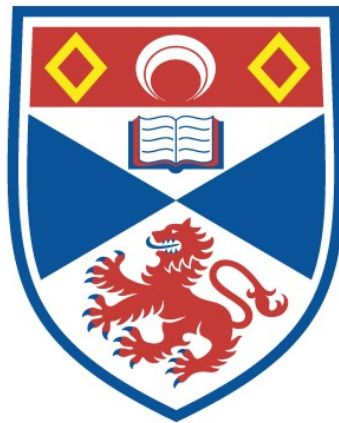# Novel signalling pathways in type III CRISPR defence systems

Haotian Chi

A thesis submitted for the degree of PhD
at the
University of St Andrews

2024

Full metadata for this item is available in
St Andrews Research Repository
at:
https://research-repository.st-andrews.ac.uk/

Identifier to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/sta/799

Candidate's declaration

I, Haotian Chi, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 40,747 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2020.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date   29/02/2024   Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date   29/02/2024   Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Haotian Chi, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Electronic copy

Embargo on all of electronic copy for a period of one year on the following ground(s):

• Publication would preclude further publication.

Supporting statement for electronic embargo request

It is necessary to delay access to a thesis until after publication of results.

Title and Abstract

• I agree to the title and abstract being published.

Date   29/02/2024   Signature of candidate

Date   29/02/2024   Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Haotian Chi, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date   29/02/2024   Signature of candidate

Permission for publication of underpinning research data or digital outputs

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

Embargo on all of electronic files for a period of one year on the following ground(s):

•   Publication would preclude further publication

Supporting statement for embargo request

It is necessary to delay access to a thesis until after publication of results.

Date   29/02/2024   Signature of candidate

Date   29/02/2024   Signature of supervisor

Contents

List of Figures

List of Tables

Abbreviations

$A_2{>}P$	2',3'-cyclic phosphate-terminated di-adenylate
$A_4{>}P$	2',3'-cyclic phosphate-terminated tetra-adenylate
2'3'-cGAMP	Cyclic 2'5'- and 3'5'-linked GMP-AMP
3'3'-cGAMP	Cyclic 3'5'-linked GMP-AMP
Abi	Abortive infection
Acr	Anti-CRISPR
Ago	Argonaute
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
Avs	Antiviral STAND
bGSDM	Bacterial gasdermin
BREX	Bacteriophage Exclusion
cA_2	Cyclic di-adenylate
cA_3	Cyclic tri-adenylate
cA_4	Cyclic tetra-adenylate
cA_6	Cyclic hexa-adenylate
cAMP	3',5'-cyclic adenosine monophosphate
Can1	CRISPR ancillary nuclease 1
Can2	CRISPR ancillary nuclease 2
Card1	cOA-activated single-stranded RNase and single-stranded DNase 1
CARF	CRISPR associated Rossman fold
Cas	CRISPR-associated system
Cascade	CRISPR-associated complex for antiviral defence
CBASS	Cyclic-oligonucleotide-based anti-phage signalling systems
CD-NTase	cGAS/DncV-like Nucleotidyltransferase
cGAS	Cyclic GMP-AMP synthase
cGAS	Cyclic GMP-AMP synthase
Cmr	Cas module RAMP
cOA	Cyclic oligoadenylates
CRISPR	Clustered regularly interspaced short palindromic repeats
Crn1	CRISPR-associated ring nuclease 1
Crn2	CRISPR-associated ring nuclease 2
Crn3	CRISPR-associated ring nuclease 3
crRNA	CRISPR RNA
Csm	Cas subtype Mtube
Csx	Cardiac-specific homeobox
CTR	Cognate Target RNA
DDM	n-dodecyl β-D-maltoside
DSRs	Defence-associated sirtuins;
Ec48	Retron found in Escherichia coli whose reverse transcribed DNA segment is 48 nt long
ECF	ExtraCytoplasmic Function
EMSA	Electrophoretic Mobility Shift Assay

FAM	Fluorescein amidites
ghmC	Glucosyl-hydroxymethylated cytosines
HD	Histidine-aspartate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEPN	Higher eukaryotes and prokaryotes nucleotide-binding
HPLC	High-performance liquid chromatography
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl-β-D-1-thiogalactoside
LC-MS/MS	Liquid Chromatography with tandem mass spectrometry
Lpa	Late Promoter Activating protein
MCS	Multiple cloning site
MGEs	Mobile genetic elements
msDNA	Multicopy single-stranded DNA
MTase	Methyltransferase
NAD$^+$	Nicotinamide adenine dinucleotide
NLRs	Nucleotide-binding oligomerization domain-like receptors
NTR	Non-Cognate Target RNA
Ocr	Overcome Classical Restriction
PAM	Protospacer-adjacent motif
PARIS	Phage anti-restriction-induced systems
Pgl	Phage growth limitation
pLDDT	Predicted local distance difference test
pre-crRNA	precursor crRNA
PT	Phosphorothioation
PYCSAR	Pyrimidine cyclase system for antiphage resistance
REase	Restriction endonuclease
RHH	ribbon-helix-helix
RM	Restriction-modification
RMSD	Root-mean-square deviation
RNAi	RNA interference
RNAP	RNA polymerase
SAVED	Second messenger oligonucleotide or dinucleotide synthetase-associated and fused to various effector domains
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
sgRNA	Synthetic guide RNA
siRNAs	Small interfering RNAs
ssDNA	Single-stranded DNA
STING	Stimulator of interferon genes
Stp	Short polypeptide
TA	Toxin-antitoxin
TBE	Tris/Borate/EDTA
TEV	Tobacco Etch Virus
TIR	Toll/interleukin-1 receptor
TLC	Thin layer chromatography
TM	Transmembrane

| tracrRNA | Trans-activating CRISPR RNA |
| v-cADPR | Variant cyclic ADP ribose |

Acknowledgements

Funding

Research Data/Digital Outputs access statement

Research data underpinning this thesis are available at:

DOI: 10.1038/s41586-023-06620-5

Chi, H., Hoikkala, V., Grüschow, S. *et al*. Antiviral type III CRISPR signalling via conjugation of ATP and SAM. *Nature* 622, 826–833 (2023). (**First author)**

DOI: 10.1038/s41586-022-05571-7

Rouillon, C., Schneberger, N., **Chi, H**. *et al*. Antiviral signalling by a cyclic nucleotide activated CRISPR protease. *Nature* 614, 168–174 (2023). (**Third author)**

ABSTRACT

CRISPR-Cas systems offer prokaryotes an adaptive defence mechanism, allowing them to respond to the invading nucleic acids. Type III CRISPR systems feature the capacity of synthesising cyclic oligoadenylate (cOA) species, which serve as second messengers to activate ancillary effectors, enhancing immune response. A diverse array of ancillary proteins is predicted to participate in cOA-mediated signalling for immunity enhancement. Nevertheless, the specific functions of many of these ancillary effectors have remained elusive. Here we have unravelled the workings of two novel type III-B CRISPR systems. The first system, from the human gut bacteria *Bacteroides fragilis* (BfrCmr), associates with an uncharacterised CorA family membrane protein and a NrN family phosphodiesterase. BfrCmr provides defence against mobile genetic elements when expressed in the heterologous host *E. coli*. A remarkable discovery was the identification of a novel signal molecule, S-adenosyl methionine (SAM)-AMP by conjugating ATP to SAM through a phosphodiester bond, when the BfrCmr system was activated. SAM-AMP in turn binds to the membrane protein CorA, presumably leading to membrane disruption and ultimately cell death. The cognate phosphodiesterase NrN or SAM lyase from *Clostridium botulinum* degrades SAM-AMP, offering two different means of regulating the signalling pathway.

The second type III CRISPR system investigated is associated with three ancillary proteins, including a Lon protease CalpL, extracytoplasmic function sigma factor CalpS and a toxin MazF homologue CalpT. CalpL consist of a SAVED sensor domain fused with a Lon protease effector domain. CalpL forms a tripartite complex with CalpS and CalpT. When SAVED domain bound to activator cA_4, CalpL oligomerises and specifically cleaves CalpT, resulting in the release of the sigma factor CalpS from the complex. This identification of a SAVED domain-containing protease that responses to cOA and triggers the transcriptional regulation provides insights into the sophisticated multi-layered defence mechanisms characterised in type III CRISPR signal-mediated immunity.

1 Introduction

1.1 The diversity of prokaryotic defence systems

Prokaryotes are surrounded by a multitude of mobile genetic elements (MGEs), such as bacteriophages, plasmids and transposable elements (Frost et al., 2005). These MGEs can transfer within or between genomes and have the capacity to confer either beneficial or negative effects on their bacterial hosts (Rankin et al., 2011). Plasmids, for example, carry resistance genes that benefit their hosts by enabling survival in the presence of antibiotics (Eberhard, 1990, Rankin et al., 2011), whereas virulent phage invade and lyse hosts to propagate themselves (Dy et al., 2014). In response to threats imposed by MGEs, cells have developed an arsenal of defence systems to defeat, control or inactivate different stages of MGEs invasion.

Phages are among the most abundant MGEs with their population (10^{31}) estimated to surpass that of bacteria (10^{29}) in the biosphere (Strange et al., 2021, Brussow and Hendrix, 2002). Furthermore, phage infections are responsible for causing 20 to 40 % of bacterial daily mortality (Suttle, 2007). Thus, phages represent a major ecological and evolutionary driver of bacterial defence system diversity. In turn, co-evolved phage counter-defences contribute to the diversity of anti-phage arsenals (Georjon and Bernheim, 2023).

The advancement of bioinformatic analysis and experimental studies has unveiled more than a hundred anti-phage defences, including early discovered restriction-modification (RM), clustered regularly interspaced short palindromic repeats–CRISPR-associated system (CRISPR-Cas) and recent emerging intracellular signal transduction-mediated defences such as cyclic-oligonucleotide-based anti-phage signalling systems (CBASS) (Kovall and Matthews, 1999, Makarova, 2015, Cohen, 2019). Notably, some bacterial defence systems share similar components or processes with the eukaryotic immune system. CBASS immunity, for instance, is systematically identified as a bacterial ancestor of the cyclic GMP-AMP synthase (cGAS) – stimulator of interferon genes (STING) pathway (Morehouse et al., 2020, Millman et al., 2020b). Recognising the structural and functional conservation of immune proteins between prokaryotes and eukaryotes, such as cGAS, STING, viperin and gasdermin, Wein and Sorek proposed an evolutionary scenario in which these proteins initialy evolved in prokaryotes as defence systems before being adopted as eukaryotic immune components during early eukaryogenesis (Wein and Sorek, 2022). Elucidating the diverse mechanisms of anti-phage defence systems will thus expand our understanding of immunity across the tree of life. The

following sections review some key bacterial defence systems, their links to eukaryotic immunity, and phage counter-defences strategies.

1.1.1 Phage infection

Bacteriophages (known as phages) were discovered independently by Frederick Twort and Félix d'Hérelle in 1915 and 1917, respectively (Twort, 1936, D'Herelle, 2007, Salmond and Fineran, 2015). The term "bacteriophage", literally meaning bacteria-eater, indicates its specific relationship with bacteria, as subsequently evidenced by increasing discoveries of diverse anti-phage mechanisms (D'Herelle, 2007, Georjon and Bernheim, 2023). Phages are considered as the most ubiquitous entities on earth and cause around 10^{25} infections per second (Dy et al., 2014, Fuhrman, 1999). Facing this intensive infectious pressure imposed by phage, bacteria have evolved multiple lines of defence that function in the different stages of phage infectious cycles.

The understanding of the phage life cycle provides insights into the mechanisms developed by the bacteria to prevent infection. Phages exhibit two distinct life cycles, a lytic cycle, and a lysogenic cycle (Fig. 1-1)(Salmond and Fineran, 2015, Dy et al., 2014, Clokie et al., 2011, Hampton et al., 2020, Stern and Sorek, 2011). Phage infection of host bacteria is initiated by interacting with specific receptors on the cell surface, a process known as adsorption. Phages subsequently puncture through the cell membrane and inject genomic material into the bacterial host. Thereafter, virulent phages exploit a lytic cycle by immediately hijacking host materials to produce their own viral progeny and ultimately killing the host cells to release progeny. In the lysogenic cycle, temperate phages are associated with hosts in a dormant state known as prophage, by integrating their genome into the host chromosome or existing in a free or plasmid-like state, potentially for thousands of generations. However, prophages may enter the lytic cycle and produce virions for release from the bacterium, often upon exposure to stress. Moreover, filamentous phages can cause a chronic infection, being secreted from cells without causing cell lysis (Rakonjac et al., 2011).

Bacteria have developed a range of strategies to protect themselves from phage infection. These include blocking adsorption and DNA injection at the start of phage injection, degrading phage nucleic acid and proteins to block phage replication and transcription or killing themselves to stop phage spreading. Here, I will mainly review defence mechanisms used after phage have injected their genome into hosts.
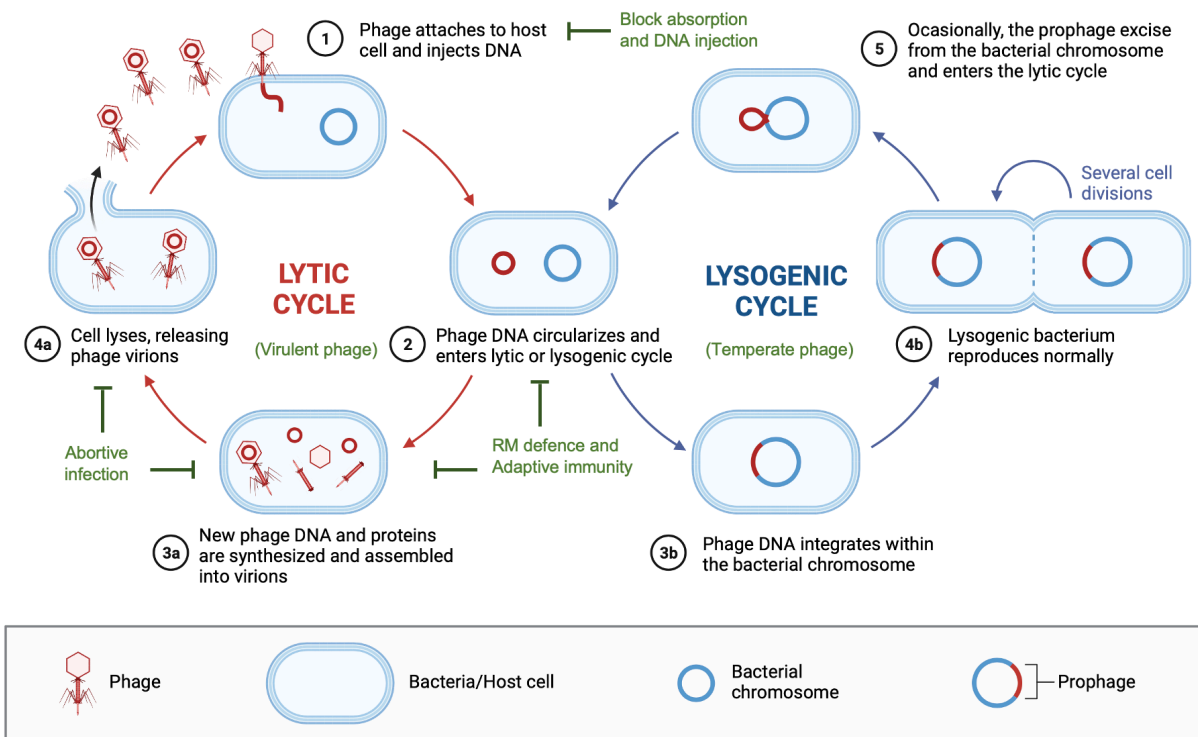
Figure 1-1 Phage life cycle

Phages initiate their life cycle by attaching to the cell surface and injecting their genetic materials into the host cell. Virulent phages proceed to the lytic cycle (left). They hijack the host machinery to facilitate their replication, transcription, and translation processes. Eventually, they release newly assembled progeny virions by lysing the cell. On the other hand, the temperate phages integrate their genomes into the bacterial chromosomes, becoming prophages, which can replicate alongside with host indefinitely (right). However, they can be induced to exit the bacterial chromosome and enter the lytic cycle. Prokaryotes have evolved numerous defence systems to target and inhibit the different stages of phage life cycle, contributing to their survival against phage infections. Abbreviations: RM, Restriction and Modification. Figure modified from the original made by Prof. Malcolm White through Biorender software.

1.1.2 Defence systems sensing MGEs

Recognition of injected phage nucleic acids is one example of an early response of antiviral defences. Two strategies are generally recruited to sense incoming MGEs. One is genome modification, allowing host to distinguish self from foreign non-self, like restriction-modification (RM) systems and RM-like systems. Another mechanism is to use nucleic acids as guides to recognise invading nucleic acids, including argonaute-dependent defence and CRISPR-Cas adaptive immunity.
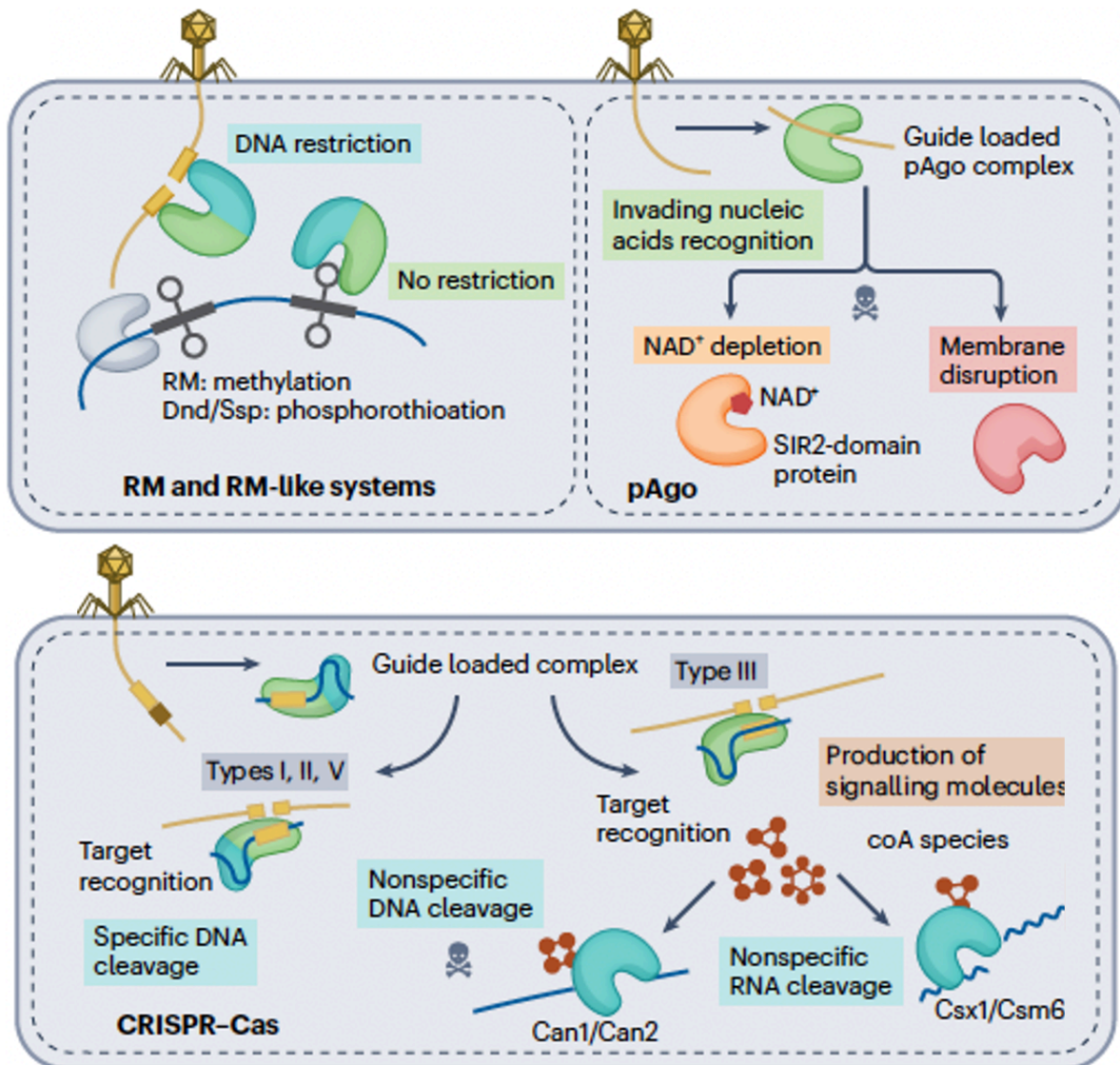
Figure 1-2 Defence systems sensing MGEs
Restriction and modification (RM) systems (left). Discrimination of self-DNA from foreign DNA through DNA modification (methylation for RM and phosphorothioate modification for DnD and Ssp) and subsequently cleave invading DNA (see **section 1.1.2.1 and 1.1.2.2** for details). CRISPR-Cas systems (middle). CRISPR systems recognise and target invading DNA or RNA using crRNA and type III CRISPR systems recruit signalling pathways for immunity (General introduction in **section 1.1.2.4**, more details in **section 1.2** and **1.3**). Prokaryotic Argonautes (pAgo) (right). Resemble eukaryotic RNA interference (RNAi) and use nucleic acid fragments as guides to sense invading nucleic acids and activate various effector for immunity (**section 1.1.2.3**). Abbreviations: CRISPR-Cas: clustered regularly interspersed short palindromic repeats-CRISPR-associated system; Can1/Can2, CRISPR ancillary nuclease 1 and 2; Csx1, cardiac-specific homeobox 1; Csm6, Cas subtype Mtube 6. cOA: cyclic oligoadenylate. Figure modified from the original of Georjon *et al*. (Georjon and Bernheim, 2023).

1.1.2.1 Restriction-Modification (RM) systems

Restriction-modification (RM) systems serve as a prokaryotic innate immune system, targeting invading nucleic acids by distinguishing self from non-self through the recognition of specific sequence motifs on viral DNA (as reviewed by (Loenen et al., 2014). These RM systems are variably distributed across 74 % of prokaryotic genomes (Oliveira et al., 2014) and classified into four families based on their subunit composition, recognition and cleavage site, and cofactor requirements (with a detailed review available in (Tock and Dryden, 2005)).

RM systems typically consist of two main components, a methyltransferase (MTase) that specifically methylates both strands of host DNA sequences and a restriction endonuclease (REase) that recognises and cleaves the same DNA sequence without methylation (Fig. 1-2) (Loenen et al., 2014, Tock and Dryden, 2005). Methylated DNA is thus recognised as self and protected, whereas foreign nonmethylated DNA is discriminated as nonself and subsequently destroyed. Type I RM enzyme complexes are encoded by three *h*ost *s*pecificity *d*eterminant (*hsd)* genes, a restriction (R), modification (M) and specificity recognition (S) gene (Fig. 1-2) (Dryden et al., 2001, Murray, 2000). The complex functions either as an REase towards unmethylated DNA or as an MTase if DNA is hemi-methylated in an ATP and Mg^{2+} dependent manner. Notably, the cleavage position is distant from the recognition site and cleavage occurs via an ATP-dependent DNA translocation (Dryden et al., 2001). Type II RM systems are the most prevalent (42%) (Oliveira et al., 2014) and are also extensively characterised due to their practical benefits (Loenen et al., 2014). These systems typically contain separate MTase and REase enzymes that share the same recognition sequence. REase cleave within or adjacent to specific DNA sequences in an Mg^{2+}-dependent manner (Pingoud and Jeltsch, 2001) and MTase usually functions as a monomer, methylating specific bases on both DNA strands (Sistla and Rao, 2004). Type III RM systems resemble type I (Dryden et al., 2001), with cleavage initiated by DNA translocation upon the formation of a hetero-oligomer with two R and two M subunits (Fig 1-2) (Janscak et al., 2001, Reich et al., 2004). Modification is independently conducted by M subunits on only one strand of DNA. Type IV RM systems operate differently, cleaving modified DNA sequences, including methylation, hydroxy-methylation, and glucosyl-hydroxyl-methylation (Roberts et al., 2003). One well-characterised type IV system is McrBC from *E. coli* K12, which specifically requires GTP for cleavage and DNA translocation processes (Raleigh and Wilson, 1986, Stewart et al., 2000).

Phages have evolved various strategies to evade bacterial RM systems. Mutations within phage genomes can lead to the removal of recognition sites, preventing cleavage (Kruger and Bickle,

1983). In some cases, phages can evade RM systems by simply reducing the number of recognition sites (Bickle and Kruger, 1993). For instance, the genomes of phage T3 and T7 are resistant to cleavage by *Eco*RII due to considerable distance between *Eco*RII binding sites on these phage genomes, preventing E*co*RII binding and cleavage (Bickle and Kruger, 1993, Kruger et al., 1988). In addition, modified bases are incorporated into phage genomes to avoid recognition by RM systems. For instance, *Bacillus subtilis* phages use hydroxymethyluracil (Warren, 1980), and T-even phages employ hydroxymethylcytosine (Kruger and Bickle, 1983). Phages can also disrupt bacterial RM processes by stimulating host MTase to modify their own DNA or by degrading host cofactors. Phage λ Ral protein enhances the modification activities of host MTases *Eco*K and *Eco*B to alleviate cognate restrictions (Zabeau et al., 1980). Phage T3, for instance, employs SAM hydrolase to degrade host RM systems cofactor SAM, thereby interfering with host MTase function (Studier and Movva, 1976).

Alternatively, phages have also developed mechanisms to directly inhibit RM enzyme activity. A well-known example is the overcome classical restriction (Ocr) protein, the first enzyme produced by phage T7 after phage DNA injection (Walkinshaw et al., 2002, Bandyopadhyay et al., 1985). Dimeric Ocr shares structural similarities with DNA and mimics DNA to interact with both *Eco*KI MTase and REase enzymes, effectively inhibiting their activities. Notably, the binding affinity of *Eco*KI enzymes for Ocr is 50-fold higher than that for DNA.

1.1.2.2 RM-like systems

Prokaryotes employ various DNA modifications, aside from methylation, to discriminate self-DNA from foreign nonself DNA (Weigele and Raleigh, 2016). One such modification is bacterial DNA phosphorothioation (PT), mediated by the DndACDE complex, which incorporates sulfur from cysteine into the DNA backbone (Fig. 1-2) (Wang et al., 2007, Eckstein, 2007, Wang et al., 2011, Xiong et al., 2015). The DndFGH complex acts as a restriction enzyme to recognise and degrade invading DNA that lacks PT modification (Gan et al., 2014, Chen et al., 2017). Similar phosphorothioation-based anti-phage defence system, known as Ssp PT systems, have also been characterised (Wang et al., 2021, Xiong et al., 2020). Another example is the 7-deazaguanine modification found in the *dpd* system, which converts guanine into 7-deazaguanine derivatives in the host DNA (Thiaville et al., 2016). Notably, phages have also evolved similar modification on their double-stranded DNA to escape from RM systems. Most recently, eight 7-deazaguanines derivatives were identified at guanine positions, including four previously uncharacterised modifications in phage genomes (Cui et al., 2023).

The Phage growth limitation (Pgl) system is another RM-like system encoded by a four-gene cluster in *Streptomyces coelicolor* A3(2) (Sumby and Smith, 2002). This cassette encodes four proteins: a predicted phosphatase (PglZ), a serine/threonine kinase (PglW), an adenine-specific DNA methyltransferase (PglX), and an ATPase (PglY). Pgl systems are predicted to confer anti-phage defence through three phases (Hoskisson et al., 2015). In uninfected cells, the Pgl proteins remain in a rest state, with the toxic activity of PglX inhibited by its interaction with PglZ. Upon detecting an infected phage from Pgl⁻ hosts, PglX becomes activated to methylate phage genomes. Subsequently, the modified phage progeny infects Pgl+ strains, triggering the activation of the restriction activity of Pgl systems, which is proposed to be mediated by PglW and PglX. However, further investigations are needed to confirm this model.

Another system that contains PglZ, known as the Bacteriophage Exclusion (BREX), was identified through the analysis of *pgl*-enriched gene cassettes in bacterial and archaeal genomic defence islands (Makarova et al., 2011b, Goldfarb, 2015). BREX systems employ methylation on the fifth position of a host non-palindromic motif, TAGGAG, to distinguish self from foreign DNA. Unmodified invading DNAs are subjected to BREX attack, which excludes them but does not degrade them, unlike RM systems (Goldfarb, 2015, Picton et al., 2021). Although the mechanism of BREX restriction remains unknown, recent studies have shown that Overcome Classical Restriction (Ocr) protein from phage T7 inhibits both methylation and restriction of BREX systems through specifically interactions with host methyltransferase (Isaev et al., 2020).

Defence ISland Associated with Restriction-Modification (DISARM) was identified as another RM-like system (Ofir et al., 2018). The DISARM cluster contains genes encoding a DNA methyltransferase (DrmM) and a helicase (DrmA), indicating a mode of action like RM systems, limiting phages that lack methylation. Recent cryo-EM structures of DrmA-DrmB complex suggest a potential phage targeting mechanism (Bravo et al., 2022). An unstructured trigger loop (TL) of DrmA, bound to the DNA binding surface of the DrmA-DrmB complex, enables the complex to distinguish between DNA structures in methylated host DNA and phage single-stranded DNA (ssDNA). Once the DrmA-DrmB complex is activated by loading with ssDNA, a nuclease DrmC from the same operon may be recruited to degrade foreign DNA. However, the detailed phage targeting mechanism requires further elucidations.

1.1.2.3 Prokaryotic Argonaute (pAgo) systems

Argonaute (Ago) proteins are ubiquitously present across all branches of life. These proteins are guided by small nucleic acids to target complementary DNA or RNA molecules, playing

essential roles in gene regulation or innate immunity. In eukaryotes, Ago proteins (eAgos) are the key components in the RNA interference (RNAi) pathways, which are well-documented in plants and animals (Peters and Meister, 2007, Vaucheret, 2008, Fang and Qi, 2016). All eAgos share structural and mechanistical similarities. They exclusively utilise small interfering RNAs (siRNAs) as guides to regulate or cleave target RNA (Bernstein et al., 2001, Hammond et al., 2001). Monomeric eAgos adopt a bi-lobed configuration, consisting of four conserved major domains (Kuhn and Joshua-Tor, 2013, Olina et al., 2018). The N- and C-terminal lobes are composed of N/PAZ domains and MID/PIWI domains, respectively. The N-domain is required for unwinding duplex RNAs and loading the guide strand to eAgos. The PAZ and MID domains functionally cooperate to protect siRNAs from degradation by binding to the 3' and 5' terminal bases of siRNAs, respectively (Wu et al., 2020). eAgos cleave target RNA by either using the active PIWI domain, which contains ribonuclease active sites, or by recruiting partner proteins when the PIWI domain is catalytically inactive (Wu et al., 2020, Pratt and MacRae, 2009).

Prokaryotic Ago (pAgo) proteins homologous to eAgos are found in approximately 9 % of sequenced bacteria and 32 % of archaea (Swarts et al., 2014). Unlike eAgos, pAgos exhibit a broader range of functions, utilising both RNA and DNA as guides to mediate either DNA or RNA interference (Fig. 1-2) (Hegge et al., 2018). pAgos are divided into long pAgos (about 40%, with long-A and long-B two subgroups), which contain six domains, four of them sharing conserved domain composition (N-PAZ-MID-PIWI) with eAgos, and short pAgos (near 60%) that only consist of the MID and PIWI domains (Ryazansky et al., 2018, Hegge et al., 2018, Kuzmenko et al., 2020). In long-A pAgos (94%), the PIWI domains are predicted to be catalytically active, as they contain key catalytic residues, whereas PIWI domains are inactive in all long-B and short pAgos due to mutations in endonuclease sites (Kuzmenko et al., 2020, Ryazansky et al., 2018). Long pAgos are the most extensively characterised pAgos, functioning not only as antiviral defence systems by degrading incoming DNA and RNA (Kuzmenko et al., 2020, Kropocheva et al., 2021), but also as key players in genome decatenation and homologous recombination (Jolly et al., 2020, Fu et al., 2019, Lee et al., 2021). Recent studies have also revealed that short pAgos provide antiviral defence through abortive infection, despite lacking endonuclease activity (Zeng et al., 2022, Koopal et al., 2022, Zaremba et al., 2022). Their immunity relies on additional effectors encoded adjacent to pAgos, including membrane proteins or NADases (Lopatina et al., 2020). The diversity of associated effectors adds another layer of complexity to pAgos-mediated immunity.

1.1.2.4 Adaptive immune systems

In prokaryotes, the only known adaptive immune system is the Clustered Regularly Interspaced Short Palindromic Repeats–CRISPR-associated systems (CRISPR-Cas) (Fig. 1-2) (Hille et al., 2018). CRISPR systems have the remarkable ability to memorise past invasions by acquiring and incorporating genetic material from invading MGE into the CRISPR array (Amitai and Sorek, 2016). They subsequently employ this acquired information to defend against future invasions by recognising and degrading the same incoming nucleic acids (Hille et al., 2018, Koonin and Makarova, 2019). These defence processes are generally considered as three stages: spacer acquisition (often referred to as "adaptation"), expression, and interference (Faure et al., 2019). CRISPR systems exhibit diversity in the composition and structure of their Cas proteins, as well as their modes of action. They are thus classified into two classes, six major types and more than 20 subtypes (Makarova et al., 2020b). Most importantly, their intrinsic sequence-specific nuclease activity has revolutionised the field of genome editing over the last decade. A more detailed review of CRISPR systems will be provided in sections 1.2 and 1.3.

1.1.3 Defence systems sensing phage proteins

Some defence systems are activated upon the detection of viral proteins, particularly when the first line of defence is less efficient or has been evaded by evolved phages (Georjon and Bernheim, 2023). Phages progress to stages of gene expression and protein synthesis if nothing breaks down viral genetic materials, providing hosts with limited time to interfere with the phage reproductive cycle (Salmond and Fineran, 2015). This explains why abortive infection (Abi) mechanisms are often triggered by defence systems recognising viral proteins (Georjon and Bernheim, 2023). These self-destructive strategies prevent the assembly of progeny virions, thereby enabling hosts to safeguard the surrounding bacterial community (Lopatina et al., 2020). Defence systems typically sense two types of viral proteins: functional proteins, and structural proteins.
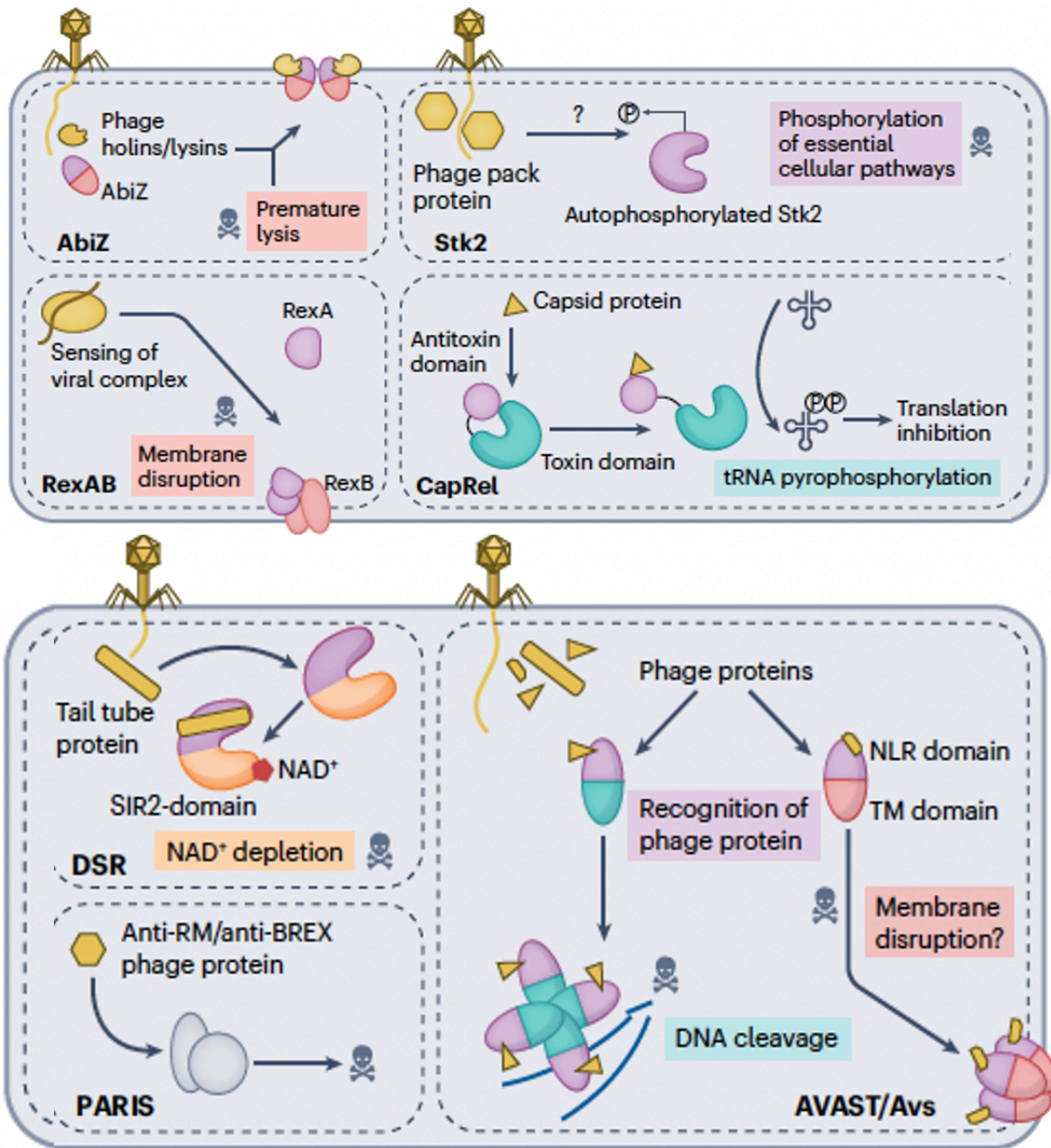
Figure 1-3 Defence systems sensing phage proteins
The AbiZ proteins sense and bind phage-encoded holins and lysins, leading to the cell lysis and interference with phage life cycles (Durmaz and Klaenhammer, 2007). The sensor RexA detects the phage infection and then activates the toxic effector RexB, resulting in cell death (Parma et al., 1992). The serine/threonine kinase Stk2 monitors phage encoded Pack proteins, leading to cell death (Depardieu et al., 2016). CapRel, DSR2 and Avs systems are activated by sensing phage structure proteins and PARIS systems are trigger by phage encoded overcome classical restriction (Ocr) proteins known to inhibit RM and BREX defence systems (**section 1.1.3**). Abbreviations: DSRs, defence-associated sirtuins; SIR2, sirtuin 2; NAD$^+$, nicotinamide adenine dinucleotide; RM, restriction and modification; BREX, bacteriophage exclusion; PARIS, phage anti-restriction-induced systems; Avs, antiviral STAND; NLR, nucleotide-binding oligomerization domain-like receptor; TM, transmembrane. Figure modified from the original of Georjon *et al.* (Georjon and Bernheim, 2023).

1.1.3.1 Sensing phage functional proteins

Prokaryotic Antiviral STAND (Avs) proteins are homologous to eukaryotic STAND NTPases, which play vital roles in immunity, cell signalling and cell death in animals, plants, and fungi (Gao et al., 2022, Koonin and Aravind, 2002, Leipe et al., 2004, Jones et al., 2016). Recent studies have unveiled that Avs systems in bacteria and archaea served as protective mechanisms against tailed phage infections by recognising conserved viral proteins (Fig. 1-3) (Gao et al., 2022, Gao et al., 2020). Prokaryotic Avs proteins share a conserved tripartite domain architecture with their eukaryotic counterparts, comprising a central NTPase domain, a C-terminal sensor domain and an N-terminal effector domain. Two Avs representatives, from *Salmonella enterica* (SeAvs3) and *Escherichia coli* (EcAvs4), have been observed to specifically recognise and bind to viral terminase subunits and portal proteins, respectively. These viral components are responsible for the DNA packaging of tailed phages. Cryo-electron microscopy (cryo-EM) structures of Avs enzymes forming complexes with their cognate viral proteins have shown that the binding of C-terminal sensors to target proteins leads to the tetramerisation of their ATPases and activation of N-terminal nuclease effectors (Gao et al., 2022). Furthermore, bioinformatic analyses have revealed that the *avs* genes are distributed in 4 to 5 % of sequenced prokaryotic genomes and the N-terminal effectors fused with Avs proteins display high diversity, including protease, sirtuins (SOR2) and Toll/interleukin-1 receptor (TIR) domains (Gao et al., 2022). Additionally, both *in vivo* and *in vitro* studies have demonstrated that phages often encode Avs inhibitors among their early expressed genes (Gao et al., 2022). However, further investigation will be necessary to fully elucidate mechanisms underlying these inhibitors.

Phage anti-restriction-induced systems (PARIS) are present in 5.2 % of sequenced prokaryotic genomes (Rousset et al., 2022). Recent studies have revealed that this system is triggered by anti-restriction proteins, leading to an abortive infection mechanism that maximises host population survival (Fig. 1-3) (Rousset et al., 2022). Anti-restriction proteins are encoded by phage and are known to inhibit RM and BREX systems (reviewed in 1.1.2.1). The activation of PARIS systems may indicate that the phages have successfully bypassed the hosts' first line of defence.

1.1.3.2 Sensing phage structural proteins

Defence-associated sirtuins (DSRs) systems have been recently documented as crucial components of innate immunity, as they recognise the phage tail tube proteins that form the structural framework of tailed phages (Fig. 1-3) (Gao et al., 2020, Garb et al., 2022). In the

case of DSR2 from *Bacillus subtilis*, it comprises an N-terminal sirtuin (SIR2) domain that becomes activated upon detection of phage tail proteins. This activation leads to the depletion of cellular nicotinamide adenine dinucleotide (NAD^+), effectively aborting phage propagation. Notably, anti-DSR2 proteins have also been identified from the DSR2-resistant phages, which can bind to DSR2, thereby inhibiting the DSR defence systems (Garb et al., 2022).

Another characterised defence system, CapRel, has been found to provide immunity upon detecting phage major capsid proteins (Fig. 1-3) (Zhang et al., 2022). CapRel[SJ46] from *E. Coli* functions through a toxin-antitoxin mechanism. The C-terminal domain of CapRel[SJ46] serves a dual role: antitoxin and phage infection sensor. Once monitoring and binding to viral capsid proteins, C-terminal domain alleviates inhibitions on the toxic N-terminal domain, which is activated to pyrophosphorylate tRNAs, thus effectively inhibiting viral translation.

The BilABCD system, which stands for bacterial ISG15-like system, encodes a prokaryotic defence system comprising E1, E2, Ubl (ubiquitin-like protein), and DUBs (deubiquitinases) (Millman et al., 2022). This system bears resemblance to eukaryotic ubiquitination and related pathways, which are essential in protein homeostasis and innate immunity (Cappadocia and Lima, 2018). Recent studies on Bil systems from *Ensifer aridi* TW10 have provided insight into the structures of E1: E2: Ubl complexes, revealing that enzymes E1 and E2 cooperate to conjugate Ubl to target proteins (Rouillon et al., 2023). DUBs are responsible for exposing the C-terminal glycine residue of Ubl, making it ready for conjugation. Simultaneously, Jens Hör and colleagues have demonstrated that the Bil system from *Collimonas sp*. OK412 specifically conjugates Ubl to the central tail fiber protein of phages Secphi27 and Secphi4 (Hör et al., 2023). This ubiquitination event either interferes with phage tail formation or prevents their infectivity due to modifications in the tail structure.

1.1.4 Defence systems sensing infection-induced cellular stress

Some defence systems are activated in response to cellular stress induced by phage infection, rather than in response to viral genetic materials or proteins. When phages infect host cells, they rapidly hijack host machinery and components to facilitate their own reproductive cycle. The sensor modules of defence systems are capable of detecting alterations in cellular processes, such as changes in host transcription or the activity of enzymes like RecBCD (Georjon and Bernheim, 2023).

1.1.4.1 Sensing the inhibition of host transcription

A type III toxin-antitoxin (TA) system known as ToxIN can monitor the inhibition of host transcription (Fig. 1-4) (Guegler and Laub, 2021). The antitoxin, *toxI*, is encoded as an RNA array featuring short tandem repeats, followed by the coding toxin *toxN* gene (Blower et al., 2011). ToxN, functioning as an endonuclease, specifically cleaves the antitoxin RNA array to generate mature *toxI*, which subsequently binds to ToxN, thereby inhibiting its toxic activity (Blower et al., 2012). Recent studies have observed that phage can lead to the shutdown of host transcription (Guegler and Laub, 2021). This event results in the rapid release of toxin ToxN, likely due to the fast turnover of antitoxin *toxI*. The liberated ToxN then directly targets and cleaves viral transcripts containing the GAAAU motif, effectively inhibiting phage particle production.

AvcID defence systems operate similar toxin-antitoxin mechanisms (Fig. 1-4) (Hsueh et al., 2022). The abundant non-coding sRNA, AvcI, functions as an antitoxin, effectively neutralising the toxic activity of the toxin AvcD. This system, as observed in *Vibrio cholerae*, provides immunity against T3 phage. Following infection, the toxin AvcD, known as a deoxycytidylate deaminase, is activated when it is released from its complex with AvcI. The liberation of AvcD is likely as a result of transcriptional inhibition, as observed in ToxIN systems (Guegler and Laub, 2021). AvcD then proceeds to deaminate host dCTP and dCMP to ultimately dUMP, presumably leading to impair phage DNA replication and virion production. However, the exact mechanisms underlying these processes require further investigation.

Simultaneously, Tal and colleagues made an intriguing discovery, identifying defensive dCTP deaminases in 2.5 % of the 38,167 analysed genomes. Additionally, dGTPase are abundant, present in around 25 % of more than 2,300 genomes (Fig. 1-4) (Tal et al., 2022). They observed that *E. coli* harbouring these nucleotide-depleting enzymes gained immunity against various types of phages, including T4, T5 and T7. However, phages that managed to overcome this mode of defence had mutated their genes responsible for shutting down host RNA polymerase (RNAP) transcription. This suggests that host nucleotide depletion-mediated defence might be triggered upon detecting the inhibition of host transcription.

1.1.4.2 Sensing changes in the activity of host enzymes

Bacterial Retrons have been recently characterised as antiviral defence systems (Fig. 1-4) (Millman et al., 2020a, Bobonis et al., 2022). The Retron system typically comprises a non-coding RNA (ncRNA), a reverse transcriptase (RT) and an effector protein (Millman et al., 2020a). The ncRNA and RT components are involved in the synthesis of multicopy single-

stranded DNA (msDNA), a distinctive branched RNA-DNA hybrid molecule that is covalently linked by a 2'-5' phosphodiester bond (Lampson et al., 2005). The effectors within the Ec48 system in *E. coli* share a similar transmembrane domain organisation with those found in the CBASS systems, which is predicted to disrupt membrane integrity, causing cell death (Cohen, 2019). Experimental validation of the Ec48 retron system has shown its effectiveness in protecting cells against phage infections through an abortive infection mechanism (Millman et al., 2020a). Further analysing of escaped phages let to the discovery of mutation in the Gam protein of phage λ and the gp5.9 protein of phage T7, in which both proteins serve as inhibition of RecBCD complex to interference host immunity. This study revealed that the Ec48 retron defence system is activated when the RecBCD is impaired, providing immunity via abortive infection. However, the precise mechanisms of how Retron msDNA sensing inhibition and in turn activating its cognate effector remain to be elucidated.
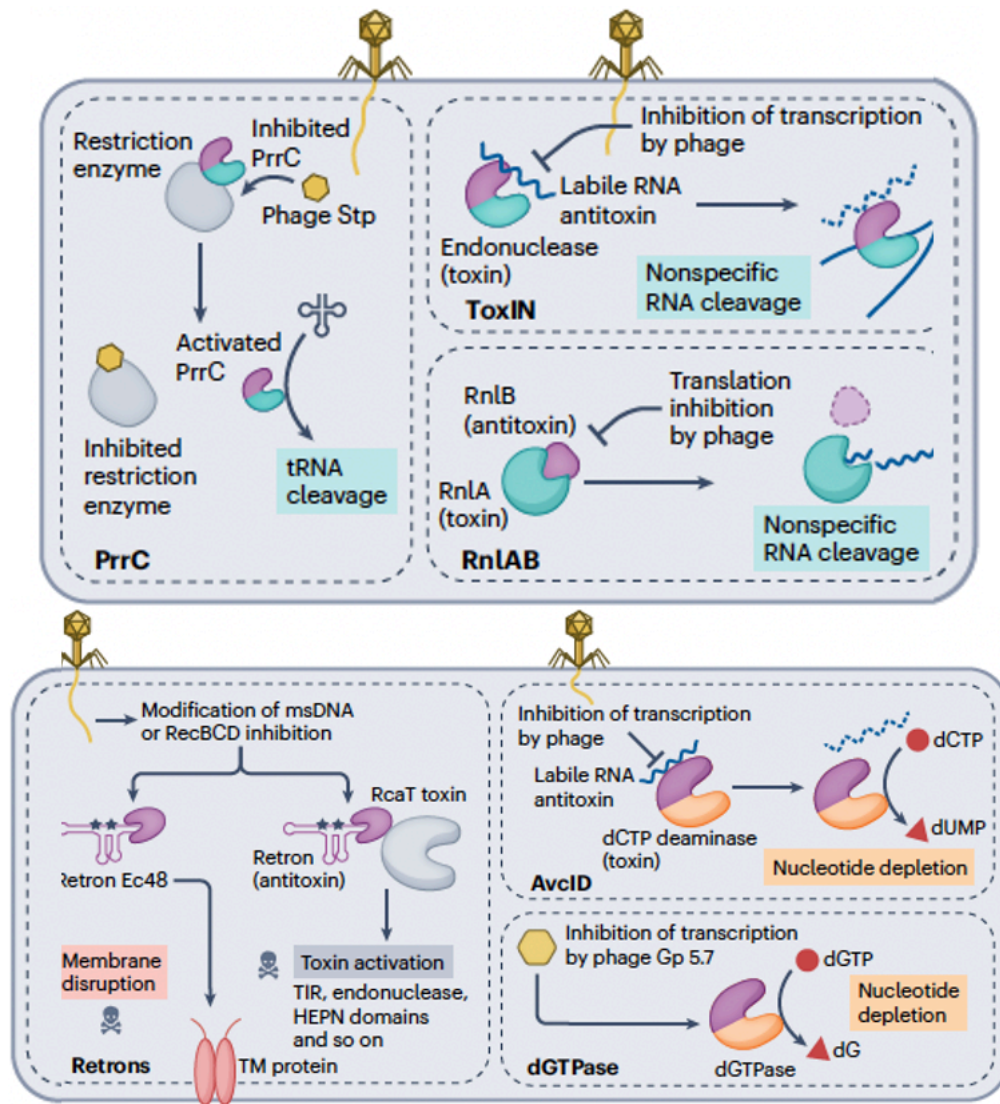
Figure 1-4 Defence systems sensing infection-induced cellular stress

The PrrC proteins remain silent by binding host restriction enzymes in the absence of phages and are released to impair protein synthesis upon detecting the phage encoded anti-restriction proteins (Kaufmann, 2000, Penner et al., 1995). In the type II toxin-antitoxin system RnlAB, antitoxin RnlB is degraded upon sensing the phage infection, releasing toxin RnlA, which mediates nonspecific RNA degradation and leads to cell death (Garcia-Rodriguez et al., 2020). ToxIN, AvcID and dGTPase defence systems are triggered by sensing the inhibition of transcription and Retron systems are activated by monitoring the inhibition of the RecBCD complex (**section 1.1.4**). Abbreviations: Stp, short polypeptide; msDNA, multicopy single-stranded DNA; Ec48, a retron from *E. coli* whose reverse transcribed DNA segment is 48 nt long; TIR, toll/interleukin-1 receptor; HEPN, higher eukaryotes and prokaryotes nucleotide-binding; TM, transmembrane. Figure modified from the original of Georjon *et al*. (Georjon and Bernheim, 2023).

1.1.5 Other defence systems

Defence systems encompass a wide array of mechanisms. These systems typically consist of a sensor module that can monitor the viral genomes, proteins, or the host cellular stresses. The activation of the sensor module subsequently triggers effector modules which interfere with every stage of the viral reproductive cycle. These sensor and effector modules can function as individual proteins, complexes, or they may be linked through signalling molecules. Moreover, some defence systems even employ antiviral molecules to impair phage replication.

1.1.5.1 Second messenger mediated defence systems

Type III CRISPR-Cas systems were the first identified prokaryotic defence systems generating signal molecules to activate effectors (Kazlauskiene et al., 2017, Niewoehner, 2017). Typically, the PALM domain of the enzymatic subunit Cas10 is allosterically activated upon detecting viral RNA, leading to the synthesis of a range of cyclic oligoadenylates (cOA), which are constituted by 2 to 6 AMP monomers with 3'-5' phosphodiester bonds. These cOA, in turn, activate various accessory proteins often found near type III CRISPR gene cassettes (Shmakov et al., 2018, Shah et al., 2019). More details are provided in Section 1.3.

CBASS (Cyclic-oligonucleotide-Based Anti-phage Signalling Systems) is another bacterial immune system that utilises cyclic nucleotides for signalling (Fig. 1-5) (Millman et al., 2020b). This system contains at least two key components. One is a signal synthetase CD-NTase (cGAS/DncV-like nucleotidyltransferase), which produces various signal molecules upon sensing invasion, including 2'3'-cGAMP, 3'3'-cGAMP, c-di-GMP, c-di-AMP, cUMP-AMP, 3'3'3'-cAAG and others (Whiteley et al., 2019). The other component is an CD-NTase associated proteins (Cap) that are activated by signal molecules to provide various antiphage immunities, such as DNA cleavage by endonuclease NucC (Lau, 2020), Cap4 (Chang et al., 2023) and Cap5 (Fatma et al., 2021), membrane disruption by phospholipase CapV (Cohen, 2019), or NAD^+ depletion by TIR-SAVED (Hogrel et al., 2022) and TIR-STING (Morehouse et al., 2022, Morehouse et al., 2020). These effectors often lead to cell death through an abortive infection mechanism, which explains why the majority of CBASS systems employ an

additional level of regulation. In type III CBASS systems, Cap7 (a peptide-binding HORMA domain protein) and Cap8 (a AAA+ ATPase TRIP13) are involved in tightly regulating CD-NTase activity (Ye et al., 2020a). Cap7 detects viral peptides and subsequently binds and actives CD-NTase to produce cyclic nucleotides, initiating signalling defence pathway. Cap8 disassembles Cap7 from the complex with CD-NTase, ensuring system robustness. Type II CBASS systems encode Cap2 and Cap3 ancillary proteins with the E1, E2 and JAB domains, respectively, related to eukaryotic ubiquitin machinery. In the *Enterobacter cloacae* CBASS system, Cap2 conjugates the CD-NTase to an unknown target, increasing the production of signal molecules, while Cap3 cleaves CD-NTase-target conjugates (Ledvina et al., 2023). Cap2 from *Bacillus cereus* has been identified to conjugate CD-NTase to phage shock protein A (PspA) and Cap3 releases CD-NTase from conjugates upon phage infection to prime defence (Krüger et al., 2023).

PYCSAR (Pyrimidine cyclase system for antiphage resistance) systems also provide defence through signalling pathways. PYCSAR is a two-gene system, encoding a cyclase and an effector (Fig. 1-5) (Tal et al., 2021). A system from *Escherichia coli* E831 can generate 3'5' cyclic cytidine monophosphate as a second messenger, which in turn activates an effector with transmembrane helices, leading to cell death, presumably by interfering with membrane integrity. The cyclase from *Burkholderia cepacia* LK29 synthesises the signal molecule 3'5' cyclic uridine monophosphate, which subsequently activates TIR effectors to deplete cellular NAD$^+$, providing immunity through abortive infection mechanisms.

Thoeris systems employ the signal molecule variant cyclic ADP ribose (V-cADP) as a second messenger to mediate the signalling defence pathway (Fig. 1-5) (Doron et al., 2018, Ofir et al., 2021). The TIR effectors sense signal molecules and are activated to provide abortive defence.

1.1.5.2 Antiviral chemicals involved in defence systems

Recently, chemical defence systems have been characterised in diverse *Streptomyces* species. Secondary metabolites, such as daunorubicin and doxorubicin, can insert into phage DNA, interfering with phage replication (Fig. 1-5) (Kronheim et al., 2018). Aminoglycoside antibiotics have also been documented to block phage life cycle before viral replication and transcription (Kever et al., 2022).

Prokaryotic viperins, which resemble their counterparts in eukaryotes, modify nucleotides to generate 3'-deoxy-3',4'-didehydro (ddh) nucleotides by catalysing the removal of the hydroxyl group at the 3' carbon of the ribose (Fig. 1-5) (Bernheim et al., 2021, Lee et al., 2023). These modified molecules have been shown to terminate viral transcription, presumably by

incorporating ddh nucleotides into the viral RNA chain (Bernheim et al., 2021). The inserted ddh nucleotides could act as chain terminators, inhibiting further polymerisation.
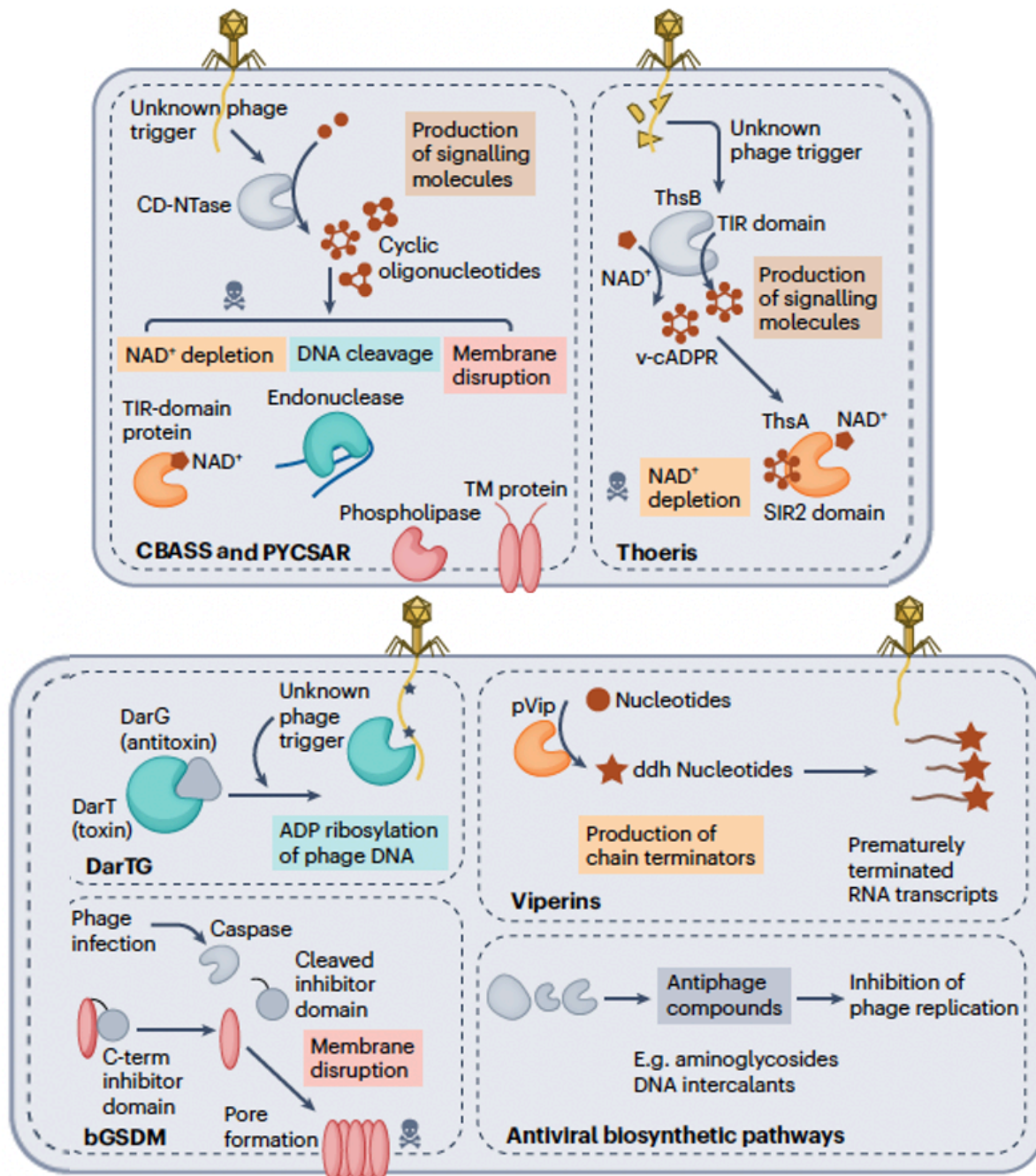
Figure 1-5 Other defence systems
CBASS, PYCSAR and Thoeris defence systems provide immunity through signalling pathways. Both Viperins and Antiviral biosynthetic pathways employ antiviral compounds to inhibit phage reproduction cycle (**section 1.1.5**). In DarTG TA systems, the toxin DarT is activated to inhibit phage replication by ADP-ribosylating phage DNA (LeRoux et al., 2022). Following phage infection, bacterial gasdermin (bGSDM) are released by removing the C-terminal inhibitor domain, leading to the formation of a large membrane pore and subsequent cell death (Johnson et al., 2022). Abbreviations: CBASS, cyclic-oligonucleotide-based anti-phage signalling systems; PYCSAR, pyrimidine cyclase system for antiphage resistance; CD-NTase, cGAS/DncV-like nucleotidyltransferase; TIR, Toll/interleukin-1 receptor; v-cADPR, variant cyclic ADP ribose; bGSDM, bacterial gasdermin; pVip, prokaryotic viperin. Figure modified from the original of Georjon *et al*. (Georjon and Bernheim, 2023).

1.2 CRISPR-Cas defence system

The CRISPR-Cas system stands as one of the most remarkable discoveries in recent decades, shedding light on the adaptive immunity employed by prokaryotes. This section will focus on the discovery, classification, molecular mechanism of CRISPR system and anti-CRISPR mechanism evolved by phage.

1.2.1 Discovery of CRISPR-Cas systems

The discovery of CRISPR-Cas systems dates back to the first observation of repeated DNA sequences in the *Escherichia coli* genome in 1987, documented by Ishino *et al*. from Osaka University (Japan) (Ishino et al., 1987). This unusual structure consisted of 29 highly conserved nucleotides arranged as repeats, separated by 32 nucleotides as spacers, what is now known as CRISPR, standing for clustered regularly interspaced short palindromic repeats.

However, it was Francisco Mojica who significantly advanced our comprehension of the biological function of CRISPR. While examining the archaeal genome of *Haloferax mediterranei*, he recognised similar repeated DNA structures, despite the fact there is no sequence similarity between bacteria and archaea (Mojica et al., 1993, Mojica et al., 1995). This intriguing commonality between such distant microbes aroused Mojica interests and inspired him to explore its purpose. In the year 2000, Mojica had discovered spaced repeat sequences in 20 different microbes (Mojica et al., 2000) and coined the term CRISPR in correspondence with Ruud Jansen (Jansen et al., 2002). By 2005, Mojica's bioinformatic work led to the discovery that near 60 spacers within CRISPR loci matched the sequence of viruses or conjugative plasmids associated with the microbes containing those spacers (Mojica et al., 2005). This important discovery led Mojica to propose that CRISPR functions to regulate or inhibit viral replication. Two other research groups reached similar conclusion around this time (Bolotin et al., 2005, Pourcel et al., 2005).

The first experimental evidence supporting CRISPR as an adaptive immune system took place in the context of the yogurt production industry. *Streptococcus thermophilus*, a bacterium commonly used in yogurt and cheese production, faced phage infection causing failures in fermentation cultures in the dairy factory. Philippe Horvath, working to address this issue, observed phage-derived sequences within the CRISPR of phage resistant strains of *S. thermophilus*. Horvath and his colleagues demonstrated increased resistance by insertion of phage sequences into CRISPR loci and showed that the Cas9 may play a crucial role in this immunity (Barrangou et al., 2007). In 2008, John van der Oost and colleagues soon

demonstrated CRISPR transcripts were processed by Cas proteins termed Cascade into CRISPR RNA (crRNA) containing spacer sequence, enabling Cascade to interfere with phages (Brouns et al., 2008).

A significant breakthrough came in 2012, Emmanuelle Charpentier and Jennifer Doudna published their work revealing that Cas9 from *Streptococcus pyogenes* could be programmed to cut DNA (Jinek et al., 2012). By fusing the crRNA (CRISPR RNA) and tracrRNA (trans-activating CRISPR RNA) into a single and synthetic guide RNA (sgRNA), they greatly simplified the system. This led to excitement regarding the potential of CRISPR-Cas systems as genome editing tools. In subsequent years, numerous research groups optimised and expanded CRISPR-Cas to enhance its efficiency and precision. This intense exploration also led to the discovery of an expanding diversity of CRISPR-Cas systems. In recognition of their pioneering contribution supporting CRISPR Cas9 for genome editing application, Emmanuelle Charpentier and Jennifer Doudna were rewarded with the Nobel Prize in Chemistry in 2020. The dedicated efforts of countless scientists have not only revolutionised genetic editing but also opened the door to a deeper understanding of the diverse CRISPR-Cas systems.

1.2.2 Classification of CRISPR-Cas systems

With the continued discovery and expansion of genomic and metagenomic databases, scientists have gained insights into the increasing number and diversity of CRISPR-Cas systems. Eugene Koonin and his colleagues developed a robust classification of CRISPR-Cas systems based on evolutionary relationships. Their latest classification, updated in 2020, builds upon their previous work from 2011 and 2015 (Makarova et al., 2011a, Makarova, 2015, Makarova et al., 2020b). This comprehensive classification considers various factors, including gene composition, genetic locus architecture, phylogenetic analysis of Cas proteins, modular structure in bipartite networks and experimental data. As a result, CRISPR-Cas systems have been classified into 2 classes, 6 types and 33 subtypes (Fig. 1-6b) (Makarova et al., 2020b).

Understanding the four distinct functional modules of CRISPR associated (Cas) proteins is essential for this classification (Fig. 1-6a) (Makarova, 2015, Makarova et al., 2013). The adaptation module is involved in spacer acquisition, mainly including Cas1 and Cas2. The expression module is responsible for crRNA processing and maturation, with Cas6 being a key component in most class 1 systems. The interference or effector module is the central component, responsible for target recognition and degradation. Two classes are distinguished in the gene composition of the interference module: class 1 systems possess multi-subunit interference complexes, while class 2 systems employ a single interference protein, such as

Cas9. The signal transduction or ancillary module is associated with core interference modules, playing an essential role in immunity (Niewoehner, 2017, Kazlauskiene et al., 2017).

Class 1 consists of type I, III and IV (Fig. 1-6b). Type I systems are the most diverse and abundant CRISPR-Cas systems. Their signature subunit is Cas3, which functions as a single-stranded DNA-stimulated helicase-nuclease. Type III systems employ the unique signature subunit Cas10, containing two polymerase-cyclase Palm domains for the synthesis of signal molecules (Niewoehner, 2017, Kazlauskiene et al., 2017). Another significant feature of type III systems is the presence of various ancillary genes located near the core *cas* genes (Shmakov et al., 2018, Shah et al., 2019). Type IV systems were derived from type III, but lack Cas10, instead including Csf1, particularly involved in mediating plasmid-plasmid conflicts (Ozcan et al., 2019, Pinilla-Redondo et al., 2020, Pinilla-Redondo et al., 2022).

Class 2 includes types II, V and VI (Fig. 1-6b). Type II systems are the most widespread among the class 2 systems. The effector Cas9 is the signature protein, featuring HNH and RuvC-like nuclease domains responsible for target DNA cleavage. Another notable feature of type II systems is tracrRNA, which is essential for pre-crRNA processing and interference. Type V systems have Cas12 as effector, with only the RuvC-like nuclease domain required for the cleavage of both stands of the target (Swarts et al., 2017). This type, although rare in bacteria, exhibits diversity in size, architecture, and molecular mechanisms of effectors. Type VI effectors only target RNA and use effectors Cas13, containing two HEPN RNase domains (Abudayyeh et al., 2016).

This classification not only serves as a vital guide for the ongoing research but also highlights the notable diversity and complexity of CRISPR-Cas systems, further enhancing our understanding of these fascinating defence mechanisms.
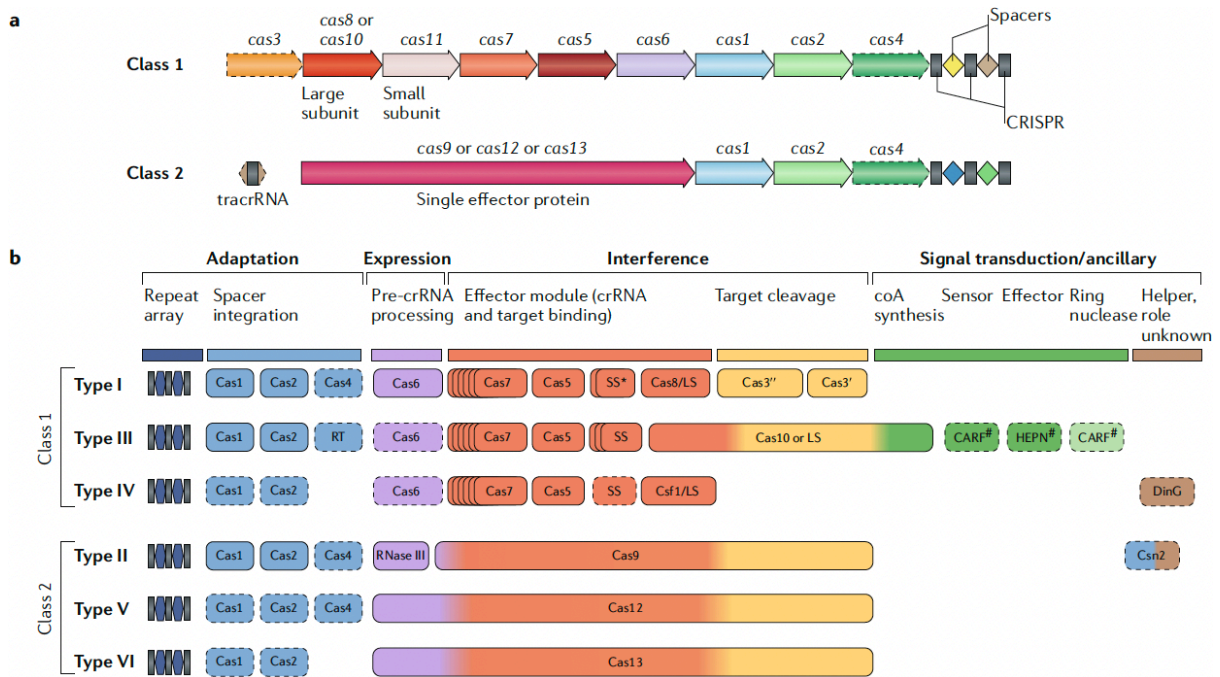
Figure 1-6 The classification of CRISPR-Cas systems
a. Genetic architecture of two classes of CRISPR loci. Class 1 interference modules encompass a collection of multiple Cas proteins, forming an interfering complex with crRNA for target RNA recognition and degradation. Class 2, on the other hand, employs a single, large and multidomain effector, guided by crRNA to execute interference. Some ancillary components are missing in some subtypes, indicated by dashed outlines.
b. The scheme of 2 classes and 6 types. The top legend provides an overview of four distinct functional modules of CRISPR-Cas systems, each corresponding to the genetic regions, distinguished by different colours. The small subunits often fused to the large subunits are indicated by asterisks. The less common components are presented with dashed outlines, maybe missing in some variants. The hash sign indicates the ancillary effectors involved in signal transduction. Figure used with permission (license number: 5718771317360) (Makarova et al., 2020b).

1.2.3 The mechanisms of CRISPR-Cas systems

The CRISPR-Cas immune response typically progresses through three primary phases: adaptation, expression, and interference. In the adaptation phase, new spacer sequences derived from viruses or plasmids are integrated into the CRISPR array. During the expression phase, the CRISPR array is transcribed and processed into mature CRISPR RNAs (crRNAs). In the interference phase, the Cas proteins are guided by crRNAs to recognise and cleave complementary target nucleic acids. Although this general process is common among CRISPR-Cas systems, the detailed molecular mechanisms in each phase can vary significantly among different types of CRISPR-Cas immunity.

1.2.3.1 Adaptation

Adaptation is mediated by Cas1 and Cas2, which are conserved across various types of CRISPR-Cas systems (Koonin et al., 2017). The core machinery of adaptation primarily involves two steps, capturing spacer sequences from invading nucleic acids (referred to as protospacers) and subsequently integrating these protospacers into the CRISPR array (McGinn and Marraffini, 2019).

For most CRISPR-Cas types, invading DNA molecules serve as the primary sources for adaptation, except for type III systems, which have been observed to convert RNA into cDNA as protospacers before integrating them into the CRISPR array. This reverse transcription process is mediated by a Cas1 nuclease fused with a reverse transcriptase, known as RT-Cas1, although the exact mechanisms remain unclear (Gonzalez-Delgado et al., 2019, Silas et al., 2016). In type I and II systems, free dsDNA ends are preferred substrates for adaptation and enriched by the host DNA repair machinery, like RecBCD in Gram-negative organisms and AddAB in Gram-positive organisms (Ivancic-Bace et al., 2015, Levy et al., 2015, Modell et al., 2017). RecBCD processes a blunt dsDNA end into a ssDNA overhang structure terminated at Chi (Cross over Hotspot Instigator) an octameric regulatory sequence that attenuates RecBCD nuclease activity (Fig. 1-7B) (Dillingham and Kowalczykowski, 2008). Chi sequences are more abundant in the host chromosome than phage or plasmid genomes (Levy et al., 2015). Furthermore, most invading DNA has a linear genome with a free dsDNA end, distinguishing it from the host circular chromosome. This machinery enables hosts to differentiate self from non-self nucleic acids, avoiding autoimmunity (Levy et al., 2015, Modell et al., 2017). However, in some cases, host DNA repair machinery is not essential for spacer acquisition, suggesting that alternative pathways are involved in spacer generation (Levy et al., 2015, Modell et al., 2017).

A critical component in selecting functional spacers is the protospacer-adjacent motif (PAM). In type I and II systems, the PAM is located downstream of the target site, allowing for target cleavage, and avoiding self-targeting. The Cas1 and Cas2 complex in the type I-E specifically recognises PAM sequences and mediates the acquisition of PAM-adjacent protospacers (Wang et al., 2015). However, in the type II, Cas9 facilitates the PAM-specific spacer acquisition through interaction with Cas1-Cas2 complex, which lacks PAM selectivity in this case (Heler et al., 2015). Several accessory proteins have been shown to associate with Cas1 and Cas2 to bias spacer selection, including Cas4 in type I and Csn2 in type II (Heler et al., 2015, Dhingra and Sashital, 2023, Dhingra et al., 2022).

The primed adaptation response to escape phages with mutations in the PAM or spacer sequences has been well-documented (McGinn and Marraffini, 2019). In this process, new spacers are acquired more efficiently from pre-encountered viral genomes compared to "naïve acquisition", which occurs when a phage or plasmid has not been previously encountered (Fig. 1-7A and B). This priming process is associated with the interference machinery, as evidenced by the fusion of Cas2 with the signature effector Cas3 in type I-F systems (Fagerlund et al., 2017). In type I-E systems, the crRNA-guided CRISPR-associated complex (Cascade) can generate spacer substrates by recruiting the nuclease-active Cas3 for target interference in a PAM-dependent manner (Redding et al., 2015). Even when the mutations affect the PAM, Cascade still can bind to the targets, but recruit a nuclease inactive Cas3, strictly depending on the Cas1 and Cas2. In this case, the Cas1 and Cas2 attenuate Cas3, allowing it to rapidly translocate along the foreign DNA and generating protospacers from adjacent target DNA sequences (Redding et al., 2015).

The integration of new spacers is mediated by Cas1-Cas2 integrase complex in a polarised manner (Fig. 1-7C). A heterohexameric complex [(Cas1$_2$-Cas2]$_2$], two Cas1 dimers connected by a central Cas2 dimer, predominantly integrates new spacers close to the leader end of the CRISPR array (Wright et al., 2017, Xiao et al., 2017). Upon loading with a protospacer, the Cas1-Cas2 complex first cleaves the leader end of first repeat and subsequent the spacer end of the repeat (Xiao et al., 2017). The 3'-OH of each strand of the protospacer is attached to each end of the repeat DNA through nucleophilic attack, resulting in the single-stranded DNA (ssDNA) repeat sequences. The integration process is completed after repeat duplication and ligation (Xiao et al., 2017, Wright et al., 2017). This polarised integration has been proposed as a bet-hedging strategy, where the latest acquisition provides more robust immunity and allows host efficiently against most recent invaders (Weinberger et al., 2012, McGinn and Marraffini, 2016). This strategy may be result from the differential expressions of crRNA. For

example, in *Streptococcus pyogenes* type II systems, the spacer sequence in the first position of the CRISPR array exabits twofold greater abundance than in the fifth position, which could significantly impact the efficiency of the immune response (McGinn and Marraffini, 2016). Cas1 and Cas2 are sufficient for this polarised integration in type II CRISPR systems, as an α-helix of Cas1 specifically interacts with the minor groove of the leader anchoring sequence (LAS) (Xiao et al., 2017). In contrast, type I systems recruit additional host factors to facilitate this process, such as integration host factor (IHF) or related DNA-bending proteins (Wright et al., 2017, Nunez et al., 2016).
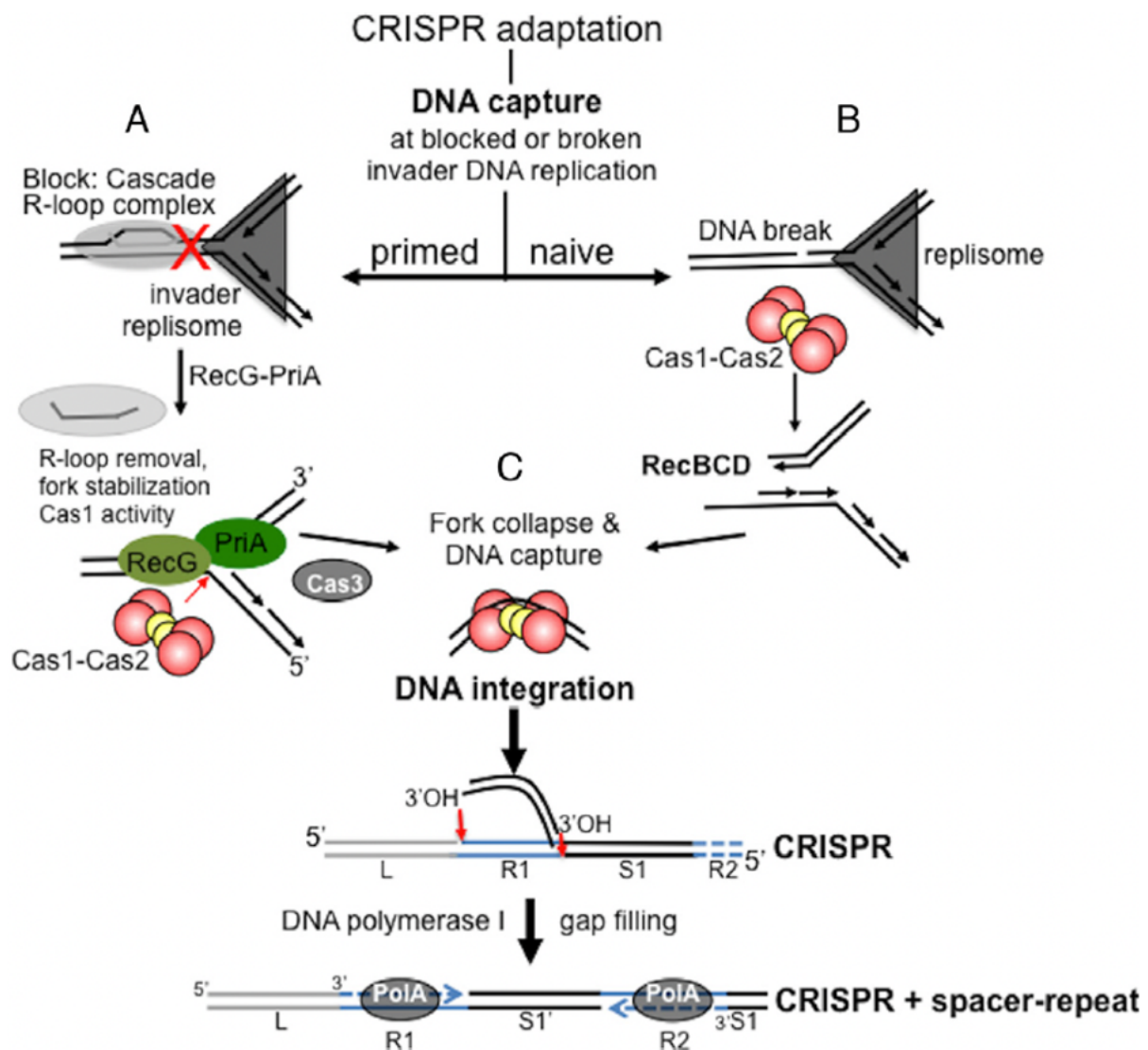
Figure 1-7 Proposed model of primed and naïve adaptation in *E. coli*
A. In primed adaptation, even when the PAM is mutated, the Cascade complex can still bind to the target, blocking replication forks of the invader as indicated by the red symbol X. The RecG and PriA proteins recognise this blockage and remodel forks to prepare them for Cas1-Cas2-mediated nicking of the substrate and subsequent DNA capture. Following this, Cas3 may intervene to release the captured DNA. **B.** In naïve adaptation, the DNA repair machinery, RecBCD, play a key role in the generation and capture of free ends dsDNA. **C.** DNA integration is catalysed through a two-step cleavage-ligations process by Cas1-Cas2 complex. DNA polymerase I (indicated as PolA) can fill the ssDNA repeat gaps, thereby facilitating the integration process. Figure is from Ivancic-Bace *et al*. (open access with unrestricted reuse, distribution and reproduction) (Ivancic-Bace et al., 2015).

