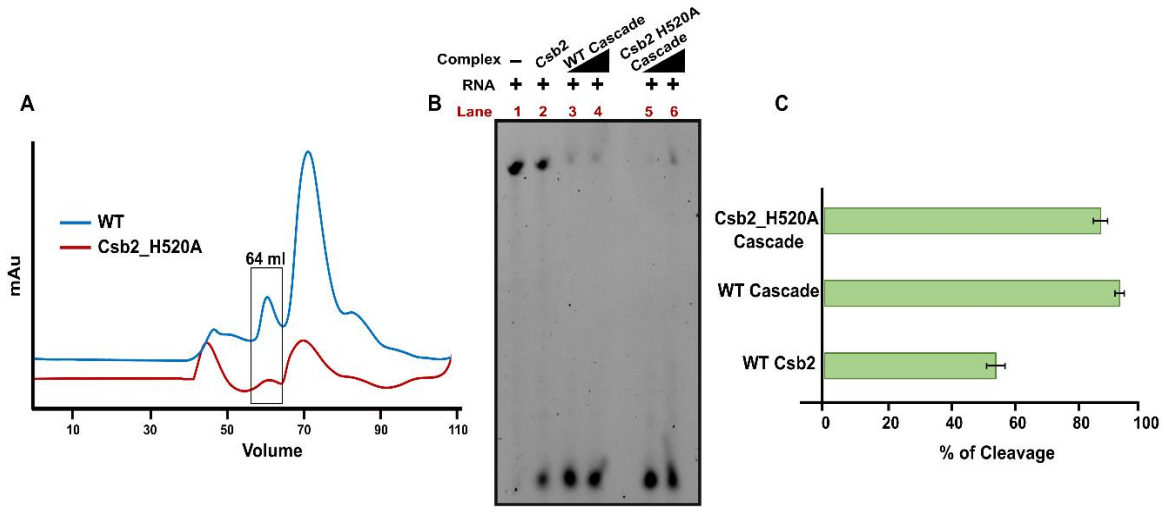


Figure 3. 3: *In vitro* nuclease assay of the Cascade complex

A. An overlay of size exclusion chromatograms of the HiLoad 16/600 Superdex 200 prep grade column for comparison of elution profiles of the WT (in blue) and the Csb2_H520A (in red) Cascade complexes. The desired complex is eluted at approximately 64 mL for both the complexes (indicated by the black box).

B. A 20% denaturing PAGE shows the RNase activity of 1 µM of Cascade complex (lane 3, 4) as well as Cascade (Csb2 H520A) complex (lane 5, 6) on 0.5 µM of 5' 6-FAM labelled repeat RNA substrate. The Csb2 that was incubated with RNA substrate is shown in lane 2 as a control.

C. A bar graph showing the densitometric analysis of the RNase activity of Wt Csb2 compared to WT Cascade complex as well as its mutant variant (Csb2_H520A), whose representative analysis is shown in panel B.

3.3.3 Identifying the additional catalytic center in type I-G effector complex

The observation of the mutant Cascade complex (Csb2 H520A) retaining RNase activity (Figure 3.3B) led us to formulate a hypothesis regarding the existence of another RNase within the Cascade complex. To explore this possibility, we turned towards investigating the nuclease activity of other Cas subunits – Csb1 and Csb3. Csb1 was predicted to be homologous to Cas7 protein, whereas, Csb3 was homologous to Cas8 protein. We then cloned and expressed both Csb1 and Csb3 separately and tested for their respective nuclease activities (Figure 3.4 A-B). *In vitro* nuclease activity of Csb1 and Csb3 against 5' 6-FAM labelled CAO RNA revealed that Csb3 exhibited no cleavage, whereas, Csb1 demonstrated RNase activity (Figure 3.4C). This observation supports our hypothesis that the Cascade complex possesses additional Cas nuclease. We therefore proceeded to characterise Csb1 further.

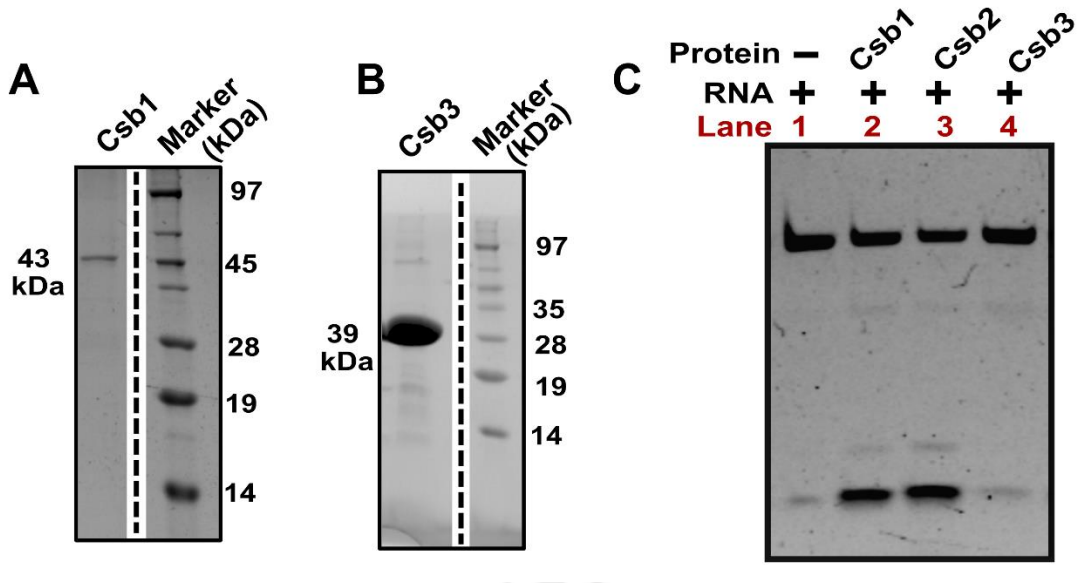


Figure 3. 4: Dual nuclease activity of type I-G Cascade subunits

A. A 15% SDS PAGE shows the purified 43 kDa Csb1 protein (marked as Csb1). A protein marker is loaded on the left to validate the molecular sizes. The vertical broken line indicates discontinuity between the lanes, which is introduced for the sake of clarity.

B. A 15% SDS PAGE shows the purified and 39 kDa Csb3 protein (marked as Csb3), respectively. A protein marker is loaded on the left to validate the molecular sizes. The vertical broken line indicates discontinuity between the lanes, which is introduced for the sake of clarity.

C. A 20% denaturing PAGE depicting the RNase activity of Csb1 (lane 2), Csb2 (lane 3), and Csb3 (lane 4) on 5' 6-FAM labelled CAO repeat RNA substrate.

3.4 Discussion

In type I CRISPR-Cas systems, the primary function of Cas6 is to process the pre-crRNA and subsequently co-assemble along with this crRNA and other Cas proteins to form the Cascade complex (Gesner et al., 2011; Sashital et al., 2011). The most widely studied type I-E Cascade complex consists of $Cas7_7$, $Cas6_1$, $Cas5_1$, $Cas8_1$ and $Cas11_2$, where the subscript indicates the stoichiometry of each subunit (Gesner et al., 2011; Sashital et al., 2011). In contrast, previous predictions (Makarova et al., 2015; Makarova et al., 2020; Shangguan et al., 2022) and our *in vivo* reconstitution experiments suggest that type I-G Cascade complex is constituted of only three components – Csb1 (Cas7 homolog), Csb2 (Cas5-Cas6 fusion), Csb3 (Cas8 homolog), thus representing a minimalistic type I effector complex (Figure 3.2 B-C). Interestingly, our work and recent reports (Shangguan et al., 2022) have confirmed that the backbone forming Cas7 homolog, Csb1, is the most enriched component of this effector complex while Csb2 and Csb3 are represented in lower stoichiometry. We could also extract the crRNA to further ascertain the assembly of crRNA with Cas proteins (Figure 3.2 D). Our reconstruction of the Cascade complex aligns with a recent study that documented the formation of a backbone in the type I-G effector complex, consisting of seven subunits of Csb1 (Shangguan et al., 2022). Our *in vitro* nuclease assays indicated that the type I-G Cascade could process crRNA in a notably more efficient manner compared to the action of Csb2 alone (Figure 3.3 B). Notably, we also discovered an additional nuclease, Csb1, which demonstrated the ability to cleave the repeat RNA similar to Csb2 (Figure 3.4 C). The presence of an additional nuclease aligned with the effective processing observed in the Cascade complex, compared to a single Cas protein. Catalytic activity of Cas7 homologs has not been previously reported in type I systems. However, type III CRISPR-Cas system harbours a catalytically active variant of Cas7 protein that cleaves target RNA (Osawa et al., 2015; Ozcan et al., 2021; Samai et al., 2015; R. H. Staals et al., 2014; D. W. Taylor et al., 2015; van Beljouw et al., 2021).

Therefore, our findings on the bidirectional transcription of the CRISPR array (as discussed in Chapter 2) and the catalytic behaviour of the Cas7 homolog Csb1 (as presented in the current chapter) bear similarity to the type III system. This suggests that the type I-G CRISPR-Cas system could potentially serve as an evolutionary link between type I and type

III systems. Moreover, it also implies that the surplus of the CRISPR array transcripts that are expressed in this system can be now attributed to the presence of dual nucleases that can process the RNA. This prompted a deeper exploration of the novel RNase activity exhibited by Csb1. Therefore, in our forthcoming Chapter 4, our objective is to meticulously characterize the RNase activity of Csb1 and compare it with Csb2.

3.5 Summary

In this chapter, we have accomplished the *in vivo* reconstitution of the type I-G Cascade complex from *B. animalis*. We evaluated the RNase activity of the Cascade complex and found that Csb2 could cleave repeat RNA both as a standalone protein and as a part of the complex. To ascertain this, we evaluated the nuclease activity of a variant of Cascade complex comprising of Csb2-point mutant (dead mutant) and found that the Cascade complex could still retain the processing ability. This prompted us to investigate the additional catalytic center. We observed that amongst Csb1 and Csb3, Csb1 harboured the additional catalytic center in the type I-G Cascade complex. With these observations, we identified an additional nuclease in type I-G effector complex, which is strikingly different from that observed previously (Shangguan et al., 2022) and proceeded further to characterize it in Chapter 4.

Chapter 4: Identification of Csb1 as an additional nuclease in Cascade complex

4 Chapter 4

4.1 Introduction

Having discovered Csb1 as an additional RNase within the type I-G system, we focused on a more in-depth characterization of the biochemical properties of this Cas nuclease. Csb1, a Cas7 homologue belonging to the RAMP superfamily that harbours a single RRM domain (Makarova, Aravind, et al., 2011; R. Wang & Li, 2012), stands out, as our studies unveil that it has acquired a catalytic trait, a feature not observed in any other type I systems. Usually, in type I systems, a sea-horse shaped Cascade complex is composed of Cas7 that primarily stabilizes its backbone by assembling with the crRNA to form a hand-like shape, the thumb region of which induces a kink in the crRNA every sixth nucleotide (Jackson et al., 2014). Cas7 also interacts with the Cas8 (large-subunit) and Cas11 (small-subunit) proteins to constitute the tail and the belly region of the Cascade complex respectively (Jackson et al., 2014; Jore et al., 2011; Mulepati et al., 2014; Nam et al., 2012; O'Brien et al., 2020; Wiedenheft et al., 2011; Zhao et al., 2014). In contrast to this, within type III CRISPR-Cas systems, a catalytic variant of Cas7 (Csm3/Cmr4) is responsible for cleaving target RNA (Osawa et al., 2015; Samai et al., 2015; R. H. Staals et al., 2014; Tamulaitis et al., 2014; D. W. Taylor et al., 2015). Initially, Cas7 in type III systems was recognized for its involvement in the interference mechanism. However, a recent discovery uncovered gRAMP Cas7-11, which was identified to play a role not only in interference but also in the crRNA maturation (Goswami et al., 2022; Ozcan et al., 2021; van Beljouw et al., 2021; Yu et al., 2022). The *cas7-11* gene is composed of four subunits of *cas7*, with an interruption in the last subunit caused by an insertion element. The first subunit is followed by *cas11*. Notably, the processing of pre-crRNA is primarily attributed to Cas7.1 and Cas7.2 subunits within the complex (Goswami et al., 2022; Ozcan et al., 2021; van Beljouw et al., 2021; Yu et al., 2022).

Our current findings indicate that Csb1, the Cas7 homolog in the type I-G CRISPR-Cas system, possesses an additional catalytic centre, setting it apart from the characteristics observed in Cas7 proteins in other type I systems. Intrigued by this, we aimed at characterizing the RNase activity of Csb1, through a series of detailed experiments that were designed to characterize Csb2 in Chapter 2. Throughout our chapter, we also try to compare our observations with that of Csb2. By doing so, we also attempted to understand the nature of this

additional catalytic centre which can eventually help us to understand the physiological basis of harbouring two catalytic centres in type I-G maturation mechanism.

4.2 Materials and methods

4.2.1 Molecular cloning

The gene encoding Csb1 from *B. animalis* was PCR amplified and cloned into pET Strep II co-transformation cloning vector, 13S-R (Addgene) vector using Gibson assembly (as described in Chapter 3, Section 3.2.1). The point mutants of Csb1, namely, E78A, H122A, R123A, R179A and R320A were created using PCR-based mutagenesis and cloned into 13S-R vector similarly.

4.2.2 Purification of proteins

Expression and purification of Csb1 and its mutant proteins were carried out as described in Chapter 3 (section 3.2.4). To study the oligomerisation pattern of Csb1, purified protein is passed through the HiLoad 16/600 Superdex 200 prep grade column (GE Healthcare) and the peak fractions were collected for all oligomeric states, followed by analysis on an SDS PAGE and utilized for biochemical assays. For TEM imaging, the affinity purified Csb1 samples were used and the TEM grids were examined using a JEOL (Model 2100F) transmission electron microscope operated at 200 kV.

4.2.3 Preparation of substrates

Repeat RNA substrates used for the studies are described in Chapter 2. Details of its synthesis and purification are described in Section 2.2.4.

4.2.4 *In vitro* RNase assay

Assays for *in vitro* RNase activity of the Csb1 mutants were carried out as described in Chapter 3 (Section 3.2.5). The time-dependent nuclease assay of Csb1 was performed similar to that of Csb2 (mentioned in Section 2.2.5 in Chapter 2).

4.2.5 Cleavage site mapping assay

To map the cleavage site on the RNA substrate by Csb1, we conducted reactions following a protocol akin to that outlined for Csb2 in Section 2.2.6 of Chapter 2.

4.2.6 Assay for characterisation of the nature of the ends of the fragment

To characterise the nature of the fragment generated by Csb1 activity on the repeat RNA, we took the same approach as described in Section 2.2.7 of Chapter 2.

4.2.7 Fluorescence anisotropy assay

The anisotropy assays were carried out as described in Section 2.2.8 in Chapter 2.

4.3 Results

4.3.1 Csb1 RNase activity

Preliminary experiments with Cascade complex harbouring a mutant version of Csb2 (H520) suggested that Csb1 was catalytic in nature and can process pre-crRNA (Figure 3.4C). This prompted us to further characterise the role of Csb1 in Cascade complex that also harbours the Csb2 nuclease. Towards this, we first assessed its metal dependency and noted that, akin to Csb2, the RNase activity of Csb1 was metal-independent (Figure 4.1A). Time-dependent cleavage assay of Csb1 at various substrate to enzyme ratios (1:4 and 1:10) was carried out. At the 1:4 (substrate: enzyme) ratio, we observed significant processing, which doesn't show any

further cleavage even after prolonged treatment, suggesting that Csb1 is an endonuclease that cleaves at a specific region within the repeat RNA (Figure 4.1B). Whereas, no significant processing was observed while using a substrate: enzyme ratio of 1:10, suggesting longer incubation or high protein concentrations are needed to elicit Csb1 mediated RNA maturation. (Figure 4.1C). Comparing this with that of Csb2 activity at the same ratio (1:10) indicated that Csb2 is a more potent nuclease than Csb1 (Figure 4.1D). We performed alkaline hydrolysis and RNase T1 based structural mapping to identify the site of cleavage within RNA substrates which revealed that Csb1 cleaves the CSO repeat RNA at 8 nt from the 3' end, that is between C_{28} and A_{29} (Figure 4.2 A, D) and the CAO repeat RNA 5 nt from the 5' end, between C_5 and A_6 (Figure 4.2 B, D). Notably, the observed cleavage site for Csb1 matched precisely with the cleavage site for Csb2. On careful analysis, we identified a conserved motif at the site of cleavage (5' U**CA**A/UU 3', where **C** and **A** are the sites of cleavage) (Figure 4.2D).

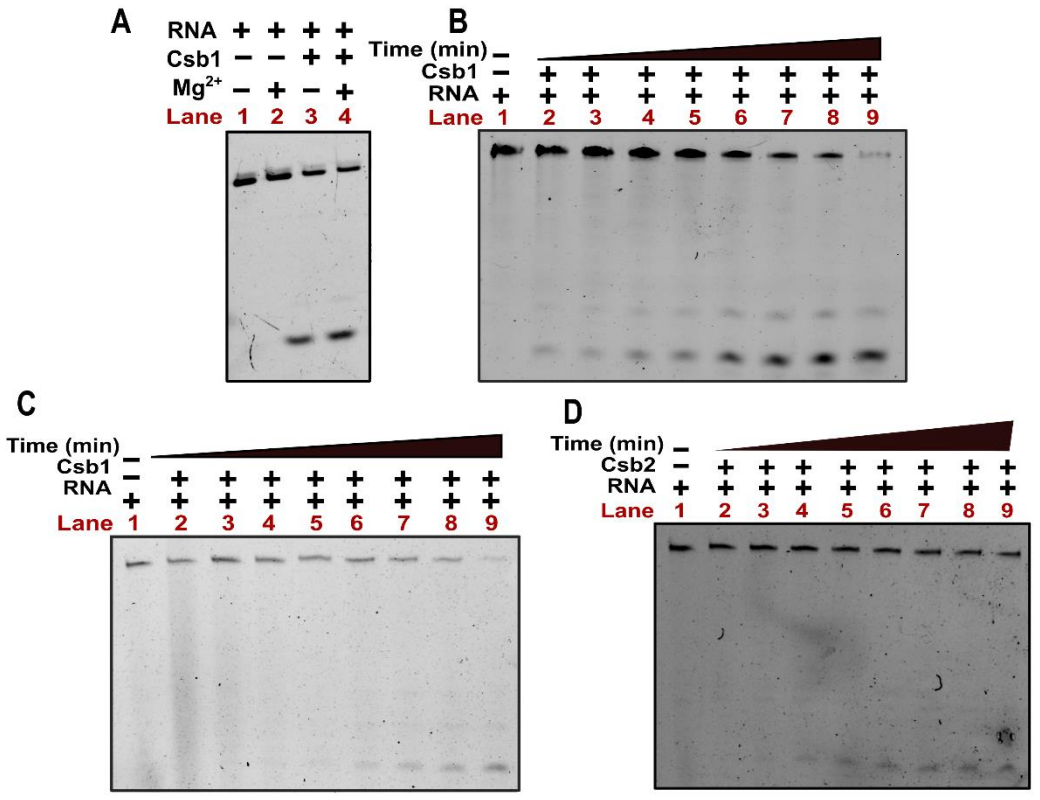


Figure 4. 1: RNase activity of Csb1

A. A 20% denaturing PAGE showing the RNase activity of EDTA treated Csb1 (1 µM) on 5' 6-FAM labelled CAO repeat RNA substrate (0.25 µM), in presence of metal.

B. A 20% denaturing PAGE demonstrating the time-dependent (0, 5, 10, 15, 30, 60, 90 and 120 min) RNase activity of Csb1 (1 µM) on 5' 6-FAM labelled CAO repeat RNA substrate (0.25 µM).

C. A 20% denaturing PAGE demonstrating the time-dependent (0, 5, 10, 15, 30, 60, 90 and 120 min) RNase activity of Csb1 (100 nM) on 5' 6-FAM labelled CAO repeat RNA substrate (10 nM).

D. A 20% denaturing PAGE showing the time-dependent (0, 5, 10, 15, 30, 60, 90 and 120 min) RNase activity of Csb2 (100 nM) on 5' 6-FAM labelled WT CAO repeat RNA substrate (10 nM).

Next, we also studied the nature of the RNA ends produced by Csb1 by treating the CAO 5' 6-FAM labelled cleavage product with T4 Polynucleotide Kinase (PNK) under acidic conditions (Das & Shuman, 2013). Similar to Csb2, T4 PNK treated fragments migrated at a slower rate compared to the untreated fragments, suggesting the conversion of the phosphate group to a less negatively charged hydroxyl group (Figure 4.2 C). Collectively, this data established parallels between modes of RNA processing between Csb1 and Csb2.