1.2.3.2 Expression

The expression phase includes crRNA biogenesis and ribonucleoprotein (RNP) interference complex formation. Mature crRNAs are essential components of the CRISPR-Cas defence response, each comprising a spacer sequence flanked by portions of repeats. They serve as guides for an interference complex to target and defend against foreign nucleic acids. crRNA maturation involves three steps (Charpentier, 2013). The CRISPR array is initially transcribed into a long precursor crRNA (pre-crRNA). This pre-crRNA is subsequently cleaved within the repeat sequence to generate intermediate crRNAs that contains intact spacer sequence flanked by parts of repeats. These first two steps are shared among the various CRISPR types (Charpentier, 2013). In some CRISPR types, the intermediate crRNAs are further processed into mature crRNAs. Here, type I and III systems exemplify class 1 CRISPR systems, while type II presents class 2.

In class 1 CRISPR-Cas systems, Cas6 or Cas6-like metal-independent endoribonucleases catalyse the initial processing event at a specific site within their cognate repeat sequences (Charpentier, 2013). In type I systems, pre-crRNA is cleaved at a conserved position, resulting in an intermediate crRNA with a central spacer sequence flanked by a 5' handle (8 repeat-derived nucleotides) and a 3' handle (21 repeat-derived nucleotides) (Fig. 1-8) (Oost, 2022). In most type I systems, no further processing is required, and Cas6-like enzymes remain associated with the hairpin of mature crRNAs as a subunit of Cascade complex (Oost, 2022, Charpentier, 2013). Furthermore, a hairpin structure at 3' end is believed to facilitate Cas6 catalysis and assist in the stable interaction between crRNA and Cascade interference complex (Charpentier, 2013).

In type III systems, intermediate crRNA undergoes a further trimming to generate the mature crRNAs (Fig. 1-8). Cas6 specifically cleaves at the base of stem-loop of type III repeats, generating an intermediate crRNA similar to that seen in the type I systems (Carte et al., 2010, Hatoum-Aslan et al., 2011). This intermediate is believed to be transferred from Cas6 to the interference complex (Csm (Cas subtype Mtube) complex in type III-A and Cmr (Cas module RAMP) in type III-B) through a transient interaction (Hatoum-Aslan et al., 2014, Sokolowski et al., 2014). Once bound to the complex, the complex backbone (Csm3/Cmr4) serves as a ruler to determine the length of mature crRNA. Unidentified host nucleases trim exposed 3' end, leading to crRNA maturation (Zhang et al., 2012, Osawa et al., 2015, Hatoum-Aslan et al., 2011, Walker et al., 2017).

In class 2 CRISPR-Cas systems, the expression machinery is much more diverse. In type II systems, crRNA maturation requires three indispensable components, Cas9, pre-crRNA, and transactivating crRNA (tracrRNA). TracrRNA, a non-protein coding RNA, contains a 25-nt stretch complementary to CRISPR repeats. It is believed that signature protein Cas9 facilitates the formation a stable tracrRNA-pre-crRNA duplex, allowing RNase III to recognise and cleave this duplex (Deltcheva et al., 2011). The further processing at 5'end of crRNAs is catalysed by unidentified nuclease, resulting in mature crRNA (around 42 nt) and tracrRNA (about 75 nt) (Deltcheva et al., 2011).
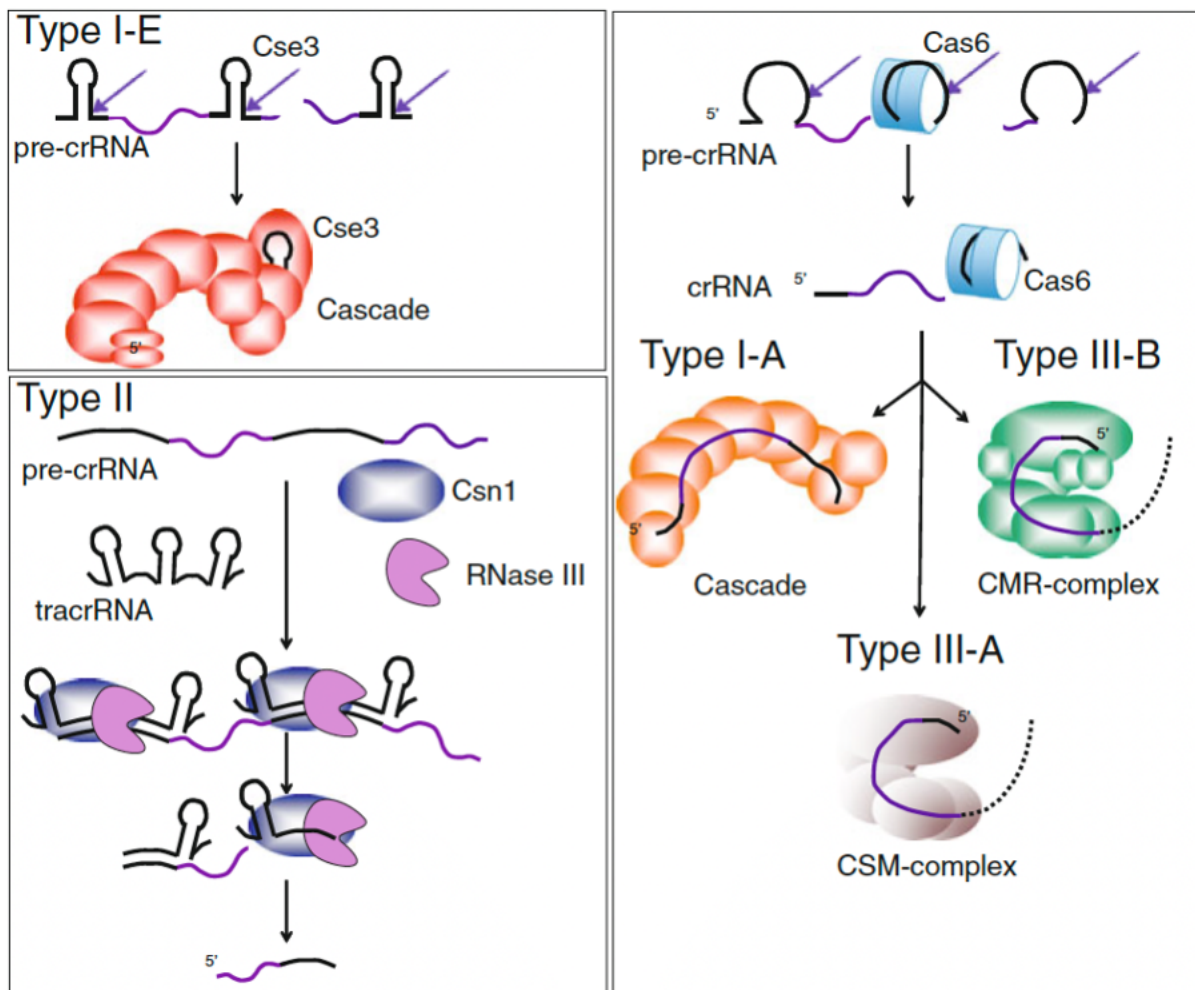
Figure 1-8 Expression in type I, II and III CRISPR-Cas systems
In the type I system, I-E as an example, the nuclease Cas6e (Cse3) recognises the hairpin structure in pre-crRNA and processes it into mature crRNA. Cas6e remains associated with hairpin of crRNA and is integrated into the Cascade complex, which is used for the recognition of invading nucleic acids. In type II systems, pre-crRNA is bound to tracrRNA that is complementary to the repeat sequence, that is recognised and cleaved by host RNase III in the presence of Cas9 (Csn1) protein. crRNA is matured with further processing by unknown nucleases. In type III-B systems, Cas6 endonuclease cleaves pre-crRNA to generate intermediate crRNA, which is transferred into Csm/Cmr complex. 3'end repeat-derived sequence is trimmed away by unknown nucleases.
Figure is used with permission (license number 5718801170088) (Charpentier, 2013).

1.2.3.3 Interference

The Interference phase involves a stepwise process in which the assembled CRISPR ribonucleoprotein (crRNP) complex distinguishes foreign nonself sequences from self-sequences and is subsequently activated to specifically degrade invading nucleic acids (Mohanraju et al., 2016). In most cases, the presence of a PAM (protospacer adjacent motif) sequence on the target is crucial for the recognition and activation of the crRNP complex.

In type I systems, the Cascade-crRNA RNP complex, comprising multiple subunits, specifically targets invading DNA by recruiting the helicase/nuclease Cas3 (Fig. 1-9). The PAM motif on a nontarget strand serves as the initial checkpoint for discrimination of self from nonself, a feature also been found in type II (Cas9) and typeV (Cas12) systems (Mojica et al., 2009, Semenova et al., 2011, Zetsche et al., 2015, Oost, 2022). PAM recognition results in the partial melting dsDNA downstream of the PAM. The crRNA's seed sequence (6-8 nt) at the 5' end of the spacer sequence then base pairs with the target, forming the second checkpoint (Semenova et al., 2011, Wiedenheft et al., 2011). Mismatches during this base-pairing process lead to interference termination (Rutkauskas et al., 2015). Perfect base pairing allows the unwinding of the dsDNA from seed sequence, forming a R-loop structure where the target strand base pairs with the crRNA guide, displacing the nontarget strand (Rutkauskas et al., 2015). Excessive mismatches downstream of the seed abort further interference, while slight mismatches are tolerated, serving as another checkpoint (Rutkauskas et al., 2015).

Cascade-crRNA complex locks the R-loop structure and subsequently recruits and activates Cas3 to degrade exposed regions of nontarget strand (Redding et al., 2015, Xiao et al., 2018, Loeff et al., 2018). Cas3, which consists of an ATP-dependent SF2 (superfamily-2)-like helicase domain and HD (histidine-aspartate)-like nuclease domain, unwinds the target dsDNA, resulting in reeling and looping of the target strand and occasional nicking of the nontarget strand. Host nucleases may further degrade the looped target strand, causing additional damage to the invading DNA.
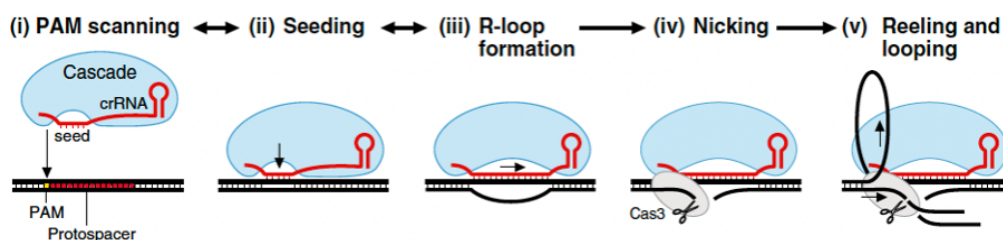
Figure 1-9 crRNA-guided dsDNA interference in type I systems
i. Cascade/crRNA scans the PAM motif on nontarget strand of target DNA. **ii.** Seed sequence of crRNA base pairs with target sequence upon recognising the PAM. **iii.** The R-loop is formed when guide sequence of crRNA completely base pairs with protospacer of target DNA. **iv.** Cas3 is recruited to nick the nontarget strand of target DNA. **v.** After initial nicking, Cas3 starts to reel and loop target strand of target DNA and keeps nicking the nontarget strand. Figure modified from the original of John van der Oost (Oost, 2022).

Type III systems recognise complementary invading transcripts (RNA) instead of DNA molecules (Fig. 1-10A). The self/nonself discrimination depends on the base-pairing potential between the 5'-repeat sequence of crRNA (8 nucleotides, known as 5'-repeat tag) and 3'-flanking sequences of target RNA (Fig. 1-10B) (Jia et al., 2019c, Wang et al., 2019). If the target sequence is complementary to the 5'-repeat tag of crRNA, it is recognised as self-transcripts from the CRISPR array, resulting in the inactivation of interference (Taylor et al., 2015, You et al., 2019). The interference complex, Csm (Csm1-5 in type III-A) or Cmr (Cmr1-6 in type III-B), is guided by crRNA to scan for complementary RNA and subsequently is activated to cleave target RNA at 6-nt interval using the catalytic activity of Csm3/Cmr4 backbone subunits (Taylor et al., 2015). Upon target RNA binding, the conformational change of Csm/Cmr complex enables the Cas10 (Csm1/Cmr2) enzymatic subunit to cleave ssDNA non-specifically using its HD nuclease domain and synthesise cyclic oligoadenylate (cOA) using its cyclase PALM domain (Kazlauskiene et al., 2017, Niewoehner, 2017). cOA act as a second messenger to activate various ancillary proteins, enhancing CRISPR immune response. Signalling pathways are one of the most unique features in type III CRISPR systems. More details are provided in section 1.3.
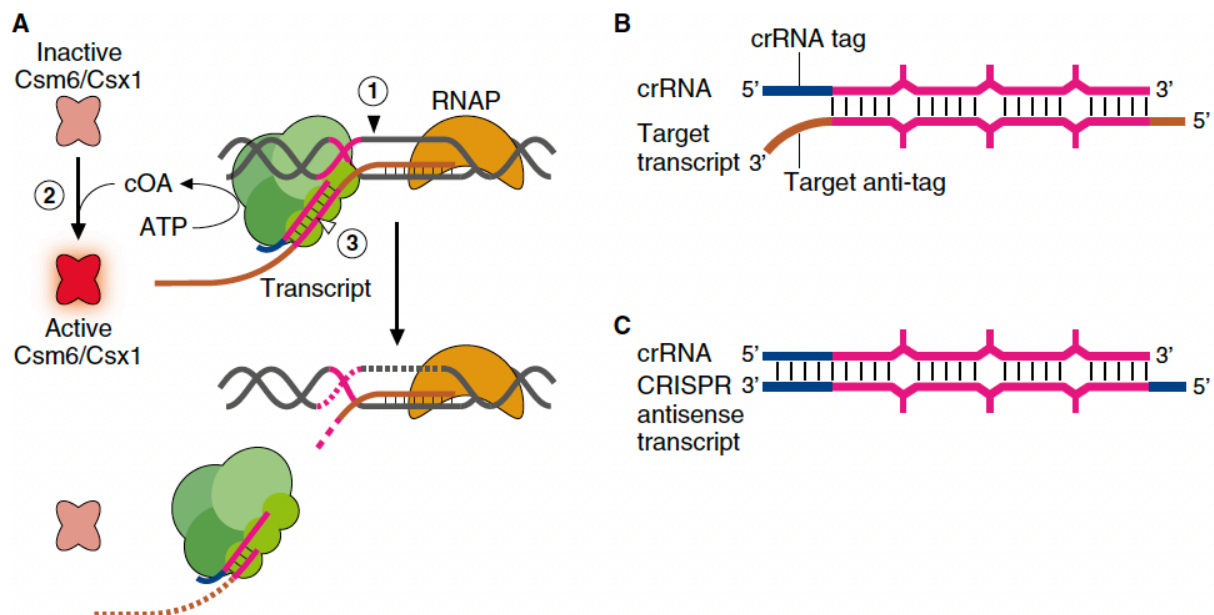
Figure 1-10 crRNA-guided RNA and DNA interference in type III systems
A. Csm/Cmr complex, guided by crRNA, provides three enzymatic activities upon recognition of target RNA, target RNA cleavage by the Csm3/Cmr4 subunit of the complex, nonspecific DNA degradation mediated by HD nuclease domain of Cas10 and cOA synthesis by cyclase domain of Cas10, which in turn activates the Csm6/Csx1 nonspecific RNase activity. **B.** The crRNA-target RNA duplex lacks base-pairing between crRNA tag and target anti-tag, which is crucial for activation of type III immunity. RNA duplex exhibits a discontinuous structure with every 6th base being flipped in the spacer region. **C.** Antisense transcription of the CRISPR array is complementary to the crRNA, which does not active type III systems, avoiding autoimmunity. Figure is used with permission (license number is 5718810454888) (Marraffini, 2022).

Class 2 systems, unlike class 1 systems, rely on a single protein for interference. In type II systems, Cas9 is the sole protein involved, binding to the crRNA-tracrRNA duplex for recognition and degradation of target dsDNA (Fig. 1-11) (Mohanraju et al., 2016). Cas9 has a bilobed architecture with nuclease (NUC) and recognition (REC) lobes (Jinek et al., 2014). The PAM interaction (PI) site of Cas9 is formed when the crRNA-tracrRNA duplex is loaded (Anders et al., 2014). Once recognising the PAM motif, the sequence-specific interaction between the PI site of Cas9 and the PAM promotes local DNA duplex melting upstream of the PAM (Anders et al., 2014). R-loop formation occurs as base pairing between seed of crRNA and target RNA strand drives propagation of target and guide DNA heteroduplex (Jiang et al., 2016a, Szczelkun et al., 2014). This R-loop triggers conformational changes in the HNH and RuvC nuclease domains of Cas9, leading to the cleavage of the complementary strand within the DNA heteroduplex by the HNH domain and the cleavage of non-complementary strand by the RuvC domain (Anders et al., 2014, Jiang et al., 2016a, Nishimasu et al., 2014, Sternberg et al., 2015). Cas9 cleaves at the PAM-proximal end of the protospacer, generating a blunt-end or 1-nt overhang double-strand break.
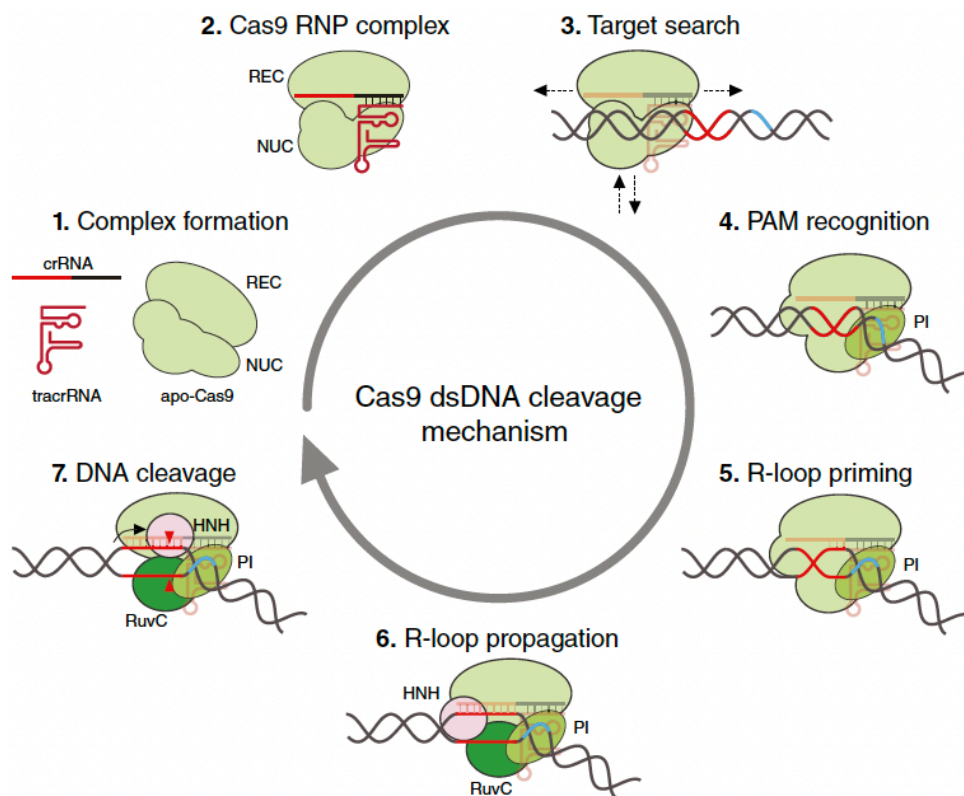
Figure 1-11 Cas9 nuclease-mediated DNA cleavage interference in type II systems
1. Cas9, comprising nuclease (NUC) and recognition (REC) lobes, is in the apo state. **2.** Once loaded with crRAN-tracrRNA duplex, apo-Cas9 turns into the DNA recognition-competent complex. **3.** Cas9 RNP complex interrogates dsDNA searching for PAM sequences. **4.** Upon recognition PAM sequences, Cas9 RNP complex interacts with PAM in a sequence-specific manner and bends dsDNA. **5.** R-loop is initiated from the PAM-proximal seed sequence. **6.** Subsequent propagation is ended at the PAM distal end. **7.** Complete R-loop formation allows concerted DNA cleavage by HNH and RuvC nuclease domains. Figure is used with permission (license number is 5718810626385) (Tautvydas Karvelis, 2022).

Type V systems require individual Cas12 effectors for dsDNA, ssDNA or ssRNA interference in a crRNA-guided manner (Beckett, 2022). Most Cas12s target dsDNA through the recognition of PAM sequences and the formation of the R-loop (Fig. 1-12A) (Liu et al., 2019, Yamano et al., 2016, Yang et al., 2016). The PAM position is at the 5' end of the target sequence, as seen in type I systems. Mutations in PAM or mismatches within the seed sequence significantly reduce the interference efficiency (Wright et al., 2016). Cas12 initially cleaves the non-complementary strand of targets at the PAM-distal end of the protospacer and then cleaves complementary strand of targets outside the complementary region. This target dsDNA cleavage generates a double-strand break with overhangs at 5' end (5-12 nt). ssDNA and ssRNA cleavage mediated by some Cas12 effectors require a tracrRNA but do not require a PAM motif (Fig. 1-12B) (Harrington et al., 2018, Yan et al., 2019).
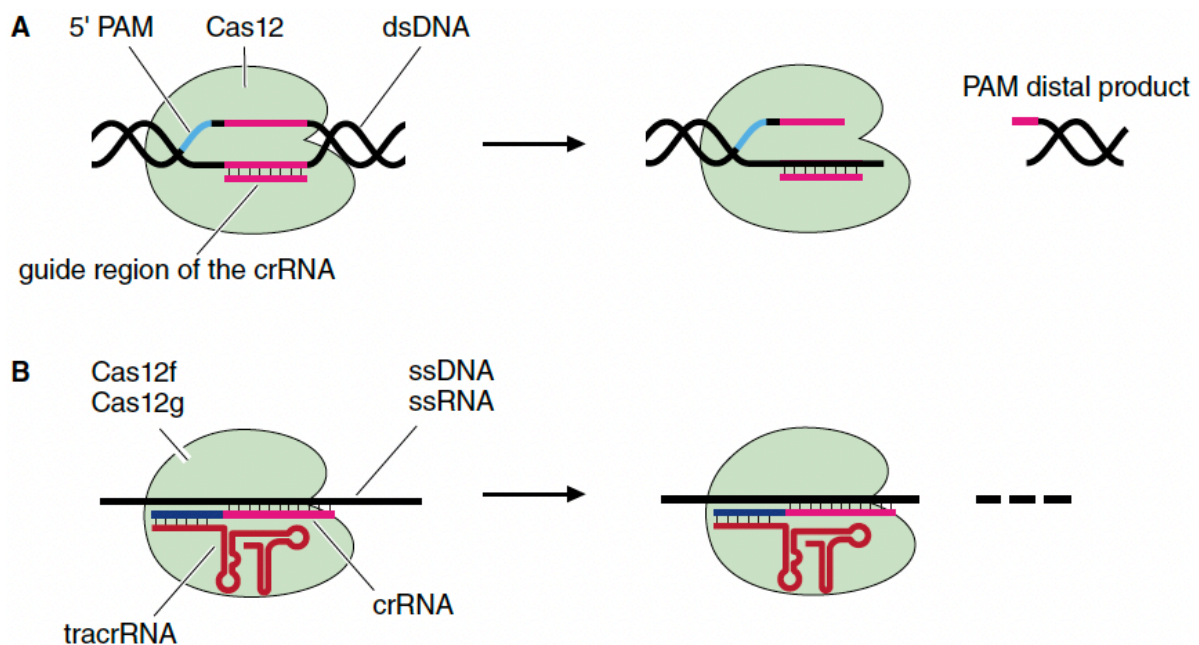
Figure 1-12 Cas12-mediated interference in type V systems
A. Most Cas12 effectors target dsDNA to produce PAM distal products with a 5' overhang. **B.** ssDNA and ssRNA cleavage are mediated by Cas12 in the presence of tracrRNA, with no PAM requirement. Figure modified from the original of Morgan Quinn Beckett (Beckett, 2022).

Type VI systems employ a unique effector, Cas13, with two HEPN RNase domains. Cas13, guided by crRNA, targets invading RNA (Fig. 1-13A) (Abudayyeh, 2022, Abudayyeh et al., 2016). Cas13 recognises a target sequence complementary to the seed region of the crRNA and cleaves single-stranded regions of targets at its preferred bases. Different Cas13 effectors have their own base preferences, such as uridines for Cas13 from *Leptotrichia shahii* and adenines for Cas13 from *Lachnospiraceae* (Abudayyeh et al., 2016, East-Seletsky et al., 2017). In addition, the 3' protospacer flanking site (PFS) motif restricts Cas13-mediated cleavage (Meeske and Marraffini, 2018, Abudayyeh et al., 2016). Complementarity between the PFS

motif and the direct repeat of crRNA prevents HEPN activation. Cas13 effectors exhibit nonspecific collateral RNase activity, which is triggered upon binding to a cognate target RNA (Fig. 1-13B) (Abudayyeh et al., 2016). This can lead to cell death or dormancy, providing population-level protection.
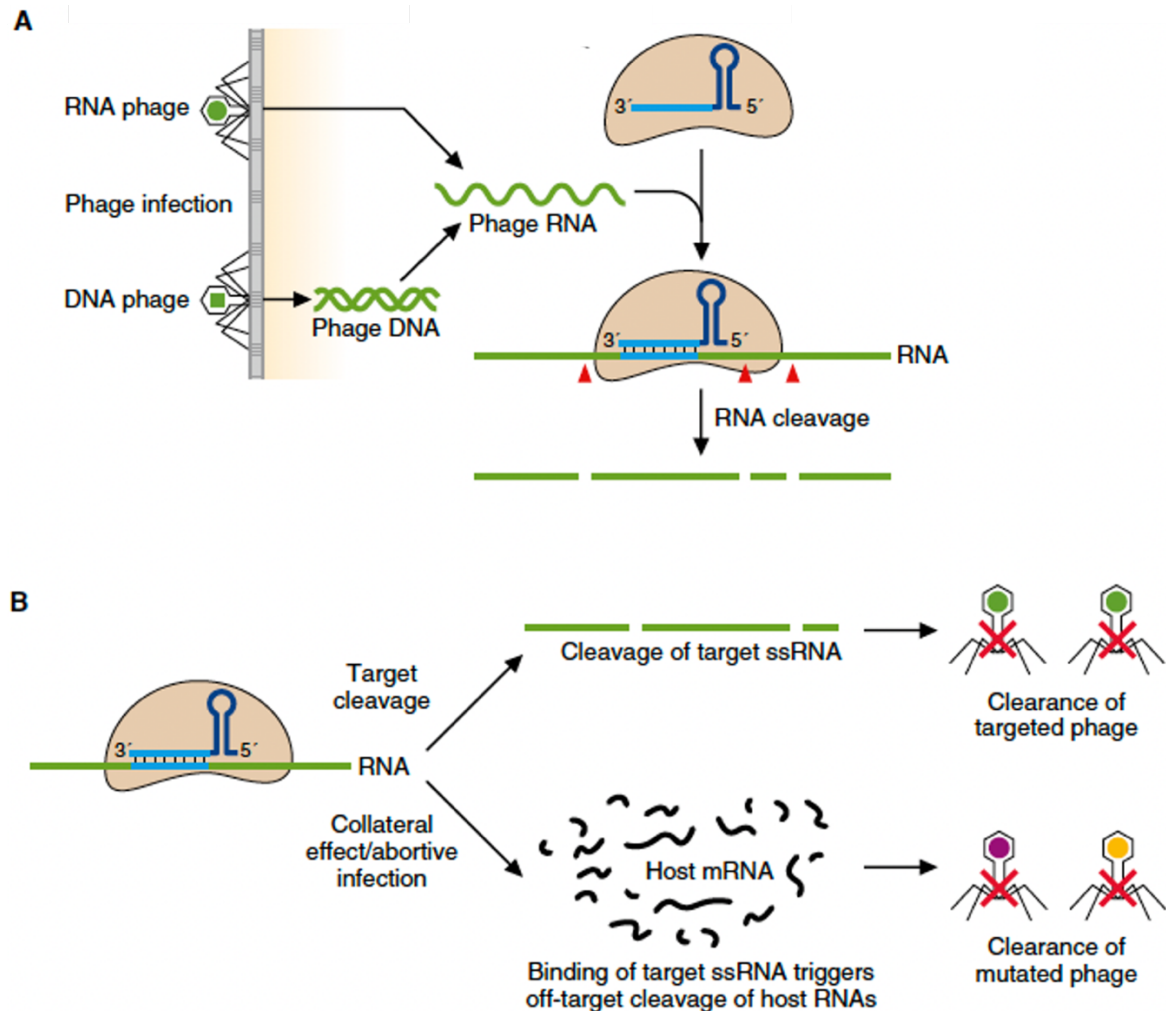
Figure 1-13 crRNA-guided RNA interference in type VI systems
A. Cas13-crRNA complex recognises and degrades invading RNA molecules. **B.** Cas13 effectors exhibit dual interference activities, target ssRNA cleavage and collateral nonspecific RNase activity, leading to cell death or dormancy. Figure modified from the original of Omar O. Abudayyeh and Jonathan S. Gootenberg (Abudayyeh, 2022).

1.2.4 Phage counter-measures

Phages have evolved a variety of counter-defence strategies to evade CRISPR-Cas systems. As described earlier, CRISPR-Cas systems initiate their defence through the base-pairing of crRNA with invading nucleic acids. Phages protect their genomes from CRISPR-Cas detection through mutations, modification and compartmentalisation (Fig. 1-14B-D) (Jenny Y. Zhang, 2022).

Analysis of phage escapers from CRISPR-Cas immunity revealed a range of mutations in PAM motif and protospacer sequences (Fig.1-14B). Even single mutations in PAM or seed sequences are sometimes sufficient to evade dsDNA interference in type I, II and V CRISPR-Cas systems, as these systems strictly rely on recognition of the PAM or seed regions (Jenny Y. Zhang, 2022, Deveau et al., 2008, Cady et al., 2012, Box et al., 2016). In contrast, phage escapers of RNA-targeting type III CRISPR systems have been found to contain large deletions including the protospacer or point mutations in the promoter to silence transcription (Pyenson et al., 2017). This is likely due to the lower specificity in base-pairing between the crRNA and target RNA and the lack of a requirement for PAM or seed sequences in type III systems.

Phage genome modification is another strategy to overcome CRISPR immunity (Fig. 1-14C). For example, ghmC (glucosyl-hydroxymethylated cytosines) modification enables *E. coli* phage T4 to evade heterologous type II and native type I-E CRISPR-Cas immunity (Bryson et al., 2015, Vlot et al., 2018). However, type V-A systems can still target T4 DNA with the same ghmC modification, suggesting that type V-A nuclease Cas12a exhibits more flexible architecture compared to type II nuclease Cas9 and type I Cascade, allowing Cas12a to bind modified target DNA (Vlot et al., 2018).

Compartmentalised phage DNA has been recently revealed as a most potent protection mechanism (Fig 1-14D). Certain phages, like jumbophage families of *Pseudomonas aeruginosa* and *Serratia* phages, have been found to assemble a nucleus-like structure to protect viral DNA from exposure to the cytoplasm (Chaikeeratisak et al., 2017, Malone et al., 2020). These phages thus are highly resistant to DNA-targeting CRISPR-Cas systems, like type I-A, II-A and V-A, but not to type III-A and V-A RNA-targeting systems (Malone et al., 2020).

Phages also encode various anti-CRISPR (Acr) proteins to specifically inhibit CRISPR-Cas systems, such as Cas proteins function, crRNA loading, DNA target binding, or nuclease activities (Fig. 1-14E-G) (Jenny Y. Zhang, 2022). The majority of Acr proteins function through inhibition of target binding. For example, AcrIF1 and AcrIF2 act as target binding inhibitors of type I-F CRISPR systems (Fig. 1-14E) (Guo et al., 2017, Chowdhury et al., 2017,

Bondy-Denomy et al., 2015). Target binding in type I is mediated by Cascade complex, where Cas7fs (Csy3) makes up the backbone and the 5' and 3' ends of crRNA are accommodated into a Cas8f:Cas5f heterodimer and a Cas6f monomer respectively (Guo et al., 2017). AcrIF1 induces a conformational change in the complex backbone by binding Cas7f, interfering with the base-pairing between crRNA and target DNA. AcrF2 has acidic charges on its surface, mimicking DNA negative charge and interacting with positively charged residues on Cas8f and Cas7f, thus inhibiting target DNA binding. AcrIF3 inhibits Cas3 nuclease activity in type I systems by blocking the DNA-binding cleft of Cas3 (Fig. 1-14G) (Bondy-Denomy et al., 2015).
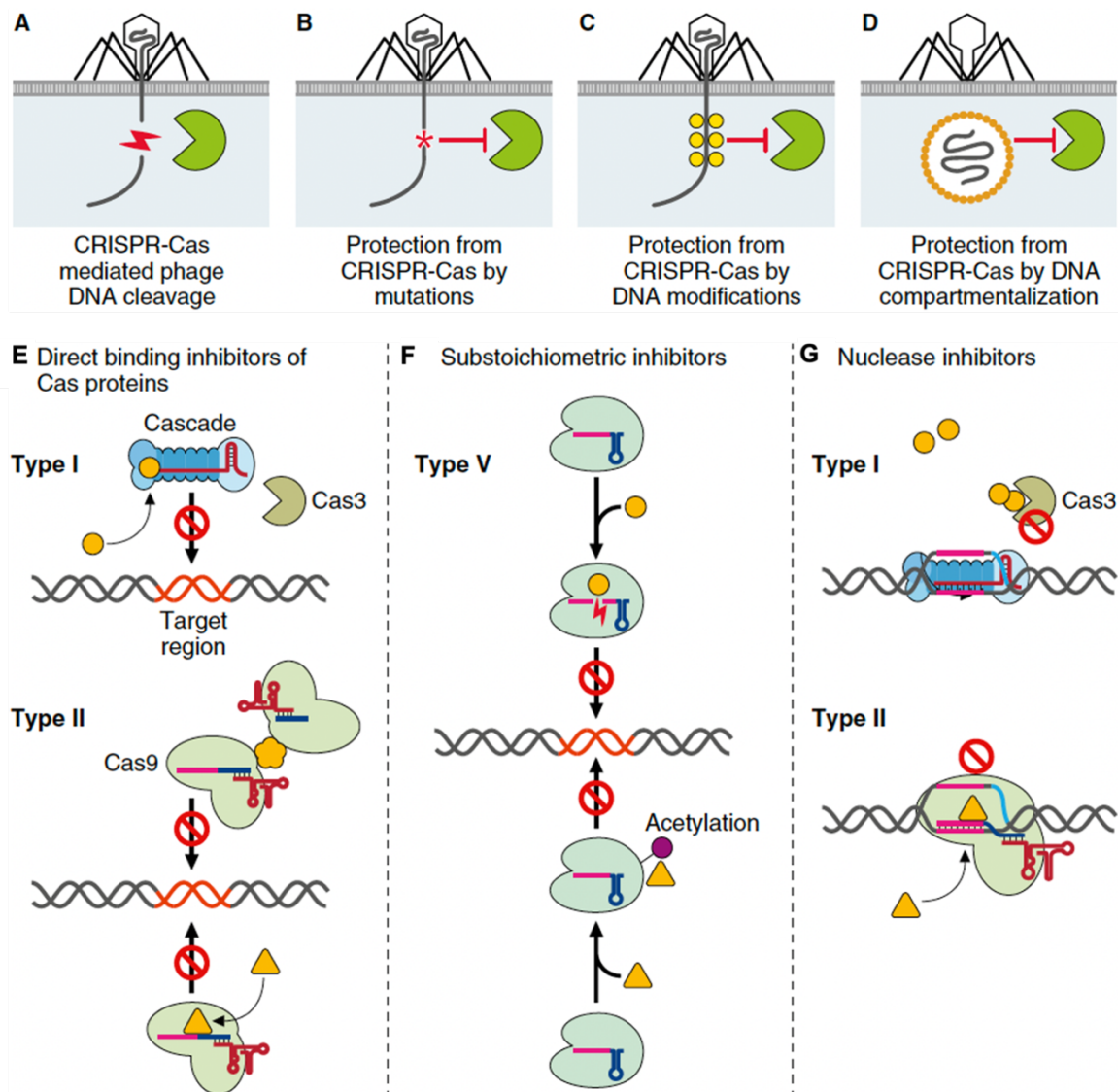
Figure 1-14 Anti-CRISPR mechanisms
A. Simplified CRISPR-Cas systems mediated invading DNA cleavage (red lightning bolt). **B.** Phage DNA mutations (indicated by star) prevent CRISPR immunity. **C.** Phage DNA modifications (yellow circles) protect from crRNA binding and cleavage. **D.** A compartmentalised structure protects the phage DNA from cleavage. **E.** Acr proteins (colored in yellow) directly bind to Cas proteins, inactivating the RNP complex. **F.** Enzymatic inhibitors catalyse crRNA cleavage and Cas protein modification to subvert CRISPR immunity. **G.** Acr proteins inhibit nuclease activity of Cas proteins. Figure modified from the original of Jenny Y. Zhang *et al*. (Jenny Y. Zhang, 2022).

1.3 Type III CRISPR-Cas system

Type III systems are categorised into six subtypes, III-A to III-F (Makarova et al., 2020b). Type III systems are generally composed of a large signature subunit Cas10, small subunit Cas11, one Cas5 protein and several paralogous Cas7 proteins. The enzymatic subunit Cas10, containing two polymerase-cyclase Palm domains, has been shown to have the capacity to synthesise cyclic oligoadenylates (cOA) primarily in subtype III-A and III-B. However, in the other subtypes, Cas10 either lacks Palm domains or is absent in subtype III-E, resulting in the absence of signalling pathways (Marraffini, 2022, Tamulaitis et al., 2017, Molina et al., 2020, Athukoralage and White, 2022). Therefore, subtypes III-A and III-B are mainly reviewed here.

1.3.1 Generation of cyclic oligoadenylate (cOA)

In subtype A, the interference complex is called Csm (Cas subtype Mtube), and in subtype B, it is referred to as Cmr (Cas module RAMP) (Makarova et al., 2020b, Makarova, 2015). These complexes share both structural and functional similarities (Molina et al., 2020, Tamulaitis et al., 2017). Their overall architectures feature a central helical backbone with a large subunit Cas10 (Csm1/Cmr2) and Cas7 family proteins (Csm5 or Cmr1/Cmr6) bound to each side. The central backbone is formed by intertwining a major filament composed of Csm4/Cmr3 and Csm3/Cmr4, with a minor filament comprising Csm2/Cmr5 and C-terminal domain of Cas10 (Fig. 1-15). The crRNA passes through the entire Csm/Cmr complex, with a trimmed spacer region kinking along the major filament until it is capped by Csm5/Cmr6 at 3' end, and a repeat-derived 5' tag attached to Csm4/Cmr3 (Fig. 1-15 a, c and d). The direct interaction between crRNA and Cas proteins indicates the indispensable role of crRNA in Csm/Cmr assembly. Upon target RNA binding, the conserved thumblike β-hairpin domain of Cmr4/Csm3 inserts itself into crRNA-target RNA duplex. This leads to the flipping out of one base pair in the opposite direction after every five base pairs, with the flipped-out base positioned adjacent to the catalytic residues of Cmr4/Csm3 for target RNA cleavage (Fig. 1-10 B and C) (Taylor et al., 2015, Jia et al., 2019c, Jia et al., 2019a).

The signature enzymatic subunit Cas10 (Csm/Cmr2) typically comprises an HD (histidine-aspartate) nuclease domain at the N-terminus and two Palm polymerase domains. However, in many cases, Cas10 lacks the HD domain but has intact Palm domains, which implies their immunity depends on signalling pathways (Gruschow et al., 2021). The comparison of Csm/Cmr-crRNA in complex with non-self RNA (Cognate Target RNA, CTR) and self-transcripts (Non-Cognate Target RNA, NTR) revealed discrimination of self from non-self

RNA relying on base pairing between 5' tag of crRNA and 3' flanking sequences of target RNA (3' anti-tag) (Jia et al., 2019c, Jia et al., 2019a, Sofos et al., 2020, You et al., 2019, Taylor et al., 2015). A non-complementary 3' anti-tag sequence triggers DNase and cOA synthase activities. For example, the structure of *Streptococcus thermophilus* (Sth) Csm with CTR bound unveiled that the 3' anti-tag region induces the formation of Csm1 Linker region and a loop in the Palm1 domain (termed L1), which is absent from the NTR-bound structure (Fig. 1-15 a and b) (You et al., 2019). Furthermore, the interaction between the non-complementary 3' anti-tag with Csm affects the DNase and cyclase activities but not RNase activity of Csm3 subunits, as substitutions of key residues in Csm1 Linker and the zinc finger exhibits significantly reduced DNA cleavage and cOA synthesis, with little effect on target RNA cleavage (You et al., 2019).

In contrast, a linker and loop L1 regions are not conserved in Cmr2, suggesting different recognition and activation mechanisms in the Cmr complex (Sofos et al., 2020). The presence of NTR in the Cmr-β complex from *Sulfolobus islandicus* (Sis) induced a large conformational change in the unique stalk loop of the Cmr3 subunit, compared with apo structure in an extended conformation (Fig. 1-15 c) (Sofos et al., 2020). This retracted configuration of stalk loop promotes the coordinated Cmr2 displacement, thus resulting in the inactivation of ssDNA cleavage and cOA synthesis. CTR-bound SisCmr-β structures visualise the different configurations of the stalk loop, alternating between an extended and retracted state (Sofos et al., 2020). This dynamic changes in Cmr3 stalk loop seems to allosterically control Cmr2 activities.

Recent studies have determined cryo-EM structures of CTR-bound Csm from *Thermococcus onnurineus* in complex with substrate ATP or its analogues, intermediate pppApApA and product cA_4, providing insights into the cOA synthesis mechanisms (Fig. 1-15 d-i) (Jia et al., 2019a). In this process, two Palm domains of Csm1 specifically accommodate an adenosine ring by forming a hydrogen bond between adenosine with side chain of Ser residues. A single GGDD motif from one of the Palm domains is positioned between two Palm domains. The 3'-OH of the acceptor ATP is activated by the side chain of Asp in the GGDD motif to perform nucleophilic attack on the α-phosphate of the donor ATP, generating a pppApA intermediate with a 3'-5' phosphodiester linkage. The pppApA subsequently occupies the donor position with the α-phosphate attacked by the 3'-OH of the incoming ATP in the acceptor position to produce a pppApApA intermediate. These intermediates can have different numbers (2-6) of AMP and can be cyclised at any states by the 3'-OH of the terminal adenosine, which

intermolecularly attacks the α-phosphate at the 5'-ppp ends. The final cyclic products are eventually released from the channel between Csm1 and Csm4.
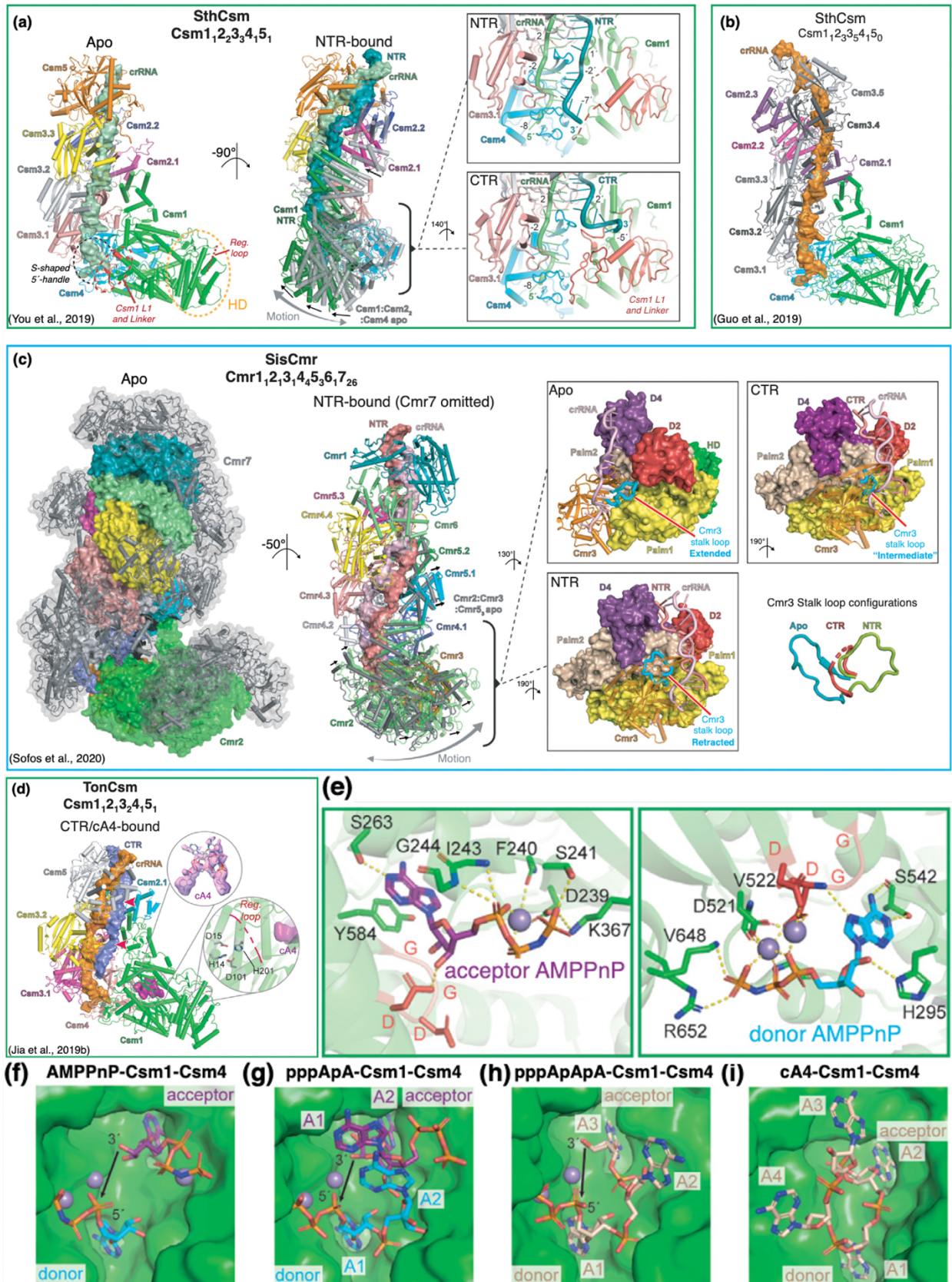
Figure 1-15 Structures of Csm/Cmr complex.

a. Structures of the SthCsm complex (You et al., 2019). The apo SthCsm is shown in a side-view orientation (left) with dashed outline highlighted regions, the 5' tag (black), the Csm1 L1 loop and Linker regions (red), the Csm1 HD domain (yellow). The proposed regulatory loop is indicated by reg loop. The NTR-bound SthCsm is superimposed on the apo structure (grey) and rotated -90°. The displacement of Csm2 and Csm1 is indicated by black arrows upon NTR binding (PDB-6IFN and PDB-6IFL). The inserts highlight close-up views of comparison of NTR (top) and CTP (bottom)-bound structures. **b.** The apo structure of SthCsm (PDB-6NUE) (Guo et al., 2019). **c.** The structure of SisCmr-β (Sofos et al., 2020). The apo structure (left) shows the Cmr1-6 core in surface representation and 13 Cmr7 in transparent surface. The NTR-bound SisCmr is rotated -50°, and superimposed with Cmr2/Cmr3/Cmr5 of apo complex (grey), with black arrow indicating the displacement of Cmr2 and Cmr5 upon NTR binding (PDB-6S6B and PDB-6D8E). Inserts show the close-up views of different configurations of Cmr3 stalk loop in the apo, NTR- and CTR- bound states. **d.** Structures of TonCsm in complex with a CTR and cA$_4$-bound state (PDB-6O7H) (Jia et al., 2019a). Target RNA cleavage sites are indicated by red arrows. Insets are close-up views of the bound cA$_4$ (top) and HD domain (bottom) with catalytic residues in grey sticks and regulatory loop in red cartoon. **e.** Bonding network of acceptor AMPPnP (left) and donor AMPPnP (right) with Palm domain residues. The GGDD motif is colored red and polar interaction is indicated by yellow (PDB-6O74). **f-i.** Structures of the Csm1-Csm4 in complex with either AMPPnP (PDB-6O74), pppApA (PDB-6O75), pppApApA (PDB-6O78), or cA$_4$ (PDB-6O7B). Figure modified from the original of Molina *et al.* (Molina et al., 2020).

1.3.2 Ancillary proteins involved in signalling pathways

Upon detecting target RNA, the Palm domain of Cas10 undergoes allosteric activation to synthesise a range of cOA from ATP. cOA in turn binds to and activates various type III CRISPR ancillary proteins to provide immunity (Fig. 1-16). The section will review experimentally characterised ancillary proteins.

1.3.2.1 Csx1 and Csm6 family ribonucleases

Csm6 in type III-A or Csx1 in type III-B is not only the first but also the most extensively studied CRISPR-associated ancillary proteins, as their encoding genes are frequently found in type III CRISPR operons (Makarova et al., 2014, Athukoralage and White, 2022). Csm6/Csx1 proteins have an N-terminal CARF (CRISPR-associated Rossmann fold) domain and a C-terminal HEPN (higher eukaryotes and prokaryotes nucleotide binding) domain. Initial studies revealed the crucial role of Csm6/Csx1 associated with type III immunity, even though they had no impact on crRNA biogenesis and complex formation (Hatoum-Aslan et al., 2014, Deng et al., 2013). Subsequent *in vivo* investigations indicated that Csm6 contributed to robust immunity by degrading phage transcripts (Jiang et al., 2016b). Furthermore, two independent studies confirmed the ribonuclease activities of *Pyrococcus furiosus* Csx1 and *Thermus thermophilus* Csm6 *in vitro*, which were mediated by the HEPN domain (Sheppard et al., 2016, Niewoehner and Jinek, 2016). The link between ancillary proteins Csx1/Csm6 and type III systems was eventually established after the discovery of cOA synthesised by Csm/Cmr complexes and the CARF domain acting as a cOA sensor (Niewoehner, 2017, Kazlauskiene et al., 2017).

Since then, further studies have shed light on the structure and molecular mechanism of Csx1/Csm6. In general, the HEPN domain, as an RNase effector, is allosterically activated to

degrade RNA non-specifically when the CARF sensory domain binds its cognate cOA (Niewoehner, 2017, Kazlauskiene et al., 2017, Rouillon et al., 2018). Importantly, the regulatory mechanisms differ in each Csm6/Csx1 RNase family. For example, a canonical Csm6 from *Thermococcus onnurineus* forms a symmetric parallel homodimer and undergoes conformational changes upon binding to cA_4 (Jia et al., 2019b). Intriguingly, cA_4 binds to both the CARF and HEPN domains of TonCsm6 and is subsequently cleaved into ApA>p (A_2>p) in the CARF domain and cAMP (A>p) in the HEPN domain. The binding and cleavage of cA_4 to A_4>p within the CARF domain activates RNA cleavage by the HEPN domain, while subsequent cleavage to A_2>p terminates RNase activity.

Another example involves the structural study of *Sulfolobus islandicus* (Sis) Csx1 in complex with cA_4 (Molina et al., 2019). SisCsx1 forms a unique hexamer, consisting of a trimer of dimers. Each dimer is formed by curling two monomers around the twofold axis and three dimers hexamerise through a unique insertion region of HEPN domains. This dimeric unit enables the formation of the cA_4 binding pocket in the CARF domain and the RNase catalytic pocket in the HEPN domain. The hexamer undergoes a conformational change upon cA_4 binding to the CARF domain, activating the RNA cleavage activity in the catalytic pockets.
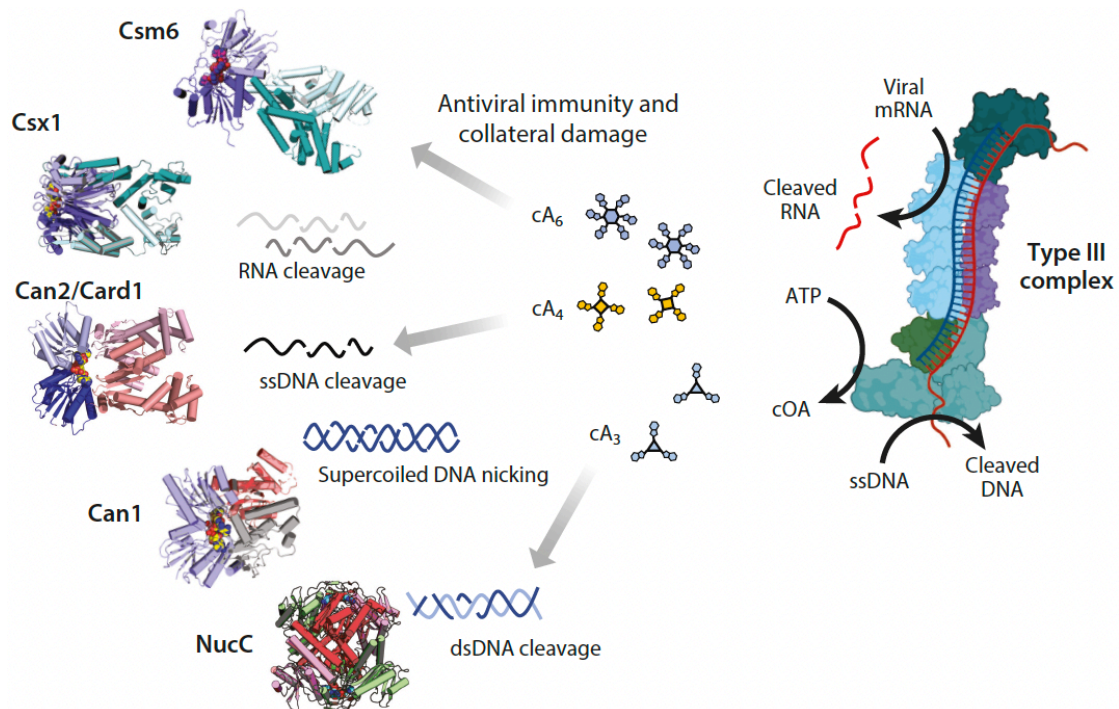
Figure 1-16 Various ancillary proteins associated with Type III CRISPR systems
Upon detecting and binding to foreign RNA, Type III CRISPR systems is activated to cleave invading RNA by Cmr4, produce cOA second messengers by Cas10 subunit PALM polymerase domains and degrade ssDNA by HD nuclease. cOA in turn bind to and activate ancillary proteins that cleave both viral and host nucleic acids. Cas10 returns to an inactive state after target RNA cleavage. Abbreviations: Csm6, Cas subtype Mtube 6; Csx1, cardiac-specific homeobox 1; Can, CRISPR ancillary nuclease; Card1, cyclic-oligoadenylate-activated single-stranded ribonuclease and single-stranded deoxyribonuclease 1; NucC, nuclease, CD-NTase associated; dsDNA, double-stranded DNA; ssDNA, single-stranded DNA. Figure is used from Athukoralage and White (open access with unrestricted reuse, distribution and reproduction) (Athukoralage and White, 2022).

1.3.2.2 Can1 and Can2/Card1 nucleases

Comparative genome analysis reveals that the CARF sensor domain is often fused with various enzymatic domains, including nucleases, proteases, or transmembrane domains, which complement or extend the type III defence process (Makarova et al., 2020a, Shmakov et al., 2018, Shah et al., 2019). Recent studies have characterised several CARF-containing ancillary proteins, offering valuable insights into their diversity (Fig. 1-16). CRISPR ancillary nuclease 1 (Can1), found within the CRISPR locus of *Thermus thermophilus*, functions as a cA_4-activited DNA nuclease, instead of an RNase (McMahon et al., 2020). Can1 exists as a unique monomer, comprising two CARF domains separated by a nuclease-like domain, a C-terminal PD-D/ExK nuclease domain. The activator cA_4 binds to the interface between two CARF domains, adopting a fused dimer conformation similar to the dimeric Csx1/Csm6 family proteins. Upon cA_4 binding, structural rearrangements of the two nuclease domains activate Can1 to randomly nick supercoiled DNA. This nicking activity is believed to interfere with viral DNA replication by causing the collapse of DNA replication forks in rapidly replicating phages.

Can2, closely related to Can1, features an N-terminal CARF domain and a C-terminal PD-D/ExK nuclease domain, forming a homodimer similar to Csx1/Csm6 (Fig. 1-16) (Zhu et al., 2021). Can2 functions as an unusual cA_4-activated nuclease that non-specifically degrades both supercoiled dsDNA and ssRNA. It provides effective immunity against both plasmid transformation and phage infection in *E. coli*. Card1 (cyclic oligoadenylate-activated single-stranded ribonuclease and single-stranded deoxyribonuclease 1), an orthologue of Can2, is activated by cA_4 to cleave both ssRNA and ssDNA but not dsDNA (Rostol et al., 2021). Activation of Card1 provide defence against plasmids and phage infection by inducing cell dormancy.

1.3.2.3 Other ancillary effectors

SAVED (second messenger oligonucleotide or dinucleotide synthetase-associated and fused to various effector domains) is another predicted sensory domain for signal molecules and is strongly associated with synthetases in CBASS systems that produce cyclic 2'-5' GMP-AMP and 2'-5' oligoadenylates (Makarova et al., 2020a). The SAVED domain is proposed as a divergent version of the CARF domain, often fused with a range of effector domains, despite limited sequence similarity between them (Makarova et al., 2020a, Shmakov et al., 2018). Recent studies have begun to illustrate the potential link between type III CRISPR systems and ancillary proteins containing SAVED domains. One such example is a TIR-SAVED, which

consists of a Toll/interleukin-1 receptor (TIR) domain fused with SAVED domain found in the type II-C CBASS operon (Hogrel et al., 2022). The activator cA_3 binds to SAVED domain, inducing the formation of extended filament, and activating the adjacent TIR domain to degrade NAD^+. The activation of TIR-SAVED in the context of with type III CRISPR system capable of producing cA_3 provides plasmid immunity. However, natural examples of ancillary proteins containing SAVED domain in type III CRISPR systems still require further investigation.

Aside from CARF and SAVED proteins, NucC (Nuclease, CD-NTase associated) features a restriction endonuclease-like fold, also functioning as a cA_3-activated endonuclease, initially studied in the context of CBASS systems (Fig. 1-16) (Lau, 2020). NucC is activated by cA_3 to form a homohexamer through the assembly of pairs of homotrimers, allowing for nonspecific dsDNA degradation. Further genomic analysis has identified 31 genes encoding NucC within type III CRISPR *loci*. One of the CRISPR-associated NucC from *Vibrio metoecus* has been shown to possess non-specific dsDNA degradation activity through sensing and binding cA_3 (Lau, 2020, Gruschow et al., 2021). The activation of NucC within CBASS systems leads to cell death through genome degradation, consistent with the pattern of CBASS immunity via abortive infection (Millman et al., 2020b). Further investigations have demonstrated that type III CRISPR systems associated with NucC provide immunity against nucleus-forming jumbo phages, also through abortive infection (Mayo-Munoz et al., 2022). The intrinsic characteristic of NucC, which enables it to degrade host genomes, confers the effective population-level protection.

1.3.3 Ring nucleases: host regulators or viral anti-CRISPR

CRISPR ancillary proteins, activated by cOA, cleave nucleic acids in a non-specific manner, aiding the host in rapidly countering invasions. However, the continual activation of ancillary proteins by existing signal molecules can be toxic to the host, even though cOA production is inactivated once target RNA is cleared. Thus, signalling pathway needs to be appropriately regulated. Recent research has identified cellular ring nucleases and self-limiting ancillary proteins capable of degrading cOA molecules, thereby halting or regulating signal transduction pathways (Fig. 1-17) (Athukoralage and White, 2022). Interestingly, phages have also evolved strategies to recruit ring nucleases, enabling them to overcome signalling-mediated defence (Fig. 1-17) (Athukoralage and White, 2022).

1.3.3.1 Self-limiting nucleases

Certain Csx1/Csm6 family proteins exhibit self-limiting activities. For example, the CARF domain serves dual functions, not only binding activators to trigger RNase activity in the HEPN domain, but also degrading its activator into $A_2>P$ products to switch off immunity (Athukoralage et al., 2019). As mentioned previously, the self-limiting CARF family protein TonCsm6 deactivates itself through stepwise degradation of activator cA_4, mediated by both the CARF and HEPN domains (Jia et al., 2019b). Similar self-regulatory mechanisms have also been identified in cA_6-driven signalling pathways, such as EiCsm6 from *Enoroccocus italicus* (Garcia-Doval et al., 2020) and StCsm6 from *Streptococcus thermophilus* (Smalakyte et al., 2020). This coordinated self-regulation enhances invader clearance and protects the host from self-toxicity.

1.3.3.2 Cellular ring nucleases

The first identified cellular ring nuclease family, Crn1 (CRISPR-associated ring nuclease 1), features a canonical CARF domain from the crenarchaeote *Sulfolobus solfataricus*. Crn1 forms a dimeric architecture and specifically cleaves cA_4 into final linear $A_2>p$ products (Athukoralage et al., 2018). Further studies have characterised the unrelated Crn2 ring nucleases and distantly related CARF family proteins Crn3/Csx3 (Brown et al., 2020, Athukoralage et al., 2020c, Samolygo et al., 2020). Crn3 from *Archaeoglobus fulgidus* specifically degrades cA_4 through active sites formed by the assembly of pairs of dimers (Athukoralage et al., 2020c). Notably, Crn2 is observed fused to the C-terminus of the Csx1 family ribonuclease found in type III CRISPR systems from *Marinitoga piezophile* (Samolygo et al., 2020). The ring nuclease activity of Crn2 regulates the cA_4-activitated RNA cleavage activity of Csx1.

1.3.3.3 Viral ring nucleases

Further studies have uncovered a viral ring nuclease, AcrIII-1, in the *Sulfolobus* virus *S. islandicus* rod-shaped virus 1 (Athukoralage et al., 2020b). AcrIII-1, homologous to Crn2, rapidly degrades cA_4, catalysing the reaction around 50-fold faster than cellular Crn1 ring nuclease, despite exhibiting similar cA_4 binding affinity (Athukoralage et al., 2020b, Athukoralage et al., 2020a). A kinetic model has demonstrated that AcrIII-1 is capable of swiftly degrading cA_4 over a wide concentration range, efficiently limiting the signalling-mediated immunity (Athukoralage et al., 2020a).
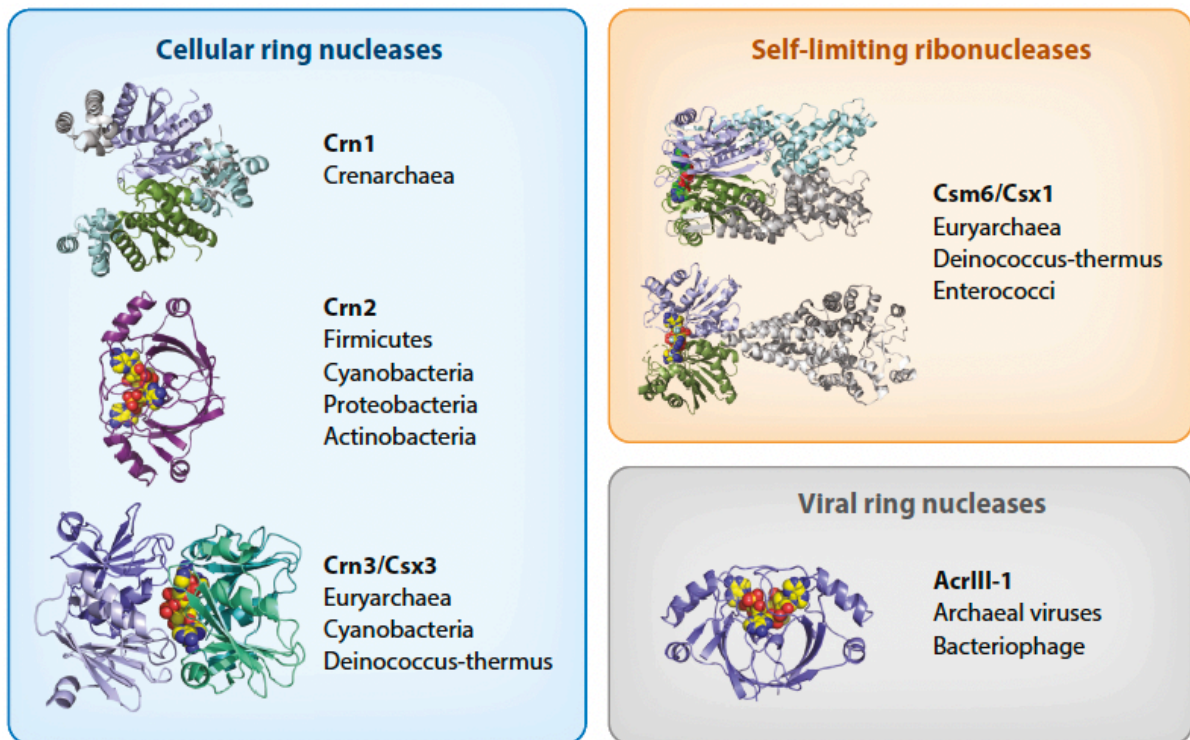
Figure 1-17 Ring nucleases
Three experimental characterised ring nucleases include cellular ring nuclease (Crn1, 2 and 3), self-limiting Csm6/Csx1 ribonucleases, cleaving their cognate activators, and the viral ring nucleases AcrIII-1, which is homologous to Crn2. Abbreviations: AcrIII-1, anti-CRISPR III-1; Crn, CRISPR ring nuclease; Csm6, Cas subtype Mtube 6; Csx, cardiac-specific homeobox. Figure is used from Athukoralage and White (open access with unrestricted reuse, distribution and reproduction) (Athukoralage and White, 2022).

Overall, the capacity to synthesise the signal molecules as second messenger is one of the most unique features in type III CRISPR systems. Various ancillary proteins are activated by these second messengers to confer diverse immunity. Beyond collateral nuclease activities, many ancillary proteins are found and predicted to function in different mechanisms. Further investigating will expand our knowledge on type III CRISPR-mediated signalling pathways.

1.4 Objectives

The large and increasing number of characterised ancillary proteins demonstrates the diversity and complexity of signalling-associated immunity in type III CRISPR systems. Through bioinformatic analysis, a significant number of ancillary genes have been identified within the type III CRISPR loci. It is worth noting that ancillary proteins lacking the CARF domain exhibit more diversity and membrane proteins represent a substantial portion of these ancillary proteins (Shmakov et al., 2018, Shah et al., 2019). However, their precise functions have remained undetermined.

We set out to investigate two type III CRISPR systems, one associated with a CorA family membrane channel and the other with a Lon protease. CorA proteins are the most observed membrane proteins encoded in type III-B CRISPR-Cas loci (Shmakov et al., 2018). In many instances, their corresponding genes are located adjacent to those encoding DHH family phosphodiesterase NrN, or are sometimes fused together in certain species, suggesting a potential functional connection (Shmakov et al., 2018). Specifically, we have characterised a CorA-associated type III-B CRISPR system from *Bacteroides fragilis*, a human gut bacterium. CHAPTER 3 is dedicated to illustrating the function of the BfrCmr complex both *in vivo* and *in vitro*. This chapter also delves into crRNA biogenesis, identification of the signal molecule produced by this system, and the biochemical characterisation of three ancillary proteins.

Additionally, we have investigated a CalpL-associated CRISPR signalling pathway, in which CalpL comprises a Lon protease and a SAVED4 domain, from the thermophilic bacterium *Sulfurihydrogenibium* spp. YO3AOP1. Our collaborators Gregor Hagelueken, Christophe Rouillon, and Niels Schneberger have characterised the function and structure of CalpL. Meanwhile, we set out to elucidate the functions of the other two ancillary proteins CalpS and CalpT, which are encoded alongside *calpL* in the same type III-B CRISPR operon. These findings are presented in CHAPTER 4 for details.

2 Materials and methods

2.1 Construction and purification

Genes encoding BfrNrN, BfrNYN, BfrCorA, BfrCas6, CboSAM-AMP lyase, truncated CaplT and CalpS were identified, and codon optimised by Prof Malcolm F. White (University of St Andrews, Scotland, UK). Dr Sabine Grüschow (University of St Andrews, Scotland) gave instruction and advice for the construction of BfrCmr and CRISPR arrays expression plasmids. Construction, expression, and purification of CboSAM-AMP lyase was carried out by Dr Shirley Graham (University of St Andrews, Scotland). The CalpT expression plasmid pET11a-CalpT was from our collaborator Dr Gregor Hagelueken (University of Bonn, Bonn, Germany). The plasmids, DNA and protein sequences used in this thesis were listed in the appendices and plasmid maps were generated by using software SnapGene.

2.1.1 Construction of BfrCmr effector complex expression plasmid

The pACE-based BfrCmr synthetic expression plasmid pBfrCmr1-6 was designed to contain six codon-optimised genes for expression in *Escherichia coli* (*E. coli*) including *cmr1* to *cmr6* that encode each subunit of the BfrCmr complex, and Cmr3 with a N-terminal polyhistidine tag for purification. The sequence of pBfrCmr1-6 was divided into five overlapping segments with similar length (designated as BfrCmr a, b, c, d, and e, listed in Appendix A) and purchased from Twist Biosciences. These segments were amplified by PCR and assembled into pBfrCmr1-6 through NEBuilder® HiFi DNA Assembly Master Kit (primers used for PCR amplification shown in Table 2-1). The obtained plasmid pBfrCmr1-6 was verified by digestion and sequencing (GATC Biotech, Eurofins Genomics, Germany. Fig. 2-1). The BfrCmr variants with D27A of BfrCmr4 and cyclase mutant (D328A:D329A), D70N, E151R and D70N/E151R of BfrCmr2 (Cas10) were generated by PCR using Phusion polymerase (Thermo Scientific) with pBfrCmr1-6 as a template in the presence of two overlapping primers containing the target mutations (primers for mutagenesis shown in Table 2-1). The correct variants were confirmed by sequencing.

Table 2-1 Primers used for construction of pBfrCmr1-6 and its variants

Name	Sequence (5'-3')	Note
BfrCmrSG1-F	CCAGACGTACCTGCCGGCATTCTTC	Forward primer
BfrCmrSG1-R	CGATTTCATTGATGCTTTCGATATTGAAGG	Reverse primer
BfrCmrSG2-F	GCGCTGTTGTCCTTCAATATCGAAAGC	Forward primer
BfrCmrSG2-R	CGATTGAATCCGACCACATAAAGTTAC	Reverse primer
BfrCmrSG3-F	CTGTCAGCTTTGTGTAACTTTATGTGG	Forward primer
BfrCmrSG3-R	CCCGAAATGGTTATTGAACGCGGCAAC	Reverse primer
BfrCmrSG4-F	CGAATCGCCTCTGGTTGCCGCGTTC	Forward primer
BfrCmrSG4-R	CTATTTCAGCAATCTGTCTATTTCGTTCG	Reverse primer
BfrCmrSG5-F	GCAACTATGGACGAACGAAATAGAC	Forward primer
BfrCmrSG5-R	GAAGTTATGACAGATGAAGAATGCCG	Reverse primer
BfrCmr4_D27A-F	CGGAGTTATTGcTAACTTGATCCAACGTGAC	Mutagenesis
BfrCmr4_D27A-R	GGATCAAGTTAgCAATAACTCCGTAGTTCACC	Mutagenesis
BfrCmr2_cyclase-F	CATTGGAGGGGCCGCTTTGCTTTGTTTTGCGC	Mutagenesis
BfrCmr2_cyclase-R	GCAAAGCGGCCCCTCCAATGAAGATCGGCTTTC	Mutagenesis
BfrCmr2_D70N-F	GCAGGGTTGTTTCCCAACCGTTATATCTTCAAG	Mutagenesis
BfrCmr2_D70N-R	CTTGAAGATATAACGGTTGGGAAACAACCCTGC	Mutagenesis
BfrCmr2_E151R-F	GTGAAAAGTACCTGAACATTATTAGAAATCAGGAGAC	Mutagenesis
BfrCmr2_E151R-R	GTCTCCTGATTTCTAATAATGTTCAGGTACTTTTCAC	Mutagenesis
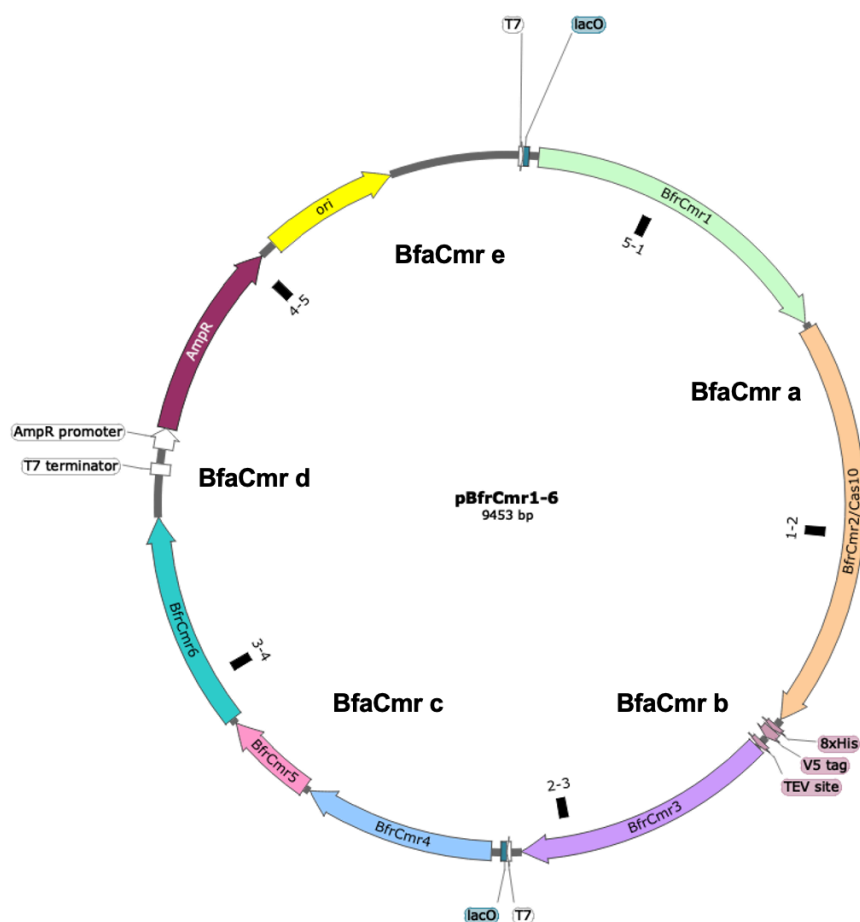
Figure 2-1 Design and construction of the BfrCmr expression plasmid pBfrCmr1-6
Overlapping sequences are indicated by black squares. The map was generated by SnapGene.

2.1.2 Construction of BfrCRISPR RNA and Cas6 expression plasmid

For the construction of BfrCRISPR RNA over-expression vector, the codon-optimised BfrCas6 gene was purchased as g-block from Integrated DNA Technologies (IDT) and inserted into the *Nde*I and *Xho*I restriction sites in MCS-2 of the vector pCDFDuet™-1 (Novagen, Missouri, USA). The synthetic gene of CRISPR pre-array with two CRISPR repeats and two divergent *Bpi*I sites between two repeats for spacer sequence insertion was cloned into 5'-*Nco*I and 3'-*Sal*I sites in MCS-1 of pCDFDuet containing BfrCas6. Designed spacer targeting the portion of tetracycline resistance gene or targeting pUC19 LacZ was annealed and constructed into the *Bpi*I sites of CRISPR pre-array to obtain the plasmid, designated as pBfrCRISPR_Tet or pBfrCRISPR_pUC for later *in vivo* or *in vitro* assay (Fig. 2-2A. Primers listed in Table 2-2). A CRISPR array, consisting of one spacer targeting the gene encoding Late Promoter Activating protein (Lpa) of phage P1, flanked by two BfrCRISPR repeat sequences, was assembled from annealing of primers Bfr-rep-5p-T, Bfr-rep-5p-C, Bfr-rep-3p-T, Bfr-rep-3p-C, Bfr-sp-phageLPA-T and Bfr-sp-phageLPA-C (Table 2-2). The array was then ligated into MCS-1 of pCDFDuet containing cas6 in MCS-2 to give pBfrCRISPR_Lpa. The successful constructs were confirmed by digestion and sequencing. Gene sequences of BfrCas6 and CRISPR pre-array were listed in Appendix A.

The BfrCas6 expression plasmid pEHisV5TEV-BfrCas6 was constructed by insertion of *cas6* synthetic gene into the plasmid pEHisV5TEV between the *Nco*I and *BamH*I restriction sites (Fig. 2-2B). The expression plasmid pEHisV5TEV contains kanamycin resistance marker for antibiotic selection and genes encoding eight histidines, followed by a V5 epitope tag for western blotting, as well as a spacer and the cleavage site of Tobacco Etch Virus (TEV) protease. The obtained plasmid pEHisV5TEV-BfrCas6 was confirmed by sequencing.