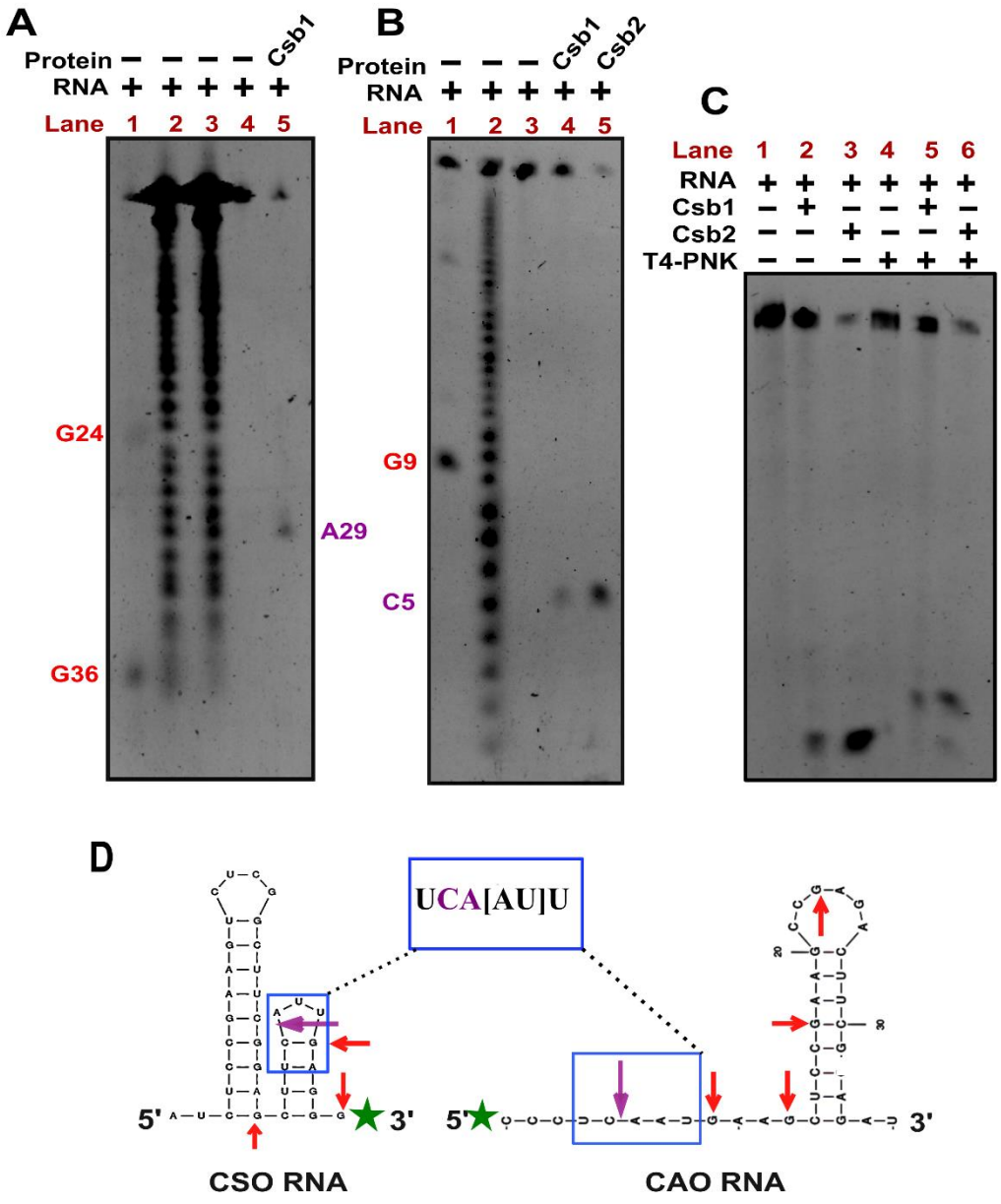


Figure 4. 2: Characterizing the RNase activity of Csb1

A. A 20% denaturing PAGE is shown for mapping the cleaved CSO RNA fragment by Csb1 (lane 5) using an RNase T1 digestion ladder (lane 1) and an alkaline hydrolysis ladder (lane 2 and 3) generated by varying the hydrolysis incubation times (5 and 10 min at 95 °C), respectively. Positions specific to cleavage by RNase T1 are indicated as G24 and G36 in red. The cleaved RNA fragment indicated as A29 (in purple) maps to the 29th nucleotide from the 5' end of the repeat RNA between C_{28} and A_{29}.

B. A 20% denaturing PAGE mapping the cleaved CAO RNA fragment by Csb1(lane 4) using an RNase T1 digestion ladder (lane 1) and an alkaline hydrolysis ladder (lane 2), is shown. The RNA substrate that was incubated with Csb2 separately in lane 5 is also indicated. Position specific to cleavage by RNase T1 is indicated as G9 in red. The cleaved RNA fragment indicated as C5 (in purple) maps to the 5th nucleotide from the 5' end of the repeat RNA between C_5 and A_6.

C. A 20% denaturing PAGE is shown for deciphering the nature of cleaved RNA product by Csb1 (lane 5). The cleaved RNA product was treated with T4 PNK at pH 5.2 in the absence of ATP to remove terminal phosphates. The difference in migration between T4 PNK treated and untreated samples is attributed to the charge difference between these groups. Similarly, Csb2 cleaved RNA fragment was also treated with T4 PNK and loaded in lane 6 for comparison.

D. A schema showing the site of cleavage by Csb1 on both the RNA substrates, along with the sites of cleavage by RNase T1 are indicated on the predicted repeat RNA structures, using purple and red arrows respectively. The conserved sequence motif 'U**CA**[AU]U' is also highlighted.

4.3.2 Substrate specificity of Csb1

Next, we wanted to test the importance of the 5' overhang region and the stem-loop region of the RNA substrate (Chapter 2, Section 2.3.4). Towards this, we tested the substrate specificity of Csb1 using similar RNA variants as described in Chapter 2. On incubating Csb1 with an RNA substrate with a smaller stem-loop region, we observed slightly reduced cleavage activity. Further, transition and transversion mutations in the sequences upstream of the cleavage site also led to an inability in the processing. However, a significant reduction in the nuclease activity was observed while using substrates that lacked the 5' overhang region, suggesting that this region is crucial for the processing of repeat RNA similar to the case in Csb2 (Figure 4.3 A, B).

To further evaluate the binding affinity of Csb1 to the repeat RNA, we performed the anisotropy-based assay employed in Chapter 2. In this assay, we incubated Csb1 with both CSO and CAO repeat RNA substrates, each carrying a fluorescent label at either 5' or 3' end. The anisotropy analysis revealed that Csb1 exhibits binding affinity to the stem-loop region of the CSO RNA but unlike Csb2, it did not demonstrate any measurable binding affinity towards the stem-loop region of the CAO RNA, suggesting that the mode of RNA binding may vary between Csb1 and Csb2 (Figure 4.3 C-F).

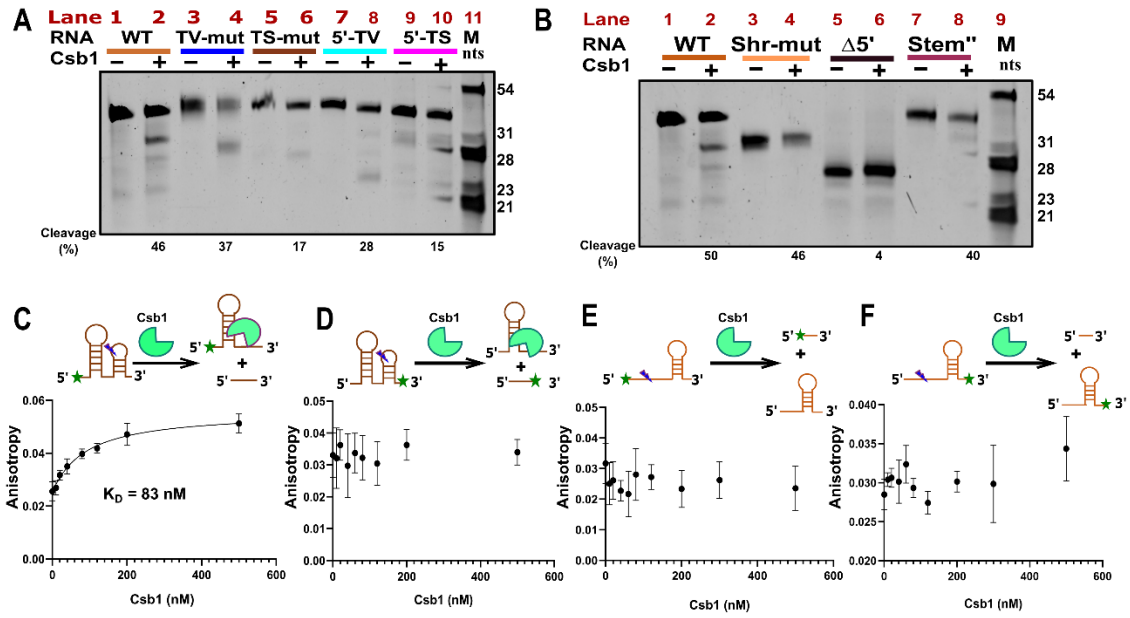


Figure 4. 3: Substrate specificity of Csb1

A. An EtBr stained 20% denaturing PAGE shows the nuclease activity of Csb1 (0.5 µM) on WT, TV-mut, TS-mut, 5'-TV and 5'-TS mutant (0.25 µM) variants of CRISPR repeat RNA as described in Figure 2.6 (Chapter 2).

B. An EtBr stained 20% denaturing PAGE shows the nuclease activity of Csb1 (0.5 µM) on WT, Shr-mut, Δ5' and Stem" mutant (0.25 µM) variants of CRISPR repeat RNA as described in Figure 2.6 (Chapter 2).

C-F. Fluorescence anisotropy analysis of the binding of Csb1 to the stem-loop region or the overhang region of the CSO (**C, D**) and CAO (**E, F**) repeat RNA. The schema above each plot indicates the probable binding preference by Csb1 (green). The FAM label is indicated by a green star. The K_d values for the samples were reported only if R^2 for the non-linear least square regression was greater than 0.8.

4.3.3 RNA-independent oligomerization of Csb1

RNA binding studies of Csb1 suggested that its mode of recognition and binding may vary as compared to Csb2. Additionally, repeated Size exclusion Chromatography (SEC) runs of highly pure Csb1 protein displayed broad elution profiles in contrast to distinct elution peaks typically expected of soluble monomeric proteins (Figure 4.4 A, B). This suggested that Csb1 may exist in different oligomeric states which in turn may affect its RNA binding and interaction properties. To test this, we first carried out a negative stain-TEM analysis of SEC purified Csb1 fractions. Here, we could identify, the presence of varied ring-like structures formed by Csb1, confirming RNA-independent oligomerization (Figure 4.4C). Further, we individually tested the nuclease activity of various fractions to test the effect of oligomerization

on RNA processing. To our surprise, fractions containing higher-order oligomeric states of Csb1 did not show significant RNA processing activity in contrast to fractions expected to contain monomeric protein fractions (Figure 4.4D). These findings suggest that Csb1 undergoes RNA-independent oligomerization, and this may influence the binding and/or processing of the RNA substrate.

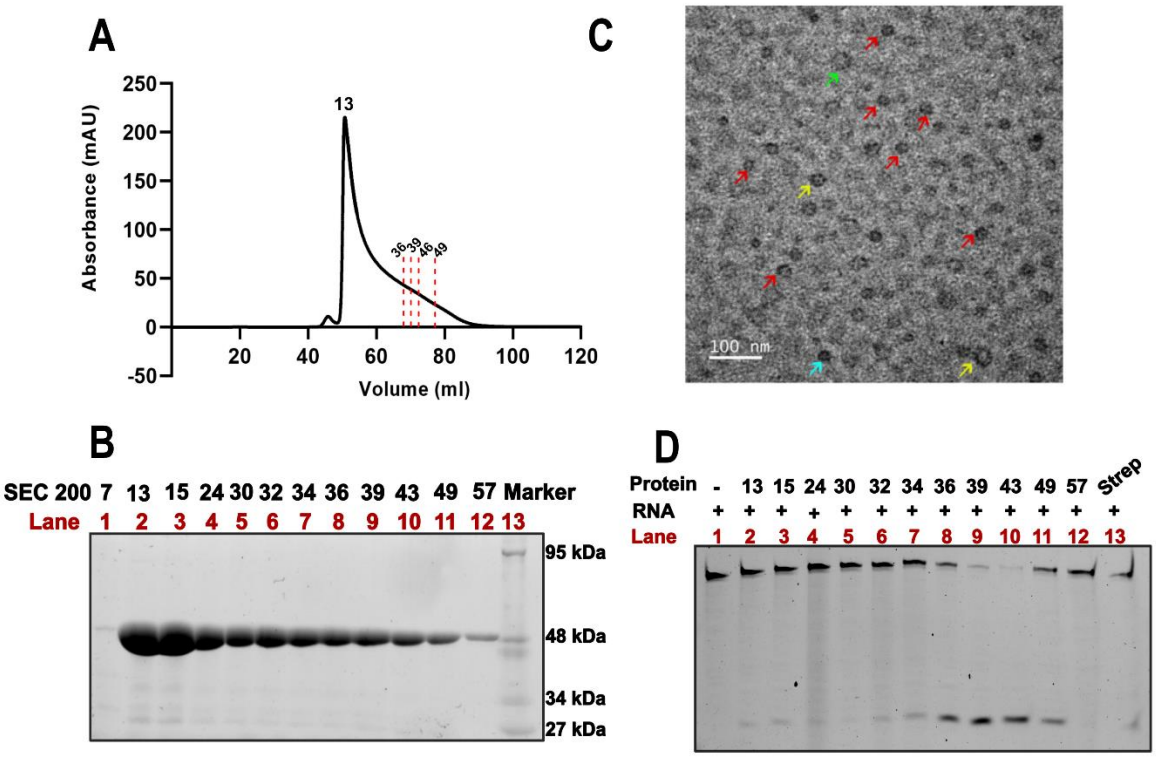


Figure 4. 4: RNA-independent oligomerisation of Csb1

A. A size exclusion chromatography (SEC) analysis indicating the profile corresponding to the dynamic oligomerization of Csb1. The fractions used for further analysis are indicated by red dashes.

B. A 12% SDS PAGE shows the fractions of Csb1 after SEC purification (**A**). A protein marker is loaded on lane 13 to validate the molecular sizes.

C. A negative staining TEM image analysis showing Csb1 in various oligomeric states, which are indicated by arrows in different colours.

D. A 20% denaturing PAGE shows the RNase activity of various Csb1 (1 µM) fractions from (**A**) and (**C**) on 5' 6-FAM labelled CAO repeat RNA substrate (0.25 µM).

4.3.4 Identifying the active site residues of Csb1

Next, we focused on identifying the catalytic residues of Csb1 protein through multiple sequence alignment of Csb1 orthologs to identify conserved charged residues that may facilitate acid-base catalysis. These residues, namely, E78, H122, R123, R179 and R320 (Figure 4.5 A-D) were substituted with alanine and subsequently, cloned, and purified for the assessment of nuclease activity. Interestingly, H122A and R123A showed significant reduction in RNase activity compared to the wild type (Figure 4.5 E-F). This observation suggests that H122 and R123 may serve as catalytic residues in Csb1.

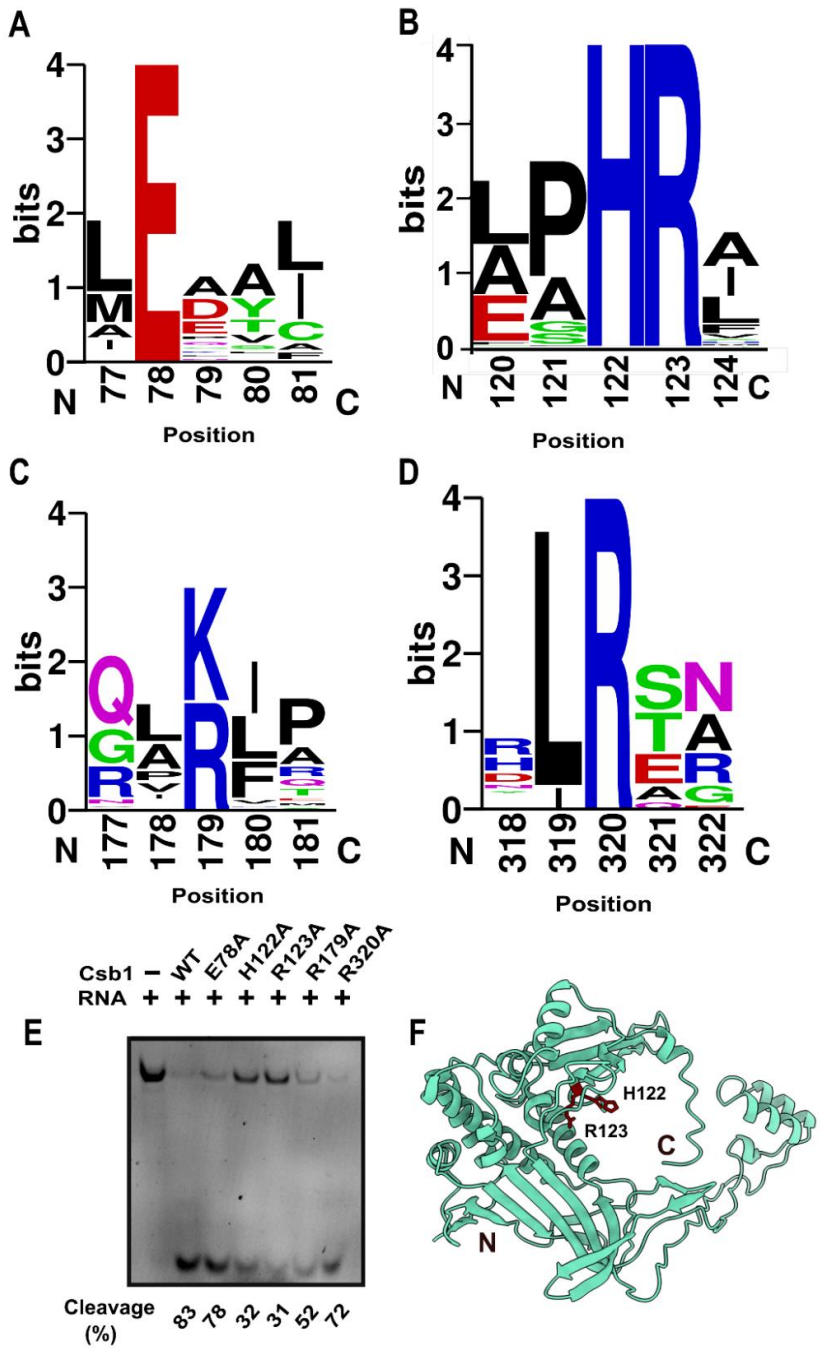


Figure 4. 5: Identifying the active site residues of Csb1

A-D. The sequence logo depicts the conservation of amino acids, namely, E78 (**A**), H122 and R123 (**B**), R179 (**C**), and R320 (**D**) in Csb1 across the type I-G CRISPR-Cas system (57 organisms), respectively. The X-axis indicates the amino acid residue number, whereas, the Y-axis indicates the extent of its conservation (in bits).

E. A 20% denaturing PAGE shows the RNase activity of the Csb1 point mutants E78A, H122A, R123A, R179A, and R320A (1 µM each) (lanes 3-7), respectively, on 0.5 µM of 5' 6-FAM labelled CAO repeat RNA substrate, in comparison to WT (lane 2). The percentage of cleaved RNA fragment is indicated below the gel.

F. An Alphafold2 predicted 3D structural model of Csb1 is shown. The catalytic residues H122 and R123 (in maroon) are shown.

4.4 Discussion

Cas7 in other type I systems (A-F) is generally inert, and forms the backbone of the Cascade complex (Brouns et al., 2008; Jackson et al., 2014; Jore et al., 2011; Lintner et al., 2011; Maier et al., 2019; Mulepati et al., 2014; O'Brien et al., 2020; Pausch et al., 2017; Zhao et al., 2014). In contrast, Cas7 in the type III system cleaves the target RNA and in case of type III-E, found to be involved even in the processing of the pre-crRNA transcript (Goswami et al., 2022; Osawa et al., 2013; Osawa et al., 2015; Ozcan et al., 2021; Samai et al., 2015; R. H. Staals et al., 2014; Tamulaitis et al., 2014; D. W. Taylor et al., 2015). Interestingly, our *in vitro* RNase assay in Chapter 3 revealed that Csb1 which is homologous to Cas7 is not inert in type I-G system (Figure 3.4C). However, a recent study on type I-G crRNA maturation could not observe any RNase activity of Csb1 at a substrate to enzyme ratio of 1:10 (Shangguan et al., 2022). It is noteworthy, using this ratio, we were able to observe RNAse activity starting from 30 mins, as opposed to 10 mins that was used in their study, suggesting that the reaction incubation time can also be a critical factor for processing by Csb1 (Figure 4.1C). This hypothesis is further strengthened by our experiments where, processing was readily observed within 5 minutes of incubation when substrate to enzyme ratios were increased to 1:4 (Figure 4.1B).

Further biochemical characterizations of Csb1 reveal that Csb1 could process repeat RNA substrates from both the orientations (CSO and CAO) analogous to Csb2 (Figure 4.2 A-B). The site of cleavage for this nuclease activity was similar to Csb2 (Figure 4.2 A, B, D). Both Csb1 and Csb2 produced RNA fragments harbouring 5’-OH and 2’,3’ cyclic phosphate ends (Figure 4.2C). A notable distinction between the two arises from their binding preferences to the stem-loop region of the RNA substrate. Csb2 exhibits higher binding affinities to the stem-loop region of CSO RNA as compared to the CAO RNA substrates (Figure 2.8). Similarly, Csb1 displays significant binding exclusively to the CSO RNA stem-loop region (Figure 4.3 C, D) whereas, binding affinities of Csb1 to the CAO RNA could not be determined reliably (Figure 4.3 E, F). We posit that this inability to determine the binding affinities for Csb1 to the CAO RNA may arise due to low binding affinities coupled with its existence of multiple oligomeric states, that could transiently bind and still cleave the substrate.

Another interesting observation of the current chapter was the fact that Csb1 could oligomerize in absence of crRNA. However, this oligomerization did not follow a particular

pattern as we observed a broad peak in the SEC chromatogram (Figure 4.4A). This was confirmed by the TEM analysis which showed several oligomeric states of Csb1 in absence of RNA (Figure 4.4B). Nuclease activity assays of these oligomers displayed differences in the processing of the repeat RNA substrate, suggesting heterogeneity in oligomeric states of Csb1 and their associated RNase activity (Figure 4.4D). We also identified that two highly conserved charged residues are important for the catalytic activity of Csb1 (H122 and R123) as their substitutions resulted in significant decrease in the RNase activity (Figure 4.5E). Interestingly, these residues are a part of the conserved RRM domain of Csb1 but are not conserved across the inert variants of Cas7. However, the catalytic Cas7 homologs of type III system harbour similar charged residues that are responsible for the nuclease activity (Goswami et al., 2022; R. H. J. Staals et al., 2013; Tamulaitis et al., 2014), suggesting the conserved residues of Csb1 may use an acid-.base catalysis based mechanism of RNA hydrolysis as observed in other active Cas7 homologs. Thus, our overall findings not only introduce Csb1 as a founding member of catalytically active Cas7 in type I CRISPR systems, but also suggest that type I-G CRISPR-Cas system could serve as a potential evolutionary intermediate between the type I and type III systems.

This chapter presents the characterization of an additional catalytic subunit, Csb1, within the type I-G effector complex, and emphasizes both the similarities and differences between Csb1 and the previously identified nuclease of the system, Csb2. However, the question of why this system requires two catalytic centres for RNA processing is still not clear. Also, the consequence of this dual nuclease activity of Cascade complex on the succeeding interference mechanism is yet to be understood. Thus, the upcoming chapter focuses on understanding the effect of dual nucleases in type I-G crRNA maturation and studying its effect on the interference mechanism.

4.5 Summary

In this chapter, we have accomplished the biochemical characterization of the RNase activity of Csb1. We evaluated the RNase activity on repeat RNA substrates in both orientations. Our findings demonstrated that Csb1 functions as a metal-independent RNase, and its cleavage pattern is similar to that of Csb2. We further examined the nature of the fragments produced during this cleavage and identified a fragment with 2',3' cyclic

phosphates. Additionally, our observations indicated that the binding of Csb1 on CSO RNA is influenced by the stem-loop region of the repeat RNA, while the nuclease activity still relies on the 5' overhang of the repeat region. We noted that Csb1 exhibits RNA-independent oligomerization, and it is noteworthy that the resulting oligomers do not possess uniform nuclease activities. Finally, we could also identify the catalytically active residues that might be responsible for its RNase activity. With these observations, we proceeded further to understand the effect of these dual nucleases on type I-G crRNA maturation and understand its impact on the interference mechanism in chapter 5.

Chapter 5: Dual nucleases influence the CRISPR interference

5 Chapter 5

5.1 Introduction

Our current investigation showed Csb1 to be an additional nuclease, capable of processing crRNA like the primary RNase Csb2. Presence of dual nucleases in the effector complex is hitherto not noted in type I CRISPR systems and is unique to type I-G. Hence, it was fascinating to explore the dynamic interaction between these two RNases within the effector complex. The fundamental question we aimed to answer was whether the activity of either Csb1 or Csb2 relied on the presence or function of the other and what would be the subsequent implications of this interaction on the interference mechanism within this system.

Typically, the interference mechanism in type I system includes recruitment of an effector nuclease, viz, Cas3 by the effector complex, for target elimination. Cas3 comprises of an N-terminal HD nuclease domain followed by a DExD/H box Superfamily 2 helicase domain and a C-terminal domain (Gong et al., 2014; Y. Huo et al., 2014; Jackson et al., 2014; Sinkunas et al., 2011). However, in contrast to Cas3 in other type I systems, the domain architecture of Cas3/I-G has undergone rearrangement, wherein the helicase domain occupies the N-terminal region and the HD nuclease domain occupies the C-terminal region (Makarova et al., 2015; Makarova et al., 2020). However, we do not understand the implications of this domain re-arrangement in Cas3 on its functionality and its interaction with the Cascade complex.

Thus, in this chapter we focus on understanding the impact of the dual nucleases on the type I-G crRNA processing as well as the interference mechanism using *in vitro* and *in vivo* experiments. Towards this, we create various mutant varieties of the Cascade complex to understand the dependency of Csb1 and Csb2 on each other, for pre-crRNA processing. Next, we also developed an *in vivo* interference system using Cas3. Using this system as a readout for efficient maturation., we investigate the impact of the dual nucleases on crRNA maturation.

5.2 Materials and Methods

5.2.1 Creation of Strains

To assess CRISPR interference *in vivo*, we integrated the CRISPR array (in CSO and CAO orientations, separately) from *B. animalis* into the *attB* sites of the *E. coli* IYK-12 genome. The integrated fragments, each approximately 1.7 kb in length, encompassed 175 nucleotides upstream and 150 nucleotides downstream of the CRISPR array. This integration was achieved using the pOSIP-CT plasmid (Addgene #45981) through a method known as clonetegration (St-Pierre et al., 2013). Consequently, two distinct strains, IG-CR_CSO and IG-CR_CAO, were generated to facilitate the investigation of CRISPR interference mechanisms.

5.2.2 Molecular cloning

The Cascade complex harbouring various point mutants like Csb1-H122A (C1H), Csb1-R123A (C1R), Csb1 H122A_Csb2 H520A (C1H_C2H), Csb1 R123A_Csb2 H520A (C1R_C2H), Csb1 H122A_R123A (C1HR) and Csb2 H520A_D530A (C2HD) were cloned into pQE2. The operon design as well the cloning approach was similar to the WT Cascade mentioned in Chapter 3 (Section 3. 2.1). To generate the target plasmid (pT/I-G), the DNA sequence of the target along with 'TTT' PAM sequence was inserted into the 13S-R vector (Addgene #48328) through Gibson assembly, whereas the empty vector backbone was used as the non-target plasmid (pNT). The gene encoding Cas3 was cloned into pET Strep II TEV LIC cloning vector, where a modification was introduced whereby the Strep II fusion tag was positioned on the C-terminal end of the protein (1R, Addgene #29664). All the clones were confirmed by Sanger sequencing.

5.2.3 *In vivo* Cascade complex reconstitution

The mutant variants of the Cascade complexes were reconstituted *in vivo* by co-expressing them with the pCR_Array plasmid and expressing them similarly as mentioned for the WT Cascade complex in Chapter 3 (Section 3.2.2).

5.2.4 *In vitro* RNase assay

To test the RNA processing activity of the mutant variants of the Cascade complex, 0.5 µM of the CAO repeat RNA (5' 6-FAM labelled) substrate was incubated with 1 µM of Cascade complex (WT or mutants) in a reaction buffer containing 20 mM Tris-Cl (pH 7.6), 100 mM KCl and 6 mM β-Mercaptoethanol (β-ME) and incubated at 37 °C for 30 min. Cleavage products were analysed on a 20% (w/v) denaturing polyacrylamide gel and viewed in Bio-Rad gel documentation system.

5.2.5 *In vivo* CRISPR interference assay

To create a functional *in vivo* CRISPR interference assay system for type I-G, *E. coli* IG-CR cells (*E. coli* IG-CR_CSO and IG-CR_CAO strains described in Section 5.2.1) were used as a surrogate. *E. coli* IG-CR cells were transformed with various combinations of constructs (pCascade, pCas3, pT, and pNT) and plated onto an LB agar plate supplemented with the respective antibiotics and 0.2 % glucose to arrest any leaky expression. A single colony was inoculated in a fresh 5 mL LB broth containing the appropriate antibiotics (25 µg/mL chloramphenicol, 25 µg/mL kanamycin, 50 µg/mL ampicillin, 50 µg/mL spectinomycin) and 0.2% glucose. The culture was allowed to grow at 37 °C until the OD_{600} reached 0.6. 500 µL of the cells were centrifuged and the resulting pellet was washed twice with fresh LB broth to eliminate any residual glucose. The pellet was then re-suspended in 500 µL of fresh LB broth containing the respective antibiotics mentioned above along with 0.02 mM IPTG and 0.2% L-arabinose. The culture was allowed to grow at 180 rpm at 37 °C for approximately 2 hours. Subsequently, the pellet was harvested, serially diluted up to 10^{-5} times, and plated onto LB agar plates supplemented with antibiotics. The plates were then incubated at 37 °C for 10-12 hours. The number of colonies in each plate was counted, and the transformation efficiency was calculated using the following equation.

$$\text{Transformation efficiency} = \frac{\text{No. of colonies x Dilution factor}}{\text{µg of DNA used}}$$

Transformation efficiency for various mutants of the Cascade complex (pC1H_Cascade, pC1HR_Cascade, pC2H_Cascade, pC2HD_Cascade and pC1H_C2H) was calculated using a similar protocol.

5.3 Results

5.3.1 *In vitro* RNase activity of the dual nuclease effector complex

Having characterised the dual nucleases namely, Csb1 and Csb2 individually, we then proceeded towards understanding their interplay as a part of the Cascade complex. Towards this, we created variants of the Cascade constructs harbouring catalytically inactive mutants of either Csb1 or Csb2 or both. The constructs namely, Csb1 H122A Cascade, Csb1 R123A Cascade, Csb1 H122A_Csb2 H520A Cascade and Csb1 R123A_Csb2 H520A Cascade were created via overlap-extension PCR and ligated into pQE2 vector via Gibson assembly (Figure 5.1 A, B). Next, these constructs were individually co-expressed with pCR_Array plasmid for *in vivo* reconstitution, similar to that for the WT Cascade complex. Upon assessing the nuclease activity, we observed that only the double mutants viz, Csb1 H122A_Csb2 H520A and Csb1 R123A_Csb2 H520A were inactive, while the single mutants (Csb1 H122A and R123A) retained nuclease activity, albeit at a reduced level (Figure 5.1C). This indicates that the absence of nuclease activity in Csb1 is compensated for by Csb2 within the Cascade complex. This provides additional evidence supporting the idea that Csb1 and Csb2 can both function as dual nucleases within the Cascade complex.

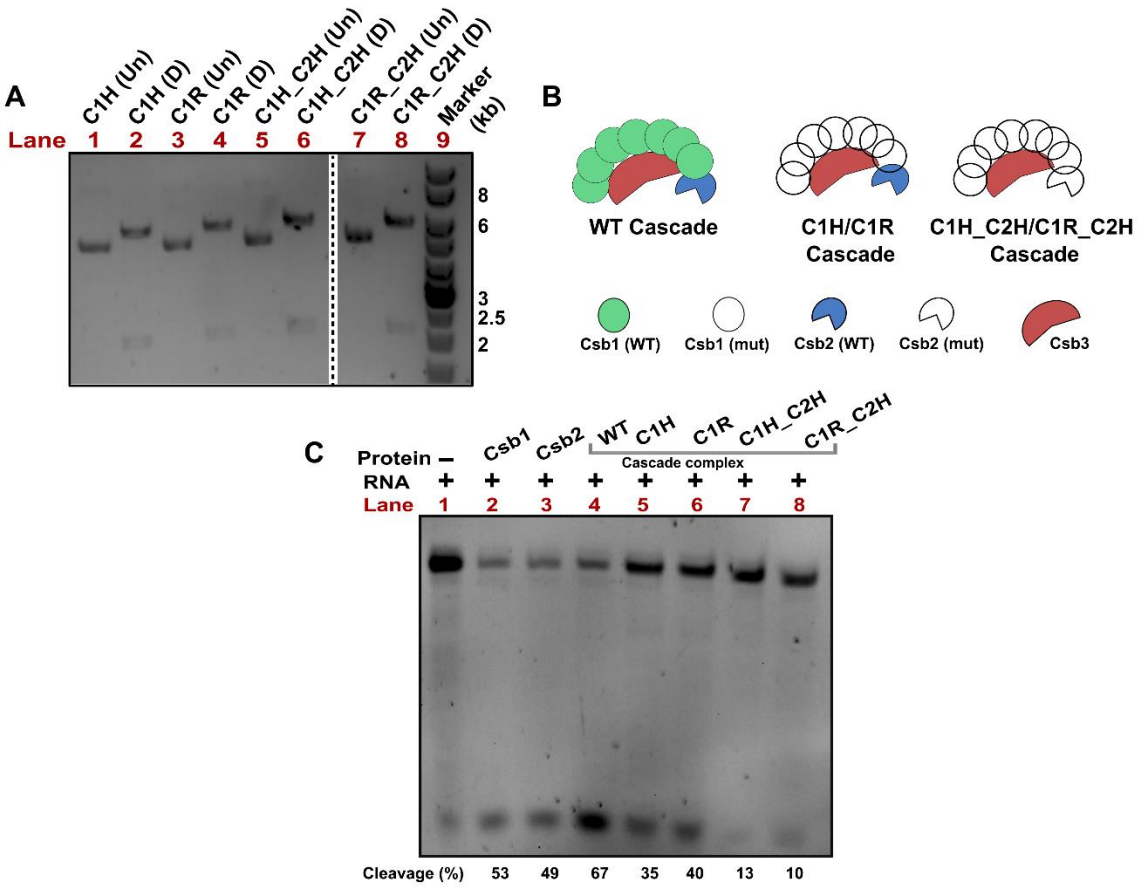


Figure 5. 1: Effect of dual nucleases on type I-G maturation

A. A 0.8% agarose gel showing the clone confirmation using restriction digestion of various Cascade mutant constructs viz., Csb1 H122A (C1H, lane 1, 2), Csb1 R122A (C1R, lane 3, 4), Csb1 H122A_Csb2 H520A (C1H_C2H, lane 5, 6) and Csb1 R123A_Csb2 H520A (C1R_C2H, lane 7, 8). UN and D indicate undigested and digested plasmids respectively. The size of the desired fragments are 6.2 kb and 2.3 kb, respectively. A DNA marker is loaded in lane 9, and the respective bands are marked for reference. The dashed line denotes discontinuity in the gel for the purpose of clarity.

B. A schematic representation of the various Cascade mutant constructs generated in (A) and tested for its nuclease activity in (C).

C. A 20% denaturing PAGE showing the RNase activity of WT (lane 4) and mutant Cascade complexes (1 µM each) on 5' 6-FAM labelled CAO repeat RNA substrate (0.5 µM). The Cascade complex mutants include Csb1 H122A (C1H, lane 5), Csb1 R123A (C1R, lane 6), Csb1 H122A_Csb2 H520A (C1H_C2H, lane 7), and Csb1 R123A_Csb2 H520A (C1R_C2H, lane 8). The RNA substrate was also incubated with Csb1 (lane 2) and Csb2 (lane 3) for a comparative analysis. The percentage of cleaved RNA fragment is indicated below the gel

5.3.2 *In vivo* CRISPR interference activity

Having established the functional redundancy of the dual nucleases on type I-G crRNA maturation, we then proceeded to comprehend the implications of this on CRISPR interference.

Towards this, we decided to ectopically reconstitute the type I-G interference machinery in *E. coli* due to inherent challenges in growing *B. animalis* in the lab. To reconstitute the functionally competent type I-G CRISPR machinery in *E. coli*, we expressed plasmid-borne genes coding for Cascade (pCascade) and Cas3 (pCas3) together with either target (pT) or non-target (pNT) (Figure 5.2). The target construct (pT) consists of 'TTT' PAM sequence (Almendros et al., 2019) followed by the spacer sequence complementary to the CRISPR array integrated into *E. coli* IG-CR_CSO and IG-CR_CAO strains. These strains express the pre-CRISPR transcript in one of the orientations from a chromosomally integrated cassette (referred in methods). A fully operational CRISPR machinery, comprising of both the Cascade complex and Cas3, significantly reduced the transformation efficiency of pT (target) compared to pNT (non-target) in both strains (Figure 5.3). This observation suggests that the type I-G CRISPR-Cas system is functionally reconstituted in a surrogate host and is also effective against the target substrate.

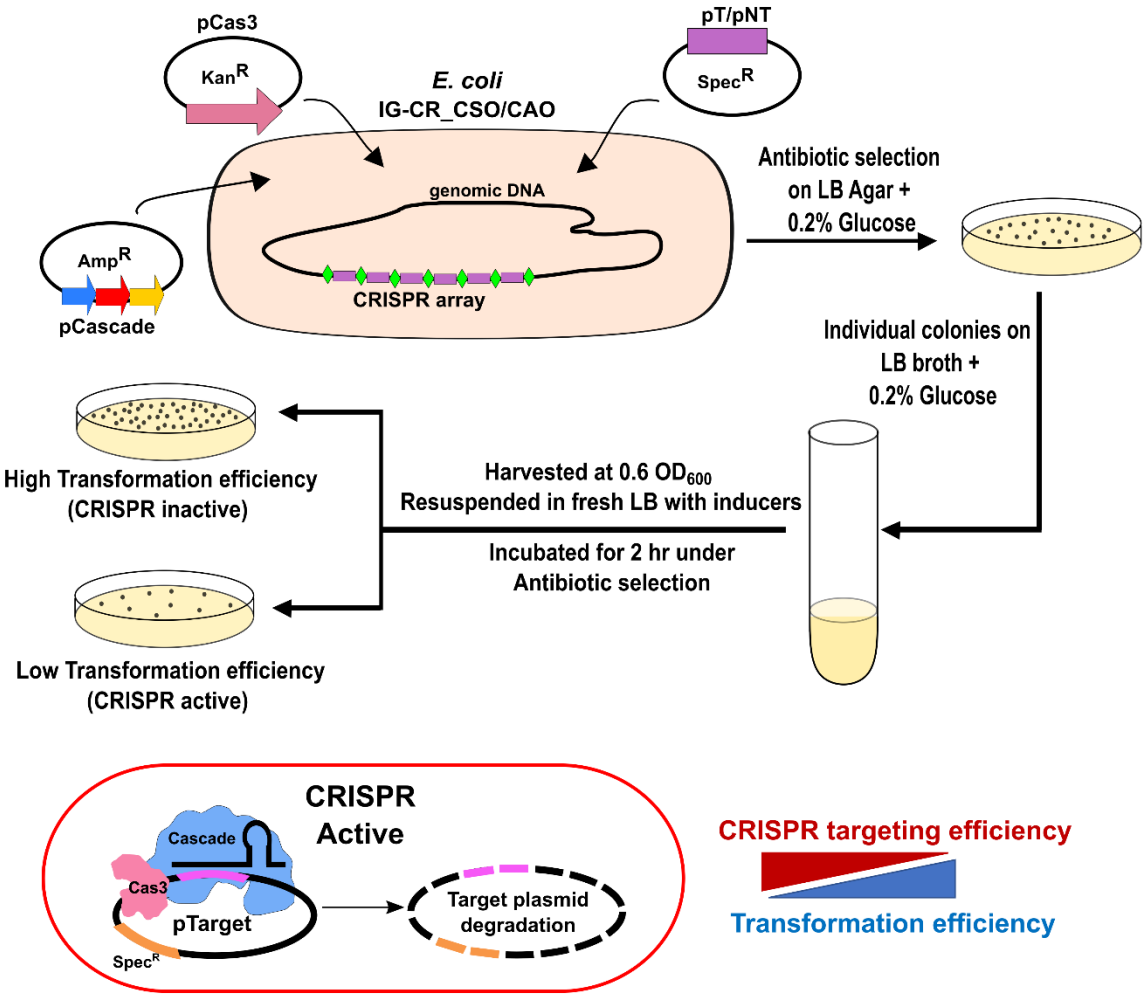


Figure 5. 2: In vivo interference assay

A schematic representation of the in vivo interference assay is shown. An *E. coli* IG-CR strain harbouring the CRISPR array from *B. animalis* in either CAO or CSO orientation was used as a surrogate host. Cascade/I-G complex (including its varied mutants) and Cas3 were co-expressed through IPTG inducible vectors harbouring ampicillin and kanamycin antibiotic resistance markers, respectively. The target sequence was inserted in 13S-R vector (T) that harbours spectinomycin resistant marker, whereas an empty 13S-R vector was used for the non-target plasmid (NT). Low transformation efficiency indicates a functional CRISPR interference resulting from target plasmid degradation that may be attributed to the successful maturation of crRNA.

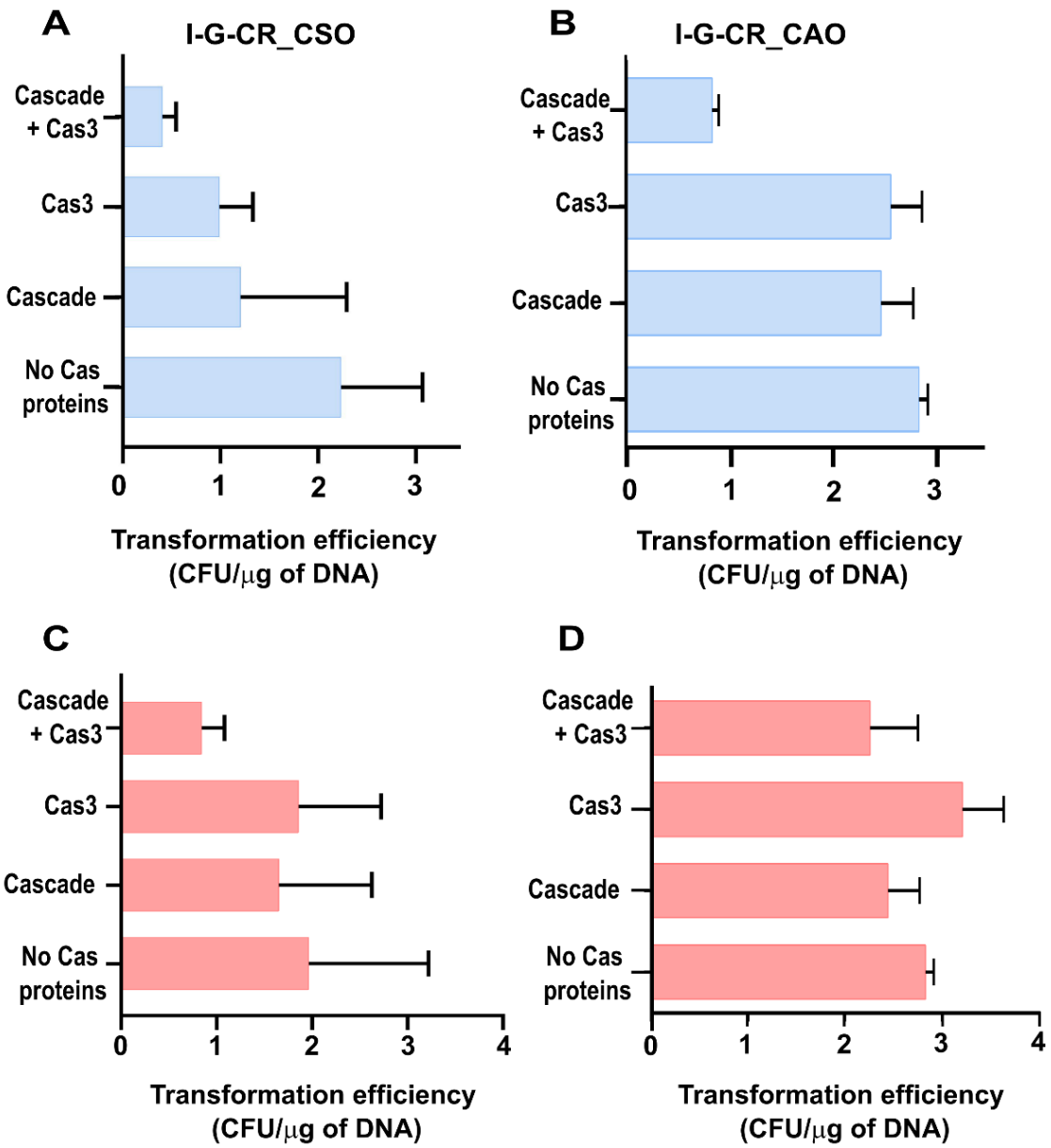


Figure 5. 3: Functionality of type I-G in vivo interference assay

A-B. Transformation efficiency is shown for *E. coli* IG-CR_CSO (A) and *E. coli* IG-CR_CAO (B) in response to the presence of the target plasmid and other components of the CRISPR interference machinery such as Cascade complex and Cas3. The error bars represent the standard deviation from three independent biological replicates.