Table 2-2 Sequences of primers used for construction of CRISPR RNA expression plasmids

Name	Sequence (5'-3')	Note
Spacer_pUC-F	AGACGAATTCGAGCTCGGTACCCGGGGATCCTCTAG	Primer
Spacer_pUC-R	ACATCTAGAGGATCCCCGGGTACCGAGCTCGAATTC	Primer
Spacer_TetR-F	AGAC TGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGA	Primer
Spacer_TetR-R	ACAT TCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCA	Primer
Bfr-rep-5p-T	CATGGAATAGTAATCTGATTATCAATAT	Primer
Bfr-rep-5p-C	ATTATACTGGAATACATCTACATATATTGATAATCAGATTACTATTC	Primer
Bfr-rep-3p-T	ATGTAGATGTATTCCAGTATAATAAGGATTAAGACTTAAATAGAG	Primer
Bfr-rep-3p-C	TCGACTCTATTTAAGTCTTAATCCTT	Primer
Bfr-sp-phageLPA-T	ATGTAGATGTATTCCAGTATAATAAGGATTAAGACATTCGTGAGTGA TTTATTTCCATGAAGTGGCGTCCCT	Primer
Bfr-sp-phagLPA-C	ATTATACTGGAATACATCTACATAGGGACGCCACTTCATGGAAATAA ATCACTCACGAATGTCTTAATCCTT	Primer
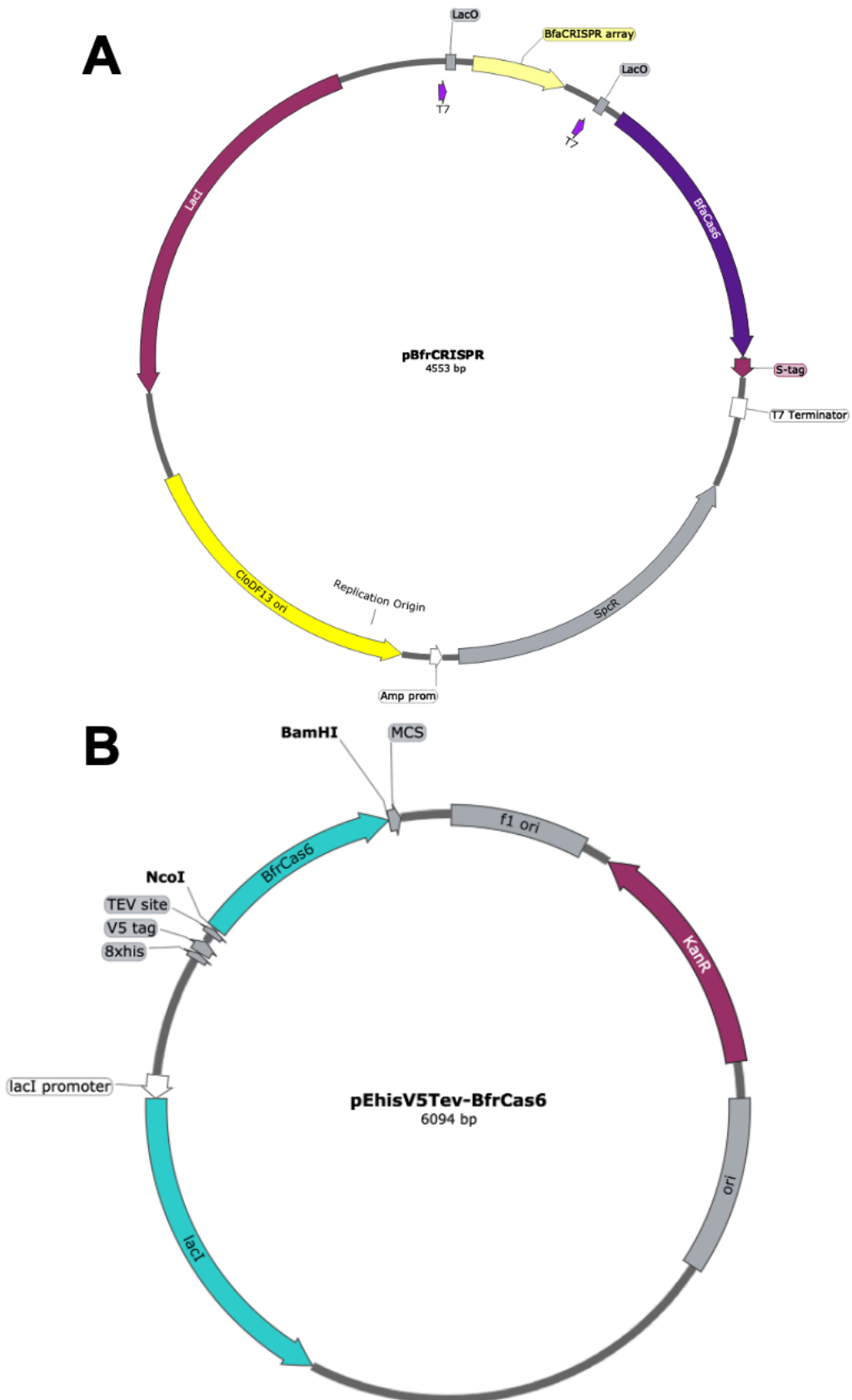
Figure 2-2 Design and construction of the CRISPR RNA expression plasmids
pBfrCRISPR (A) and BfrCas6 expression plasmid pEHisV5TEV-BfrCas6 (B).

2.1.3 Construction of ancillary proteins expression plasmid

The synthetic genes encoding membrane protein CorA, phosphodiesterase NrN and nuclease NYN from *B. fragilis* and SAM lyase from *Clostridium botulinum* had been codon optimised to express in *E. coli* and purchased from IDT. Genes were cloned between the *Nco*I and *BamH*I restriction sites of vector pEHisV5TEV (Sequences of synthetic genes listed in Appendix A). Truncated membrane protein CorA^{tr} (aa 1-428), inactive variants NrN$^\Delta$ (D85A/H86A/H87A) and NYN variants (D13A and D72A) were constructed by PCR using Q5$^®$ High-Fidelity DNA Polymerase (New England Biolabs (NEB)) with plasmids constructed above as templates and two overlapping primers containing the target mutations (primers for mutagenesis shown in Table 2-3). All successful constructs were verified by enzymatic digestion and sequencing.

To construct pRATDuet-based plasmids for *in vivo* assay. Plasmid pRATDuet was constructed by Dr Sabine Grüschow as described in Athukoralage *et al.*(Athukoralage et al., 2020b). For single ancillary protein expression, the synthetic gene encoding BfrCorA, BfrNrN, BfrNYN or their variants was inserted between *Nco*I and *EcoR*I restriction sites in MCS-1 under control of pBAD promoter (Fig. 2-3A). The synthetic gene of NrN or its variants was cloned into *Nde*I and *Xho*I sites in MCS-2 of pRATDuet containing *corA* in MCS-1 for two effectors co-expression (Fig. 2-3B).

Table 2-3 Sequences of primers used for mutagenesis

Name	Sequence (5'-3')	Note
CorA^{tr}-F	GGCGTCCTAGTTGAATGATATTGCAACTCTTTTCC	Mutagenesis
CorA^{tr}-R	GCAATATCATTCAACTAGGACGCCTTTTTGTTGCG	Mutagenesis
NrN$^\Delta$-F	TCGCCGCCGCCAATGAGTATGCCACGTATCCAAGTG	Mutagenesis
NrN$^\Delta$-R	CATACTCATTGGCGGCGGCGATACGAATGTAATTGGTTGGAGG	Mutagenesis
NYN^{D13A}-F	CGTCAATTGGAATTTTCATTGCTGGAGGCTACTTTACC	Mutagenesis
NYN^{D13A}-R	GGTAAAGTAGCCTCCAGCAATGAAAATTCCAATTGACG	Mutagenesis
NYN^{D72A}-F	GCGCTACCGCGTGAACGCTGCCAACAACAAGCACC	Mutagenesis
NYN^{D72A}-R	GGTGCTTGTTGTTGGCAGCGTTCACGCGGTAGCGC	Mutagenesis
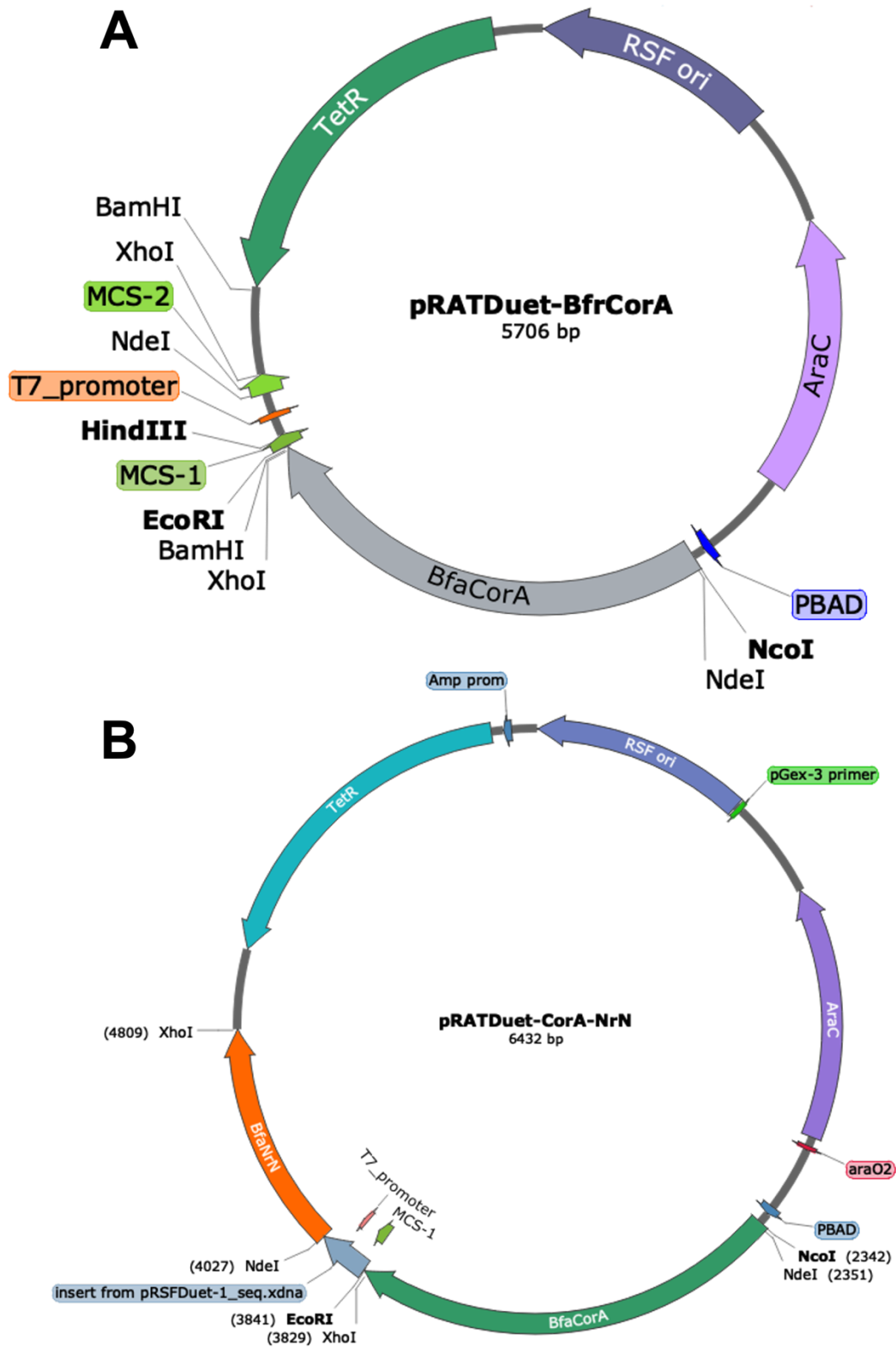
Figure 2-3 Design and construction of pRATDuet-based plasmid
A. Sigle ancillary protein expression plasmid map. Synthetic gene encoding BfrCorA was inserted between *Nco*I and *Eco*RI sites in MCS-1 of pRATDuet, which is shown as an example. **B.** Ancillary proteins co-expression plasmid map (generated by SnapGene.).

2.1.4 Expression and purification of BfrCmr complex and its variants

E. coli BL21 Star (DE3) competent cells (Invitrogen) were co-transformed with both constructs of BfrCmr complex pBfrCmr1-6 or its variants and of CRISPR RNA pBfrCRISPR_Lpa. After overnight incubation at 37 °C, a single colony was selected and grown in the LB (lysogeny broth) liquid media containing 100 µg/ml ampicillin and 50 µg/ml spectinomycin at 37 °C with overnight shaking. Next day, overnight culture was 100-fold diluted into fresh LB containing equivalent antibiotics (2 L in total), incubating at 37°C with shaking until OD_{600} was between 0.6-0.8. The expression was then induced with 0.2 mM isopropyl-β-D-1-thiogalactoside (IPTG) and continued growing at 25 °C overnight with shaking at 180 rpm. The cell pellet was harvested by centrifuging at 5,000 rpm at 4 °C for 15 min and then frozen at -70 °C until needed. For purification, the cell pellet was resuspended in the lysis buffer A (50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 20 mM imidazole, and 10 % glycerol) with an additional EDTA-free protease inhibitor (Roche, Switzerland) and 1 mg/ml lysozyme (Sigma-Aldrich, Missouri, USA), followed by sonication for six times 1 min with 1 min interval rest. Lysed cells were ultracentrifuged at 40,000 rpm for 30 min at 4 °C (70 Ti rotor, Beckman Coulter Optima L-90K) to spin down debris and unbroken cells. The clear supernatant was filtered with 0.45 µM filter and then loaded onto a 5 mL HisTrap FF column (GE Healthcare) pre-equilibrated with lysis buffer. The bound his-tagged target protein was washed with lysis buffer and eluted in a gradient elution with increasing the concentration of imidazole to 250 mM. Fractions containing the target protein were then pooled and concentrated using a 30 kDa molecular mass cut-off centrifugal filter (Merck Amicon™ Ultra-15) (Fig. 3-6A). The concentrated target protein was dialysing overnight at room temperature to reduce the concentration of imidazole in the dialysis buffer A (20 mM Tris-HCl pH 7.5, 0.25 M NaCl and 10 % glycerol). If needed, the TEV protease (1 mg/10 mg protein) was incubated with target protein to remove the his-tag during dialysis. The TEV-cleaved protein was separated from the uncleavable his-tagged TEV protease using the HisTrap FF column. The target protein was further purified by size exclusion chromatography (SEC, HiLoad® 16/600 Superdex® 200 prep grade, Cytiva, Massachusetts, USA) in the SEC buffer A (20 mM Tris-HCl pH 7.5, 0.25 M NaCl, 10 % glycerol and 1 mM DTT) (Fig. 3-6 B). Fractions containing the target protein from SEC were pooled and concentrated using centrifugal filter, followed by flashing frozen aliquots and stored at -70 °C.

2.1.5 Expression and purification of BfrCas6 and ancillary proteins

E. coli C43 (DE3) cells were transformed with the constructs containing the genes encoding BfrCas6, BfrNrN, BfrNYN and CboSAM-AMP lyase. These proteins followed the same expression and purification steps. For expression, single recombinant *E. coli* strain for each protein was selected and grown overnight in the LB containing 50 µg/ml kanamycin at 37 °C. Overnight cell culture was 100-fold diluted into fresh LB with the same antibiotic (2 L in total for each protein) and cultivated at 37 °C until OD_{600} was reached 0.6-0.8, followed by induction with 0.2 mM IPTG at 18 °C for 16 h. The cell pellets were collected and stored at -70 °C until needed.

The purification steps were similar as for the BfrCmr complex, except that BfrCas6 kept its tag so purification steps of incubation with TEV protease and dialysis were omitted. Briefly, the clear cell lysate was loaded onto a HisTrap FF column pre-equilibrated with lysis buffer A, and the bound his-tagged protein was eluted through gradient elution buffer A (50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 0.5 M imidazole, and 10 % glycerol) (Fig. 3-4A (BfrCas6), 3-23A (BfrNrN wild type and its variant) and 3-28 A, D and G (BfrNYN wild type and its variants)). Fractions containing the target protein were analysed by SDS-PAGE and then pooled to dialyse overnight with TEV protease removing the his-tag in the dialysis buffer A. The TEV-cleaved target proteins were then recovered using the HisTrap FF column for the second time. Proteins were further purified using SEC column (Fig. 3-4B (BfrCa6), 3-23B (BfrNrN wild type and its variant) and 3-28 B, E and H (BfrNYN wild type and its variants)), and their identities and purity were confirmed by SDS-PAGE. Aliquots of concentrated proteins were flash frozen and stored at -70 °C. Subsequently, a series of BfrNrN and BfrNYN variants were expressed and purified following the same procedure as the wild types.

2.1.6 Expression and purification of membrane protein BfrCorA and its variant

E. coli C41 (DE3) strain containing the membrane protein expression plasmids was grown in 2 L LB with 50 µg/mL kanamycin at 37 ºC to an OD_{600} of 0.6-0.8. 0.2 mM IPTG was then added to induce the expression of protein BfrCorA. The cells were grown for an additional 18 h at 18 ºC with the shake of 180 rpm, harvested by centrifugation and stored at -70 ºC.

For purification, the pellet was resuspended in the ice-cold lysis buffer B (50 mM HEPES, pH 7.5, 250 mM NaCl, 5 % glycerol and 10 mM imidazole) with an additional EDTA-free protease inhibitor and lysozyme and then lysed by cell disruptor (Constant System) at 30 psi. The

unbroken cells and debris were spun down at 20,000 g (JLA 25.50 rotor) for 10 min. The membranes in the supernatant were then collected by ultra-centrifugation at 41,000 rpm at 4 ºC for 2 h (70 Ti rotor). The membrane pellet was resuspended in the lysis buffer B with additional 1 % DDM (n-dodecyl β-D-maltoside, GLYCON Biochemicals) and incubated at 4 ºC overnight, followed by centrifugation again at 40,000 rpm at 4 ºC for 30 min. The target protein in the clear supernatant was isolated by using the 5 ml HisTrap FF column with lysis buffer and elution buffer B (50 mM HEPES, pH 7.5, 250 mM NaCl, 5 % glycerol, 250 mM imidazole) with addition of 0.1 % DDM (Fig. 3-17A). The fraction containing target protein was analysed by SDS-PAGE and collected to be further purified by SEC (Superose® 6 Increase 10/300 GL, Cytiva, Massachusetts, USA) using the SEC buffer B (20 mM HEPES, pH 7.5, 150 mM NaCl, 5 % glycerol) with an additional 0.03 % DDM (Fig. 3-17B). Purified target protein was flash frozen and stored at -70 ºC.

2.1.7 Construction, expression, and purification of CalpS and truncated CalpT

The gene of CalpS and truncated CalpT (CalpT^{tr}, aa 1-173) were codon-optimized and purchased from IDT as a G-Bock with flanking restriction sites for cloning (Sequences are listed in Appendix A). *CalpS* and *CalpT^{tr}* were constructed into *Nco*I and *BamH*I restriction sites of vector pEHisV5TEV, allowing expressed proteins with an N-terminal polyhistidine-TEV tag (Fig. 2-7A and B). All successful constructs were verified by restriction-digestion and sequencing.

For expression, *E. coli* C43(DE3) cells transformed with constructs were incubated at 37 °C with shaking at 180 rpm until OD_{600} was reached between 0.6 and 0.8. The cell culture was grown at 16 °C overnight after inducing with 0.2 mM IPTG. The cell pellet was collected and stored at -70 °C.

For purification, the cell pellet was resuspended into lysis buffer A and lysed by sonication. The cleared cell lysate was loaded onto a 5mL HisTrap FF column pre-equilibrated with lysis buffer A. The his-tagged proteins were eluted in a linear gradient with elution buffer A (Fig. 4-2A and Fig. 4-5C). The his-tag was removed by incubating with TVE protease during overnight dialysis at room temperature. The TEV-cleaved proteins were recovered by a HisTrap FF column again and further purified by size-exclusion chromatography in the SEC buffer A (Fig. 4-2B and Fig. 4-5D). Truncated CalpT had further purification over a HiTrap Heparin column (Cytiva) with a NaCl gradient from 10 to 500 mM in the Heparin buffer of 20 mM Tris-HCl, pH 7.0, 10 % glycerol (Fig. 4-2C). The identity of proteins was evaluated on

the SDS-PAGE at each purification step (Fig. 4-2D and Fig. 4-5B). Aliquots of concentrated proteins were flash-frozen and stored at -70 °C.
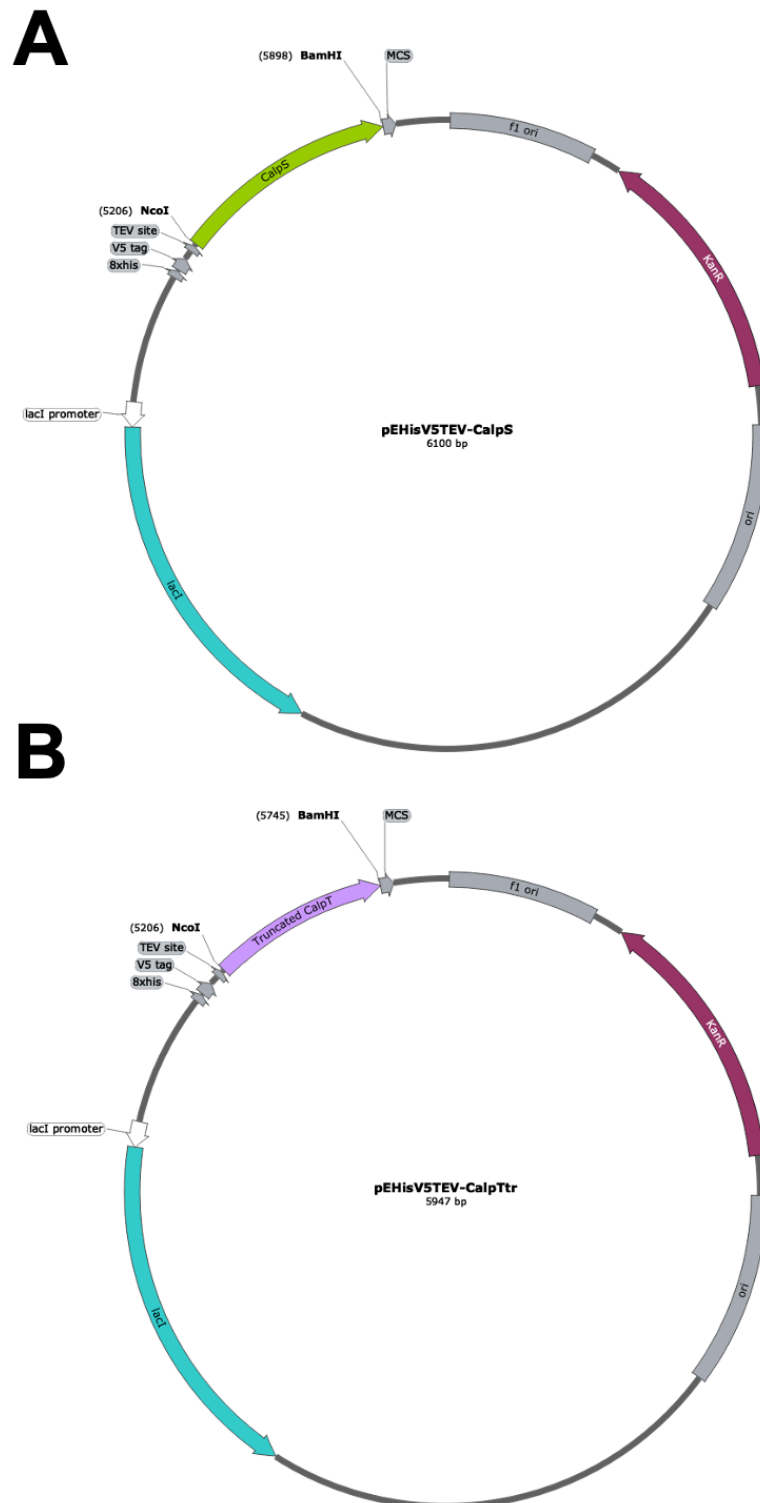
Figure 2-4 Expression constructs of CalpS and truncated CalpT
A, and B. CalpS and CalpT^{tr} expression plasmid map, respectively. Synthetic genes encoding CalpS and CalpT^{tr} were inserted between *Nco*I and *BamH*I sites of pEHisV5TEV.

2.1.8 Co-expression and co-purification of CalpS and CalpT

For the co-expression of the his-tagged CalpS with CalpT, the fragment of CalpT from pET11a-CalpT (Fig. 2-8C) was constructed into the *Nde*I and *Xho*I sites of MCS-2 of vector pCDFDuet™-1 (Fig. 2-8A, the sequence of plasmid pET11a-CalpT is listed in Appendix A). The successful construct was confirmed by sequencing. The *E. coli* C43(DE3) cells were transformed with both constructs pEHisV5TEV-CalpS and pCDFDuet-CalpT. A single transformant was inoculated into LB medium plus antibiotics (50 µg/ml kanamycin and 50 µg/ml spectinomycin) for overnight cultivation at 37 °C. Overnight culture was 100-fold diluted into fresh LB (2 L in total) and grown at 37 °C with shaking at 180 rpm until OD_{600} was between 0.6-0.8. The expression was induced with 0.2 mM IPTG and continued overnight growth at 16 °C. The co-purification of his-tagged CalpS with CalpT was followed the same purification procedure as of CalpS, and CalpT^{tr}, except for the his-tag removal steps. Only first immobilized metal affinity chromatography (IMAC) was performed before size-exclusion chromatography to keep his tag. Concentrated proteins were flash frozen and stored at -70 °C (Fig. 4-7A, B and C).

For the co-expression of his-tagged CalpT with CalpS, a g-Block of CalpS flanking the *Nco*I and *BamH*I sites was constructed into the MCS-1 of vector pCDFDuet™-1 (Fig. 2-8B). *E. coli* C43(DE3) cells were co-transformed with both constructs of pET11a-CalpT and pCDFDuet-CalpS. The cell culture was induced by 0.2 mM IPTG while OD_{600} was between 0.6-0.8 and grown overnight at 16 °C with shaking at 180 rpm. The purification procedures are the same as for the CalpS using his-tagged CalpT to pull down CalpS (Fig. 4-7D, E and F).
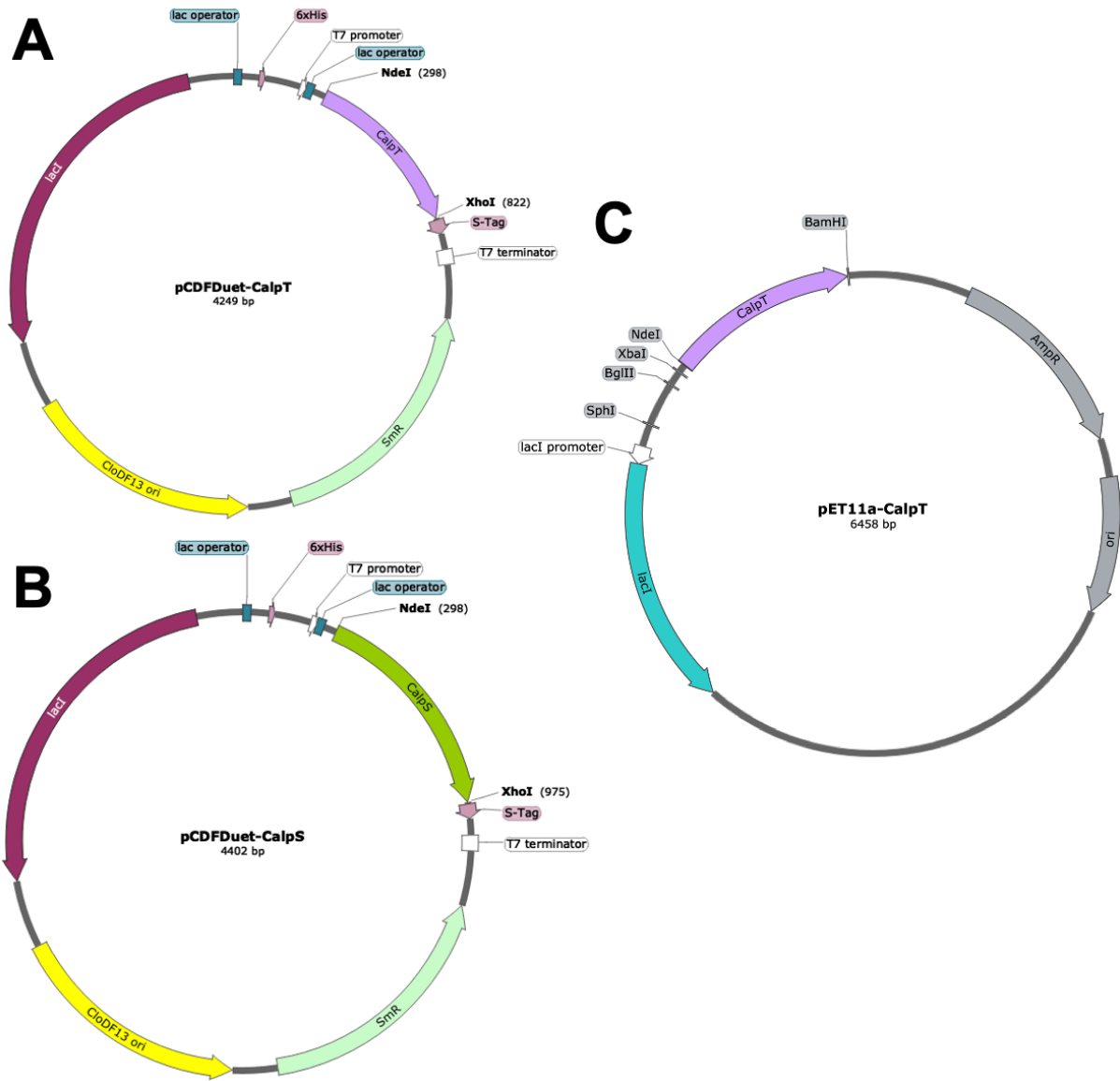
Figure 2-5 Expression constructs and the Co-purification of CalpS and CalpT
A, and B. CalpS and CalpT expression plasmid map, respectively. Synthetic gene encoding CalpT was inserted into *Nde*I and *Xho*I of pCDFDuet and CalpT from pET11a-CalpT was inserted between *Nco*I and *BamH*I sites of pCDFDuet. **C.** Plasmid map pET11a-CalpT.

2.2 Biochemical investigation

LC-MS/MS was conducted by Dr Sally Shirran (University of St Andrews, Scotland, UK). Purified CalpL and CalpT were obtained from our collaborator Dr Gregor Hagelueken and Niels Schneberger (University of Bonn, Bonn, Germany). The images of acrylamide gel were analysed by using software Fiji. The figures of HPLC and protein purification were generated by using software Prism and structures of proteins and molecules were generated by using software PyMOL and ChemDraw, respectively.

2.2.1 BfrCas6 nuclease assay

Nuclease activity of Cas6 was assayed by incubating 1.2 µM BfrCas6 with 300 nM 5' end FAM-labelled BfrCRISPR repeat RNA (purchased from Integrated DNA Technologies (IDT), listed in Table 2-4) in a buffer of 20 mM Tris-HCl pH 7.5, 50 mM NaCl and 1 mM EDTA, at 37 °C for 5 min. The reaction was stopped by heating at 95 °C for 5 min and then analysed by 20 % acrylamide, 7 M urea, 1X TBE denaturing gel, which was run at 30 W, 45 °C for 2 h. Alkaline hydrolysis ladder was generated by incubating RNA in the buffer of 5 mM $NaHCO_3$, pH 9.5 at 95 °C for 5 min. The gel was finally imaged by Typhoon FLA 7000 imager (GE Healthcare) at a wavelength of 532 nm (pmt 600~700).

An internally radio-labelled transcript RNA containing two BfrCRISPR repeats and one guide sequence (Table 2-4) was incubated with 2 µM BfrCas6 in the same condition mentioned above and the reaction products were checked on a 20 % polyacrylamide gel at different time points. The transcript RNA was generated by following the instructions of MEGAscript®Kit (Invitrogen). The template used in transcription was obtained by PCR of plasmid pBfrCRISPR_Lpa using primer Duet-up and Duet-Down. PCR product (120 ng) mixed with ATP, GTP, UTP, CTP solution and 133 nM $\alpha^{32}P$-ATP as a tracer was incubated at 37 °C for 4 h in the 1X reaction buffer with T7 enzyme mix. Transcript was then purified by phenol: chloroform extraction and isopropanol precipitation.

2.2.2 Target RNA cleavage assay of BfrCmr effector complex

RNA cleavage assays using 1 µM wild type BfrCmr (or variant with Cmr4 D27A) and a 5' end-labelled target RNA-Lpa substrate (Table 2-4) were conducted in the reaction buffer (20 mM Tris-HCl, pH 7.5, 10 mM NaCl, 1 % glycerol and 5 mM $MnCl_2$, 0.1 U ul^{-1} SUPERase•In™ (Thermo Scientific)) at 37 °C. The reaction was stopped at indicated time points by adding EDTA (pH 8.0) and extracted with phenol-chloroform to remove protein. After adding equal volume 100 % formamide, the samples were loaded onto 20 % denaturing polyacrylamide

sequencing gel. The gel electrophoresis was carried out at 90 W for 3-4 h. Visualization was achieved by phosphorimaging (Typhoon FLA 7000 imager). A 5' end-labelled target RNA-Lpa substrate was subjected to alkaline hydrolysis generating a single nucleotide resolution ladder for RNA size determination.

A 5' end labelled target RNA-Lpa substrate was generated by incubating 10 µM 5' end dephosphorylated RNA with T4 polynucleotide kinase (Thermofisher) and 1 µl [γ-^{32}P]-ATP (10 mCi/ml) according to the manufacturer's instructions. RNA was purified by denaturing polyacrylamide gel electrophoresis (7 M urea, 1× TBE, 20 % polyacrylamide). The gel pieces with RNA band were excised and then soaked in 500 µl buffer (10 mM Tris, pH 8.0, 1 mM EDTA) overnight at 4 °C. RNA was precipitated with adding ammonium acetate to the final concentration of 2.5 M and 2 volumes of cold ethanol (100 %) stored at -20 °C overnight. The RNA pellets were collected by centrifuging at 13,000 rpm 4 °C for 90 min. The air-dried pellets were resuspended in RNase-free water, after washing with 70 % ethanol.

2.2.3 Synthesis of the signal molecule SAM-AMP and its analogues

2 µM wild type BfrCmr was incubated with ATP and SAM or SAH or sinefungin or the mixture of GTP/CTP/UTP (0.1 mM each for radio-labelled products or 0.5 mM each for HPLC analysis) respectively in the reaction buffer (20 mM Tris-HCl, pH7.5, 10 mM NaCl, 1 % glycerol and 5 mM MnCl$_2$). The reaction was initiated by adding 5 µM target RNA-Lpa (using non-target RNA-pUC as negative control, listed in table 2-4) and carried out at 37 °C for 1 h or different time points in time course assay. 4 nM α^{32}P-ATP as a tracer was added in each reaction to generate radio-labelled products, if needed.

For *in vivo* production, a single colony of *E. coli* BL21star harbouring the plasmids pBfrCmr1-6, pBfrCRISPR_Tet (or pBfrCRISPR_pUC) and pRATDuet was inoculated into 10 mL of L-broth with antibiotic (50 µg/ml ampicillin, 50 µg/ml spectinomycin and 12.5 µg/ml tetracycline) and grown overnight at 37 °C with shaking at 180 rpm. The overnight culture was recultivated with 20-fold dilution into 20 ml fresh L-broth with same antibiotics and then incubated at 37 °C. The cell culture was adding 0.2 % (w/v) D-lactose and 0.2 % (w/v) L-arabinose to fully induce BfrCmr complex expression after OD$_{600}$ of the cells was between 0.4 and 0.6. After overnight induction at 25 °C, the cell culture was mixed with 4 volumes of cold PBS and then centrifuged at 4,000 rpm for 10 min at 4 °C. Cell pellet was resuspended into 2 ml cold extraction solvent [acetonitrile/methanol/water (2/2/1, v/v/v)], vortexed for 30 s and stored at -20 °C overnight. The supernatant was obtained by centrifuged at 13,200 rpm for 10 min at 4 °C, followed by

evaporation. Samples were completely dried then resuspended in water and analysed by HPLC or LC-MS.

2.2.4 Liquid chromatography and Mass Spectrometry analysis

Enzymatic reactions were analysed by UltiMate 3000 UHPLC system (ThermoFisher scientific) with absorbance monitoring at 260 nm. Samples were injected into a C18 column (Kinetex EVO 2.1 X 50 mm, the particle size of 2.6 μm) at 40 °C. Gradient elution was performed with solvent A (10 mM ammonium bicarbonate) and solvent B (Acetonitrile plus 0.1 % TFA) at a flow rate of 0.3 ml/min as follow: 0-0.5 min, 0 % B; 0.5-3.5 min, 20 % B; 3.5-5 min, 50 % B; 5-7 min, 100 % B.

Preliminary LC-MS analysis was conducted on a Thermo Scientific LCQ Fleet Ion Trap LC/MS. Gradient elution was performed as the same as above and the flow from the column sprayed into the ESI of MS. Data were collected in positive ionization mode from 100-2000 m/z.

LC-MS and LC-MS/MS analysis were carried out on a Eksigent 400 LC coupled to Sciex 6600 QT of mass spectrometer. A YMC Triart C18 trap cartridge (0.5 x 5.0 mm) was used to analyse samples in trap and elute configuration in 99.95 % water and 0.05 % TFA at 10 μl/min. The trap was then switched in-line with the analytical column (a YMC Triart 150 x 0.075 mm), when the salts were washed into the waste at beginning 3 min. Gradient elution was performed with solvent A (99.9 % water, 0.1 % FA) and solvent B (20 % water 80 % acetonitrile 0.1 % FA) at a flow rate of 5 μl/min as follows: 0-6 min, 3 % to 95 % B; 6-8 min, 95 % B; 8-9 min, 3 % B; 9-13 min, 3 % B. The flow from the column sprayed directly into the ESI turbospray orifice of the MS, which data were collected in positive ionization mode from 120-1000 m/z. Ions of interest were selected for CID fragmentation at collision voltages of 25-45 V and the fragmentation spectra collected from 50-1000 m/z. The mass spectrometer was externally calibrated prior to analysis with Sciex tuning solution 4457953.

2.2.5 Thin layer chromatography analysis

Radio-labelled SAM-AMP and its analogous were separated by TLC. 1 μl reaction solution was analysed on the 20 X 20 cm Silica gel TLC aluminium plate (sigma-Aldrich) with 0.5 cm of TLC buffer (0.2 M ammonium bicarbonate, 70 % ethanol, and 30 % water pH 9.3) at 35 °C. TLC plate was removed from TLC chamber until the solvent front is approximately 5 cm from the top of the TLC plate and finally imaged by Typhoon FLA 7000 imager (GE Healthcare) using phosphorimaging (PMT = 700-900).

2.2.6 Treatment with nuclease P1

100 μM of compound was incubated with 0.02 units P1 nuclease (New England Biolabs) in the P1 reaction buffer (50 mM sodium acetate (pH 5.5 (25 °C)) at 37 °C for 1 h. Each reaction solution was purified with spin filter (Pall Nanosep®, MWCO 3kDa) followed by HPLC or LC-MS analysis.

2.2.7 BfrNrN and CboSAM-AMP lyase cleavage activity

SAM-AMP cleavage activity was carried out by incubating 1 μM wild type NrN or CboSAM-AMP lyase and their inactive variants with 100 μM purified SAM-AMP or its analogues in the buffer of 20 mM Tris-HCl pH 7.5, 20 mM NaCl, 1 % glycerol and 0.5 mM $MnCl_2$ at 37 °C for 1 h or at indicated time points for time course assay. The reaction was stopped by mixing with 2 volumes pre-cold methanol and vortex for 30 s, before centrifuging at 13,000 rpm at 4 °C for 20 min to remove denatured protein. The supernatant was dried using speed vacuum (Thermo Scientific Savant SPD1010) and then resuspended in the RNase-free H_2O, followed by HPLC analysis.

2.2.8 Electrophoretic Mobility Shift Assay of BfrCorA

40 nM ^{32}P-radiolabelled-signal molecules were incubated with different amounts of BfrCorA in the binding buffer (12.5 mM Tris-HCl, pH 8.0, 10 % (v/v) glycerol, 0.5 mM EDTA) at 25 ºC for 15 min. Reactions were mixed with ficoll loading buffer and then analysed on the native polyacrylamide gel (8 % (w/v) 19:1 acrylamide:bis-acrylamide). Electrophoresis was carried out at 200 V for 2 h at room temperature in the running buffer (1X TBE buffer), followed by phosphor imaging (Typhoon FLA 7000 imager (GE Healthcare), PMT = 700-900).

2.2.9 Western blot of wild type and variants of BfrCorA

E. coli C41 (DE3) was transformed with the plasmid pEHisV5TEV encoding CorA wild type and variants, respectively. A single colony was picked into 5 ml LB (50 μg/ml kanamycin) at 37 ºC with overnight shaking. Overnight cells were 100-fold diluted into 5 ml selective LB and grown at 37 ºC until OD_{600} reached 0.6-0.8, before induced with 0.2 mM IPTG at 16 ºC for 16 h. Cells were collected by centrifuging at 4,000 rpm for 10 min and resuspended in lysis buffer B (50 mM HEPES, pH 7.5, 250 mM NaCl, 5 % glycerol and 10 mM imidazole). After sonication to lyse cells, 10 μl 20-fold diluted lysate was loaded onto the NuPAGE Bis-Tris Gel (Thermo Fisher Scientific) for separation and then transferred into a nitrocellulose membrane using iBlot^{TM} Dry Blotting System (Invitrogen). Membranes were blocked for 1 h with shaking

in TBST (20 mM Tris, 150 mM NaCl, pH 7.6, 0.1 % Tween-20) with 0.03 % milk and then incubated with mouse anti-V5 antibody (Invitrogen) at 1:10,000 dilution in TBST with 0.03 % milk at 4 ºC overnight. Membranes were washed three times in TBST with 0.03 % milk and then incubated with anti-mouse IgG antibody (LI-COR) at 1:20,000 dilution in TBST 0.03 % milk for 1 h at room temperature with shaking. The membranes were washed again with in TBST with 0.03 % milk twice and TBST once, before imaging on an Odyssey® imager system (LI-COR).

2.2.10 BfrNYN ribonuclease assay

The ribonuclease activity of BfrNYN was assayed by incubating 1 or 5 µM BfrNYN with 400 nM fluorescent FAM labelled BfrCRISPR repeat RNA (listed in the table 2-4), in the buffer of 20 mM Tris-HCl, pH 7.5, 50 mM NaCl and 5 mM $MnCl_2$ (or $MgCl_2$) at 37 °C for 1 h. 200 nM BfrNYN was incubated with 40 nM fluorescent FAM labelled RNA D (listed in the table 2-4) in the buffer mentioned above at different time points. 70 nM BfrNYN coupled with 1 µM BfrCas6 were incubated with an internally radio-labelled transcript RNA (listed in the table 2-4) in the same buffer condition at different time points. The reactions were stopped by adding 10 mM EDTA, before heating at 95 °C for 10 min, followed by 20 % acrylamide, 7 M urea, 1×TBE denaturing gel, which was run at 30 W, 45 °C for 2h. The gel was finally imaged by Typhoon FLA 7000 imager (GE Healthcare) at a wavelength of 532 nm (pmt 600~700). Alkaline hydrolysis ladder used for cleavage sites mapping was generated by incubating RNA in the buffer of 5 mM $NaHCO_3$, pH 9.5 at 95 °C for 5 min.

2.2.11 Size exclusion chromatography (SEC) of CalpS, CalpT and CalpL

To determine the interaction of the complex of CalpS and CalpT with CalpL, the SEC assay was carried out on a Superose6 increase 10/300 chromatography column (GE Healthcare) equilibrated with SEC buffer (20 mM Tris, 0.25 M NaCl, 1 mM DTT, 10 % glycerol, pH 8.0). The 200 µl tested sample solution was analysed with a flow rate of 0.5 ml/min. The final concentrations were set to c(CalpL) = 63.3 µM, c (complex of CalpS and CalpT) = 115.8 µM, and $c(cA_4)$ = 60 µM (diluted with SEC buffer). All samples were incubated at 60 °C for 60 min, before cooling down to room temperature and loading onto the column.

2.2.12 Pull-down assay of CalpS, CalpT and CalpL

The magnetic nickel beads-based immobilized metal affinity chromatography (IMAC) was performed to detect releasing of CalpS from the CalpL-CalpT-CaplS complex. The complex of His-tagged CalpS and CalpT complex was incubated with CalpL in binding buffer (20 mM Tris-HCl, pH 7.5, 60 mM NaCl, 0.01 % Tween^{TM}-20) at 60 °C for 60 min in the presence or absence of cA$_4$. After cooling down to room temperature, the sample solution was mixed with pre-equilibrated beads (MagneHis™ Ni particle, Promega) in the binding buffer on a roller for 20 min at 4 °C. The beads were washed three times with 300 µl wash buffer (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 mM imidazole, 0.01% Tween^{TM}-20) before eluting twice with 25 µl elution buffer (20 mM Tris-HCl, pH 8.0, 120 mM NaCl, 300 mM imidazole, 0.01 % Tween^{TM}-20). The sample from the first elution and 20 % of input were analysed on the SDS-PAGE. The final concentrations were set to c(complex of CalpS and CalpT) = 0.208 mg/ml, c(CRISPR–Lon) = 0.127 mg/ml, , and c(cA$_4$) = 2.5 µM, which were diluted by binding buffer.

2.2.13 Truncated CalpT and Cleaved CalpT$_{23}$ ribonuclease assay

Ribonuclease activity of cleaved CalpT$_{23}$ (23 kDa fragment) was assayed by incubating full-length CRISPR-T with CRISPR-Lon and five different fluorescent-labelled RNA substrates, which were synthesised with the fluorescent dye (6-FAM) attached at 5' end or at 3' end (purchased from Integrated DNA Technologies (IDT), Table 2-4). The mixture of CalpL (5.5 µM) and CalpT (5.5 µM) was incubated at 60 °C in 20 mM Tris-HCl, pH 8.0, 50 mM NaCl and 1 mM EDTA for 15 min, cA$_4$ (10 µM) was then added and the mixture was incubated for another 15 min at 60 °C, followed by adding one of the above RNA substrates into the mixture, incubating for an additional 30 min at 60°C. Finally, 6 µl of the sample was analysed on SDS-PAGE (NuPAGE Bis-Tris Gel, Thermo Fisher Scientific) by heating at 95 °C for 5 min with 2 µL of SDS-PAGE loading dye (Thermo Fisher Scientific; NuPAGE Sample Reducing Agent and LDS Sample Buffer). The remaining 14 µl of the sample were loaded to 20% acrylamide, 7 M urea, 1×TBE denaturing gel, which was run at 30W, 45 °C for 2h. The gel was finally imaged by Typhoon FLA 7000 imager (GE Healthcare) at a wavelength of 532 nm (pmt 600~700).

2.3 *In vivo* assay

2.3.1 Plasmid challenge assay

E. coli Bl21star cells were co-transformed with both pBfrCmr1–6 and pBfrCRISPR_Tet (or pBfrCRISPR_pUC). Single colony of transformants was picked up for competent cells preparation into L-Broth (100 µg/ml ampicillin and 50 µg/ml spectinomycin) and cultivated at 37 °C overnight. 50-fold overnight culture was diluted into 20 ml selective LB medium and grown at 37 °C with shaking of 220 rpm until the OD600 reached 0.8-1.0. Cell pellets were collected and then resuspended in an equal volume of pre-chilled competent cells solution (60 mM CaCl$_2$, 25 mM MES, pH 5.8, 5 mM MgCl$_2$, 5 mM MnCl$_2$). Cells were incubated on ice for 1 h and collected pellet was resuspended in 0.1 volumes of the same buffer containing 10 % glycerol. Aliquots (100 µl) were flash frozen by liquid nitrogen and then stored at -80 ºC. The competent cells were transformed with 50 ng pRATDuet or pRATDuet derived plasmids encoding ancillary proteins, respectively, incubated on ice for 30 min and transformed by heat shock at 42 °C. Following addition of 0.5 ml LB medium, the transformation mixture was incubated at 37 °C for 2.5 h. 3 µl of a 10-fold serial dilution was applied in duplicate to LB agar plates (supplemented with 100 µg/ml ampicillin and 50 µg/ml spectinomycin) for recipients' selection. The transformants were selected on LB agar containing ampicillin, spectinomycin and plus 12.5 µg/ml tetracycline. The additional 0.2 % (w/v) - lactose and 0.2 % (w/v) - L-arabinose was used for fully induction. Plates were incubated at 37 ºC overnight. The experiment was performed as two independent experiments with two biological replicates and at least two technical replicates.

Table 2-4 DNA and RNA sequences used in biochemical investigation

Name	Sequence (5'-3')	Note
BfrCRISPR repeat	AUGUAGAUGUAUUCCAGUAUAAUAAGGAUUAAGAC	5' 6-FAM™
Target RNA_Lpa	CACAAGGGACGCCACUUCAUGGAAAUAAAUCACUCACGAAUAGACACGA	RNA
Non-target RNA_pUC	AACGACUCUAGAGGAUCCCCGGGUACCGAGCUCGAAUUCCAAAGGCA	RNA
An internally radio-labelled transcript RNA	GGGGAAUUGUGAGCGGAUAACAAUUCCCCUGUAGAAAUAAUUUUGUUUAACUUUAAUAAGGAGAUAUACCAUGGAAUAGUAAUCUGAUUAUCAAUAUAUGUAGAUGUAUUCCAGUAUAAUAAGG/AUUAAGACAUUCGUGAGUGAUUUAUUUCCAUGAAGUGGCGUCCCUAUGUAGAUGUAUUCCAGUAUAAUAAGG/AUUAAGACUUAAAUAGAGUCGACAAGCUUGCGGCCGCAUAAUGCUUAAGUCGAACAGAAAGUAAUCGUAUUGUACACGGCCGCAUAAUC	T7 transcript
RNA D	AUUGAAAGACCAUACCCAACUUCUAACAACGUCGUUCUUAACAACGGAUUAAUCCCAAAA-OH	5' 6-FAM™/60 nt
RNA F	CUUUCAAUUCUAUAGUAGAUUAGC-OH	5' 6-FAM™/24 nt
RNA U	UUUUUUUUUU-OH	3' 6-FAM™/10 nt
RNA B	UGAUAAUCUCUUAUAGA-P	5' 6-FAM™/17 nt
RNA C	UGUCGUCAGACCCAAAACCCCGAGAGGGGACGGAAAC-OH	5' 6-FAM™/37 nt

3 Antiviral Type III CRISPR signalling via conjugation of ATP and AdoMet

3.1 Introduction

As introduced in the first Section, experimental studies have increasingly characterised the biochemical and structural mechanism of CRISPR signalling immunity, where the diversity and complexity of signalling associated defence pathways have been recognised. Apart from well-known Csx1/Csm6 family ancillary effectors, Shmakov *et al*. showed there are a considerable number of CRISPR-relevant genes encoding membrane proteins located in the vicinity of the type III CRISPR loci, of which the detailed functions remain elusive. This suggests the potential membrane connection of type III CRISPR system and the existence of uncharacterised signal transduction pathways (Shmakov et al., 2018).

Here, we set out to investigate the membrane associated type III-B CRISPR system from the gut microorganism *Bacteroides fragilis* (Bfr) (Fig. 3-1A). *B. fragilis* is a gram-negative and anaerobic bacterium. It's colonised as part of the normal microbiota in the human colon but can lead to clinically significant infection once spreading into the bloodstream or surrounding tissues (Murphy et al., 2011). Bioinformatic analyses have revealed that three CRISPR-Cas systems type I-B, II-C and III-B are present in *B. fragilis* strains and the type III-B are the most common one (Tajkarimi and Wexler, 2017). The interference complex BfrCmr contains six subunits, Cmr1-6, where the main enzymatic subunit Cmr2/Cas10 lacks an HD nuclease domain but has an intact palm cyclase domain. This suggests the signalling pathway may play an essential role in the defence, as observed in the *Vibrio metoecus* type III-B system (Gruschow et al., 2021). The genes adjacent to the CRISPR loci encode non-characterised ancillary effectors, a DHH/DHHA1-family phosphodiesterase, a CorA divalent cation channel and a NYN family nuclease, denoted as BfrNrN, BfrCorA and BfrNYN respectively. We first constructed the BfrCmr system to test if it functioned in the heterologous host *E. coli*. Then, we detected what signal molecule this system employs and what role ancillary effectors play in the BfrCmr signalling pathway.

3.2 Results

3.2.1 BfrCmr system provides immunity against Mobile Genetic Elements (MGEs) in *E. coli*

First, we set out to investigate if the type III-B system from *B. fragilis* could provide immunity in *E. coli*. To test it, three plasmids were constructed as described in the section 2.1.1, 2.1.2 and 2.1.3. Plasmid pBfrCmr1-6 contained the codon optimised BfrCmr interference complex genes *cmr1* to *cmr6*, expressing the functional target RNA dependent Cmr complex as previously confirmed in other CRISPR systems, like *M. tuberculosis* Csm system (Gruschow et al., 2019, Athukoralage et al., 2020b) and *V. metoecus* Cmr system (Gruschow et al., 2021). A control plasmid expressing the cyclase defective large subunit (Cmr2 (Cas10) D328A:D329A) was unable to produce any signal molecules, thus incapable to active signal dependent downstream effectors. The second plasmid pBfrCRISPR encoded BfrCas6 and a mini CRISPR array to help produce a mature CRISPR RNA (crRNA). pBfrCRISPR_Tet encoded crRNA targeting a portion of tetracycline resistance gene, while pBfrCRISPR_pUC as a non-targeting control targeted pUC19 LacZ which is irrelevant in this system. The last construct was derived from pRATDuet (Athukoralage et al., 2020b), which contains genes encoding one of the BfrCorA, BfrNrN and BfrNYN or both BfrCorA and BfrNrN ancillary proteins. Plasmid pRATDuet had a tetracycline resistance gene for activation of the BfrCmr targeting system.

E. coli Bl21star (DE3) expressing a targeting or non-targeting BfrCmr interference complex was challenged by transformation with a pRATDuet encoding variable ancillary proteins (Fig. 3-1B). Cells were 10-fold serial diluted before applying onto selective and fully inducing LB agar plates to determine the colony-forming units (cfu's) of transformants. Reduced cfu's were expected when signal dependent downstream effectors were present in an activated BfrCmr targeting system. No differences were observed in the number of transformants with the pRATDuet vector control, suggesting that target RNA cleavage induced by activation of BfrCmr complex did not provide immunity in the heterologous host *E. coli*. This vector control served as a baseline for transformation (Fig. 3-2A). When only BfrNrN or BfrNYN was present, there was no reduction in cfu's, which suggested they didn't exhibit signal induced activity. When only BfrCorA was expressed, a reduction in cfu's was observed in BfrCmr wild type non-target control and the suppression of cell growth was observed in BfrCmr wild type target system and both Δcyclase BfrCmr target and non-target system, suggesting some toxicity of

the membrane protein BfrCorA (Fig. 3-2A and Fig. 3-2B). When both the BfrCorA and BfrNrN proteins were present, decreased cfu's was only observed in the BfrCmr wild type target system. In addition, the number of transformants was recovered back to the baseline as the vector control when BfrNrN was replaced with an inactive variant NrN$^{\Delta}$ (D85A:H86A:H87A) or when membrane protein BfrCorA was truncated to remove transmembrane domain (Fig. 3-2C). It was thus evident that both functional NrN and CorA are essential for immunity. These findings indicate that Cmr system from *B. fragilis* provides defence against MGEs in *E. coli* in presence of both BfrCorA and BfrNrN, which function in a signal dependent manner.
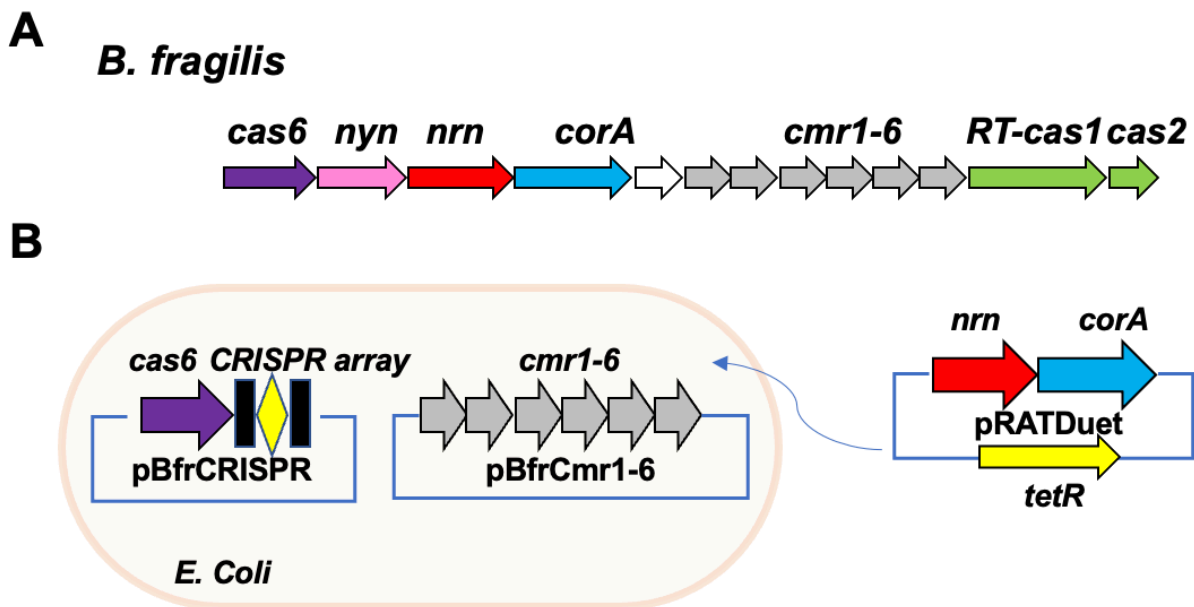
Figure 3-1 Type III-B CRISPR-Cas loci of *B. fragilis* and schematic description of plasmid challenge assay
A. *B. fragilis* type III-B CRISPR loci. The cas genes *cmr1-6* are shown in grey, with *cas6* in purple and the adaptation genes *cas1* (or a gene encoding a fused reverse transcriptase-cas1 protein) and *cas2* in green. The *corA* gene (blue) encodes a putative ion channel, which is adjacent to or fused with the gene encoding PDEs NrN (red). The gene *nyn* (pink) encode a NYN family nuclease. **B.** The schematic shows three plasmids used in the plasmid challenge assay. Cells co-transformed with plasmids pBfrCmr1-6 (wild type or cyclase variant) and pBfrCRISPR (target (TetR) or non-target (pUC)) as recipients were challenged by pRATDuet vector or encoding variable effectors.
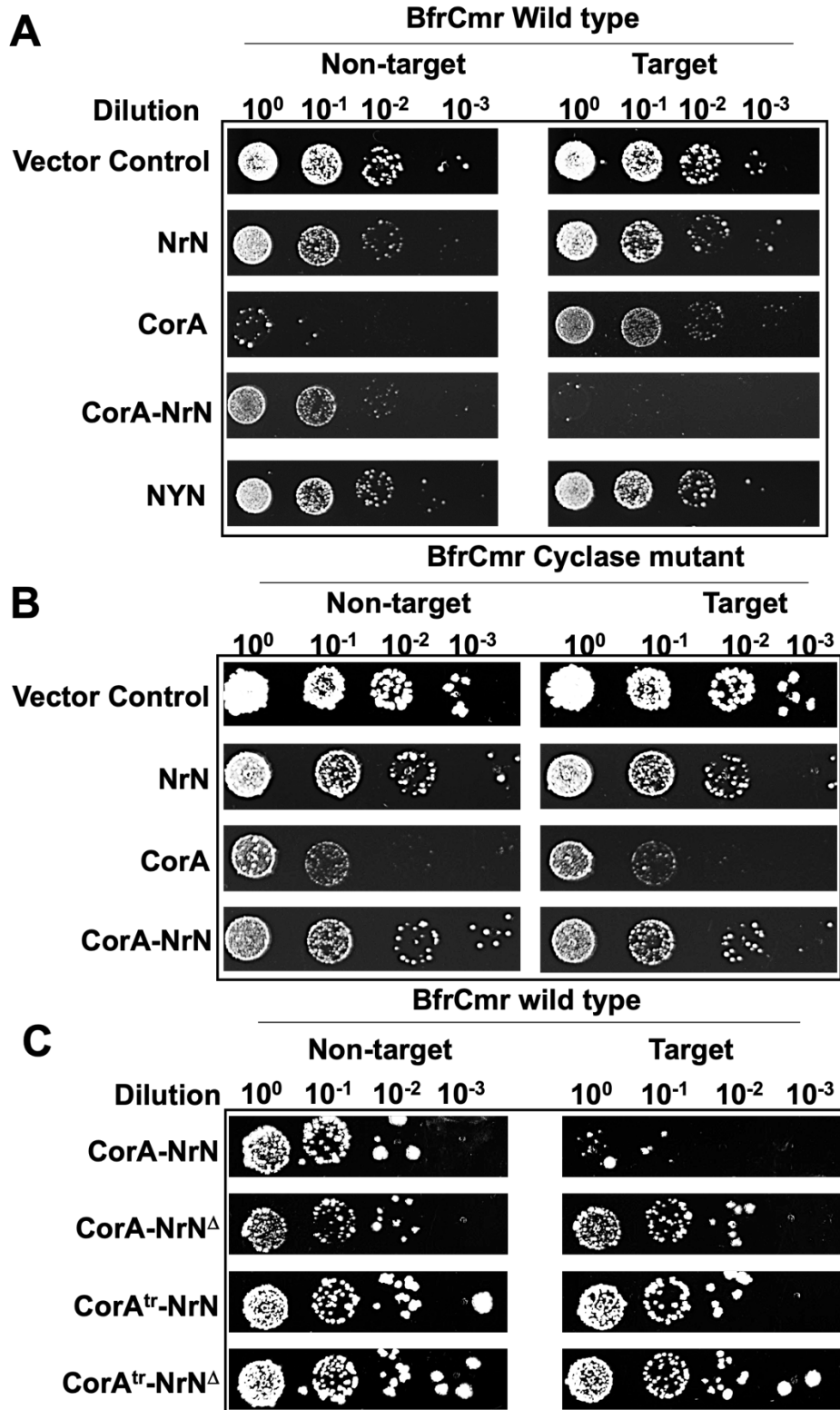
Figure 3-2 Plasmid immunity of BfrCmr wild type, cyclase defective variant and effectors variants
A. Plasmid challenge assay of the BfrCmr wild type system. *E. coli* BL21 Star cells expressing BfrCmr wild type programmed with target (tetR) or non-target (pUC19) CRISPR RNA (crRNA) were transformed with a pRATDuet plasmids that expressed various ancillary effectors and carried a tetracycline resistance gene. Resistance was only observed when a targeting crRNA and both BfrCorA and BfrNrN ancillary proteins were present. B. Plasmid challenge assay of the BfrCmr cyclase defective variant system. Cells expressing BfrCmr cyclase variant and both effectors have similar transformation efficiency as vector control. **C.** Different ancillary effector variants were tested in the BfrCmr wild type system. CorA^{tr} is the truncated membrane protein CorA (aa 1-428) and NrN$^{\Delta}$ is mutated in the active site motif (D85A:H86A:H87A).

3.2.2 BfrCmr system combined with BfrCorA and BfrNrN confers immunity via a non-canonical signal transduction in *E. coli*

To further investigate if the BfrCmr system employs typical cA_{2-6} as its second messengers to regulate signalling defence pathway, BfrCorA and BfrNrN were tested in the recombinant CRISPR system from *M. tuberculosis* type III-A Csm system (MtbCsm) (Gruschow et al., 2019) or from *V. metoecus* type III-B Cmr system (VmeCmr) (Gruschow et al., 2021). MtbCsm system had been observed to produce not only cA_{3-6} but also linear intermediates and provide plasmid immunity combined with its own cA_6-mediated effector MtbCsm6 (Gruschow et al., 2019). MtbCsm has previously been used to test the cOA dependent ancillary effectors or even ring nuclease in immune response. Csx1 from *Thioalkalivibrio sulfidiphilus* (TsuCsx1) showed specific cA_4 dependent RNase activity *in vitro* assay. TsuCsx1 expressed in the activated MtbCsm system in *E. coli* exhibits the similar level of immunity as cognate MtbCsm6. The same immune response was also observed when another cA_4-mediated nuclease, Can2 from *T. sulfidiphilus* was tested with the MtbCsm system in *E. coli* (Zhu et al., 2021). In addition, the viral ring nuclease AcrIII-1 from archaeal virus SSeV degrades cA_4 rapidly and combats immunity conferred by TsuCsx1 in the MtbCsm system in *E. coli* (Athukoralage et al., 2020b). Here, BfrCorA and BfrNrN were tested in MtbCsm system by Dr Sabine Grüschow (University of St Andrews, Scotland) of our laboratory. This assay followed the same procedure as above and used TsuCsx1 as a positive control (Fig. 3-3A). No immunity was detected when expressing BfrCorA or both BfrCorA and BfrNrN in MtbCsm system in *E. coli* C43 (DE3), suggesting that these effectors may use different signal molecules from cA_{3-6} (Fig. 3-3A).

We also swapped BfrCorA and BfrNrN into *V. metoecus* type III-B (VmeCmr) system, which has been confirmed to synthesise predominantly cA_3 and cA_4 (Gruschow et al., 2021). Plasmid pACE-vmeCmr (containing the VmeCmr1-6 genes) and pCDF-target-CRISPR (containing *Vibrio metoecus cas6f* and a CRISPR array targeting a portion of tetracycline resistance gene) were designed and constructed by Dr Sabine Grüschow (University of St Andrews, Scotland). *E. coli* Bl21 star co-transformed with plasmids pACE-vmeCmr and pCDF-target-CRISPR (or pCDF-nontarget-CRISPR as a control) was challenged by pRATDuet encoding BfrCorA or both BfrCorA and BfrNrN effectors. cA_3 activated nuclease VmeNucC acted as a positive control (Fig. 3-3B). There was no difference in cfu's in VmeCmr target and non-target system when in presence of both BfrCorA and BfrNrN effectors, suggesting that effectors from *B. fragilis* may not use cA_3 or cA_4 as their activators. VmeNucC had been verified as a cA_3 specific DNase with high sensitivity. VmeNucC was thus selected to test in BfrCmr system in

E. coli and cells grew properly in both target and non-target system, confirming that BfrCmr may not employ cA_3 as its second messenger (Fig. 3-3C). The data strongly imply that Cmr system from *B. fragilis* provides protection against MGEs via a different defence mechanism from the canonical type III CRISPR signalling pathway.
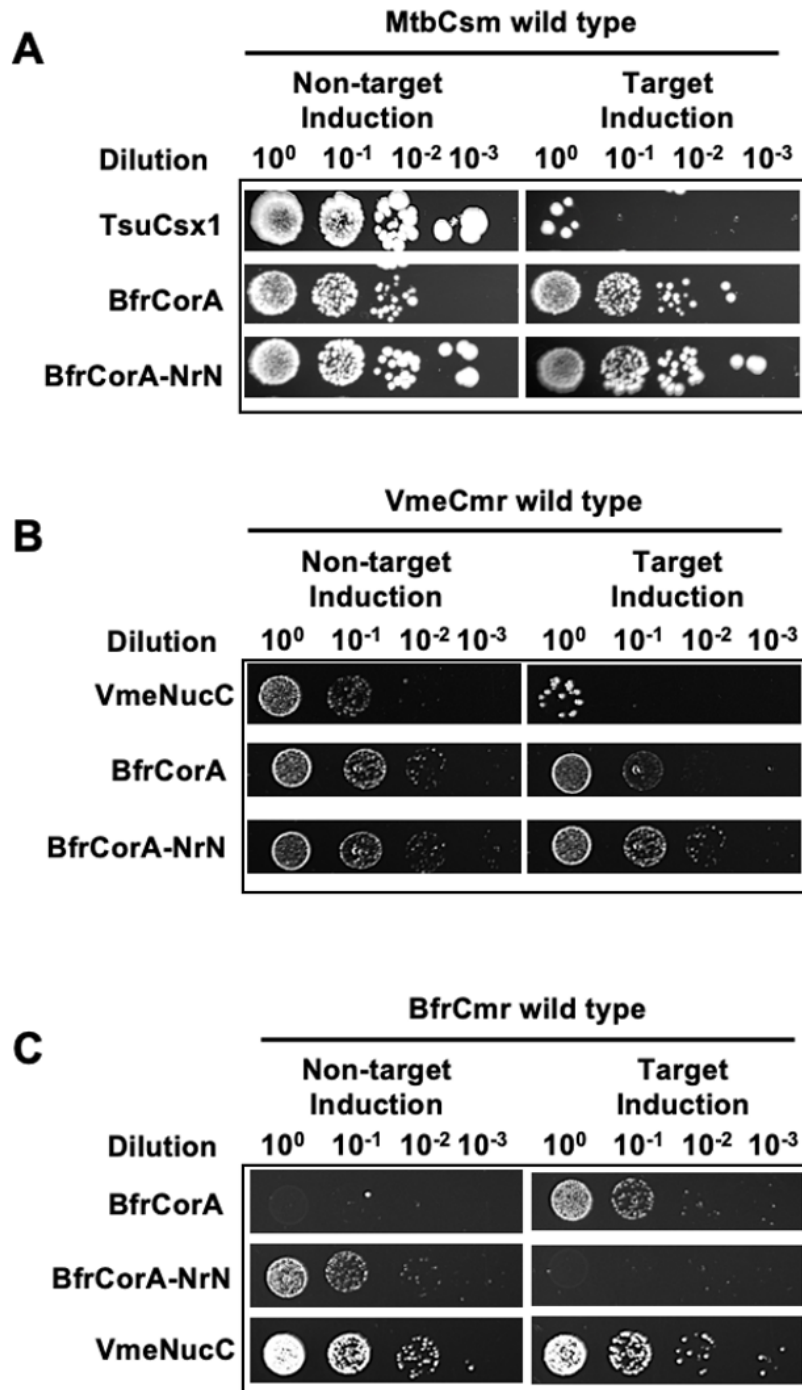
Figure 3-3 Effectors from *B. fragilis* were tested in MtbCsm and VmeCmr CRISPR systems
A. Ancillary proteins BfrCorA and BfrNrN were assayed in the *Mycobacterium tuberculosis* (Mtb) Csm wild type system. *E. coli* Bl21Star cells harboring the MtbCsm wild type systems programmed with target (tetR) and non-target (pUC) crRNA were transformed with pRATDuet plasmids that expressed the NrN with and without CorA proteins and carried a tetracycline resistance gene. cA_4-mediated nuclease TuCsx1 from *Thioalkalivibrio sulfidiphilus* acted as a positive control. No resistance was observed. **B.** BfrCorA and BfrNrN were tested in the *Vibrio metoecus* (Vme) Cmr wild type system. cA_3 specific DNase VmeNucC acted as a positive control. **C.** VmeNucC was assayed in BfrCmr wild type system, and no immunity was observed.

3.2.3 Bfr CRISPR RNA processing mediated by purified BfrCas6

Cas6 has been recognised as a CRISPR-associated endoribonuclease responsible for crRNA processing, which enables Csm/Cmr subunits to assemble as a functional interference complex. Before reconstituting a functional BfrCmr system, we firstly studied the function of BfrCas6. BfrCas6 was codon optimised and expressed in *E. coli* C43 (DE3) using expression vector pEHisV5TEV (Fig. 2-2B) and purified into homogeneity (Fig. 3-4A) by immobilised metal affinity and size exclusion chromatography (Fig. 3-4B and C). The recombinant BfrCas6 cleaved a synthetic FAM-labelled BfrCRISPR repeat substrate at the base of the predicted 2 bp stem to generate a canonical 8 nt 5'-handle away from 3' end of repeat (Fig. 3-5A). This recognition and processing of a non-stem-loop CRISPR RNA had also been observed with Cas6b from *Methanococcus maripaludis* (Shao et al., 2016). The capacity of BfrCas6 to process crRNA was further investigated by incubating with an *in vitro* radio-labelled transcript consisting of two repeats flanking one spacer. According to the identified cleavage site of BfrCas6 in the repeat sequence, the length of cleavage products was predicted, and the final processed crRNA was 72 nt. A set of cleavage products with expected sizes was observed from denaturing polyacrylamide gel electrophoresis (Fig. 3-5B) crRNA processing was catalysed by BfrCas6 in a metal-independent manner, consistent with all other studied Cas6 enzymes (Hochstrasser and Doudna, 2015, Li, 2015).
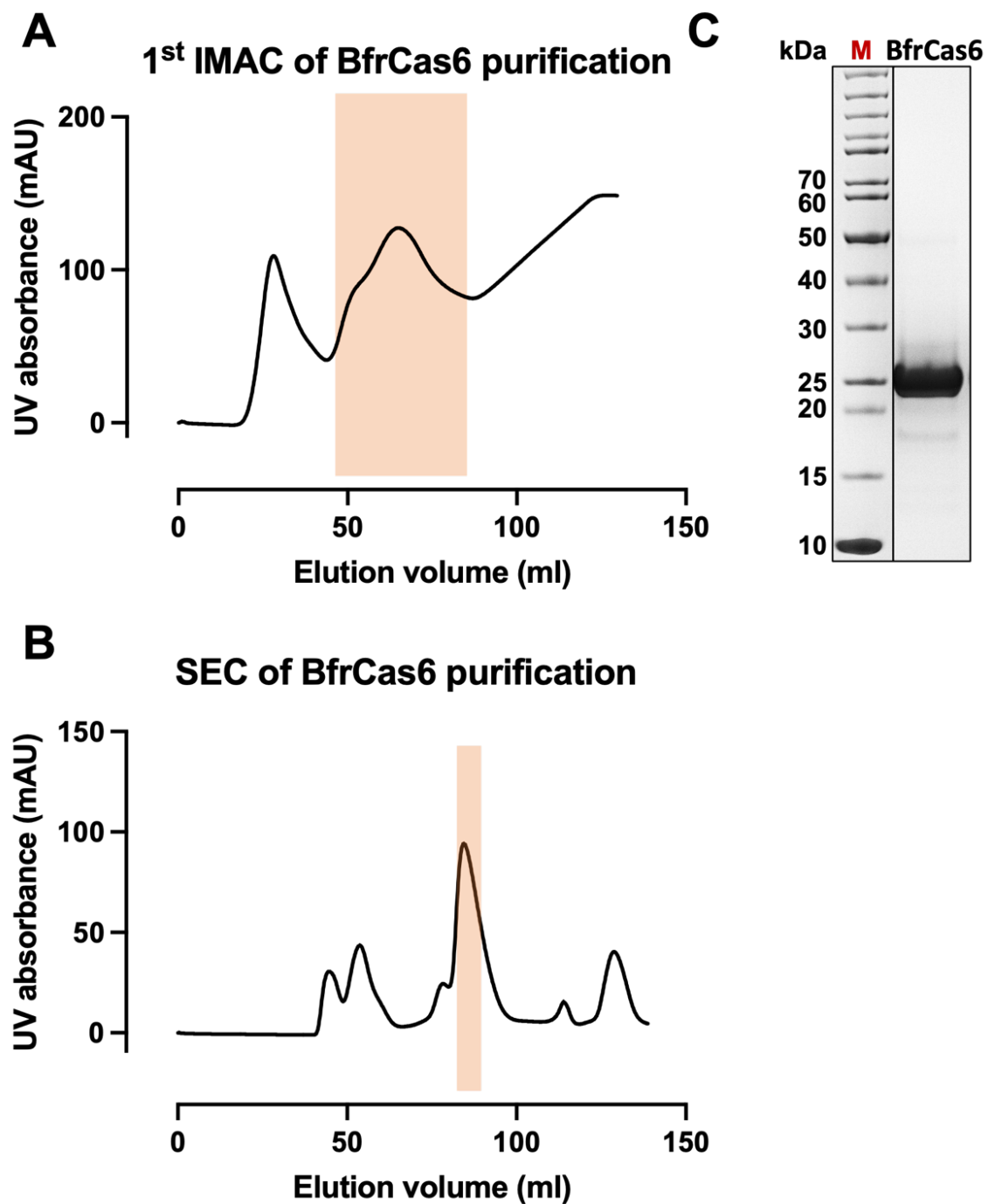
Figure 3-4 Purification of BfrCas6
A. The first immobilised metal affinity chromatography (1st IMAC) step of BfrCas6 purification. The fractions containing target protein highlighted with a red rectangle was eluted with 50% elution buffer and pooled for SEC. **B.** Superdex200 SEC profile of BfrCas6. Pooled samples were then subjected to SEC. Fractions indicated by a red rectangle were collected and concentrated for further enzymatic analysis. **C.** SDS-PAGE analysis of purity of BfrCas6. The monomer mass is approximately 26 kDa, in agreement with the theoretical mass of BfrCas6. M is the marker to indicate the size on the gel.
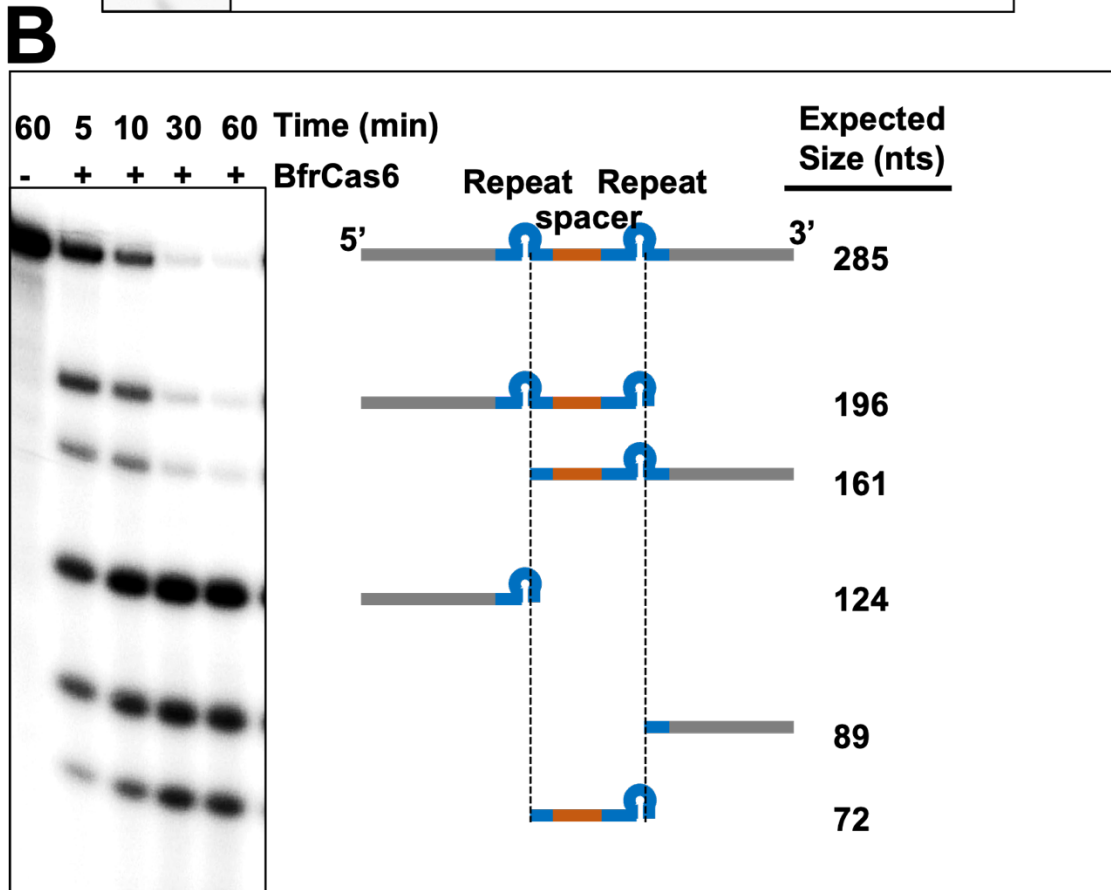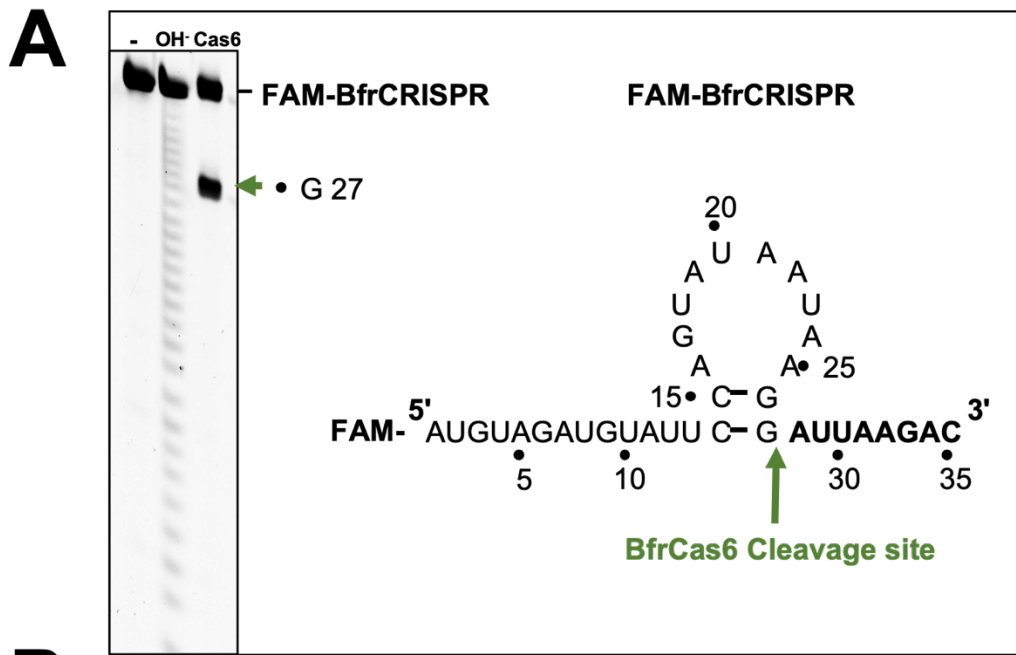
Figure 3-5 BfrCRISPR RNA processing of BfrCas6
A. The cleavage site of Cas6 within the CRISPR repeat was mapped by incubating 5' end FAM-labelled repeat (300 nM) with Cas6 nuclease (1.2 µM). Alkaline hydrolysis (OH-) ladder was used to mark the size of 5' RNA cleavage products (green arrow). Potential secondary structure of CRISPR repeat RNA with cleavage site was indicated (green arrow). **B.** An internally radio-labelled transcript RNA containing two CRISPR repeats (blue) and one guide (targeting Phage P1) sequence (orange) was incubated with BfrCas6 (2 µM). Samples were collected at the indicated time points and analysed by denaturing gel. The expected sizes and compositions of cleavage products are indicated based on the specific cleavage site of Cas6 within each repeat.