C-D. Transformation efficiency is shown for *E. coli* IG-CR_CSO (C) and *E. coli* IG-CR_CAO (D) in response to the presence of the non-target plasmid and other components of the CRISPR interference machinery such as Cascade complex and Cas3. The error bars represent the standard deviation from three independent biological replicates.

5.3.3 Functional implications of dual nucleases on the *in vivo* CRISPR interference activity

Encouraged by the above findings, we next aimed at understanding the implication of Csb1 and Csb2 functionality on CRISPR interference. To study this, we created single and double mutants of Csb1 and Csb2 and tested for its impact on interference via *in vivo* interference assay. Here, the high transformation efficiency indicates that the CRISPR interference is less active or vice versa. We observed that the single mutants, such as Csb1 H122A and Csb2 H520A, exhibited increased transformation efficiency compared to the WT, while the double mutant (Csb1 H122A_Csb2 H520A) displayed even higher transformation efficiency than the single mutants (Figure 5.4). This suggests that mutations in both Csb1 and Csb2 abolish CRISPR interference, likely due to defective crRNA maturation in the respective constructs. To further dissect the roles of Csb1 and Csb2 in pre-crRNA processing, we generated dual mutants for a single Cas subunit of the Cascade complex, namely, Csb1 H122A/R123A and Csb2 H520A/D530A. Both mutants showed better transformation efficiencies than the WT. Remarkably, the dual mutant variant of Csb2 (Csb2 H520A/D530A) exhibited significantly elevated transformation efficiency, indicating the pivotal role of Csb2 and its extended involvement in the interference mechanism (Figure 5.4). Collectively, these results further underscore that the presence of dual nucleases (Csb1 and Csb2) enhances the efficiency of CRISPR-mediated target elimination, potentially by expediting the processing of the pre-CRISPR transcript.

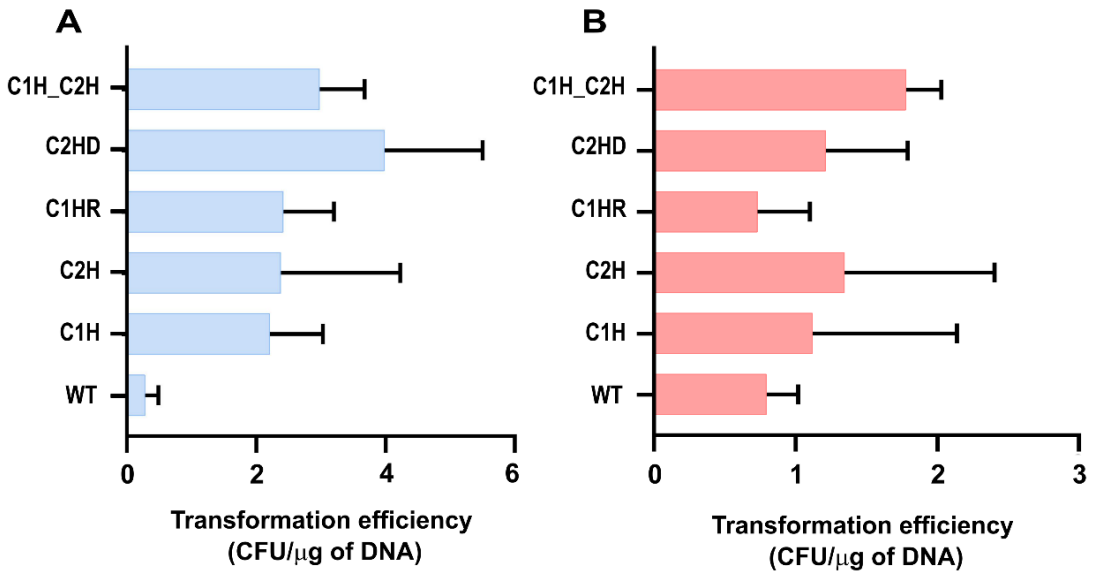


Figure 5. 4: Functional implications of dual nucleases on interference efficiency

A. Transformation efficiency of *E. coli* IG-CR_CAO was tested in response to target plasmid in the presence of varied mutant forms of Cascade complex such as C1H (Csb1 H122A), C2H (Csb2 H520A), C1HR (Csb1 H122A_R123A), C2HD (Csb2 H520A_D530A) and C1H_C2H (Csb1 H122A and Csb2 H520A). The error bars represent standard deviation from three independent biological replicates.

B. Transformation efficiency of *E. coli* IG-CR_CAO was tested in response to non-target plasmid in the presence of varied mutant forms of Cascade complex such as C1H (Csb1 H122A), C2H (Csb2 H520A), C1HR (Csb1 H122A_R123A), C2HD (Csb2 H520A_D530A) and C1H_C2H (Csb1 H122A and Csb2 H520A). The error bars represent standard deviation from three independent biological replicates.

5.4 Discussion

Csb1 stands out as the sole studied catalytically active Cas7 homolog within type I systems. Moreover, it is a component of the only investigated effector complex known to host two nucleases. Thus, we aimed at understanding the interplay between these nucleases for processing the crRNA. Our *in vitro* nuclease assay of the mutant variants of the Cascade complex showed that the Cascade complex harbouring mutations in both Csb1 and Csb2 had significantly lost the processing ability while complex harbouring mutation in only Csb1 had reduced nuclease activity (Figure 5.1C). Similar observations were also made in Chapter 3 where we observed processing by the Csb2 mutant complex variant. This suggests that the nuclease activity of Csb1 is compensated by Csb2 within the Cascade complex and vice versa, supporting the notion that Csb1 and Csb2 function as dual nucleases within the complex.

Thereafter, we aimed at comprehending the influence of the dual nucleases on the interference mechanism. For this, we successfully designed an *in vivo* interference assay, where we noted substantial interference efficiency against the target substrate using CRISPR array in both orientations. Notably, this efficacy for both orientations has not yet been reported in any CRISPR systems and indicates that type I-G CRISPR systems can actively engage and neutralize incoming MGEs using crRNA guides originating from both orientations of the CRISPR array. (Figure 5.3). Interestingly, our experiments with mutant Cascade complex exhibited markedly lower targeting efficiencies compared to the wild-type variant (Figure 5.4). Further, it is also interesting to note that despite some differences in biochemical properties of individual Cas nucleases (Figure 2.2B and Figure 4.4) and their RNA processing activities (Figure 2.7, 4.3 and Figure 4.1), their cumulative nuclease activity was instrumental in efficient

targeting of DNA targets (Figure 5.1C). These findings not only underscore the significance of dual nucleases in CRISPR RNA maturation but also emphasizes that a non-functional maturation step can consequently result in reduced interference efficiency.

Collectively, it is interesting to note that our study comprehensively established the presence of dual nucleases in type I effector complex. Further, we also show the importance of these dual nucleases in assembled Cascade complex and their interplay with respect to the interference machinery in enhancing the targeting efficiency of type 1-G system.

5.5 Summary

This chapter embarks on understanding the role of dual Cas nucleases on the type I-G crRNA maturation and its impact on the overall targeting efficiency of a dual nuclease effector complex. Firstly, we investigated the interplay between the nucleases Csb1 and Csb2, within the Cascade complex through the creation of catalytically inactive Cascade variants harbouring mutations in either Csb1 or Csb2 or both, and testing for their processing activity. Here, we found that only double mutants, Csb1 H122A_Csb2 H520A and Csb1 R123A_Csb2 H520A, were inactive, while single mutants (Csb1 H122A and R123A) retained nuclease activity, though at a reduced level. This functional redundancy suggested that Csb2 compensates for the loss of Csb1 nuclease activity within the Cascade complex. Subsequently, we reconstituted the type I-G CRISPR machinery in *E. coli* as a surrogate host and observed effective CRISPR interference against target substrates. Further, we explored the implications of Csb1 and Csb2 functionality on CRISPR interference by creating single and double mutants and conducting *in vivo* interference assays. The results indicated that mutations in both Csb1 and Csb2 abolished CRISPR interference, and the presence of dual nucleases enhanced the efficiency of CRISPR-mediated target elimination, potentially by accelerating pre-CRISPR transcript processing. Overall, our findings shed light on the intricate dynamics and functional roles of Csb1 and Csb2 in the type I-G CRISPR-Cas system, offering valuable insights into maturation and interference mechanisms.

Chapter 6: Conclusion and future directions

6 Chapter 6

6.1 Conclusion

The abundant variety of living organisms sustaining in competitive and most challenging environmental niches stand as evidence for the remarkable influence of natural selection in guiding the course of life's evolution over countless millennia. In this ongoing contest for survival of the fittest, every single-celled prokaryotes to multi-cellular eukaryotes compete, adapt, and evolve to thrive. One such life form that are present with extreme abundance in every environment, the phages, are devoid of cellular machinery, thereafter parasitize on other hosts for its fundamental survival tasks such as replication or protein synthesis (Edwards & Rohwer, 2005). In response to such threats, bacteria have evolved and refined diverse defense systems, with the CRISPR-Cas system standing out as the sole recognized adaptive immune system in prokaryotes (Bolotin et al., 2005; Jansen et al., 2002; F. J. Mojica et al., 2005; Pourcel et al., 2005). CRISPR-Cas based defence system operates through 3 steps – adaptation (acquisition of foreign MGE into its CRISPR array), maturation (processing of the CRISPR array to generate specific crRNA that are complementary to the target MGE) and interference (recognition and degradation of the target MGE). The core aspect of CRISPR-Cas defence lies in the utilization of crRNAs for precise identification and targeting of invading genetic elements. The maturation process plays a crucial role in ensuring the accurate generation of these crRNA molecules, thereby serving as a pivotal step in the CRISPR-based targeting system. In the extensively studied type I CRISPR systems, Cas6 (Cas5, in case of type I-C) generates guide crRNA by cleaving the repeat region of pre-CRISPR transcripts (Garside et al., 2012; Haurwitz et al., 2010; Nam et al., 2012; R. Wang et al., 2011). Notably, type I-G system harbours Csb2, that was predicted to be a fusion of Cas5-like and Cas6-like domains (Makarova et al., 2015; Makarova et al., 2020). The present work has attempted to unravel the molecular mechanism of type I-G CRISPR maturation by probing the primary maturase Csb2 and other components that assist crRNA maturation.

Using a type 1-G CRISPR array we also showed that the catalytic activity of Csb2 resides in the C-terminal Cas6-like domain, while the N-terminal Cas5-like domain is catalytically inert. This observation falls in line with most type I systems, except I-C, where Cas5 is generally inactive and plays a structural role in RNA binding and stabilizing other Cas

proteins in the Cascade complex (Brouns et al., 2008; Garside et al., 2012; Jore et al., 2011; Nam et al., 2012; Wiedenheft et al., 2011). Interestingly, the Cas6-like domain of Csb2 is inherently prone to self-degradation (Figure 2.2C). Further, structural predictions using AlphaFold (Jumper et al., 2021) confirms the existence of two independently folded regions of Csb2 which suggest a head-to-head fusion of the two Cas proteins, where the characteristic motifs like the ferredoxin folds of Cas5 and Cas6 and the Cas6 specific glycine-rich loop are conserved. Interestingly, other type I-G proteins involved in adaptation and interference have been shown to undergo significant fusion or rearrangement during evolution (Almendros et al., 2019; Makarova et al., 2015; Shangguan & White, 2023). Thorough biochemical and structural studies on these Cas proteins can further help unveil the overall functioning of the type I-G system and factors driving its evolution.

The next interesting observation comes from the type I-G CRISPR array that harbour promoter-like elements upstream of both strands giving rise to pre-crRNA transcripts with distinct structures and sequences (Figure 2.3). Notably, bidirectional expression of the CRISPR array that gives rise to complementary transcripts was observed previously in type III systems. These complementary transcripts were hypothesised to regulate the crRNA-mediated self-targeting in the absence of an invading genetic element (Lillestol et al., 2006; Lillestol et al., 2009). Alternately, the bidirectional CRISPR array may also help achieve higher crRNA expression levels for neutralizing invading genetic elements. Fascinatingly, Csb2 can process RNA transcripts from both strands, a unique feature among reported CRISPR-Cas systems (Figure 6.1). Structure mapping studies revealed that the cleavage site is conserved in both CSO and CAO repeat RNA, and the generated RNA fragment end that possesses 2',3' cyclic phosphate shows characteristics of metal-independent nuclease activity of Cas5 and Cas6. Furthermore, Csb2 displays specificity towards the stem-loop region for binding and cleaves at the overhang region, indicating an unconventional mode of repeat RNA specificity which is unlike other type I systems, where Cas6 recognises and binds the stem-loop region of the repeat RNA and cleaves at the base of it (Gesner et al., 2011; Haurwitz et al., 2010; Sashital et al., 2011). Thus, comprehensive structural analysis of Csb2 bound to the crRNA (holoenzyme) may help us understand this distinctive recognition and cleavage mechanism adopted by Csb2.

Next, we advanced towards understanding the type 1-G effector complex by *in vivo* reconstitution using Csb1 (Cas7 homolog), Csb2 (Cas5-Cas6 fusion), Csb3 (Cas8 homolog) and crRNA. The compositional analysis of the purified Cascade complex showed the expected stoichiometry suggesting it to be a minimalistic type 1 effector complex consisting of only

three proteins. Interestingly, like Csb2, the type I-G Cascade efficiently processed crRNA and contains an additional nuclease, Csb1, which exhibited similar repeat RNA cleavage activity unlike its counterparts in other type I systems. Csb1 exhibited RNase activity *in vitro*, processing repeat RNA substrates from both orientations (CSO and CAO) akin to Csb2 (Figure 6.1). Interestingly, Csb1 demonstrated RNA-independent oligomerisation, displaying a dynamic pattern observed through SEC chromatography and TEM analysis. The varied oligomeric states of Csb1 correlated with differences in RNase activity, suggesting the dynamic nature of Csb1-mediated processing. The RNA cleavage pattern for both CSO and CAO RNA were consistent for both Csb1 and Csb2. Interestingly, the binding affinities for the CAO RNA were significantly lower for Csb2 and could not be determined reliably for Csb1, likely due to the presence of varied oligomeric states. It is noteworthy that Cas7 in other type I systems is inert and plays major structural role in the Cascade assembly by forming the backbone of the complex via interacting with the crRNA (Brouns et al., 2008; Jackson et al., 2014; Jore et al., 2011; Lintner et al., 2011; Maier et al., 2019; Mulepati et al., 2014; O'Brien et al., 2020; Pausch et al., 2017). In contrast, type III systems feature catalytically active Cas7 homologs engaged in RNA targeting, except for type III-E systems with a gRAMP Cas7-11 protein, where Cas7 subunits are additionally involved in pre-crRNA processing (Goswami et al., 2022; Osawa et al., 2015; Ozcan et al., 2021; Samai et al., 2015; R. H. Staals et al., 2014; Tamulaitis et al., 2014; D. W. Taylor et al., 2015). Given these observations, we propose that Csb1, acting as a catalytically active Cas7, challenges the conventional inert role of Cas7 in type I systems, suggesting a potential evolutionary link between type I and type III CRISPR systems.

To further understand the functional basis of dual nucleases in the Cascade complex, we aimed to study crRNA processing by individually perturbing each nuclease in the Cascade complex. Here, the absence of one of the nuclease activities was compensated by the other within the Cascade complex, suggesting that the dual nucleases may act complementarily or cumulatively enhance crRNA processing for better MGE targeting. This observation leads us to suggest that the potential rise in the pre-crRNA population due to bidirectional transcription of the array can now be efficiently processed by the dual nucleases. However, the mechanism underlying their preference for either strands remains to be elucidated. Moving forward, we delved into understanding the impact of these dual nucleases on the interference mechanism. Successfully transplanting the type I-G machinery into a heterologous system allowed us to design an *in vivo* interference assay, as a readout for efficient maturation of pre-crRNAs arising from both orientations of the CRISPR array. Here, we also observed that cells with mutant

variants of the Cascade complex, exhibited markedly lower targeting efficiencies, emphasizing the crucial role of dual nucleases in CRISPR maturation (Figure 6.1). Furthermore, it underscores that a non-functional maturation step leads to a subsequent reduction in interference efficiency. Overall, the thesis makes important advances towards understanding the distinctly evolved type I-G CRISPR-Cas system. This work throws light on the disparate maturases that participate in the processing of crRNA and their interplay with the interference machinery.

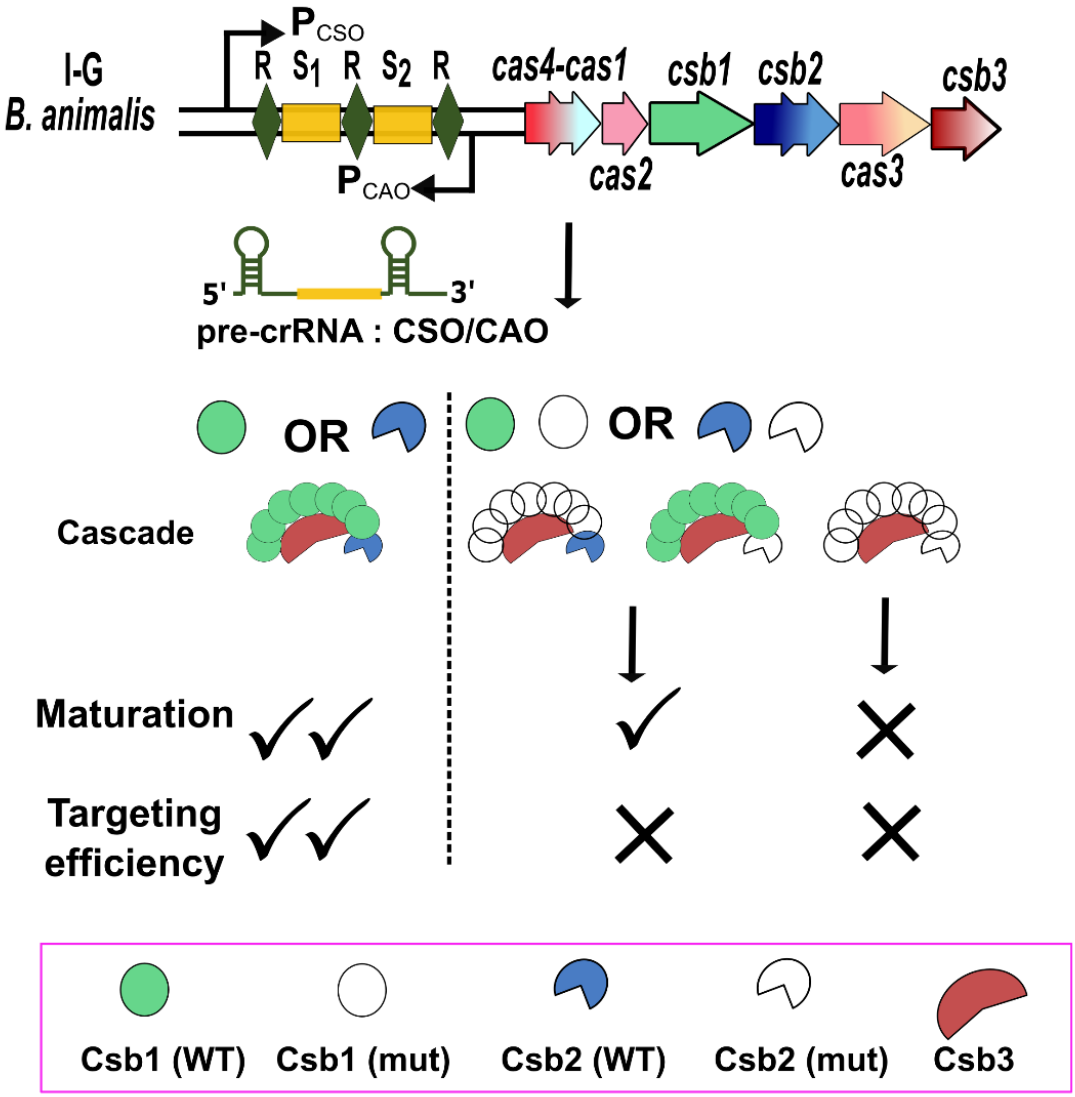


Figure 6. 1: Disparate Cas nucleases participate in the maturation of crRNA in type I-G CRISPR-Cas system.

A schematic representation of functional interplay between various nucleases and the effector complex driving crRNA maturation in type I-G system. The bidirectionally encoded crRNA substrates can be processed by either Csb1 (green), Csb2 (blue) or Cascade complex, whose cumulative action enables the highest targeting efficiency in the type 1-G system.

6.2 Future directions

In this study, we have attempted to unveil the molecular events associated with pre-crRNA processing in the type I-G CRISPR-Cas system. Nevertheless, numerous questions remain unanswered. Currently, the type I-G CRISPR-Cas system lacks high-resolution structural information, posing a challenge to scrutinize its functional mechanism at the atomic level. Structural determination of both apo and holo forms of both Csb1 and Csb2 would assist in addressing the underlying interactions between the substrate and the enzyme. Recent structural investigations of type I-G Cascade complex reveal the partial structures of the Cascade backbone formed by Csb1-crRNA complex and the PAM recognising protein, Csb3 (Shangguan et al., 2022). However, additional structural details highlighting the interactions between these Cas proteins remain unexplored. Thus, obtaining the structural details of the type I-G Cascade complex would provide comprehensive insights into the interaction among Cas subunits and the collaborative role of Csb1 and Csb2 in pre-crRNA processing.

The present study also aims to understand the influence of the maturation mechanism on CRISPR interference. A distinctive feature of the type I-G effector nuclease, Cas3, is its unique domain rearrangement as compared to other type I systems. Typically, the N-terminal HD domain, responsible for target degradation has been relocated to the C-terminal region. This domain rearrangement also places the DEAD-box helicase domain on distal N-terminus end (Makarova et al., 2015). Future biochemical and structural investigations of Cas3/I-G could provide insights into the interference mechanism and contribute to understanding the impact of domain rearrangement on interference. Further elucidation of the structural aspects of the interference module (Cascade: crRNA: target: Cas3) promises to uncover a wealth of information pertaining to the functionality of type I-G CRISPR defence.

Since the advent of CRISPR-based genome engineering, Cas9, Cas12 and Cas13 from Class 2 CRISPR-Cas systems, have been employed as the primary tools because of the ease of deploying a single effector protein for gene editing in higher organisms. Nonetheless, recent studies have demonstrated the efficacy of utilizing type I systems for genome editing in prokaryotes (X. Han et al., 2024; Hidalgo-Cantabrana et al., 2019; C. Hu et al., 2024; Z. Xu et al., 2021). This could be possible due to longer spacer length ranging from a minimum of approximately 25 nucleotides to as long as around 36 nucleotides, a notable exemption compared to Cas9 (~20 nucleotides). The longer spacer length may exhibit increased specificity

towards the target site and thus holds promise to reduce off-target effects, thereby enhancing the overall efficiency of the system. In this context, the type I-G CRISPR-Cas system emerges as a promising contender. Its minimal three-component Cascade complex renders it an especially worthy candidate for further exploration and potential utilization in genetic manipulation studies (Shangguan & White, 2023).

References

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B., . . . Zhang, F. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science, 353*(6299), aaf5573. doi:10.1126/science.aaf5573

Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S., & Shinkai, A. (2010). Transcription profile of Thermus thermophilus CRISPR systems after phage infection. *J Mol Biol, 395*(2), 270-281. doi:10.1016/j.jmb.2009.10.057

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*(10), 5223-5230. doi:10.1093/nar/gkz217

Anantharaman, V., Makarova, K. S., Burroughs, A. M., Koonin, E. V., & Aravind, L. (2013). Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biol Direct, 8*, 15. doi:10.1186/1745-6150-8-15

Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature, 513*(7519), 569-573. doi:10.1038/nature13579

Anton, T., Karg, E., & Bultmann, S. (2018). Applications of the CRISPR/Cas system beyond gene editing. *Biol Methods Protoc, 3*(1), bpy002. doi:10.1093/biomethods/bpy002

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., . . . Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature, 576*(7785), 149-157. doi:10.1038/s41586-019-1711-4

Aravind, L., Watanabe, H., Lipman, D. J., & Koonin, E. V. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc Natl Acad Sci U S A, 97*(21), 11319-11324. doi:10.1073/pnas.200346997

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., . . . Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science, 315*(5819), 1709-1712. doi:10.1126/science.1138140

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

Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., & Yakunin, A. F. (2011). Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J, 30*(22), 4616-4627. doi:10.1038/emboj.2011.377

Bertozzi Silva, J., Storms, Z., & Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett, 363*(4). doi:10.1093/femsle/fnw002

Bikard, D., Hatoum-Aslan, A., Mucida, D., & Marraffini, L. A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe, 12*(2), 177-186. doi:10.1016/j.chom.2012.06.003

Biswas, A., Fineran, P. C., & Brown, C. M. (2014). Accurate computational prediction of the transcribed strand of CRISPR non-coding RNAs. *Bioinformatics, 30*(13), 1805-1813. doi:10.1093/bioinformatics/btu114

Blosser, T. R., Loeff, L., Westra, E. R., Vlot, M., Kunne, T., Sobota, M., . . . Joo, C. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol Cell, 58*(1), 60-70. doi:10.1016/j.molcel.2015.01.028

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading), 151*(Pt 8), 2551-2561. doi:10.1099/mic.0.28048-0

Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature, 493*(7432), 429-432. doi:10.1038/nature11723

Borges, A. L., Castro, B., Govindarajan, S., Solvik, T., Escalante, V., & Bondy-Denomy, J. (2020). Bacterial alginate regulators and phage homologs repress CRISPR-Cas immunity. *Nat Microbiol, 5*(5), 679-687. doi:10.1038/s41564-020-0691-3

Bourniquel, A. A., & Bickle, T. A. (2002). Complex restriction enzymes: NTP-driven molecular motors. *Biochimie, 84*(11), 1047-1059. doi:10.1016/s0300-9084(02)00020-2

Brasier, M., McLoughlin, N., Green, O., & Wacey, D. (2006). A fresh look at the fossil evidence for early Archaean cellular life. *Philos Trans R Soc Lond B Biol Sci, 361*(1470), 887-902. doi:10.1098/rstb.2006.1835

Briner, A. E., Lugli, G. A., Milani, C., Duranti, S., Turroni, F., Gueimonde, M., . . . Barrangou, R. (2015). Occurrence and Diversity of CRISPR-Cas Systems in the Genus Bifidobacterium. *PLoS One, 10*(7), e0133661. doi:10.1371/journal.pone.0133661

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., . . . van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science, 321*(5891), 960-964. doi:10.1126/science.1159689

Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., . . . Banfield, J. F. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat Commun, 7*, 10613. doi:10.1038/ncomms10613

Camara-Wilpert, S., Mayo-Munoz, D., Russel, J., Fagerlund, R. D., Madsen, J. S., Fineran, P. C., . . . Pinilla-Redondo, R. (2023). Bacteriophages suppress CRISPR-Cas immunity using RNA-based anti-CRISPRs. *Nature, 623*(7987), 601-607. doi:10.1038/s41586-023-06612-5

Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M., & Terns, M. P. (2010). Binding and cleavage of CRISPR RNA by Cas6. *RNA, 16*(11), 2181-2188. doi:10.1261/rna.2230110

Carte, J., Wang, R., Li, H., Terns, R. M., & Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev, 22*(24), 3489-3496. doi:10.1101/gad.1742908

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell, 136*(4), 642-655. doi:10.1016/j.cell.2009.01.035

Charpentier, E., Richter, H., van der Oost, J., & White, M. F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol Rev, 39*(3), 428-441. doi:10.1093/femsre/fuv023

Chen, L. X., Al-Shayeb, B., Meheust, R., Li, W. J., Doudna, J. A., & Banfield, J. F. (2019). Candidate Phyla Radiation Roizmanbacteria From Hot Springs Have Novel and Unexpectedly Abundant CRISPR-Cas Systems. *Front Microbiol, 10*, 928. doi:10.3389/fmicb.2019.00928

Cheng, R., Huang, F., Wu, H., Lu, X., Yan, Y., Yu, B., . . . Zhu, B. (2021). A nucleotide-sensing endonuclease from the Gabija bacterial defense system. *Nucleic Acids Res, 49*(9), 5216-5229. doi:10.1093/nar/gkab277

Chevez-Guardado, R., & Pena-Castillo, L. (2021). Promotech: a general tool for bacterial promoter recognition. *Genome Biol, 22*(1), 318. doi:10.1186/s13059-021-02514-9

Chylinski, K., Makarova, K. S., Charpentier, E., & Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res, 42*(10), 6091-6105. doi:10.1093/nar/gku241

Chyou, T. Y., & Brown, C. M. (2019). Prediction and diversity of tracrRNAs from type II CRISPR-Cas systems. *RNA Biol, 16*(4), 423-434. doi:10.1080/15476286.2018.1498281

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Neron, B., . . . Pourcel, C. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res, 46*(W1), W246-W251. doi:10.1093/nar/gky425

Cui, N., Zhang, J. T., Liu, Y., Liu, Y., Liu, X. Y., Wang, C., . . . Jia, N. (2023). Type IV-A CRISPR-Csf complex: Assembly, dsDNA targeting, and CasDinG recruitment. *Mol Cell, 83*(14), 2493-2508 e2495. doi:10.1016/j.molcel.2023.05.036

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*(18), 5012-5019. doi:10.1128/JB.00843-12

Danna, K., & Nathans, D. (1971). Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae. *Proc Natl Acad Sci U S A, 68*(12), 2913-2917. doi:10.1073/pnas.68.12.2913

Darty, K., Denise, A., & Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics, 25*(15), 1974-1975. doi:10.1093/bioinformatics/btp250

Das, U., & Shuman, S. (2013). Mechanism of RNA 2',3'-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase. *Nucleic Acids Res, 41*(1), 355-365. doi:10.1093/nar/gks977

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. doi:10.1038/ncomms1937

Dedrick, R. M., Guerrero-Bustamante, C. A., Garlena, R. A., Russell, D. A., Ford, K., Harris, K., . . . Spencer, H. (2019). Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. *Nat Med, 25*(5), 730-733. doi:10.1038/s41591-019-0437-z

Deep, A., Gu, Y., Gao, Y. Q., Ego, K. M., Herzik, M. A., Jr., Zhou, H., & Corbett, K. D. (2022). The SMC-family Wadjet complex protects bacteria from plasmid transformation by recognition and cleavage of closed-circular DNA. *Mol Cell, 82*(21), 4145-4159 e4147. doi:10.1016/j.molcel.2022.09.008

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., . . . Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature, 471*(7340), 602-607. doi:10.1038/nature09886

Depardieu, F., Didier, J. P., Bernheim, A., Sherlock, A., Molina, H., Duclos, B., & Bikard, D. (2016). A Eukaryotic-like Serine/Threonine Kinase Protects Staphylococci against Phages. *Cell Host Microbe, 20*(4), 471-481. doi:10.1016/j.chom.2016.08.010

Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., . . . Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J Bacteriol, 190*(4), 1390-1400. doi:10.1128/JB.01412-07

Devi, V., Harjai, K., & Chhibber, S. (2022). Self-targeting spacers in CRISPR-array: Accidental occurrence or evolutionarily conserved phenomenon. *J Basic Microbiol, 62*(1), 4-12. doi:10.1002/jobm.202100514

Diez-Villasenor, C., Guzman, N. M., Almendros, C., Garcia-Martinez, J., & Mojica, F. J. (2013). CRISPR-spacer integration reporter plasmids reveal distinct genuine

acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. *RNA Biol, 10*(5), 792-802. doi:10.4161/rna.24023

Dillingham, M. S., & Kowalczykowski, S. C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev, 72*(4), 642-671, Table of Contents. doi:10.1128/MMBR.00020-08

Domingues, S., Chopin, A., Ehrlich, S. D., & Chopin, M. C. (2004). The Lactococcal abortive phage infection system AbiP prevents both phage DNA replication and temporal transcription switch. *J Bacteriol, 186*(3), 713-721. doi:10.1128/JB.186.3.713-721.2004

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., . . . Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science, 359*(6379). doi:10.1126/science.aar4120

East-Seletsky, A., O'Connell, M. R., Burstein, D., Knott, G. J., & Doudna, J. A. (2017). RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes. *Mol Cell, 66*(3), 373-383 e373. doi:10.1016/j.molcel.2017.04.008

East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H., Tjian, R., & Doudna, J. A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature, 538*(7624), 270-273. doi:10.1038/nature19802

Edgar, R., & Qimron, U. (2010). The Escherichia coli CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. *J Bacteriol, 192*(23), 6291-6294. doi:10.1128/JB.00644-10

Edwards, R. A., & Rohwer, F. (2005). Viral metagenomics. *Nat Rev Microbiol, 3*(6), 504-510. doi:10.1038/nrmicro1163

Ellinger, P., Arslan, Z., Wurm, R., Tschapek, B., MacKenzie, C., Pfeffer, K., . . . Smits, S. H. (2012). The crystal structure of the CRISPR-associated protein Csn2 from Streptococcus agalactiae. *J Struct Biol, 178*(3), 350-362. doi:10.1016/j.jsb.2012.04.006

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*(4), 447-459. doi:10.1101/gad.272153.115

Erdmann, S., & Garrett, R. A. (2012). Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol Microbiol, 85*(6), 1044-1056. doi:10.1111/j.1365-2958.2012.08171.x

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*(4), 460-470. doi:10.1101/gad.273722.115

Fagerlund, R. D., Wilkinson, M. E., Klykov, O., Barendregt, A., Pearce, F. G., Kieper, S. N., . . . Fineran, P. C. (2017). Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. *Proc Natl Acad Sci U S A, 114*(26), E5122-E5128. doi:10.1073/pnas.1618421114

Faure, G., Shmakov, S. A., Yan, W. X., Cheng, D. R., Scott, D. A., Peters, J. E., . . . Koonin, E. V. (2019). CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nat Rev Microbiol, 17*(8), 513-525. doi:10.1038/s41579-019-0204-7

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science, 281*(5374), 237-240. doi:10.1126/science.281.5374.237

Filipowicz, W. (2005). RNAi: the nuts and bolts of the RISC machine. *Cell, 122*(1), 17-20. doi:10.1016/j.cell.2005.06.023

Fineran, P. C., Blower, T. R., Foulds, I. J., Humphreys, D. P., Lilley, K. S., & Salmond, G. P. (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A, 106*(3), 894-899. doi:10.1073/pnas.0808832106

Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A., & Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature, 532*(7600), 517-521. doi:10.1038/nature17945

Foss, D. V., Hochstrasser, M. L., & Wilson, R. C. (2019). Clinical applications of CRISPR-based genome editing and diagnostics. *Transfusion, 59*(4), 1389-1399. doi:10.1111/trf.15126

Freije, C. A., Myhrvold, C., Boehm, C. K., Lin, A. E., Welch, N. L., Carter, A., . . . Sabeti, P. C. (2019). Programmable Inhibition and Detection of RNA Viruses Using Cas13. *Mol Cell, 76*(5), 826-837 e811. doi:10.1016/j.molcel.2019.09.013

Fujita, T., & Fujii, H. (2013). Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem Biophys Res Commun, 439*(1), 132-136. doi:10.1016/j.bbrc.2013.08.013

Gao, P., Yang, H., Rajashankar, K. R., Huang, Z., & Patel, D. J. (2016). Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res, 26*(8), 901-913. doi:10.1038/cr.2016.88

Garneau, J. E., Dupuis, M. E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., . . . Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature, 468*(7320), 67-71. doi:10.1038/nature09523

Garside, E. L., Schellenberg, M. J., Gesner, E. M., Bonanno, J. B., Sauder, J. M., Burley, S. K., . . . MacMillan, A. M. (2012). Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *RNA, 18*(11), 2020-2028. doi:10.1261/rna.033100.112

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A, 109*(39), E2579-2586. doi:10.1073/pnas.1208507109

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature, 551*(7681), 464-471. doi:10.1038/nature24644

Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M., & Macmillan, A. M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat Struct Mol Biol, 18*(6), 688-692. doi:10.1038/nsmb.2042

Goldberg, G. W., Jiang, W., Bikard, D., & Marraffini, L. A. (2014). Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature, 514*(7524), 633-637. doi:10.1038/nature13637

Gong, B., Shin, M., Sun, J., Jung, C. H., Bolt, E. L., van der Oost, J., & Kim, J. S. (2014). Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc Natl Acad Sci U S A, 111*(46), 16359-16364. doi:10.1073/pnas.1410806111

Gonzalez-Delgado, A., Mestre, M. R., Martinez-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*(19), 10202-10211. doi:10.1093/nar/gkz746

Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., . . . Zhang, F. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science, 356*(6336), 438-442. doi:10.1126/science.aam9321

Goren, M. G., Yosef, I., Auster, O., & Qimron, U. (2012). Experimental definition of a clustered regularly interspaced short palindromic duplicon in Escherichia coli. *J Mol Biol, 423*(1), 14-16. doi:10.1016/j.jmb.2012.06.037

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*. doi:10.7554/eLife.81678

Gregory, A. C., Zayed, A. A., Conceicao-Neto, N., Temperton, B., Bolduc, B., Alberti, A., . . . Sullivan, M. B. (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell, 177*(5), 1109-1123 e1114. doi:10.1016/j.cell.2019.03.040

Guadarrama, C., Medrano-Lopez, A., Oropeza, R., Hernandez-Lucas, I., & Calva, E. (2014). The Salmonella enterica serovar Typhi LeuO global regulator forms tetramers: residues involved in oligomerization, DNA binding, and transcriptional regulation. *J Bacteriol, 196*(12), 2143-2154. doi:10.1128/JB.01484-14

Hale, C., Kleppe, K., Terns, R. M., & Terns, M. P. (2008). Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA, 14*(12), 2572-2579. doi:10.1261/rna.1246808

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., . . . Terns, M. P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell, 139*(5), 945-956. doi:10.1016/j.cell.2009.07.040

Hammond, S. M. (2005). Dicing and slicing: the core machinery of the RNA interference pathway. *FEBS Lett, 579*(26), 5822-5829. doi:10.1016/j.febslet.2005.08.079

Hampton, H. G., Watson, B. N. J., & Fineran, P. C. (2020). The arms race between bacteria and their phage foes. *Nature, 577*(7790), 327-336. doi:10.1038/s41586-019-1894-8

Han, D., & Krauss, G. (2009). Characterization of the endonuclease SSO2001 from Sulfolobus solfataricus P2. *FEBS Lett, 583*(4), 771-776. doi:10.1016/j.febslet.2009.01.024

Han, X., Chang, L., Chen, H., Zhao, J., Tian, F., Ross, R. P., . . . Yang, B. (2024). Harnessing the endogenous Type I-C CRISPR-Cas system for genome editing in Bifidobacterium breve. *Appl Environ Microbiol*, e0207423. doi:10.1128/aem.02074-23

Harrington, L. B., Burstein, D., Chen, J. S., Paez-Espino, D., Ma, E., Witte, I. P., . . . Doudna, J. A. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science, 362*(6416), 839-842. doi:10.1126/science.aav4294

Harrington, L. B., Ma, E., Chen, J. S., Witte, I. P., Gertz, D., Paez-Espino, D., . . . Doudna, J. A. (2020). A scoutRNA Is Required for Some Type V CRISPR-Cas Systems. *Mol Cell, 79*(3), 416-424 e415. doi:10.1016/j.molcel.2020.06.022

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., & Doudna, J. A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science, 329*(5997), 1355-1358. doi:10.1126/science.1192272

Hayes, R. P., Xiao, Y., Ding, F., van Erp, P. B., Rajashankar, K., Bailey, S., . . . Ke, A. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature, 530*(7591), 499-503. doi:10.1038/nature16995

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*(7542), 199-202. doi:10.1038/nature14245

Hermans, P. W., van Soolingen, D., Bik, E. M., de Haas, P. E., Dale, J. W., & van Embden, J. D. (1991). Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains. *Infect Immun, 59*(8), 2695-2705. doi:10.1128/iai.59.8.2695-2705.1991

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*(1), 21-33. doi:10.1016/j.jmb.2018.09.013

Hille, F., & Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. *Philos Trans R Soc Lond B Biol Sci, 371*(1707). doi:10.1098/rstb.2015.0496

Hille, F., Richter, H., Wong, S. P., Bratovic, M., Ressel, S., & Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. *Cell, 172*(6), 1239-1259. doi: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*(1), 58-66. doi:10.1016/j.tibs.2014.10.007

Hochstrasser, M. L., Taylor, D. W., Bhat, P., Guegler, C. K., Sternberg, S. H., Nogales, E., & Doudna, J. A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc Natl Acad Sci U S A, 111*(18), 6618-6623. doi:10.1073/pnas.1405079111

Hoyland-Kroghsbo, N. M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-Denomy, J., & Bassler, B. L. (2017). Quorum sensing controls the Pseudomonas aeruginosa CRISPR-Cas adaptive immune system. *Proc Natl Acad Sci U S A, 114*(1), 131-135. doi:10.1073/pnas.1617415113

Hu, C., Almendros, C., Nam, K. H., Costa, A. R., Vink, J. N. A., Haagsma, A. C., . . . Ke, A. (2021). Mechanism for Cas4-assisted directional spacer acquisition in CRISPR-Cas. *Nature, 598*(7881), 515-520. doi:10.1038/s41586-021-03951-z

Hu, C., Myers, M. T., Zhou, X., Hou, Z., Lozen, M. L., Nam, K. H., . . . Ke, A. (2024). Exploiting activation and inactivation mechanisms in type I-C CRISPR-Cas3 for genome-editing applications. *Mol Cell, 84*(3), 463-475 e465. doi:10.1016/j.molcel.2023.12.034