3.2.4 Purification of BfrCmr interference complex

To explore the activation mechanism of the BfrCmr system, we set out to reconstitute BfrCmr system *in vitro*. To test it, two plasmids were constructed as described in the section 2.1.1 and 2.1.2. Plasmid pBfrCmr1-6 contains the codon optimised BfrCmr interference complex genes *cmr1* to *cmr6* and pBfrCRISPR_Lpa encodes BfrCas6 and CRISPR array consisting of two BfrCRISPR repeats flanking one spacer targeting the gene encoding Late Promoter Activating protein (Lpa) of phage P1 (Lobocka et al., 2004)(Fig. 2-1 and Fig. 2-2A). BfrCmr complex was purified by co-transforming plasmids pBfrCmr1-6 and pBfrCRISPR_Lpa into *E. coli* Bl21 star (DE3), followed by the expression and purification procedure mentioned in the section 2.1.4. We also constructed and purified their variants BfrCmr_ΔCy (Cmr2 D328A:D329A), BfrCmr_DN (Cmr2 D70N), BfrCmr_ER (Cmr2 E151R) and the double mutant BfrCmr_DNER (Cmr2 D70N:E151R). The purity of all BfrCmr recombinants was checked by SDS-PAGE (Fig. 3-6A, B and C).

Family	Cmr	Size (kDa)	Subunit
Cas10	Cmr2	69.17	Large
Cas11	Cmr5	15.26	Small
Cas7	Cmr4	31.55	Backbone
Cas5	His-Cmr3	48.78	Bottom cap
Cas7	Cmr1	55.26	Top cap
Cas7	Cmr6	35.91	Top cap
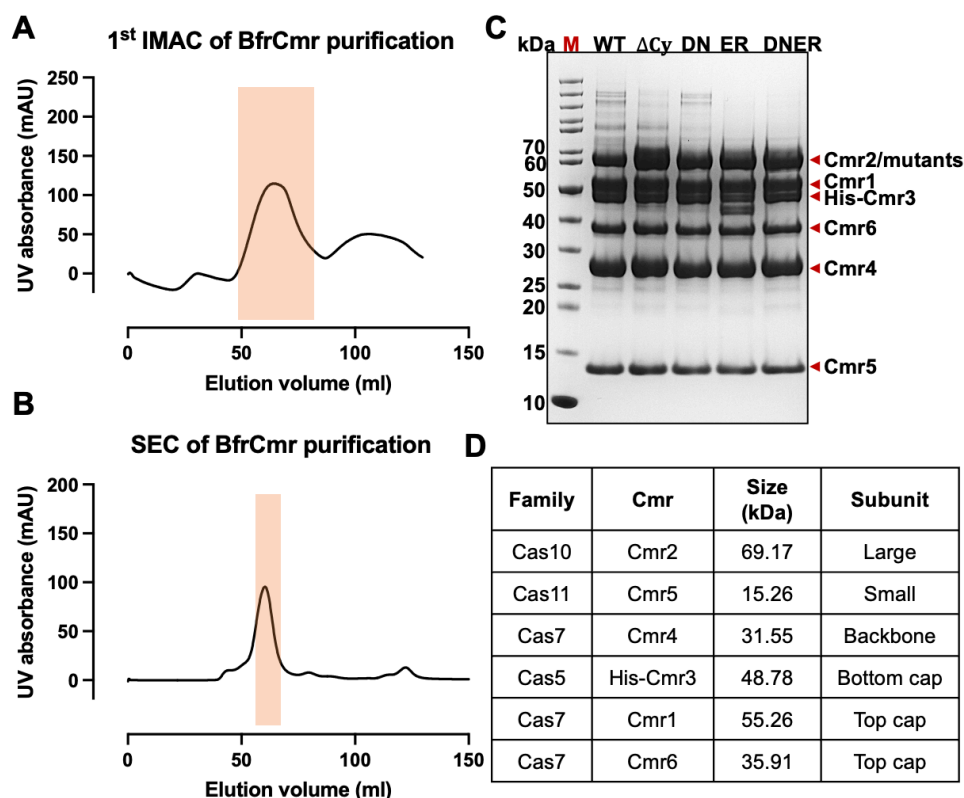
Figure 3-6 Purification of BfrCmr complex
A. The first immobilised metal affinity chromatography (1st IMAC) for BfrCmr purification. The fractions containing target protein highlighted by a red rectangle was eluted with 50 % elution buffer and pooled for his tag removal. **B.** Superdex200 SEC profiles. The TEV-cleaved protein was recovered from the nickel column and then subjected to SEC. Fractions indicated by a red rectangle were collected and concentrated for further enzymatic analysis. The figures made by the Prism was plotted the absorption at 280 nm against elution volume. **C.** The purified wild type (WT) and variants of Cmr (1-6) complex were analysed by the SDS-PAGE gel, which include Cmr2 D328A:D329A (Δcy), Cmr2 D70N (DN), Cmr2 E151R (ER) and the double mutant (D70N:E151R, DNER). Each subunit of BfrCmr was indicated by red arrow, where the monomer mass of Cmr1 to Cmr6 (Cmr3 with his tag) is approximately 55, 69, 48, 31, 15 and 35 kDa respectively, in agreement with their theoretical mass. M is the marker to indicate the size on the gel. **D.** Components and size of BfrCmr complexes.

3.2.5 The crRNA content of BfrCmr complex

Subunits of interference complex are assembled around Cas7 (Csm3/Cmr4) backbone that binds a Cas6-processed crRNA. Unprotected regions of crRNA are trimmed from 3' end by unknown cellular trimming nucleases to obtain the mature crRNA, which is essential for the target RNA cleavage (Hatoum-Aslan et al., 2013, Hatoum-Aslan et al., 2011, Hale et al., 2009, Garneau et al., 2010, Walker et al., 2017, Chou-Zheng and Hatoum-Aslan, 2022, Chou-Zheng and Hatoum-Aslan, 2019). We thus determined the crRNA content present in the BfrCmr complex purified from *E. coli*. This was done by isolation of the crRNA from the purified BfrCmr complex. Isolated crRNA species were subsequently labelled at 5' end using γ-^{32}P-ATP and polynucleotide kinase, before analysing on the denaturing polyacrylamide gel. Three major crRNA species were detected with a variation of 6 nt in length, which are all shorter than the BfrCas6 processed crRNA, suggestive of the trimming from 3' end to remove the repeat derived sequence and partial spacer sequence (Fig. 3-7A and B). The mature crRNA differing in the length also indicated the variable composition of BfrCmr complex with different number of ruler protein cas7, consistent with other type III systems (reviewed in (Tamulaitis et al., 2017)).
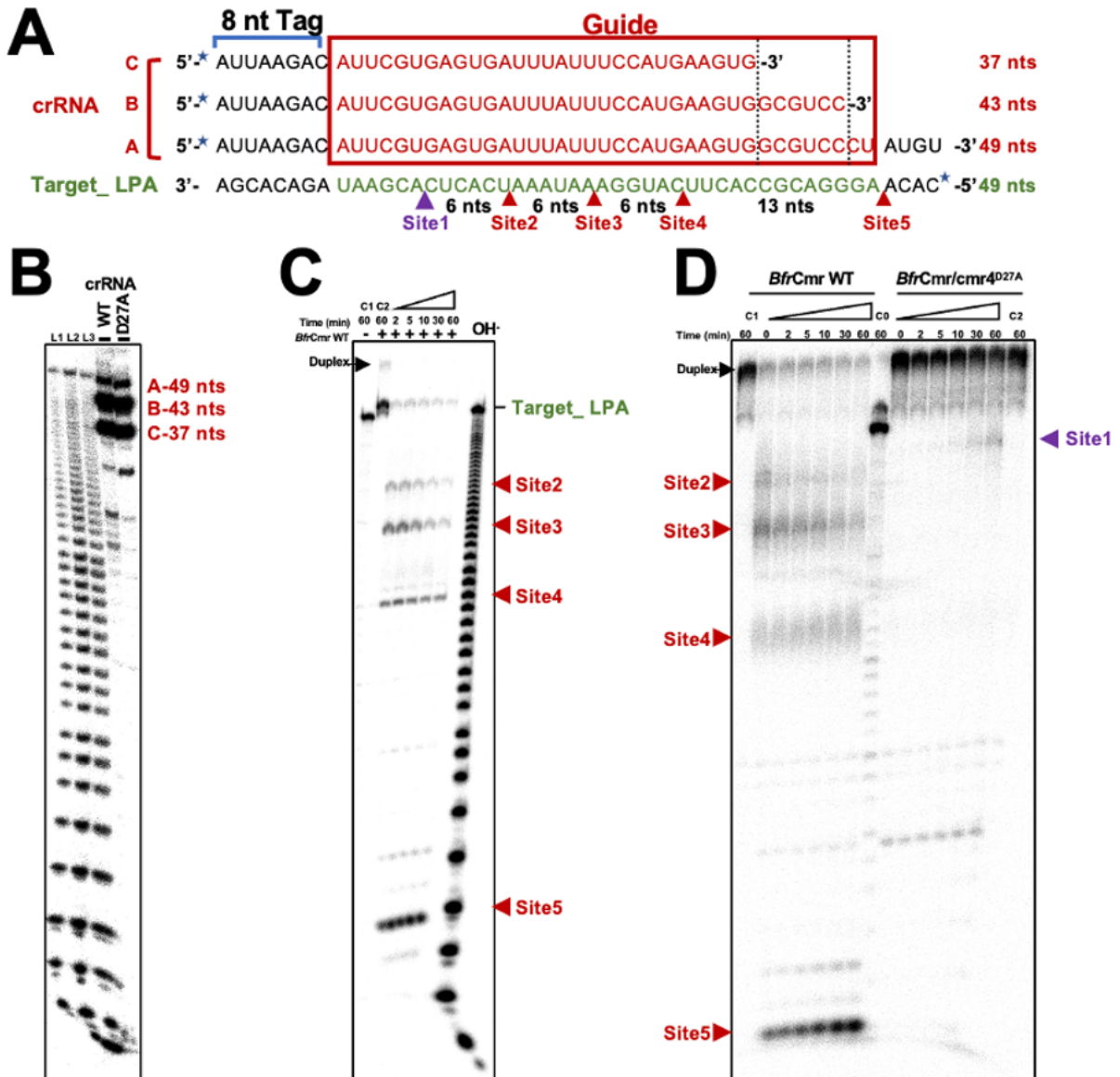
Figure 3-7 BfrCmr crRNA composition and target RNA degradation
A. The sequence of crRNA species extracted from purified Cmr and the target RNA substrate used in the activity assay. The repeat-derived sequence (8 nt tag), spacer-derived sequence (guide) and the sequence complementary to guide RNA are coloured black, red, and green, respectively. Five putative cleavage sites are indicated by arrows (Site1 is indicated by purple arrows, while sites 2 to 5 by red arrows). Extracted crRNAs and the target RNA substrate were 5'-labelled with ^{32}P (blue star). **B.** The size of extracted crRNAs from wild type and mutant Cmr (Cmr4 D27A) was mapped by comparing with alkaline hydrolysis ladder of Target_LPA substrates (L1-3 with increased concentration of substrates). **C.** The indicated Target_LPA was incubated with (+) or without (-/C1) wild type Cmr in the presence of Mn^{2+}(no Mn^{2+} in buffer C2). The cleavage sites were mapped by comparing with alkaline hydrolysis (OH-) ladder and indicated by red arrows. **D.** Time course of cleavage on the 5'-radio-labelled Target_LPA by wild type or mutant Cmr (Cmr4 D27A). The buffer of C1 and C2 are in absence of Mn^{2+}, while C0 is in absence of Cmr.

3.2.6 An efficient target RNA cleavage activity of BfrCmr

We proceeded to investigate whether the purified BfrCmr complex is functional. Csm/Cmr interference complexes commonly exhibit three enzymatic activities, including target RNA cleavage, non-specific ssDNA degradation and second messenger production. Briefly, the mature crRNA guided Cmr complex detects invading MGEs via complementarity to the spacer region of the crRNA, activating target RNA cleavage activity of Cas7 (Csm3/Cmr4), non-specific ssDNA cleavage activity of Cas10 HD nuclease domain and signal molecule synthesis activity of Cas10 palm domain (Samai et al., 2015, Elmore et al., 2016, Kazlauskiene et al., 2016, Kazlauskiene et al., 2017, Niewoehner, 2017). Cas10 will be deactivated after bound target RNA is cleaved by the Cmr4 subunit (Kazlauskiene et al., 2016, Rouillon et al., 2018). However, the Cmr complex from *B. fragilis* lacks the capacity to cleave ssDNA as the Cmr2 (Cas10) lacks a HD domain. Thus, we test if purified BfrCmr complex exhibits target RNA cleavage activity.

To detect the target RNA cleavage activity, 5' end radio labelled target RNA was designed complementary to the spacer region of crRNA and was incubated with wild type BfrCmr complex in the presence of Mn^{2+}. The reaction products were analysed via polyacrylamide gel electrophoresis and cleavage sites were determined by mapping alongside a target RNA alkaline hydrolysis ladder. A 49 nt target RNA was cleaved at three sites (site 2 to 4) with 6 nt intervals to give 17, 23 and 29 nt products, respectively (Fig. 3-7A and C). According to the time course assay, degradation was extremely rapid, initiating from the 3' end internal sites of target RNA and extending towards 5' end with 6 nt spacing and nearly complete after 2 min, the first time point (Fig. 3-7C). The 6 nt periodic products also indicated the number of Cas7 (Cmr4) in the backbone.

To further investigate the active sites of BfrCmr4, the sequence of BfrCmr4 was aligned with its structural homologues, including Cmr4 from *Archaeoglobus fulgidus* DSM 4304 (PDB: 3X1L), Cmr4 from *Pyrococcus furiosus* DSM 3638 (PDB：4WNZ) and Cmr4 from *Saccharolobus islandicus* (PDB: 6S8B). Target RNA degradation was completely abolished by the D26A mutation of PfuCmr4, D31A variants of AfuCmr4 and SisCmr4 (Zhu and Ye, 2015, Osawa et al., 2015, Sofos et al., 2020). The equivalent active site residue of BfrCmr4 is Asp27, according to the sequence alignment (Fig. 3-8). A BfrCmr/Cmr4 D27A variant was thus constructed and purified using the same procedure as the wild type BfrCmr complex. It was hard to observe the degradation products cleaved at sites 2 to 5 in the Cmr4 D27A variant, when followed the same time course assay as the wild type (Fig. 3-7D). A site 1 cleavage

product was observed 6 nt away from site 2 and slowly accumulated in the D27A variant. A further product, only observed for the wild type of complex, was cleaved at site 5, at the boundary of the crRNA: target RNA duplex (Fig. 3-5D), suggesting that this activity may be due to the Cmr4 subunit. As target RNA binding and clearance are known to regulate the HD nuclease and cOA synthesis activities of Csm/Cmr effectors, these data suggest BfrCmr may only be briefly activated, thanks to the fast degradation of target RNA. This has also been observed in the type III system from *Streptococcus thermophilus* and *Thermotoga maritima* (Estrella et al., 2016, Kazlauskiene et al., 2016).
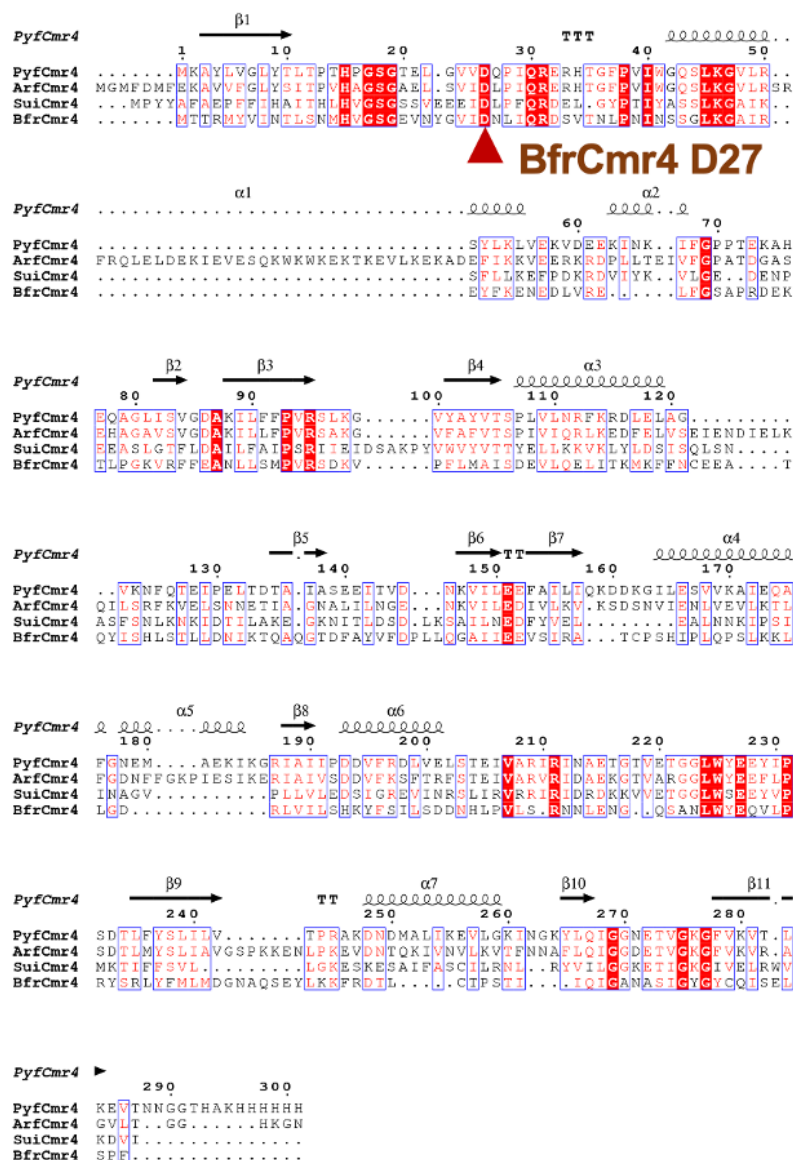
Figure 3-8 Alignment of BfrCmr4 and its structural homologues
Multi sequence alignment of BfrCmr4 with PyfCmr4 from *Pyrococcus furiosus* DSM 3638 (WP_011012266.1 and PDB: 4WNZ), ArfCmr4 from *Archaeoglobus fulgidus* DSM 4304 (WP_086976053.1 and PDB: 3X1L),) and SuiCmr4 from *Sulfolobus islandicus* (WP_014513657.1 and PDB: 6S8B). NCBI reference sequence and PDB ID are indicated in round brackets. Sequences were aligned using MUSCLE of EMBL-EBI (Madeira et al., 2019). ESPript 3.0 (Robert and Gouet, 2014) was used for secondary structure depiction, where secondary structure of PyfCmr4 is schematically annotated above alignment. Blue boxes indicate conserved regions and highly conserved residues are written in the red. The red background highlights identical residues. The potential active site of BfrCmr4 is indicated by a red arrow.

3.2.7 BfrCmr possesses capability to generate a new second messenger

B. fragilis Cas10 lacks a HD domain, suggestive of an essential role of second messenger mediated signalling pathway in BfrCmr immune system. We next set out to identify the target RNA-activated signal molecule synthesis activity of the BfrCmr effector. The cyclase domain of Csm/Cmr complexes generally produce a range of cyclic oligoadenylates from cA_2 to cA_6 upon target RNA binding. The wild type BfrCmr was thus incubated with ATP in the presence of Mn^{2+} by adding target RNA to initiate this reaction and keeping activation by addition of extra target RNA every 15 min, followed by LC-MS analysis. Only two peaks were observed from HPLC, and they are all linear intermediates identified by MS (Fig. 3-9A, C and D). The less efficient generation of final cyclic products may be caused by a short-lived active state of the BfrCmr complex, driven by the fast target RNA cleavage. Slower target RNA degradation as observed in the Cmr4 D27A variant, was investigated in the hope that it would increase production of cOA by extending the activation time. However, no cOA products were observed when D27A variant was incubated with ATP (Fig. 3-9A).

Apart from cyclic oligoadenylates made by the type III CRISPR system (Kazlauskiene et al., 2017, Niewoehner, 2017), diverse cyclic anti-phage signalling molecules have been recently discovered, like cUMP and cCMP from the PYCSAR system and cyclic di- and tri- nucleotides from the CBASS system (Tal et al., 2021, Whiteley et al., 2019). One possibility is that the Cmr from *B. fragilis* may produce other types of molecules than cOAs. We then incubated wild type and Cmr4 D27A variant of BfrCmr with the mixture of four ribonucleotides ATP, UTP, CTP and GTP. Still no cyclic products other than linear intermediates were observed (Fig. 3-9B, C and D). These results imply that BfrCmr does not produce any known or previously identified second messengers.
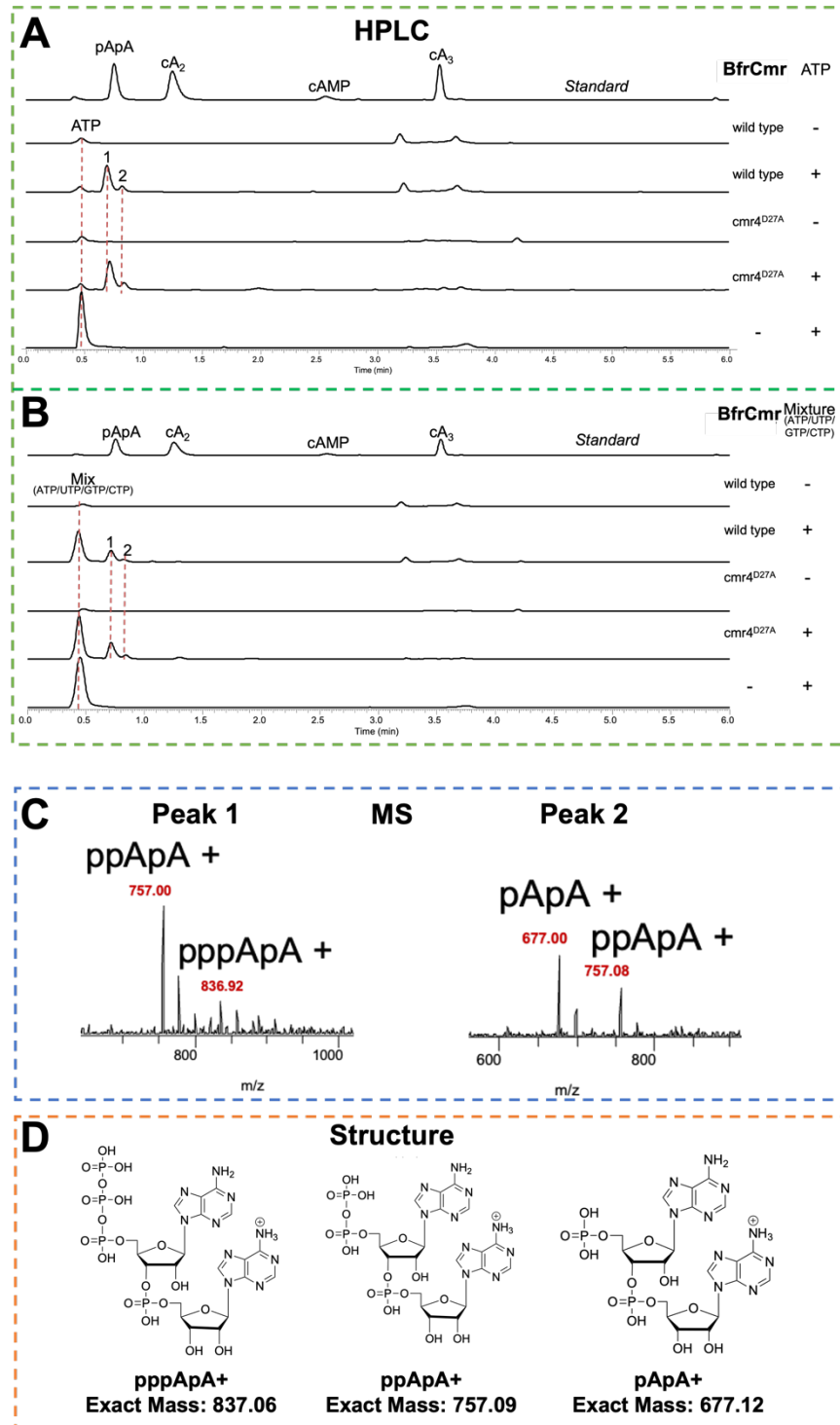
Figure 3-9 HPLC-MS analysis of *in vitro* products generated by incubating BfrCmr wild type and Cmr4D27A variant with nucleotide triphosphates

A. HPLC analysis of the products synthesized by incubating BfrCmr wild type and Cmr4 D27A variant with ATP *in vitro*, respectively. **B.** The *in vitro* products were analysed on the HPLC after BfrCmr wild type and Cmr4 D27A variant incubated with the four ribonucleotide triphosphates mixture. The pApA, cA_2, 3'5'-cAMP and cA_3 standards were aligned at the top trace. The red dashed lines were indicated reaction products. The absence of enzymes or the substrates was set as controls, indicated by minus sign. **C.** MS analysis of products performed on LCQ Fleet Ion trap LC/MS in the positive mode. The experimental mass of products was highlighted in the red text and the possible molecules with one positive ion were written above the mass. The experimental mass of standard pApA is 677.08 (mass error is 60 ppm). **D.** The predicted structures with chemical formula and exact mass were listed according to the MS data.

3.2.8 Isolation and identification of a novel signal molecule, SAM-AMP

We postulated that *E. coli* harbouring the activated BfrCmr system in the absence of ancillary proteins might potentially accumulate Cas10-derived signal molecules. *E. coli* BL21star was thus co-transformed with three plasmids: pBfrCmr1-6, pBfrCRISPR_Tet (or pBfrCRISPR_pUC as an inactivated control) and pRATDuet vector and was grown until OD_{600} of 0.6-0.8 before full induction. The nucleotide products were purified and isolated from overnight induced cell lysates followed by HPLC analysis. We also extracted nucleotide products from BfrCmr cyclase variant as a negative control and wild type *V. meteocus* Cmr system (VmeCmr) as a positive control. A significant HPLC peak was observed from the extracts of activated BfrCmr wild type system (Target) but not in the absence of target (Non-T) or the cyclase variant (BfrΔCy) system (Fig. 3-10A trace i-iv). The retention time of this peak was about 2.5 min, different from the retention time of cA_3 (3.5 min) and cA_4 (3.7 min) extracted from VmeCmr system (Fig. 3-10A trace i, v and vi). The molecule from the peak at 2.5 min was then analysed by MS in positive ionization mode, yielding a m/z value of 728.1963, which didn't match any known signal molecule or any other preciously characterised metabolite (Fig.3-10B). Tandem MS/MS was performed to identify this molecule by fragmentation, where we found the fragments of AMP and methionine (Fig. 3-10C), indicating the molecule isolated from BfrCmr system was S-adenosyl methionine (AdoMet, SAM) adenylated on the ribose moiety (Fig. 3-10D). Henceforward, it was designated as SAM-AMP. The lack of any information on SAM-AMP in both chemical and enzymatic synthesis methods indicates a novel and unexplored class of signalling molecule.
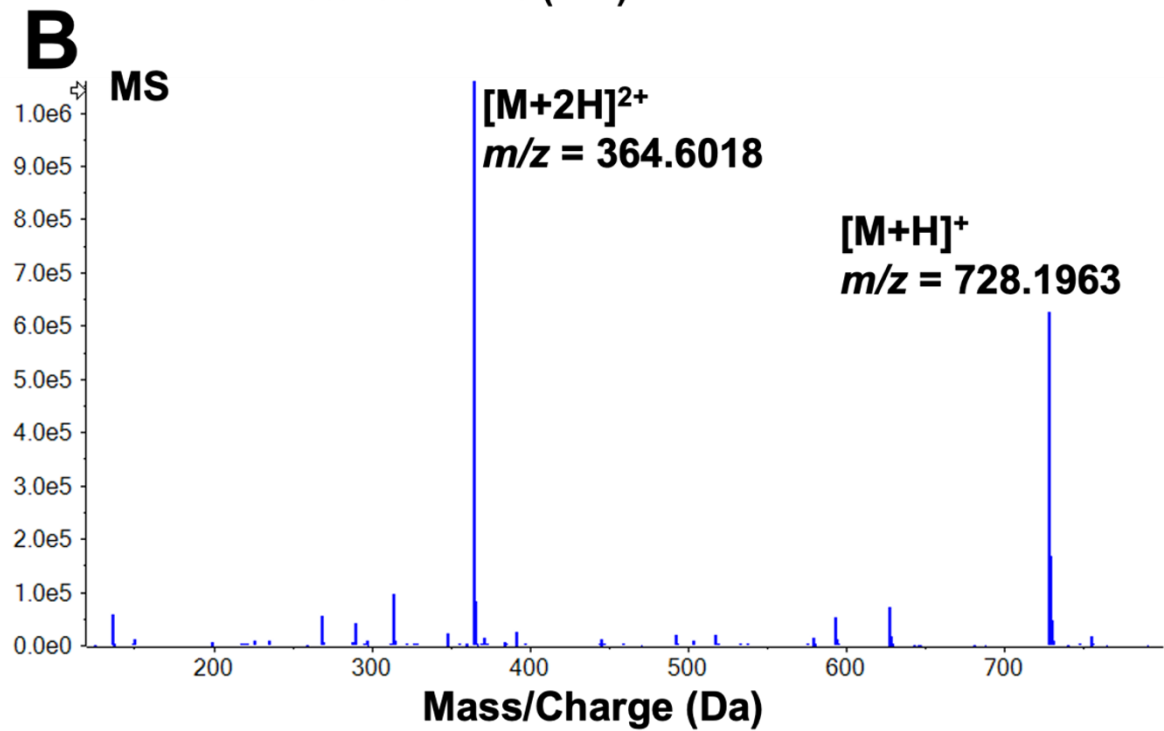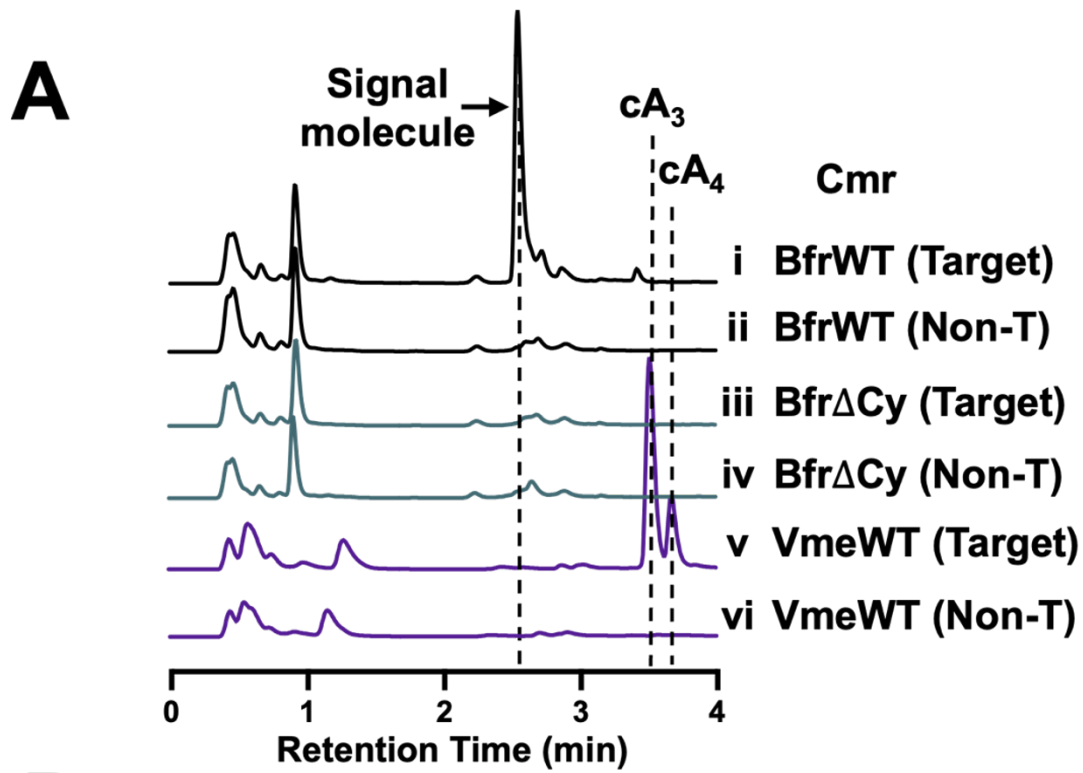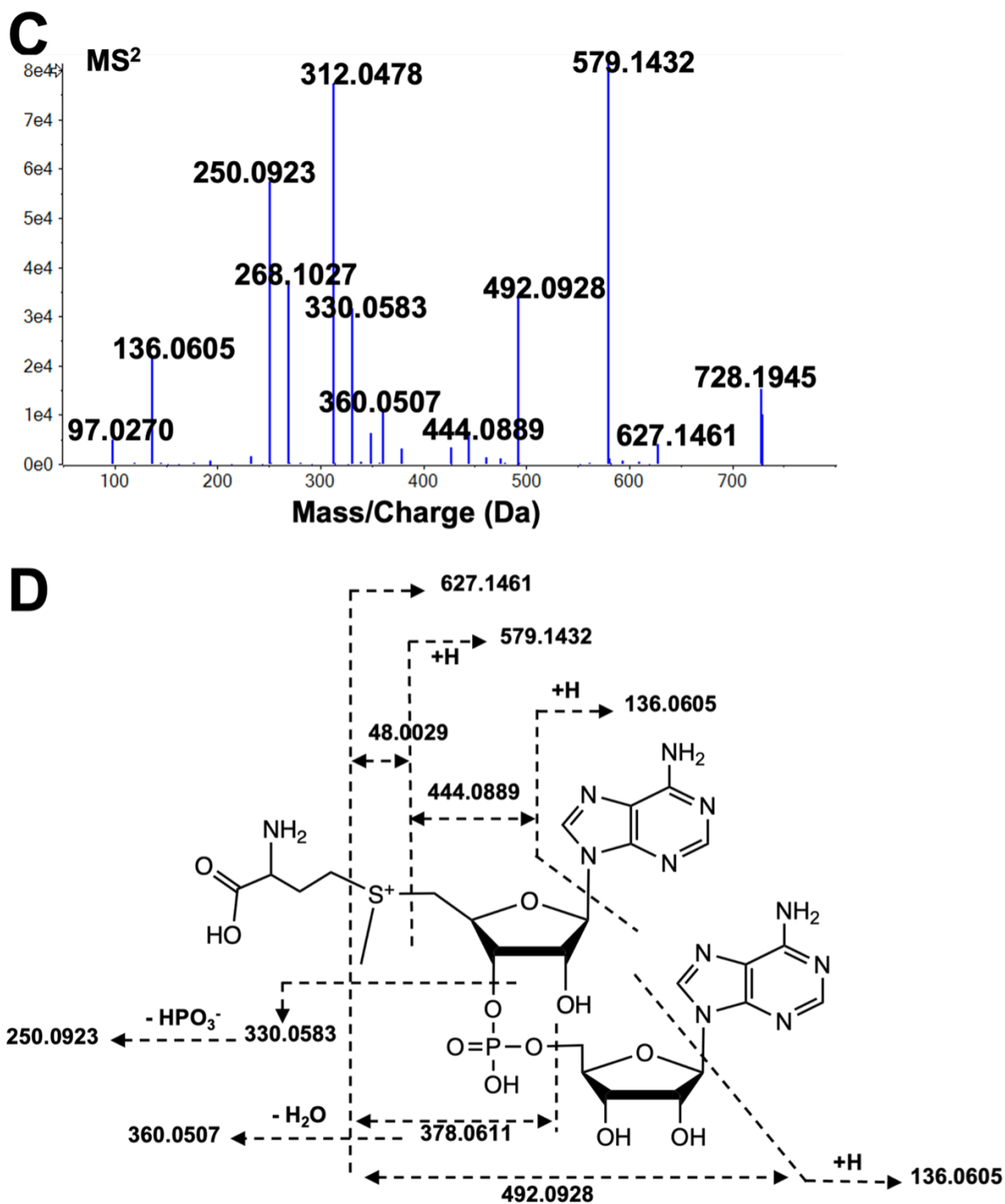
Figure 3-10 Isolation and identification of SAM-AMP from cells harbouring the activated BfrCmr complex
A. Extracted and purified nucleotide products were analysed by HPLC. Right side panel was indicated the analysed samples extracted from the wild type or mutant (ΔCy) *B. fragilis* Cmr system or wild type *V. meteocus* cmr system with target or non-target crRNA. The black arrow indicates the putative signal molecule only observed for the activated BfrCmr wild type system (trace i). Signal molecule cA₃ and cA₄ extracted from activated wild type VmeCmr system were highlighted by black dashed lines. **B.** LC-MS analysis of extracted signal molecule from wild type BfrCmr system. LC-MS was performed on a Eksigent 400 LC coupled to Sciex 6600 QT of MS in positive ionization mode. [M+H]⁺ and [M+2H]²⁺ are two different ionization forms. **C.** MS/MS analysis of the signal molecule with m/z value of 728.1963. The calibration was conducted with analysis of standard cA₂ with an error of -1.3 ppm. **D.** The proposed structure of SAM-AMP, whose exact mass is 728.1970 and fragmentation pattern is shown by dotted arrows. Linkage of SAM and AMP cannot be identified by LC-MS/MS data. SAM-AMP are more likely in a 3'-5' phosphodiester bonds shown here, but a 2'-5' bond cannot be completely ruled out presently.

3.2.9 BfrCmr synthesises SAM-AMP *in vitro*

To confirm that SAM-AMP is the signalling molecule generated by BfrCmr system, the purified wild type BfrCmr complex was incubated with ATP and AdoMet and the reaction was initiated by adding target RNA, followed by both HPLC and TLC analysis (Fig 3-11A and B). SAM-AMP was observed when both SAM and ATP were present, while there were no significant products in the presence of only ATP (Fig 3-11A trace i and B). The evidence presented here strongly suggested that BfrCmr system produces a previously uncharacterised conjugate of AdoMet and ATP, distinct from cOA or other cyclic nucleotides.

We also substituted AdoMet with S-adenosyl homocysteine (SAH) or the AdoMet analogue sinefungin (Vedel et al., 1978), which differ in the sulfur centre (Fig. 3-11C). SAH and sinefungin were also observed to conjugate with ATP (Fig. 3-11A, B and C), suggesting that BfrCmr exhibits tolerance towards the sulfur centre. Only SAM-AMP and not SAH-AMP was detected from *E. coli* cell extracts, probably due to the higher concentration of SAM (0.4 mM) than SAH (1.4 μM) in *E. coli* (Halliday et al., 2010). Additionally, the wild type BfrCmr complex demonstrated a rapid generation of SAM-AMP and SAH-AMP. After just 2 min of the reaction, 3 μM BfrCmr efficiently conjugated almost all of 500 μM SAM or SAH with 500 μM ATP (Fig. 3-12A and B).

The phosphodiester linkage of SAM-AMP could not be distinguished by LC-MS/MS, despite the likelihood of 3'-5' phosphodiester bond formation, as the Cas10 family enzymes catalyse the attachment of the 3'-OH group of a ribose unit onto the α-phosphate of a nucleotide 5'-triphosphate with the release of pyrophosphate (Kazlauskiene et al., 2017, Niewoehner, 2017). Both chemical and enzymatic methods are efficient ways to identify the linkage. Here, nuclease P1 as a specific 3′-phosphomonoesterase had been used as an enzymatic way to test the bond. The cleavage of SAH-AMP by nuclease P1 was complete, while only a small amount of SAM-AMP was degraded (Fig. 3-13). It's possible that the activity of nuclease P1 might be sensitive to the positive sulfur centre of SAM. This provides an explanation for the stability of SAM-AMP in cellular environments which have numerous nucleases. Hence, we favour the likelihood that SAM-AMP has a 3'-5' phosphodiester bond, although the possibility of a 2'-5' bond cannot be entirely excluded. Further confirmation through chemical methods such as NMR are necessary to resolve this definitively.
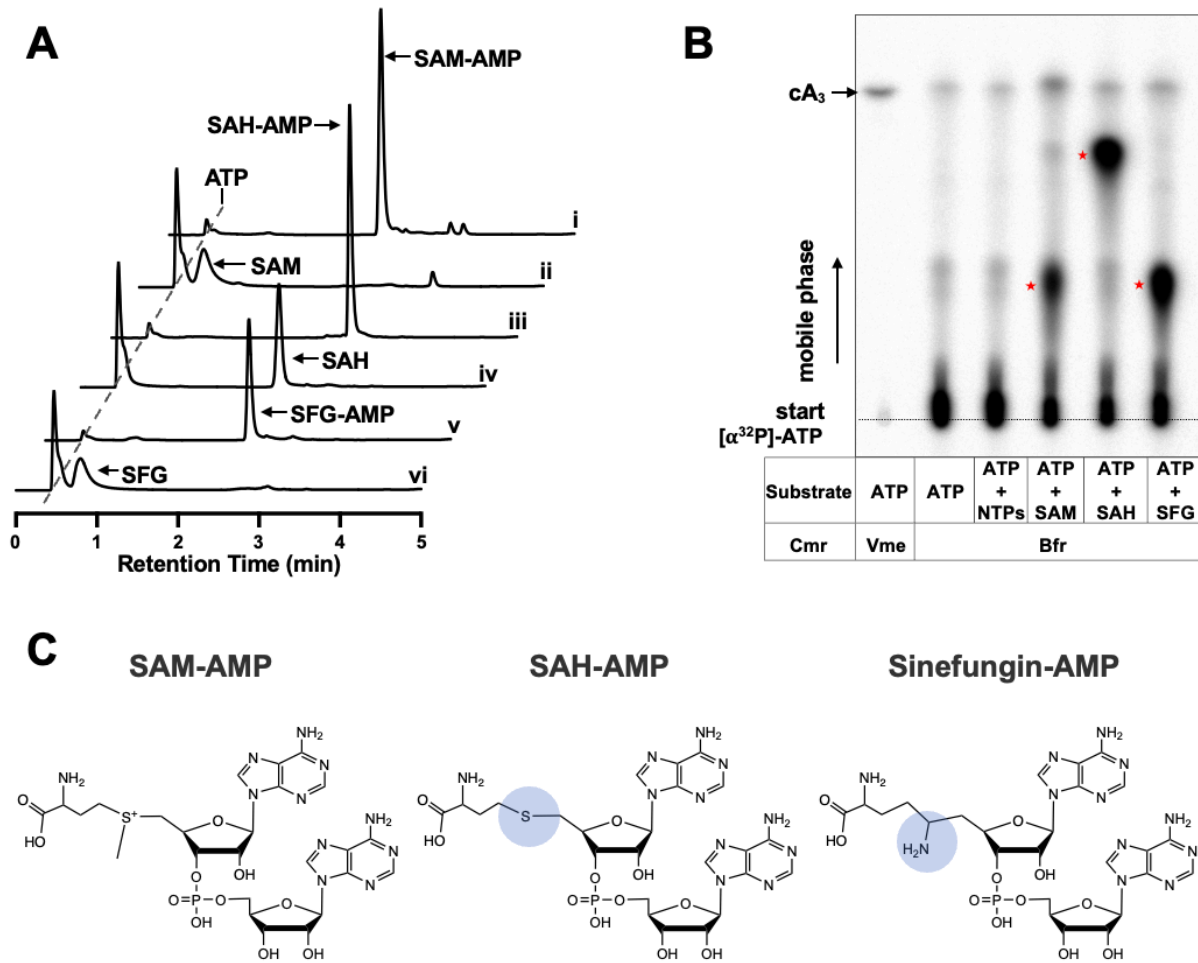
Figure 3-11 Reconstitution of SAM-AMP synthesis *in vitro*
A. HPLC analysis of *in vitro* reaction products. The purified wild type *B. fragilis* Cmr complex synthesises the signal molecule adenylyl-AdoMet (SAM-AMP) from ATP and S-adenosyl methionine (trace i). BfrCmr also conjugates S-Adenosyl-L-homocysteine (SAH) and sinefungin (SFG) with ATP (traces iii and v, respectively). Traces ii, iv and vi are control reactions in the absence of enzymes. **B.** TLC analysis of *in vitro* reaction products. SAM, SAH and sinefungin plus ATP yielded radioactive products (red stars) but ATP alone did not. cA_3 generated by wild type *V.* metoecus Cmr complex is shown for comparison. **C.** The proposed structure of SAM-AMP, SAH-AMP, and sinefungin-AMP with blue circles to highlight the differ sulfur centre.
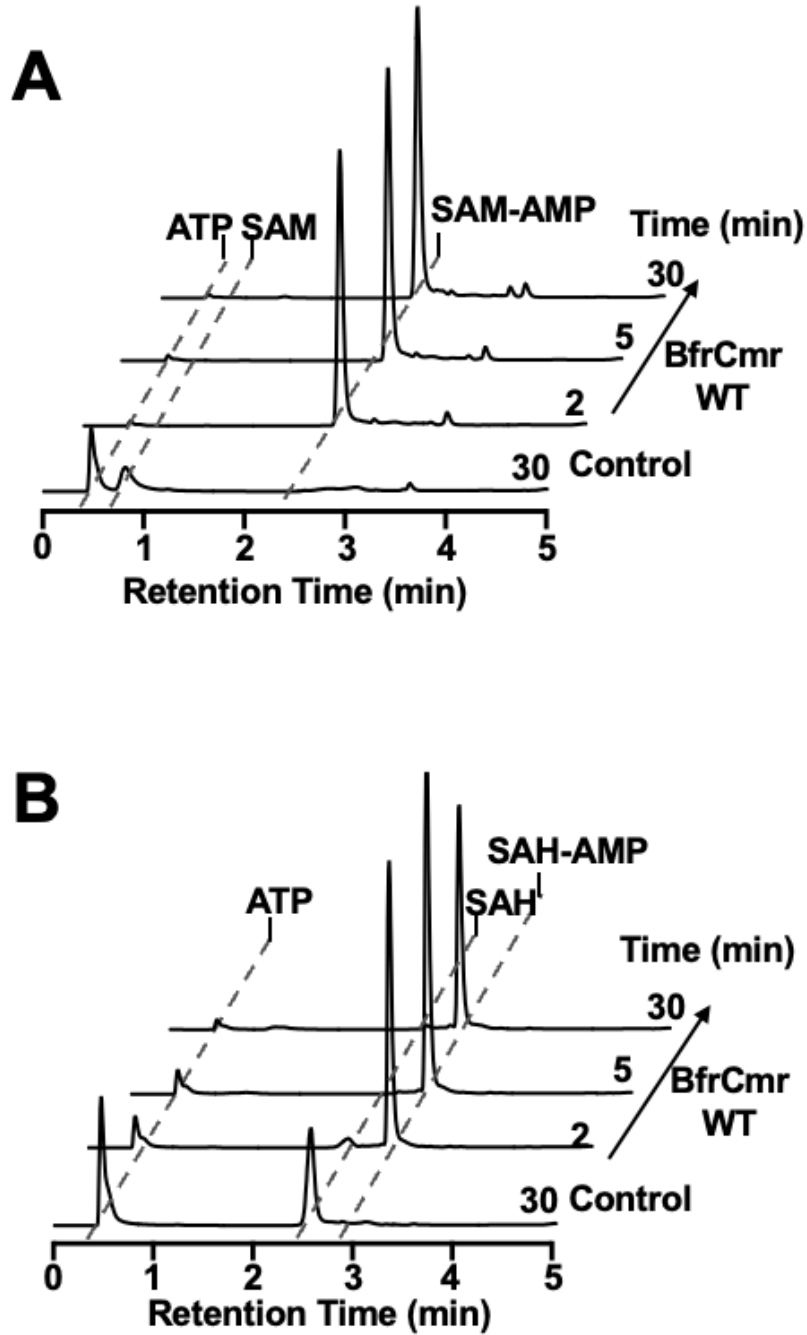
Figure 3-12 SAM-AMP synthesis rate of wild-type BfrCmr complex
A. 0.5 mM ATP and SAM were incubated with purified wild type BfrCmr (3 μM) in presence of Mn^{2+}. Samples were collected at the indicated time points and analysed by HPLC. BfrCmr was absence in control samples. **B.** as for (A), with substitution of SAH for SAM.

A
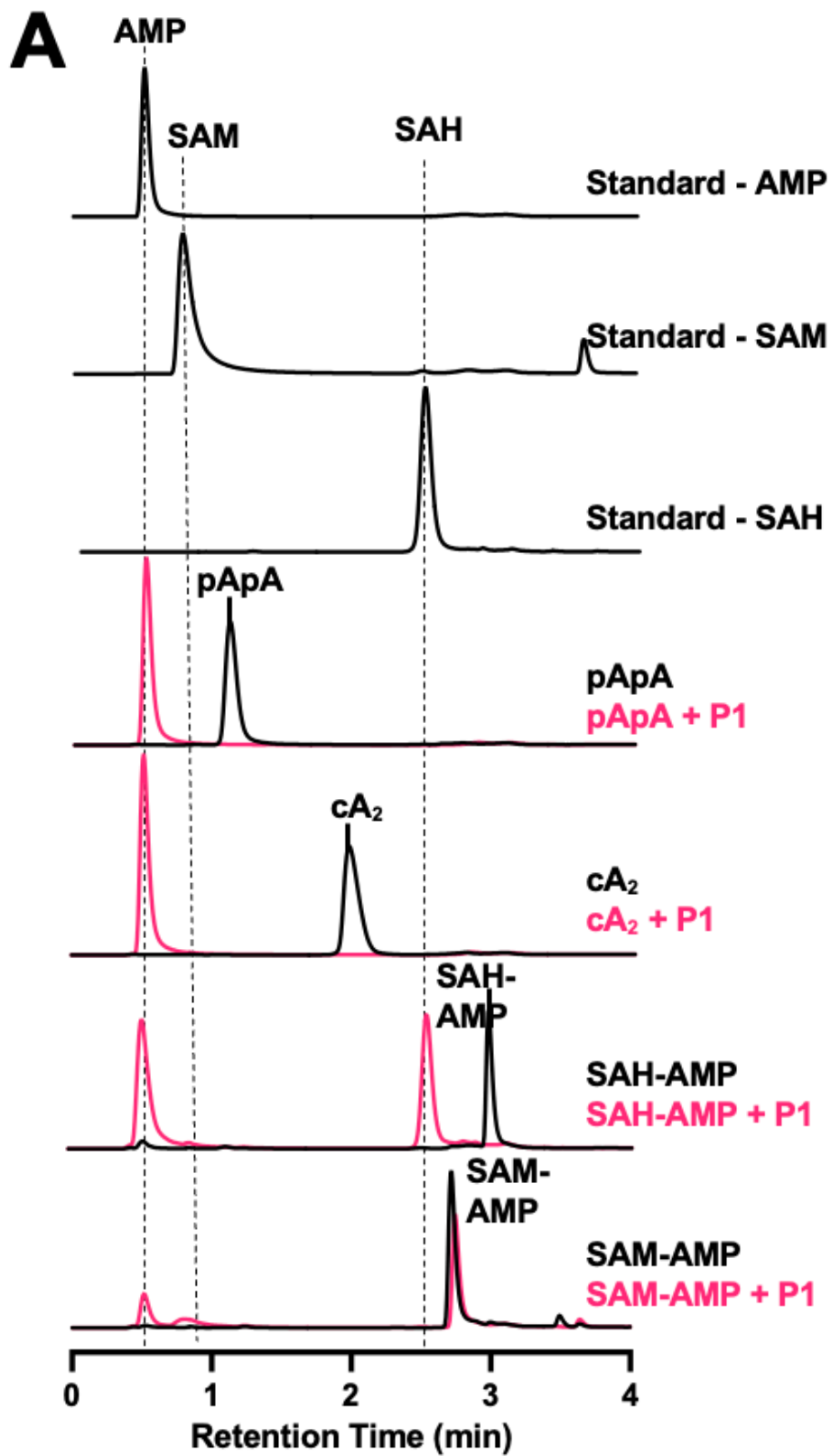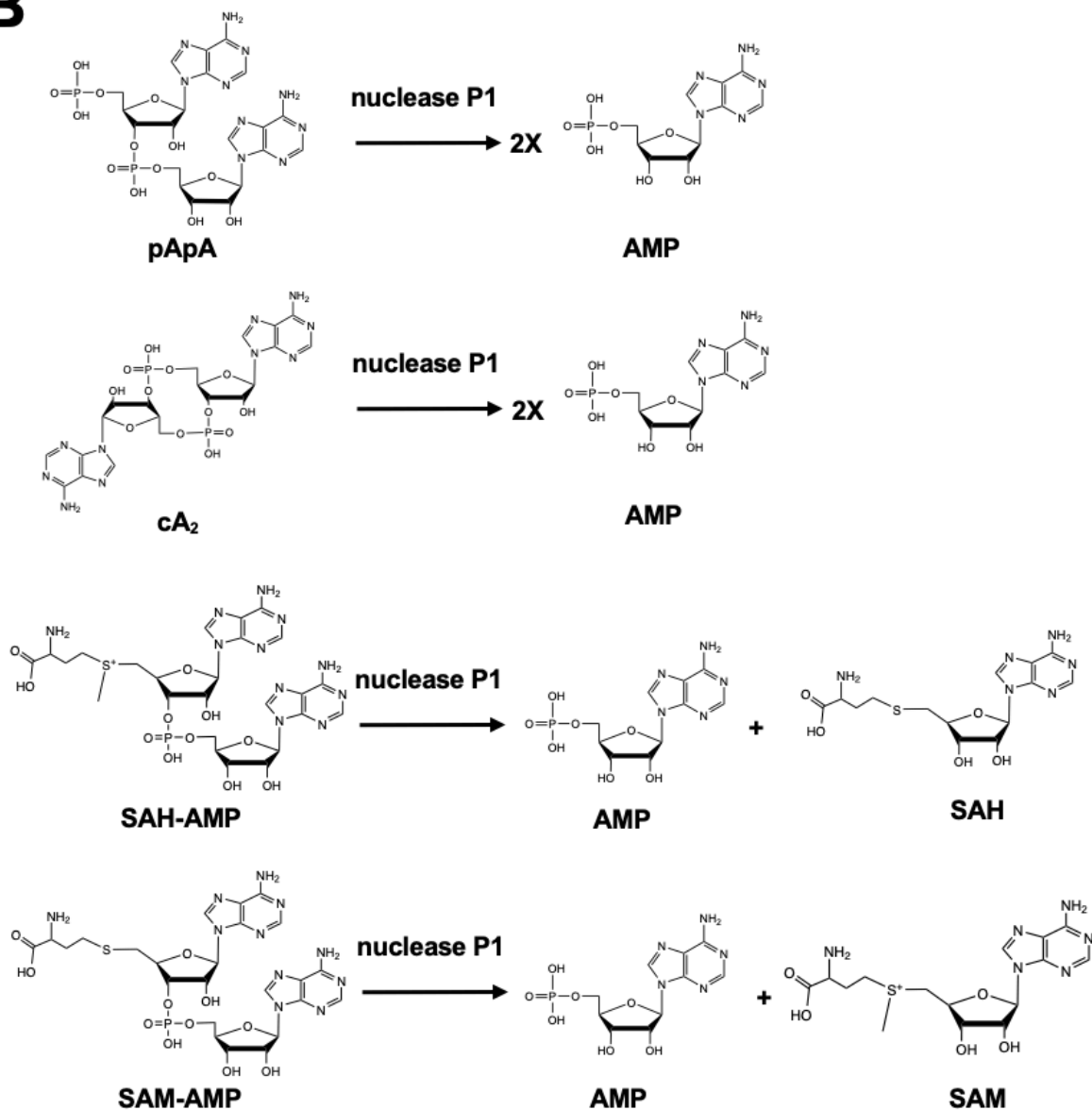
Figure 3-13 Analysing the linkage of SAM-AMP by nuclease P1-mediated degradation
A. HPLC analysis of nuclease P1-mediated hydrolysis reactions towards Cmr's products. cA₂ (cyclic di-3',5'-adenylate) and pApA (5'-phosphoadenylyl- (3'-5')-adenosine) were used as controls. **B.** The proposed reactions were shown on the right.

3.2.10 The key residues in BfrCas10 for SAM recognition

The acceptance of both SAM and ATP as substrates by BfrCmr complex, rather than ATP alone, raised questions that motivated us to conduct further investigation. The sequence and structure alignments were conducted by Prof. Malcolm White (University of St Andrews) to explore the possible sites for SAM recognition.

The Cmr2dHD-Cmr3 complex from *P. furiosus* was the first structure to show the accommodation of two ATP molecules within the crevasse formed between the two proteins (Osawa et al., 2013). One ATP (ATP1) had been observed binding in the donor Palm pocket, where the ribose moiety is recognised by D274 of Cmr2dHD, and the triphosphate group is on the opposite side of the GGDD motif. The second ATP (ATP2) molecule in the acceptor Palm pockets is loosely recognised by the complex as compared to ATP1, since only one hydrogen bond has been observed between the nucleobase and Cmr2dHD. Superposition of the structures of apo-form and one ATP-bound form of the Csm1-Csm4 complex from *Thermococcus onnurineus* further demonstrated that ATP1 was cooperative binding in the donor Palm pocket of the complex, and subsequently caused acceptor pocket a significant conformational change to accept ATP2 (Jia et al., 2019a). A pair of Palm domains hosting two ATP molecules enable the 3'-hydroxyl group (3'-OH) of acceptor ATP to target the α-phosphate of donor ATP to eliminate pyrophosphate, forming a 3'-5' phosphodiester bond. The structure underlying the mechanism of cOA formation indicated that the molecule SAM more likely occupies the acceptor palm pocket, with ATP in the donor position (Fig. 3-15B). However, the major structural difference between ATP and SAM is the replacement of the triphosphate group by the methionine moiety, resulting in the local charge shift from -4 to +1 in the ligand. This variation suggests the involvement of less basic protein residues in methionine moiety recognition. According to the sequence alignment of BfrCas10 with its orthologues, two highly conserved acidic residues, D70 and E151, were observed with a potential role in recognising SAM in the acceptor site (Fig. 3-14). The predicted BfrCas10 structure also suggested that these two residues were located adjacent to the methionine moiety of SAM (Fig. 3-15B). In comparison to the structure of PfuCas10, D70 of BfrCas10 corresponds to the N300 in PfuCas10 which is in the vicinity of the β- phosphate of the acceptor ATP1 ligand. Likewise, the residue E151 in BfrCas10 occupies a position equivalent to R436 in PfuCas10 which forms bidentate hydrogen bond with the γ-phosphate (Fig. 3-15A and B).

We thus constructed BfrCas10 variants with mutations D70N, E151R and the double mutant in the text of BfrCmr complex, following the same expression and purification procedure as

the wild type. The purity of variants was verified by SDS-PAGE analysis (Fig. 3-6). The SAM-AMP synthesis ability of the wild type and variants of BfrCmr complex were assessed by incubating them with SAM and ATP or ATP alone. The mutation in E151 had a limited impact on the SAM-AMP generation ability. In contrast, the D70N variant's synthesis ability was reduced to half of the wild type, and the double mutant had nearly lost its activity (Fig. 3-16A and B). Moreover, a higher amount of pppApA was detected in the double mutant when incubated with ATP alone, as compared with wild type. This finding indicated that mutations in residues D70 and E151 subtly influenced the preference for ATP over SAM (Fig. 3-16C). Further structural analysis is required for a deeper understanding the reaction mechanism and substrate specificity.
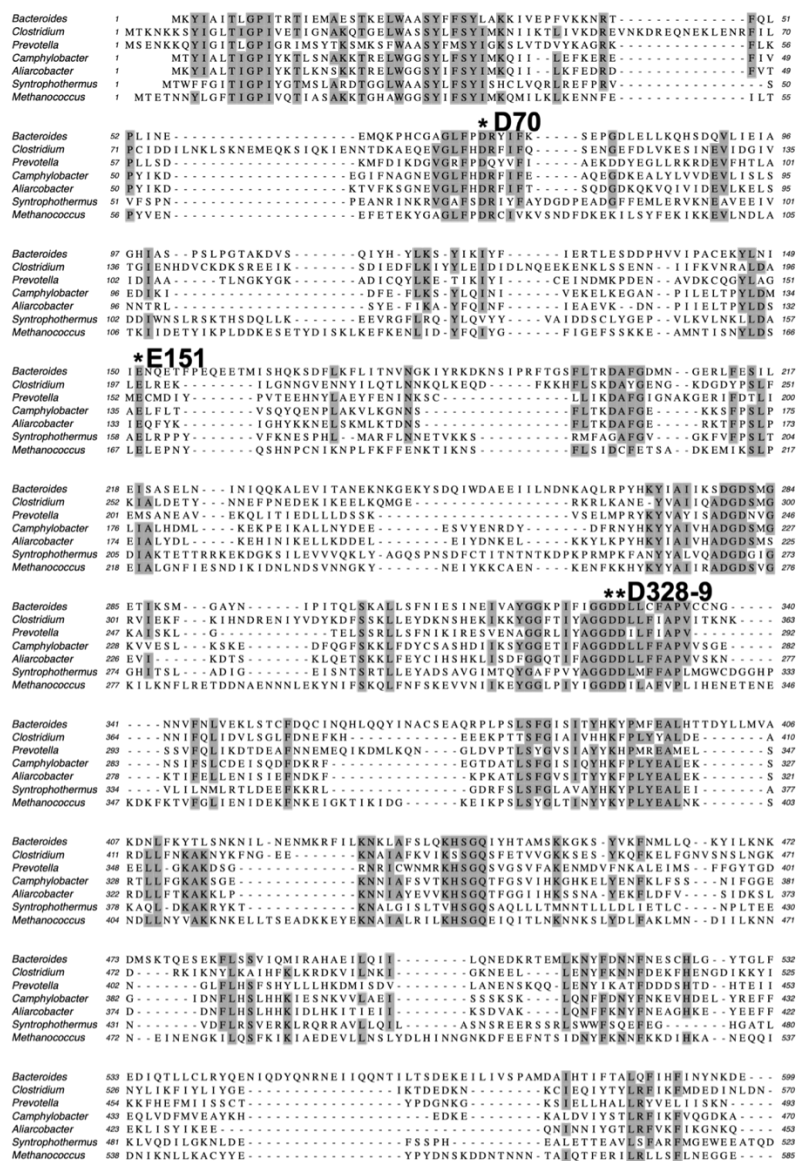
Figure 3-14 Sequence alignment of BfrCas10 with its orthologues
Sequence IDs are: *Bacteroides fragilis* ANQ60746.1; *Clostridium botulinum* WP_011986674; *Prevotella - Xylanibacter muris* WP_172276208; *Camphylobacterales bacterium* HIP52383.1; *Aliarcobacter butzleri* WP_260918755; *Syntrophothermus lipocalidus* WP_013175521; *Methanococcus voltae* WP_209731901.

88
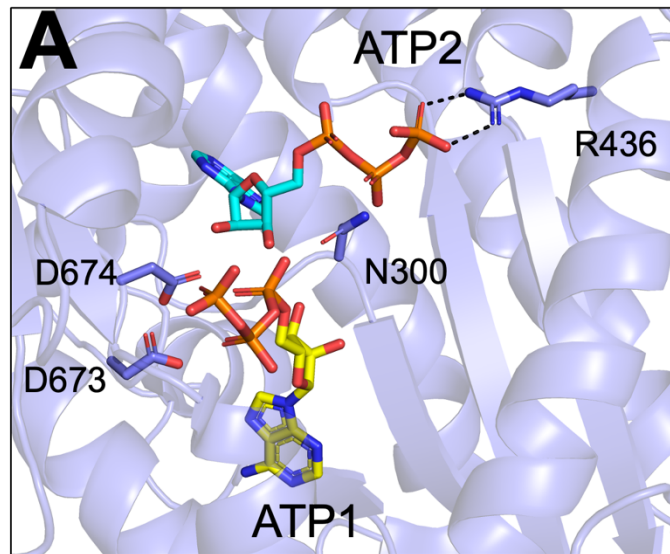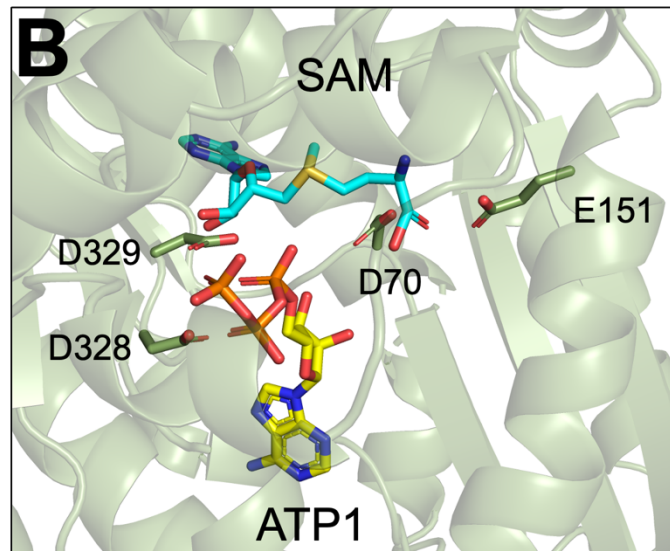
Figure 3-15 The potential key sites of BfrCas10 for SAM-AMP synthesis
A. The crystal structure of the *P. furiosus* (Pfu) Cas10 subunit with 2 ATP molecules bound (Osawa et al., 2013). Side chains for the two metal binding aspartate residues of the "DD" motif, together with residues N300 and R436 that interact with ATP2, are shown. **B.** Equivalent view of the AF2 model (Jumper et al., 2021) of the BfrCas10 structure with ATP1 from the PfuCas10 structure and ATP2 replaced by SAM. The precise conformation and position of SAM is unknown. The conserved acidic residues D70, E151, D328 and D329 are shown.
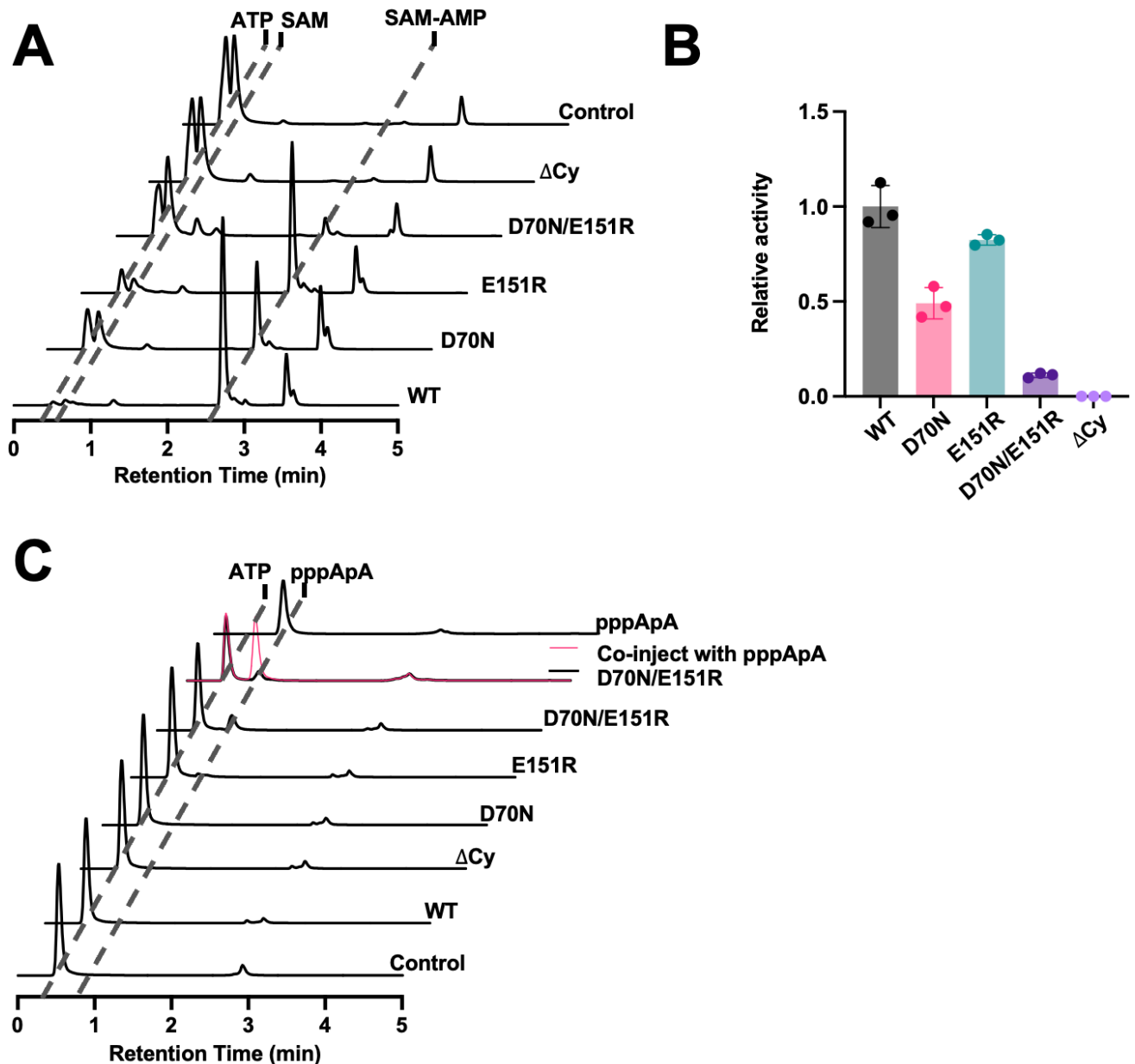
Figure 3-16 Assessment of SAM-AMP synthesis ability among the wild-type and variant BfrCmr complexes
A. SAM-AMP synthase activity of purified wt and variants *B. fragilis* Cmr. The reaction products were analysed by HPLC following incubation of 2 µM Cmr with 0.5 mM ATP and SAM for 30 min. **B.** Relative SAM-AMP synthase activity of Cmr variants. Three independent experiments were carried out, with the mean and standard deviation shown. The relative activity was measured by quantification of peak area of product SAM-AMP and then comparing with BfrCmr wild type that is normalized as 1. **C.** pppApA synthase activity of purified wt and variants *B. fragilis* Cmr, analysed by HPLC following incubation of 2 µM Cmr with 0.5 mM ATP alone for 30 min. The D70N/E151R double mutant synthesises pppApA but not SAM-AMP.

3.2.11 Production of the membrane protein BfrCorA and its SAM-AMP binding affinity

CorA family proteins function as the major cation channels responsible for transporting magnesium ions (Mg^{2+}) in prokaryotes. Structural and biochemical investigations have revealed that CorA family proteins form homo-pentamers, possessing a substantial cytoplasmic domain with the regulatory function and a membrane-spanning domain that creates an extracellular pore entrance. Each protomer consists of two transmembrane helices connected by a short loop bearing the signature motif GxN, which is believed to serve as the selectivity filter (Eshaghi et al., 2006, Matthies et al., 2016, Lerche et al., 2017). CorA from *B. fragilis* (accession number is WP_005787774.1) shares structural homology with CorA family proteins mainly because of the presence of two transmembrane helices at the C terminus (from 387 to 488), which contains a conserved signature motif GxN essential for the Mg^{2+} uptake. Conversely, the cytoplasmic domain shares limited sequence or structural similarity to any known proteins. Interestingly, the gene encoding BfrCorA is positioned adjacent to the CRISPR loci of *B. fragilis* where it employs SAM-AMP as its second messenger to confer immunity. BfrCorA thus can be predicted to function as an effector regulated by this second messenger.

To test this hypothesis, BfrCorA was first expressed and purified from *E. coli* to near homogeneity in the presence of detergent DDM (Fig. 3-17A, B and C). The purified BfrCorA was then incubated with radio labelled SAM-AMP, SAH-AMP, Sinefungin-AMP, cA$_3$, or BfrCmr-mediated ATP reaction products, respectively, followed by electrophoretic mobility shift assay (EMSA) (Fig. 3-18). The presence of retarded species near the wells was noted when SAM-AMP and SAH-AMP were subjected to incubation with increasing concentrations of BfrCorA, whereas no such observation occurred with cA$_3$. This finding suggests the intriguing possibility that BfrCorA may specifically bind to the SAM-AMP second messenger to provide immunity.
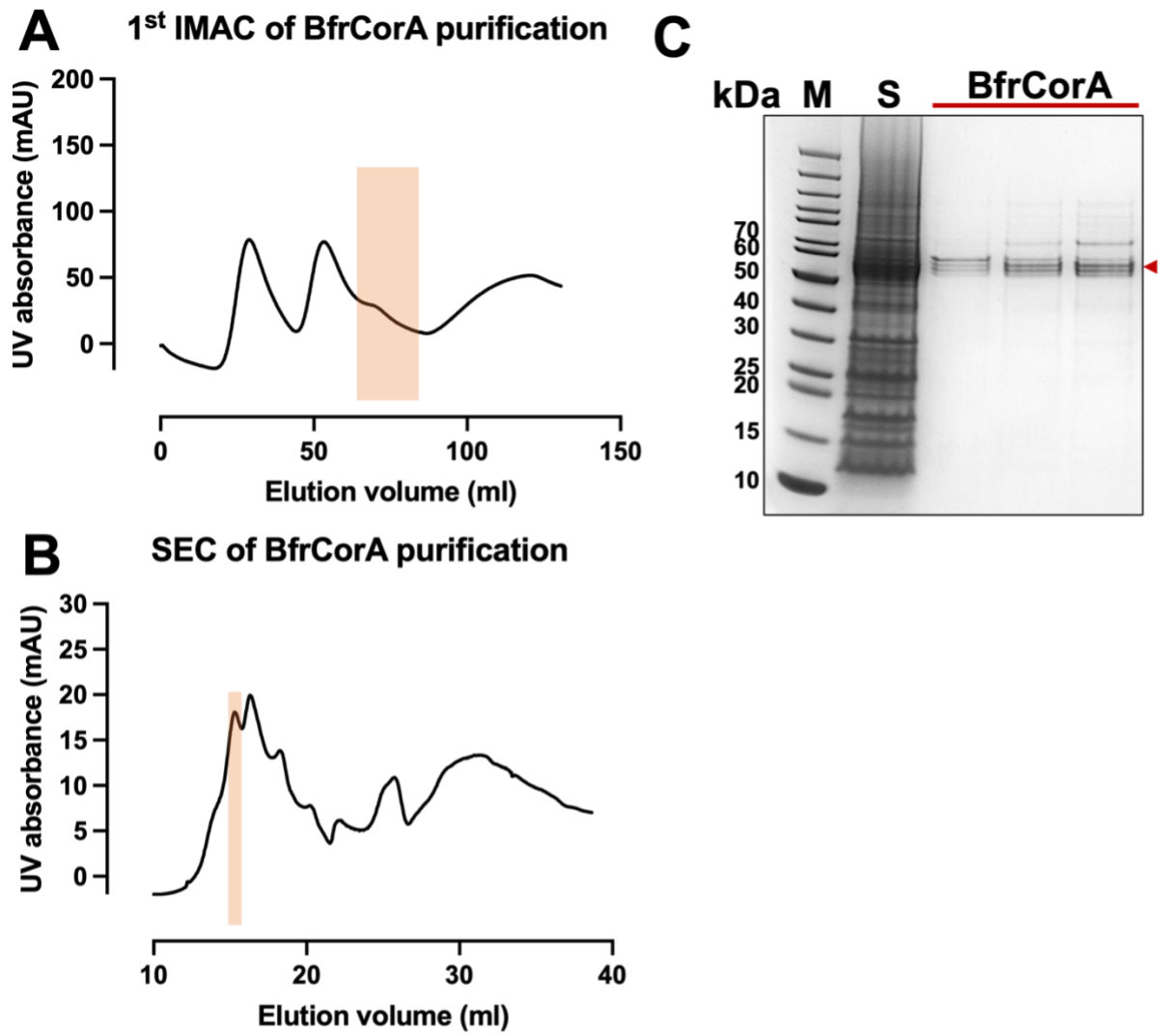
Figure 3-17 Purification of BfrCorA wild type
A. The first immobilised metal affinity chromatography (1st IMAC) of BfrCorA. The fractions containing target protein highlighted with a red rectangle was eluted with 50% elution buffer and pooled for his tag removal. **B.** Superdex200 SEC profile of BfrCorA. The TEV-cleaved protein was recovered from the nickel column and then subjected to SEC. Fractions indicated by a red rectangle were collected and concentrated for further enzymatic analysis. **C.** SDS-PAGE analysis of the final SEC step for purification of BfrCorA, which was used for SAM-AMP binding assays. The four tightly spaced protein bands in the gel all correspond to CorA, perhaps indicating limited proteolysis of the termini. M is the marker to indicate the size on the gel. S indicates the sample applied to SEC. The horizontal red bar indicates the three fractions pooled for further analysis.
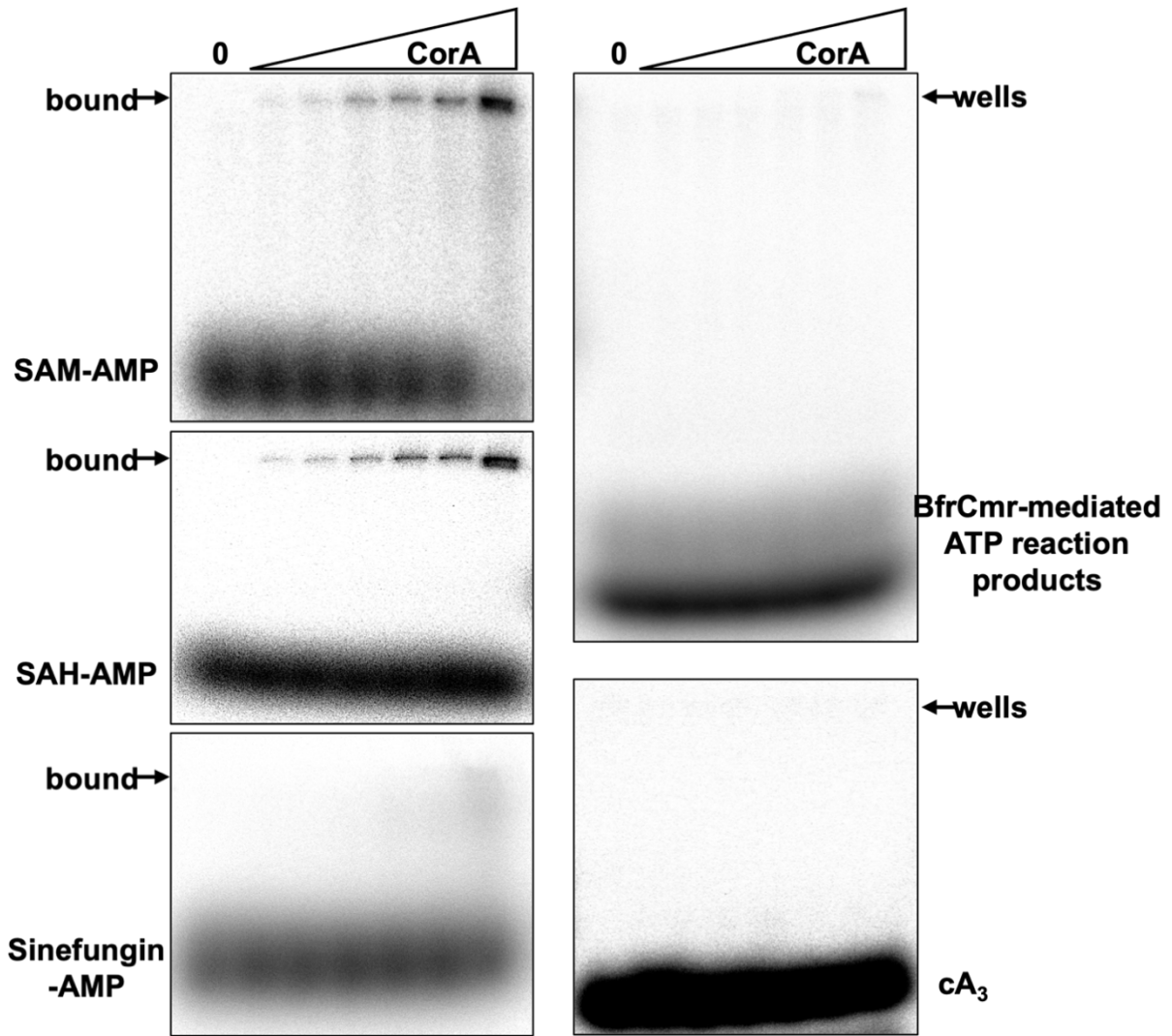
Figure 3-18 SAM-AMP binding of membrane protein BfrCorA
CorA binds SAM-AMP and SAH-AMP, but not Sinefungin-AMP, cA_3 or BfrCmr-mediated ATP reaction products (1 µM ^{32}P-labelled ligand incubated with BfrCorA at an increasing concentration 0, 0.0625, 0.125, 0.35, 0.75, 1.5, 3.3 µM), illustrated by acrylamide gel electrophoresis and phosphor imaging.