Hu, C., van Beljouw, S. P. B., Nam, K. H., Schuler, G., Ding, F., Cui, Y., . . . Ke, A. (2022). Craspase is a CRISPR RNA-guided, RNA-activated protease. *Science, 377*(6612), 1278-1285. doi:10.1126/science.add5064

Hu, Y., & Li, W. (2022). Development and Application of CRISPR-Cas Based Tools. *Front Cell Dev Biol, 10*, 834646. doi:10.3389/fcell.2022.834646

Huo, G., Shepherd, J., & Pan, X. (2023). Craspase: A novel CRISPR/Cas dual gene editor. *Funct Integr Genomics, 23*(2), 98. doi:10.1007/s10142-023-01024-0

Huo, Y., Nam, K. H., Ding, F., Lee, H., Wu, L., Xiao, Y., . . . Ke, A. (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat Struct Mol Biol, 21*(9), 771-777. doi:10.1038/nsmb.2875

Ishino, Y., Krupovic, M., & Forterre, P. (2018). History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. *J Bacteriol, 200*(7). doi:10.1128/JB.00580-17

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*(12), 5429-5433. doi:10.1128/jb.169.12.5429-5433.1987

Ivancic-Bace, I., Cass, S. D., Wearne, S. J., & Bolt, E. L. (2015). Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity. *Nucleic Acids Res, 43*(22), 10821-10830. doi:10.1093/nar/gkv1213

Iwakawa, H. O., & Tomari, Y. (2022). Life of RISC: Formation, action, and degradation of RNA-induced silencing complex. *Mol Cell, 82*(1), 30-43. doi:10.1016/j.molcel.2021.11.026

Jackson, R. N., Golden, S. M., van Erp, P. B., Carter, J., Westra, E. R., Brouns, S. J., . . . Wiedenheft, B. (2014). Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. *Science, 345*(6203), 1473-1479. doi: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*(5), 722-728. doi:10.1016/j.molcel.2015.05.023

Jansen, R., Embden, J. D., Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol, 43*(6), 1565-1575. doi:10.1046/j.1365-2958.2002.02839.x

Jaroszewski, L., Li, Z., Cai, X. H., Weber, C., & Godzik, A. (2011). FFAS server: novel features and applications. *Nucleic Acids Res, 39*(Web Server issue), W38-44. doi:10.1093/nar/gkr441

Jesser, R., Behler, J., Benda, C., Reimann, V., & Hess, W. R. (2019). Biochemical analysis of the Cas6-1 RNA endonuclease associated with the subtype I-D CRISPR-Cas system in Synechocystis sp. PCC 6803. *RNA Biol, 16*(4), 481-491. doi:10.1080/15476286.2018.1447742

Jia, N., Jones, R., Yang, G., Ouerfelli, O., & Patel, D. J. (2019). CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise cA(4) Cleavage with ApA>p Formation Terminating RNase Activity. *Mol Cell, 75*(5), 944-956 e946. doi:10.1016/j.molcel.2019.06.014

Jiang, F., & Doudna, J. A. (2015). The structural biology of CRISPR-Cas systems. *Curr Opin Struct Biol, 30*, 100-111. doi:10.1016/j.sbi.2015.02.002

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol, 31*(3), 233-239. doi:10.1038/nbt.2508

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, 337*(6096), 816-821. doi:10.1126/science.1225829

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *Elife, 2*, e00471. doi:10.7554/eLife.00471

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., . . . Doudna, J. A. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science, 343*(6176), 1247997. doi:10.1126/science.1247997

Jore, M. M., Brouns, S. J., & van der Oost, J. (2012). RNA in defense: CRISPRs protect prokaryotes against mobile genetic elements. *Cold Spring Harb Perspect Biol, 4*(6). doi:10.1101/cshperspect.a003657

Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., . . . Brouns, S. J. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol, 18*(5), 529-536. doi:10.1038/nsmb.2019

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583-589. doi:10.1038/s41586-021-03819-2

Kaufmann, G., Klein, T., & Littauer, U. Z. (1974). T4 RNA ligase: substrate chain length requirements. *FEBS Lett, 46*(1), 271-275. doi:10.1016/0014-5793(74)80385-6

Kazlauskiene, M., Kostiuk, G., Venclovas, C., Tamulaitis, G., & Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science, 357*(6351), 605-609. doi:10.1126/science.aao0100

Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, C., & Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Mol Cell, 62*(2), 295-306. doi:10.1016/j.molcel.2016.03.024

Kieper, S. N., Almendros, C., Behler, J., McKenzie, R. E., Nobrega, F. L., Haagsma, A. C., . . . Brouns, S. J. J. (2018). Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Rep, 22*(13), 3377-3384. doi:10.1016/j.celrep.2018.02.103

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A, 93*(3), 1156-1160. doi:10.1073/pnas.93.3.1156

Kim, Y. G., & Chandrasegaran, S. (1994). Chimeric restriction endonuclease. *Proc Natl Acad Sci U S A, 91*(3), 883-887. doi:10.1073/pnas.91.3.883

Kim, Y. G., Shi, Y., Berg, J. M., & Chandrasegaran, S. (1997). Site-specific cleavage of DNA-RNA hybrids by zinc finger/FokI cleavage domain fusions. *Gene, 203*(1), 43-49. doi:10.1016/s0378-1119(97)00489-7

Klompe, S. E., & Sternberg, S. H. (2018). Harnessing "A Billion Years of Experimentation": The Ongoing Exploration and Exploitation of CRISPR-Cas Immune Systems. *CRISPR J, 1*(2), 141-158. doi:10.1089/crispr.2018.0012

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*(10), 825-833. doi:10.1038/nsmb.3466

Kolesnik, M. V., Fedorova, I., Karneyeva, K. A., Artamonova, D. N., & Severinov, K. V. (2021). Type III CRISPR-Cas Systems: Deciphering the Most Complex Prokaryotic Immune System. *Biochemistry (Mosc), 86*(10), 1301-1314. doi:10.1134/S0006297921100114

Konermann, S., Lotfy, P., Brideau, N. J., Oki, J., Shokhirev, M. N., & Hsu, P. D. (2018). Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell, 173*(3), 665-676 e614. doi:10.1016/j.cell.2018.02.033

Koo, Y., Ka, D., Kim, E. J., Suh, N., & Bae, E. (2013). Conservation and variability in the structure and function of the Cas5d endoribonuclease in the CRISPR-mediated microbial immune system. *J Mol Biol, 425*(20), 3799-3810. doi:10.1016/j.jmb.2013.02.032

Koonin, E. V., Makarova, K. S., & Wolf, Y. I. (2017). Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu Rev Microbiol, 71*, 233-261. doi:10.1146/annurev-micro-090816-093830

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

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nat Rev Microbiol, 8*(5), 317-327. doi:10.1038/nrmicro2315

Le Rhun, A., Escalera-Maurer, A., Bratovic, M., & Charpentier, E. (2019). CRISPR-Cas in Streptococcus pyogenes. *RNA Biol, 16*(4), 380-389. doi:10.1080/15476286.2019.1582974

Lee, H., Dhingra, Y., & Sashital, D. G. (2019). The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. *Elife, 8*. doi: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*(1), 48-59 e45. doi:10.1016/j.molcel.2018.03.003

Lee, K. H., Lee, S. G., Eun Lee, K., Jeon, H., Robinson, H., & Oh, B. H. (2012). Identification, structural, and biochemical characterization of a group of large Csn2 proteins involved in CRISPR-mediated bacterial immunity. *Proteins, 80*(11), 2573-2582. doi:10.1002/prot.24138

LeRoux, M., Srikant, S., Teodoro, G. I. C., Zhang, T., Littlehale, M. L., Doron, S., . . . Laub, M. T. (2022). The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA. *Nat Microbiol, 7*(7), 1028-1040. doi:10.1038/s41564-022-01153-5

Letarov, A. V., & Kulikov, E. E. (2017). Adsorption of Bacteriophages on Bacterial Cells. *Biochemistry (Mosc), 82*(13), 1632-1658. doi:10.1134/S0006297917130053

Levy, A., Goren, M. G., Yosef, I., Auster, O., Manor, M., Amitai, G., . . . Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature, 520*(7548), 505-510. doi:10.1038/nature14302

Li, H. (2015). Structural Principles of CRISPR RNA Processing. *Structure, 23*(1), 13-20. doi:10.1016/j.str.2014.10.006

Li, Y., & Bondy-Denomy, J. (2021). Anti-CRISPRs go viral: The infection biology of CRISPR-Cas inhibitors. *Cell Host Microbe, 29*(5), 704-714. doi:10.1016/j.chom.2020.12.007

Lieber, M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem, 79*, 181-211. doi:10.1146/annurev.biochem.052308.093131

Lillestol, R. K., Redder, P., Garrett, R. A., & Brugger, K. (2006). A putative viral defence mechanism in archaeal cells. *Archaea, 2*(1), 59-72. doi:10.1155/2006/542818

Lillestol, R. K., Shah, S. A., Brugger, K., Redder, P., Phan, H., Christiansen, J., & Garrett, R. A. (2009). CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. *Mol Microbiol, 72*(1), 259-272. doi:10.1111/j.1365-2958.2009.06641.x

Lin, J., Fuglsang, A., Kjeldsen, A. L., Sun, K., Bhoobalan-Chitty, Y., & Peng, X. (2020). DNA targeting by subtype I-D CRISPR-Cas shows type I and type III features. *Nucleic Acids Res, 48*(18), 10470-10478. doi:10.1093/nar/gkaa749

Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., . . . 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*(24), 21643-21656. doi:10.1074/jbc.M111.238485

Lisitskaya, L., Aravin, A. A., & Kulbachinskiy, A. (2018). DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. *Nat Commun, 9*(1), 5165. doi:10.1038/s41467-018-07449-7

Liu, J. J., Orlova, N., Oakes, B. L., Ma, E., Spinner, H. B., Baney, K. L. M., . . . Doudna, J. A. (2019). CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature, 566*(7743), 218-223. doi:10.1038/s41586-019-0908-x

Liu, L., Chen, P., Wang, M., Li, X., Wang, J., Yin, M., & Wang, Y. (2017). C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism. *Mol Cell, 65*(2), 310-322. doi:10.1016/j.molcel.2016.11.040

Liu, L., Li, X., Ma, J., Li, Z., You, L., Wang, J., . . . Wang, Y. (2017). The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell, 170*(4), 714-726 e710. doi:10.1016/j.cell.2017.06.050

Liu, L., Li, X., Wang, J., Wang, M., Chen, P., Yin, M., . . . Wang, Y. (2017). Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell, 168*(1-2), 121-134 e112. doi:10.1016/j.cell.2016.12.031

Liu, M., Deora, R., Doulatov, S. R., Gingery, M., Eiserling, F. A., Preston, A., . . . Miller, J. F. (2002). Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. *Science, 295*(5562), 2091-2094. doi:10.1126/science.1067467

Loenen, W. A., & Raleigh, E. A. (2014). The other face of restriction: modification-dependent enzymes. *Nucleic Acids Res, 42*(1), 56-69. doi:10.1093/nar/gkt747

Maharajan, A. D., Hjerde, E., Hansen, H., & Willassen, N. P. (2022). Quorum Sensing Controls the CRISPR and Type VI Secretion Systems in Aliivibrio wodanis 06/09/139. *Front Vet Sci, 9*, 799414. doi:10.3389/fvets.2022.799414

Mahas, A., Aman, R., & Mahfouz, M. (2019). CRISPR-Cas13d mediates robust RNA virus interference in plants. *Genome Biol, 20*(1), 263. doi:10.1186/s13059-019-1881-2

Maier, L. K., Stachler, A. E., Brendel, J., Stoll, B., Fischer, S., Haas, K. A., . . . Marchfelder, A. (2019). The nuts and bolts of the Haloferax CRISPR-Cas system I-B. *RNA Biol, 16*(4), 469-480. doi:10.1080/15476286.2018.1460994

Majumdar, S., & Terns, M. P. (2019). CRISPR RNA-guided DNA cleavage by reconstituted Type I-A immune effector complexes. *Extremophiles, 23*(1), 19-33. doi:10.1007/s00792-018-1057-0

Makarova, K. S., Aravind, L., Grishin, N. V., Rogozin, I. B., & Koonin, E. V. (2002). A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res, 30*(2), 482-496. doi:10.1093/nar/30.2.482

Makarova, K. 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. doi: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. doi:10.1186/1745-6150-1-7

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., . . . Koonin, E. V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol, 9*(6), 467-477. doi:10.1038/nrmicro2577

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., . . . Koonin, E. V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol, 13*(11), 722-736. doi:10.1038/nrmicro3569

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., . . . Koonin, E. V. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol, 18*(2), 67-83. doi:10.1038/s41579-019-0299-x

Makarova, K. S., Wolf, Y. I., Snir, S., & Koonin, E. V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J Bacteriol, 193*(21), 6039-6056. doi:10.1128/JB.05535-11

Makarova, K. S., Wolf, Y. I., van der Oost, J., & Koonin, E. V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol Direct, 4*, 29. doi:10.1186/1745-6150-4-29

Malone, L. M., Birkholz, N., & Fineran, P. C. (2021). Conquering CRISPR: how phages overcome bacterial adaptive immunity. *Curr Opin Biotechnol, 68*, 30-36. doi:10.1016/j.copbio.2020.09.008

Malone, L. M., Warring, S. L., Jackson, S. A., Warnecke, C., Gardner, P. P., Gumy, L. F., & Fineran, P. C. (2020). A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nat Microbiol, 5*(1), 48-55. doi:10.1038/s41564-019-0612-5

Marraffini, L. A., & Sontheimer, E. J. (2010a). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet, 11*(3), 181-190. doi:10.1038/nrg2749

Marraffini, L. A., & Sontheimer, E. J. (2010b). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature, 463*(7280), 568-571. doi:10.1038/nature08703

McGinn, J., & Marraffini, L. A. (2016). CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. *Mol Cell, 64*(3), 616-623. doi:10.1016/j.molcel.2016.08.038

Medina-Aparicio, L., Rebollar-Flores, J. E., Gallego-Hernandez, A. L., Vazquez, A., Olvera, L., Gutierrez-Rios, R. M., . . . Hernandez-Lucas, I. (2011). The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in Salmonella enterica serovar Typhi. *J Bacteriol, 193*(10), 2396-2407. doi:10.1128/JB.01480-10

Meisel, A., Bickle, T. A., Kruger, D. H., & Schroeder, C. (1992). Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature, 355*(6359), 467-469. doi:10.1038/355467a0

Mendoza, S. D., Nieweglowska, E. S., Govindarajan, S., Leon, L. M., Berry, J. D., Tiwari, A., . . . Bondy-Denomy, J. (2020). A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. *Nature, 577*(7789), 244-248. doi:10.1038/s41586-019-1786-y

Mitic, D., Radovcic, M., Markulin, D., & Ivancic-Bace, I. (2020). StpA represses CRISPR-Cas immunity in H-NS deficient Escherichia coli. *Biochimie, 174*, 136-143. doi:10.1016/j.biochi.2020.04.020

Modell, J. W., Jiang, W., & Marraffini, L. A. (2017). CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature, 544*(7648), 101-104. doi:10.1038/nature21719

Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., Koonin, E. V., & van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science, 353*(6299), aad5147. doi:10.1126/science.aad5147

Mojica, F. J., 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*(2), 174-182. doi:10.1007/s00239-004-0046-3

Mojica, F. J., Diez-Villasenor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol, 36*(1), 244-246. doi:10.1046/j.1365-2958.2000.01838.x

Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading), 155*(Pt 3), 733-740. doi:10.1099/mic.0.023960-0

Molina, R., Stella, S., Feng, M., Sofos, N., Jauniskis, V., Pozdnyakova, I., . . . Montoya, G. (2019). Structure of Csx1-cOA(4) complex reveals the basis of RNA decay in Type III-B CRISPR-Cas. *Nat Commun, 10*(1), 4302. doi:10.1038/s41467-019-12244-z

Mulepati, S., Heroux, A., & Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science, 345*(6203), 1479-1484. doi:10.1126/science.1256996

Murray, N. E. (2000). Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev, 64*(2), 412-434. doi:10.1128/MMBR.64.2.412-434.2000

Nakata, A., Amemura, M., & Makino, K. (1989). Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome. *J Bacteriol, 171*(6), 3553-3556. doi:10.1128/jb.171.6.3553-3556.1989

Nam, K. H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M. P., & Ke, A. (2012). Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure, 20*(9), 1574-1584. doi:10.1016/j.str.2012.06.016

Navarre, W. W., McClelland, M., Libby, S. J., & Fang, F. C. (2007). Silencing of xenogeneic DNA by H-NS-facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. *Genes Dev, 21*(12), 1456-1471. doi:10.1101/gad.1543107

Niewoehner, O., Garcia-Doval, C., Rostol, J. T., Berk, C., Schwede, F., Bigler, L., . . . Jinek, M. (2017). Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature, 548*(7669), 543-548. doi:10.1038/nature23467

Nilsen, T. W. (2013). RNA structure determination using nuclease digestion. *Cold Spring Harb Protoc, 2013*(4), 379-382. doi:10.1101/pdb.prot072330

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*(5), 2486-2501. doi:10.1093/nar/gkz1218

Nishimasu, H., Cong, L., Yan, W. X., Ran, F. A., Zetsche, B., Li, Y., . . . Nureki, O. (2015). Crystal Structure of Staphylococcus aureus Cas9. *Cell, 162*(5), 1113-1126. doi:10.1016/j.cell.2015.08.007

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., . . . Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell, 156*(5), 935-949. doi:10.1016/j.cell.2014.02.001

Nunez, 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*(6), 824-833. doi:10.1016/j.molcel.2016.04.027

Nunez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N., & Doudna, J. A. (2015). Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature, 527*(7579), 535-538. doi:10.1038/nature15760

Nunez, J. K., Lee, A. S., Engelman, A., & Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature, 519*(7542), 193-198. doi:10.1038/nature14237

Nussenzweig, P. M., & Marraffini, L. A. (2020). Molecular Mechanisms of CRISPR-Cas Immunity in Bacteria. *Annu Rev Genet, 54*, 93-120. doi:10.1146/annurev-genet-022120-112523

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*(4), 515-526 e516. doi: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*(1), 5931. doi:10.1038/s41467-020-19785-8

O'Connell, M. R. (2019). Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR-Cas Systems. *J Mol Biol, 431*(1), 66-87. doi:10.1016/j.jmb.2018.06.029

O'Sullivan, L., Bolton, D., McAuliffe, O., & Coffey, A. (2019). Bacteriophages in Food Applications: From Foe to Friend. *Annu Rev Food Sci Technol, 10*, 151-172. doi:10.1146/annurev-food-032818-121747

Ofir, G., Herbst, E., Baroz, M., Cohen, D., Millman, A., Doron, S., . . . Sorek, R. (2021). Antiviral activity of bacterial TIR domains via immune signalling molecules. *Nature, 600*(7887), 116-120. doi:10.1038/s41586-021-04098-7

Osawa, T., Inanaga, H., & Numata, T. (2013). Crystal structure of the Cmr2-Cmr3 subcomplex in the CRISPR-Cas RNA silencing effector complex. *J Mol Biol, 425*(20), 3811-3823. doi:10.1016/j.jmb.2013.03.042

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*(3), 418-430. doi:10.1016/j.molcel.2015.03.018

Ozcan, A., Krajeski, R., Ioannidi, E., Lee, B., Gardner, A., Makarova, K. S., . . . Gootenberg, J. S. (2021). Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. *Nature, 597*(7878), 720-725. doi:10.1038/s41586-021-03886-5

Ozcan, A., Pausch, P., Linden, A., Wulf, A., Schuhle, K., Heider, J., . . . Randau, L. (2019). Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nat Microbiol, 4*(1), 89-96. doi:10.1038/s41564-018-0274-8

Paez-Espino, D., Morovic, W., Sun, C. L., Thomas, B. C., Ueda, K., Stahl, B., . . . Banfield, J. F. (2013). Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun, 4*, 1430. doi:10.1038/ncomms2440

Park, J. U., Tsai, A. W., Rizo, A. N., Truong, V. H., Wellner, T. X., Schargel, R. D., & Kellogg, E. H. (2023). Structures of the holo CRISPR RNA-guided transposon integration complex. *Nature, 613*(7945), 775-782. doi:10.1038/s41586-022-05573-5

Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., . . . Brockhurst, M. A. (2010). Antagonistic coevolution accelerates molecular evolution. *Nature, 464*(7286), 275-278. doi:10.1038/nature08798

Patsali, P., Kleanthous, M., & Lederer, C. W. (2019). Disruptive Technology: CRISPR/Cas-Based Tools and Approaches. *Mol Diagn Ther, 23*(2), 187-200. doi:10.1007/s40291-019-00391-4

Patterson, A. G., Chang, J. T., Taylor, C., & Fineran, P. C. (2015). Regulation of the Type I-F CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. *Nucleic Acids Res, 43*(12), 6038-6048. doi:10.1093/nar/gkv517

Patterson, A. G., Jackson, S. A., Taylor, C., Evans, G. B., Salmond, G. P. C., Przybilski, R., . . . Fineran, P. C. (2016). Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems. *Mol Cell, 64*(6), 1102-1108. doi:10.1016/j.molcel.2016.11.012

Pausch, P., Al-Shayeb, B., Bisom-Rapp, E., Tsuchida, C. A., Li, Z., Cress, B. F., . . . Doudna, J. A. (2020). CRISPR-CasPhi from huge phages is a hypercompact genome editor. *Science, 369*(6501), 333-337. doi:10.1126/science.abb1400

Pausch, P., Muller-Esparza, H., Gleditzsch, D., Altegoer, F., Randau, L., & Bange, G. (2017). Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. *Mol Cell, 67*(4), 622-632 e624. doi:10.1016/j.molcel.2017.06.036