3.2.12 The key SAM-AMP binding residues in BfrCorA

Sequence alignment and structure modelling were conducted by Prof. Malcolm White to facilitate further detailed investigation. Conserved residues were identified by comparison of sequences of BfrCorA and its orthologues and the structure of BfrCorA was predicted using Alphafold (Jumper et al., 2021) (Fig. 3-20 A, B and C). Upon mapping the conserved residues in the structural model, a pair of arginine at positions 152 and 153, and a pair of aspartic acid at positions 219 and 220, were observed at the interdomain surface, indicating their potential binding ability (Fig. 3-19A, B, C and D).

BfrCorA variants were then constructed by introducing alanine mutation into two pairs of conserved residues (R152/R153 and D219/D220) to investigate the impact on their SAM-AMP binding affinity and plasmid immunity. The expression and purification of BfrCorA variants followed the same procedure as the wild type, but no production of purified proteins could be obtained. Due to the challenging nature of purifying the wild type BfrCorA, investigating the underlying reason for this issue proved to be difficult. Western blotting demonstrated similar *in vivo* expressions level among the BfrCorA variants and wild type (Fig. 3-21A). The variants were thus tested in the plasmid challenge assay. The gene encoding wild type BfrCorA in pRATDuet derived plasmids was replaced by the genes of BfrCorA variants, before challenging the *E. coli* BL21 star harbouring pBfrCmr1-6 and pBfrCRISPR_Tet (or pBfrCRISPR_pUC as a non-target control). No immunity was provided with the presence of BfrCorA variants with wild type BfrNrN together in the activated BfrCmr system (Fig. 3-21B). When only variant R152A/R153A was introduced into the wild type BfrCmr system, the cell growth in the inactivated BfrCmr system (non-target induction) was back to the similar level as in the target system (Fig. 3-21B). This observation suggests the possibility of a loss of the toxicity observed in the presence of the wild-type BfrCorA. Conversely, less observing number of transformants in the activated BfrCmr system was shown when challenged with BfrCorA variant D219A/D220A (Fig. 3-21B). These data implied that these two pairs of conserved residues play a vital role in the signalling-mediated immunity. Although the mechanism of BfrCorA effector has not been determined, BfrCorA is more likely a ligand-regulated ion channel, leading to membrane disruption and cell death or dormancy upon SAM-AMP activation to confer immunity.

```
                         *
MBV4191239.1    1   MIYSYHIFYFPFKWEIMG-----LENQAFSDQVNLDNIQ--YNR-NSHWERSQ-KP----------DPGEEESLYNE   58
HJC97925.1      1   MTHSYHIFYFPFRWDITG-----VRLRTLSERIDLERIH--YKP-HSEWVHTQ-NPM---------SDNEAVDLYNE   59
WP_172276200.1  1   MTYSYHIFYFPFKWKLYN-----GKELLFSEQTDLAKIN--IRN-DSFWIRQQ-EPV----------NYDELKSLYNE   59
MBR1964303.1    1   MKQYSSHIFLFFPFKWDLRG----IE-YCM-PSYWERMH-KPD--------NIEDEKVLYNE   60
GHU64252.1      1   MYSYHIFYFPFKWELKD-----KKYRLFSERTNLKNIG--FDETFSNWLPNP-APL--------NKQEADELYNE   59
MEH5455609.1    1   MEHSCHILLFPFQWRIND-----LDKKDFSKQISLKNIA--FDP-NSMWERSITQ----------ATDEEKMLFNE   58
MCD7916322.1    1   MSKPYSYHIFYFPFKWEIEN-----RKDKLFSEQVNLESLH--FSE-NDGWERVQYNQPGDFRLMNEQQKQEVRELFAE   71
WP_186928207.1  1   MSTYSYHVFYFPFKWELPG-----EDKKFFSEQIDLKHIP--INE-YTLWDRVQYDPAEKYIPVVGEDTKDREELFGE   70
WP_077492690.1  1   MGAETYSRHIFIFPFQWDYIE----KRGYASYNERTDIKKFDKDLFEA-VSPLKRQYF-----------EIGASEERYNE   63
PID93465.1      1   MGYHSKHTFMFPFQWDYID----KKRDFYNNRTNLADFQRLFEAIDNGFQLTPF-----------KIGYDVDKYNE   61
WP_036792078.1  1   METPKHSRHIFIFPFQWDYIDHPSGKKKNTSYGKRTDLDAFDRLFTK-VSPLKRRYF-----------EIGLSAEHYSE   67

                    **    *
MBV4191239.1    59  KNYYYTFVHNILYDE-EHSPLNLIHHFERKEPKLSN-HIYYYIKKKGRNNPYK-----LIVDAMNINLYATGVGFLSFYL  131
HJC97925.1      60  KNYYYRFVHNILYDE-PGNKFNLIRHYERKEPQEST-NVRYLIKVKNRQQPYE-----LKVDAMNINLYSTGVGFLSFYL  132
WP_172276200.1  60  KNYYYKFVHNILYDEADGSSSAMIYHFERDEPRQSD-KVRYVIKVKGRAKPYS-----LAVDAMNINLYATGVGFLSFYL  133
MBR1964303.1    61  KNYFYKFVHPILYD--EGGSKSLIHHFERKEPKLSNGNVTYLIKTKNRKEPYR-----LRIDAMNVNLYDTGVGVLSLYL  133
GHU64252.1      60  KNYYYQFVHPVLYD--TGEPDSILKHFERKEPQINE--VFYKIARRG-DKTYS-----LKVDALNLNLYATGVGMLTFFL  129
MBP5455609.1    59  KNYFYEFAHGALFDD-GEEEGNLMQHYERKETDSGN---VLFSFGHGKKTYT-----LHLKYLNLNLYETGVGVLSFYL  128
MCD7916322.1    72  QNYYFDFVHPVLYDI-QGKPDPMIAHFERKETKEESRKVTYQIKVKD-KQEYV-----LDLEAINLNLYSTGVGVLSFFL  144
WP_186928207.1  71  LQYYFDFVHPVLYDI-KGAQNPIIYHYERREPKAGN--VEYRIAVGD--KEYV-----LKVDAINLNLYATGVGILSFFL  140
WP_077492690.1  64  YTYFHPFARRALF---GTKENPSVYHYELSDTQGSY---NISVKIDGVSTTYS-----LQLEKIAVSVFDTGVGVTAYFL  132
PID93465.1      62  FSYFHPFARKALY---YTEDNRLLRYYEIEHPNGEY-HIEY-LKEGNNTKAYENQLLALTLDNICVHVYATGVGVVTFNI  136
WP_036792078.1  68  YVYFHPFARKALY---GTKENRSIYFYELEVDGGRY---CIDINTGSLSVTYE-----LRLESVSLHAFNTGVGVISFSL  136

                    **   R152R153
MBV4191239.1    132 KNEDCTQNSPEDILAINQYGRRIMPPFFNDT--RLRNEISEYIRIEGLNQTV----YFEDF----------KSYTP----  191
HJC97925.1      133 KNERDDQQAPDDILKINQYGRRIMPPFFADI--KFRNETSEYIAIQGLLPNN---EYQEDF----------KDHEP----  193
WP_172276200.1  134 RNDDSSQKEPDDILRINQYGRRIMPPFFDDI--KNRYELSESIGIEGLWPVK---EYKDDF----------SSYSV----  194
MBR1964303.1    135 MNEIESQSSPEDILAINQYGRRIMPPFFGDI--QYRDEIAEYLSIEGLQSIG---DYYEDF----------NGYKT----  194
GHU64252.1      130 KNEREDQKSPEDILAINQYGRRIFPPFIADI--DLRGEIAEYLSIEGLYGDPI--RYKEDF----------SGYTN----  191
MBP5455609.1    129 YNDT--YEDPDDILTINQFARRVYPPYWEEV--ENRKLLADFYLNEGLGNGS-----YEENF---------QNYTL----  185
MCD7916322.1    145 ANKATTHCEEEHIRDINQFGRRIMPPHYGEI--ISRGMLSEYICIEGLNGYAG--KYREDF----------SGYTP----  206
WP_186928207.1  141 ANNEESQKEELDVRNINQFGRRIMPPHSGEFSAENRSMLSKSIKIEGLHGNPS--KYMDEFNYKIHTTGQDSEFGL----  214
WP_077492690.1  133 RNEK--HTSFEDILRINDSGRRIYPQYMAK---GNRLKAKDSFLADRIYGSLGVLDFDDDF---------RQFEEDISH   197
PID93465.1      137 TNYK--YPNKEAILAINEFGRRMYPQFMSK---GNRLAAKGVFLPNRLWGRIADFEFDEDF---------SAYQLDIQT   201
WP_036792078.1  137 YNDK--YVEEEDVLRINDFGRRLYPQFMAS---DDRLKAKTTFLANSISGSIGDMSFYDDF---------SQFEQLINH   201

                                      **  D219D220
MBV4191239.1    192 YDSWQPSSSIKKLIC-------------------ELVTNLSIDPIIDDRMFVATWYKNNQLSQQFTNNAKAYFDSQDP   250
HJC97925.1      194 TDSWKPASFIHTLIM-------------------ELADNIEIAPIIDDRLFVICWYKNDKLAAHFTEHPQAYCDTNDD   262
WP_172276200.1  195 DDSWKPASFVNELIM-------------------ELADNIVFEPVIDDRMFVASWYKNDKLADMFTENDEAYMNPADG   253
MBR1964303.1    195 TDSWHEASFLTKLIY-------------------ELADNIDIEPIIDDRMYVLSWYKHASFATQFSSNPNAIFKSSK-   252
GHU64252.1      192 KKSWQPSCFIRNLIG-------------------DLSETLKIEPIIDDRMFVNCWYGNDELSDKIKNGKLNIDKSCSG   250
MBP5455609.1    186 ADDNIPASYITHLIK-------------------DLATNIDIIPVVDDRMYVLSWYKNDKLSKEFSASEE--IEPLK-   241
MCD7916322.1    207 SDVWKPASFICQLIA-------------------DLHPGMIATPVIDDRMFVNSCYVNKDLSKEIAVDDPSLEKERLK   265
WP_186928207.1  215 DLSPDMKVTPVIDDRMLVNCWYGNDELSKQVKDMTGNLKDGEFV   273
WP_077492690.1  198 TNIFLPPDHIRQVFGYNKMDRTEDDGCQKFVFHQESERKDIIRISPVMDDRMFFITYYHSPYYAKKLRSKPKNFRRKREY   277
PID93465.1      202 DGIFLPPDHIKKVFGYKGKEQIGDHGQ-LVFVFRKREEQKGTIKISPVTDDRMFFLSHYNKADLVDQLAKRTDPEKEPIAY   280
WP_036792078.1  202 TDIFLPPDHIRQVFGYKPTDKLGDDRC-KFVFRNSDEETGRIRISPVMDDRMYFLSCFYSRELAESLKESERNFSREEKI   280

MBV4191239.1    251 -------------------FSDYWYRFLFIDGSN-ATCQNEKMKKELLEEHTYYRWQ-------QWSSLYGISKYSLVYLT   304
HJC97925.1      253 -------------------FSTFWYKFLFIDGTD-ETCQNDTMKTELLKRHTYLRWQ-------KWTSLYGVSKYSFVYLT   306
WP_172276200.1  254 -------------------FSSFWYKYLFIDGSF-ETCQNDRMKKELLDSHTYTRWQ-------KYSSLYGVTRYSFMLLT   307
MBR1964303.1    253 -------------------LSSFWYRFVFVDGKS-VTCQNDMKASLLNKATYMRWQ-------NLGSLYGVSRYSMVYLA   306
GHU64252.1      251 DFS-------------GEDDFWYKYIYVGVTY-ASCQNSELRKKLTYSQTYTRWQ-------KYGTLYGISRYSFVSIS   308
MBP5455609.1    242 -------------QNDFWYKYVFIEAEE-MICHNRKMRNRQIEQSTYLRWQ-------GKGAFYGCSRYSFVFLT   295
MCD7916322.1    266 EFI-------------LGKFWYKYVYVDNGQWDMCQHDPMKEQLLNRATYFRWQ-------KYGTLFGISRYSFVLLK   323
WP_186928207.1  274 -------------MGNFWYKYVFVDDGNDDTCRNNEMKKKLLEESTYYRWQ-------EQGTLYGISRYSMVALT   338
WP_077492690.1  278 NIY-----------EEGRFGDLWFRYLFGDGGD-KTVANTSVQREDVRRTTYARWS-------GSGTLFGLTKDSLVGIS   338
PID93465.1      281 AFE-----------LDNFWYAYIFGDKEG-PSIADDRMQRNDILKATYTRWLDTEKRGWGSGSLYGISRDSFVSIG   344
WP_036792078.1  281 QGGLCFLPEMGYAYEVGKKRDFWYRYLFGDGLF-KGVASDRLQLEEIKRCSYERWI-------EAGTLIGITRDSAVYTI   352

                                                                                  *    *
MBV4191239.1    305 N--NEVPDYLIE-YFQTIYARMAELVLVQRASMLRFSGEITKVSQLS--NQDVEVVSKRVSSLYKEYIRFVNQIYFREIT   379
HJC97925.1      307 N--STVPAYLLS-YVQTIYARMVELVLAQRASMLRFSGEVTNVSNLS--HQNVDEISSLISSLYKEYIRPMNQIYFREVT   381
WP_172276200.1  308 N--SSAPDHLLE-CFQTIYARMAELVLVQRASLLCFSDEVTRVSNLS--NQEVGVISERISSLYKEYIRFINIVYFRDVT   382
MBR1964303.1    307 G--EGVPDFLLS-NFNGVYARMAEIVIVQRATMLRFSGEITNVSHLS--NKNIPDVFERINSVYKEYIRFINQIYHRDIT   381
GHU64252.1      309 DESDYSKNKLTAVHLRTVYARMIELVIVQRASILKFSEEVTRVSSLSKNKQTDTQLVDKISSLYNEYIRFVNQIYFREVT   388
MBP5455609.1    296 N--SRCPDFLIS-NFETEYVRMAELVLVQRASILRFSEEISVLSSD---KFDNTYLSKRVNSLRKEFIRFVNQVHFREVT   369
MCD7916322.1    324 G--IG-AEYIET-HMATLYSRMIELALVQRASMLRFSSEVTSLSNLT--GQDRQKLLQKISSLYKEYIRFVNQVYFREVT   397
WP_186928207.1  398 DTSWFASNILAM-HMRTIYSRMFELIIRTDRRLAERIGSLYQEYIRFINQIYFQHVT   405
WP_077492690.1  339 D--W-----YLIGDHITTMYYQMATLSLVQRSTILRFSNEIASITQLILDEKTKRKTEKQVKELYGNYIRFLNETYFREVT   412
PID93465.1      345 T------DFIGQ-HMQTMYYQMAILCLVQRASILRFSWEISQIINRLFANKKVKP-QAIRELYENYIRFINQIYFREIT   416
WP_036792078.1  353 G--W-----DVIEG-HFMSMYYQMAILCLVQRASVLRFSGEISYLTSQI--RRGKRKAGSNIKELYENYIRFLNELYFREVT   424

                    *  *                                                      *    *
MBV4191239.1    380 AQDQGIEMYNKLHSCLQMESYIKDLDGEIEELHQYISLMEDRERNKKASLLNDIATLFLPITVITGFWGM-NQISEVMEE   458
HJC97925.1      382 AQDQGIELYDLLHSCLKMETYIKDLDDEIEELHQYVSLKEDRSRNKKASVLNDIATLFLPLTVITGFFGM-NLWEDIFDK   460
WP_172276200.1  383 AQDQGIELYHKLHSFLNMDECIHDLDNEIAETLHQYVSLNDEKATRLNDIATFLPVSVITGFWGM-NSIEAVT--   460
MBR1964303.1    382 AQDQGIELYAMLHECYKMDDCIKDLDHEFQELHEYVSLNEDRIRNNKATILNYVATVFLPISIITGFFGM-NSWCNVLCE   460
GHU64252.1      389 TQDQGIELYDLLMSKTLKTEEYIKDLQEIELYLQYVSLVEDKERNRNGAMLNTIAAIFLPATLIAGLFGM-NKSDDLG--   466
MBP5455609.1    370 AQDQGIEIYEKLYKFADLQRHVEKLDAEIEELDNYVSMVEDRNNINRIAGNLSMVASFFVPATVLSGIFGMNNTWANGLND   449
MCD7916322.1    398 AQEQGIELYDLLLQQLNSEKQIEYLESEIRELHQYVAIVEEQTRTGRMEWFSWLAIIFVPVSLIIGVLGM-NSIFDKEYL   476
WP_186928207.1  406 VQDQGIEIYDMLMQQFASSEKIKELDGEIGELHQYITLLDQKRRNEGEWLNWLAAVFLPATVITGIFGM-NPFGA----   481
WP_077492690.1  413 PQIQGIEMYDMLQSNMNIEKDAKALDEEMGELFNYTTLEEQ------GVLNKTANLYLPITLLVGFLGI-NTFVEKGCI   484
PID93465.1      417 PQIQGIELYNLFQKEMNLEKEAKDLQEMQELFNYLSVKEQ------THLGQMANLFLPLTLLIGFFSM-HSINDHFFN   488
WP_036792078.1  425 SQIQGMEMYEMFQKNMNIERDTKALEGEMTELFGYLDIDEQ------SKLNRIANLYLPLGLLTGVLGI-NTFAKGGIF   496

MBV4191239.1    459 N----GELSTGFIIQ-----SLLLIIGTLCA--ICIIYKRKR-KL   491
HJC97925.1      461 ------ETPNGLFWK-----TFILLVGFICA-LIVIRNRNK-KL   491
WP_172276200.1  460 ------GEQESFKWQ-----CLLIFVGVIAS-LLVIYNKKK-RL   491
MBR1964303.1    461 ---------NCGLEWQ-----IITIFAGVLISVGVLLIMNKKL-KV   491
GHU64252.1      466 -----------SFWWQ-----LLIVIGGSLLAGIFFLRKKTQFKIK   495
MBP5455609.1    450 ------GEFSSRLWQNPLFELGLIIIVSILFVLGFYLMKKRK-K   486
MCD7916322.1    477 ---------FLWD-YFLYHLLFLGGMIY-LTYLIFKRK-KW   505
WP_186928207.1  481 ---------NMLWQ-----TLVIVGISA-LLYNNIKDR-RIK   506
WP_077492690.1  485 TRFFNIGDPCPSTWGLLGGDIITLIVFAICLYFPIRHYFKYR-K   527
PID93465.1      489 RNLSSIGINANLTLY----DLVTTILAIWTIIVFMLFIFRNKFKH   529
WP_036792078.1  497 SEF--LGSDLYIQWH--AGDIVTLLVMALCLFVPIRYYLTHK-K   535
```

Figure 3-19 Multiple sequence alignment of CorA and its analogues

Conserved residues present in the interdomain interface are indicated by asterisks, and the positions of the RR and DD motifs probed by site directed mutagenesis are shown.
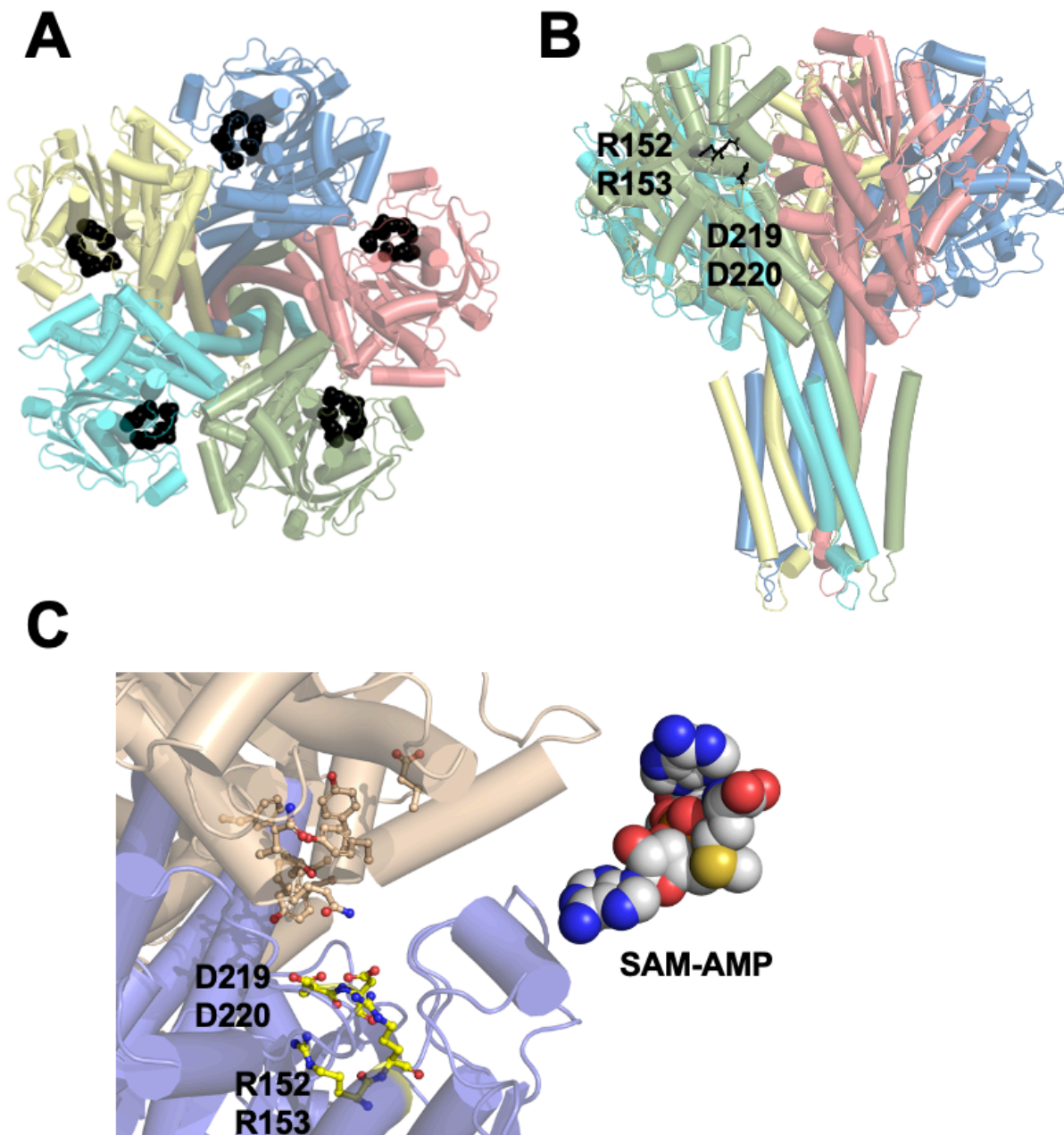
Figure 3-20 Modelling membrane protein BfrCorA
A. Top-down view of the pentameric BfrCorA model with individual subunits coloured differently and the conserved R152/R152/D219/D220 residues indicated by black spheres. **B.** Orthogonal view of the BfrCorA model showing the TM helical bundle at the bottom. **C.** Close up of the inter-subunit interface for CorA, with conserved residues shown and a model of SAM-AMP in sphere representation included for scale.
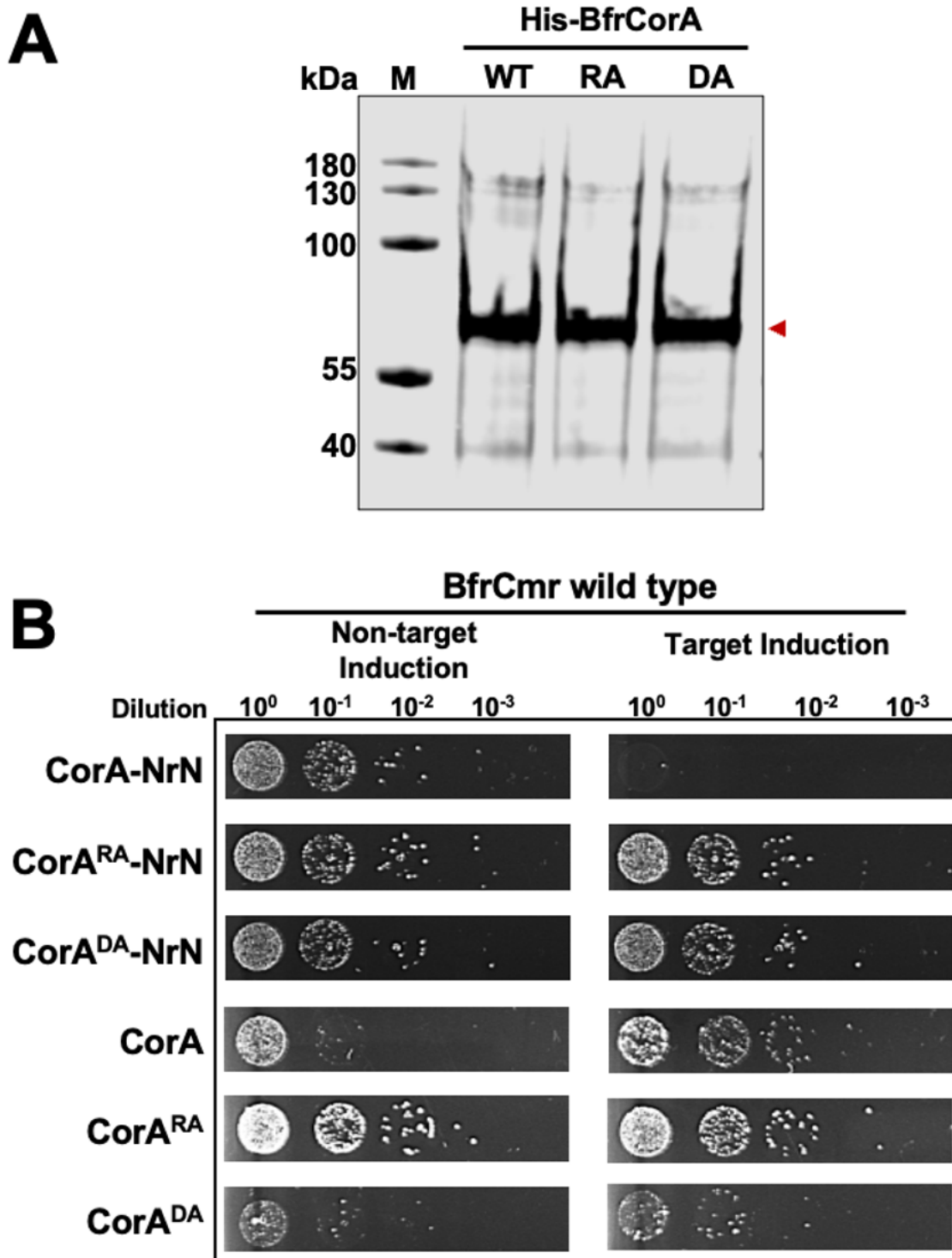
Figure 3-21 Plasmid challenge assay of BfrCorA wild type and its variants in the context of BfrCmr system
A. Western blot using the V5 antibody to detect expression of the wild-type (WT), R152A/R153A (RA) and D219A/D220A (DA) variants in *E. coli*. M is the marker to indicate the size on the gel. **B.** Plasmid challenge assay, showing that wild-type CorA in conjunction with NrN provides immunity from plasmids carrying a target sequence, but neither CorA variants does.

3.2.13 Production of BfrNrN and its variants

The gene of ancillary protein BfrNrN was consistently found adjacent to the gene encoding the membrane effector BfrCorA in most type III CRISPR loci (Shmakov et al., 2018). Additionally, when both BfrCorA and BfrNrN were present in the BfrCmr system, they exhibited plasmid immunity in *E. coli*. This intriguing observation motivated us to investigate the function of BfrNrN.

To gain deeper understanding of BfrNrN, we began by generating the structural model of BfrNrN protein (accession number is WP_005787771.1) using Alphafold2 (Jumper et al., 2021). This predicted structure was then compared with the PDB database of experimentally determined structures to identify structural homologues using FoldSeek (van Kempen et al., 2023). Notably, the two best predicted structural matches for BfrNrN were two bacterial phosphodiesterases (PDE) with well-established roles in signalling regulation: pGpG-specific PDE PggH from *Vibrio cholerae*, which participates in turnover of c-di-GMP (Heo et al., 2022) and GdpP from *Staphylococcus aureus*, specifically degrading c-di-AMP into 5'-pApA (Wang et al., 2018). Subsequently, we performed a sequence alignment of BfrNrN with its structural homologues, well-characterised bacterial DHH/DHHA1 family proteins, including PggH from *V. cholerae* (Heo et al., 2022), phosphatidate phosphatase PAP from *Methanothermococcus thermolithotrophicus* , exonuclease RecJ from *Thermus thermophilus* (Yamagata et al., 2002*)*, and NanoRNase NrnA from *Bacillus subtilis* (Schmier et al., 2017) (Fig. 3-22). The presence of the conserved DHH active site motif (D85:H86:H87) in BfrNrN strongly suggests that it belongs to the DHH/DHHA1-family phosphodiesterases (PDE). These findings indicate that BfrNrN may play a vital role in SAM-AMP signalling regulation within type III CRISPR systems.

We therefore expressed BfrNrN wild type and variant mutated in the DHH active sites (D85A:H86A:H87A) in the *E. coli* C43 (DE3). The purification procedure is detailed in the section 2.1.5. Briefly, BfrNrN wild-type and variant were purified by immobilised metal affinity chromatography (1st IMAC), followed by N-terminal poly-histidine affinity tag removal (Fig. 3-23A). The proteins were then isolated from TEV protease by a second IMAC and further purified by size-exclusion chromatography (Fig. 3-23B). The purity of all proteins was analysed by SDS-PAGE (Fig. 3-23C).
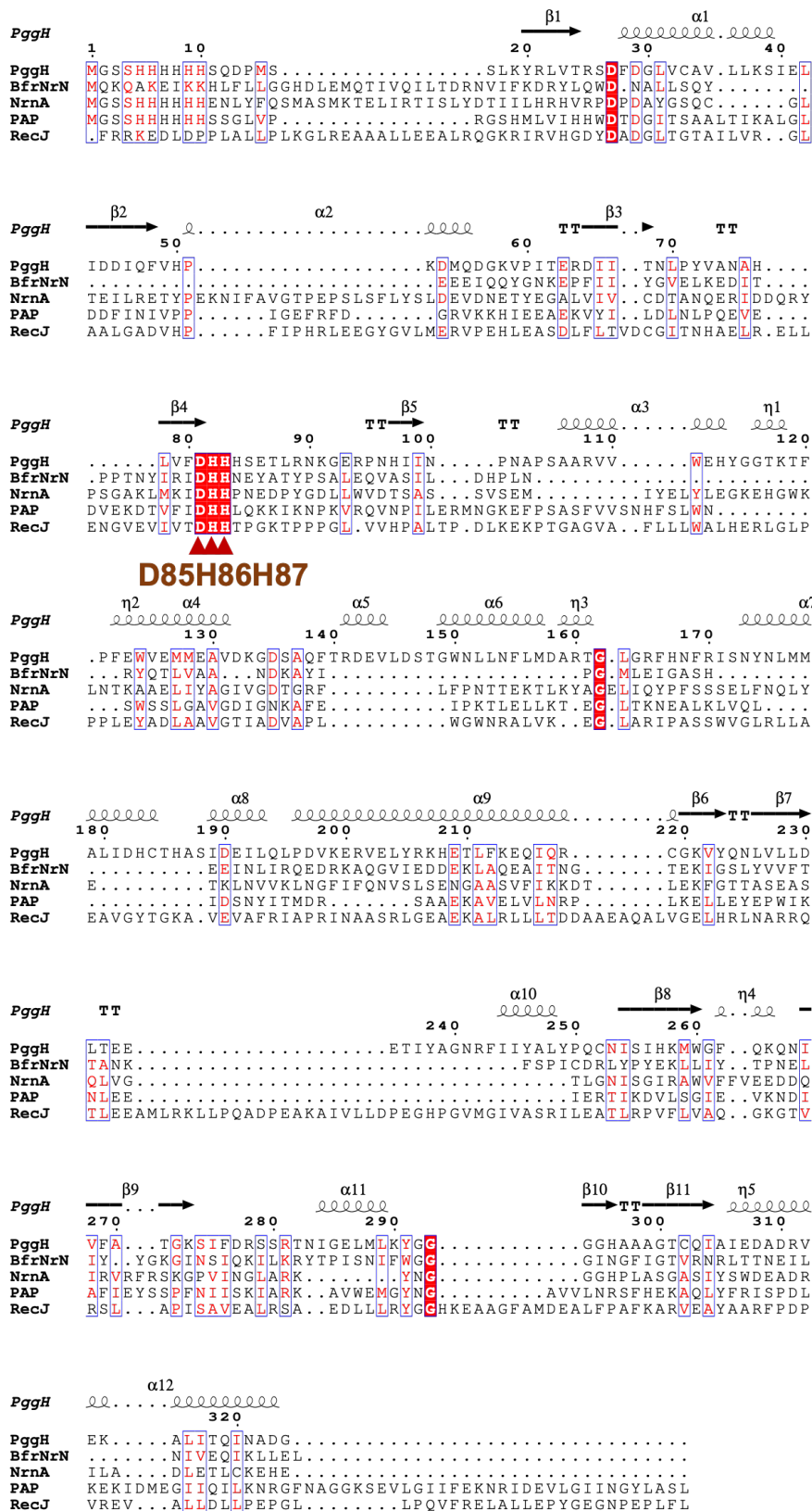
Figure 3-22 Alignment of BfrNrN and its structural homologues
Multi sequence alignment of BfrNrN with PggH from *V. cholerae* (PDB ID: 7D62), PAP from *M. thermolithotrophicus* (PDB ID: 8A8K), RecJ from *T. thermophilus* (PDB ID: 1IR6), and NrnA from *B. subtilis* (PDB ID: 5IUF). Sequences were aligned using MUSCLE of EMBL-EBI (Madeira et al., 2019). ESPript 3.0 (Robert and Gouet, 2014) was used for secondary structure depiction, where the secondary structure of PggH is schematically annotated above the alignment. The conserved active sites of BfrNrN are indicated by red arrows.
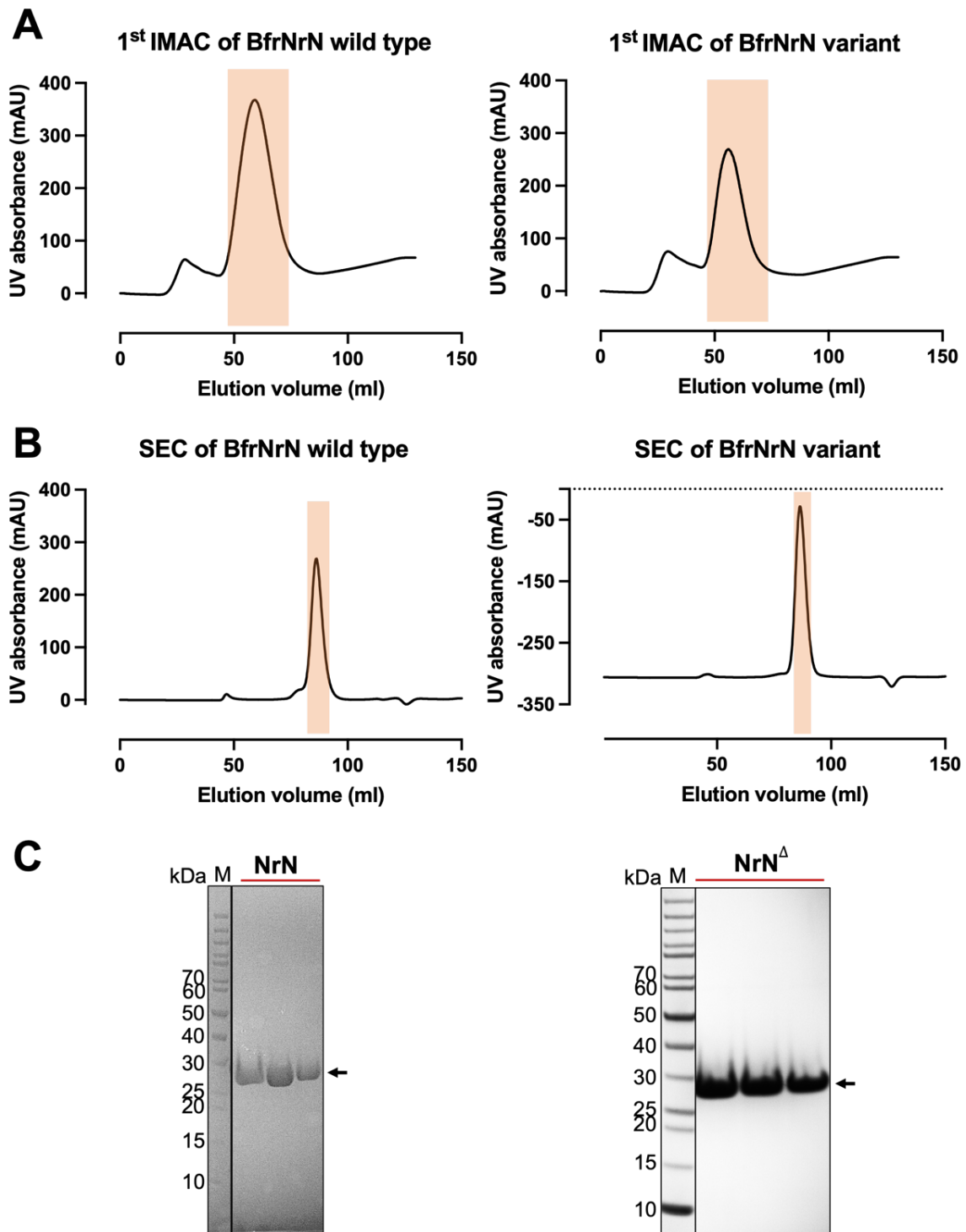
Figure 3-23 The purification of BfrNrN wild type and variant
A. The first immobilised metal affinity chromatography (1st IMAC) trace. The fractions containing target protein highlighted with a red rectangle were eluted with 50% elution buffer and pooled for his tag removal. **B.** Superdex200 SEC profiles. The TEV-cleaved protein was recovered from the nickel column and then subjected to SEC. Fractions indicated by a red rectangle were collected and concentrated for further enzymatic analysis. **C.** SDS-PAGE analysis of purity of BfrNrN wild type and variant. The monomer mass is approximately 30 kDa, consistent with the theoretical mass of BfrNrN. M is the marker to indicate the size on the gel.

3.2.14 SAM-AMP degradation by BfrNrN

Based on the structure and sequence alignment, we assumed that BfrNrN might degrade the signal molecule SAM-AMP. We first tested the function of BfrNrN *in vivo*. As previously observed, SAM-AMP was isolated from the activated BfrCmr system (with target RNA activation). Subsequently, the genes of BfrNrN wild type and its variant were constructed onto the vector pRATDuet, which was co-transformed with pBfrCmr1-6 and pBfrCRISPR_Tet into the *E. coli* BL21star. The nucleotide products were then purified and isolated from overnight induced cell lysates followed by HPLC analysis, as described previously. No production of SAM-AMP was observed when the BfrNrN wild type was present along with the activated BfrCmr system, whereas the presence of the variant NrN$^\Delta$ had no effect on the generation of SAM-AMP (Fig. 3-24A and E). The purified BfrNrN wild type and variant were subsequently incubated with purified SAM-AMP in the presence of Mn^{2+}. SAM-AMP was specifically degraded into AMP and SAM by the BfrNrN wild type, while the cleavage of SAM-AMP was completely abolished in the DHH mutated variant NrN$^\Delta$ (Fig. 3-24B). Additionally, no cleavage was detected when BfrNrN was incubated with cyclic oligoadenylates (cA_{2-6}) or linear dinucleotides (pppApA or pApA) (Fig. 3-24C). Furthermore, the kinetics of degradation of SAM-AMP by BfrNrN were investigated. 100 µM SAM-AMP was completely cleaved by 1.2 µM BfrNrN within the first 2 min of the reaction, consistent with rapid, multiple-turnover catalysis (Fig. 3-24D).

These findings demonstrate that specialised NrN PDEs function to degrade SAM-AMP generated by activated BfrCmr. One potential function is as an "off-switch" to regulate the signalling pathway, similar to the ring nucleases responsible for degrading cyclic oligoadenylates in canonical type III CRISPR system (Athukoralage and White, 2021).
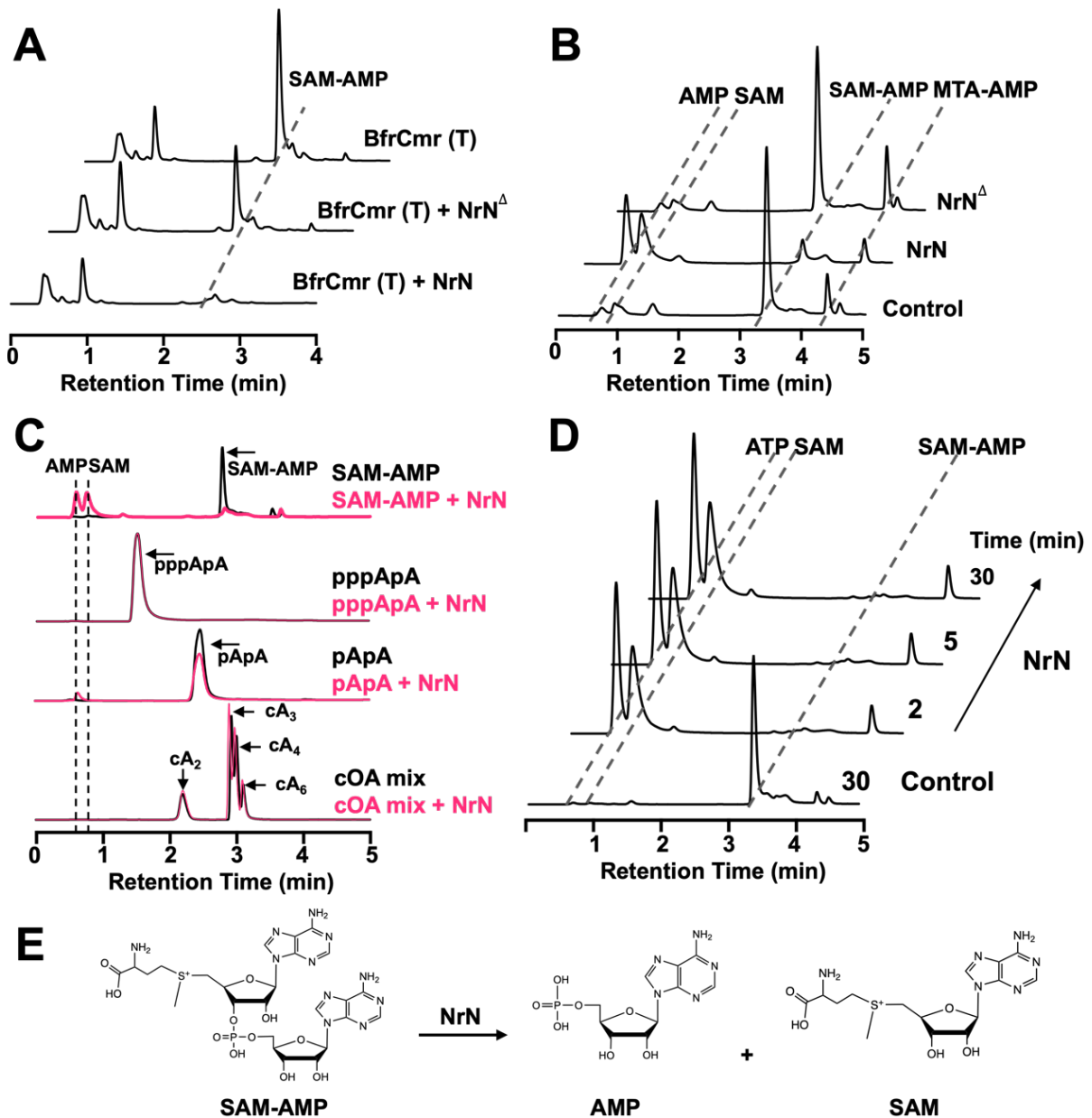
Figure 3-24 SAM-AMP degradation by BfrNrN
A. HPLC analysis of extracted nucleotide products from *E. coli*. The analysed samples extracted from the wild type BfrCmr system with target in the presence of BfrNrN wild type (+ NrN) and an inactive variant (+ NrN$^\Delta$) respectively. The black dashed line indicates the signal molecule SAM-AMP. **B.** NrN specifically degrades SAM-AMP to SAM and AMP *in vitro*. Purified SAM-AMP was incubated with NrN and NrN$^\Delta$, an inactive variant (D85A/H86A/H87A), followed by HPLC analysis. **C.** HPLC analysis of reaction products when BfrNrN was incubated with cA2, cA3, cA4, cA6, pppApA, pApA and SAM-AMP for 30 min. **D.** *In vitro* characterisation of BfrNrN-catalysed reaction, SAM-AMP (0.1 mM) was incubated with wild type of NrN (1.2 μM). Samples were collected at the indicated time points and analysed by HPLC. BfrNrN was omitted in control samples. **E.** Schematic representation of the reactions catalysed by NrN.

3.2.15 Phylogenetic analysis of CorA-associated type III CRISPR systems

The discovery of SAM-AMP as a new class of signal molecules within *B. fragilis* type III CRISPR system aroused our interest to explore the possibility of other type III CRISPR systems that might provide SAM-AMP mediated signalling defence. Thus, the phylogenetic analysis of Cas10 proteins across type III CRIPSR-Cas loci was performed by Dr Ville Hoikkala (University of St Andrews) to investigate the diversity of Cas10 proteins and their associations with CorA.

A phylogenetic tree of Cas10 proteins linked with CorA was constructed through the analysis of 745 type III CRISPR loci from 613 genomes, in association with representative Cas10 sequences. There are three distinct phylogenetic clades of CorA-associated CRISPR systems. The largest cluster (CorA-1) was associated with type III-B system, while the other two clusters (CorA-2 and 3) was linked to the type III-D (Fig. 3-25A). Analysing the genomic context of the corA-associated type III-B CRISPR loci (Fig. 3-25B) showed a consistent pattern wherein the gene encoding CorA was commonly found adjacent to the *nrn* gene in the case of *B. fragilis* and *Methanococcus vanielii* or sometimes even fused together in the genome of *Aliarcobacter butzleri* and related species. This suggests that a functional correlation between them. This relationship can also be substantiated by the requirement of both proteins for plasmid immunity (Fig. 3-1C). The gene encoding NrN is occasionally replaced by another phosphodiesterase (PDE) - a DEDD family nuclease in the genome of *Streptococcus oralis* and *Syntrophothermus lipocalidus*. Analysis of the predicted structure of this protein indicated its resemblance to proteins with small RNA and DNA degradation activities. For instance, RNase T engages in short 3' end trimming (Hsiao et al., 2012), Oligoribonuclease (ORN) possesses small RNA hydrolysis activity (Lee et al., 2019) and the mammalian REXO2 for dinucleotide degradation is required for regulation of transcription (Nicholls et al., 2019).
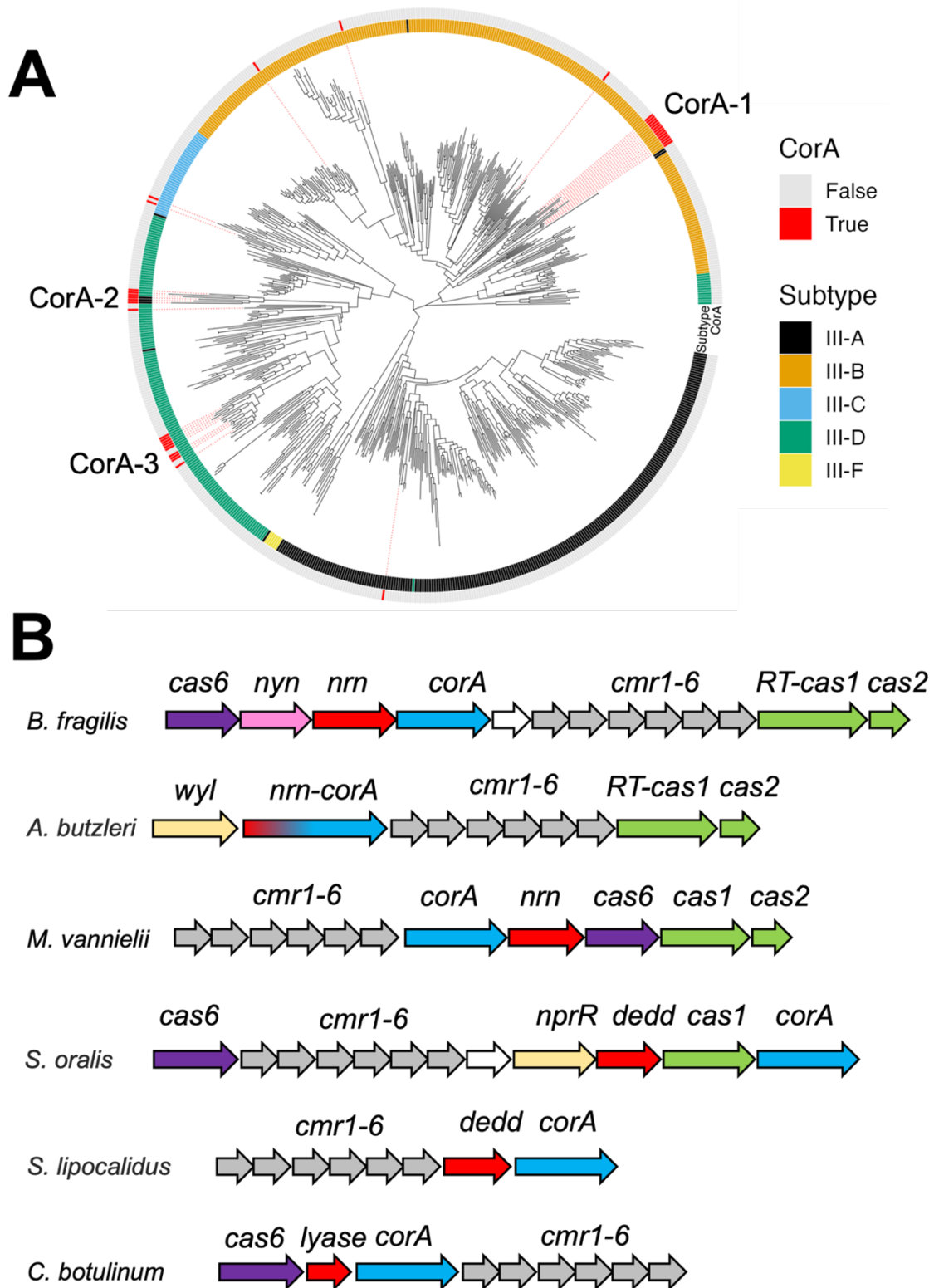
Figure 3-25 Type III CRISPR systems with a CorA effector
A. Phylogenetic tree of Cas10 proteins from type III CRISPR-Cas systems of complete bacterial and archaeal genomes, colour coded by subtype (Russel et al., 2020). Red bars on the outer ring indicate systems associated with a CorA family effector protein. Three main clusters of CorA-associated Cas10s are observed, labelled CorA-1, -2 and -3. **B.** Genome context and effectors of selected type III-B CRISPR systems with a *corA* gene (cluster CorA-1). The type III-B cas genes *cmr1-6* are shown in grey, with *cas6* in purple and the adaptation genes *cas1* (or a gene encoding a fused reverse transcriptase-cas1 protein) and *cas2* in green. The putative membrane channel protein is encoded by the *corA* gene (blue), which is adjacent to or fused with the genes encoding PDEs NrN or DEDD (red). In *C. botulinum*, the PDE is replaced with a predicted SAM lyase. The *wyl* and *nprR* genes encode predicted transcriptional regulators.

3.2.16 SAM-AMP cleavage activity of SAM lyase from *Clostridium botulinum*.

The phylogenetic analysis of CorA-associated type III CRISPR systems led to an intriguingly finding - a type III systems from *Clostridium botulinum*, where the gene encoding PDE NrN is substituted with a gene encoding a protein predicted to resemble a family of phage SAM lyase enzymes. It's worth noting that enzymes from this family are recognised for their role in evading host immune systems by efficiently depleting host SAM pools, thereby inactivating the restriction-modification system (Guo et al., 2021, Simon-Baram et al., 2021) (Fig. 3-26A). This suggests that Type III CRISPR loci encoding a SAM lyase may employ an alternative mechanism for degrading the SAM-AMP signalling molecule.

C. botulinum SAM lyase was thus expressed and purified followed the same procedure as NrN, performed by Dr Shirley Graham (University of St Andrews, Scotland). The enzyme was analysed by SDS-PAGE to confirm its purity (Fig. 3-26B). We then incubated purified SAM lyse with SAM-AMP and SAM, and the reaction samples were subjected to HPLC analysis. We observed an efficient degradation of SAM-AMP into 5'-methylthioadenosine (MTA), while the other degradation product, L-homoserine lactone (HL), was not detectable in the HPLC analysis as it is not UV visible (Fig. 3-26D). Additionally, the SAM lyase degrades SAM-AMP more efficiently than SAM (Fig. 3-26C), suggesting a specialised role in defence.

Figure 3-26 SAM-AMP degradation by SAM lyase
A. *C. botulinum* type III-B CRISPR loci with genes encoding a CorA and a predicted SAM lyase. **B.** SDS-PAGE analysis of purified *C. botulinum* SAM lyase. The monomer mass is approximately 15 kDa, consistent with the theoretical mass of SAM lyase. M is the marker to indicate size on the gel. **C.** *C. botulinum* lyase degrades SAM-AMP to generate methylthioadenosine (MTA) and L-homoserine lactone (not UV visible), but not degrading SAM. **D.** Schematic representation of the reactions catalysed by *C. botulinum* SAM lyase.

3.2.17 Modelling and production of BfrNYN and its variants

The gene encoding ancillary protein BfrNYN was found in *B. fragilis* type III CRISPR loci (Shmakov et al., 2018), adjacent to the genes of membrane effector BfrCorA and SAM-AMP signalling regulator BfrNrN. When BfrNYN was tested in the plasmid challenge assay, no cfu changes were observed, indicating that BfrNYN might not be an effector regulated by SAM-AMP (Fig. 3-1C). Thus, the involvement of BfrNYN in *B. fragilis* type III CRISPR system aroused our interest and led us to investigate its function.

To gain more understanding of BfrNYN, a structural model of BfrNYN was generated using Alphafold (Jumper et al., 2021) and its structural homologues were identified using Foldseek (van Kempen et al., 2023). The closest structural matches for BfrNYN are the N-terminal NYN (Nedd4-BP1/YacP nuclease) domain of MARF1 (meiosis regulator and mRNA stability factor 1), which exhibits ribonuclease activity to control oocyte meiosis and genome integrity in mice (Yao et al., 2018). Another match is Rae1/YacP from *Bacillus subtilis*, which is an endoribonuclease involved in translation-dependent RNA processing (Leroy et al., 2017). The active site of the NYN domain have been shown to have a common set of 4 acidic conserved residues which are essential for degradation activity (Anantharaman and Aravind, 2006). We thus mapped the active sites of BfrNYN by conducting a structural alignment with MARF1, revealing that D13, D72 and D118 in BfrNYN are equivalent to conserved aspartate residues: D178, D246 and D272 in MARF1 (Fig. 3-27A), which are essential for the RNase activity of MARF1 (Yao et al., 2018). These findings implied the BfrNYN might exhibit ribonuclease activity.

To investigate the function of BfrNYN, we first expressed and purified the BfrNYN wild type and two variants (D13A and D72A), following the same purification steps as BfrNrN. Briefly, clear cell lysate was loaded onto a HisTrap FF column and the bound his-tagged target protein was eluted through gradient elution buffer (Fig. 3-28A, D and G). Fractions containing the target proteins were pooled and dialysed overnight with TEV protease to remove the tag. The TEV-cleaved target proteins were then recovered using the HisTrap FF column for the second time and were finally purified by size exclusion chromatography (Fig. 3-28B, E and H). Their identities and purity were confirmed by SDS-PAGE (Fig. 3-28C, F and I).
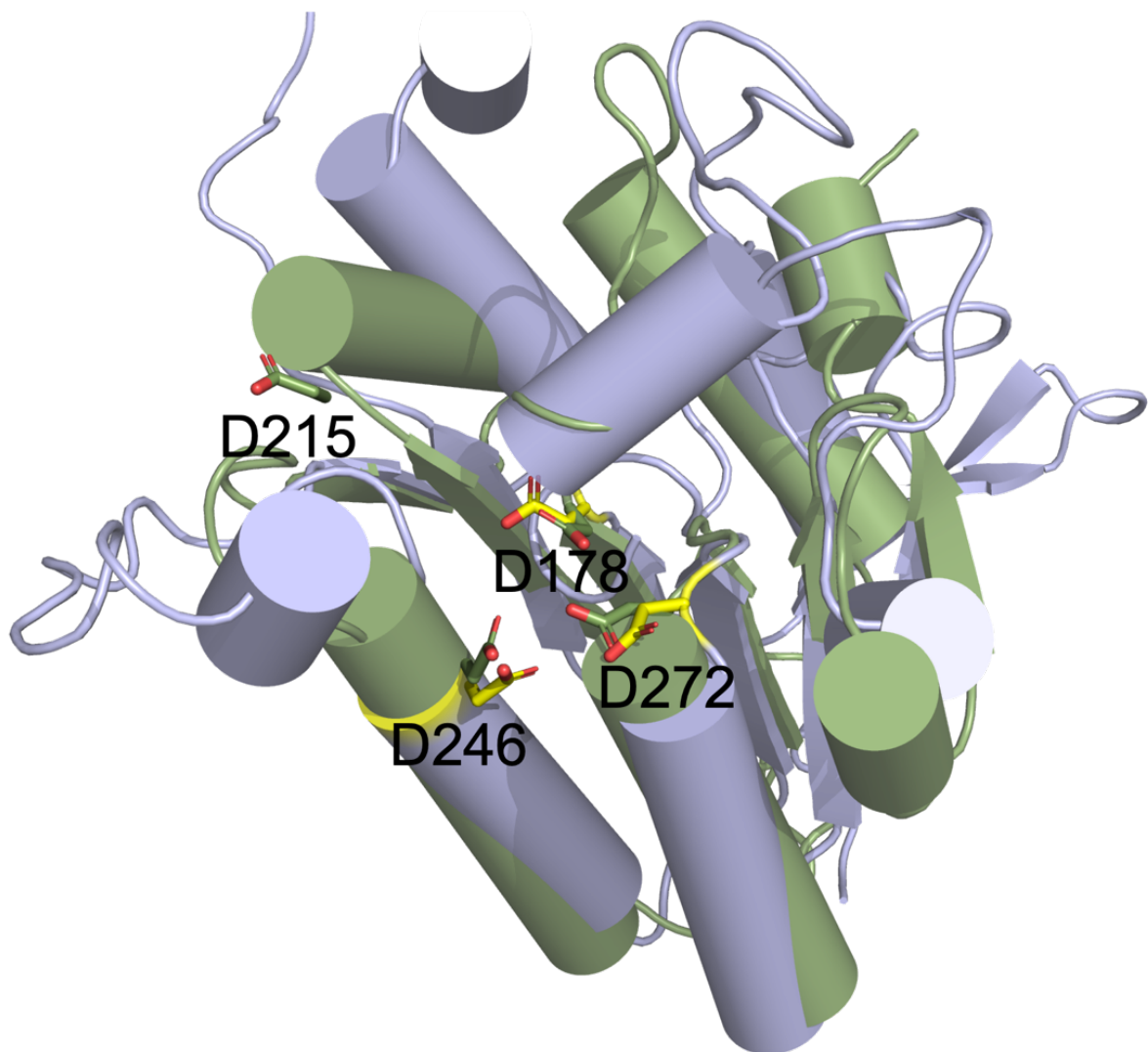
Figure 3-27 Structural alignment of BfrNYN with NYN domain of MARF1
Crystal structure of MARF1 NYN domain from *Mus musculus* (PDB ID: 5YAA) (Yao et al., 2018) is coloured in green and its conserved D178, D215, D256 and D272 residues are shown. The AF2 model (Jumper et al., 2021) of the BfrNYN structure is shown in purple with conserved aspartate residues in yellow. An RMSD (root-mean-square deviation) value between the crystal structure MARF1 NYN and the predicted structure BfrNYN is 3.43 over 144 residues.
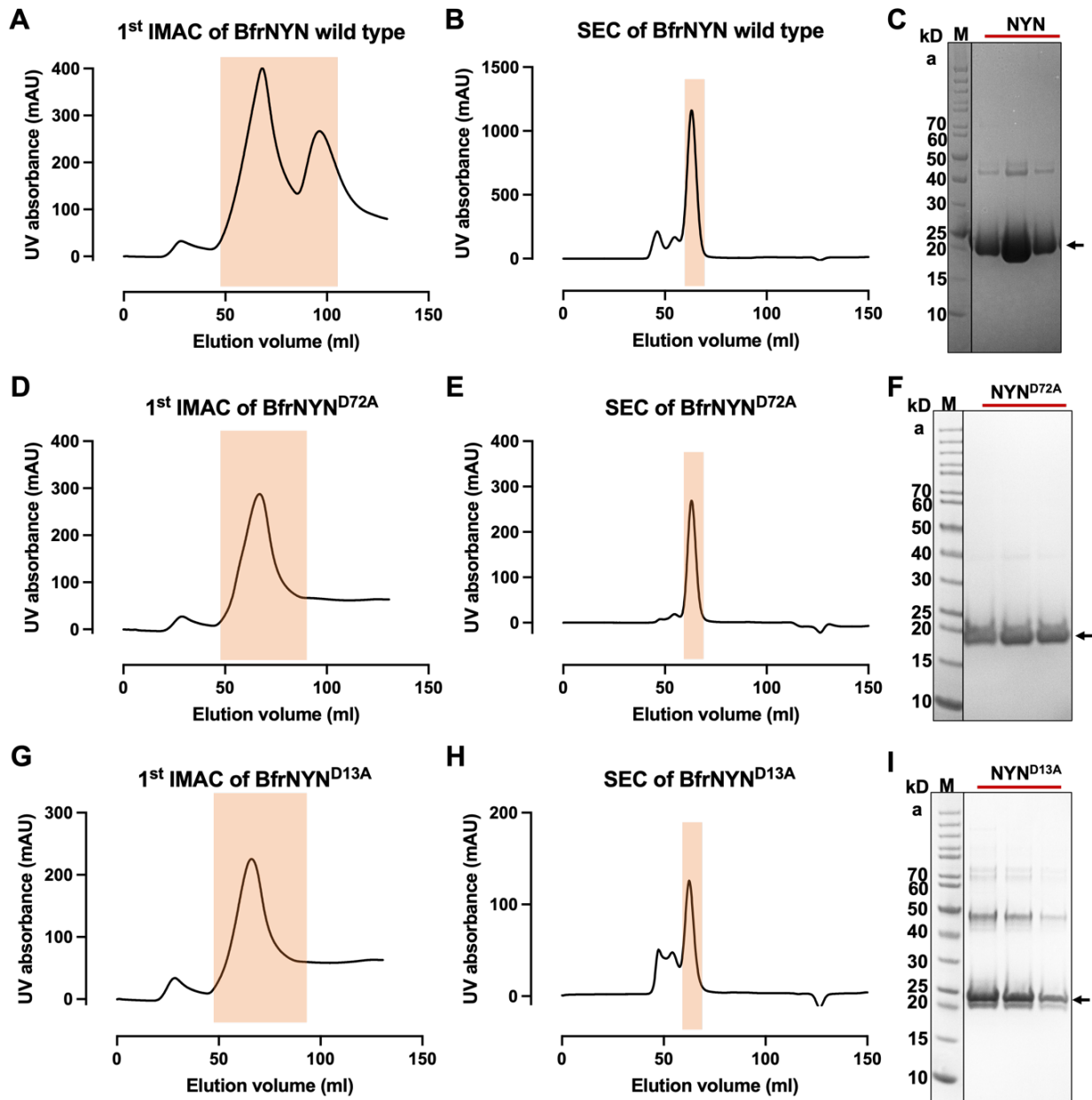
Figure 3-28 The purification of BfrNYN wild type and two variants
A, D and G. The first immobilise metal affinity chromatography (1st IMAC) of BfrNYN wild type and two variants. The fractions containing target protein highlighted with a red rectangle was eluted with 50-100 % elution buffer and pooled for his tag removal. **B, E and H.** Superdex200 SEC profiles. The TEV-cleaved protein was recovered from the nickel column and then subjected to SEC. Fractions indicated by a red rectangle were collected and concentrated for further enzymatic analysis. **C, F and I.** SDS-PAGE analysis of purity of BfrNYN wild type and two variants. The monomer mass is approximately 25 kDa, consistent with the theoretical mass of BfrNYN. M is the marker to indicate size on the gel.

3.2.18 Mn^{2+}-dependent ribonuclease activity of BfrNYN – a potential role in crRNA maturation in *B. fragilis*

We hypothesised that BfrNYN functions as a ribonuclease based on analysis of the structural model of BfrNYN. We then tested the RNase activity of BfrNYN by performing an *in vitro* ribonuclease assay. The purified BfrNYN protein was incubated with 5' FAM end labelled CRISPR repeat (35 nt) in the presence of either Mn^{2+} or Mg^{2+}. Notably, as the concentration of BfrNYN increased from 1 to 5 µM, smaller RNA degradation products accumulated in the presence of Mn^{2+} (Fig. 3-29A). In addition, ribonuclease assay was conducted using another 5'-FAM labelled RNA oligonucleotide "D" (60 nt) in the presence of Mn^{2+} or Mg^{2+}. In this case, RNA D was also cleaved only when Mn^{2+} was present (Fig. 3-29B). Moreover, the predicted conserved D13 and D72 residues in the NYN domain of BfrNYN were mutated to alanine to test its putative catalytic mechanism. The ribonuclease assay of these two variants showed that only residue D13 was essential for the RNA degradation activity of BfrNYN (Fig. 3-29C). These data indicate that the BfrNYN displays Mn^{2+}-dependent ribonuclease activity that is not activated by SAM-AMP. Notably, the active centre of BfrNYN may partially differ from that of other NYN family proteins.

We then set out to test whether SAM-AMP could serve as a substrate of BfrNYN. BfrNYN was thus constructed into the pRATDuet vector, which was co-transformed with pBfrCmr1-6 and pBfrCRISPR_Tet into *E. coli* BL21 star (DE3). The resultant transformant was grown at 37 °C with full induction, followed by extraction and purification of cellular nucleotides. HPLC analysis of isolated nucleotides showed that the presence of BfrNYN in the activated BfrCmr system did not affect the production of SAM-AMP (Fig. 3-29D). Additionally, purified BfrNYN was incubated with SAM-AMP *in vitro* in the presence of Mn^{2+}, and subsequent HPLC analysis revealed that no observable degradation of SAM-AMP in the presence of BfrNYN (Fig. 3-29E). These findings indicate that BfrNYN does not exhibit SAM-AMP degradation activity.

As we showed before, BfrCas6 processed the CRISPR array within CRISPR repeat to generate a processed crRNA of 72 nt in length (Fig. 3-5B), which is longer than the mature crRNAs (37, 43, and 49 nt) extracted from purified BfrCmr complex (Fig. 3-7B). This suggested that there are unknown ribonucleases to assist with crRNA maturation in *E. coli*. We thus hypothesized that BfrNYN might play a role in crRNA maturation in *B. fragilis*. We proceeded to set up a ribonuclease assay to test whether BfrNYN with BfrCas6 together could process CRISPR array into the mature crRNA. Both BfrNYN and BfrCas6 were incubated with radiolabelled CRISPR

array in the presence or absence of Mn^{2+}. Either BfrCas6 or BfrNYN itself was incubated with this CRISPR array as a control. The reactions were stopped following incubation periods of 5, 10, 30 and 60 min by heating at 95 °C and then analysed by denaturing polyacrylamide gel electrophoresis (PAGE). No expected degradation products with the same size as the mature crRNA was observed, while a smear-like degradation pattern of the CRISPR array emerged when both BfrNYN and Mn^{2+} were present (Fig. 3-30). Additionally, the presence of BfrCas6 had no effect on the ribonuclease activity of BfrNYN. These results reveal that BfrNYN functions as a Mn^{2+}-dependent ribonuclease without a specific RNA substrate recognition motif, potentially contributing to trimming processed crRNA intermediates into the mature crRNA in *B. fragilis* CRISPR system. Further analysis would require inclusion of the apo-BfrCmr complex subunits, allowing reconstitution and crRNA trimming *in vitro*. Unfortunately, these are not available.
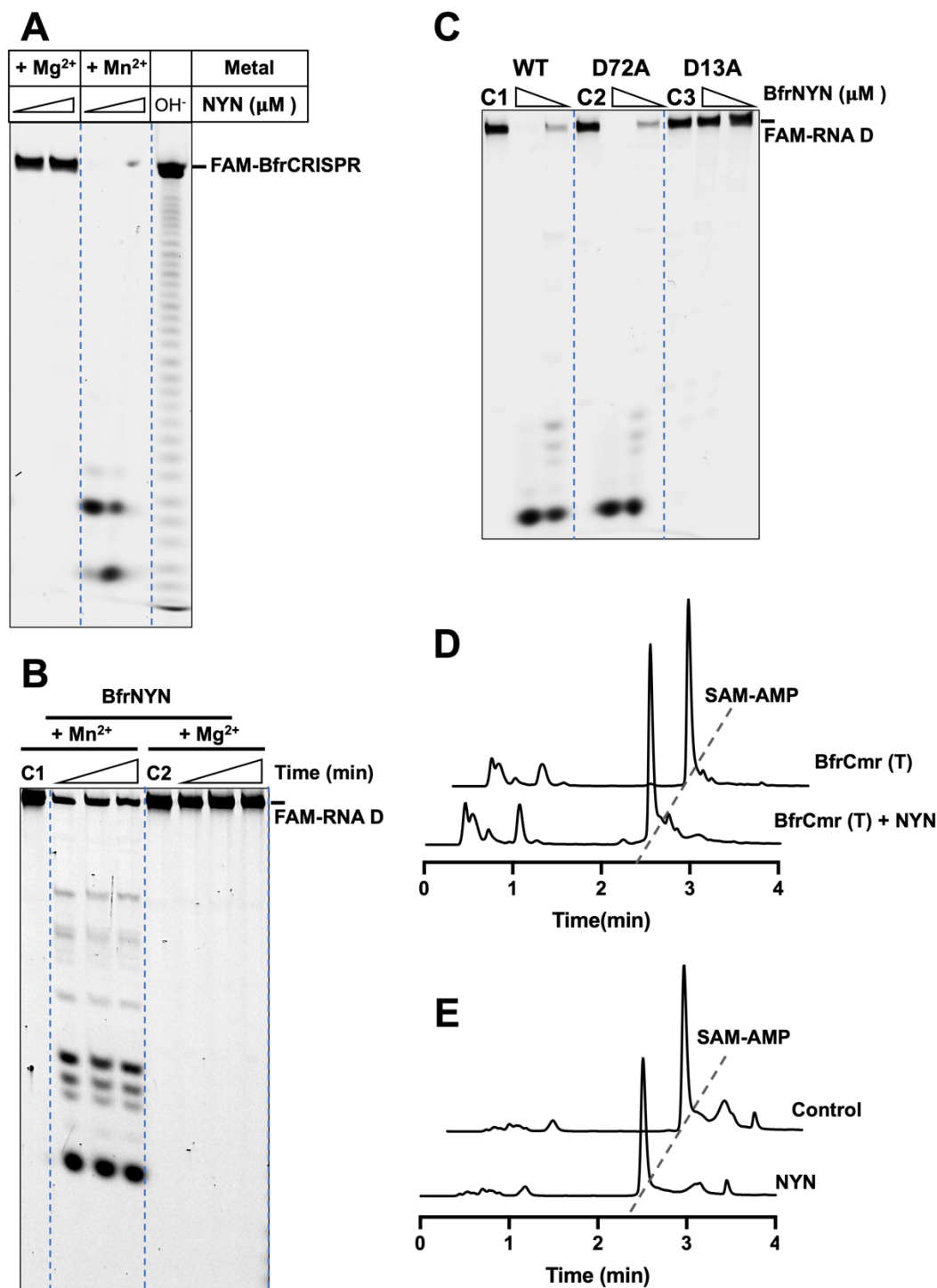
Figure 3-29 Mn^{2+}-dependent ribonuclease activity of BfrNYN
A. Mn^{2+}-dependent ribonuclease activity of BfrNYN. 400 nM 5'-FAM labelled BfrCRISPR repeat RNA was incubated with BfrNYN (1 or 5 µM) in the presence of 5 mM MnCl$_2$ or MgCl$_2$ at 37 °C for 1 h. Lanes labeled OH$^-$ is an alkaline hydrolysis of corresponding RNA under denaturing conditions. **B.** Ribonuclease activity of BfrNYN. 40 nM 5'-FAM labelled RNA D was incubated with 200 nM BfrNYN in the presence of 5 mM MnCl$_2$ or MgCl$_2$ at 37 °C for varying durations of 5, 10 and 15 min. C1 and C2 are control samples in the absence of BfrNYN. **C.** *In vitro* ribonuclease assay of BfrNYN wild type and two variants D13A and D72A. Each of three proteins (0.2 or 1 µM) was incubated with 400 nM 5'-FAM labelled RNA D at 37 °C for 15 min in the presence of MnCl$_2$ (5 mM). No enzymes were added into the control sample. **D.** HPLC analysis of cellular nucleotides extracted from *E. Coli*. Analysed samples extracted from the wild type *B. fragilis* Cmr system with target crRNA in the absence or presence of BfrNYN. The black dash line indicates the retention time of SAM-AMP. **(E)** HPLC analysis of *in vitro* reaction products of BfrNYN. 5 µM BfrNYN was incubated with 100 µM SAM-AMP in the presence of MnCl$_2$ at 37 °C for 1 h. The retention time of SAM-AMP is indicated by a black dash line.
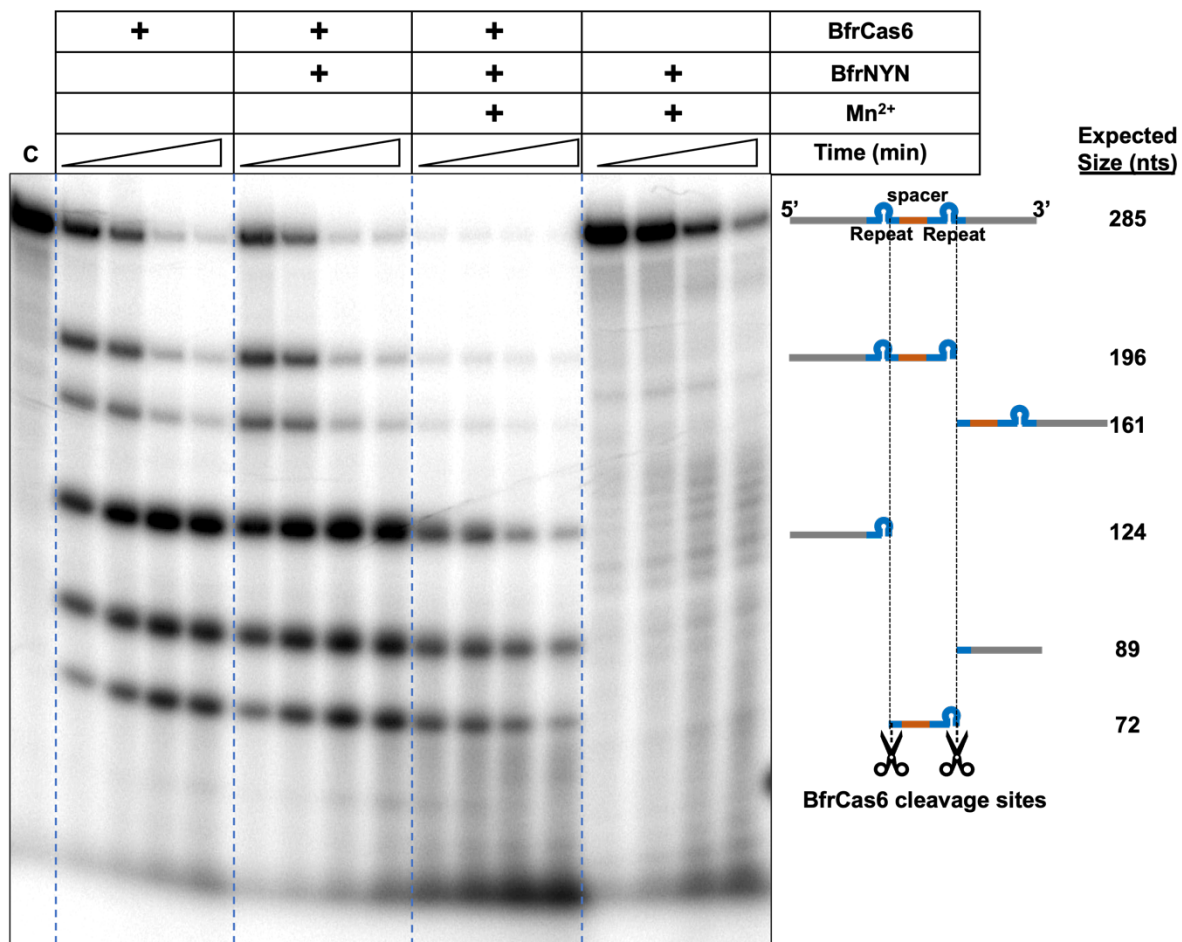
Figure 3-30 The potential role in crRNA maturation of BfrNYN
An internally radio-labelled transcript RNA containing two CRISPR repeats (blue) and one guide (targeting Phage P1) sequence (orange) was incubated with both BfrCas6 (1 μM) and BfrNYN (70 nM) in the absence or presence of $MnCl_2$. Samples were collected at different time points 5, 10, 30 and 60 min and then analysed by denaturing gel. The expected sizes and compositions of cleavage products are indicated based on the specific cleavage site of Cas6 within each repeat (indicated by cartoon of scissors). C is the control samples in the absence of any enzymes.

3.3 Discussion

Type III-A/B (Csm/Cmr) CRISPR systems are well-known for their antiviral response, involving the synthesis of cyclic oligoadenylates (cOA) upon detection of invading RNA (Niewoehner, 2017, Kazlauskiene et al., 2017). The responsibility for cOA synthesis lies with the two conserved Palm polymerase domains of the signature Cas10 subunit. In this study, Cas10 from *B. fragilis* with intact Palm domains showed the capacity to synthesise a novel signal molecule, SAM-AMP (Fig. 3-10D). This unique molecule is formed by conjugating ATP to S-adenosyl methionine (SAM) and is previously unreported either in nature or as a synthetic product. Analysing the sequence of BfrCas10 revealed limited sequence divergence and a conserved GGDD motif in Palm2 (the donor palm pocket) which was found to be crucial for ATP binding (Athukoralage and White, 2022). The comparison of the structural model of BfrCas10 with the crystal structure of Cas10 from *P. furiosus* (Fig. 3-15) suggested that SAM occupies the acceptor palm pocket. This pocket's relatively lower stringency for ATP recognition could have facilitated the evolutionary adaptation to accept SAM (Osawa et al., 2013, Jia et al., 2019a). In consideration of the mechanism for cOA formation, a similar chemical mechanism can be proposed for SAM-AMP synthesis, which the 3'-hydroxyl of SAM attacks the α-phosphate of donor ATP to create 5'-3' phosphodiester bond and release PPi. However, the replacement of triphosphate group of ATP with the methionine moiety of SAM eliminated the possibility of intramolecular nucleophilic attack required for the cyclisation, which is the final step in the cOA formation. Furthermore, the accommodation of SAM in the acceptor palm pocket also eliminates the potential for further polymerisation. In a type III CRISPR system, a new class of signal molecule SAM-AMP has emerged as a linear second messenger, indicating the dynamic evolution of microbial defence systems in response to pressure from viral anti-CRISPRs, perhaps the increased prevalence of viral ring nuclease that degrade cOA. Our phylogenetic analysis of CorA associated type III CRISPR systems revealed three distinct clades of Cas10 widespread in the members of *bacteroidetes*, *firmicutes*, δ and ε-*proteobacteria* and *euryarchaea* (Fig. 3-25A). This implies that the SAM-AMP signalling pathways have a broad distribution, potentially involving other ancillary effectors beyond CorA that could be regulated by SAM-AMP.