Perculija, V., Lin, J., Zhang, B., & Ouyang, S. (2021). Functional Features and Current Applications of the RNA-Targeting Type VI CRISPR-Cas Systems. *Adv Sci (Weinh), 8*(13), 2004685. doi:10.1002/advs.202004685

Pinilla-Redondo, R., Mayo-Munoz, D., Russel, J., Garrett, R. A., Randau, L., Sorensen, S. J., & Shah, S. A. (2020). Type IV CRISPR-Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Res, 48*(4), 2000-2012. doi:10.1093/nar/gkz1197

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*(10), 2491-2500. doi:10.1128/JB.00206-12

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*(6), 1367-1379. doi: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 (Reading), 151*(Pt 3), 653-663. doi:10.1099/mic.0.27437-0

Pourcel, C., Touchon, M., Villeriot, N., Vernadet, J. P., Couvin, D., Toffano-Nioche, C., & Vergnaud, G. (2020). CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. *Nucleic Acids Res, 48*(D1), D535-D544. doi:10.1093/nar/gkz915

Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., & Wagner, R. (2010). Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. *Mol Microbiol, 75*(6), 1495-1512. doi:10.1111/j.1365-2958.2010.07073.x

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*(6), 3846-3856. doi:10.1093/nar/gkt1335

Qiu, J., Zhai, Y., Wei, M., Zheng, C., & Jiao, X. (2022). Toxin-antitoxin systems: Classification, biological roles, and applications. *Microbiol Res, 264*, 127159. doi:10.1016/j.micres.2022.127159

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*(11), 3403-3414. doi:10.1074/jbc.RA119.012196

Ramalingam, S., Annaluru, N., Kandavelou, K., & Chandrasegaran, S. (2014). TALEN-mediated generation and genetic correction of disease-specific human induced pluripotent stem cells. *Curr Gene Ther, 14*(6), 461-472. doi:10.2174/1566523214666140918101725

Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., . . . Greene, E. C. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell, 163*(4), 854-865. doi:10.1016/j.cell.2015.10.003

Reeks, J., Sokolowski, R. D., Graham, S., Liu, H., Naismith, J. H., & White, M. F. (2013). Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing. *Biochem J, 452*(2), 223-230. doi:10.1042/BJ20130269

Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L., & Corn, J. E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol, 34*(3), 339-344. doi:10.1038/nbt.3481

Richter, H., Lange, S. J., Backofen, R., & Randau, L. (2013). Comparative analysis ofCas6b processing and CRISPR RNA stability. *RNA Biol, 10*(5), 700-707. doi:10.4161/rna.23715

Riede, I., & Eschbach, M. L. (1986). Evidence that TraT interacts with OmpA of Escherichia coli. *FEBS Lett, 205*(2), 241-245. doi:10.1016/0014-5793(86)80905-x

Rollie, C., Graham, S., Rouillon, C., & White, M. F. (2018). Prespacer processing and specific integration in a Type I-A CRISPR system. *Nucleic Acids Res, 46*(3), 1007-1020. doi:10.1093/nar/gkx1232

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*. doi:10.7554/eLife.08716

Rollins, M. F., Schuman, J. T., Paulus, K., Bukhari, H. S., & Wiedenheft, B. (2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. *Nucleic Acids Res, 43*(4), 2216-2222. doi:10.1093/nar/gkv094

Rousseau, C., Gonnet, M., Le Romancer, M., & Nicolas, J. (2009). CRISPI: a CRISPR interactive database. *Bioinformatics, 25*(24), 3317-3318. doi:10.1093/bioinformatics/btp586

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*(5), 1164-1174. doi:10.1016/j.cell.2015.04.027

Samson, J. E., Magadan, A. H., Sabri, M., & Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol, 11*(10), 675-687. doi:10.1038/nrmicro3096

Sashital, D. G., Jinek, M., & Doudna, J. A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat Struct Mol Biol, 18*(6), 680-687. doi:10.1038/nsmb.2043

Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., & Severinov, K. (2013). High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. *RNA Biol, 10*(5), 716-725. doi:10.4161/rna.24325

Scadden, A. D. (2005). The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat Struct Mol Biol, 12*(6), 489-496. doi:10.1038/nsmb936

Scholz, I., Lange, S. J., Hein, S., Hess, W. R., & Backofen, R. (2013). CRISPR-Cas systems in the cyanobacterium Synechocystis sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. *PLoS One, 8*(2), e56470. doi:10.1371/journal.pone.0056470

Seed, K. D. (2015). Battling Phages: How Bacteria Defend against Viral Attack. *PLoS Pathog, 11*(6), e1004847. doi:10.1371/journal.ppat.1004847

Serbanescu, M. A., Cordova, M., Krastel, K., Flick, R., Beloglazova, N., Latos, A., . . . Cvitkovitch, D. G. (2015). Role of the Streptococcus mutans CRISPR-Cas systems in immunity and cell physiology. *J Bacteriol, 197*(4), 749-761. doi:10.1128/JB.02333-14

Shangguan, Q., Graham, S., Sundaramoorthy, R., & White, M. F. (2022). Structure and mechanism of the type I-G CRISPR effector. *Nucleic Acids Res, 50*(19), 11214-11228. doi:10.1093/nar/gkac925

Shangguan, Q., & White, M. F. (2023). Repurposing the atypical type I-G CRISPR system for bacterial genome engineering. *Microbiology (Reading), 169*(8). doi:10.1099/mic.0.001373

Shao, Y., & Li, H. (2013). Recognition and cleavage of a nonstructured CRISPR RNA by its processing endoribonuclease Cas6. *Structure, 21*(3), 385-393. doi:10.1016/j.str.2013.01.010

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*(4), 547-554. doi:10.1016/j.str.2016.02.009

Shiimori, M., Garrett, S. C., Graveley, B. R., & Terns, M. P. (2018). Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci. *Mol Cell, 70*(5), 814-824 e816. doi:10.1016/j.molcel.2018.05.002

Shipman, S. L., Nivala, J., Macklis, J. D., & Church, G. M. (2016). Molecular recordings by directed CRISPR spacer acquisition. *Science, 353*(6298), aaf1175. doi:10.1126/science.aaf1175

Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., . . . Koonin, E. V. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell, 60*(3), 385-397. doi:10.1016/j.molcel.2015.10.008

Shmakov, S., Savitskaya, E., Semenova, E., Logacheva, M. D., Datsenko, K. A., & Severinov, K. (2014). Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res, 42*(9), 5907-5916. doi:10.1093/nar/gku226

Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc Natl Acad Sci U S A, 115*(23), E5307-E5316. doi:10.1073/pnas.1803440115

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*(5). doi:10.1128/mBio.01397-17

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol, 7*, 539. doi:10.1038/msb.2011.75

Silas, S., Mohr, G., Sidote, D. J., Markham, L. M., Sanchez-Amat, A., Bhaya, D., . . . Fire, A. Z. (2016). Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science, 351*(6276), aad4234. doi:10.1126/science.aad4234

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J, 30*(7), 1335-1342. doi:10.1038/emboj.2011.41

Sinkunas, T., Gasiunas, G., & Siksnys, V. (2015). Cas3 nuclease-helicase activity assays. *Methods Mol Biol, 1311*, 277-291. doi:10.1007/978-1-4939-2687-9_18

Skennerton, C. T., Angly, F. E., Breitbart, M., Bragg, L., He, S., McMahon, K. D., . . . Tyson, G. W. (2011). Phage encoded H-NS: a potential achilles heel in the bacterial defence system. *PLoS One, 6*(5), e20095. doi:10.1371/journal.pone.0020095

Smargon, A. A., Cox, D. B. T., Pyzocha, N. K., Zheng, K., Slaymaker, I. M., Gootenberg, J. S., . . . Zhang, F. (2017). Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol Cell, 65*(4), 618-630 e617. doi:10.1016/j.molcel.2016.12.023

Sontheimer, E. J. (2005). Assembly and function of RNA silencing complexes. *Nat Rev Mol Cell Biol, 6*(2), 127-138. doi:10.1038/nrm1568

St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., & Shearwin, K. E. (2013). One-step cloning and chromosomal integration of DNA. *ACS Synth Biol, 2*(9), 537-541. doi:10.1021/sb400021j

Staals, R. H., Jackson, S. A., Biswas, A., Brouns, S. J., Brown, C. M., & Fineran, P. C. (2016). Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat Commun, 7*, 12853. doi:10.1038/ncomms12853

Staals, R. H., Zhu, Y., Taylor, D. W., Kornfeld, J. E., Sharma, K., Barendregt, A., . . . van der Oost, J. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. *Mol Cell, 56*(4), 518-530. doi:10.1016/j.molcel.2014.10.005

Staals, R. H. J., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D. W., van Duijn, E., . . . Shinkai, A. (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. *Mol Cell, 52*(1), 135-145. doi:10.1016/j.molcel.2013.09.013

Stanley, S. Y., & Maxwell, K. L. (2018). Phage-Encoded Anti-CRISPR Defenses. *Annu Rev Genet, 52*, 445-464. doi:10.1146/annurev-genet-120417-031321

Stella, S., Alcon, P., & Montoya, G. (2017). Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature, 546*(7659), 559-563. doi:10.1038/nature22398

Stern, A., Keren, L., Wurtzel, O., Amitai, G., & Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet, 26*(8), 335-340. doi:10.1016/j.tig.2010.05.008

Sternberg, S. H., Haurwitz, R. E., & Doudna, J. A. (2012). Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA, 18*(4), 661-672. doi:10.1261/rna.030882.111

Sternberg, S. H., LaFrance, B., Kaplan, M., & Doudna, J. A. (2015). Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature, 527*(7576), 110-113. doi:10.1038/nature15544

Steven, A. C., Trus, B. L., Maizel, J. V., Unser, M., Parry, D. A., Wall, J. S., . . . Studier, F. W. (1988). Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. *J Mol Biol, 200*(2), 351-365. doi:10.1016/0022-2836(88)90246-x

Stoebel, D. M., Free, A., & Dorman, C. J. (2008). Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology (Reading), 154*(Pt 9), 2533-2545. doi:10.1099/mic.0.2008/020693-0

Stratmann, T., Pul, U., Wurm, R., Wagner, R., & Schnetz, K. (2012). RcsB-BglJ activates the Escherichia coli leuO gene, encoding an H-NS antagonist and pleiotropic regulator of virulence determinants. *Mol Microbiol, 83*(6), 1109-1123. doi:10.1111/j.1365-2958.2012.07993.x

Strecker, J., Demircioglu, F. E., Li, D., Faure, G., Wilkinson, M. E., Gootenberg, J. S., . . . Zhang, F. (2022). RNA-activated protein cleavage with a CRISPR-associated endopeptidase. *Science, 378*(6622), 874-881. doi:10.1126/science.add7450

Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., & Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. *Science, 365*(6448), 48-53. doi:10.1126/science.aax9181

Sulakvelidze, A., Alavidze, Z., & Morris, J. G., Jr. (2001). Bacteriophage therapy. *Antimicrob Agents Chemother, 45*(3), 649-659. doi:10.1128/AAC.45.3.649-659.2001

Swarts, D. C., Jore, M. M., Westra, E. R., Zhu, Y., Janssen, J. H., Snijders, A. P., . . . van der Oost, J. (2014). DNA-guided DNA interference by a prokaryotic Argonaute. *Nature, 507*(7491), 258-261. doi:10.1038/nature12971

Swarts, D. C., Mosterd, C., van Passel, M. W., & Brouns, S. J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS One, 7*(4), e35888. doi:10.1371/journal.pone.0035888

Swarts, D. C., van der Oost, J., & Jinek, M. (2017). Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol Cell, 66*(2), 221-233 e224. doi:10.1016/j.molcel.2017.03.016

Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, C., Nwokeoji, A. O., Dickman, M. J., . . . Siksnys, V. (2014). Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus. *Mol Cell, 56*(4), 506-517. doi:10.1016/j.molcel.2014.09.027

Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., . . . Huttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proc Natl Acad Sci U S A, 99*(11), 7536-7541. doi:10.1073/pnas.112047299

Tang, T. H., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., . . . Huttenhofer, A. (2005). Identification of novel non-coding RNAs as potential antisense regulators

in the archaeon Sulfolobus solfataricus. *Mol Microbiol, 55*(2), 469-481. doi:10.1111/j.1365-2958.2004.04428.x

Taylor, D. W., Zhu, Y., Staals, R. H., Kornfeld, J. E., Shinkai, A., van der Oost, J., . . . Doudna, J. A. (2015). Structural biology. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science, 348*(6234), 581-585. doi:10.1126/science.aaa4535

Taylor, H. N., Laderman, E., Armbrust, M., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning Diverse Type IV Structures and Functions Within Class 1 CRISPR-Cas Systems. *Front Microbiol, 12*, 671522. doi:10.3389/fmicb.2021.671522

Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., & Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. *RNA Biol, 16*(10), 1438-1447. doi:10.1080/15476286.2019.1634965

Teklemariam, A. D., Al-Hindi, R. R., Qadri, I., Alharbi, M. G., Ramadan, W. S., Ayubu, J., . . . Harakeh, S. (2023). The Battle between Bacteria and Bacteriophages: A Conundrum to Their Immune System. *Antibiotics (Basel), 12*(2). doi:10.3390/antibiotics12020381

Tock, M. R., & Dryden, D. T. (2005). The biology of restriction and anti-restriction. *Curr Opin Microbiol, 8*(4), 466-472. doi:10.1016/j.mib.2005.06.003

Tong, B., Dong, H., Cui, Y., Jiang, P., Jin, Z., & Zhang, D. (2020). The Versatile Type V CRISPR Effectors and Their Application Prospects. *Front Cell Dev Biol, 8*, 622103. doi:10.3389/fcell.2020.622103

Tuminauskaite, D., Norkunaite, D., Fiodorovaite, M., Tumas, S., Songailiene, I., Tamulaitiene, G., & Sinkunas, T. (2020). DNA interference is controlled by R-loop length in a type I-F1 CRISPR-Cas system. *BMC Biol, 18*(1), 65. doi:10.1186/s12915-020-00799-z

Unterholzner, S. J., Poppenberger, B., & Rozhon, W. (2013). Toxin-antitoxin systems: Biology, identification, and application. *Mob Genet Elements, 3*(5), e26219. doi:10.4161/mge.26219

van Beljouw, S. P. B., Haagsma, A. C., Rodriguez-Molina, A., van den Berg, D. F., Vink, J. N. A., & Brouns, S. J. J. (2021). The gRAMP CRISPR-Cas effector is an RNA endonuclease complexed with a caspase-like peptidase. *Science, 373*(6561), 1349-1353. doi:10.1126/science.abk2718

van Houte, S., Ekroth, A. K., Broniewski, J. M., Chabas, H., Ashby, B., Bondy-Denomy, J., . . . Westra, E. R. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature, 532*(7599), 385-388. doi:10.1038/nature17436

Venkatesh, G. R., Kembou Koungni, F. C., Paukner, A., Stratmann, T., Blissenbach, B., & Schnetz, K. (2010). BglJ-RcsB heterodimers relieve repression of the Escherichia coli bgl operon by H-NS. *J Bacteriol, 192*(24), 6456-6464. doi:10.1128/JB.00807-10

Vlot, M., Houkes, J., Lochs, S. J. A., Swarts, D. C., Zheng, P., Kunne, T., . . . Brouns, S. J. J. (2018). Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes. *Nucleic Acids Res, 46*(2), 873-885. doi:10.1093/nar/gkx1264

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., & Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell, 163*(4), 840-853. doi:10.1016/j.cell.2015.10.008

Wang, R., & Li, H. (2012). The mysterious RAMP proteins and their roles in small RNA-based immunity. *Protein Sci, 21*(4), 463-470. doi:10.1002/pro.2044

Wang, R., Preamplume, G., Terns, M. P., Terns, R. M., & Li, H. (2011). Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure, 19*(2), 257-264. doi:10.1016/j.str.2010.11.014

Wang, X., Yu, G., Wen, Y., An, Q., Li, X., Liao, F., . . . Zhang, H. (2022). Target RNA-guided protease activity in type III-E CRISPR-Cas system. *Nucleic Acids Res, 50*(22), 12913-12923. doi:10.1093/nar/gkac1151

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*(4), 356-361. doi:10.1101/gad.257550.114

Weissman, J. L., Laljani, R. M. R., Fagan, W. F., & Johnson, P. L. F. (2019). Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy. *ISME J, 13*(10), 2589-2602. doi:10.1038/s41396-019-0411-2

Weissman, J. L., Stoltzfus, A., Westra, E. R., & Johnson, P. L. F. (2020). Avoidance of Self during CRISPR Immunization. *Trends Microbiol, 28*(7), 543-553. doi:10.1016/j.tim.2020.02.005

Westra, E. R., Nilges, B., van Erp, P. B., van der Oost, J., Dame, R. T., & Brouns, S. J. (2012). Cascade-mediated binding and bending of negatively supercoiled DNA. *RNA Biol, 9*(9), 1134-1138. doi:10.4161/rna.21410

Westra, E. R., Pul, U., Heidrich, N., Jore, M. M., Lundgren, M., Stratmann, T., . . . Brouns, S. J. (2010). H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. *Mol Microbiol, 77*(6), 1380-1393. doi:10.1111/j.1365-2958.2010.07315.x

Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., & Brouns, S. J. (2013). Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet, 9*(9), e1003742. doi:10.1371/journal.pgen.1003742

Wiedenheft, B., Sternberg, S. H., & Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature, 482*(7385), 331-338. doi:10.1038/nature10886

Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., . . . Doudna, J. A. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A, 108*(25), 10092-10097. doi:10.1073/pnas.1102716108

Wilkinson, M., Drabavicius, G., Silanskas, A., Gasiunas, G., Siksnys, V., & Wigley, D. B. (2019). Structure of the DNA-Bound Spacer Capture Complex of a Type II CRISPR-Cas System. *Mol Cell, 75*(1), 90-101 e105. doi:10.1016/j.molcel.2019.04.020

Wright, A. V., Nunez, J. K., & Doudna, J. A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell, 164*(1-2), 29-44. doi:10.1016/j.cell.2015.12.035

Xiao, Y., Luo, M., Hayes, R. P., Kim, J., Ng, S., Ding, F., . . . Ke, A. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell, 170*(1), 48-60 e11. doi:10.1016/j.cell.2017.06.012

Xiao, Y., Ng, S., Nam, K. H., & Ke, A. (2017). How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature, 550*(7674), 137-141. doi:10.1038/nature24020

Xu, Y., & Li, Z. (2020). CRISPR-Cas systems: Overview, innovations and applications in human disease research and gene therapy. *Comput Struct Biotechnol J, 18*, 2401-2415. doi:10.1016/j.csbj.2020.08.031

Xu, Z., Li, Y., Li, M., Xiang, H., & Yan, A. (2021). Harnessing the type I CRISPR-Cas systems for genome editing in prokaryotes. *Environ Microbiol, 23*(2), 542-558. doi:10.1111/1462-2920.15116

Xue, C., Whitis, N. R., & Sashital, D. G. (2016). Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Mol Cell, 64*(4), 826-834. doi:10.1016/j.molcel.2016.09.033

Yan, W. X., Chong, S., Zhang, H., Makarova, K. S., Koonin, E. V., Cheng, D. R., & Scott, D. A. (2018). Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. *Mol Cell, 70*(2), 327-339 e325. doi:10.1016/j.molcel.2018.02.028

Yan, W. X., Hunnewell, P., Alfonse, L. E., Carte, J. M., Keston-Smith, E., Sothiselvam, S., . . . Scott, D. A. (2019). Functionally diverse type V CRISPR-Cas systems. *Science, 363*(6422), 88-91. doi:10.1126/science.aav7271

Yang, C. D., Chen, Y. H., Huang, H. Y., Huang, H. D., & Tseng, C. P. (2014). CRP represses the CRISPR/Cas system in Escherichia coli: evidence that endogenous CRISPR spacers impede phage P1 replication. *Mol Microbiol, 92*(5), 1072-1091. doi:10.1111/mmi.12614

Yang, H., Gao, P., Rajashankar, K. R., & Patel, D. J. (2016). PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell, 167*(7), 1814-1828 e1812. doi:10.1016/j.cell.2016.11.053

Yang, H., & Patel, D. J. (2022). A type III-E CRISPR Craspase exhibiting RNase and protease activities. *Cell Res, 32*(12), 1044-1046. doi:10.1038/s41422-022-00739-2

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*(52), 20039-20053. doi:10.1074/jbc.RA119.009438

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*(12), 5569-5576. doi:10.1093/nar/gks216

Yu, G., Wang, X., Zhang, Y., An, Q., Wen, Y., Li, X., . . . Zhang, H. (2022). Structure and function of a bacterial type III-E CRISPR-Cas7-11 complex. *Nat Microbiol, 7*(12), 2078-2088. doi:10.1038/s41564-022-01256-z

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., . . . Zhang, F. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol, 35*(1), 31-34. doi:10.1038/nbt.3737

Zhang, Y., Heidrich, N., Ampattu, B. J., Gunderson, C. W., Seifert, H. S., Schoen, C., . . . Sontheimer, E. J. (2013). Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. *Mol Cell, 50*(4), 488-503. doi:10.1016/j.molcel.2013.05.001

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., . . . Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. *Nature, 515*(7525), 147-150. doi:10.1038/nature13733

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*(3), 102201. doi:10.1016/j.isci.2021.102201

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res, 31*(13), 3406-3415. doi:10.1093/nar/gkg595

Appendix

1. Table 1. Sequences of the various oligonucleotides used in the study:

Oligo name	Sequence (5'-3')	Description
Csb1-13S-R FP	AACCTGTACTTCCAATCCAATGCAATGATGCGCAAACTCACAG TACAAG	Amplification of gene encoding Csb1 from *B. animalis* with restriction sites SspI for 13S-R plasmid.
Csb1-13S-R RP	TTATCCACTTCCAATGTTATTATTATCATCTATCATCGTCGTCA GCAGCG	
BA_Csb2-1R FP	TACTTCCAATCCAATGCAATGACGTTCGCGATTCGTATCCAC	Amplification of gene encoding Csb2 from *B. animalis* with restriction sites SspI for 1R plasmid.
BA_Csb2-1R RP	TTATCCACTTCCAATGTTATTATCATCTCCATTTCGGTTGCCCT TT	
Csb3-pQE2 FP	ACATCACCATCACCATCACCATATG ATGAGCGTTCTGCGAATTCC	Amplification of gene encoding Csb3 from *B. animalis* with restriction sites NdeI and HindIII for pQE2 plasmid
Csb3-pQE2 RP	AGTCCAAGCTCAGCTAATTAAGCTT CTAGAGCCGTATGAGCTGCC	
Cascade-pQE2 FP	ACATCACCATCACCATCACCATATGATGACGTTCGCGATTCGT ATCCAC	Amplification of gene encoding Csb2 to assemble the Cascade operon from *B. animalis* with restriction sites NdeI and HindIII for pQE2 plasmid
Csb2-Csb1 overlap RP	ACTGTGAGTTTGCGCATCATGGTATATCTCCTTCTTAAATTATC ATCTCCATTTCGGTT	Amplification of gene encoding Csb2 and overlapping region of Csb1, to assemble the Cascade operon from *B. animalis* in pQE2 vector
Csb2-1 overlap FP	AACCGAAATGGAGATGATAATTTAAGAAGGAGATATACCATG ATGCGCAAACTCACAGT	Amplification of gene encoding Csb1 and overlapping region of Csb2, to assemble the Cascade operon from *B. animalis* in pQE2 vector
Csb1-Csb3 overlap RP	GGAATTCGCAGAACGCTCATGGTATATCTCCTTCTTAAATTAT CATCTATCATCGTCG	Amplification of gene encoding Csb1 and overlapping region of Csb3, to assemble the Cascade operon from *B. animalis* in pQE2 vector
Csb1-Csb3 overlap FP	ACGACGATGATAGATGATAATTTAAGAAGGAGATATACCATG AGCGTTCTGCGAATTCC	Amplification of gene encoding Csb3 and overlapping region of Csb1, to assemble the Cascade operon from *B.*

		animalis in pQE2 vector
Csb2 ΔN-terminal-1R FP	TACTTCCAATCCAATGCAACTCGTATGATTATCATTCCCGCGC	Amplification of gene encoding Csb2 ΔN-terminal from *B. animalis* with restriction sites SspI for 1R plasmid
Csb2 ΔN-terminal-1R RP	TTATCCACTTCCAATGTTATTATCACCCTTTCGTTGTGAAACAG CTTTCG	
Csb2 ΔC-terminal-1R FP	TACTTCCAATCCAATGCAATGACGTTCGCGATTCGTATCC	Amplification of gene encoding Csb2 ΔC-terminal from *B. animalis* with restriction sites SspI for 1R plasmid
Csb2 ΔC-terminal-1R RP	TTATCCACTTCCAATGTTATTATCAAGTCCATGGCAGCTGTACT GC	
Csb1 E78A-13S-R RP	TCGGCTGATGTAGTCTGCCAAACGATTGGCCTG	Amplification of gene encoding various point mutants of Csb1 from *B. animalis* with restriction sites SspI for 13S-R plasmid.
Csb1 H122A-13S-R RP	GTCCACCGCACGAGCTGGCAATTGAACATCATCGAAATAC	
Csb1 R123A-13S-R RP	GCGTCCACCGCAGCATGTGGCAATTGAACATCATCG	
Csb1 R179A-13S-R RP	TTGAAGCTGGCGGGAATAGCCAGTTGGTTCTTGTTTCG	
Csb1 R320A-13S-R FP	AAGCAATGCGGAATTGCATCTAGCGGAAAACTGCTTCCTTG	
Csb2 Y14A-1R RP	ATATTCGCTTGCTCCTTGAGCGGAAGCAAGAAGGAAGTGG	Amplification of gene encoding various point mutants of Csb2 from *B. animalis* with restriction sites SspI for 1R plasmid.
Csb2 E24A-1R RP	TGGGAGTTGGAAAAGATGCCTTCTCCCCATATTCG	
Csb2 R31A-1R RP	GATACCATTGCCTGATATAAAGCCATGGGAGTTGGAAAAGAT TC	
Csb2 E65A-1R RP	TCCGGCGGATTCGATGCCAACCATTCGAGTG	
Csb2 Y90A-1R RP	CTTTGTCTGCTTTACGCCTAGCAGCAATGGCATTATGGGAGG	
Csb2 Y236A-1R RP	GCTTGTTTCTGAGGGCTGGCGGTACTGCGATGTACACA	
Csb2 E410A-1R FP	TATGCCATTTTCGGTCAATGCAACACAAGCTGAAAAAAGCC	
Csb2 H520A-1R FP	CGGGCAGAGCAGGGCCTTTGGCGGGGGC	
Csb2 D530A-1R FP	GCTTACTCATTCCTATGGCTTGTCCCGAAAGCTGTT	
Array_CI_CSO-pOSIP-CT FP	GAATTCGAGCTCGGTACCCGGGGGATCCTAATACGACTCACTATAGG GCAGCATAAAGCGGGGAATCTCTCGC	For integration of the CSO CRISPR array from *B. animalis* using clonetegration into P21 *attB* site of IYB5101.
Array_CI_CSO-pOSIP-CT RP	GCGCCATGCATCTCGAGGCATGCCTGCAGGTCACCTTGGATTTCTAAC CATGCC	
Array_CI_CAO-pOSIP-CT FP	GAATTCGAGCTCGGTACCCGGGGGATCCTAATACGACTCACTATAGG GGTCACCTTGGATTTCTAACCATGCC	For integration of the CAO CRISPR array from *B. animalis* using clonetegration into P21 *attB* site of IYB5101.
Array_CI_CAO-pOSIP-CT RP	GCGCCATGCATCTCGAGGCATGCCTGCAGCAGCATAAAGCGG GGAATCTCTCGC	
TTT_Target-13S-R FP	GTGTGAAGCTTGCATGCCTGCAGGTCGACTCTAGATTTGACAG TGCGAACACAGTGCAGT	Amplification of TTT PAM containing Target sequence from *B. animalis* with restriction
TTT_Target-13S-R RP	CACACGAATTCGAGCTCGGTACCCGGCGCGATCGTCACCGACT GCACTGTGTTCGCACTG	

		sites PstI and KpnI for 13S-R plasmid.
Cas3-1R FP	TACTTCCAATCCAATGCAATGGAGATGAATGCAACAACCCCAA	Amplification of gene encoding Cas3 from *B. animalis* with restriction sites SspI for 1R plasmid.
Cas3-1R RP	TTATCCACTTCCAATGTTATTATCATCGTCCTTCCATGGATATCGTC	
T7 Promoter FP	GAAATTAATACGACTCACTATAGG	For annealing the T7 promoter to the DNA constructs of various mutant RNA substrates to be synthesised via *in vitro* RNA transcription
TV-Mut RNA RP	ATCTCCGAAGTCTCGGCTTCGGAGCTTCTAACTCCCCCTATAGTGAGTCGTATTAATTTC	For annealing the DNA template to the T7 promoter region for synthesis of mutant RNA substrates via *in vitro* RNA transcription (RNA constructs are described in table S2)
TS-Mut RNA RP	ATCTCCGAAGTCTCGGCTTCGGAGCTTCGCCAGAAACCTATAGTGAGTCGTATTAATTTC	
Shr-Mut RNA RP	ATCTCCGAAGTCTCGGCTTCGGAGCTTCCCTATAGTGAGTCGTATTAATTTC	
5'TV RNA RP	ATCTCCGAAGTCTCGGCTTCGGACGAAGTATGTCCCCCTATAGTGAGTCGTATTAATTTC	
5'TS RNA RP	ATCTCCGAAGTCTCGGCTTCGGACGAAGGCTGGAAACCTATAGTGAGTCGTATTAATTTC	
Δ5' RNA RP	ATCTCCGAAGTCTCGGCTTCGGAGCCTATAGTGAGTCGTATTAATTTC	
Stem'' RNA RP	ATCTCCGAAGTCTCGGCTTCTTTTCTTCATTGAGGGCCTATAGTGAGTCGTATTAATTTC	
CR_Array	GCTCTTCCCCTGTAGATTAATTAAGCGGCCGCTAATACGACTCACTATAGGGCCCTCAATGAAGCTCCGAAGCCGAGACTTCGGAGATGACAGTGCGAACACAGTGCAGTCGGTGACGATCGCGCCCTCAATGAAGCTCCGAAGCCGAGACTTCGGAGATGACAGTGCGAACACAGTGCAGTCGGTGACGATCGCGCCCTCAATGAAGCTCCGAAGCCGAGACTTCGGAGATGACAGTGCGAACACAGTGCAGTCGGTGACGATCGCGCCCTCAATGAAGCTCCGAAGCCGAGACTTCGGAGATGACAGTGCGAACACAGTGCAGTCGGTGACGATCGCGCCCTCAATGAAGCTCCGAAGCCGAGACTTCGGAGATCCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGGGTACCACGCGTGCGCGCTGATCCGGAAGAGC	DNA sequence encoding type I-G CRISPR array harbouring 7 repeat units interspersed by 6 identical spacer units from *B. animalis* flanking with T7 promoter and T7 terminator, followed by 13S-R vector backbone and BspQ I recognition sites.
GS_Csb2-1R FP	TACTTCCAATCCAATGCAATGAGTATGTACTTTGTATTGACAATCGCC	Amplification of gene encoding Csb2 from *G. sulfurreducens* with restriction sites SspI for 1R plasmid.
GS_Csb2-1R RP	TTATCCACTTCCAATGTTATTATCAGTCAACGGGGGCGAAAAGACC	
KT_Csb2-1R FP	TACTTCCAATCCAATGCAATGGAAGAGCGGTGGGCCA	Amplification of gene encoding Csb2 from *K. tusciae* with restriction sites SspI for 1R plasmid.,
KT_Csb2-1R RP	TTATCCACTTCCAATGTTATTATTAAAACCCGGATTCACCGTTGCG	

2. Table 2. Sequences of the various RNA constructs used in the study:

Name	Sequence (5'-3')	Description
WT CSO RNA	AUCUCCGAAGUCUCGGCUUCGGAGCUUCAUUGAGGG	5'/3' 6-FAM labelled CSO CRISPR repeat RNA from *B. animalis*
WT CAO RNA	CCCUCAAUGAAGCUCCGAAGCCGAGACUUCGGAGAU	5'/3' 6-FAM labelled CAO CRISPR repeat RNA from *B. animalis*
TV-Mut RNA	**GGGAGUUA**GAAGCUCCGAAGCCGAGACUUCGGAGAU	Mutant CAO RNA construct where transversion mutations have been introduced in the 8 nt from the 5'end (in bold)
TS-Mut RNA	**UUUCUGGC**GAAGCUCCGAAGCCGAGACUUCGGAGAU	Mutant CAO RNA construct where transition mutations have been introduced in the 8 nt from the 5'end (in bold)
Shr-Mut RNA	GAAGCUCCGAAGCCGAGACUUCGGAGAU	Mutant CAO RNA construct where 8 nt from the 5'end are deleted.
5'TV-Mut RNA	GGGA**CA**UACUUCGUCCGAAGCCGAGACUUCGGAGAU	Mutant CAO RNA construct where transversion mutations have been introduced in the 12 nt from the 5'end, keeping the site of cleavage (in bold) intact.
5'TS-Mut RNA	UUUC**CA**GCCUUCGUCCGAAGCCGAGACUUCGGAGAU	Mutant CAO RNA construct where transition mutations have been introduced in the 12 nt from the 5'end, keeping the site of cleavage (in bold) intact.
Δ5' RNA	CUCCGAAGCCGAGACUUCGGAGAU	Mutant CAO RNA construct where 12 nt from the 5'end are deleted
Stem'' RNA	CCCUCAAUGAAGAAAAGAAGCCGAGACUUCGGAGAU	Mutant CAO RNA construct where the stem region is deleted keeping the overall length of the repeat RNA intact.

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

Strain	Genotype	Source
E. coli IYB5101	F-Δ(*araD-araB*)567 Δ*lacZ4787* (::rrnB-3) λ- rph-1 Δ(*rhaD-rhaB*)568 *hsdR514 araB*::T7-RNAp-*tetA*, TetR	Kind gift from Prof. Udi Qimron
E. coli BL21(DE3)	F- *ompT hsdsB*(rB-, mB-) *gal dcm* λ(DE3)	NEB
E. coli TOP10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80*lacZ*ΔM15 Δ *lac*X74 *recA1 araD139* Δ(*araleu*)7697 *galU galK rpsL endA1 nupG*, Str^R	Invitrogen
E. coli IG-CR_CSO	IYB5101 P21:: *I-G CSO array*, Cam^R	This study
E. coli IG-CR_CAO	IYB5101 P21:: *I-G CAO array*, Cam^R	This study

4. Table 4. List of plasmids used in this study:

Plasmid name	Description	Source
pOSIP-CT	*ori* Rγ, *ori* pUC, Cam^R, *attP* P21, *ccdB*, λ (*cI857*) encodes P21 integrase under the control of λ promoter (λ pR).	Addgene #45981
pUC19	*ori* PBR322, Amp^R, for gene insertion under the control of lac promoter.	NEB
pQE2	*ori* ColE1, Amp^R, expresses gene of interest to synthesize N-terminal 6xHis tagged protein under the IPTG inducible T5 promoter.	
pET StrepII TEV LIC cloning vector (p1R)	*ori* pMB1, Kan^R, *lacI*, expresses the gene of interest to synthesize N-terminal StrepII tagged protein under the control IPTG inducible promoter (PT7lac).	Addgene #29664 (Scott Gradia)
pET StrepII TEV co-transformation cloning vector (p13SR) / pNT	*ori* CloDF13, Spc^R, *lacI*, expresses the gene of interest to synthesize N-terminal StrepII tagged protein under the control IPTG inducible promoter (PT7lac).	Addgene #48328 (Scott Gradia)
pCsb2/I-G	*csb2* gene from *B. animalis* encoding N-terminal Strep tagged protein inserted in p1R plasmid. Kanamycin resistance plasmid (Kan^R)	This study
pGS_Csb2/I-G	*csb2* gene from *G. sulfurreducens* encoding N-terminal Strep tagged protein inserted in p1R plasmid. Kanamycin resistance plasmid (Kan^R)	This study
pKT_Csb2/I-G	*csb2* gene from *K. tusciae* encoding N-terminal Strep tagged protein inserted in p1R plasmid. Kanamycin resistance plasmid (Kan^R)	This study

pCsb2 Y14A	Csb2/I-G with alanine mutation at Y14 amino acid, encoding N-terminal Strep-II tagged protein expressed using p1R plasmid. Kan^R	This study
pCsb2 E24A	Csb2/I-G with alanine mutation at E24 amino acid, encoding N-terminal Strep-II tagged protein expressed using p1R plasmid. Kan^R	This study
pCsb2 R31A	Csb2/I-G with alanine mutation at R31 amino acid, encoding N-terminal Strep-II tagged protein expressed using p1R plasmid. Kan^R	This study
pCsb2 E65A	Csb2/I-G with alanine mutation at E65 amino acid, encoding N-terminal Strep-II tagged protein expressed using p1R plasmid. Kan^R	This study
pCsb2 Y90A	Csb2/I-G with alanine mutation at Y90 amino acid, encoding N-terminal Strep-II tagged protein expressed using p1R plasmid. Kan^R	This study
pCsb2 Y236A	Csb2/I-G with alanine mutation at Y236 amino acid, encoding N-terminal Strep-II tagged protein expressed using p1R plasmid. Kan^R	This study
pCsb2 E410A	Csb2/I-G with alanine mutation at E410 amino acid, encoding N-terminal Strep-II tagged protein expressed using p1R plasmid. Kan^R	This study
pCsb2 H520A	Csb2/I-G with alanine mutation at H520 amino acid, encoding N-terminal Strep-II tagged protein expressed using p1R plasmid. Kan^R	This study
pCsb2 D530A	Csb2/I-G with alanine mutation at D530 amino acid, encoding N-terminal Strep-II tagged protein expressed using p1R plasmid. Kan^R	This study
pCsb2 ΔC	Cas5-like domain of Csb2/I-G without the C-terminal regions (from 251 – 545 amino acids), encoding N-terminal Strep-II tagged protein expressed using p1R plasmid. Kan^R	This study
pCsb2 ΔN	Cas6-like domain of Csb2/I-G without the N-terminal regions (from 1– 250 amino acids), encoding N-terminal Strep-II tagged protein expressed using p1R plasmid. Kan^R	This study
pCsb1/I-G	*csb1* gene from *B. animalis* encoding N-terminal Strep tagged protein inserted in p13S-R plasmid. Spectinomycin resistance plasmid (Spc^R)	This study
pCsb1 E78A	Csb1/I-G with alanine mutation at E78 amino acid, encoding N-terminal Strep-II tagged protein expressed using p13S-R plasmid. Spc^R	This study
pCsb1 H122A	Csb1/I-G with alanine mutation at H122 amino acid, encoding N-terminal Strep-II tagged protein expressed using p13S-R plasmid. Spc^R	This study
pCsb1 R123A	Csb1/I-G with alanine mutation at R123 amino acid, encoding N-terminal Strep-II tagged protein expressed using p13S-R plasmid. Spc^R	This study
pCsb1 R179A	Csb1/I-G with alanine mutation at R179 amino acid, encoding N-terminal Strep-II tagged protein expressed using p13S-R plasmid. Spc^R	This study
pCsb1 R320A	Csb1/I-G with alanine mutation at R320 amino acid, encoding N-terminal Strep-II tagged protein expressed using p13S-R plasmid. Spc^R	This study
pCsb3/I-G	*csb3* gene from *B. animalis* encoding N-terminal 6xHis tagged protein inserted in pQE2 plasmid. Ampilcilin resistance plasmid (Amp^R)	This study

pCascade/I-G	Cascade operon (Csb2-Csb1-Csb3) encoding N-terminal 6xHis tagged protein inserted in pQE2 plasmid. Amp^R	This study
pC1H_Cascade	Cascade operon (Csb2-Csb1-Csb3) with mutation at Csb1 H122 amino acid, encoding N-terminal 6xHis tagged protein inserted in pQE2 plasmid. Amp^R	This study
pC2H_Cascade	Cascade operon (Csb2-Csb1-Csb3) with mutation at Csb2 H520 amino acid, encoding N-terminal 6xHis tagged protein inserted in pQE2 plasmid. Amp^R	This study
pC1HR_Cascade	Cascade operon (Csb2-Csb1-Csb3) with mutation at Csb1 HR122/3 amino acid, encoding N-terminal 6xHis tagged protein inserted in pQE2 plasmid. Amp^R	This study
pC2HD_Cascade	Cascade operon (Csb2-Csb1-Csb3) with mutation at Csb2 H520 and D530 amino acid, encoding N-terminal 6xHis tagged protein inserted in pQE2 plasmid. Amp^R	This study
pC1H_C2H Cascade	Cascade operon (Csb2-Csb1-Csb3) with mutation at Csb1 H122 and Csb2 H520 amino acids, encoding N-terminal 6xHis tagged protein inserted in pQE2 plasmid. Amp^R	This study
pCR_Array/I-G	I-G CRISPR array harbouring 7 repeat units interspersed by 6 identical spacer units from *B. animalis*, inserted in 13S-R plasmid, Spc^R	This study
pCas3/I-G	*cas3* gene from *B. animalis* encoding N-terminal Strep tagged protein inserted in p1R plasmid. Kanamycin resistance plasmid (Kan^R)	This study
pT/I-G	Target DNA sequence with TTT PAM inserted in 13S-R plasmid. Spc^R	This study

List of Publications

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. Preprint at ***BioRxiv*** 2022.08.19.504415 DOI: https://doi.org/10.1101/2022.08.19.504415 (manuscript under revision)

List of workshops and posters presented

Posters

- **Sunanda Chhetry** and B. Anand, "Unravelling the Maturation of CRISPR RNA in an atypical CRISPR-Cas System". International lecture course on RNA binding proteins: From RNA binding to condensation and Aggregation, 08 – 11 February 2022, National Center for Cell Science (Pune), (Virtual mode)
- **Sunanda Chhetry** and B. Anand, "Exploring CRISPR Maturation in an atypical CRISPR-Cas System" 10th RNA Group Meeting 2-4 May. 2019, RGCB, Thiruvananthapuram, Kerala, India.
- **Sunanda Chhetry** and B. Anand, "Deciphering the mechanism of CRISPR Maturation in an atypical CRISPR-Cas System". Structure Assisted Development of Novel Therapeutics. 12-16 February 2019, Regional Center for Biotechnology, Faridabad, Haryana, India.
- **Sunanda Chhetry** and B. Anand, "Investigating the mechanism of CRISPR Maturation in Type I-U CRISPR-Cas". 9th RNA Group Meeting. 26-28 October 2017, Banaras Hindu University, Varanasi, India.

Short Talks

- **Sunanda Chhetry** and B. Anand, "Investigating the Maturation of CRISPR RNA in an atypical CRISPR-Cas system". Research and Industrial Conclave 2022, 20 – 23 January 2022, Indian Institute of Technology, Guwahati (Virtual mode)
- **Sunanda Chhetry** and B. Anand, "Unravelling the CRISPR RNA maturation in an atypical CRISPR-Cas system". 1st National conference on CRISPR/Cas: From Biology to Technology, 25- 27 November 2021, SRM University (Andhra Pradesh) and Institute of Bioinformatics and Applied Biotechnology (Bengaluru), (Virtual Mode)

Workshops:

- Structure Assisted Development of Novel Therapeutics. 12-16 February 2019, Regional Center for Biotechnology, Faridabad, Haryana, India.