A systematic analysis of CRISPR-associated genes present in type III CRISPR-*cas* loci has revealed that genes encoding CorA family proteins, which likely function as a divalent cation channel, are the most abundant uncharacterized effector, and that genes encoding the NrN PDE commonly appear adjacent to, or sometimes even fused with, the gene of CorA (Shmakov et

al., 2018). Our data demonstrated that the type III-B CRISPR system from *B. fragilis* provides anti-MGEs immunity in *E. coli* in the presence of both membrane protein CorA and phosphodiesterase NrN (Fig. 3-1C). While CorA family proteins are recognised as major cation channels for magnesium ion (Mg^{2+}) transport in prokaryotes and eukaryotic mitochondria, CRISPR associated CorA only shares structural similarity in the membrane-spanning domain, which bears the signature motif GxN known to function as the selectivity filter (Pfoh et al., 2012, Guskov et al., 2012, Dalmas et al., 2014, Stetsenko and Guskov, 2020, Lerche et al., 2017, Matthies et al., 2016).

Our data showed that CorA specifically binds the signal molecule SAM-AMP, while not interacting with cA_3 (Fig. 3-18). We hypothesise that CorA could potentially be activated by SAM-AMP, resulting in the opening of the channel. This activation might lead to cell death or dormancy to prevent the spread of phages. However, an alternate possibility is that SAM-AMP mediates membrane disruption by binding to CorA, which had been observed in other membrane linked defence systems (Duncan-Lowey et al., 2021, Georjon and Bernheim, 2023). Thus, further investigations are necessary to elucidate the biochemical and structural mechanism involved here.

Our finding also unveiled that the phosphodiesterase NrN exhibits specific degradation of SAM-AMP (Fig. 3-24). Furthermore, NrN's association with the effector CorA is essential for plasmid immunity and CorA can be toxic in the absence of NrN. However, the underlying reason for this phenomenon remains unknown. One possibility is that the critical degradative function of NrN in the SAM-AMP mediated signalling pathway could help the host avoid unnecessary cell death once evasion has been cleared, which can be supported by the presence of ring nucleases (Crn1-3, Csx3) frequently associated with cOA signalling CRISPR systems (Athukoralage and White, 2021). Another potential explanation is that the degradation of SAM-AMP might be necessary to desensitize the CorA ion channel. This phenomenon has been observed in other ligand-gated ion channels when the concentrations of activator remain high (Velisetty and Chakrapani, 2012). To uncover the underlying cause, further investigation on this system is required in a native host at its natural expression levels, coupled with structure and function studies of the phosphodiesterase NrN and CorA ion channel.

A diverse range of signalling molecules have been discovered recently from prokaryotic defence systems (reviewed in (Georjon and Bernheim, 2023)), including cyclic nucleotides from CBASS system (Whiteley et al., 2019), cUMP and cCMP from PYCSAR system (Tal et al., 2021), and cyclic oligoadenylates generated by CRISPR system (Kazlauskiene et al., 2017, Niewoehner, 2017). The identification of SAM-AMP as a new type of signalling molecule

expands the range of nucleotide-based second messengers. This discovery also opens venues for potential implication in broader immune signalling systems, given that family B polymerases are commonly found in all branches of the tree of life.

Overall, the discovery of the new type of signal molecule indicates the diversity of second messengers and expands our understanding of type III CRISPR-Cas-guided immunity (Fig. 3-31).
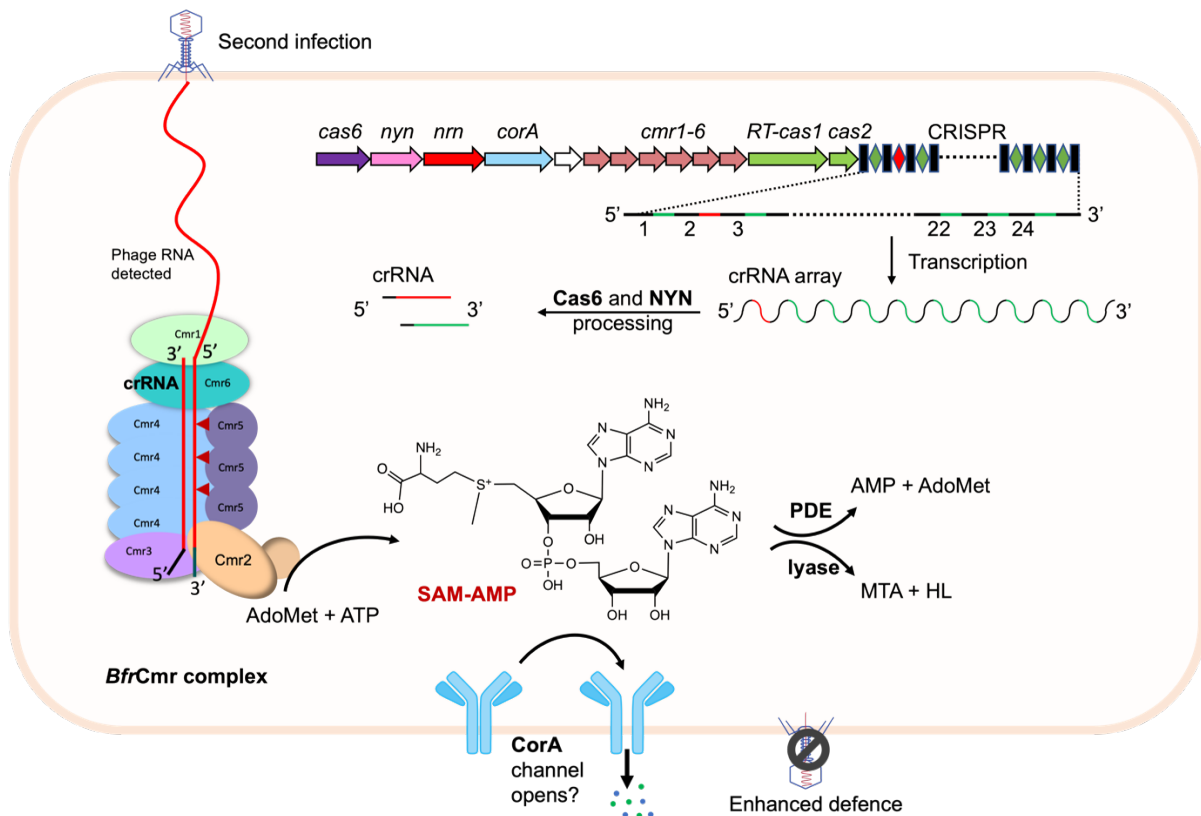
Figure 3-31 Model of the SAM-AMP immune signalling pathway
The CRISPR array was transcribed and processed by Cas6 and NYN into the mature crRNA. Each subunit of the *B. fragilis* Cmr complex assembled around the crRNA. Upon detection of the transcription of the infecting phage genome, Cmr complex becomes active, leading to the generation of the SAM-AMP second messenger. SAM-AMP binds to the CorA membrane protein, resulting in the opening of a pore that disrupts the host membrane to combat infection. SAM-AMP is degraded by specialised PDE enzymes that hydrolyse the phosphodiester bond, generating AMP and AdoMet or lyases that target the methionine moiety, generating MTA and homoserine lactone (HL). These enzymes likely deactivate the signalling molecule to reset the system once phage have been eliminated.

4 Antiviral signalling by a cyclic nucleotide activated CRISPR protease

4.1 Introduction

The signalling pathway involved in the type III CRISPR interference system is one of the most unique features. The enzymatic subunit Cas10 synthesises a variety of cyclic oligoadenylates (cOA) upon detecting invading MGEs (Niewoehner, 2017, Kazlauskiene et al., 2017). These cOAs, in turn, bind to proteins containing a CRISPR-associated Rossmann-fold (CARF) domain, thus allosterically activating linked effector domains, leading to RNA or dsDNA cleavage, supercoiled DNA nicking or transcription modulation (Athukoralage and White, 2021, Lau, 2020, Ye et al., 2020b, McMahon et al., 2020, Zhu et al., 2021, Rostol et al., 2021, Garcia-Doval et al., 2020). These second messenger-regulated effects can lead to cell dormancy or cell death, thereby clearing invading MGEs (Rostol et al., 2021, Athukoralage and White, 2021, Meeske et al., 2019).

Bioinformatic analysis has unveiled the diversity of effectors regulated by signal molecules, including CARF family proteins, membrane proteins like CorA and proteins harbouring a SMODS associated and fused to various effector domain (SAVED) (Burroughs et al., 2015, Shmakov et al., 2018, Shah et al., 2019). The SAVED domain, as a signal sensor domain, has been found to fused to a diverse range of effector domains in both type III CRISPR and CBASS systems. TIR-SAVED effectors from the type II CBASS system have been found to confer immunity through NAD^+ degradation following cA_3 activation and filamentation (Hogrel et al., 2022).

In this chapter, we focus on a type III B CRISPR system involving CalpL which contains a SAVED signal sensor domain and a Lon protease domain. MazF homologue CalpT and extracytoplasmic Sigma factor homologue CalpS are encoded by adjacent genes in the same operon of the thermophilic bacterium *Sulfurihydrogenibium* spp. YO3AOP1. However, their function in CRISPR defence was unknown. With our collaborators, we demonstrated that CalpL forms a stable ternary complex with CalpT and CalpS. Upon activation by cyclic tetra-adenylate cA_4, CalpL oligomerises and specifically cleaves CalpT, resulting in the release of the sigma factor CalpS-CalpT$_{23}$ from the complex. It is predicted that after the degradation of cleaved CalpT$_{23}$, CalpS could be completely released to interact with RNA polymerase, enabling adaption to phage attack.

In this study, our collaborators Christophe Rouillon, Niels Schneberger and Gregor Hagelueken investigated the structure and function of the SAVED-containing protein CalpL. Despite predictions suggesting it to be a transmembrane protein, CalpL was expressed and purified from *E. coli* as a soluble monomer. CalpL was subsequently crystalised, and its structure (PDB ID: 7QDA) was solved to 2.1 Å and refined to a final R and R_{free} values of 19.3 and 22.5, respectively (Zwart et al., 2008, Chen et al., 2010, Rouillon et al., 2023). The structure revealed that the Lon protease domain, with a hallmark catalytic Ser-Lys dyad, lies at the end of a narrow channel that presumably binds substrate peptide. The SAVED domain exhibits an extensive, positively charged cavity on its surface, suitable for cOA ligand binding. Subsequent surface plasmon resonance (SPR) assays revealed that CalpL selectively bound cA_4 with a dissociation constant (K_d) of approximately 1 nM. In addition, a 2.2 Å crystal structure of CalpL in complex with cA_4 (PDB ID: 8B0R) confirmed that the ligand binds to the SAVED domain (Fig. 4-1A).

To explore its protease activity, CalpL was incubated with CalpT and cA_4, resulting in the specific cleavage of CalpT into two distinct products with molecular weights of 23 and 10 kDa, respectively (Fig. 4-1C). They also identified the protease active site by repeating the cleavage assay with an S152A variant of CalpL. Additionally, peptide sequencing of the two cleavage products and mutagenesis of predicted cleavage sites showed that A195 of CalpT is highly possible to be the P1 residue. Furthermore, they revealed that CalpL and CalpT form a stable complex at a 1:1 ratio using multi-angle light scattering coupled with SEC (SEC-MALS) (Fig. 4-1C). cA_4 induced cleavage of CalpT results in observation of two peaks containing a CalpL-$CalpT_{10}$ complex and a $CalpT_{23}$, respectively, during the SEC-MALS analysis (Fig. 4-1C). A 3.3 Å crystal structure of the CalpL-$CalpT_{10}$ complex indicated that $CalpT_{10}$ binds to the N-terminal domain of CalpL with the interface formed by the residues W28, L6, V14, L18, E20, E13, K8 and H2 of CalpL and the residues K200, Y210, Y203 and E222 of CalpT (Fig. 4-1C). The crystal structure of the CalpL-$CalpT_{10}$ complex also showed that CalpL cleavage site is more than 35 Å away from protease active site, indicating the occurrence of cA4 induced structural rearrangement of CalpL to allow cleavage of CalpT. Dynamic light scattering (DLS) and SAXS experiments confirmed that CalpL oligomerises in a cA_4 induced and protein-concentration dependent manner (Fig. 4-1B).

My main contribution to this study was the investigation of the function of CalpT and CalpS. The following pages of this chapter provide detailed formation about the formation of a stable ternary complex among CalpL, CalpT and CalpS, as well as the release of the ECF sigma factor

CalpS from complex after cA4-induced cleavage, which is presumably involved in transcriptional regulation to provide immunity against MGEs.
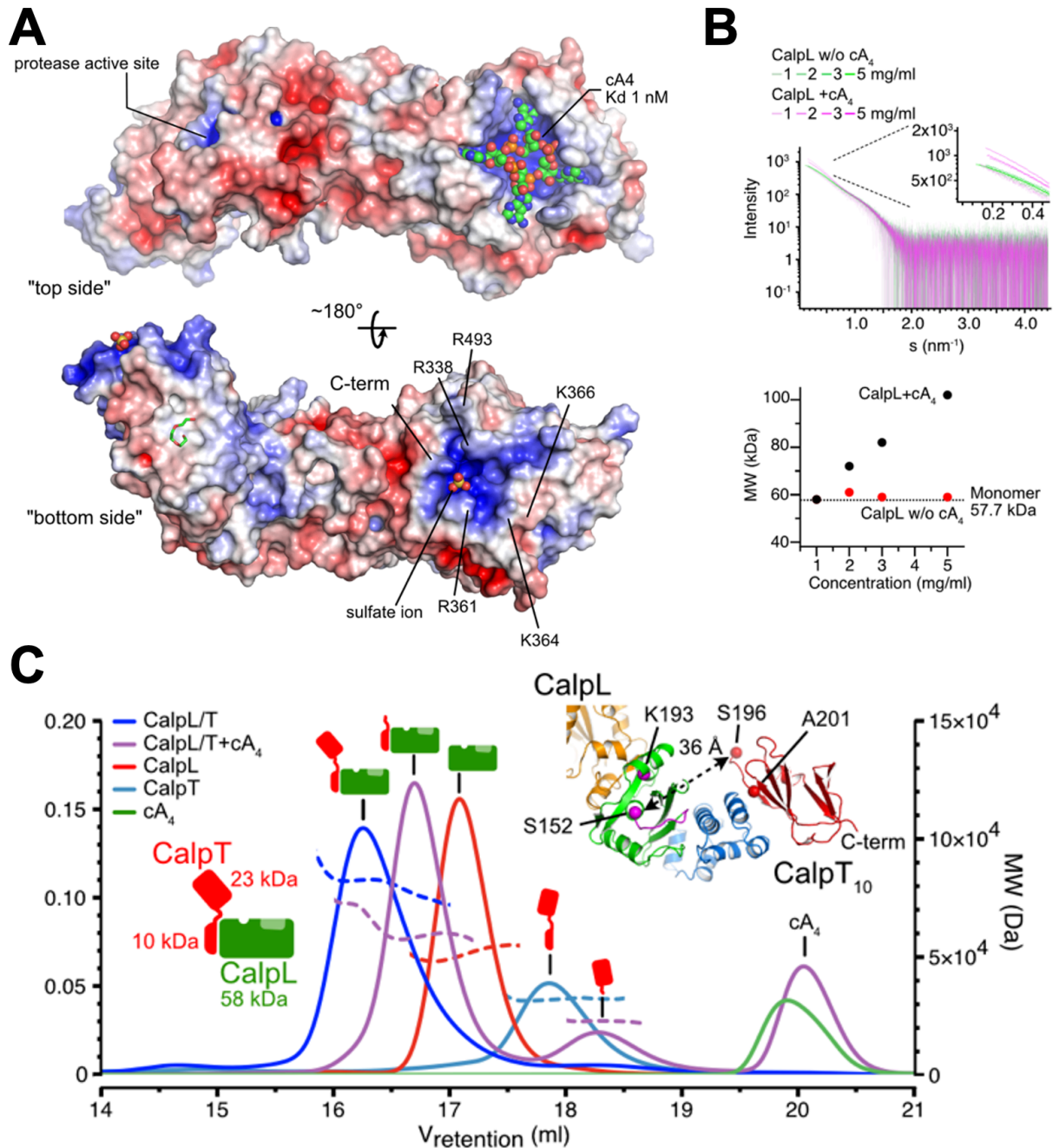
Figure 4-1 The structure and mechanism of CalpL
A. Crystal structure of CalpT in the complex with cA4. A surface model of CalpL is shown in both top and bottom views with the electrostatic potential, which blue and red represent positive and negative, respectively. The protease active site and the bound cA4 molecule are highlighted. Molecule cA4 and sulfate ions are shown as sphere and a polyethylene glycol molecule is shown as sticks. **B.** The small angle X-ray scattering (SAXS) data were gathered at four different concentrations of CalpL in either the presence or absence of cA4 (Top figure). Subsequently, molecular weights from forward scattering (I_0) values, were plotted against the concentrations (Bottom). Notably, in the presence of cA4, the molecular weight of the protein apparently increases as the concentration rises. **C.** SEC-MALS traces (solid lines: UV_{280}, dashed lines: MW_{MALS}) of proteolysis reactions involving various combinations of CalpL wild type (wt), CalpT wild type (wt), and cA4. The schematic represents the molecular species behind the individual peaks. The crystal structure of the CalpL-T_{10} complex is showed inset, with highlighted indication of the 35 Å distance between the P-1 position (S196) and the protease active site.

4.2 Results

4.2.1 Modelling of CalpT

Our collaborators demonstrated that CalpT forms a stable 1:1 complex with CalpL and, upon the introduction of the signalling molecule cA_4, is cleaved by CalpL into two distinct products with molecular weights of 23 and 10 kDa, designated as $CalpT_{23}$ and $CalpT_{10}$, respectively. Both sequence and structural alignments of CalpT conducted using HHpred (Zimmermann et al., 2018), AlphaFold2 (Jumper et al., 2021) and DALIsever (Guo et al., 2021) indicated homology with the MazF toxin in the N-terminal region ($CalpT_{23}$) (Fig. 4-2). Conversely, the C-terminal half ($CalpT_{10}$) shared a weak similarity to DUF2080, a domain of unknown function containing an immunoglobulin fold (Fig. 4-2). Intriguingly, the C-terminal fold resembles the ribbon-helix-helix (RHH) motif found in MazE from *Bacillus subtilis*, which acts as an anti-toxin, forming a complex with MazF to regulate the mRNA interferase activity of MazF (Simanshu et al., 2013). Thus, we hypothesise that CalpT may function in a toxin and anti-toxin manner, in which MazF-like toxin half, $CalpT_{23}$, is expected to exhibit ribonuclease activity once released from the anti-toxin half, $CalpT_{10}$.

Experiments conducted by Niels Schneberger have identified A195 as the most likely cleavage residue by CalpL (Fig. 4-2). Additionally, given that the structural model suggests two fragments $CalpT_{23}$ and $CalpT_{10}$ connected by a flexible linker, we thus created truncated version of CalpT ($CalpT^{tr}$) including amino acids 1 to 173 to retain the complete MazF-like fragment.
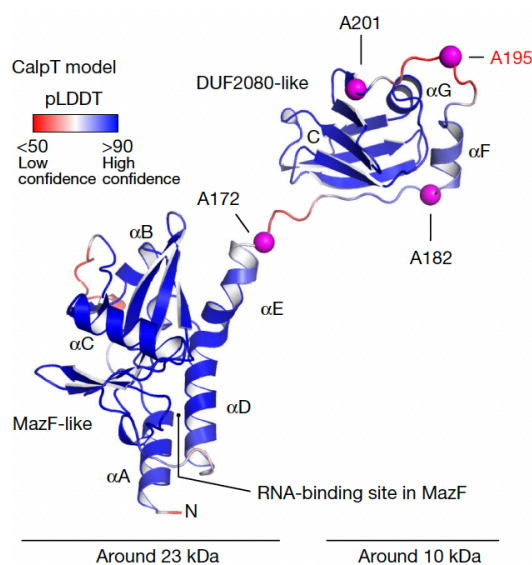
Figure 4-2 Structural prediction of CalpT
Structural prediction is conducted by using AlphaFold2. Predicted protein structure is shown as a cartoon and prediction confidence is indicated by color (predicted local distance difference test (pLDDT)). The P1 residue A195 is highlighted in red and other predicted cleavage residues are marked out.

4.2.2 Purification of truncated CalpT

To investigate the biochemical activity of the MazF-like toxin in the N-terminal region of CalpT, we designed a synthetic gene encoding truncated CalpT (N-terminal fragment, aa 1-173) and cloned it into the *E. coli* expression vector pEHisV5TEV (Fig. 2-4B). Subsequently, we conducted purification of the truncated CalpT through immobilised metal affinity chromatography (IMAC). The target protein was eluted using imidazole at a concentration of around 0.25 M (Fig. 4-3A). The fractions containing the protein were collected and then subjected to dialysis with TEV protease at room temperature overnight, followed by a second round of IMAC to recover his tag-removed protein. The further purification of the proteins was accomplished using SEC column, followed by a Heparin column (Fig. 4-3B and C). The identity and purity were confirmed through SDS-PAGE analysis (Fig. 4-3D).
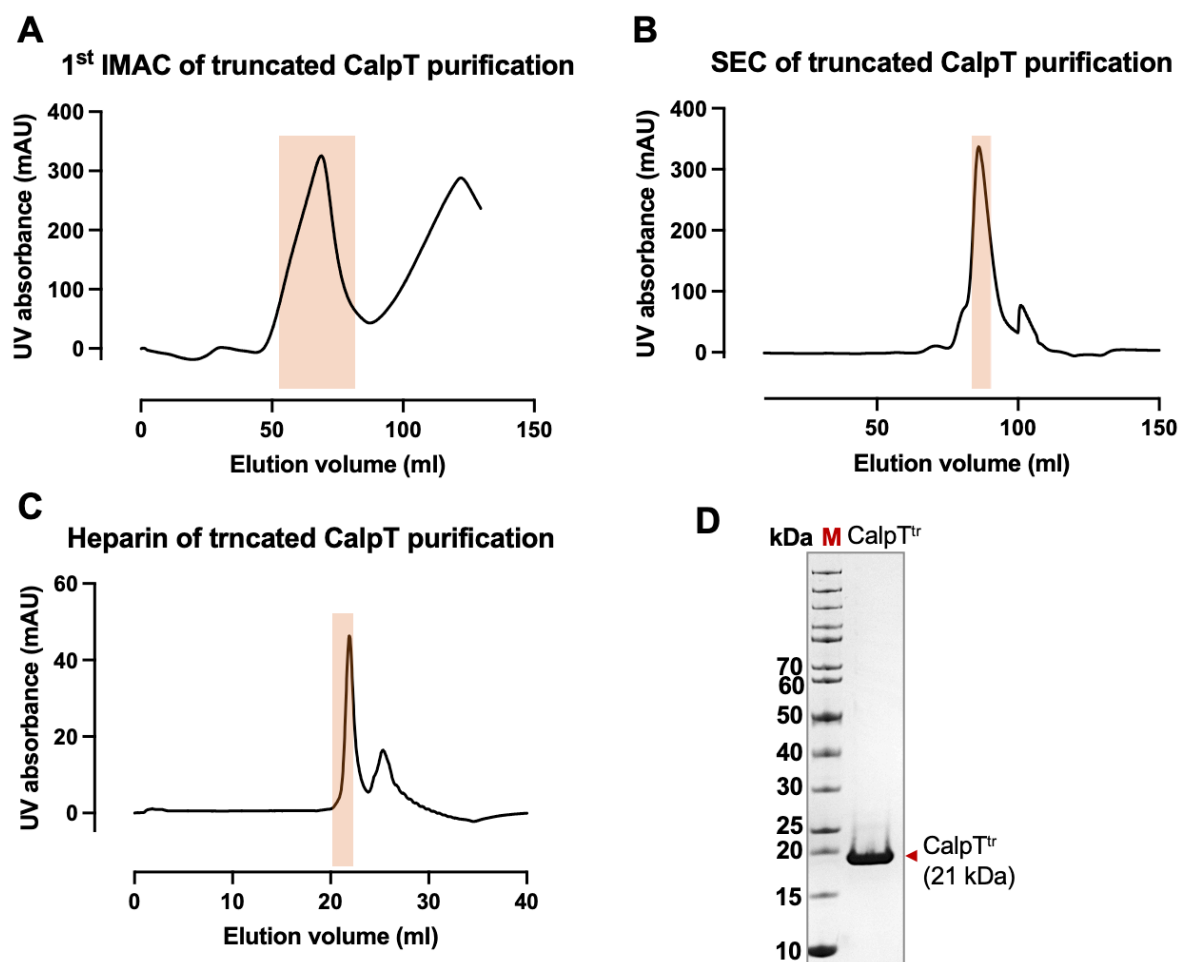
Figure 4-3 Purification of part C-terminus truncated CalpT (aa 1-173)
A. First immobilised metal affinity chromatography (1st IMAC) for truncated CalpT purification. The fractions containing target protein highlighted by a red rectangle was evaluated by SDS-PAGE and pooled for his tag removal. **B.** Superdex200 SEC profile for truncated CalpT. The TEV-cleaved target proteins were recovered from the nickel column and then subjected to SEC. Fractions indicated by a red rectangle were collected and concentrated for further purification. **C.** Histrap Heparin profile. Fractions indicated by a red rectangle were collected for the further enzymatic analysis. **D.** SDS-PAGE analysis of purity of truncated CalpT. The monomer mass is approximately 21 kDa, consistent with the theoretical mass of CalpT. M is the marker to indicate size on the gel.

4.2.3 Ribonuclease Activity of truncated CalpT and cleaved CalpT$_{23}$

Modelling indicated that the structural homologue of N-terminal region (CalpT$_{23}$) is the MazF toxin and the C-terminal half (CalpT$_{10}$) is predicted as a potential anti-toxin, presumably regulating the ribonuclease activity of CalpT$_{23}$ (Fig. 4-2).

We set out to test whether CalpT$_{23}$ was a ribonuclease, the purified CalpT^{tr} was thus incubated with five, 5'- end FAM labelled RNA substrates (B-F) of varying lengths from 17 to 60 nt in the presence or absence of metals (either Mn^{2+} or Mg^{2+}) at 60 °C for 1 h (Fig. 4-4A, RNA sequence shown in the table 2-4). No obvious signs of ribonuclease activity were observed, despite the presence of some random degradation which was more likely introduced from protein contamination. We then assessed the optimal conditions for CalpT^{tr} ribonuclease activity with the substrate RNA D, exploring a pH range from 6 to 9, a NaCl concentration from 10 to 250 mM, and the use of MES, HEPES or CAPS buffers (Fig. 4-4B). We didn't detect any cleavage activity under any of the tested condition.

One possibility was that our truncated construct no longer retained ribonuclease activity. Therefore, in addition, we conducted experiments to test the ribonuclease activity of CalpT after cleavage by activated CalpL (CalpT$_{23}$) by incubating six different FAM-labelled RNA substrates with CalpT and CalpL complex in the presence or absence of cA$_4$ (Fig. 4-5A and B). No ribonuclease activity was identified in the conditions where CalpT was efficiently cleaved by CalpL upon activation by cA$_4$. Furthermore, our collaborators Katja Blumenstock and Jonathan L, Schmind-Burgk performed RNase screening experiments by incubating the CalpL/T complex with random ssRNA libraries in the presence or absence of cA$_4$. No signs of ribonuclease activity were observed. Considering the tolerance of the expression of truncated CalpT in *E. coli* cells, our data suggests that CalpT may not function as a MazF-like ribonuclease.
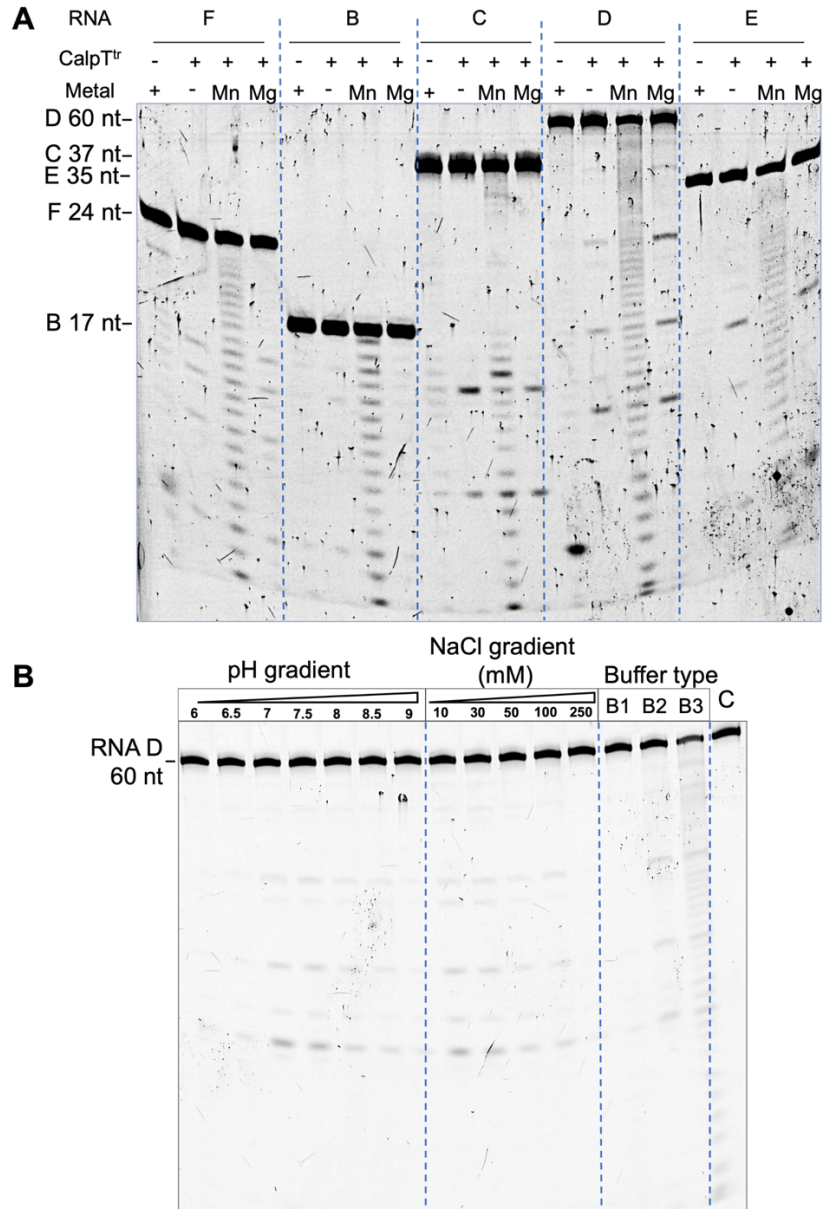
Figure 4-4 Investigation of ribonuclease activity of truncated CalpT
A. Fluorescence image of the denaturing polyacrylamide gel electrophoresis to determine ribonuclease activity of the reactions by incubating CalpT[tr] with five fluorescent-labelled RNA substrates (RNAs listed in Table 2-4) in the presence of different metals. Some cleavage reactions were observed after 60 min incubation at 60 °C, but these may come from the protein contamination. **B.** The optimal reaction condition screen. CalpT[tr] was incubated with fluorescent-labelled RNA substrate D (60 nt), in the buffer of 20 mM Tris-HCl, 50 mM NaCl with a pH range from 6 to 9, in the buffer of 20 mM Tris-HCl, pH8.0 with a NaCl range from 10 to 250 mM, and in three different buffers: 20 mM MES, pH 6.0, 50 mM NaCl (B1), 20 mM HEPES-KOH, pH 7.5, 50 mM KCl (B2) and 50 mM CAPS, pH 9.4, 50 mM KCl (B3). Reaction mixture was incubating for 60 min at 60 °C. Control reaction only contains the RNA substrate.
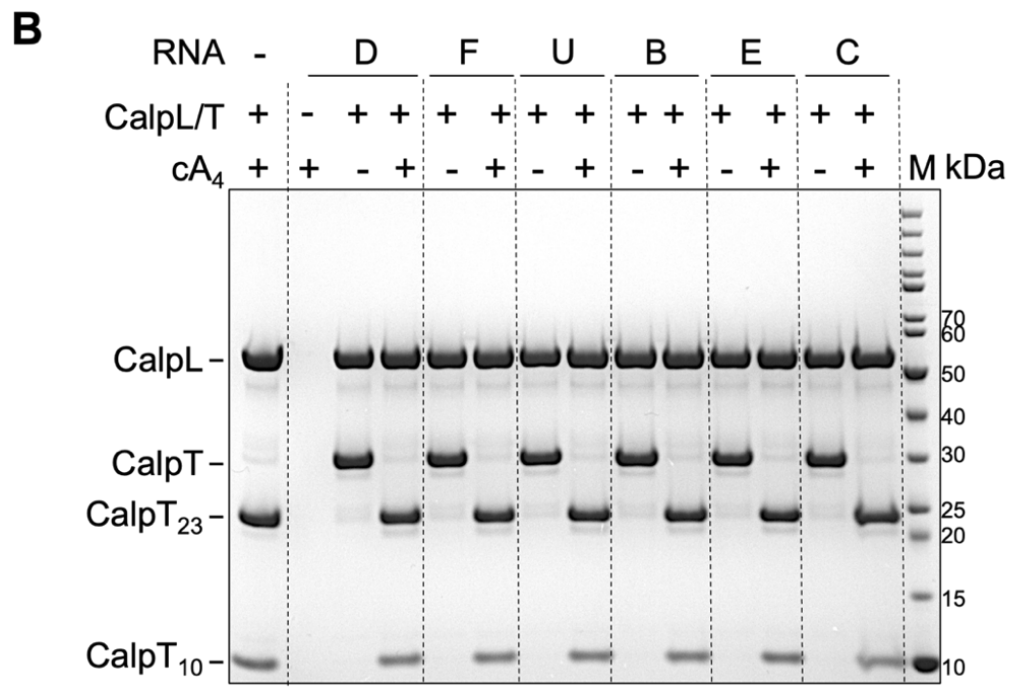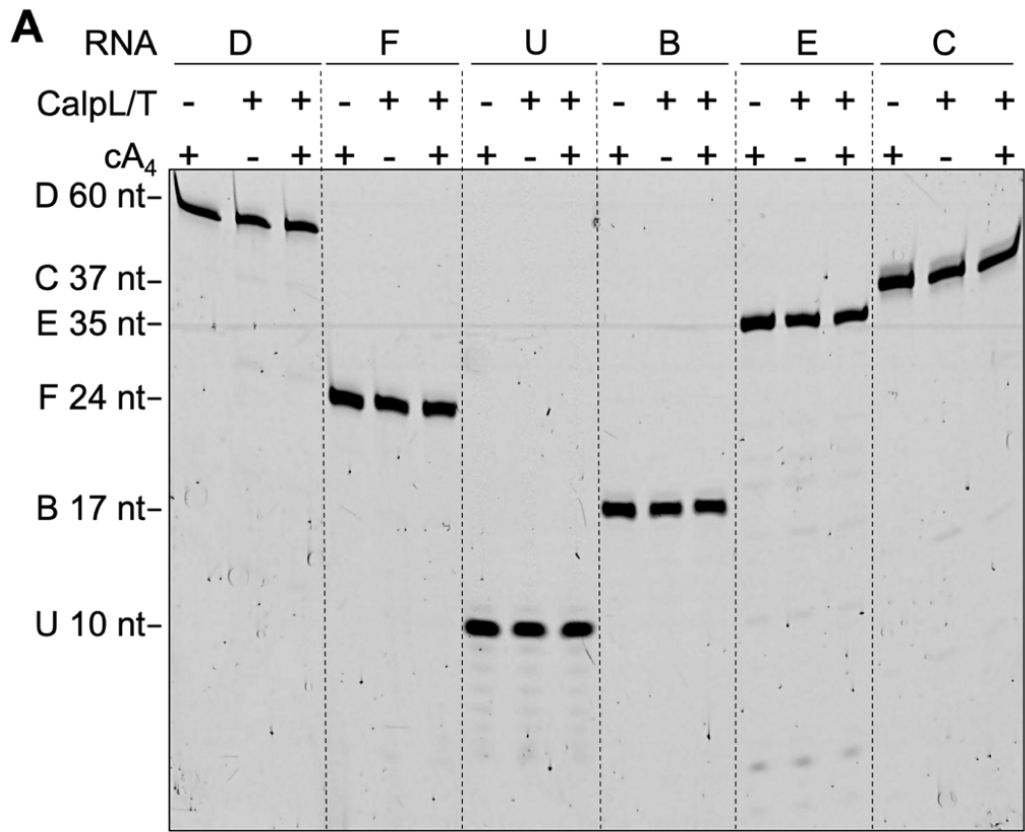
Figure 4-5 Probing the ribonuclease activity of the activated toxin CalT23
A. Fluorescence image of the denaturing polyacrylamide gel electrophoresis to determine ribonuclease activity of CalpT23. CalpT (5.5 μM) and CalpL (5.5 μM) were incubated with six 40 nM FAM-labelled RNA substrates (RNAs listed in Table 2-4) in the presence or absence of 10 μM cA4. Some cleavage reactions were observed after 60 min incubation at 60 °C, but these were not dependent on the presence of cA4 activator. **B.** SDS-PAGE analysis of cA4-induced cleavage of CalpT (33 kDa) by CalpL for each condition in part A. Cleavage was complete after 60 min at 60 °C, confirming the lack of RNase activity of CalpT23.

4.2.4 Purification of the ECF sigma factor CalpS

Despite conducting both biochemical and RNase screening assays against an RNA library, we were not able to confirm the presence of MazF-like nuclease activity in either the truncated or cleaved CalpT. These data strongly suggested that CalpT may exhibit an alternative activity. Furthermore, we observed the presence of a third conserved gene next to the *calpT* gene (Fig. 4-6A). While the precise role of this gene within the CRISPR system was unconfirmed, HHpred analysis revealed that the protein was homologous to extracytoplasmic function (ECF) family σ factors, which play a vital role in promoter recognition and transcription initiation (Zimmermann et al., 2018, Sineva et al., 2017, Paget, 2015). Thus, we designated this third conserved protein as CalpS. Additionally, the activity of ECF σ factor were frequently found to be negatively regulated by an anti-sigma factor (Paget, 2015, Sineva et al., 2017). We thus hypothesized that CalpS and CalpT may function in a manner analogous to a sigma factor and its corresponding anti-sigma factor within the CRISPR system.

To explore the potential relationship between CalpS and CalpT, we cloned the codon-optimised gene of CalpS into a *E. coli* expression vector pEHisV5TEV (Rouillon et al., 2019) (Fig. 2-7A, protein sequence listed in Appendix A). The his-tagged CalpS was expressed and purified with a first step of immobilised metal affinity chromatography (Fig. 4-6C). Fractions containing CalpS were collected and incubated with TEV overnight to remove the His tag at room temperature, followed by second IMAC to recover TEV-cleaved CalpS. Size-exclusion chromatography (SEC) was conducted to further purify CalpS (Fig. 4-6D) The purity and integrity of CalpS were analysed by SDS-PAGE (Fig. 4-6B). Unexpectedly, a substantial amount of CalpS co-purified with alpha and beta subunits of the DNA-directed RNA polymerase (RNAP) from *E. coli* (Fig. 4-6B). Sequence analysis unveiled a 44% sequence identity between beta subunits of RNAP from *Sulfurihydrogenibium* spp. and *E. coli*. This result provides support of our earlier hypothesis that CalpS functions as a sigma factor.
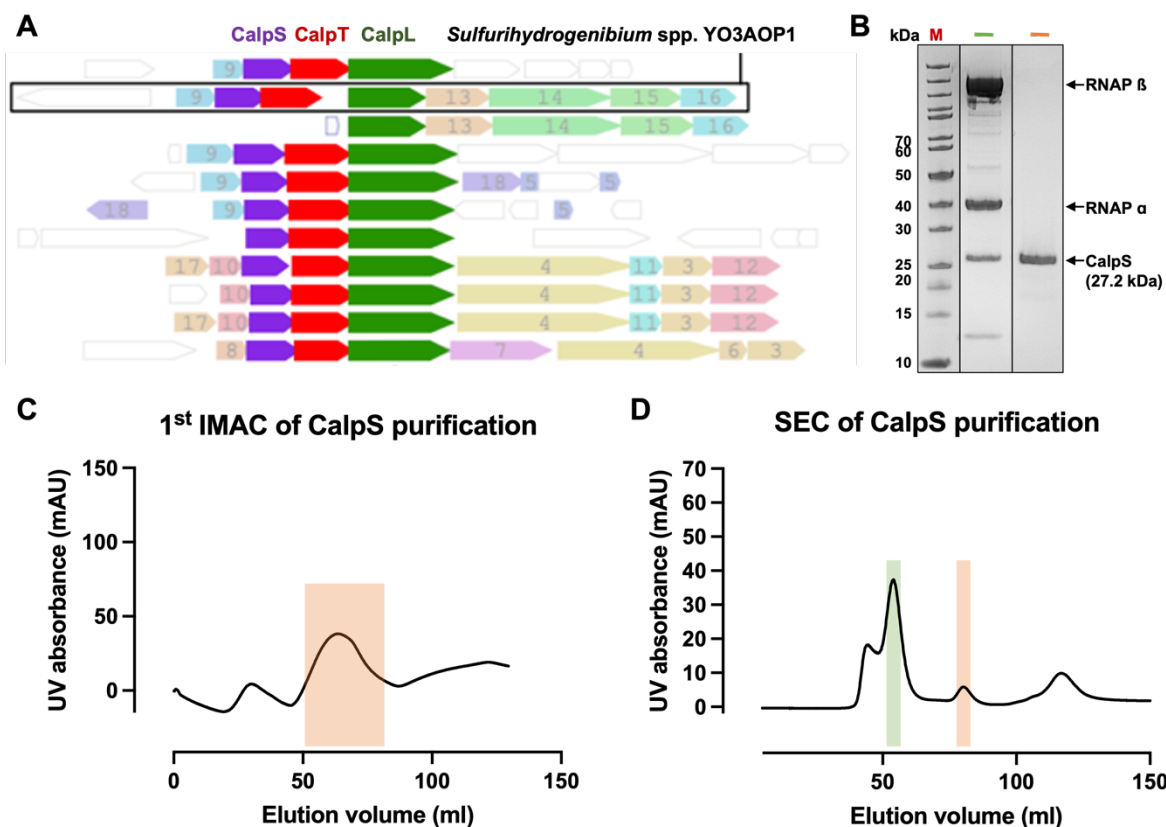
Figure 4-6 Purification of ECF sigma factor CalpS
A. The CRISPR loci of selected type III CRISPR systems with CalpS, CalpT and CalpL effector proteins. The genomic context of genes adjacent to CalpL gene (green) was investigated by using the WebFLAGs server (Saha et al., 2021), indicating its neighborhood *calpT* gene (red) and *calpS* gene (blue). **B.** SDS-PAGE analysis of purified CalpS. The monomer mass was approximately 27 kDa, consistent with the theoretical mass of CalpS. The bars above the image represent fractions obtained from SEC (D), with each color corresponding to the respective peaks of the same color. M is the marker to indicate size on the gel. **C.** First immobilised metal affinity chromatography (1st IMAC) for CalpS purification. The fractions containing target protein highlighted by a red rectangle was evaluated by SDS-PAGE and pooled for his tag removal. **D.** Superdex200 SEC profiles for CalpS. The TEV-cleaved CalpS was recovered from the nickel column and then subjected to SEC. Fractions indicated by a red rectangle were collected and concentrated for further enzymatic analysis, and fractions indicated by a green rectangle was analysed by mass spec to identify proteins co-purified with CalpS.

4.2.5 Formation of a ECF σ factor CalpS and its anti-sigma factor CalpT complex

Crystallisation studies have elucidated several structures of ECF sigma factor bound to their cognate anti-sigma factor, including σ^E-RseA from *E. coli* (Campbell et al., 2003), σ^E-ChrR from *Rhodobacter sphaeroides* (Campbell et al., 2007), σ^W-RsiW from *Bacillus subtilis* (Devkota et al., 2017) and σ^K-RskA from *Mycobacterium tuberculosis* (Shukla et al., 2014). The expression and activity of sigma factor are typically regulated by anti-sigma factors, which bind to sigma factors to prevent their interaction with RNA polymerase. Anti-sigma factors release sigma factors from this inhibition in response to the specific stimuli (Sineva et al., 2017, Paget, 2015).

To investigate the potential binding of proposed anti-sigma factor CalpT to sigma factor CalpS, size exclusion chromatography was performed on a Superose6 increase 10/300 column (GE Healthcare), equilibrated with SEC buffer. As control samples, we separately loaded CalpS and CalpT to the column. The SEC analysis revealed peaks corresponding to their respective protein sizes, with CalpS at 27 kDa and CalpT at 32 kDa (Fig. 4-7A). Next, a mixture of CalpS and CalpT in a 1:1 ratio was incubated at room temperature for 15 min before loading onto the SEC column. Notably, a single peak was observed for the CalpS and CalpT mixture, eluting earlier than either CalpS or CalpT alone (Fig. 4-7A). This observation suggests the formation of a stable CalpS and CalpT complex. Furthermore, CalpS was incubated with His-tagged CalpT (His-CalpT) or His-tagged truncated CalpT (His-CalpT^{tr}) in a binding buffer with magnetic nickel beads at room temperature. We observed co-elution of CalpS with both His-CalpT and His-Calp^{tr} (Fig. 4-7B). These data support the possible interaction between CalpS and CalpT, with a particular contribution from the N-terminal region of CalpT.

To gain a deeper understanding of the formation of the CalpS and CalpT complex, our collaborators, Gregor Hagelueken and Niels Schneberger, preformed structural modelling of a potential heterotrimeric complex involving CalpS, CalpT and CalpL (Fig. 4-7C). The predicted structures of CalpS and CalpT were generated by Alphafold2 and the crystal structure of CalpL was solved at a resolution of 2.1 Å. Notably, the modelling suggested that CalpT$_{23}$ interacts with CalpS, forming an interface with a combined buried surface area with high confidence scores about 4,000 Å (Krissinel and Henrick, 2007) and credible side chain interactions (Fig. 4-7C). Furthermore, CalpT was observed to specifically interact with σ_2 and σ_4 domains of CalpS. This interaction effectively blocks most of the -10 region interface (Fig. 4-7C), thus preventing interaction of CalpS with the RNAP complex, which is consistent with other characterised sigma factors (Sineva et al., 2017, Paget, 2015).
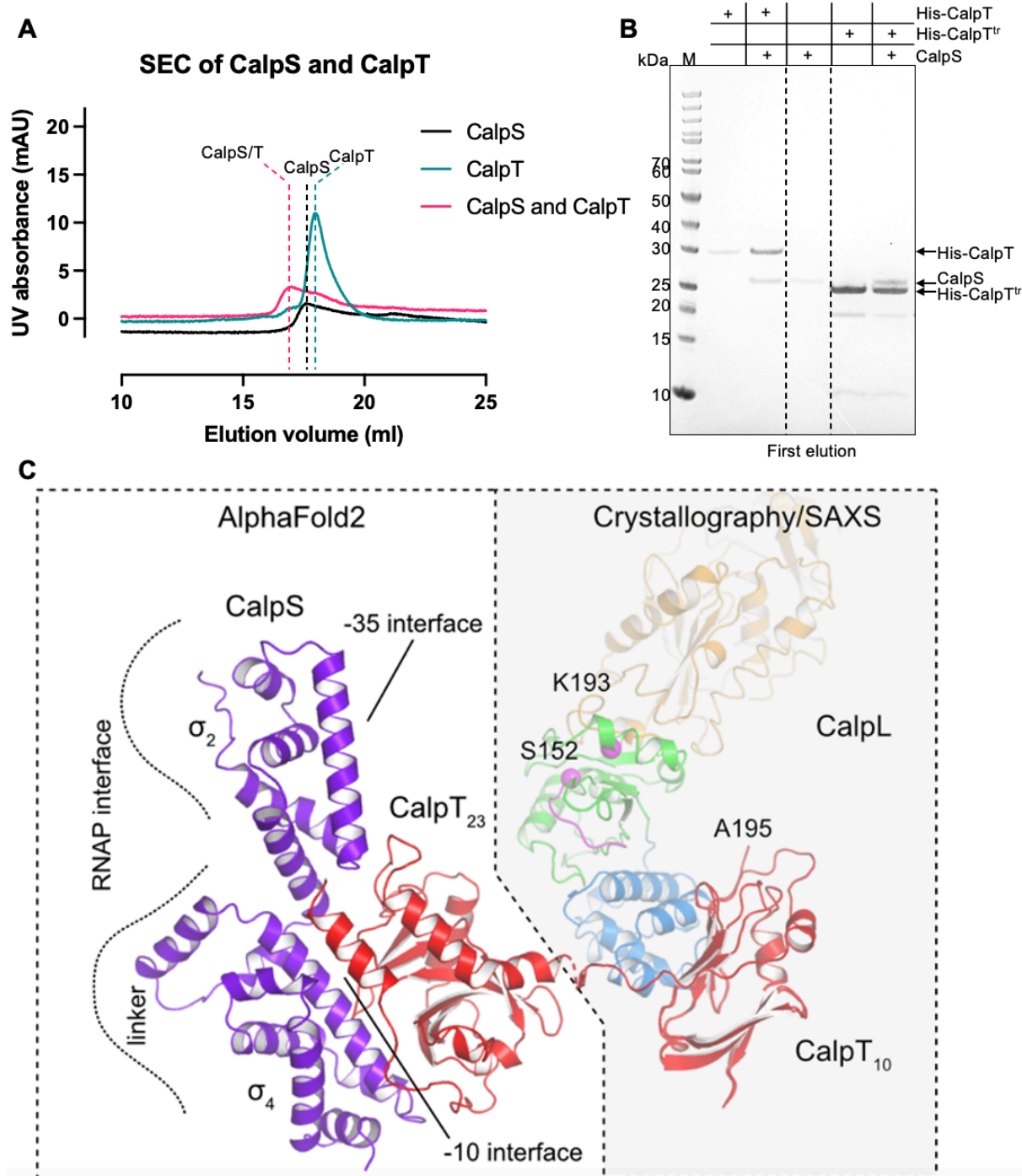
Figure 4-7 Formation of a stable CalpS and CalpT complex
A. SEC profiles of CalpS and CalpT with Superose6 increase column. A single peak of CalpS and CalpT mixture (red) was eluted out earlier than CalpT alone (green) and CalpS alone (black) during SEC, suggesting the formation of a stable complex. **B.** Pulldown assay involving CalpT and CalpS. Following incubation with His-CalpT and His-CalpT^{tr}, CalpS was pulled down in both cases, indicating the interaction between them. **C.** High confidence structural model of the CalpL-CalpT-CalpS complex was obtained by a combination of crystallography, SAXS and Alphafold2. Predicted CalpS structure is shown in purple with indicating the predicted RNAP interface and σ_2 and σ_4 domains. A structural prediction of CalpT$_{23}$ is presented in red and positioned interacting with CalpS, which blocks most of the -10 region interface to interference with the interaction between CalpS and RNAP.

4.2.6 Co-purification of CalpS and CalpT complex

To further confirm the formation of the CalpS and CalpT complex, we conducted co-purification assays in two ways, using his-tagged CalpS to pull down CalpT, or the other way around. Specifically, *E. coli* C43 (DE3) cells were co-transformed with the plasmids pEHisTEV-CalpS and pCDFDuet-CalpT for the purpose of using his-tagged CalpS to pull down CalpT (Fig. 2-4A and Fig. 2-5A). Additionally, *E. coli* C43 (DE3) cells were co-transformed with plasmids pET11a-CalpT and pCDFDuet-CalpS to test if CalpS could be pulled down by his-tagged CalpT (Fig. 2-5B and C). Both co-purification processes followed similar procedures, except that his tag was not removed in the co-purification of his-tagged CalpS with CalpT.

Briefly, the co-expression and co-purification of CalpS and CalpT initially involved immobilised metal affinity chromatography (IMAC) (Fig. 4-8A and D). Fractions containing HisCalpT and CalpS were collected and incubated with TEV overnight at room temperature to remove the His tag. Subsequently a second round of IMAC was performed to recover TEV-cleaved CalpT/S. Further purification was achieved through size-exclusion chromatography (SEC) (Fig. 4-8B and E). The purity and integrity of the purified proteins were evaluated using SDS-PAGE (Fig. 4-8C and F). However, the sizes of the his-tagged CalpS and CalpT were quite similar, both round 32 kDa, resulting in their co-migration on SDS-PAGE. Their identities were finally confirmed through mass spectrometry by analysis of the gel bands. Thus, the co-purified CalpS and CalpT strongly support that CalpS and CalpT can form a complex that is stable through both IMAC and SEC processes.
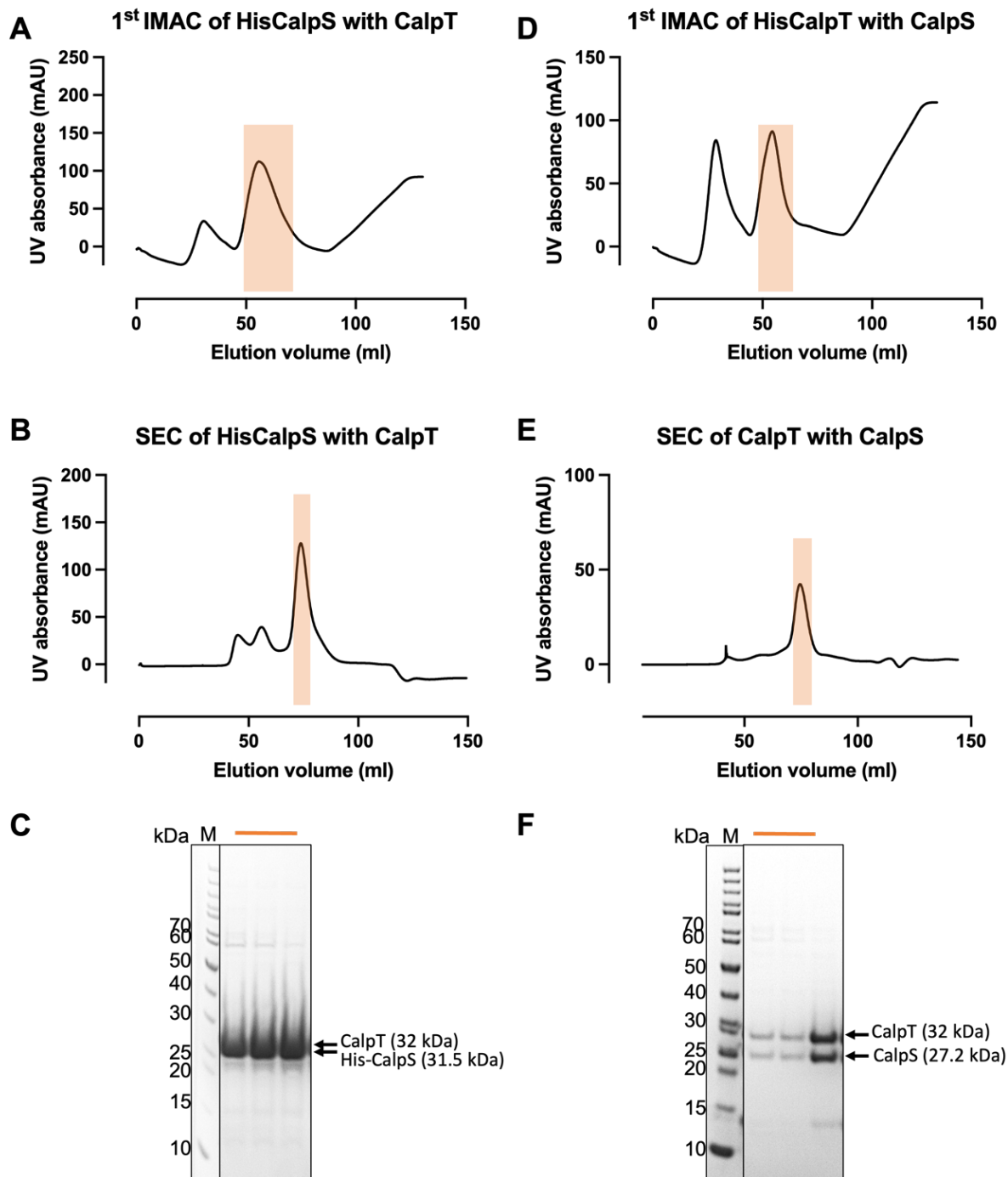
Figure 4-8 Co-purification of CalpS and CalpT complex
A and D. First immobilised metal affinity chromatography (1st IMAC). The fractions containing target protein highlighted by a red rectangle was evaluated by SDS-PAGE and pooled for his tag removal, except that his tag was kept for his tagged CalpS pulling down CalpT. **B and E**. Superdex200 SEC profiles. The TEV-cleaved CalpT pulling down CalpS was recovered from the nickel column and then subjected to SEC. His tagged CalpS with CalpT was directly subjected to SEC after 1st IMAC. Fractions indicated by a red rectangle were collected and concentrated for further enzymatic analysis. **C and F.** SDS-PAGE analysis. M is the marker to indicate size on the gel. Red bars above represent the tested fractions from corresponding SEC.

4.2.7 CalpL, CalpT and CalpS form a tripartite complex

Structural modelling predicted CalpL, CalpT and CalpS might form a trimeric complex, incorporating experimental evidence for the formation of both a CalpL and CalpT complex and a CalpT and CalpS complex. This led us to investigate if this ternary complex assembles and explore the regulation of the system.

We analysed CalpL and the complex of CalpT and CalpS and their equimolar mixtures using size exclusion chromatography (SEC). CalpL alone eluted as a single peak at 17.5 ml, while the complex of CalpT and CalpS eluted at 17.1 ml (Fig. 4-9A), with protein identities confirmed by SDS-PAGE analysis (Fig. 4-9B). Notably, when all three proteins were mixed in a 1:1:1 ratio, they eluted as a single elution peak at 16 ml (Fig. 4-9C) and the identities from this peak were validated through SDS-PAGE analysis (Fig. 4-9D). These results provide evidence for the formation of a tripartite complex involving CalpL, CalpT and CalpS. Given that CalpT is specifically cleaved by protease CalpL upon the induction of cA_4, we subsequently investigated if CalpS could be released from this ternary complex following the addition of cA_4. To test this, SEC was used to analyse the sample containing the ternary complex and cA_4, resulting in a shift in the elution peak from 16 ml to 17.4 ml, compared with the SEC profile in the absence of cA_4 (Fig. 4-9C). SDS-PAGE analysis suggested that CalpT was cleaved into $CalpT_{23}$ and $CalpT_{10}$ after the addition of cA_4 (Fig. 4-9D). Considering the similarity in the protein sizes and the appearance of a single peak after cleavage, it's highly possible that the ternary complex was divided into two components, CalpL - $CalpT_{10}$ with a molecular weight (MW) of approximately 67.7 kDa and CalpS - $CalpT_{23}$ with a MW of about 50.3 kDa. To further investigate this, we conducted a pull-down assay by incubating a complex of his-tagged CalpS and CalpT with CalpL at 60 °C for 60 min in the presence or absence of cA_4. In the presence of cA4, only his tagged CalpS and $CalpT_{23}$ were observed after washing and eluting from the nickel beads (Fig. 4-9E). These finding strongly support the notion that CalpL, CalpT and CalpS form a stable complex, preventing sigma factor CalpS from interaction with RNA polymerase (RNAP), and releasing CalpS from the complex in response to the signal molecule cA4.
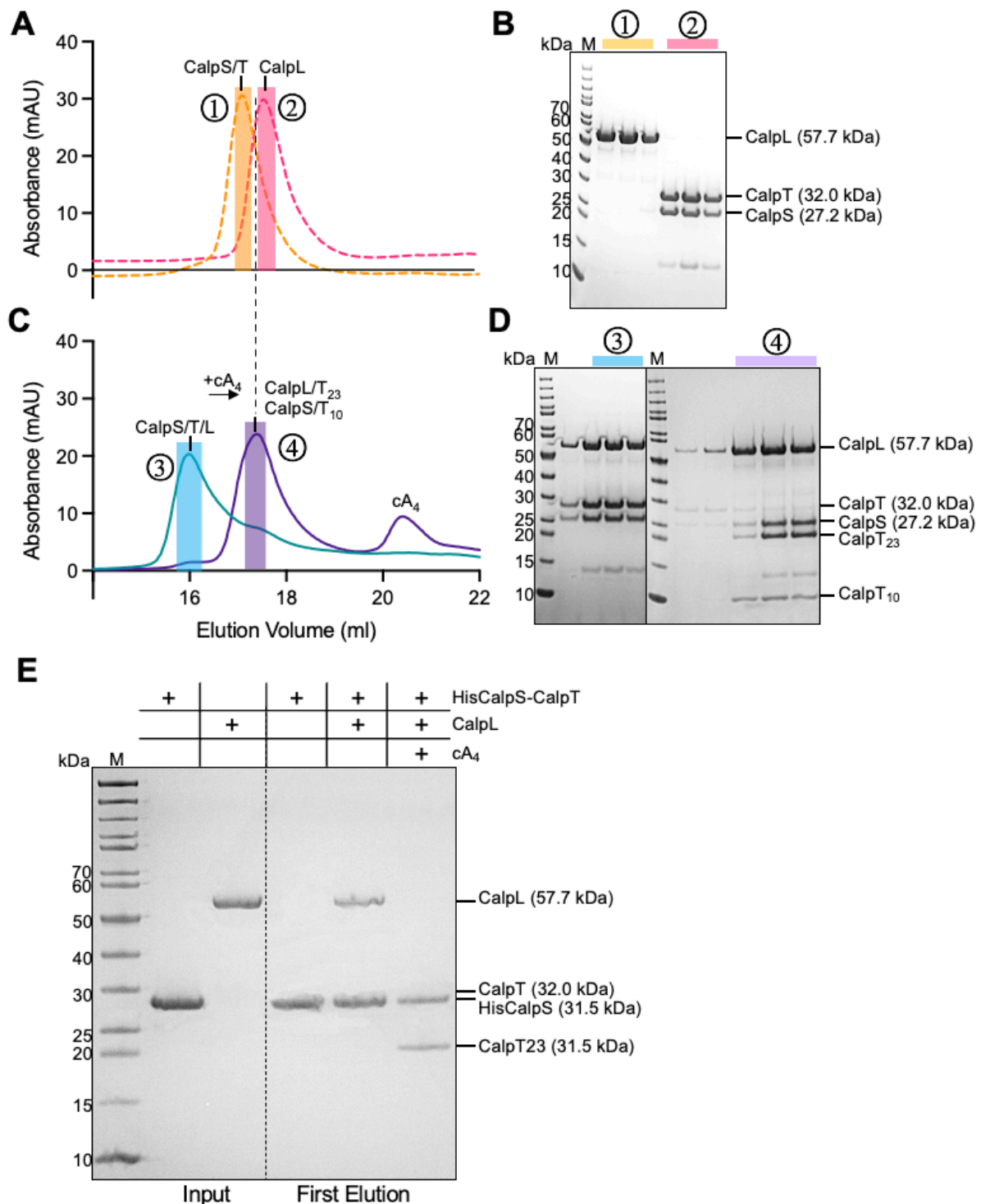
Figure 4-9 Formation of a ternary complex of CalpL, CalpT and CalpS
A. SEC profiles of CalpL alone and a complex of CalpT and CalpS. CalpT (63 µM, red) and a complex of CalpS/T (116 µM, yellow) were analysed on a pre-equilibrated Superose6 increase 10/300 chromatography column. **B.** SDS-PAGE analysis of proteins. The colour-coded bars above the image correspond to fractions eluted from SEC, highlighted with the matching colour. **C.** SEC profiles of a complex of CalpL, CalpT and CalpS with and without cA_4. The addition of cA_4 (60 µM) induced the cleavage of CalpT. **D.** SDS-PAGE analysis of proteins eluted from SEC (C). **E.** Pulldown assays. The dissociation of CalpS - $CalpT_{23}$ from the ternary complex was observed after induction by cA_4.

4.3 Discussion

The vast and diverse range of ancillary proteins linked to CRISPR system were discovered through an in-depth analysis of the genomic neighbourhoods flanking the core *cas* genes (Shmakov et al., 2018, Shah et al., 2019). This came shortly after addressing one of the most significant discoveries regarding the function of the Cas10 Palm domain for the generation of cOA (Niewoehner, 2017, Kazlauskiene et al., 2017). Among the ancillary effectors, those harbouring a signal sensor CARF domain have emerged as the most prevalent and extensively characterised, particularly those fused to a nuclease, like RNases of the HEPN, PIN, and RelE, and PD-D/ExK endonucleases families (Makarova et al., 2020a). However, the effectors containing another key sensor SAVED domain fused to a Lon protease domain aroused our interest, as neither of these domains had been investigated within the CRISPR system. Our collaborators have determined crystal structure of apo CalpL, the complex of CalpL-cA$_4$ and CalpL-CalpT$_{10}$. CalpL forms a 1:1 complex with CalpT and specifically cleaves CalpT upon cA$_4$ binding in the SAVED domain of CalpL with a nanomolar affinity (1 nM) (Fig. 4-1 and Fig. 4-10). This binding enables oligomerisation of CalpL, a phenomenon commonly observed in SAVED-containing effectors within CBASS defence systems (Lowey et al., 2020, Fatma et al., 2021, Hogrel et al., 2022), but not previously determined in CRISPR systems. It's noteworthy that while some CARF domain proteins exhibit dual functions by degrading their activators, this degradation activity has not been observed for SAVED domains and remains untested for CalpL.

The observation of the CalpL-CalpT-CalpS cascade signifies the establishment of a multifaceted signalling network tightly regulated by cA$_4$ (Fig. 4-10). This network incorporates elements from CRISPR adaptive immune systems with diverse innate defence systems, including proteolysis, the TA systems, as well as sigma and anti-sigma systems. Initially, CalpT was presumed to exhibit MazF-like nuclease activity, as the N-terminal half of CalpT is homologous to the MazF toxin. Subsequent experiments revealed the formation of a stable complex with CalpS, suggesting its role as an anti-sigma factor. CalpS belongs to the ExtraCytoplasmic Function (ECF) family of sigma factors, renowned for their role in sensing and responding to extracellular stresses like envelope, iron transport, or oxidative stress (Sineva et al., 2017, Paget, 2015). Typically, ECF σ factors are controlled by their cognate anti σ factor which hinders RNAP binding by occluding the crucial RNAP binding sites of σ factors (Campbell et al., 2008). This inhibition mechanism has been observed in a predicted CalpT-CalpS complex, where CalpT blocks the major RNAP binding determinants in σ$_2$ and σ$_4$ of

CalpS (Fig. 4-7C). The release of ECF σ factor from its anti σ factor is commonly regulated by proteolysis in response to various stimuli. For instance, σ^E from *E. coli* is entirely released from its anti σ factor RseA through a series of proteolysis progresses that degrade RseA (Paget, 2015). Based on these established patterns, CalpS is possibly released through further proteolysis of $CalpT_{23}$, initiating transcription to provide immunity (Fig. 4-10). However, the identity of the proteases responsible for this proteolysis, and the specific genes regulated by CalpS remain unclear. Further investigations are therefore required.

Crosstalk between CRISPR systems and proteases have also been discovered in the form of a CRISPR-guided caspase (Craspase) in the type III-E CRISPR-Cas system. Specifically, the Cas7-11 CRISPR complex interacts with the protease Csx29 (also known as TPR-CHAT), which becomes activated upon the binding of invading RNA to the interference effector Cas7-11 (Hu et al., 2022, van Beljouw et al., 2021). Notably, ECF sigma factor RpoE (termed CASP-σ) has been identified as a key component involved in this Cas7-11-Csx29 mediated Craspase pathway. CASP-σ experiences inhibition due to its binding to Csx30 and this inhibition is relieved upon the proteolytic cleavage of Csx30 by the activated protease Csx29. Furthermore, binding motifs of CASP-σ were identified within the CRISPR locus related to CRISPR adaptation, including Cas1 and Cas2 (Strecker et al., 2022). It's worth noting that the protease Csx29 belongs to the caspase family, unrelated to CalpL (from the Lon family), and is not regulated by signal molecules. Nevertheless, both proteolytic processes within CRISPR systems revealed a remarkable level of complexity and finely tuned regulation.
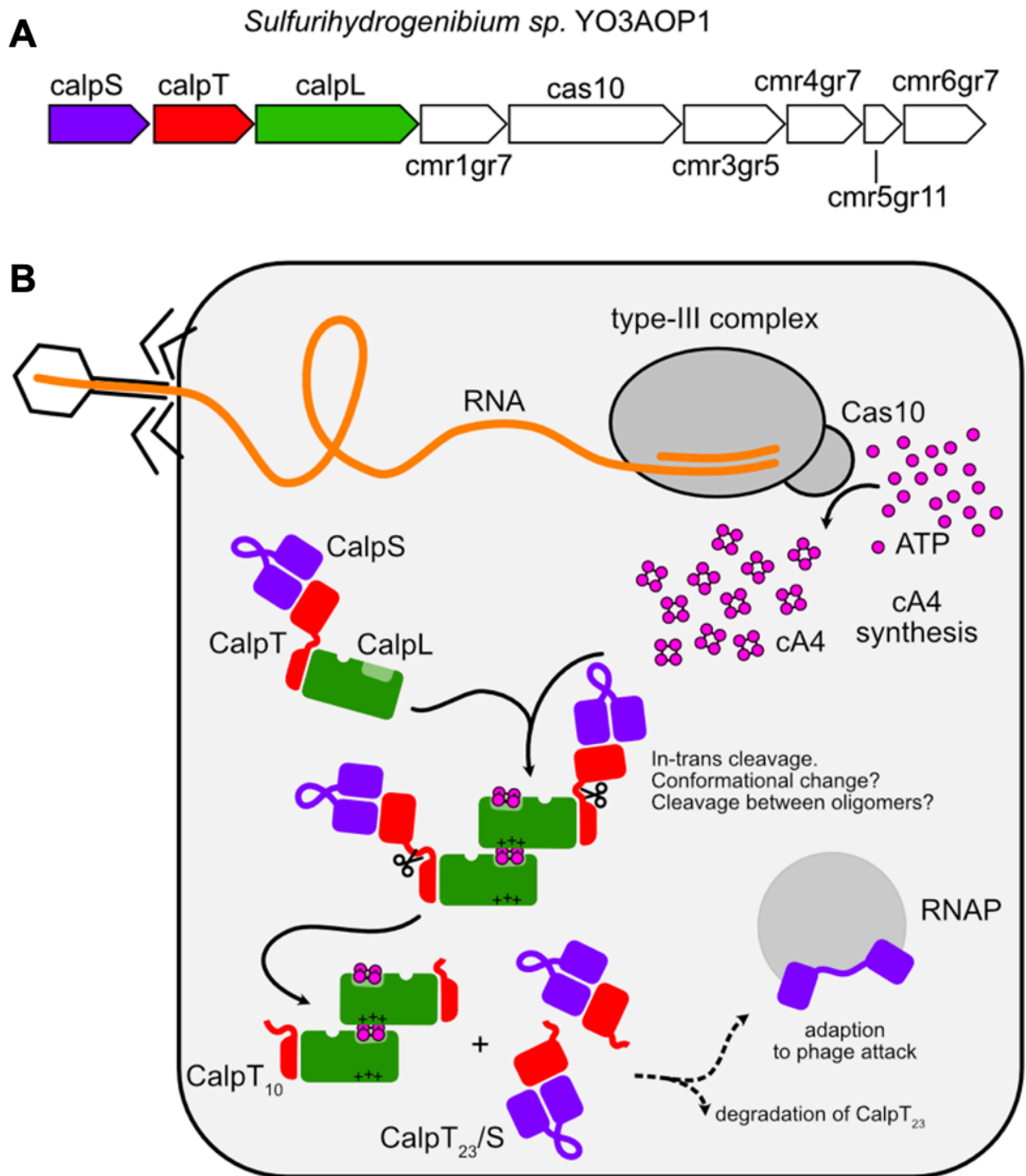
Figure 4-10 Model of CalpL-CalpT-CalpS mediated antiviral defense
A. The genomic context of CRISPR locus in *Sulfurihydrogenibium* spp. YO3AOP1. The genes of *calpL* (green), *calpT* (red), *calpS* (blue) are adjacent to the core *cas* genes. **B.** Upon detection of invading RNA, the Cas10 subunit of the RNP is activated to synthesise cA4 from ATP. The second messenger cA4 subsequently binds to preformed CalpL-CalpT-CalpS ternary complex, resulting in the oligomerization of CalpL. CalpL is thereby activated to cleave CalpT, releasing the CalT23-CalpS fragment. CalpT23 is probably degraded by proteases, which in turn enables CalpS to associate with the RNAP.

135

5 Conclusions and future work

5.1 *B. fragilis* Cmr functions as a novel membrane channel protein-associated type III-B system

Type III CRISPR systems exhibit the distinctive ability to synthesise cyclic oligoadenylates as second messengers (Niewoehner, 2017, Kazlauskiene et al., 2017). These signals can be amplified over 1000 times upon detecting one invading RNA and activate collateral activities of large numbers of ancillary effectors, resulting in cell dormancy or death (Athukoralage et al., 2020a, Rouillon et al., 2018). Almost all characterised cOA-activated ancillary effectors function as nucleases, like Cax1/Csm6 family ribonucleases, Can1/Can2 nucleases and NucC endonuclease (Athukoralage et al., 2020a). Recent bioinformatic analysis showed numerous membrane proteins encoded within the type III CRISPR loci, suggesting the membrane is tightly associated with antiviral signalling pathways (Shmakov et al., 2018). However, their functions in the CRIPSR systems remain unclear. We are thus interested in the most abundant membrane protein CorA associated type III-B CRISPR systems from *B. fragilis*.

We initially reconstituted and expressed the BfrCmr system in the heterologous host *E. coli*. BfrCmr restricted plasmid transformation in the presence of both ancillary proteins, membrane protein BfrCorA and phosphodiesterase BfrNrN. This plasmid immunity was abolished when either the cyclase domain of the BfrCmr complex or catalytic sites of BfrNrN were mutated or when the transmembrane domain of BfrCorA was removed. These findings suggested that the BfrCmr system function was dependent on a signal molecule-mediated pathway. Furthermore, no plasmid immunity was observed when both BfrCorA and BfrNrN were introduced into another well-characterised type III CRISPR system which had been proven to generate a range of cOA (cA_{2-6}). These data imply that BfrCmr system exhibits a different defence mechanism from the canonical type III CRISPR systems.

BfrCmr systems were subsequently characterised *in vitro*. BfrCas6 was first shown to process pre-crRNA into crRNA intermediates by cleaving the CRISPR repeat sequence to generate a canonical 8 nt 5' tag. BfrCmr complex was then purified successfully from *E. coli* in the presence of BfrCas6 and a mini CRISPR array. Three major crRNA species were isolated from purified BfrCmr complex, which varied in the length by 6 nt increments, suggesting a variable composition of BfrCmr complex with different numbers of the crRNA-binding backbone protein Cas7. Moreover, extracted crRNA species were all shorter than the BfrCas6 processed crRNA, indicating unknown cellular nucleases in *E. coli* were involved in trimming from 3'

end to help crRNA maturation. These features have been observed previously in other type III CRISPR effectors (Shao et al., 2016, Tamulaitis et al., 2017). We subsequently detected the rapid cleavage of target RNA with 6 nt intervals by BfrCmr and confirmed by mutagenesis that this cleavage activity was mediated by the BfrCmr4 subunit. Collectively, purified BfrCmr complex was functional in many respects. However, no cOA signal molecules were detected when BfrCmr complex was incubated with ATP or even the mixture of four ribonucleotides (ATP, UTP, CTP and GTP).

5.2 BfrCmr systems produce a new class of signalling molecule, SAM-AMP

To investigate the signal molecules synthesised by BfrCmr system, the BfrCmr complex was activated *in vivo* in *E. coli* in the absence of ancillary proteins, after which the nucleotide products were purified and isolated from cell lysates. A significant HPLC peak was detected from the activated BfrCmr wild type systems, but not from the inactive or the cyclase mutant systems. In addition, the retention time of this peak in HPLC was different from those of cOA (cA$_3$ and cA$_4$) standards. The subsequent MS analysis identified a m/z value of 728.1963, which didn't match any known natural or synthetic molecules. The further fragmentation by MS/MS analysis enabled us to identify the fragments of AMP and methionine. These data indicated the isolated molecule is S-adenosyl methionine (AdoMet, SAM) adenylated on the ribose moiety, which is designated as SAM-AMP. Most importantly, the SAM-AMP synthesis pathway was successfully reconstituted *in vitro* by incubating BfrCmr complex with both ATP and SAM.

BfrCas10 accepts both SAM and ATP as substrates, instead of ATP alone, indicating differences in the acceptor PALM pocket compared to canonical Cas10. The different local charges of methionine moiety of SAM (+1) and the triphosphate group of ATP (-4) suggests the involvement of less basic protein residues in the recognition of the methionine moiety. The comparison of the structure model of BfrCas10 (Cmr2) and the crystal structure of Cmr2dHD-Cmr3 complex from *P. furiosus* (Osawa et al., 2013) highlighted two highly conserved acidic residues in BfrCas10, D70 and E151, which correspond to the N300 and R436 residues of PfuCas10, potentially involved in SAM binding. BfrCas10 variants with mutations D70N or E151R were defective in SAM-AMP synthesis, whereas the double mutant was virtually inactive. In addition, the double mutant could generate a slight amount of pppApA when incubated with ATP alone, as compared with wild type. These data hint at the evolutionary steps required to evolve from a cOA to a SAM-AMP specific Cas10.

5.3 Ancillary proteins for SAM-AMP signalling

Three ancillary proteins were next purified from *E. coli* to near homogeneity. The purified membrane protein BfrCorA showed specific binding affinity to the signal molecule SAM-AMP. The predicted structural model of BfrCorA indicates the potential binding sites are two pairs of conserved residues R152/R153 and D219/D220. Two variants of BfrCorA (R152A/R153A and D219A/D220A) could be expressed but could not be purified and thus were tested *in vivo* in the plasmid challenge assay. Both variants abolished plasmid immunity in the context of activated BfrCmr system. BfrCorA may therefore function as a ligand-regulated ion channel, conferring immunity upon SAM-AMP binding.

The phosphodiesterase BfrNrN specifically cleaves SAM-AMP into SAM and AMP. No cleavage activities were detected when incubated with cOA ($cA_{2/3/4/6}$) or linear dinucleotides. A variant of BfrNrN (D85:H86:H87) eliminated this cleavage activity. An alternative SAM-AMP degradation mechanism was observed to utilise a SAM-AMP lyase, which cleaves SAM-AMP into 5'-methylthioadenosine (MTA) and L-homoserine lactone (HL). SAM-AMP lyase is structurally similar to a family of phage SAM lyases, which neutralise host RM systems by depleting SAM pools (Guo et al., 2021, Simon-Baram et al., 2021). However, SAM-AMP lyase from *C. botulinum* degrades SAM-AMP more efficiently than SAM.

BfrNYN showed constitutive, Mn^{2+}-dependent ribonuclease activity in a signal molecule-independent manner. A smear-like degradation pattern was observed when BfrNYN was incubated with the CRISPR array (285 nt) and the presence or absence of BfrCas6 had no effects on the ribonuclease activity of BfrNYN. Furthermore, BfrNYN was unable to cleave SAM-AMP either *in vivo* or *in vitro*. These findings suggest BfrNYN is likely involved in the crRNA maturation in the cognate *B. fragilis* host.

5.4 The antiviral signalling connects CRISPR-based detection of foreign nucleic acids and transcriptional regulation

CHAPTER 4 focused on a Lon protease CalpL-associated type III B CRISPR system. CalpL contains a SAVED signal sensor domain and a Lon protease domain. The MazF homologue CalpT and extracytoplasmic Sigma factor homologue CalpS are encoded by adjacent genes in the same operon in the thermophilic bacterium *Sulfurihydrogenibium* spp. YO3AOP1. Our collaborators Christophe Rouillon, Niels Schneberger and Gregor Hagelueken investigated the structure and function of the SAVED-containing protein CalpL. They solved crystal structures of apo and cA_4-bound CalpL, demonstrating cA_4 binding to the SAVED domain. They showed

CalpL forms a stable complex with CalpT and is activated by cA_4 to cleave CalpT into two distinct products $CalpT_{23}$ and $CalpT_{10}$ with molecular weights of 23 and 10 kDa, respectively. This is the first experimentally characterised SAVED sensor domain binding to cA_4 in a CRISPR effector and one of the first examples where the effector functions as a protease, instead of nuclease.

The cA_4-mediated interference mechanism remained unclear, even though the protease activity of CalpL had been identified. Thus, we set out to investigate the function of CalpT and CalpS. The N-terminal MazF-like toxin in CalpT was released upon cA_4-activated CalpL cleavage, which was presumed to function as a RNase like MazF. However, we failed to detect any ribonuclease activities of either truncated or cleaved $CalpT_{23}$ upon screening RNA substrates with various lengths and sequences, or even random ssRNA libraries. These data strongly suggested that CalpT might exhibit an alternative activity. In addition, we observed the presence of a third conserved gene, *calpS* encoding a predicted extracytoplasmic function (ECF) family σ factor. When conducting CalpS purification, the alpha and beta subunits of the DNA-directed RNA polymerase (RNAP) from *E. coli* were co-purified, consistent with the hypothesis that CalpS may function as a sigma factor.

CalpS was next co-purified with CalpT. Interestingly, no RNAP alpha and beta subunits co-purified under these conditions. CalpT thus is proposed to function as anti-sigma factor, as the activity of a sigma factor is typically regulated by a cognate anti-sigma factor through interaction between them (Paget, 2015). We then detected the formation of a stable ternary complex among CalpL, CalpT and CalpS. Two components, CalpL - $CalpT_{10}$ and CalpS - $CalpT_{23}$, were observed after the addition of cA_4. It is predicted that after the degradation of cleaved $CalpT_{23}$, CalpS could be completely released to interact with RNA polymerase, enabling adaption to phage attack by transcriptional reprogramming.

5.5 Future work

We have conducted biochemical investigation of the BfrCmr system and revealed a novel signalling pathway. Further structure analysis would allow deeper understanding of this system. For example, the substrate preference of Cas10 and the details about the assembly mechanism of SAM-AMP could be investigated through cryo-EM analysis of the BfrCmr complex. The regulatory and interference mechanism of the membrane protein BfrCorA mediated by SAM-AMP could be elucidated via further structural analysis or biological physics. Structural analysis of phosphodiesterase BfrNrN and the ribonuclease BfrNYN would also be important to provide insight into the molecular mechanism of the BfrCmr system.

It would also be interesting to investigate the BfrCmr system in the cognate host *B. fragilis*, to investigate whether SAM-AMP signalling has other effects *in vivo*. Species of Bacteroides account for around 25 % anaerobic microbiome colonised in the human colon, where possesses a complex ecosystem in the body, including bacteria, fungi, archaea, viruses and protozoa (Martin et al., 2014). These communities play vital roles in human health and disease (Chatterjee and Duerkop, 2018). *B. fragilis* are usually beneficial to their host when as part of colon flora. However, it can cause significant polymicrobial infection once the spread of *B. fragilis* into blood or adjacent tissue (Wexler, 2007). Their virulence is mostly due to toxin production which results in inflammatory disease (Sears, 2009). The reasons underlying this transition remain unclear. Considering that CRISPR-Cas systems are associated with acquisition of invading genetic elements, including virulence and antimicrobial genes, investigating the CRISPR-Cas system in *B. fragilis* could expand our understanding of opportunistic pathogen. Furthermore, the treatments of *B. fragilis* infection mainly include source control and targeted antimicrobial therapy (Bogdan et al., 2018). However, antimicrobial resistance (AMR) has raised as a major concern. Phage therapy is thus rising as an alternative treatment, as phage infects and lyses specific bacteria to restore the gut microbiome balance and control disease progression (El Haddad et al., 2022). Finally, investigating the defence systems in anaerobes will enhance our understanding in the phage mode of action, thus contributing to phage therapy.

Additionally, type III CRISPR-Cas system had recently been revealed to provide defence against nucleus-forming jumbo phages via abortive infection, protecting the population of bacterial colony (Mayo-Munoz et al., 2022). Considering the complex communities in human colon, it would be interesting to explore if SAM-AMP as a signalling molecule in *B. fragilis* type III-B CRISPR-Cas systems has crosstalk with other microbial species or even human host, once exposure to the gut microbial communities. If so, it will be interesting to explore SAM-AMP derived compounds for potential application in biochemistry or biomedicine.

References

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E. S., Koonin, E. V. and Zhang, F. (2016) 'C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector', *Science,* 353(6299), pp. aaf5573. doi: 10.1126/science.aaf5573

Abudayyeh, O. O. a. G., J.S. (2022) *CRISPR-Cas13: Biology, Mechanism, and Applications of RNA-Guided, RNA-Targeting CRISPR Systems. In Crispr.*

Amitai, G. and Sorek, R. (2016) 'CRISPR-Cas adaptation: insights into the mechanism of action', *Nat Rev Microbiol,* 14(2), pp. 67-76. doi: 10.1038/nrmicro.2015.14

Anantharaman, V. and Aravind, L. (2006) 'The NYN domains: novel predicted RNAses with a PIN domain-like fold', *RNA Biol,* 3(1), pp. 18-27. doi: 10.4161/rna.3.1.2548

Anders, C., Niewoehner, O., Duerst, A. and Jinek, M. (2014) 'Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease', *Nature,* 513(7519), pp. 569-73. doi: 10.1038/nature13579

Athukoralage, J. S., Graham, S., Gruschow, S., Rouillon, C. and White, M. F. (2019) 'A Type III CRISPR Ancillary Ribonuclease Degrades Its Cyclic Oligoadenylate Activator', *Journal of Molecular Biology,* 431(15), pp. 2894-2899. doi: 10.1016/j.jmb.2019.04.041

Athukoralage, J. S., Graham, S., Rouillon, C., Gruschow, S., Czekster, C. M. and White, M. F. (2020a) 'The dynamic interplay of host and viral enzymes in type III CRISPR-mediated cyclic nucleotide signalling', *Elife,* 9. doi: 10.7554/eLife.55852

Athukoralage, J. S., McMahon, S. A., Zhang, C., Gruschow, S., Graham, S., Krupovic, M., Whitaker, R. J., Gloster, T. M. and White, M. F. (2020b) 'An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity', *Nature,* 577(7791), pp. 572-575. doi: 10.1038/s41586-019-1909-5

Athukoralage, J. S., McQuarrie, S., Gruschow, S., Graham, S., Gloster, T. M. and White, M. F. (2020c) 'Tetramerisation of the CRISPR ring nuclease Crn3/Csx3 facilitates cyclic oligoadenylate cleavage', *Elife,* 9. doi: 10.7554/eLife.57627

Athukoralage, J. S., Rouillon, C., Graham, S., Gruschow, S. and White, M. F. (2018) 'Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate', *Nature,* 562(7726), pp. 277-280. doi: 10.1038/s41586-018-0557-5

Athukoralage, J. S. and White, M. F. (2021) 'Cyclic oligoadenylate signalling and regulation by ring nucleases during type III CRISPR defence', *RNA,* 27(8), pp. 855-67. doi: 10.1261/rna.078739.121

Athukoralage, J. S. and White, M. F. (2022) 'Cyclic Nucleotide Signaling in Phage Defense and Counter-Defense', *Annu Rev Virol,* 9(1), pp. 451-468. doi: 10.1146/annurev-virology-100120-010228

Bandyopadhyay, P. K., Studier, F. W., Hamilton, D. L. and Yuan, R. (1985) 'Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7', *J Mol Biol,* 182(4), pp. 567-78. doi: 10.1016/0022-2836(85)90242-6

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. and Horvath, P. (2007) 'CRISPR provides acquired resistance against viruses in prokaryotes', *Science,* 315(5819), pp. 1709-12. doi: 10.1126/science.1138140

Beckett, M. Q., Ramachandran, A. and Bailey, S. (2022) *Type V CRISPR-Cas Systems. In Crispr*

Bernheim, A., Millman, A., Ofir, G., Meitav, G., Avraham, C., Shomar, H., Rosenberg, M. M., Tal, N., Melamed, S., Amitai, G. and Sorek, R. (2021) 'Prokaryotic viperins produce diverse antiviral molecules', *Nature,* 589(7840), pp. 120-124. doi: 10.1038/s41586-020-2762-2

Bernstein, E., Caudy, A. A., Hammond, S. M. and Hannon, G. J. (2001) 'Role for a bidentate ribonuclease in the initiation step of RNA interference', *Nature,* 409(6818), pp. 363-6. doi: 10.1038/35053110

Bickle, T. A. and Kruger, D. H. (1993) 'Biology of DNA restriction', *Microbiol Rev,* 57(2), pp. 434-50. doi: 10.1128/mr.57.2.434-450.1993

Blower, T. R., Pei, X. Y., Short, F. L., Fineran, P. C., Humphreys, D. P., Luisi, B. F. and Salmond, G. P. (2011) 'A processed noncoding RNA regulates an altruistic bacterial antiviral system', *Nat Struct Mol Biol,* 18(2), pp. 185-90. doi: 10.1038/nsmb.1981

Blower, T. R., Short, F. L., Rao, F., Mizuguchi, K., Pei, X. Y., Fineran, P. C., Luisi, B. F. and Salmond, G. P. (2012) 'Identification and classification of bacterial Type III toxin-antitoxin systems encoded in chromosomal and plasmid genomes', *Nucleic Acids Res,* 40(13), pp. 6158-73. doi: 10.1093/nar/gks231

Bobonis, J., Mitosch, K., Mateus, A., Karcher, N., Kritikos, G., Selkrig, J., Zietek, M., Monzon, V., Pfalz, B., Garcia-Santamarina, S., Galardini, M., Sueki, A., Kobayashi, C., Stein, F., Bateman, A., Zeller, G., Savitski, M. M., Elfenbein, J. R., Andrews-Polymenis, H. L. and Typas, A. (2022) 'Bacterial retrons encode phage-defending tripartite toxin-antitoxin systems', *Nature,* 609(7925), pp. 144-150. doi: 10.1038/s41586-022-05091-4

Bogdan, M., Peric, L., Ordog, K., Vukovic, D., Urban, E. and Soki, J. (2018) 'The first characterized carbapenem-resistant Bacteroides fragilis strain from Croatia and the case study for it', *Acta Microbiol Immunol Hung,* 65(3), pp. 317-323. doi: 10.1556/030.65.2018.024

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

Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M. F., Hidalgo-Reyes, Y., Wiedenheft, B., Maxwell, K. L. and Davidson, A. R. (2015) 'Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins', *Nature,* 526(7571), pp. 136-9. doi: 10.1038/nature15254

Box, A. M., McGuffie, M. J., O'Hara, B. J. and Seed, K. D. (2016) 'Functional Analysis of Bacteriophage Immunity through a Type I-E CRISPR-Cas System in Vibrio cholerae and Its Application in Bacteriophage Genome Engineering', *J Bacteriol,* 198(3), pp. 578-90. doi: 10.1128/JB.00747-15

Bravo, J. P. K., Aparicio-Maldonado, C., Nobrega, F. L., Brouns, S. J. J. and Taylor, D. W. (2022) 'Structural basis for broad anti-phage immunity by DISARM', *Nat Commun,* 13(1), pp. 2987. doi: 10.1038/s41467-022-30673-1

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V. and van der Oost, J. (2008) 'Small CRISPR RNAs guide antiviral defense in prokaryotes', *Science,* 321(5891), pp. 960-4. doi: 10.1126/science.1159689

Brown, S., Gauvin, C. C., Charbonneau, A. A., Burman, N. and Lawrence, C. M. (2020) 'Csx3 is a cyclic oligonucleotide phosphodiesterase associated with type III

CRISPR-Cas that degrades the second messenger cA(4)', *Journal of Biological Chemistry,* 295(44), pp. 14963-14972. doi: 10.1074/jbc.RA120.014099

Brussow, H. and Hendrix, R. W. (2002) 'Phage genomics: small is beautiful', *Cell,* 108(1), pp. 13-6. doi: 10.1016/s0092-8674(01)00637-7

Bryson, A. L., Hwang, Y., Sherrill-Mix, S., Wu, G. D., Lewis, J. D., Black, L., Clark, T. A. and Bushman, F. D. (2015) 'Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9', *mBio,* 6(3), pp. e00648. doi: 10.1128/mBio.00648-15

Burroughs, A. M., Zhang, D., Schaffer, D. E., Iyer, L. M. and Aravind, L. (2015) 'Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling', *Nucleic Acids Res,* 43(22), pp. 10633-54. doi: 10.1093/nar/gkv1267

Cady, K. C., Bondy-Denomy, J., Heussler, G. E., Davidson, A. R. and O'Toole, G. A. (2012) 'The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages', *J Bacteriol,* 194(21), pp. 5728-38. doi: 10.1128/JB.01184-12

Campbell, E. A., Greenwell, R., Anthony, J. R., Wang, S., Lim, L., Das, K., Sofia, H. J., Donohue, T. J. and Darst, S. A. (2007) 'A conserved structural module regulates transcriptional responses to diverse stress signals in bacteria', *Mol Cell,* 27(5), pp. 793-805. doi: 10.1016/j.molcel.2007.07.009

Campbell, E. A., Tupy, J. L., Gruber, T. M., Wang, S., Sharp, M. M., Gross, C. A. and Darst, S. A. (2003) 'Crystal structure of Escherichia coli sigmaE with the cytoplasmic domain of its anti-sigma RseA', *Mol Cell,* 11(4), pp. 1067-78. doi: 10.1016/s1097-2765(03)00148-5

Campbell, E. A., Westblade, L. F. and Darst, S. A. (2008) 'Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective', *Curr Opin Microbiol,* 11(2), pp. 121-7. doi: 10.1016/j.mib.2008.02.016

Cappadocia, L. and Lima, C. D. (2018) 'Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism', *Chem Rev,* 118(3), pp. 889-918. doi: 10.1021/acs.chemrev.6b00737

Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M. and Terns, M. P. (2010) 'Binding and cleavage of CRISPR RNA by Cas6', *RNA,* 16(11), pp. 2181-8. doi: 10.1261/rna.2230110

Chaikeeratisak, V., Nguyen, K., Khanna, K., Brilot, A. F., Erb, M. L., Coker, J. K., Vavilina, A., Newton, G. L., Buschauer, R., Pogliano, K., Villa, E., Agard, D. A. and Pogliano, J. (2017) 'Assembly of a nucleus-like structure during viral replication in bacteria', *Science,* 355(6321), pp. 194-197. doi: 10.1126/science.aal2130

Chang, J. J., You, B. J., Tien, N., Wang, Y. C., Yang, C. S., Hou, M. H. and Chen, Y. (2023) 'Specific recognition of cyclic oligonucleotides by Cap4 for phage infection', *Int J Biol Macromol,* 237, pp. 123656. doi: 10.1016/j.ijbiomac.2023.123656

Charpentier, E., van der Oost, J., White, M.F. (2013) *crRNA Biogenesis. CRISPR-Cas Systems*: Springer, Berlin, Heidelberg.

Chatterjee, A. and Duerkop, B. A. (2018) 'Beyond Bacteria: Bacteriophage-Eukaryotic Host Interactions Reveal Emerging Paradigms of Health and Disease', *Front Microbiol,* 9, pp. 1394. doi: 10.3389/fmicb.2018.01394

Chen, C., Wang, L., Chen, S., Wu, X., Gu, M., Chen, X., Jiang, S., Wang, Y., Deng, Z., Dedon, P. C. and Chen, S. (2017) 'Convergence of DNA methylation and phosphorothioation epigenetics in bacterial genomes', *Proc Natl Acad Sci U S A,* 114(17), pp. 4501-4506. doi: 10.1073/pnas.1702450114

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) 'MolProbity: all-atom structure validation for macromolecular crystallography', *Acta Crystallogr D Biol Crystallogr,* 66(Pt 1), pp. 12-21. doi: 10.1107/S0907444909042073

Chou-Zheng, L. and Hatoum-Aslan, A. (2019) 'A type III-A CRISPR-Cas system employs degradosome nucleases to ensure robust immunity', *Elife,* 8. doi: 10.7554/eLife.45393

Chou-Zheng, L. and Hatoum-Aslan, A. (2022) 'Critical roles for 'housekeeping' nucleases in type III CRISPR-Cas immunity', *Elife,* 11. doi: 10.7554/eLife.81897

Chowdhury, S., Carter, J., Rollins, M. F., Golden, S. M., Jackson, R. N., Hoffmann, C., Nosaka, L., Bondy-Denomy, J., Maxwell, K. L., Davidson, A. R., Fischer, E. R., Lander, G. C. and Wiedenheft, B. (2017) 'Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex', *Cell,* 169(1), pp. 47-57 e11. doi: 10.1016/j.cell.2017.03.012

Clokie, M. R., Millard, A. D., Letarov, A. V. and Heaphy, S. (2011) 'Phages in nature', *Bacteriophage,* 1(1), pp. 31-45. doi: 10.4161/bact.1.1.14942

Cohen, D. (2019) 'Cyclic GMP–AMP signalling protects bacteria against viral infection', *Nature,* 574. doi: 10.1038/s41586-019-1605-5

Cui, L., Balamkundu, S., Liu, C. F., Ye, H., Hourihan, J., Rausch, A., Hauss, C., Nilsson, E., Hoetzinger, M., Holmfeldt, K., Zhang, W., Martinez-Alvarez, L., Peng, X., Tremblay, D., Moinau, S., Solonenko, N., Sullivan, M. B., Lee, Y. J., Mulholland, A., Weigele, P. R., de Crecy-Lagard, V., Dedon, P. C. and Hutinet, G. (2023) 'Four additional natural 7-deazaguanine derivatives in phages and how to make them', *Nucleic Acids Res,* 51(17), pp. 9214-9226. doi: 10.1093/nar/gkad657

D'Herelle, F. (2007) 'On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux. 1917', *Res Microbiol,* 158(7), pp. 553-4. doi: 10.1016/j.resmic.2007.07.005

Dalmas, O., Sandtner, W., Medovoy, D., Frezza, L., Bezanilla, F. and Perozo, E. (2014) 'A repulsion mechanism explains magnesium permeation and selectivity in CorA', *Proc Natl Acad Sci U S A,* 111(8), pp. 3002-7. doi: 10.1073/pnas.1319054111

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J. and Charpentier, E. (2011) 'CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III', *Nature,* 471(7340), pp. 602-7. doi: 10.1038/nature09886

Deng, L., Garrett, R. A., Shah, S. A., Peng, X. and She, Q. (2013) 'A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus', *Mol Microbiol,* 87(5), pp. 1088-99. doi: 10.1111/mmi.12152

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

Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P. and Moineau, S. (2008) 'Phage response to CRISPR-encoded resistance in Streptococcus thermophilus', *J Bacteriol,* 190(4), pp. 1390-400. doi: 10.1128/JB.01412-07

Devkota, S. R., Kwon, E., Ha, S. C., Chang, H. W. and Kim, D. Y. (2017) 'Structural insights into the regulation of Bacillus subtilis SigW activity by anti-sigma RsiW', *PLoS One,* 12(3), pp. e0174284. doi: 10.1371/journal.pone.0174284

Dhingra, Y. and Sashital, D. G. (2023) 'Cas4/1 dual nuclease activities enable prespacer maturation and directional integration in a type I-G CRISPR-Cas system', *J Biol Chem,* 299(9), pp. 105178. doi: 10.1016/j.jbc.2023.105178

Dhingra, Y., Suresh, S. K., Juneja, P. and Sashital, D. G. (2022) 'PAM binding ensures orientational integration during Cas4-Cas1-Cas2-mediated CRISPR adaptation', *Mol Cell,* 82(22), pp. 4353-4367 e6. doi: 10.1016/j.molcel.2022.09.030

Dillingham, M. S. and Kowalczykowski, S. C. (2008) 'RecBCD enzyme and the repair of double-stranded DNA breaks', *Microbiol Mol Biol Rev,* 72(4), pp. 642-71, Table of Contents. doi: 10.1128/MMBR.00020-08

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

Dryden, D. T., Murray, N. E. and Rao, D. N. (2001) 'Nucleoside triphosphate-dependent restriction enzymes', *Nucleic Acids Res,* 29(18), pp. 3728-41. doi: 10.1093/nar/29.18.3728

Duncan-Lowey, B., McNamara-Bordewick, N. K., Tal, N., Sorek, R. and Kranzusch, P. J. (2021) 'Effector-mediated membrane disruption controls cell death in CBASS antiphage defense', *Mol Cell,* 81(24), pp. 5039-5051 e5. doi: 10.1016/j.molcel.2021.10.020

Durmaz, E. and Klaenhammer, T. R. (2007) 'Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected Lactococcus lactis', *J Bacteriol,* 189(4), pp. 1417-25. doi: 10.1128/JB.00904-06

Dy, R. L., Richter, C., Salmond, G. P. and Fineran, P. C. (2014) 'Remarkable Mechanisms in Microbes to Resist Phage Infections', *Annu Rev Virol,* 1(1), pp. 307-31. doi: 10.1146/annurev-virology-031413-085500

East-Seletsky, A., O'Connell, M. R., Burstein, D., Knott, G. J. and Doudna, J. A. (2017) 'RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes', *Mol Cell,* 66(3), pp. 373-383 e3. doi: 10.1016/j.molcel.2017.04.008

Eberhard, W. G. (1990) 'Evolution in bacterial plasmids and levels of selection', *Q Rev Biol,* 65(1), pp. 3-22. doi: 10.1086/416582

Eckstein, F. (2007) 'Phosphorothioation of DNA in bacteria', *Nat Chem Biol,* 3(11), pp. 689-90. doi: 10.1038/nchembio1107-689

El Haddad, L., Mendoza, J. F. and Jobin, C. (2022) 'Bacteriophage-mediated manipulations of microbiota in gastrointestinal diseases', *Front Microbiol,* 13, pp. 1055427. doi: 10.3389/fmicb.2022.1055427

Elmore, J. R., Sheppard, N. F., Ramia, N., Deighan, T., Li, H., Terns, R. M. and Terns, M. P. (2016) 'Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system', *Genes Dev,* 30(4), pp. 447-59. doi: 10.1101/gad.272153.115

Eshaghi, S., Niegowski, D., Kohl, A., Martinez Molina, D., Lesley, S. A. and Nordlund, P. (2006) 'Crystal structure of a divalent metal ion transporter CorA at 2.9 angstrom resolution', *Science,* 313(5785), pp. 354-7. doi: 10.1126/science.1127121

Estrella, M. A., Kuo, F. T. and Bailey, S. (2016) 'RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex', *Genes Dev,* 30(4), pp. 460-70. doi: 10.1101/gad.273722.115

Fagerlund, R. D., Wilkinson, M. E., Klykov, O., Barendregt, A., Pearce, F. G., Kieper, S. N., Maxwell, H. W. R., Capolupo, A., Heck, A. J. R., Krause, K. L., Bostina, M., Scheltema, R. A., Staals, R. H. J. and 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), pp. E5122-E5128. doi: 10.1073/pnas.1618421114

Fang, X. and Qi, Y. (2016) 'RNAi in Plants: An Argonaute-Centered View', *Plant Cell,* 28(2), pp. 272-85. doi: 10.1105/tpc.15.00920

Fatma, S., Chakravarti, A., Zeng, X. and Huang, R. H. (2021) 'Molecular mechanisms of the CdnG-Cap5 antiphage defense system employing 3',2'-cGAMP as the second messenger', *Nat Commun,* 12(1), pp. 6381. doi: 10.1038/s41467-021-26738-2

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

Frost, L. S., Leplae, R., Summers, A. O. and Toussaint, A. (2005) 'Mobile genetic elements: the agents of open source evolution', *Nat Rev Microbiol,* 3(9), pp. 722-32. doi: 10.1038/nrmicro1235

Fu, L., Xie, C., Jin, Z., Tu, Z., Han, L., Jin, M., Xiang, Y. and Zhang, A. (2019) 'The prokaryotic Argonaute proteins enhance homology sequence-directed recombination in bacteria', *Nucleic Acids Res,* 47(7), pp. 3568-3579. doi: 10.1093/nar/gkz040

Fuhrman, J. A. (1999) 'Marine viruses and their biogeochemical and ecological effects', *Nature,* 399(6736), pp. 541-8. doi: 10.1038/21119

Gan, R., Wu, X., He, W., Liu, Z., Wu, S., Chen, C., Chen, S., Xiang, Q., Deng, Z., Liang, D., Chen, S. and Wang, L. (2014) 'DNA phosphorothioate modifications influence the global transcriptional response and protect DNA from double-stranded breaks', *Sci Rep,* 4, pp. 6642. doi: 10.1038/srep06642

Gao, L., Altae-Tran, H., Bohning, F., Makarova, K. S., Segel, M., Schmid-Burgk, J. L., Koob, J., Wolf, Y. I., Koonin, E. V. and Zhang, F. (2020) 'Diverse enzymatic activities mediate antiviral immunity in prokaryotes', *Science,* 369(6507), pp. 1077-1084. doi: 10.1126/science.aba0372

Gao, L. A., Wilkinson, M. E., Strecker, J., Makarova, K. S., Macrae, R. K., Koonin, E. V. and Zhang, F. (2022) 'Prokaryotic innate immunity through pattern recognition of conserved viral proteins', *Science,* 377(6607), pp. eabm4096. doi: 10.1126/science.abm4096

Garb, J., Lopatina, A., Bernheim, A., Zaremba, M., Siksnys, V., Melamed, S., Leavitt, A., Millman, A., Amitai, G. and Sorek, R. (2022) 'Multiple phage resistance systems inhibit infection via SIR2-dependent NAD(+) depletion', *Nat Microbiol,* 7(11), pp. 1849-1856. doi: 10.1038/s41564-022-01207-8

Garcia-Doval, C., Schwede, F., Berk, C., Rostol, J. T., Niewoehner, O., Tejero, O., Hall, J., Marraffini, L. A. and Jinek, M. (2020) 'Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6', *Nature Communications,* 11(1). doi: 10.1038/s41467-020-15334-5

Garcia-Rodriguez, G., Talavera Perez, A., Konijnenberg, A., Sobott, F., Michiels, J. and Loris, R. (2020) 'The Escherichia coli RnlA-RnlB toxin-antitoxin complex: production, characterization and crystallization', *Acta Crystallogr F Struct Biol Commun,* 76(Pt 1), pp. 31-39. doi: 10.1107/S2053230X19017175

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

Georjon, H. and Bernheim, A. (2023) 'The highly diverse antiphage defence systems of bacteria', *Nat Rev Microbiol*. doi: 10.1038/s41579-023-00934-x

Goldfarb, T. (2015) 'BREX is a novel phage resistance system widespread in microbial genomes', *EMBO J.,* 34. doi: 10.15252/embj.201489455

Gonzalez-Delgado, A., Mestre, M. R., Martinez-Abarca, F. and 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), pp. 10202-10211. doi: 10.1093/nar/gkz746

Gruschow, S., Adamson, C. S. and White, M. F. (2021) 'Specificity and sensitivity of an RNA targeting type III CRISPR complex coupled with a NucC endonuclease effector', *Nucleic Acids Res,* 49(22), pp. 13122-13134. doi: 10.1093/nar/gkab1190

Gruschow, S., Athukoralage, J. S., Graham, S., Hoogeboom, T. and White, M. F. (2019) 'Cyclic oligoadenylate signalling mediates Mycobacterium tuberculosis CRISPR defence', *Nucleic Acids Research,* 47(17), pp. 9259-9270. doi: 10.1093/nar/gkz676

Guegler, C. K. and Laub, M. T. (2021) 'Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection', *Mol Cell,* 81(11), pp. 2361-2373 e9. doi: 10.1016/j.molcel.2021.03.027

Guo, M., Zhang, K., Zhu, Y., Pintilie, G. D., Guan, X., Li, S., Schmid, M. F., Ma, Z., Chiu, W. and Huang, Z. (2019) 'Coupling of ssRNA cleavage with DNase activity in type III-A CRISPR-Csm revealed by cryo-EM and biochemistry', *Cell Res,* 29(4), pp. 305-312. doi: 10.1038/s41422-019-0151-x

Guo, T. W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., Eng, E. T., Raczkowski, A. M., Fox, T., Earl, L. A., Patel, D. J. and Subramaniam, S. (2017) 'Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex', *Cell,* 171(2), pp. 414-426 e12. doi: 10.1016/j.cell.2017.09.006

Guo, X., Soderholm, A., Kanchugal, P. S., Isaksen, G. V., Warsi, O., Eckhard, U., Triguis, S., Gogoll, A., Jerlstrom-Hultqvist, J., Aqvist, J., Andersson, D. I. and Selmer, M. (2021) 'Structure and mechanism of a phage-encoded SAM lyase revises catalytic function of enzyme family', *Elife,* 10. doi: 10.7554/eLife.61818

Guskov, A., Nordin, N., Reynaud, A., Engman, H., Lundback, A. K., Jong, A. J., Cornvik, T., Phua, T. and Eshaghi, S. (2012) 'Structural insights into the mechanisms of Mg2+ uptake, transport, and gating by CorA', *Proc Natl Acad Sci U S A,* 109(45), pp. 18459-64. doi: 10.1073/pnas.1210076109

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M. and Terns, M. P. (2009) 'RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex', *Cell,* 139(5), pp. 945-56. doi: 10.1016/j.cell.2009.07.040

Halliday, N. M., Hardie, K. R., Williams, P., Winzer, K. and Barrett, D. A. (2010) 'Quantitative liquid chromatography-tandem mass spectrometry profiling of activated methyl cycle metabolites involved in LuxS-dependent quorum sensing in Escherichia coli', *Anal Biochem,* 403(1-2), pp. 20-9. doi: 10.1016/j.ab.2010.04.021

Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. and Hannon, G. J. (2001) 'Argonaute2, a link between genetic and biochemical analyses of RNAi', *Science,* 293(5532), pp. 1146-50. doi: 10.1126/science.1064023

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

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

Hatoum-Aslan, A., Maniv, I. and Marraffini, L. A. (2011) 'Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site', *Proc Natl Acad Sci U S A,* 108(52), pp. 21218-22. doi: 10.1073/pnas.1112832108

Hatoum-Aslan, A., Maniv, I., Samai, P. and Marraffini, L. A. (2014) 'Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system', *J Bacteriol,* 196(2), pp. 310-7. doi: 10.1128/JB.01130-13

Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W. and Marraffini, L. A. (2013) 'A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs', *J Biol Chem,* 288(39), pp. 27888-97. doi: 10.1074/jbc.M113.499244

Hegge, J. W., Swarts, D. C. and van der Oost, J. (2018) 'Prokaryotic Argonaute proteins: novel genome-editing tools?', *Nat Rev Microbiol,* 16(1), pp. 5-11. doi: 10.1038/nrmicro.2017.73

Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D. and Marraffini, L. A. (2015) 'Cas9 specifies functional viral targets during CRISPR-Cas adaptation', *Nature,* 519(7542), pp. 199-202. doi: 10.1038/nature14245

Heo, K., Lee, J. W., Jang, Y., Kwon, S., Lee, J., Seok, C., Ha, N. C. and Seok, Y. J. (2022) 'A pGpG-specific phosphodiesterase regulates cyclic di-GMP signaling in Vibrio cholerae', *J Biol Chem,* 298(3), pp. 101626. doi: 10.1016/j.jbc.2022.101626

Hille, F., Richter, H., Wong, S. P., Bratovic, M., Ressel, S. and Charpentier, E. (2018) 'The Biology of CRISPR-Cas: Backward and Forward', *Cell,* 172(6), pp. 1239-1259. doi: 10.1016/j.cell.2017.11.032

Hochstrasser, M. L. and Doudna, J. A. (2015) 'Cutting it close: CRISPR-associated endoribonuclease structure and function', *Trends Biochem Sci,* 40(1), pp. 58-66. doi: 10.1016/j.tibs.2014.10.007

Hogrel, G., Guild, A., Graham, S., Rickman, H., Gruschow, S., Bertrand, Q., Spagnolo, L. and White, M. F. (2022) 'Cyclic nucleotide-induced helical structure activates a TIR immune effector', *Nature,* 608(7924), pp. 808-812. doi: 10.1038/s41586-022-05070-9

Hör, J., Wolf, S. G. and Sorek, R. (2023) 'Bacteria conjugate ubiquitin-like proteins to interfere with phage assembly', *bioRxiv*. doi: 10.1101/2023.09.04.556158

Hoskisson, P. A., Sumby, P. and Smith, M. C. M. (2015) 'The phage growth limitation system in Streptomyces coelicolor A(3)2 is a toxin/antitoxin system, comprising enzymes with DNA methyltransferase, protein kinase and ATPase activity', *Virology,* 477, pp. 100-109. doi: 10.1016/j.virol.2014.12.036

Hsiao, Y. Y., Duh, Y., Chen, Y. P., Wang, Y. T. and Yuan, H. S. (2012) 'How an exonuclease decides where to stop in trimming of nucleic acids: crystal structures of RNase T-product complexes', *Nucleic Acids Res,* 40(16), pp. 8144-54. doi: 10.1093/nar/gks548

Hsueh, B. Y., Severin, G. B., Elg, C. A., Waldron, E. J., Kant, A., Wessel, A. J., Dover, J. A., Rhoades, C. R., Ridenhour, B. J., Parent, K. N., Neiditch, M. B., Ravi, J., Top, E. M. and Waters, C. M. (2022) 'Phage defence by deaminase-mediated depletion of deoxynucleotides in bacteria', *Nat Microbiol,* 7(8), pp. 1210-1220. doi: 10.1038/s41564-022-01162-4

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

Isaev, A., Drobiazko, A., Sierro, N., Gordeeva, J., Yosef, I., Qimron, U., Ivanov, N. V. and Severinov, K. (2020) 'Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence', *Nucleic Acids Res,* 48(10), pp. 5397-5406. doi: 10.1093/nar/gkaa290

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) 'Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product', *J Bacteriol,* 169(12), pp. 5429-33. doi: 10.1128/jb.169.12.5429-5433.1987

Ivancic-Bace, I., Cass, S. D., Wearne, S. J. and Bolt, E. L. (2015) 'Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity', *Nucleic Acids Res,* 43(22), pp. 10821-30. doi: 10.1093/nar/gkv1213

Janscak, P., Sandmeier, U., Szczelkun, M. D. and Bickle, T. A. (2001) 'Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I', *J Mol Biol,* 306(3), pp. 417-31. doi: 10.1006/jmbi.2000.4411

Jansen, R., Embden, J. D., Gaastra, W. and Schouls, L. M. (2002) 'Identification of genes that are associated with DNA repeats in prokaryotes', *Mol Microbiol,* 43(6), pp. 1565-75. doi: 10.1046/j.1365-2958.2002.02839.x

Jenny Y. Zhang, S. G., Joseph Bondy-Denomy (2022) *Evasion Tactics Manifested by Bacteriophages against Bacterial Immunity. In Crispr.*

Jia, N., Jones, R., Sukenick, G. and Patel, D. J. (2019a) 'Second Messenger cA(4) Formation within the Composite Csm1 Palm Pocket of Type III-A CRISPR-Cas Csm Complex and Its Release Path', *Mol Cell,* 75(5), pp. 933-943 e6. doi: 10.1016/j.molcel.2019.06.013

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

Jia, N., Mo, C. Y., Wang, C., Eng, E. T., Marraffini, L. A. and Patel, D. J. (2019c) 'Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity', *Mol Cell,* 73(2), pp. 264-277 e5. doi: 10.1016/j.molcel.2018.11.007

Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., Nogales, E. and Doudna, J. A. (2016a) 'Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage', *Science,* 351(6275), pp. 867-71. doi: 10.1126/science.aad8282

Jiang, W., Samai, P. and Marraffini, L. A. (2016b) 'Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity', *Cell,* 164(4), pp. 710-21. doi: 10.1016/j.cell.2015.12.053

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science,* 337(6096), pp. 816-21. doi: 10.1126/science.1225829

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E. and Doudna, J. A. (2014) 'Structures of Cas9 endonucleases

reveal RNA-mediated conformational activation', *Science,* 343(6176), pp. 1247997. doi: 10.1126/science.1247997

Johnson, A. G., Wein, T., Mayer, M. L., Duncan-Lowey, B., Yirmiya, E., Oppenheimer-Shaanan, Y., Amitai, G., Sorek, R. and Kranzusch, P. J. (2022) 'Bacterial gasdermins reveal an ancient mechanism of cell death', *Science,* 375(6577), pp. 221-225. doi: 10.1126/science.abj8432

Jolly, S. M., Gainetdinov, I., Jouravleva, K., Zhang, H., Strittmatter, L., Bailey, S. M., Hendricks, G. M., Dhabaria, A., Ueberheide, B. and Zamore, P. D. (2020) 'Thermus thermophilus Argonaute Functions in the Completion of DNA Replication', *Cell,* 182(6), pp. 1545-1559 e18. doi: 10.1016/j.cell.2020.07.036

Jones, J. D., Vance, R. E. and Dangl, J. L. (2016) 'Intracellular innate immune surveillance devices in plants and animals', *Science,* 354(6316). doi: 10.1126/science.aaf6395

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. and Hassabis, D. (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature,* 596(7873), pp. 583-589. doi: 10.1038/s41586-021-03819-2

Kaufmann, G. (2000) 'Anticodon nucleases', *Trends Biochem Sci,* 25(2), pp. 70-4. doi: 10.1016/s0968-0004(99)01525-x

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

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

Kever, L., Hardy, A., Luthe, T., Hunnefeld, M., Gatgens, C., Milke, L., Wiechert, J., Wittmann, J., Moraru, C., Marienhagen, J. and Frunzke, J. (2022) 'Aminoglycoside Antibiotics Inhibit Phage Infection by Blocking an Early Step of the Infection Cycle', *mBio,* 13(3), pp. e0078322. doi: 10.1128/mbio.00783-22

Koonin, E. V. and Aravind, L. (2002) 'Origin and evolution of eukaryotic apoptosis: the bacterial connection', *Cell Death Differ,* 9(4), pp. 394-404. doi: 10.1038/sj.cdd.4400991

Koonin, E. V. and Makarova, K. S. (2019) 'Origins and evolution of CRISPR-Cas systems', *Philos Trans R Soc Lond B Biol Sci,* 374(1772), pp. 20180087. doi: 10.1098/rstb.2018.0087

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

Koopal, B., Potocnik, A., Mutte, S. K., Aparicio-Maldonado, C., Lindhoud, S., Vervoort, J. J. M., Brouns, S. J. J. and Swarts, D. C. (2022) 'Short prokaryotic Argonaute systems trigger cell death upon detection of invading DNA', *Cell,* 185(9), pp. 1471-1486 e19. doi: 10.1016/j.cell.2022.03.012

Kovall, R. A. and Matthews, B. W. (1999) 'Type II restriction endonucleases: structural, functional and evolutionary relationships', *Curr Opin Chem Biol,* 3(5), pp. 578-83. doi: 10.1016/s1367-5931(99)00012-5

Krissinel, E. and Henrick, K. (2007) 'Inference of macromolecular assemblies from crystalline state', *J Mol Biol,* 372(3), pp. 774-97. doi: 10.1016/j.jmb.2007.05.022

Kronheim, S., Daniel-Ivad, M., Duan, Z., Hwang, S., Wong, A. I., Mantel, I., Nodwell, J. R. and Maxwell, K. L. (2018) 'A chemical defence against phage infection', *Nature,* 564(7735), pp. 283-286. doi: 10.1038/s41586-018-0767-x

Kropocheva, E., Kuzmenko, A., Aravin, A. A., Esyunina, D. and Kulbachinskiy, A. (2021) 'A programmable pAgo nuclease with universal guide and target specificity from the mesophilic bacterium Kurthia massiliensis', *Nucleic Acids Res,* 49(7), pp. 4054-4065. doi: 10.1093/nar/gkab182

Kruger, D. H., Barcak, G. J., Reuter, M. and Smith, H. O. (1988) 'EcoRII can be activated to cleave refractory DNA recognition sites', *Nucleic Acids Res,* 16(9), pp. 3997-4008. doi: 10.1093/nar/16.9.3997

Kruger, D. H. and Bickle, T. A. (1983) 'Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts', *Microbiol Rev,* 47(3), pp. 345-60. doi: 10.1128/mr.47.3.345-360.1983

Krüger, L., Gaskell-Mew, L., Graham, S., Shirran, S., Hertel, R. and White, M. F. (2023) 'Reversible conjugation of a CBASS nucleotide cyclase regulates immune response to phage infection', *bioRxiv*. doi: 10.1101/2023.09.29.560134

Kuhn, C. D. and Joshua-Tor, L. (2013) 'Eukaryotic Argonautes come into focus', *Trends Biochem Sci,* 38(5), pp. 263-71. doi: 10.1016/j.tibs.2013.02.008

Kuzmenko, A., Oguienko, A., Esyunina, D., Yudin, D., Petrova, M., Kudinova, A., Maslova, O., Ninova, M., Ryazansky, S., Leach, D., Aravin, A. A. and Kulbachinskiy, A. (2020) 'DNA targeting and interference by a bacterial Argonaute nuclease', *Nature,* 587(7835), pp. 632-637. doi: 10.1038/s41586-020-2605-1

Lampson, B. C., Inouye, M. and Inouye, S. (2005) 'Retrons, msDNA, and the bacterial genome', *Cytogenet Genome Res,* 110(1-4), pp. 491-9. doi: 10.1159/000084982

Lau, R. K. (2020) 'Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity', *Mol. Cell,* 77. doi: 10.1016/j.molcel.2019.12.010

Ledvina, H. E., Ye, Q., Gu, Y., Sullivan, A. E., Quan, Y., Lau, R. K., Zhou, H., Corbett, K. D. and Whiteley, A. T. (2023) 'An E1-E2 fusion protein primes antiviral immune signalling in bacteria', *Nature,* 616(7956), pp. 319-325. doi: 10.1038/s41586-022-05647-4

Lee, C. W., Park, S. H., Jeong, C. S., Cha, S. S., Park, H. and Lee, J. H. (2019) 'Structural basis of small RNA hydrolysis by oligoribonuclease (CpsORN) from Colwellia psychrerythraea strain 34H', *Sci Rep,* 9(1), pp. 2649. doi: 10.1038/s41598-019-39641-0

Lee, J. H., Wood, J. M., Almo, S. C., Evans, G. B., Harris, L. D. and Grove, T. L. (2023) 'Chemoenzymatic Synthesis of 3'-Deoxy-3',4'-didehydro-cytidine triphosphate (ddhCTP)', *ACS Bio Med Chem Au,* 3(4), pp. 322-326. doi: 10.1021/acsbiomedchemau.3c00014

Lee, K. Z., Mechikoff, M. A., Kikla, A., Liu, A., Pandolfi, P., Fitzgerald, K., Gimble, F. S. and Solomon, K. V. (2021) 'NgAgo possesses guided DNA nicking activity', *Nucleic Acids Res,* 49(17), pp. 9926-9937. doi: 10.1093/nar/gkab757

Leipe, D. D., Koonin, E. V. and Aravind, L. (2004) 'STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by

horizontal gene transfer', *J Mol Biol,* 343(1), pp. 1-28. doi: 10.1016/j.jmb.2004.08.023

Lerche, M., Sandhu, H., Flockner, L., Hogbom, M. and Rapp, M. (2017) 'Structure and Cooperativity of the Cytosolic Domain of the CorA Mg(2+) Channel from Escherichia coli', *Structure,* 25(8), pp. 1175-1186 e4. doi: 10.1016/j.str.2017.05.024

LeRoux, M., Srikant, S., Teodoro, G. I. C., Zhang, T., Littlehale, M. L., Doron, S., Badiee, M., Leung, A. K. L., Sorek, R. and Laub, M. T. (2022) 'The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA', *Nat Microbiol,* 7(7), pp. 1028-1040. doi: 10.1038/s41564-022-01153-5

Leroy, M., Piton, J., Gilet, L., Pellegrini, O., Proux, C., Coppée, J. Y., Figaro, S. and Condon, C. (2017) 'Rae1/YacP, a new endoribonuclease involved in ribosome-dependent mRNA decay in Bacillus subtilis', *Embo j,* 36(9), pp. 1167-1181. doi: 10.15252/embj.201796540

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

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

Liu, J. J., Orlova, N., Oakes, B. L., Ma, E., Spinner, H. B., Baney, K. L. M., Chuck, J., Tan, D., Knott, G. J., Harrington, L. B., Al-Shayeb, B., Wagner, A., Brotzmann, J., Staahl, B. T., Taylor, K. L., Desmarais, J., Nogales, E. and Doudna, J. A. (2019) 'CasX enzymes comprise a distinct family of RNA-guided genome editors', *Nature,* 566(7743), pp. 218-223. doi: 10.1038/s41586-019-0908-x

Lobocka, M. B., Rose, D. J., Plunkett, G., 3rd, Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M. B. and Blattner, F. R. (2004) 'Genome of bacteriophage P1', *J Bacteriol,* 186(21), pp. 7032-68. doi: 10.1128/JB.186.21.7032-7068.2004

Loeff, L., Brouns, S. J. J. and Joo, C. (2018) 'Repetitive DNA Reeling by the Cascade-Cas3 Complex in Nucleotide Unwinding Steps', *Mol Cell,* 70(3), pp. 385-394 e3. doi: 10.1016/j.molcel.2018.03.031

Loenen, W. A., Dryden, D. T., Raleigh, E. A., Wilson, G. G. and Murray, N. E. (2014) 'Highlights of the DNA cutters: a short history of the restriction enzymes', *Nucleic Acids Res,* 42(1), pp. 3-19. doi: 10.1093/nar/gkt990

Lopatina, A., Tal, N. and Sorek, R. (2020) 'Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy', *Annu Rev Virol,* 7(1), pp. 371-384. doi: 10.1146/annurev-virology-011620-040628

Lowey, B., Whiteley, A. T., Keszei, A. F. A., Morehouse, B. R., Mathews, I. T., Antine, S. P., Cabrera, V. J., Kashin, D., Niemann, P., Jain, M., Schwede, F., Mekalanos, J. J., Shao, S., Lee, A. S. Y. and Kranzusch, P. J. (2020) 'CBASS Immunity Uses CARF-Related Effectors to Sense 3'-5'- and 2'-5'-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection', *Cell,* 182(1), pp. 38-49.e17. doi: 10.1016/j.cell.2020.05.019

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D. and Lopez, R. (2019) 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic Acids Res,* 47(W1), pp. W636-W641. doi: 10.1093/nar/gkz268

Makarova, K. S. (2015) 'An updated evolutionary classification of CRISPR–Cas systems', *Nat. Rev. Microbiol.,* 13. doi: 10.1038/nrmicro3569

Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V. and Aravind, L. (2014) 'CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems', *Front Genet,* 5, pp. 102. doi: 10.3389/fgene.2014.00102

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F., van der Oost, J. and Koonin, E. V. (2011a) 'Evolution and classification of the CRISPR-Cas systems', *Nat Rev Microbiol,* 9(6), pp. 467-77. doi: 10.1038/nrmicro2577

Makarova, K. S., Timinskas, A., Wolf, Y. I., Gussow, A. B., Siksnys, V., Venclovas, C. and Koonin, E. V. (2020a) 'Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense', *Nucleic Acids Res,* 48(16), pp. 8828-8847. doi: 10.1093/nar/gkaa635

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

Makarova, K. S., Wolf, Y. I. and Koonin, E. V. (2013) 'The basic building blocks and evolution of CRISPR-CAS systems', *Biochem Soc Trans,* 41(6), pp. 1392-400. doi: 10.1042/BST20130038

Makarova, K. S., Wolf, Y. I., Snir, S. and Koonin, E. V. (2011b) 'Defense islands in bacterial and archaeal genomes and prediction of novel defense systems', *J Bacteriol,* 193(21), pp. 6039-56. doi: 10.1128/JB.05535-11

Malone, L. M., Warring, S. L., Jackson, S. A., Warnecke, C., Gardner, P. P., Gumy, L. F. and 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), pp. 48-55. doi: 10.1038/s41564-019-0612-5

Marraffini, L. A. (2022) *Mechanism of Type III CRISPR - Cas Immunity. In Crispr.*

Martin, R., Miquel, S., Ulmer, J., Langella, P. and Bermudez-Humaran, L. G. (2014) 'Gut ecosystem: how microbes help us', *Benef Microbes,* 5(3), pp. 219-33. doi: 10.3920/BM2013.0057

Matthies, D., Dalmas, O., Borgnia, M. J., Dominik, P. K., Merk, A., Rao, P., Reddy, B. G., Islam, S., Bartesaghi, A., Perozo, E. and Subramaniam, S. (2016) 'Cryo-EM Structures of the Magnesium Channel CorA Reveal Symmetry Break upon Gating', *Cell,* 164(4), pp. 747-56. doi: 10.1016/j.cell.2015.12.055

Mayo-Munoz, D., Smith, L. M., Garcia-Doval, C., Malone, L. M., Harding, K. R., Jackson, S. A., Hampton, H. G., Fagerlund, R. D., Gumy, L. F. and Fineran, P. C. (2022) 'Type III CRISPR-Cas provides resistance against nucleus-forming jumbo phages via abortive infection', *Mol Cell,* 82(23), pp. 4471-4486 e9. doi: 10.1016/j.molcel.2022.10.028

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

McGinn, J. and Marraffini, L. A. (2019) 'Molecular mechanisms of CRISPR-Cas spacer acquisition', *Nat Rev Microbiol,* 17(1), pp. 7-12. doi: 10.1038/s41579-018-0071-7

McMahon, S. A., Zhu, W., Graham, S., Rambo, R., White, M. F. and Gloster, T. M. (2020) 'Structure and mechanism of a Type III CRISPR defence DNA nuclease

activated by cyclic oligoadenylate', *Nat Commun,* 11(1), pp. 500. doi: 10.1038/s41467-019-14222-x

Meeske, A. J. and Marraffini, L. A. (2018) 'RNA Guide Complementarity Prevents Self-Targeting in Type VI CRISPR Systems', *Mol Cell,* 71(5), pp. 791-801 e3. doi: 10.1016/j.molcel.2018.07.013

Meeske, A. J., Nakandakari-Higa, S. and Marraffini, L. A. (2019) 'Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage', *Nature,* 570(7760), pp. 241-245. doi: 10.1038/s41586-019-1257-5

Millman, A., Bernheim, A., Stokar-Avihail, A., Fedorenko, T., Voichek, M., Leavitt, A., Oppenheimer-Shaanan, Y. and Sorek, R. (2020a) 'Bacterial Retrons Function In Anti-Phage Defense', *Cell,* 183(6), pp. 1551-1561 e12. doi: 10.1016/j.cell.2020.09.065

Millman, A., Melamed, S., Amitai, G. and Sorek, R. (2020b) 'Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems', *Nat Microbiol,* 5(12), pp. 1608-1615. doi: 10.1038/s41564-020-0777-y

Millman, A., Melamed, S., Leavitt, A., Doron, S., Bernheim, A., Hor, J., Garb, J., Bechon, N., Brandis, A., Lopatina, A., Ofir, G., Hochhauser, D., Stokar-Avihail, A., Tal, N., Sharir, S., Voichek, M., Erez, Z., Ferrer, J. L. M., Dar, D., Kacen, A., Amitai, G. and Sorek, R. (2022) 'An expanded arsenal of immune systems that protect bacteria from phages', *Cell Host Microbe,* 30(11), pp. 1556-1569 e5. doi: 10.1016/j.chom.2022.09.017

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

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

Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. and Soria, E. (2005) 'Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements', *J Mol Evol,* 60(2), pp. 174-82. doi: 10.1007/s00239-004-0046-3

Mojica, F. J., Diez-Villasenor, C., Soria, E. and Juez, G. (2000) 'Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria', *Mol Microbiol,* 36(1), pp. 244-6. doi: 10.1046/j.1365-2958.2000.01838.x

Mojica, F. J., Ferrer, C., Juez, G. and Rodriguez-Valera, F. (1995) 'Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning', *Mol Microbiol,* 17(1), pp. 85-93. doi: 10.1111/j.1365-2958.1995.mmi_17010085.x

Mojica, F. J., Juez, G. and Rodriguez-Valera, F. (1993) 'Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites', *Mol Microbiol,* 9(3), pp. 613-21. doi: 10.1111/j.1365-2958.1993.tb01721.x

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

Molina, R., Sofos, N. and Montoya, G. (2020) 'Structural basis of CRISPR-Cas Type III prokaryotic defence systems', *Curr Opin Struct Biol,* 65, pp. 119-129. doi: 10.1016/j.sbi.2020.06.010

Molina, R., Stella, S., Feng, M., Sofos, N., Jauniskis, V., Pozdnyakova, I., López-Méndez, B., She, Q. and Montoya, G. (2019) 'Structure of Csx1-cOA4 complex reveals the basis of RNA decay in Type III-B CRISPR-Cas', *Nature Communications,* 10(1), pp. 4302. doi: 10.1038/s41467-019-12244-z

Morehouse, B. R., Govande, A. A., Millman, A., Keszei, A. F. A., Lowey, B., Ofir, G., Shao, S., Sorek, R. and Kranzusch, P. J. (2020) 'STING cyclic dinucleotide sensing originated in bacteria', *Nature,* 586(7829), pp. 429-433. doi: 10.1038/s41586-020-2719-5

Morehouse, B. R., Yip, M. C. J., Keszei, A. F. A., McNamara-Bordewick, N. K., Shao, S. and Kranzusch, P. J. (2022) 'Cryo-EM structure of an active bacterial TIR-STING filament complex', *Nature,* 608(7924), pp. 803-807. doi: 10.1038/s41586-022-04999-1

Murphy, E. C., Morgelin, M., Cooney, J. C. and Frick, I. M. (2011) 'Interaction of Bacteroides fragilis and Bacteroides thetaiotaomicron with the kallikrein-kinin system', *Microbiology (Reading),* 157(Pt 7), pp. 2094-2105. doi: 10.1099/mic.0.046862-0

Murray, N. E. (2000) 'Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle)', *Microbiol Mol Biol Rev,* 64(2), pp. 412-34. doi: 10.1128/MMBR.64.2.412-434.2000

Nicholls, T. J., Spahr, H., Jiang, S., Siira, S. J., Koolmeister, C., Sharma, S., Kauppila, J. H. K., Jiang, M., Kaever, V., Rackham, O., Chabes, A., Falkenberg, M., Filipovska, A., Larsson, N. G. and Gustafsson, C. M. (2019) 'Dinucleotide Degradation by REXO2 Maintains Promoter Specificity in Mammalian Mitochondria', *Mol Cell,* 76(5), pp. 784-796 e6. doi: 10.1016/j.molcel.2019.09.010

Niewoehner, O. (2017) 'Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers', *Nature,* 548. doi: 10.1038/nature23467

Niewoehner, O. and Jinek, M. (2016) 'Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6', *RNA,* 22(3), pp. 318-29. doi: 10.1261/rna.054098.115

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F. and Nureki, O. (2014) 'Crystal structure of Cas9 in complex with guide RNA and target DNA', *Cell,* 156(5), pp. 935-49. doi: 10.1016/j.cell.2014.02.001

Nunez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. and Doudna, J. A. (2016) 'CRISPR Immunological Memory Requires a Host Factor for Specificity', *Mol Cell,* 62(6), pp. 824-833. doi: 10.1016/j.molcel.2016.04.027

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

Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S. and Sorek, R. (2018) 'DISARM is a widespread bacterial defence system with broad anti-phage activities', *Nat Microbiol,* 3(1), pp. 90-98. doi: 10.1038/s41564-017-0051-0

Olina, A. V., Kulbachinskiy, A. V., Aravin, A. A. and Esyunina, D. M. (2018) 'Argonaute Proteins and Mechanisms of RNA Interference in Eukaryotes and Prokaryotes', *Biochemistry (Mosc),* 83(5), pp. 483-497. doi: 10.1134/S0006297918050024

Oliveira, P. H., Touchon, M. and Rocha, E. P. (2014) 'The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts', *Nucleic Acids Res,* 42(16), pp. 10618-31. doi: 10.1093/nar/gku734

Oost, J. v. d. (2022) *Molecular Mechanisms of Type I CRISPR-Cas Systems. In CRISPR.*

Osawa, T., Inanaga, H. and Numata, T. (2013) 'Crystal structure of the Cmr2-Cmr3 subcomplex in the CRISPR-Cas RNA silencing effector complex', *J Mol Biol,* 425(20), pp. 3811-23. doi: 10.1016/j.jmb.2013.03.042

Osawa, T., Inanaga, H., Sato, C. and Numata, T. (2015) 'Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog', *Mol Cell,* 58(3), pp. 418-30. doi: 10.1016/j.molcel.2015.03.018

Ozcan, A., Pausch, P., Linden, A., Wulf, A., Schuhle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G. and Randau, L. (2019) 'Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum', *Nat Microbiol,* 4(1), pp. 89-96. doi: 10.1038/s41564-018-0274-8

Paget, M. S. (2015) 'Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution', *Biomolecules,* 5(3), pp. 1245-65. doi: 10.3390/biom5031245

Parma, D. H., Snyder, M., Sobolevski, S., Nawroz, M., Brody, E. and Gold, L. (1992) 'The Rex system of bacteriophage lambda: tolerance and altruistic cell death', *Genes Dev,* 6(3), pp. 497-510. doi: 10.1101/gad.6.3.497

Penner, M., Morad, I., Snyder, L. and Kaufmann, G. (1995) 'Phage T4-coded Stp: double-edged effector of coupled DNA and tRNA-restriction systems', *J Mol Biol,* 249(5), pp. 857-68. doi: 10.1006/jmbi.1995.0343

Peters, L. and Meister, G. (2007) 'Argonaute proteins: mediators of RNA silencing', *Mol Cell,* 26(5), pp. 611-23. doi: 10.1016/j.molcel.2007.05.001

Pfoh, R., Li, A., Chakrabarti, N., Payandeh, J., Pomes, R. and Pai, E. F. (2012) 'Structural asymmetry in the magnesium channel CorA points to sequential allosteric regulation', *Proc Natl Acad Sci U S A,* 109(46), pp. 18809-14. doi: 10.1073/pnas.1209018109

Picton, D. M., Luyten, Y. A., Morgan, R. D., Nelson, A., Smith, D. L., Dryden, D. T. F., Hinton, J. C. D. and Blower, T. R. (2021) 'The phage defence island of a multidrug resistant plasmid uses both BREX and type IV restriction for complementary protection from viruses', *Nucleic Acids Res,* 49(19), pp. 11257-11273. doi: 10.1093/nar/gkab906

Pingoud, A. and Jeltsch, A. (2001) 'Structure and function of type II restriction endonucleases', *Nucleic Acids Res,* 29(18), pp. 3705-27. doi: 10.1093/nar/29.18.3705

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

Pinilla-Redondo, R., Russel, J., Mayo-Munoz, D., Shah, S. A., Garrett, R. A., Nesme, J., Madsen, J. S., Fineran, P. C. and Sorensen, S. J. (2022) 'CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids', *Nucleic Acids Res,* 50(8), pp. 4315-4328. doi: 10.1093/nar/gkab859

Pourcel, C., Salvignol, G. and Vergnaud, G. (2005) 'CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies', *Microbiology (Reading),* 151(Pt 3), pp. 653-663. doi: 10.1099/mic.0.27437-0

Pratt, A. J. and MacRae, I. J. (2009) 'The RNA-induced silencing complex: a versatile gene-silencing machine', *J Biol Chem,* 284(27), pp. 17897-901. doi: 10.1074/jbc.R900012200

Pyenson, N. C., Gayvert, K., Varble, A., Elemento, O. and Marraffini, L. A. (2017) 'Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape', *Cell Host Microbe,* 22(3), pp. 343-353 e3. doi: 10.1016/j.chom.2017.07.016

Rakonjac, J., Bennett, N. J., Spagnuolo, J., Gagic, D. and Russel, M. (2011) 'Filamentous bacteriophage: biology, phage display and nanotechnology applications', *Curr Issues Mol Biol,* 13(2), pp. 51-76. doi: 10.21775/cimb.013.051

Raleigh, E. A. and Wilson, G. (1986) 'Escherichia coli K-12 restricts DNA containing 5-methylcytosine', *Proc Natl Acad Sci U S A,* 83(23), pp. 9070-4. doi: 10.1073/pnas.83.23.9070

Rankin, D. J., Rocha, E. P. and Brown, S. P. (2011) 'What traits are carried on mobile genetic elements, and why?', *Heredity (Edinb),* 106(1), pp. 1-10. doi: 10.1038/hdy.2010.24

Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., Wiedenheft, B., Doudna, J. A. and Greene, E. C. (2015) 'Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System', *Cell,* 163(4), pp. 854-65. doi: 10.1016/j.cell.2015.10.003

Reich, S., Gossl, I., Reuter, M., Rabe, J. P. and Kruger, D. H. (2004) 'Scanning force microscopy of DNA translocation by the Type III restriction enzyme EcoP15I', *J Mol Biol,* 341(2), pp. 337-43. doi: 10.1016/j.jmb.2004.06.031

Robert, X. and Gouet, P. (2014) 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Res,* 42(Web Server issue), pp. W320-4. doi: 10.1093/nar/gku316

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S., Dryden, D. T., Dybvig, K., Firman, K., Gromova, E. S., Gumport, R. I., Halford, S. E., Hattman, S., Heitman, J., Hornby, D. P., Janulaitis, A., Jeltsch, A., Josephsen, J., Kiss, A., Klaenhammer, T. R., Kobayashi, I., Kong, H., Kruger, D. H., Lacks, S., Marinus, M. G., Miyahara, M., Morgan, R. D., Murray, N. E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao, D. N., Reich, N., Repin, V. E., Selker, E. U., Shaw, P. C., Stein, D. C., Stoddard, B. L., Szybalski, W., Trautner, T. A., Van Etten, J. L., Vitor, J. M., Wilson, G. G. and Xu, S. Y. (2003) 'A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes', *Nucleic Acids Res,* 31(7), pp. 1805-12. doi: 10.1093/nar/gkg274

Rostol, J. T., Xie, W., Kuryavyi, V., Maguin, P., Kao, K., Froom, R., Patel, D. J. and Marraffini, L. A. (2021) 'The Card1 nuclease provides defence during type III CRISPR immunity', *Nature,* 590(7847), pp. 624-629. doi: 10.1038/s41586-021-03206-x

Rouillon, C., Athukoralage, J. S., Graham, S., Gruschow, S. and White, M. F. (2018) 'Control of cyclic oligoadenylate synthesis in a type III CRISPR system', *Elife,* 7. doi: 10.7554/eLife.36734

Rouillon, C., Athukoralage, J. S., Graham, S., Gruschow, S. and White, M. F. (2019) 'Investigation of the cyclic oligoadenylate signaling pathway of type III CRISPR systems', in Bailey, S. (ed.) *Crispr-Cas Enzymes Methods in Enzymology*, pp. 191-218.

Rouillon, C., Schneberger, N., Chi, H., Blumenstock, K., Da Vela, S., Ackermann, K., Moecking, J., Peter, M. F., Boenigk, W., Seifert, R., Bode, B. E., Schmid-Burgk, J. L., Svergun, D., Geyer, M., White, M. F. and Hagelueken, G. (2023) 'Antiviral signalling by a cyclic nucleotide activated CRISPR protease', *Nature,* 614(7946), pp. 168-174. doi: 10.1038/s41586-022-05571-7

Rousset, F., Depardieu, F., Miele, S., Dowding, J., Laval, A. L., Lieberman, E., Garry, D., Rocha, E. P. C., Bernheim, A. and Bikard, D. (2022) 'Phages and their satellites encode hotspots of antiviral systems', *Cell Host Microbe,* 30(5), pp. 740-753 e5. doi: 10.1016/j.chom.2022.02.018

Russel, J., Pinilla-Redondo, R., Mayo-Munoz, D., Shah, S. A. and Sorensen, S. J. (2020) 'CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci', *CRISPR J,* 3(6), pp. 462-469. doi: 10.1089/crispr.2020.0059

Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M. S., Siksnys, V. and Seidel, R. (2015) 'Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection', *Cell Rep,* 10(9), pp. 1534-1543. doi: 10.1016/j.celrep.2015.01.067

Ryazansky, S., Kulbachinskiy, A. and Aravin, A. A. (2018) 'The Expanded Universe of Prokaryotic Argonaute Proteins', *mBio,* 9(6). doi: 10.1128/mBio.01935-18

Saha, C. K., Sanches Pires, R., Brolin, H., Delannoy, M. and Atkinson, G. C. (2021) 'FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation', *Bioinformatics,* 37(9), pp. 1312-1314. doi: 10.1093/bioinformatics/btaa788

Salmond, G. P. and Fineran, P. C. (2015) 'A century of the phage: past, present and future', *Nat Rev Microbiol,* 13(12), pp. 777-86. doi: 10.1038/nrmicro3564

Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A. and Marraffini, L. A. (2015) 'Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity', *Cell,* 161(5), pp. 1164-1174. doi: 10.1016/j.cell.2015.04.027

Samolygo, A., Athukoralage, J. S., Graham, S. and White, M. F. (2020) 'Fuse to defuse: a self-limiting ribonuclease-ring nuclease fusion for type III CRISPR defence', *Nucleic Acids Res,* 48(11), pp. 6149-6156. doi: 10.1093/nar/gkaa298

Schmier, B. J., Nelersa, C. M. and Malhotra, A. (2017) 'Structural Basis for the Bidirectional Activity of Bacillus nanoRNase NrnA', *Sci Rep,* 7(1), pp. 11085. doi: 10.1038/s41598-017-09403-x

Sears, C. L. (2009) 'Enterotoxigenic Bacteroides fragilis: a rogue among symbiotes', *Clin Microbiol Rev,* 22(2), pp. 349-69, Table of Contents. doi: 10.1128/CMR.00053-08

Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., van der Oost, J., Brouns, S. J. and Severinov, K. (2011) 'Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence', *Proc Natl Acad Sci U S A,* 108(25), pp. 10098-103. doi: 10.1073/pnas.1104144108

Shah, S. A., Alkhnbashi, O. S., Behler, J., Han, W., She, Q., Hess, W. R., Garrett, R. A. and Backofen, R. (2019) 'Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes

reveals 39 new cas gene families', *RNA Biol,* 16(4), pp. 530-542. doi: 10.1080/15476286.2018.1483685

Shao, Y., Richter, H., Sun, S., Sharma, K., Urlaub, H., Randau, L. and Li, H. (2016) 'A Non-Stem-Loop CRISPR RNA Is Processed by Dual Binding Cas6', *Structure,* 24(4), pp. 547-554. doi: 10.1016/j.str.2016.02.009

Sheppard, N. F., Glover, C. V., 3rd, Terns, R. M. and Terns, M. P. (2016) 'The CRISPR-associated Csx1 protein of Pyrococcus furiosus is an adenosine-specific endoribonuclease', *RNA,* 22(2), pp. 216-24. doi: 10.1261/rna.039842.113

Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V. and 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), pp. E5307-e5316. doi: 10.1073/pnas.1803440115

Shukla, J., Gupta, R., Thakur, K. G., Gokhale, R. and Gopal, B. (2014) 'Structural basis for the redox sensitivity of the Mycobacterium tuberculosis SigK-RskA sigma-anti-sigma complex', *Acta Crystallogr D Biol Crystallogr,* 70(Pt 4), pp. 1026-36. doi: 10.1107/S1399004714000121

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

Simanshu, D. K., Yamaguchi, Y., Park, J. H., Inouye, M. and Patel, D. J. (2013) 'Structural basis of mRNA recognition and cleavage by toxin MazF and its regulation by antitoxin MazE in Bacillus subtilis', *Mol Cell,* 52(3), pp. 447-58. doi: 10.1016/j.molcel.2013.09.006

Simon-Baram, H., Kleiner, D., Shmulevich, F., Zarivach, R., Zalk, R., Tang, H., Ding, F. and Bershtein, S. (2021) 'SAMase of Bacteriophage T3 Inactivates Escherichia coli's Methionine S-Adenosyltransferase by Forming Heteropolymers', *mBio,* 12(4), pp. e0124221. doi: 10.1128/mBio.01242-21

Sineva, E., Savkina, M. and Ades, S. E. (2017) 'Themes and variations in gene regulation by extracytoplasmic function (ECF) sigma factors', *Curr Opin Microbiol,* 36, pp. 128-137. doi: 10.1016/j.mib.2017.05.004

Sistla, S. and Rao, D. N. (2004) 'S-Adenosyl-L-methionine-dependent restriction enzymes', *Crit Rev Biochem Mol Biol,* 39(1), pp. 1-19. doi: 10.1080/10409230490440532

Smalakyte, D., Kazlauskiene, M., Havelund, J. F., Ruksenaite, A., Rimaite, A., Tamulaitiene, G., Faergeman, N. J., Tamulaitis, G. and Siksnys, V. (2020) 'Type III-A CRISPR-associated protein Csm6 degrades cyclic hexa-adenylate activator using both CARF and HEPN domains', *Nucleic Acids Research,* 48(16), pp. 9204-9217. doi: 10.1093/nar/gkaa634

Sofos, N., Feng, M., Stella, S., Pape, T., Fuglsang, A., Lin, J., Huang, Q., Li, Y., She, Q. and Montoya, G. (2020) 'Structures of the Cmr-beta Complex Reveal the Regulation of the Immunity Mechanism of Type III-B CRISPR-Cas', *Mol Cell,* 79(5), pp. 741-757 e7. doi: 10.1016/j.molcel.2020.07.008

Sokolowski, R. D., Graham, S. and White, M. F. (2014) 'Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system', *Nucleic Acids Res,* 42(10), pp. 6532-41. doi: 10.1093/nar/gku308

Stern, A. and Sorek, R. (2011) 'The phage-host arms race: shaping the evolution of microbes', *Bioessays,* 33(1), pp. 43-51. doi: 10.1002/bies.201000071

Sternberg, S. H., LaFrance, B., Kaplan, M. and Doudna, J. A. (2015) 'Conformational control of DNA target cleavage by CRISPR-Cas9', *Nature,* 527(7576), pp. 110-3. doi: 10.1038/nature15544

Stetsenko, A. and Guskov, A. (2020) 'Cation permeability in CorA family of proteins', *Sci Rep,* 10(1), pp. 840. doi: 10.1038/s41598-020-57869-z

Stewart, F. J., Panne, D., Bickle, T. A. and Raleigh, E. A. (2000) 'Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme', *J Mol Biol,* 298(4), pp. 611-22. doi: 10.1006/jmbi.2000.3697

Strange, J. E. S., Leekitcharoenphon, P., Moller, F. D. and Aarestrup, F. M. (2021) 'Metagenomics analysis of bacteriophages and antimicrobial resistance from global urban sewage', *Sci Rep,* 11(1), pp. 1600. doi: 10.1038/s41598-021-80990-6

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

Studier, F. W. and Movva, N. R. (1976) 'SAMase gene of bacteriophage T3 is responsible for overcoming host restriction', *J Virol,* 19(1), pp. 136-45. doi: 10.1128/JVI.19.1.136-145.1976

Sumby, P. and Smith, M. C. (2002) 'Genetics of the phage growth limitation (Pgl) system of Streptomyces coelicolor A3(2)', *Mol Microbiol,* 44(2), pp. 489-500. doi: 10.1046/j.1365-2958.2002.02896.x

Suttle, C. A. (2007) 'Marine viruses--major players in the global ecosystem', *Nat Rev Microbiol,* 5(10), pp. 801-12. doi: 10.1038/nrmicro1750

Swarts, D. C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R. F., Koonin, E. V., Patel, D. J. and van der Oost, J. (2014) 'The evolutionary journey of Argonaute proteins', *Nat Struct Mol Biol,* 21(9), pp. 743-53. doi: 10.1038/nsmb.2879

Swarts, D. C., van der Oost, J. and Jinek, M. (2017) 'Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a', *Mol Cell,* 66(2), pp. 221-233 e4. doi: 10.1016/j.molcel.2017.03.016

Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V. and Seidel, R. (2014) 'Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes', *Proc Natl Acad Sci U S A,* 111(27), pp. 9798-803. doi: 10.1073/pnas.1402597111

Tajkarimi, M. and Wexler, H. M. (2017) 'CRISPR-Cas Systems in Bacteroides fragilis, an Important Pathobiont in the Human Gut Microbiome', *Front Microbiol,* 8, pp. 2234. doi: 10.3389/fmicb.2017.02234

Tal, N., Millman, A., Stokar-Avihail, A., Fedorenko, T., Leavitt, A., Melamed, S., Yirmiya, E., Avraham, C., Brandis, A., Mehlman, T., Amitai, G. and Sorek, R. (2022) 'Bacteria deplete deoxynucleotides to defend against bacteriophage infection', *Nat Microbiol,* 7(8), pp. 1200-1209. doi: 10.1038/s41564-022-01158-0

Tal, N., Morehouse, B. R., Millman, A., Stokar-Avihail, A., Avraham, C., Fedorenko, T., Yirmiya, E., Herbst, E., Brandis, A., Mehlman, T., Oppenheimer-Shaanan, Y., Keszei, A. F. A., Shao, S., Amitai, G., Kranzusch, P. J. and Sorek, R. (2021) 'Cyclic CMP and cyclic UMP mediate bacterial immunity against phages', *Cell,* 184(23), pp. 5728-5739 e16. doi: 10.1016/j.cell.2021.09.031

Tamulaitis, G., Venclovas, C. and Siksnys, V. (2017) 'Type III CRISPR-Cas Immunity: Major Differences Brushed Aside', *Trends Microbiol,* 25(1), pp. 49-61. doi: 10.1016/j.tim.2016.09.012

Tautvydas Karvelis, V. S. (2022) *Molecular Mechanisms of Type II CRISPR - Cas Systems. In Crispr.*

Taylor, D. W., Zhu, Y., Staals, R. H., Kornfeld, J. E., Shinkai, A., van der Oost, J., Nogales, E. and Doudna, J. A. (2015) 'Structural biology. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning', *Science,* 348(6234), pp. 581-5. doi: 10.1126/science.aaa4535

Thiaville, J. J., Kellner, S. M., Yuan, Y., Hutinet, G., Thiaville, P. C., Jumpathong, W., Mohapatra, S., Brochier-Armanet, C., Letarov, A. V., Hillebrand, R., Malik, C. K., Rizzo, C. J., Dedon, P. C. and de Crecy-Lagard, V. (2016) 'Novel genomic island modifies DNA with 7-deazaguanine derivatives', *Proc Natl Acad Sci U S A,* 113(11), pp. E1452-9. doi: 10.1073/pnas.1518570113

Tock, M. R. and Dryden, D. T. (2005) 'The biology of restriction and anti-restriction', *Curr Opin Microbiol,* 8(4), pp. 466-72. doi: 10.1016/j.mib.2005.06.003

Twort, F. W. (1936) 'Further Investigations on the Nature of Ultra-Microscopic Viruses and their Cultivation', *J Hyg (Lond),* 36(2), pp. 204-35. doi: 10.1017/s0022172400043606

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

van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., Soding, J. and Steinegger, M. (2023) 'Fast and accurate protein structure search with Foldseek', *Nat Biotechnol.* doi: 10.1038/s41587-023-01773-0

Vaucheret, H. (2008) 'Plant ARGONAUTES', *Trends Plant Sci,* 13(7), pp. 350-8. doi: 10.1016/j.tplants.2008.04.007

Vedel, M., Lawrence, F., Robert-Gero, M. and Lederer, E. (1978) 'The antifungal antibiotic sinefungin as a very active inhibitor of methyltransferases and of the transformation of chick embryo fibroblasts by Rous sarcoma virus', *Biochem Biophys Res Commun,* 85(1), pp. 371-6. doi: 10.1016/s0006-291x(78)80052-7

Velisetty, P. and Chakrapani, S. (2012) 'Desensitization mechanism in prokaryotic ligand-gated ion channel', *J Biol Chem,* 287(22), pp. 18467-77. doi: 10.1074/jbc.M112.348045

Vlot, M., Houkes, J., Lochs, S. J. A., Swarts, D. C., Zheng, P., Kunne, T., Mohanraju, P., Anders, C., Jinek, M., van der Oost, J., Dickman, M. J. and 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), pp. 873-885. doi: 10.1093/nar/gkx1264

Walker, F. C., Chou-Zheng, L., Dunkle, J. A. and Hatoum-Aslan, A. (2017) 'Molecular determinants for CRISPR RNA maturation in the Cas10-Csm complex and roles for non-Cas nucleases', *Nucleic Acids Res,* 45(4), pp. 2112-2123. doi: 10.1093/nar/gkw891

Walkinshaw, M. D., Taylor, P., Sturrock, S. S., Atanasiu, C., Berge, T., Henderson, R. M., Edwardson, J. M. and Dryden, D. T. (2002) 'Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA', *Mol Cell,* 9(1), pp. 187-94. doi: 10.1016/s1097-2765(02)00435-5

Wang, F., He, Q., Su, K., Wei, T., Xu, S. and Gu, L. (2018) 'Structural and biochemical characterization of the catalytic domains of GdpP reveals a unified hydrolysis mechanism for the DHH/DHHA1 phosphodiesterase', *Biochem J,* 475(1), pp. 191-205. doi: 10.1042/BCJ20170739

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M. and Wang, Y. (2015) 'Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems', *Cell,* 163(4), pp. 840-53. doi: 10.1016/j.cell.2015.10.008

Wang, L., Chen, S., Vergin, K. L., Giovannoni, S. J., Chan, S. W., DeMott, M. S., Taghizadeh, K., Cordero, O. X., Cutler, M., Timberlake, S., Alm, E. J., Polz, M. F., Pinhassi, J., Deng, Z. and Dedon, P. C. (2011) 'DNA phosphorothioation is widespread and quantized in bacterial genomes', *Proc Natl Acad Sci U S A,* 108(7), pp. 2963-8. doi: 10.1073/pnas.1017261108

Wang, L., Chen, S., Xu, T., Taghizadeh, K., Wishnok, J. S., Zhou, X., You, D., Deng, Z. and Dedon, P. C. (2007) 'Phosphorothioation of DNA in bacteria by dnd genes', *Nat Chem Biol,* 3(11), pp. 709-10. doi: 10.1038/nchembio.2007.39

Wang, L., Mo, C. Y., Wasserman, M. R., Rostol, J. T., Marraffini, L. A. and Liu, S. (2019) 'Dynamics of Cas10 Govern Discrimination between Self and Non-self in Type III CRISPR-Cas Immunity', *Mol Cell,* 73(2), pp. 278-290 e4. doi: 10.1016/j.molcel.2018.11.008

Wang, S., Wan, M., Huang, R., Zhang, Y., Xie, Y., Wei, Y., Ahmad, M., Wu, D., Hong, Y., Deng, Z., Chen, S., Li, Z. and Wang, L. (2021) 'SspABCD-SspFGH Constitutes a New Type of DNA Phosphorothioate-Based Bacterial Defense System', *mBio,* 12(2). doi: 10.1128/mBio.00613-21

Warren, R. A. (1980) 'Modified bases in bacteriophage DNAs', *Annu Rev Microbiol,* 34, pp. 137-58. doi: 10.1146/annurev.mi.34.100180.001033

Weigele, P. and Raleigh, E. A. (2016) 'Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses', *Chem Rev,* 116(20), pp. 12655-12687. doi: 10.1021/acs.chemrev.6b00114

Wein, T. and Sorek, R. (2022) 'Bacterial origins of human cell-autonomous innate immune mechanisms', *Nat Rev Immunol,* 22(10), pp. 629-638. doi: 10.1038/s41577-022-00705-4

Weinberger, A. D., Sun, C. L., Plucinski, M. M., Denef, V. J., Thomas, B. C., Horvath, P., Barrangou, R., Gilmore, M. S., Getz, W. M. and Banfield, J. F. (2012) 'Persisting viral sequences shape microbial CRISPR-based immunity', *PLoS Comput Biol,* 8(4), pp. e1002475. doi: 10.1371/journal.pcbi.1002475

Wexler, H. M. (2007) 'Bacteroides: the good, the bad, and the nitty-gritty', *Clin Microbiol Rev,* 20(4), pp. 593-621. doi: 10.1128/CMR.00008-07

Whiteley, A. T., Eaglesham, J. B., de Oliveira Mann, C. C., Morehouse, B. R., Lowey, B., Nieminen, E. A., Danilchanka, O., King, D. S., Lee, A. S. Y., Mekalanos, J. J. and Kranzusch, P. J. (2019) 'Bacterial cGAS-like enzymes synthesize diverse nucleotide signals', *Nature,* 567(7747), pp. 194-199. doi: 10.1038/s41586-019-0953-5

Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., Westphal, W., Heck, A. J., Boekema, E. J., Dickman, M. J. and 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), pp. 10092-7. doi: 10.1073/pnas.1102716108

Wright, A. V., Liu, J. J., Knott, G. J., Doxzen, K. W., Nogales, E. and Doudna, J. A. (2017) 'Structures of the CRISPR genome integration complex', *Science,* 357(6356), pp. 1113-1118. doi: 10.1126/science.aao0679

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

Wu, J., Yang, J., Cho, W. C. and Zheng, Y. (2020) 'Argonaute proteins: Structural features, functions and emerging roles', *J Adv Res,* 24, pp. 317-324. doi: 10.1016/j.jare.2020.04.017

Xiao, Y., Luo, M., Dolan, A. E., Liao, M. and Ke, A. (2018) 'Structure basis for RNA-guided DNA degradation by Cascade and Cas3', *Science,* 361(6397). doi: 10.1126/science.aat0839

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

Xiong, W., Zhao, G., Yu, H. and He, X. (2015) 'Interactions of Dnd proteins involved in bacterial DNA phosphorothioate modification', *Front Microbiol,* 6, pp. 1139. doi: 10.3389/fmicb.2015.01139

Xiong, X., Wu, G., Wei, Y., Liu, L., Zhang, Y., Su, R., Jiang, X., Li, M., Gao, H., Tian, X., Zhang, Y., Hu, L., Chen, S., Tang, Y., Jiang, S., Huang, R., Li, Z., Wang, Y., Deng, Z., Wang, J., Dedon, P. C., Chen, S. and Wang, L. (2020) 'SspABCD-SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities', *Nat Microbiol,* 5(7), pp. 917-928. doi: 10.1038/s41564-020-0700-6

Yamagata, A., Kakuta, Y., Masui, R. and Fukuyama, K. (2002) 'The crystal structure of exonuclease RecJ bound to Mn2+ ion suggests how its characteristic motifs are involved in exonuclease activity', *Proc Natl Acad Sci U S A,* 99(9), pp. 5908-12. doi: 10.1073/pnas.092547099

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I. M., Li, Y., Fedorova, I., Nakane, T., Makarova, K. S., Koonin, E. V., Ishitani, R., Zhang, F. and Nureki, O. (2016) 'Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA', *Cell,* 165(4), pp. 949-62. doi: 10.1016/j.cell.2016.04.003

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

Yang, H., Gao, P., Rajashankar, K. R. and Patel, D. J. (2016) 'PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease', *Cell,* 167(7), pp. 1814-1828 e12. doi: 10.1016/j.cell.2016.11.053

Yao, Q., Cao, G., Li, M., Wu, B., Zhang, X., Zhang, T., Guo, J., Yin, H., Shi, L., Chen, J., Yu, X., Zheng, L., Ma, J. and Su, Y. Q. (2018) 'Ribonuclease activity of MARF1 controls oocyte RNA homeostasis and genome integrity in mice', *Proc Natl Acad Sci U S A,* 115(44), pp. 11250-11255. doi: 10.1073/pnas.1809744115

Ye, Q., Lau, R. K., Mathews, I. T., Birkholz, E. A., Watrous, J. D., Azimi, C. S., Pogliano, J., Jain, M. and Corbett, K. D. (2020a) 'HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity', *Mol Cell,* 77(4), pp. 709-722 e7. doi: 10.1016/j.molcel.2019.12.009

Ye, Q., Zhao, X., Liu, J., Zeng, Z., Zhang, Z., Liu, T., Li, Y., Han, W. and Peng, N. (2020b) 'CRISPR-Associated Factor Csa3b Regulates CRISPR Adaptation and Cmr-Mediated RNA Interference inSulfolobus islandicus', *Frontiers in Microbiology,* 11. doi: 10.3389/fmicb.2020.02038

You, L., Ma, J., Wang, J., Artamonova, D., Wang, M., Liu, L., Xiang, H., Severinov, K., Zhang, X. and Wang, Y. (2019) 'Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference', *Cell,* 176(1-2), pp. 239-253 e16. doi: 10.1016/j.cell.2018.10.052

Zabeau, M., Friedman, S., Van Montagu, M. and Schell, J. (1980) 'The ral gene of phage lambda. I. Identification of a non-essential gene that modulates restriction and modification in E. coli', *Mol Gen Genet,* 179(1), pp. 63-73. doi: 10.1007/BF00268447

Zaremba, M., Dakineviciene, D., Golovinas, E., Zagorskaite, E., Stankunas, E., Lopatina, A., Sorek, R., Manakova, E., Ruksenaite, A., Silanskas, A., Asmontas, S., Grybauskas, A., Tylenyte, U., Jurgelaitis, E., Grigaitis, R., Timinskas, K., Venclovas, C. and Siksnys, V. (2022) 'Short prokaryotic Argonautes provide defence against incoming mobile genetic elements through NAD(+) depletion', *Nat Microbiol,* 7(11), pp. 1857-1869. doi: 10.1038/s41564-022-01239-0

Zeng, Z., Chen, Y., Pinilla-Redondo, R., Shah, S. A., Zhao, F., Wang, C., Hu, Z., Wu, C., Zhang, C., Whitaker, R. J., She, Q. and Han, W. (2022) 'A short prokaryotic Argonaute activates membrane effector to confer antiviral defense', *Cell Host Microbe,* 30(7), pp. 930-943 e6. doi: 10.1016/j.chom.2022.04.015

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V. and Zhang, F. (2015) 'Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system', *Cell,* 163(3), pp. 759-71. doi: 10.1016/j.cell.2015.09.038

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S. V., Naismith, J. H., Spagnolo, L. and White, M. F. (2012) 'Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity', *Mol Cell,* 45(3), pp. 303-13. doi: 10.1016/j.molcel.2011.12.013

Zhang, T., Tamman, H., Coppieters 't Wallant, K., Kurata, T., LeRoux, M., Srikant, S., Brodiazhenko, T., Cepauskas, A., Talavera, A., Martens, C., Atkinson, G. C., Hauryliuk, V., Garcia-Pino, A. and Laub, M. T. (2022) 'Direct activation of a bacterial innate immune system by a viral capsid protein', *Nature,* 612(7938), pp. 132-140. doi: 10.1038/s41586-022-05444-z

Zhu, W., McQuarrie, S., Gruschow, S., McMahon, S. A., Graham, S., Gloster, T. M. and White, M. F. (2021) 'The CRISPR ancillary effector Can2 is a dual-specificity nuclease potentiating type III CRISPR defence', *Nucleic Acids Res,* 49(5), pp. 2777-2789. doi: 10.1093/nar/gkab073

Zhu, X. and Ye, K. (2015) 'Cmr4 is the slicer in the RNA-targeting Cmr CRISPR complex', *Nucleic Acids Res,* 43(2), pp. 1257-67. doi: 10.1093/nar/gku1355

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A. N. and Alva, V. (2018) 'A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core', *J Mol Biol,* 430(15), pp. 2237-2243. doi: 10.1016/j.jmb.2017.12.007

Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C. and Adams, P. D. (2008) 'Automated structure solution with the PHENIX suite', *Methods Mol Biol,* 426, pp. 419-35. doi: 10.1007/978-1-60327-058-8_28

Appendices

Plasmids used in this study

Name	Description	Reference or source
pBfrCmr1-6	pACE-based construct was assembled with five PCR products of BfrCmra, b, c, d and e; This plasmid is used for expression of 6 subunits of type III B Cmr complex from *Bacteroides fragilis* (Bfr). Primers BfrCmrSG1-F/R, BfrCmrSG2-F/R, BfrCmrSG3-F/R, BfrCmrSG4-F/R and BfrCmrSG5-F/R were used for this construction. Ap[r]	This work
pCDFDuet	Vector used for pBfrCRISPR construction and co-expression and co-purification of CaplS and CalpT; Sp[r]	Novagen, Missouri, USA
pBfrCRISPR_Tet	Genes encoding BfrCas6 and mini-CRISPR array targeting the portion of tetracycline resistance gene were inserted in MCS-2 and MCS-1 of pCDFDuet, respectively. Primers Spacer_TetR-F and R were used for this construction. Sp[r]	This work
pBfrCRISPR_pUC	Genes encoding BfrCas6 and mini-CRISPR array targeting the portion of pUC19 LacZ gene were inserted in MCS-2 and MCS-1 of pCDFDuet, respectively. Primers Spacer_pUC-F and R were used for this construction. Sp[r]	This work
pBfrCRISPR_Lpa	Genes encoding BfrCas6 and mini-CRISPR array targeting the gene encoding Late Promoter Activating protein (Lpa) of phage P1 were inserted in MCS-2 and MCS-1 of pCDFDuet, respectively. Primers Bfr-rep-5p-T, Bfr-rep-5p-C, Bfr-rep-3p-T, Bfr-rep-3p-C, Bfr-sp-phageLPA-T and Bfr-sp-phageLPA-S were used for this construction. Sp[r]	This work
pCDFDuet-CalpT	Gene encoding CalpT from *Sulfurihydrogenibium* spp. was inserted into the vector pCDFDuet. Sp[r]	This work
pCDFDuet-CalpS	Gene encoding CalpS from *Thermosipho* was inserted into the vector pCDFDuet. Sp[r]	This work
pEHisV5TEV	Vector used for proteins expression with a cleavable eight histidines tag, followed by a V5 epitope tag. Km[r]	Rouillon et al., 2019
pEHisV5TEV-BfrCas6	Gene encoding BfrCas6 was inserted into the vector pEHisV5TEV. Km[r]	This work
pEHisV5TEV-BfrNrN	Gene encoding BfrNrN was inserted into the vector pEHisV5TEV. Km[r]	This work
pEHisV5TEV-BfrCorA	Gene encoding BfrCorA was inserted into the vector pEHisV5TEV. Km[r]	This work
pEHisV5TEV-BfrNYN	Gene encoding BfrNYN was inserted into the vector pEHisV5TEV. Km[r]	This work
pEHisV5TEV-Cbolyase	Gene encoding Cbolyase from *Clostridium botulinum* was inserted into the vector pEHisV5TEV. Km[r]	This work
pEHisV5TEV-CalpS	Gene encoding CalpS was inserted into the vector pEHisV5TEV. Km[r]	This work

pEHisV5TEV-CalpT^{tr}	Gene encoding truncated CalpT (CalpT^{tr}, aa1-173) was inserted into the vector pEHisV5TEV. Km^r	This work
pRATDuet	Vector for cloning used in the plasmid challenge assay. Tc^r	Athukoralage et al., 2020b
pRATDuet-BfrCorA	Gene encoding BfrCorA was inserted in MCS-1 of vector pRATDuet. Tc^r	This work
pRATDuet-BfrNrN	Gene encoding BfrNrN was inserted in MCS-1 of vector pRATDuet. Tc^r	This work
pRATDuet-BfrNYN	Gene encoding BfrNYN was inserted in MCS-1 of vector pRATDuet. Tc^r	This work
pRATDuet-BfrCorA-BfrNrN	Genes encoding BfrNrN and BfrCorA were inserted in MCS-2 and MCS-1 of vector pRATDuet, respectively. Tc^r	This work
pET11a-CalpT	Gene encoding CalpT were inserted into vector pET11a. Ap^r	Constructed by our collaborators
pACE-vmeCmr	This plasmid is used for expression of 6 subunits of type III B Cmr complex from *Vibrio metoecus* (Vme). Ap^r	Gruschow et al., 2021
pCDF-target-CRISPR	Genes encoding VmeCas6 and mini-CRISPR array targeting the portion of tetracycline resistance gene were inserted in vector pCDFDuet, respectively. Sp^r	Gruschow et al., 2021
pCDF-nontarget-CRISPR	Genes encoding VmeCas6 and mini-CRISPR array targeting the portion of pUC19 LacZ gene were inserted in vector pCDFDuet, respectively. Sp^r	Gruschow et al., 2021
pRATDuet-vmeNucC	Gene encoding VmeNucC was inserted in the vector pRATDuet. Tc^r	Gruschow et al., 2021
pBfrCmr1-6_Cmr2Δcyclase	Derived from pBfrCmr1-6 by site-directed mutagenesis using primers BfrCmr2_cyclase-F and R	This work
pBfrCmr1-6_Cmr4D27A	Derived from pBfrCmr1-6 by site-directed mutagenesis using primers BfrCmr4_D27A-F and R	This work
pBfrCmr1-6_Cmr2D70N	Derived from pBfrCmr1-6 by site-directed mutagenesis using primers BfrCmr2_D70N-F and R	This work
pBfrCmr1-6_Cmr2E151R	Derived from pBfrCmr1-6 by site-directed mutagenesis using primers BfrCmr2_E151R-F and R	This work
pBfrCmr1-6_Cmr2D70NE151R	Derived from pBfrCmr1-6 by site-directed mutagenesis using primers BfrCmr2_D70N-F and R and BfrCmr2_E151R-F and R	This work
pEHisV5TEV-BfrNrN^Δ	Derived from pEHisV5TEV-BfrNrN for expression and purification of inactive variant BfrNrN^Δ (D85A/H86A/H87A) by site-directed mutagenesis using primers NrN^Δ-F and R	This work
pEHisV5TEV-BfrCorA^{tr}	Derived from pEHisV5TEV-BfrCorA for expression and purification of truncated BfrCorA (BfrCorA^{tr}, aa 1-428) by site-directed mutagenesis using primers CorA^{tr}-F and R	This work
pEHisV5TEV-BfrNYND13A	Derived from pEHisV5TEV-BfrNYN by site-directed mutagenesis using primers NYN^{D13A}-F and R	This work
pEHisV5TEV-BfrNYND72A	Derived from pEHisV5TEV-BfrNYN by site-directed mutagenesis using primers NYN^{D72A}-F and R	This work
pRATDuet-BfrCorA-BfrNrN^Δ	Derived from pRATDuet-BfrCorA and pEHisV5TEV-BfrNrN^Δ by insertion of genes encoding variant NrN^Δ from pEHisV5TEV-BfrNrN^Δ into vector pRATDuet-BfrCorA	This work
pRATDuet-BfrCorA^{tr}-BfrNrN	Derived from pRATDuet-BfrNrN and pEHisV5TEV-BfrCorA^{tr} by insertion of genes encoding truncated CorA from pEHisV5TEV-BfrCorA^{tr} into vector pRATDuet-BfrNrN	This work

pRATDuet-BfrCorA^{tr}-BfrNrN^Δ	Derived from pRATDuet-BfrCorA-BfrNrN^Δ and pRATDuet-BfrCorA^{tr}-BfrNrN by insertion of genes encoding variant NrN^Δ and CorA^{tr} into vector pRATDuet	This work

Synthetic gene of BfrCmra

CCAGACGTACCTGCCGGCATTCTTCATCTGTCATAACTTCGGGACCCGTAATAAT
AAGGGATTCGGTAGCTTCACGGTGGAGTACATCAATAACCAAAAAATATCTGT
AATGTCGAGGACACATTGAAAGAAAATTTTGCGTTCGTATATAAGAAAAAGATC
GCTCTTTCGCGTCAATCCACACTGGACTTTATTTATATTTATAATCAGATCTTTAG
TACAATCAAAAAGGACTATCAAATTCTTAAGAGTGGCTATAATTTTCGTAATGAG
TATATCAAATCCTTGCTTTTTTGCTACTTTGTGTCCAAGTATCCAAATTATCGCTG
GGAAAAACGCAAGATGAAACAGCTTATTAAGGCCCGTGGCTATGAATTGAAAGG
AGATCATTCGCCAATCAGTGGGATTCGTGAAAACGACAATTCTTGGAACGACCCT
AATCCCAACGGGTATAATTATGCGTATATTCGTGCTATTCTTGGCCTTGCTGAGC
AGTACGAGTTCCAGTTGGAAACACCCTACCAGAAGGCAATTGTTAAAATCAAGT
CGGCCAATAACTGCATCTCACGTTATAAATCCCCTTTACTTTTCAAAATCATTAAT
AACTCCATCTACTTGGTGGGGAACGAGATCAATACGGAAATTTTGAATAAGCCG
TTTCAGTATTCATACATTGAACAAACGAAAAACAAGAACATGCGCACTGGAAAG
TCAGAGATTACGGAACGTACAATGCATATCAATGAAATCGAGATGAACTACAAG
AACCGTATTAACTATCACTACACTCCAACAAGCTTTTCATTGATCGACTTTATGC
AATACGCGATGTCATACAAGAAGAACGGTAAAAACATCTTAAATTATATTCCCTT
AAAACAGTAAGACTAAGAAGGAGATATACATATGAAATACATCGCGATCACTCT
TGGTCCGATTACTCGCACCATCGAGATGGCAGAATCCACGAAGGAGTTGTGGGC
GGCGTCTTACTTTTTTTCGTATCTTGCCAAGAAAATTGTAGAACCCTTTGTCAAAA
AGAATCGCACGTTTCAATTACCTCTTATTAACGAGGAGATGCAGAAGCCCCACTG
CGGTGCAGGGTTGTTTCCCGACCGTTATATCTTCAAGTCGGAACCTGGAGACCTG
GAGTTACTTAAGCAACATTCCGACCAAGTACTTATCGAGATCGCGGGCCATATCG
CGAGCCCCAGTTTACCTGGGACAGCGAAAGATGTGTCGCAAATTTACCATTACCT
GAAGAGTTATATCAAGATCTATTTCATCGAGCGCACACTGGAATCCGATGACCCT
CATGTAGTCATCCCGGCCTGTGAAAAGTACCTGAACATTATTGAAAATCAGGAG
ACTTTTCCGGAGCAGGAGGAAACCATGATTTCCCACCAGAAAAGTGATTTCCTTA
AATTCTTAATTACAAACGTTAATGGTAAAATCTACCGCAAAGACAAGAATAGTAT
TCCACGCTTTACTGGCTCATTCTTGACTCGCGACGCTTTCGGAGACATGAATGGA
GAGCGCCTTTTTGAGAGCATCTTAGAAATCTCTGCGAGTGAGCTTAACATTAACA
TTCAGCAGAAGGCGTTGGAGGTTATCACTGCAAACGAGAAGAACAAAGGCGAA

AAGTATAGTGACCAAATCTGGGACGCAGAAGAAATTATCCTTAACGATAACAAA

GCACAATTACGCCCCTACCACAAATACATCGCTATTATTAAAAGTGACGGAGATT

CTATGGGAGAAACGATTAAGAGCATGGGTGCATACAACATCCCAATTACTCAGC

TTTCAAAAGCGCTGTTGTCCTTCAATATCGAAAGCATCAATGAAATCG

Synthetic gene of BfrCmrb

GCGCTGTTGTCCTTCAATATCGAAAGCATCAATGAAATCGTTGCCTATGGCGGAA

AGCCGATCTTCATTGGAGGGGACGATTTGCTTTGTTTTGCGCCGGTATGTTGCAA

CGGTAATAACGTTTTCAATTTGGTCGAGAAACTGAGCACTTGTTTCGACCAGTGT

ATTAATCAACATCTTCAACAATACATTAATGCTTGCAGCGAGGCGCAGCGTCCCT

TACCAAGCTTGTCTTTCGGTATCAGCATCACGTATCATAAATACCCTATGTTTGA

AGCCCTTCACACTACCGACTATCTTTTAGAAATGGTGGCCAAGGACAACTTGTTC

AAGTATACCTTGAGCAATAAAAACATTCTGAATGAAAATATGAAGCGCTTTATTT

TGAAAAATAAATTGGCGTTCTCTCTTCAAAAGCATAGTGGACAGATCTACCATAC

CGCTATGTCGAAAAAGGGAAAGTCCTACGTGAAGTTTAACATGCTTCTTCAAAA

GTACATTCTGAAGAACAAGGACATGAGTAAGACCCAGGAATCTGAAAAATTTTT

ATCATCCGTAATCCAAATGATTCGTGCTCATGCTGAGATCCTTCAAATCATTTTGC

AGAATGAAGACAAACGTACCGAAATGTTAAAAAACTACTTTGATAACAACTTCA

ATGAGAGTTGTCACCTTGGGTACACGGGATTGTTTGAGGATATCCAAACCTTGCT

GTGTTTACGCTACCAAGAAAATATTCAAGATTACCAAAACCGTAATGAAATTATT

CAGCAGAACACTATCCTGACGAGTGACGAGAAGGAGATTCTGATCGTGTCACCG

GCCATGGATGCAATTCATACGATTTTCACAGCGTTGCAATTTATCCACTTCATTA

ATTATAATAAAGATGAGTAACCTTAAGAAGGAGATATAACCATGTCGCATCACC

ATCATCACCATCATCACGATGGCAAACCGATTCCGAACCCGCTGCTGGGCCTGGA

TAGCACCGGCAGCGACCAGACCGAGAACAGCGGCGAAAACCTGTATTTTCAGGG

CGCAAACGCCATGAACCGTCATTACTTAATCACTCTGACCCCGATGGATTGGTTT

TTTTTTGGCGGTGAGCGTACGCTGGATGACGGTAAGTCCGCTGACTATATCAGCC

ATTCCAACAAGTTTCCTCAGCAGTCTGCCCTGTTAGGGATGATTCGTTATCAGTT

GCTTAAACAGCATAACTTATTGTCACAATTCCCATATACAGAAAATAAGCCCACC

GAGAAAGAGATTATGAAAACCCTGATCGGGGAGCAAAGCTTCCGCATGACAGAG

CGCAAGGCGAAAGTCTGGGATTGGGCGTAATTAAACAGATTAGTCCATTAATG

CTGATCGAGTGTAAAGATGACACTTCCTCCCGTAGTATCTATTTCCCCTTACCTCT

GGACGACGGTTATAAGGTGTCGTTCAATGAGACGAGCAATGAGGACAAAGTGTT

CTATAACGGGATTGAATGCCCCATTCCTAACGTATATCCTGCATCGGAGGAGCAG

GACAGCGGGAATCAGAACGCAAGTTTTTCGATCATAAAACATACAACAACTAT

CTGTTTTGGTGCACTCAAGGAAACAACCAGATTAAGAAGCTGTTGTCAGACGAG

ATCTGGATTTCTAAGATGCAAATCGGCATTACAAAGCACGTTGAAGAAGGAGAG

GATAACGATAAATCGTTTTATAAACAAGAGTTTCTGCAGCTTAAAAAGAGCTTCA

TCTATGCCTTCTATATCACATTATCGGGCGAAAGCGAATTGTCATCGGACATTAT

TCAACTTGGCGGTCAACGTTCCGTATTTCGTATGGAGGTTGAAAGTATTGAAGAA

AATTCCGACATTCAGGAGAAATATCAGACTGCAGCACAGTTTTTGACCCAGTCCG

ACCGTCTTCTTATCCTTTCTCCTACGTATGTAGACAATTTAAAAGAACTGTCAGCT

TTGTGTAACTTTATGTGGTCGGATTCAATCG

Synthetic gene of BfrCmrc

CTGTCAGCTTTGTGTAACTTTATGTGGTCGGATTCAATCGTATTTCGCAATATCCA

AACGACGAATGCCTCTAACTTCTATGGGAAACCTATCAAATCGTCCTCCAAATAT

CACTTTTTAAAGCCGGGGTCGGTCTTATATTTCAAACAAGGCAAACGTAAAGAG

GTTGAGAAGCTGCTGATGGATTATACTTATCTGCGCCTTTCCGGCTACAACATTT

ATATCTAAGAACAGAAAGTAATCGTATTGTACACGGCCGCATAATCGAAATTAA

TACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCATCTTAGTATAT

TAGTTAAGTATAAGAAGGAGATATAACCATGACCACGTATGTACGTCATTAA

CACGTTAAGCAATATGCACGTGGGTAGCGGGGAGGTGAACTACGGAGTTATTGA

TAACTTGATCCAACGTGACTCTGTAACGAACTTACCAAACATTAATTCTTCGGGG

TTAAAAGGCGCGATTCGTGAATACTTCAAGGAGAATGAAGACCTGGTACGCGAG

TTGTTCGGATCAGCTCCACGTGACGAGAAGACGTTGCCAGGAAAAGTCCGCTTTT

TCGAAGCGAACTTACTTTCGATGCCGGTCCGCTCCGATAAGGTCCCCTTTTTGAT

GGCTATCAGTGATGAGGTATTGCAAGAGCTGATTACCAAAATGAAATTCTTTAAT

TGTGAAGAGGCGACTCAGTACATTTCCCATTTGAGCACTTTACTTGATAACATTA

AAACACAAGCGCAAGGTACCGACTTTGCCTACGTCTTTGACCCTTTATTGCAGGG

TGCTATCATTGAGGAAGTATCGATCCGTGCAACTTGCCCGTCGCACATCCCTCTT

CAGCCTTCACTTAAGAAATTACTTGGTGATCGTCTTGTGATCTTGTCCCATAAGTA

TTTTTCAATTTTGAGTGATGATAATCATTTGCCGGTTCTTTCTCGCAACAATCTGG

AGAACGGGCAGTCAGCCAATCTTTGGTATGAGCAAGTGCTGCCGCGTTATAGTC

GCTTGTACTTCATGTTGATGGATGGTAATGCTCAATCCGAGTACCTTAAAAAATT

TCGTGACACATTGTGTACCCCGTCTACCATCATTCAGATTGGAGCTAATGCGTCG

ATTGGATATGGGTATTGCCAAATTAGCGAACTGTCGCCGTTCTAACTATAAGAAG

GAGATATACACATGAAGATTAGTAAGAAACAAATCGAATATGCCATCGAAGCAC

TGCGTGCAAATAACATTATCACAAATGACAACCAGTACCCGAAGGTTTTCAAAG

GATACATTTCTTCCTTCGGCGCGGCAGTGATCCAGAGCGGACTTATTCCTGCAAT

CATTTTTTTCGAGAACGAAGATAACGACGCCAACGCGGACCGTCACAAGATTATT

GGTGTGTTGAAGGACATTATCAACGCAATGCGTCAGCAATACACAGTTACTGAT

GCGACTATTTTGGTCTCTTCACAAATTCCCGCAAATTATAGTATGGCGCAGTACA

TTATTGAACATGGCAATACGGATCAATTACTTAAGGAAATTACAGAGGCTGCTGT

GGCAATGAAGTTAGCGTTACGTATGTATAAGAGCGAGTGAGTCTAAGAAGGAGA

TATAACTATGCCGAAAAATTATACGTTACAAAATGCGTCAAACCTTGGATGGCTG

TTCTACAAGGATTACTACCGCCAGGAACCAAACGTAGACTTCATCTCGACTCAGG

GTAAGGAGTCCGATACCACGGCCGACTTTTTTCGTAAGACAAACCAGCGTATTAC

GGCCTACCAGCTGAATAGCGAATCGCCTCTGGTTGCCGCGTTCAATAACCATTTC

GGG

Synthetic gene of BfrCmrd

CGAATCGCCTCTGGTTGCCGCGTTCAATAACCATTTCGGGACTCCCCTTCAACTG

AAGACCATTTATCCTGGGTTGATCACAGGGTCGGGTCTGCCACACCAAACAGGG

TCCAAAGGCGAATTTAAACTTGGATTCCAATTTGACTATACGACGGGCTTACCCT

ACATTCCCGGAAGTAGTATTAAAGGAACTCTTCGCTCTATGTTCCCGTTTTCGCTT

AAGGACAAAGGTTCTACAAAGCGTATTTTACCAGAGTATCGCAAGGAACGTATG

GAATACATCCGTGACTTGATTATCGAGGTAACCAACATTAACGAAATTTCTGACA

CGGAAATTCAAGCTCTGGAATACGCTATCTTCACCAATTCCACACCGTCTGGAAA

AACAATTGAGTTTAGCTTAGAGGAAAAGATGTTTTCTATGACGCGTTCGTCGCT

GATTCCAAAGATGGGGTAATGCTTTCAGATGATTATATTACTCCTCACGGGGAGA

ATCCCCTGAAAGATCCAAAACCCATTTTGTTCTTGAAGATTCGCCCGGATGTAAC

AATCAACTTCTACTTTAAGTTGTGCACAACACACTTGTATAAGGAAAAAGTCTGT

AGCTCCAAGCAGATCGAAGAGATTAAGAAACAAAATGATTTTAGTTCCAGCGAT

TACAAAATGATTACGGCCCATCAGAAGCGCAACTTGTTCGAGAAGATTTTACTTT

GCATTGGTATTGGGGCTAAAACTAACATTGGATACGGGCAGTTGAAGAAATTAT

GACCTGTAAAACGACGGCCAGTGAATTCCCCGGGAAGCTTCGCCAGGGTTTTCCC

AGTCGAGCTCGATATCGGTACCAGCGGATAACAATTTCACATCCGGATCGCGAA

CGCGTCTCGAGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCT

GCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCT

TGAGGGGTTTTTTGGTTTAAACCCATCTAATTGGACTAGTAGCCCGCCTAATGAG

CGGGCTTTTTTTAATTCCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTA

TCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAA

GAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTT

GCCTTCCTGTTTTTGCTCACCCAGAAACGCTCGTGAAAGTAAAAGACGCAGAGG

ACCAATTGGGGGGCACGAGTGGGATACATAGAACTGGACTTGAATAGCGGTAAAA

TCCTTGAGAGTTTTCGCCCTGAAGAGCGTTTTCCAATGATGAGCACTTTCAAAGT

TCTGCTATGTGGAGCAGTATTATCCCGTGTAGATGCGGGGCAAGAGCAACTCGG

ACGACGAATACACTATTCGCAGAATGACTTGGTTGAATACTCCCCAGTGACAGA

AAAGCACCTTACGGACGGAATGACGGTAAGAGAATTATGTAGTGCCGCCATAAC

GATGAGTGATAACACTGCGGCGAACTTACTTCTGACAACCATCGGTGGACCGAA

GGAATTAACCGCTTTTTTGCACAATATGGGAGACCATGTAACTCGCCTTGACCGT

TGGGAACCAGAACTGAATGAAGCCATACCAAACGACGAGCGAGACACCACAAT

GCCTGCGGCAATGGCAACAACATTACGCAAACTATTAACTGGCGAACTACTTACT

CTGGCTTCACGGCAACAATTAATAGACTGGCTTGAAGCGGATAAAGTTGCAGGA

CCACTACTGCGTTCGGCACTTCCTGCTGGCTGGTTTATTGCTGATAAATCTGGGG

CAGGAGAGCGTGGTTCACGGGGTATCATTGCCGCACTTGGACCAGATGGTAAGC

CTTCCCGTATCGTAGTTATCTACACGACGGGTAGTCAGGCAACTATGGACGAACG

AAATAGACAGATTGCTGAAATAG

Synthetic gene of BfrCmre

GCAACTATGGACGAACGAAATAGACAGATTGCTGAAATAGGGGCTTCACTGATT

AAGCATTGGTAAACCGATACAATTAAAGGCTCCTTTTGGAGCCTTTTTTTTTGGA

CGGACCGGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTA

ATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGG

ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGAT

ACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCT

GTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA

GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATA

AGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGC

GAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCA

CGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGA

ACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGT

CCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGG

GGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGC

CTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGA

TAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACC

GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTT

CTCCTTACGCATCTGTGCGGTATTTCACACCGCAATGGTGCACTCTCAGTACAAT

CTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACT

GGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGG
CTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTG
CATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGGGGGAATT
CCAGATAACTTCGTATAATGTATGCTATACGAAGTTATGGTGTCCGGGATCTCGA
CGCTCTCCCTTATGCGACTCCTGCATTAGGAAATTAATACGACTCACTATAGGGG
AATTGTGAGCGGATAACAATTCCCCTGTAGAAATAATTTTGTTTAACTTTAATAA
GGAGATATACCATGAACCAACTTACTGCCATTTTGAAGCAGCATACCCCAATGAT
TCATTTCCAGCACAACGAGTCAGGGGCTACCCTTCGTGCGTCCGAGGTTAAACCC
CTGCTTGATAAGTTTATTTTAACTAAGCTGGGGAATGGAGACATCCGCGAGGGAC
GTTTGTATGCTAAAAAAATAATTGGTTGATTGACAATGAGAAGAACTACGCTCT
GAATTACAAATTGTCAATTTCGCTGCAGAAGAAAGCCGTCTGGAATACCTTATC
ACCTCTAGCACATTTCCTTTGCCAACAGAACGTCCGTCAAATTTCTTTACCATCCA
AAACTCTCCTTACTTTGCACAAGAAAGTGCGTCGGAATTAATACGAACAGCAC
CATCATTTTGAAGAAGTCAAACTCGGACCCGCGCAAAAAGGAGGCGGAGTTCAA
AGAGAAAACTGGTCACAGATTGATAAAAGGGTCTGGAATGGCAAGATTTTAC
TATTAAAATCTTCTCTCTTAAGGGGGATTTAATCAATAAAATCCAGACGTACCTG
CCGGCATTCTTCATCTGTCATAACTTC

Synthetic gene of BfrCas6

GCGCCCATGGCACATATGAAAAATACACATGTCTTACTGATTAAATTTAAGAACA
AGATTAGTGACGACGAAGTTCAGTTCTTTCGTAGCTCAATCATTCAGAAGTTAGG
GGACCAACCAGATATTTTATACCATAACCATGTGGAAAAGAACAAATATCGCTA
TTCCTACCCCTTGATCCAGTATAAAAATATCGAACAGCAAGCCACAATTGTGTGC
ATTGACCAGGGTACAAAAGCCATTGAGAAGTTTTTCAGCCAGTGTGATTTCAACT
TTCAGCTTGGTAACCGCAAAGTTAATATGAAGTTCGCATCAGTGACACCCTATAA
GTTGCTGATCGAACGTCAATCACGCATGATTAATTACCATATTCATAACTGGTTG
CCCTTGAACTCTGACAACTATAAGAAGTATCAAAATATTAGTATCTTGTCAGAAC
GCATCAACTTCCTTGAGAAATCTTAGTTGGCAACATTCTGTCATTTACCAAGGG
AGTAAATTATTTCATTGACTTTCCTTTGCAGTGCAAGTTGCTGCAGCTTTCTTTTG
CCAAGTTAATTTCTAATAAAAACATCAAGTTGATGAGCTTTGATGCGGACTTCCA
ATGCAACTTGAACTTACCCGACTATATTGGTATTGGCAAACACACATCTATCGGC
TATGGGACGATCACTCGCAACTGACTCGAGGGATCCCGCG

Synthetic gene of CRISPR pre-array

CGCGCCATGGTAAAAATACAATTTTTACCCTAACTGACTGTTGTAACTTACTTTTA
TAGATTTATTCTATAATGTAGATGTATTCCAGTATAATAAGGATTAAGACGTGTC

TTCGTACCTTGAAGACCAATGTAGATGTATTCCAGTATAATAAGGATTAAGACAA

TCTTTATATATCTTATGGTTGCAGATCTAAAAAGTTGGGATTATATAAATGACAG

TCGACGCGC

Synthetic gene of BfrCorA

GCGCCCATGGCACATATGATTTATTCTTATCACATTTTCTACTTCCCGTTTAAGTG

GGAGATTATGGGGTTAGAGAACCAGGCATTTTCAGACCAGGTGAACTTAGATAA

CATTCAGTATAACCGCAACTCACATTGGGAGCGTTCCCAGAAACCTGATCCAGGC

GAAGAAGAAAGCCTGTACAATGAGAAAAATTATTACTATACGTTTGTACATAAC

ATCTTATACGACGAGGAACACTCGCCTTTGAACTTAATTCATCACTTTGAACGCA

AAGAACCGAAGTTGAGTAATCATATCTACTACTATATCAAAAAAAGGGTCGCA

ACAACCCTTATAAACTTATTGTGGACGCTATGAACATTAATCTGTACGCTACAGG

CGTAGGGTTTTTGTCGTTTTACTTGAAAAATGAGGATTGTACACAAACTCTCCC

GAGGATATTTTGGCAATCAATCAATACGGACGTCGCATTATGCCTCCATTCTTTA

ATGACACCCGTTTACGCAATGAAATTTCTGAATATATCCGTATTGAGGGGGTTGAA

TCAGACGGTTTACTTTGAGGATTTCAAATCCTATACACCCTATGATTCATGGCAA

CCGAGTTCCTCTATCAAGAAACTTATCTGTGAATTGGTTACTAATTTGTCAATCGA

TCCCATCATTGATGACCGTATGTTCGTAGCGACATGGTATAAAAATAATCAGCTT

TCACAACAATTCACTAATAACGCAAAGCATATTTTGATTCCCAAGATCCGTTCT

CCGATTATTGGTATCGTTTTTTGTTCATCGATGGCTCGAATGCCACATGTCAAAAT

GAAAAAATGAAAAAGAATTATTGGAGGAGCACACATATTATCGCTGGCAACAA

TGGTCATCCTTATACGGGATCTCAAAATATAGTCTTGTTTATTTGACGAATAATG

AGGTGCCTGATTATCTGATCGAGTATTTCCAGACAATTTATGCACGCATGGCCGA

GCTGGTCTTAGTCCAACGCGCTAGCATGTTACGTTTTTCTGGTGAGATTACTAAA

GTGTCACAATTGTCCAATCAAGATGTTGAAGCCGTGTCGAAGCGCGTCTCTAGCC

TGTACAAGGAATATATTCGTTTTGTTAATCAGATCTACTTCCGCGAGATTACCGC

GCAGGACCAGGGGATTGAGATGTACAATAAACTTCACTCGTGTTTGCAGATGGA

GTCTTATATCAAGGACTTGGATGGCGAGATCGAGGAATTACACCAGTATATTTCA

TTGATGGAAGACCGCGAACGCAACAAAAAGGCGTCCCTGTTGAATGATATTGCA

ACTCTTTTCCTTCCAATTACCGTGATTACTGGGTTTTGGGGGATGAACCAGATCTC

GGAAGTGATGGAAGAAACGGCGAGCTTAGTACAGGTTTTATTATTCAAAGTTT

ACTTCTTATTATTGGCACGCTTTGCGCGATTTGCATCATCTATAAACGCAAACGT

AAATTGTGACTCGAGGGATCCCGCG

Synthetic gene of BfrNrN

GCGCCCATGGCACATATGCAAAAACAGGCGAAAGAAATCAAGAAGCATTTGTTC
CTTTTGGGTGGTCACGATCTTGAGATGCAGACCATTGTGCAAATCTTAACAGATC
GCAACGTCATTTTCAAAGATCGTTATCTTCAATGGGACAATGCATTGTTATCGCA
ATACGAGGAAGAAATCCAACAATACGGGAATAAGGAACCATTCATTATTTATGG
CGTCGAGCTGAAAGAAGACATTACACCTCCAACCAATTACATTCGTATCGACCAC
CACAATGAGTATGCCACGTATCCAAGTGCCCTTGAACAGGTCGCGTCAATCTTAG
ACCACCCTCTGAACCGTTATCAAACACTGGTTGCTGCAAATGACAAGGCCTACAT
TCCGGGTATGCTTGAAATTGGAGCGAGCCATGAAGAGATTAACTTAATTCGCCA
GGAGGATCGCAAAGCCCAAGGCGTTATCGAGGATGATGAGAAATTGGCGCAAG
AGGCTATCACAAATGGGACTGAAAAGATTGGTAGCTTGTATGTCGTCTTTACTAC
CGCTAACAAATTTTCTCCGATCTGTGACCGTTTATATCCGTACGAGAAATTGTTG
ATTTACACTCCAAATGAGTTAATCTATTATGGAAAGGGAATCAATAGTATTCAAA
AGATCCTGAAGCGCTATACTCCAATCAGCAACATTTTTTGGGGCGGCGGGATCAA
TGGCTTTATCGGGACAGTACGCAATCGCCTGACTACGAATGAGATCTTAAATATC
GTTGAGCAGATTAAGCTGCTGGAGCTGTGACTCGAGGGATCCCGCG

Synthetic gene of BfrNYN

GCGCCCATGGCACATATGATTGAGTCAATCACGTCAATTGGAATTTTCATTGATG
GAGGCTACTTTACCAAAATCAACCAGGCTCTGGAGGAAAAATTAAGTTTGAACA
TCGATATTACTTTTTTCTTTAAGTTTATTAAAGAAAAGATCGCCTATGAGTATAAC
TTGAATACGGAGTTTTGTCAGATCACAGAATCGCATTACTTCCGTGGGCGCTACC
GCGTGAACGATGCCAACAACAAGCACCTGCTGTTTTCCGAACGCAAATTCGAAG
ACTCGCTTATTGAGAATGATGTGATTTTTCACTATAAGCACTTACGTGAGATCCA
GAAGGAAGGCGAGATCAATGTTATTGAGAAAGGCATCGATGTGTGGTTTGCTCTT
GAGGCCTATGAGCTGTCCTTATTCCGCAAATTTGACTTTGTTATTCTTATTACCGG
TGACGCGGATCATGAGATGTTGATTAAAAAATTAAAGGCATTAAAGATCCACAC
CATCTTATTAACCTGGGACCTGAGCCCAGAAAGTGCAACTGCTCGCCTTCTGCGT
GAGGAAGCATGTAAACACATTGAGCTTAGTGAGATTGCCATCGAGGACAAAGAC
CTGATTAAAAAAATTTGCCGTAGTAAACAAAGCGTTGACTCGAGGGATCCCGC
G

Synthetic gene of CboSAM-AMP lyase

GCGCCCATGGCACATATGGGGAAGACCTTACGCTTCGAGATTGTGTCGGGTGTG
AATAAGGGATATTTTCATACGAACTCACAGTCGGAATCACTGGACCTGGTAGGG
GGTATCTGGCAGAAGATCGCTAAAGAAGAATTTGAGAAATCCAATATCTACGTC
AGCGCAGTTATTAAACCCAGCAAGACTGTATATAACCAGGAGTGGGGCTGTCCC

GAGAATGGAGAAGAAACAGTGGTGTTAACTGGAGTTGCCAATGAAGAGTTCGTT
GACGATATTGAGAAATGGAAGGATACGGTAATCAAATTGGCCAAGGAGCTGAAG
AACCAAATGAAACAGTCAACGTTAACGTGCGAGTTTATCGAGACAGAATTGCAC
TACTTCAAGTGACTCGAGGGATCCCGCG

Synthetic gene of CalpS

GCGCCCATGGCACATATGTCAGGCAATGACTTGATTTTCAATTTCTTCTCTGAGA
ACGACCCTAAGGGGCTGGAAGTCATCAAAAACATTTTCTTTAAAATTATTAGCAG
CCCGACATACCAGCTTATTTTGAATTATTACGATAAAGAGGACGTATTTCAAGAG
TTTCTGGCAACAAAAATCTTACCACATCGCAATCACATTGTTGATAAGTTCTTCG
AACAGCAAAGTGGCTTGGTAAGCTATATCCAACGTATGACAAAGAACTTCCTGG
CCGATGTCTACGCTTCGGTCAAACTTATGTCGGAAAATGAGATTTCCGAGGTGAT
TATCTCCAAAGAAGAAGATGACGAGGACGAGGTAAAGTCCTACTTTGATTTAATT
GGAAAGCGCGAAAATTACACCCTTTCGATCGAAGTTGAGGAACTTAAAATTGCG
TTTACCAAGCGCTTATCAGACAATGAAATGTTAATGTTTTGCTACCAAATTTCAG
ACTCTAAAGAGCTTTATAAAAGTAAGTACTTCAATGACTTGAGTGATGATGCGCT
TTATAAACGTGTTGAGCGTATGAAGACAAAAATCAAAGAAATCTTAAAAGAATA
CTCTTTTTCAGCCGAGGCATTCGAGAAATTTTTGAAGGAACAGTCTTACGAGATT
TGCAAAAAGTTTGAGGTCAACAGCAACGGTTGACTCGAGGGATCCCGCG

Synthetic gene of truncated CalpT

GCGCCCATGGCACATATGGCGAAATGGCTGAAGGACCTGTACAACGAGTACATT
GAGGAGGAATTGGAAGAAGACTTAACGTCGCATATCAGTCGTTCTACCTTTCCTG
TGATCGGTGGTGTTTATTTCGGAAGTTTGAAATCCTTAAATAAAGAAAAACCGAA
TAAACCATTGTACTTCTTGGTACTTCGCAAGATTGACAACAACCTTTACGAGATC
ATGAAAGTAAGCGACTGGCATCACTTTGCTTCAAACACTGAAATCTTTATTGAGT
TGCCAACTATGACTTTAATCATTGAAACAACGAATAACTTTTATCTGACCTCCGA
GGAAATTTCCAAGTTCATTCTGATTGATATTCTGTCGAAGGAAGATCTGACGAAC
ATTTTGAAATTTCGCCGCGGCCACGAGATCCCTGGATTAAAGAAGGGTTTCACAC
CAATCTTCGAGGATGACATCCGCAACAAGTTCAAGAAAGAGGAATTCAATCAGA
TTAAAGAGTTCCACACCCGTATTTTTGAAATCCTGGCAGAATGACTCGAGGGATC
CCGCG

Sequence of plasmid pET11a-CalpT (The sequence of CalpT was highlighted with underline)

TTTTGCTGAAAGGAGGAACTATATCCGGATATCCCGCAAGAGGCCCGGCAGTAC
CGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGAC
GATGAGCGCATTGTTAGATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAAC

TGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAAT

TCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGA

TAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAAC

CCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAAT

AACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAAC

ATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTC

ACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAG

TGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCC

CGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTA

TTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTC

AGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCA

TGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGG

CCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCA

CAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGA

AGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAAC

GTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTA

ATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTT

CCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCG

GTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTA

CACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGAT

AGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATA

CTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCT

TTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGT

CAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGT

AATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG

GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG

ATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACT

CTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGC

CAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGAT

AAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAG

CGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCC

ACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGG

AACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAG

TCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAG

GGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGG
CCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGG
ATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGAC
CGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTT
TCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTAC
AATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTG
ACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGAC
GGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAG
CTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCG
GTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCC
GCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGC
GGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGG
GGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCAC
GATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAA
ACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATG
CCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCT
GCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGA
CTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGAC
GTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTA
ACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGAT
CATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAGATGCGCCGCGTGCGGCT
GCTGGAGATGGCGGACGCGATGGATATGTTCTGCCAAGGGTTGGTTTGCGCATTC
ACAGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAG
CGAGGTGCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCGGCTCCATGCACCGCG
ACGCAACGCGGGGAGGCAGACAAGGTATAGGGCGGCGCCTACAATCCATGCCA
ACCCGTTCCATGTGCTCGCCGAGGCGGCATAAATCGCCGTGACGATCAGCGGTCC
AGTGATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGATCCTTGAAGCTGTCCCTG
ATGGTCGTCATCTACCTGCCTGGACAGCATGGCCTGCAACGCGGGCATCCCGATG
CCGCCGGAAGCGAGAAGAATCATAATGGGGAAGGCCATCCAGCCTCGCGTCGCG
AACGCCAGCAAGACGTAGCCCAGCGCGTCGGCCGCCATGCCGGCGATAATGGCC
TGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGG
GCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAG
CGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACG
AGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGC

GCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGAT
CCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCG
CTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGC
GGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGA
GACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAA
GCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAAC
GGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATAT
CCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCA
TCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTG
CATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATC
GGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGC
GCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAAT
GCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATA
CTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTG
CAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCA
GCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGAC
GCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGA
GATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTG
GCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGG
GAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAA
ACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCA
TACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACT
CTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTG
TCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGT
AGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAG
ATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAA
CAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCG
GCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATG
CGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTAT
AGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTT
AAGAAGGAGATATACAT<u>ATGGCCAAATGGCTGAAAGATCTGTATAATGAATACA</u>
<u>TTGAGGAGGAACTGGAAGAAGATCTGACCAGTCATATTAGTCGCAGCACCTTTC</u>
<u>CGGTTATTGGTGGTGTTTATTTTGGTAGCCTGAAAAGTCTGAATAAGGAAAAACC</u>
<u>GAATAAGCCGCTGTATTTTCTGGTTCTGCGCAAAATTGATAATAATCTGTATGAA</u>

ATCATGAAGGTGAGTGATTGGCATCATTTTGCCAGCAATACCGAAATTTTTATTG
AACTGCCGACCATGACCCTGATTATTGAAACCACCAATAATTTTTACCTGACCAG
CGAAGAAATTAGTAAATTCATTCTGATCGACATCCTGAGCAAAGAAGATCTGAC
AAATATTCTGAAATTCCGCCGTGGCCATGAAATTCCGGGCCTGAAAAAAGGTTTT
ACCCCGATTTTTGAAGATGATATTCGCAACAAATTCAAAAAGGAAGAATTCAAC
CAGATCAAAGAATTTCATACCCGCATTTTTGAAATCCTGGCCGAACCGGAAGAA
CAGGTTATTGAAATTGCACCGGAACGTATTAGTGAATTTGTGCTGCGCCATGTTG
CCAGTACCAGTCAGAAAGCAACCTATACCGATGATTTTGTTCTGTATCGTGGCGA
TGATTTTATTGAAATTATCATTGACGAGAAGTACCTGAATAAGAAAGTGAAAATT
CTGCTGGATAACGATACCATTTTTAATGGCATTCTGAAAGATACCAGTATTTTTAT
TCCGGTGAAAGAACAGATTGATCTGGAAGAACTGGCCAAACATATTAGCATTCT
GCCGGAAGGTTAAGGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTG
GCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGG
GTCTTGAGGGGTT

Amino acid sequence of BfrCas6

MKNTHVLLIKFKNKISDDEVQFFRSSIIQKLGDQPDILYHNHVEKNKYRYSYPLIQYK
NIEQQATIVCIDQGTKAIEKFFSQCDFNFQLGNRKVNMKFASVTPYKLLIERQSRMIN
YHIHNWLPLNSDNYKKYQNISILSERINFLEKILVGNILSFTKGVNYFIDFPLQCKLLQ
LSFAKLISNKNIKLMSFDADFQCNLNLPDYIGIGKHTSIGYGTITRN

Amino acid sequence of BfrCmr1

MNQLTAILKQHTPMIHFQHNESGATLRASEVKPLLDKFILTKLGNGDIREGRLYAKK
NNWLIDNEKNYALNYKLSISLQKKSRLEYLITSSTFPLPTERPSNFFTIQNSPYFAQEK
CVGINTNSTIILKKSNSDPRKKEAEFKEKNWSQIDKKGLEWQDFTIKIFSLKGDLINKI
QTYLPAFFICHNFGTRNNKGFGSFTVEYINNQKNICNVEDTLKENFAFVYKKKIALSR
QSTLDFIYIYNQIFSTIKKDYQILKSGYNFRNEYIKSLLFCYFVSKYPNYRWEKRKMK
QLIKARGYELKGDHSPISGIRENDNSWNDPNPNGYNYAYIRAILGLAEQYEFQLETPY
QKAIVKIKSANNCISRYKSPLLFKIINNSIYLVGNEINTEILNKPFQYSYIEQTKNKNMR
TGKSEITERTMHINEIEMNYKNRINYHYTPTSFSLIDFMQYAMSYKKNGKNILNYIPL
KQ

Amino acid sequence of BfrCmr2

MKYIAITLGPITRTIEMAESTKELWAASYFFSYLAKKIVEPFVKKNRTFQLPLINEEMQ
KPHCGAGLFPDRYIFKSEPGDLELLKQHSDQVLIEIAGHIASPSLPGTAKDVSQIYHYL
KSYIKIYFIERTLESDDPHVVIPACEKYLNIIENQETFPEQEETMISHQKSDFLKFLITNV
NGKIYRKDKNSIPRFTGSFLTRDAFGDMNGERLFESILEISASELNINIQQKALEVITAN

EKNKGEKYSDQIWDAEEIILNDNKAQLRPYHKYIAIIKSDGDSMGETIKSMGAYNIPIT
QLSKALLSFNIESINEIVAYGGKPIFIGGDDLLCFAPVCCNGNNVFNLVEKLSTCFDQC
INQHLQQYINACSEAQRPLPSLSFGISITYHKYPMFEALHTTDYLLEMVAKDNLFKYT
LSNKNILNENMKRFILKNKLAFSLQKHSGQIYHTAMSKKGKSYVKFNMLLQKYILK
NKDMSKTQESEKFLSSVIQMIRAHAEILQIILQNEDKRTEMLKNYFDNNFNESCHLGY
TGLFEDIQTLLCLRYQENIQDYQNRNEIIQQNTILTSDEKEILIVSPAMDAIHTIFTALQF
IHFINYNKDE

Amino acid sequence of BfrCmr3

MNRHYLITLTPMDWFFFGGERTLDDGKSADYISHSNKFPQQSALLGMIRYQLLKQH
NLLSQFPYTENKPTEKEIMKTLIGEQSFRMTERKAKSLGLGVIKQISPLMLIECKDDTS
SRSIYFPLPLDDGYKVSFNETSNEDKVFYNGIECPIPNVYPASEEQDSGNQKRKFFDH
KTYNNYLFWCTQGNNQIKKLLSDEIWISKMQIGITKHVEEGEDNDKSFYKQEFLQLK
KSFIYAFYITLSGESELSSDIIQLGGQRSVFRMEVESIEENSDIQEKYQTAAQFLTQSDR
LLILSPTYVDNLKELSALCNFMWSDSIVFRNIQTTNASNFYGKPIKSSSKYHFLKPGSV
LYFKQGKRKEVEKLLMDYTYLRLSGYNIYI

Amino acid sequence of BfrCmr4

MTTRMYVINTLSNMHVGSGEVNYGVIDNLIQRDSVTNLPNINSSGLKGAIREYFKEN
EDLVRELFGSAPRDEKTLPGKVRFFEANLLSMPVRSDKVPFLMAISDEVLQELITKM
KFFNCEEATQYISHLSTLLDNIKTQAQGTDFAYVFDPLLQGAIIEEVSIRATCPSHIPLQ
PSLKKLLGDRLVILSHKYFSILSDDNHLPVLSRNNLENGQSANLWYEQVLPRYSRLYF
MLMDGNAQSEYLKKFRDTLCTPSTIIQIGANASIGYGYCQISELSPF

Amino acid sequence of BfrCmr5

MKISKKQIEYAIEALRANNIITNDNQYPKVFKGYISSFGAAVIQSGLIPAIIFFENEDND
ANADRHKIIGVLKDIINAMRQQYTVTDATILVSSQIPANYSMAQYIIEHGNTDQLLKEI
TEAAVAMKLALRMYKSE

Amino acid sequence of BfrCmr6

MPKNYTLQNASNLGWLFYKDYYRQEPNVDFISTQGKESDTTADFFRKTNQRITAYQ
LNSESPLVAAFNNHFGTPLQLKTIYPGLITGSGLPHQTGSKGEFKLGFQFDYTTGLPYI
PGSSIKGTLRSMFPFSLKDKGSTKRILPEYRKERMEYIRDLIIEVTNINEISDTEIQALEY
AIFTNSTPSGKTIEFSLEEKDVFYDAFVADSKDGVMLSDDYITPHGENPLKDPKPILFL
KIRPDVTINFYFKLCTTHLYKEKVCSSKQIEEIKKQNDFSSSDYKMITAHQKRNLFEKI
LLCIGIGAKTNIGYGQLKKL

Amino acid sequence of BfrNrN

MQKQAKEIKKHLFLLGGHDLEMQTIVQILTDRNVIFKDRYLQWDNALLSQYEEEIQQ
YGNKEPFIIYGVELKEDITPPTNYIRIDHHNEYATYPSALEQVASILDHPLNRYQTLVA
ANDKAYIPGMLEIGASHEEINLIRQEDRKAQGVIEDDEKLAQEAITNGTEKIGSLYVV
FTTANKFSPICDRLYPYEKLLIYTPNELIYYGKGINSIQKILKRYTPISNIFWGGGINGFI
GTVRNRLTTNEILNIVEQIKLLEL

Amino acid sequence of BfrCorA

MIYSYHIFYFPFKWEIMGLENQAFSDQVNLDNIQYNRNSHWERSQKPDPGEEESLYN
EKNYYYTFVHNILYDEEHSPLNLIHHFERKEPKLSNHIYYYIKKKGRNNPYKLIVDA
MNINLYATGVGFLSFYLKNEDCTQNSPEDILAINQYGRRIMPPFFNDTRLRNEISEYIR
IEGLNQTVYFEDFKSYTPYDSWQPSSSIKKLICELVTNLSIDPIIDDRMFVATWYKNNQ
LSQQFTNNAKAYFDSQDPFSDYWYRFLFIDGSNATCQNEKMKKELLEEHTYYRWQ
QWSSLYGISKYSLVYLTNNEVPDYLIEYFQTIYARMAELVLVQRASMLRFSGEITKVS
QLSNQDVEAVSKRVSSLYKEYIRFVNQIYFREITAQDQGIEMYNKLHSCLQMESYIK
DLDGEIEELHQYISLMEDRERNKKASLLNDIATLFLPITVITGFWGMNQISEVMEENG
ELSTGFIIQSLLLIIGTLCAICIIYKRKRKL

Amino acid sequence of CboSAM-AMP lyase

MGKTLRFEIVSGVNKGYFHTNSQSESLDLVGGIWQKIAKEEFEKSNIYVSAVIKPSKT
VYNQEWGCPENGEETVVLTGVANEEFVDDIEKWKDTVIKLAKELKNQMKQSTLTC
EFIETELHYFK

Amino acid sequence of BfrNYN

MIESITSIGIFIDGGYFTKINQALEEKLSLNIDITFFFKFIKEKIAYEYNLNTEFCQITESH
YFRGRYRVNDANNKHLLFSERKFEDSLIENDVIFHYKHLREIQKEGEINVIEKGIDVW
FALEAYELSLFRKFDFVILITGDADHEMLIKKLKALKIHTILLTWDLSPESATARLLRE
EACKHIELSEIAIEDKDLIKKICRSKQKR

Amino acid sequence of CalpT

MAKWLKDLYNEYIEEELEEDLTSHISRSTFPVIGGVYFGSLKSLNKEKPNKPLYFLVL
RKIDNNLYEIMKVSDWHHFASNTEIFIELPTMTLIIETTNNFYLTSEEISKFILIDILSKED
LTNILKFRRGHEIPGLKKGFTPIFEDDIRNKFKKEEFNQIKEFHTRIFEILAEPEEQVIEIA
PERISEFVLRHVASTSQKATYTDDFVLYRGDDFIEIIIDEKYLNKKVKILLDNDTIFNGI
LKDTSIFIPVKEQIDLEELAKHISILPEG

Amino acid sequence of CalpS

MKKDLFRKELLDYIVNNKVSEKFLNSVKGIVISIISKNKTYQTGIKACYGSIEDAINDI
LNDILIKIKNKAHIFKNLSDNHGAYLYTMIKNHIVDVLRNYRFNISLDNESDDFENRIE

YFLHSDDLNDSFESVIVSQYFFKELKKINDKYLCFYLYKVLYSEEICFSEKTKDAKYK
INQRTKEKLKELVQENGVTEKEFLLAIRIYMSEICEKLRNNK

Article

Antiviral type III CRISPR signalling via conjugation of ATP and SAM

Haotian Chi[1], Ville Hoikkala[1,2], Sabine Grüschow[1], Shirley Graham[1], Sally Shirran[1] & Malcolm F. White[1✉]

CRISPR systems are widespread in the prokaryotic world, providing adaptive immunity against mobile genetic elements[1,2]. Type III CRISPR systems, with the signature gene *cas10*, use CRISPR RNA to detect non-self RNA, activating the enzymatic Cas10 subunit to defend the cell against mobile genetic elements either directly, via the integral histidine–aspartate (HD) nuclease domain[3–5] or indirectly, via synthesis of cyclic oligoadenylate second messengers to activate diverse ancillary effectors[6–9]. A subset of type III CRISPR systems encode an uncharacterized CorA-family membrane protein and an associated NrN family phosphodiesterase that are predicted to function in antiviral defence. Here we demonstrate that the CorA-associated type III-B (Cmr) CRISPR system from *Bacteroides fragilis* provides immunity against mobile genetic elements when expressed in *Escherichia coli*. However, *B. fragilis* Cmr does not synthesize cyclic oligoadenylate species on activation, instead generating *S*-adenosyl methionine (SAM)-AMP (SAM is also known as AdoMet) by conjugating ATP to SAM via a phosphodiester bond. Once synthesized, SAM-AMP binds to the CorA effector, presumably leading to cell dormancy or death by disruption of the membrane integrity. SAM-AMP is degraded by CRISPR-associated phosphodiesterases or a SAM-AMP lyase, potentially providing an 'off switch' analogous to cyclic oligoadenylate-specific ring nucleases[10]. SAM-AMP thus represents a new class of second messenger for antiviral signalling, which may function in different roles in diverse cellular contexts.

Bacteroides spp. are Gram-negative, anaerobic bacteria that constitute a significant portion of the human gut microbiome[11]. The *Bacteroidales* are host to the most widespread and abundant phage found in the human digestive system, CrAssphage[12]. *B. fragilis* is an opportunistic pathogen, and is responsible for more than 70% of *Bacteroides* infections[13]. Bioinformatic analyses have revealed the presence of three CRISPR types—I-B, II-C and III-B—in *B. fragilis* strains, with the type III-B system being the most common[14]. Sequence analysis shows that *B. fragilis* Cas10, the main enzymatic subunit of type III effectors, lacks an HD nuclease domain but has an intact cyclase domain, similar to the *Vibrio metoecus* Cas10[15]. This suggests that the system functions via cyclic oligoadenylate (cOA) signalling to associated ancillary effectors. In *B. fragilis* and more generally in the *Cytophaga–Bacteroides–Flavobacterium* bacterial phylum these type III CRISPR systems are strongly associated with an uncharacterized gene encoding a divergent member of the CorA-family of divalent cation channel proteins[16,17] (Fig. 1a). The CRISPR-associated CorA proteins have not been studied biochemically but are predicted to consist of a C-terminal membrane spanning helical domain fused to a larger N-terminal domain with a unique fold. To investigate this further, we first generated a phylogenetic tree of Cas10 proteins and identified those associated with a gene encoding the CorA protein. Three phylogenetically distinct clusters of CorA-associated type III CRISPR systems were apparent,

with the largest (CorA-1) being associated with type III-B systems (Fig. 1a).

The genomic context of CorA-containing type III CRISPR loci from cluster CorA-1 (Fig. 1b) reveals that the *corA* gene is typically found next to a gene encoding a phosphodiesterase (PDE)—the DHH-family nuclease NrN in the case of *B. fragilis* and *M. vanielii*, and a DEDD-family nuclease in the case of *S. oralis* and *S. lipocalidus*. In the genome of *A. butzleri* and related species, the *nrn* and *corA* genes are fused, suggestive of a close functional relationship. The closest predicted structural matches for *B. fragilis* NrN are to the pGpG-specific PDE PggH from *Vibrio cholerae*, which has a role in the turnover of the cyclic nucleotide c-di-GMP[18] and the GdpP PDE from *Staphylococcus aureus*, which degrades pApA molecules as a component of c-di-AMP signalling systems[19]. Analysis of the DEDD protein suggests structural matches to RNaseT[20], oligoribonuclease[21], and the mammalian REXO2 protein, which degrades linear RNA and DNA dinucleotides[22]. Thus, the CRISPR-associated NrN and DEDD proteins appear to be homologous to protein families that degrade small RNA and DNA species. Of note, in some CorA-containing type III systems including *Clostridium botulinum*, the PDE is replaced by a protein that is predicted to resemble a family of phage SAM lyase enzymes involved in evasion of host immune systems[23,24] (Fig. 1b).

[1]Biomedical Sciences Research Complex, School of Biology, University of St Andrews, St Andrews, UK. [2]University of Jyväskylä, Department of Biological and Environmental Science and Nanoscience Center, Jyväskylä, Finland. ✉e-mail: mfw2@st-andrews.ac.uk
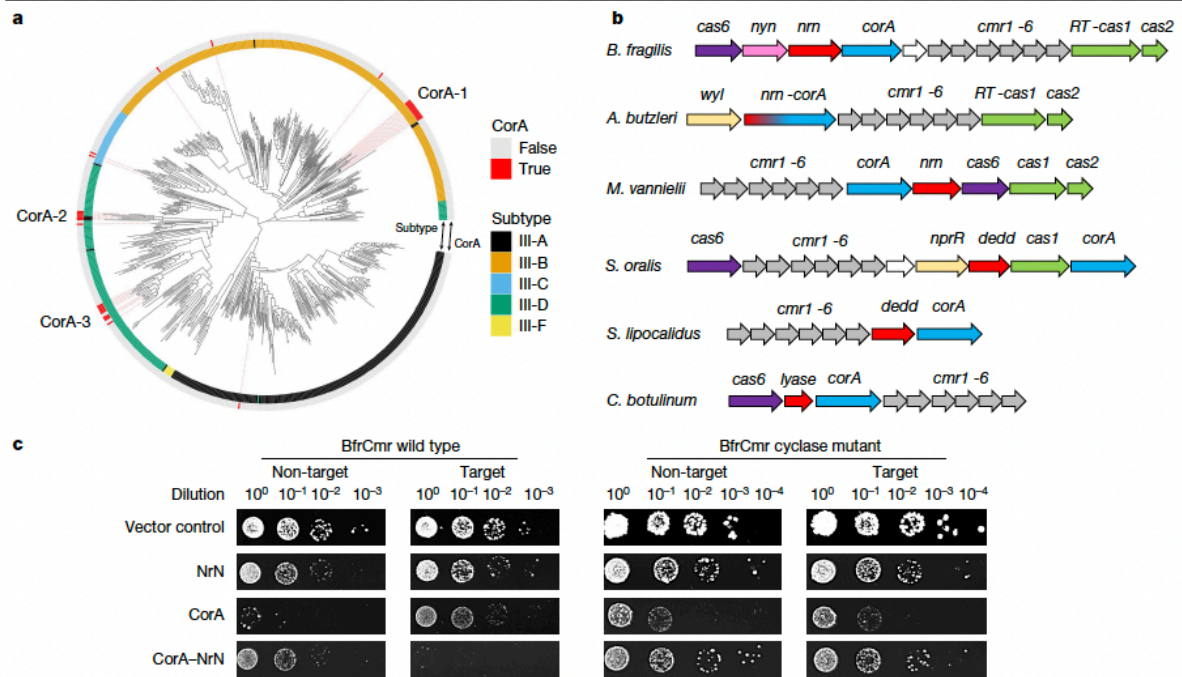
Fig. 1 | Type III CRISPR systems with a CorA effector. a, A phylogenetic tree of Cas10 proteins from type III CRISPR systems of complete bacterial and archaeal genomes, colour coded by subtype[42]. Red bars on the outer ring indicate systems associated with a CorA-family effector protein. There are three main clusters of CorA-associated Cas10s, labelled CorA-1, CorA-2 and CorA-3. **b,** Genome context and effectors of selected type III-B CRISPR systems with a *corA* gene (cluster CorA-1: *B. fragilis, Aliarcobacter butzleri, Methanococcus vannielii, Streptococcus oralis, Snytrophothermus lipocalidus, Clostridium botulinum*). The type III-B *cas* genes *cmr1–6* are shown in grey, with *cas6* in purple and the adaptation genes *cas1* (or a gene encoding a fused reverse transcriptase–Cas1 protein) and *cas2* in green. The putative membrane channel protein is encoded by the *corA* gene

(blue), which is adjacent to or fused with the genes encoding the PDEs NrN or DEDD (red). In *C. botulinum*, the PDE is replaced with a predicted SAM lyase. The *wyl* and *nprR* genes encode predicted transcriptional regulators. **c,** Plasmid challenge assay. *E. coli* BL21 Star cells expressing *B. fragilis* Cmr (wild-type or cyclase-defective variant) programmed with target (tetR) or non-target (pUC19) CRISPR RNA (crRNA) species were transformed with a pRAT plasmid that expressed the NrN and/or CorA proteins and carried a tetracycline resistance gene. Resistance was observed only when a targeting crRNA, active cyclase and both effector proteins were all present. Raw data are presented in Supplementary Data Fig. 1.

B. fragilis Cmr is active in vivo

To investigate the activity of the *B. fragilis* type III CRISPR system, two plasmids were constructed. Plasmid pBfrCmr1-6, built using Gibson assembly[25], expresses synthetic versions of the codon-optimized genes *cmr1–6* and plasmid pBfrCRISPR encodes Cas6 and a mini-CRISPR array (Extended Data Fig. 1a). We expressed the complex in *E. coli* with a targeting (pCRISPR-Tet) or non-targeting (pCRISPR-pUC) crRNA and challenged cells by transformation with a pRAT-Duet plasmid expressing one or both of the CorA and NrN effector proteins (Fig. 1c). The pRAT-Duet plasmid also contains the *tetR* gene for activation of *B. fragilis* Cmr carrying the targeting crRNA. Cells were transformed with the pRAT-Duet vectors and grown in the presence of tetracycline to select for transformants. We included vectors expressing wild-type and cyclase-defective (Cas10 D328A/D329A variant) Cmr for comparison. We previously used this experimental design to investigate the *V. metoecus* Cmr system[15]. In conditions in which the Cmr system was activated and had the required ancillary effector proteins, lower numbers of colony-forming units were expected. The vector control (no effectors) served as a baseline for transformation. When only the NrN effector was present, no reduction in colonies was observed, suggesting no active targeting. When only the CorA protein was expressed, fewer colonies were observed in both target and non-target conditions, for both the wild-type and cyclase-deficient mutant (Δcyclase) Cmr,

suggesting some toxicity of the CorA protein. When both the CorA and NrN effectors were expressed, immunity—indicated by a markedly reduced number of colonies—was observed only for the wild-type Cmr system with *tetR* targeting. Immunity was lost when wild-type NrN was substituted with a variant mutated in the DHH active site motif (D85A/H86A/H87A), or when the CorA protein was truncated to remove the transmembrane domain (Extended Data Fig. 1b).

These data suggest that the *B. fragilis* Cmr system is functional in *E. coli* and requires the activity of the Cas10 cyclase domain and the presence of both effector proteins. The toxicity of CorA appears to be reduced by the presence of the NrN protein, regardless of activation of the type III system, suggestive of a strong functional link. Intriguingly, the type III-B complex from *Mycobacterium tuberculosis*, which synthesizes cyclic oligoadenylate 3–6 (cA$_{3-6}$) in vitro[26], did not provide immunity when combined with the NrN and CorA effectors, hinting at a non-canonical activation mechanism (Extended Data Fig. 1c). The strict requirement for the SAM-AMP degrading NrN protein in addition to CorA for plasmid immunity is an unusual aspect of the system and is discussed further below.

RNA processing and degradation

Co-transformation of the expression plasmids into *E. coli* strain BL21 (DE3) enabled the expression of the *B. fragilis* Cmr effector and

purification by immobilized metal-affinity and size-exclusion chromatography (Extended Data Fig. 2a). We also purified *B. fragilis* Cas6 individually using the same chromatography steps. We first confirmed that Cas6 processed crRNA in the expected manner. The recombinant Cas6 enzyme cleaved synthetic fluorescein (FAM)-labelled crRNA at the base of a predicted hairpin with a 2-bp stem, reminiscent of *Methanococcus maripaludis* Cas6b[27]. This generates a canonical 8-nucleotide (nt) 5′ handle, (Extended Data Fig. 2b). Cleavage of an in vitro transcript comprising 2 repeats flanking one spacer generated the expected set of reaction products, culminating in a processed crRNA of 72 nt (Extended Data Fig. 2c). To investigate the composition of the crRNA present in the effector complex purified from *E. coli*, we isolated and labelled the crRNA using γ-^{32}P-ATP and polynucleotide kinase. This revealed 3 major crRNA species differing in length by 6 nt (Extended Data Fig. 3a,b). These products correspond to 3′ end trimming of the crRNA to remove the repeat-derived sequence and probably reflect effector complexes that differ in the number of Cas7 subunits and thus length of backbone, as has been seen for other type III systems (reviewed in ref. 28).

Type III CRISPR systems also cleave bound target RNA using the Cas7 subunit, either for direct defence against mobile genetic elements[29] or for regulatory purposes[8]. We proceeded to test for cleavage of target RNA bound to the crRNA in the effector. The 5′-end-labelled target RNA was cleaved at 4 positions with 6-nt spacing, corresponding to the placement of the Cas7 active sites in the backbone[29]. Cleavage was extremely rapid and was essentially complete after 2 min, the first time point (Extended Data Fig. 3c). As these sites interconvert and site 1 is furthest from the 5′ label, this cleavage was only observed for the Cmr4 or Cas7 D27A variant, which cleaves target RNA more slowly (Extended Data Fig. 3d). We also observed cleavage of target RNA at the boundary of the crRNA–target RNA duplex. This activity, which has not been observed for other type III systems, appeared to be due to the Cmr4 subunit, as it was not observed for the D27A variant. As target RNA cleavage has been shown to correlate with the deactivation of the Cas10 subunit[8,30], this suggests that the Cmr complex remains active for a very short time after target RNA binding. This groups *B. fragilis* Cmr together with the type III effectors from *Streptococcus thermophilus* and *Thermotoga maritima*, which cleave target RNA rapidly[5,30]. By contrast, the type III systems from *S. solfataricus* and *V. metoecus* have much slower RNA cleavage kinetics[8,15]. In the absence of structural data to define the number and positions of Cas7 subunits in the complex, we could not analyse the cleavage pattern further with any degree of certainty.

Identification of the signalling molecule

As *B. fragilis* Cmr lacks a HD nuclease domain in the Cas10 subunit, immune function would be expected to be mediated by the cyclase domain via the generation of nucleotide second messengers. However, although the system provided cyclase-dependent immunity in vivo, activation of the wild-type Cmr in vitro resulted in very low yields of any observable product when incubated with ATP, in contrast to the Cmr complex from *V. metoecus*, which synthesizes cA[3][15]. This hinted at the possibility that a vital component was missing in the in vitro assays. Accordingly, we activated *B. fragilis* Cmr in *E. coli* using a plasmid to express target RNA and then processed cell lysates to allow isolation of nucleotide products. These were purified and analysed by high-performance liquid chromatography (HPLC). A prominent peak was observed following HPLC of extracts with activated Cmr, which was absent in the absence of target RNA or when the cyclase activity was knocked out by mutagenesis (ΔCy) (Fig. 2a). Mass spectrometry yielded a *m/z* value of 728.196 for the positive ion; to our knowledge, this *m/z* value did not correspond with any known cyclic nucleotide or indeed any other previously characterized metabolite (Fig. 2b). To identify the product, we fragmented the purified molecule using

tandem mass spectrometry (MS/MS). This enabled the identification of fragments characteristic of AMP and methionine (Fig. 2c). Further examination suggested that the molecule under study was SAM that was adenylated on the ribose moiety (Fig. 2d), a molecule that we hereafter refer to as SAM-AMP. To our knowledge, SAM-AMP has not previously been described in the literature—from either chemical or enzymatic synthesis perspectives—suggesting that it is a previously undiscovered class of signalling molecule.

To confirm that *B. fragilis* Cmr synthesized SAM-AMP, we reconstituted the reaction in vitro with ATP and SAM, analysing reaction products by HPLC and thin-layer chromatography (TLC) (Fig. 2e,f). We observed SAM-AMP production when SAM and ATP were present in vitro. *S*-adenosyl-L-homocysteine (SAH) and the SAM analogue sinefungin[31], which differ at the sulfur centre, were also conjugated with ATP by Cmr (Fig. 2e,f and Extended Data Fig. 4c). No significant products were observed in the presence of ATP or all four ribonucleotides. The synthesis of SAM-AMP and SAH-AMP by *B. fragilis* Cmr were consistent with rapid, multiple-turnover kinetics that were essentially complete within the first 2 min of the reaction (Extended Data Fig. 4a,b). The observation of only SAM-AMP, and not SAH-AMP, in *E. coli* cell extracts is probably the result of the much higher concentration of SAM than SAH in *E. coli*[32] (0.4 mM versus 1.3 μM). Overall, these data provide strong evidence that the *B. fragilis* Cmr system generates a previously undescribed conjugate of SAM and ATP, rather than cOA.

Since Cas10 family enzymes synthesize 3′–5′ phosphodiester bonds[6,7], we considered it likely that SAM was fused to AMP at the 3′ position on the ribose ring, but the mass spectrometry data did not rule out a 2′–5′ phosphodiester bond. To address this, we incubated SAM-AMP and SAH-AMP with nuclease P1, which is specific for 3′–5′ phosphodiester bonds. Whereas SAH-AMP was completely degraded by nuclease P1, we observed only partial degradation of SAM-AMP (Extended Data Fig. 5). Thus, although we consider a 3′–5′ phosphodiester linkage to be likely, we cannot rule out a 2′–5′ linkage completely. Final confirmation of the linkage will require further analysis—for example, by NMR.

The crystal structure of *Pyrococcus furiosus* Cas10–Cas5 bound to two ATP molecules[33] shows one ATP in the 'donor' ATP1 site next to the GGDD cyclase catalytic motif and another in the 'acceptor' site (Fig. 3a). For the enzymes that synthesize SAM-AMP, SAM must bind in the acceptor ATP2 binding site, next to ATP1 in the donor site[34]. This arrangement would allow nucleophilic attack from the 3′-hydroxyl group of SAM to the α-phosphate of ATP1, resulting in the formation of a 3′–5′ phosphodiester bond linking SAM with AMP, and the release of pyrophosphate. The reaction chemistry is essentially the same as the one that takes place in canonical type III CRISPR systems that synthesize cOA species[34]. The major difference is that the triphosphate of ATP2 in the acceptor site is replaced by the methionine moiety of SAM, resulting in a change in local charge of the ligand from net negative to net positive, raising the possibility that Cas10s binding SAM will have a less basic binding site in this area. Examination of sequence conservation in the Cas10s associated with a CorA-1 cluster (Extended Data Fig. 6) revealed the presence of two absolutely conserved acidic residues, D70 and E151. Modelling of the *B. fragilis* Cas10 structure places these two residues in the vicinity of the methionine moiety of SAM in the acceptor site (Fig. 3b). D70 occupies the position equivalent to N300 in *P. furiosus* Cas10, which neighbours the β-phosphate of the acceptor ATP2 ligand, while E151 is in a similar position to R436, which forms a bidentate hydrogen bond with the γ-phosphate (Fig. 3a). We created variants of Cas10 with D70N, E151R and D70N/E151R mutations, which were expressed and purified as for the wild-type protein (Extended Data Fig. 2a), and assessed them for their ability to synthesize nucleotide products (Fig. 3c). The E151R variant had a limited effect on SAM-AMP synthase activity, but the D70N variant was significantly compromised and the double mutant showed no detectable
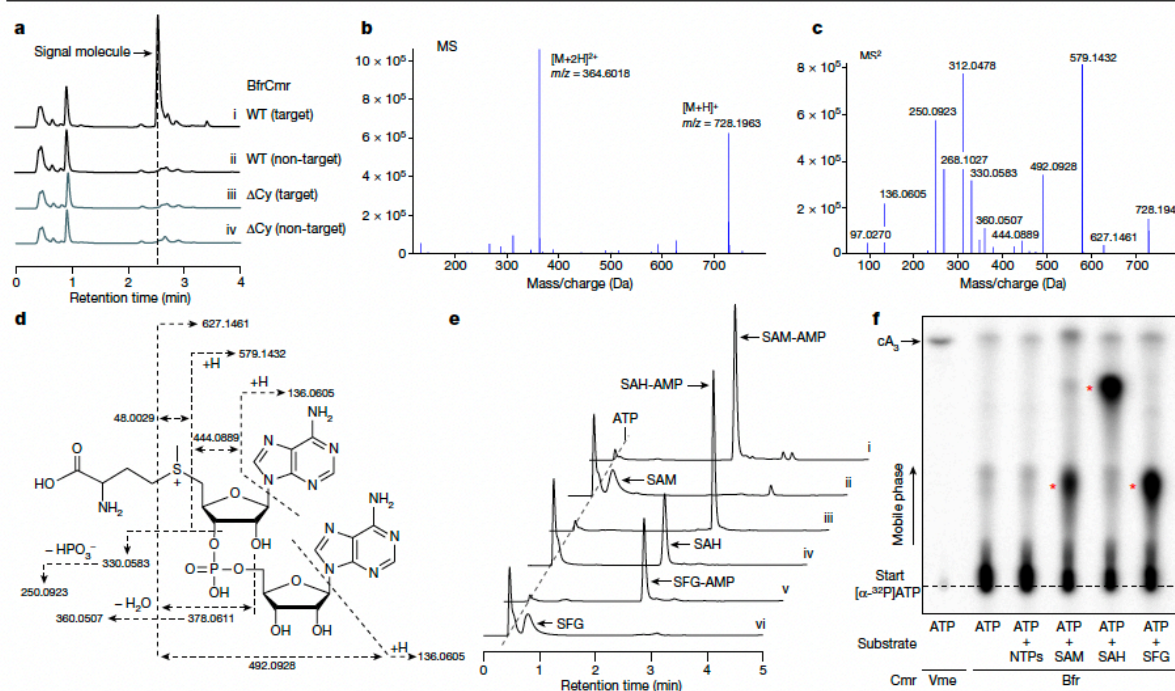
Fig. 2 | Identification of SAM-AMP from cells containing the activated Cmr complex. a, HPLC analysis of *E.coli* extracts expressing the wild-type (WT) or mutant (ΔCy) *B. fragilis* Cmr system with target or non-target crRNA. The putative signal molecule was only observed for the activated system (trace i). **b**, Characterization of the extracted signal molecule by liquid chromatography–mass spectrometry (LC–MS) in positive mode. [M + H]+ and [M + 2H]2+ represent two different ionization forms. **c**, MS/MS analysis of the signal molecule with *m/z* 728.1963. **d**, The proposed structure of the signalling molecule, whose fragmentation pattern is shown by dashed arrows. The MS/MS data cannot distinguish between 2′–5′ and 3′–5′ phosphodiester bonds.

The 3′–5′ phosphodiester bonds are more likely and is shown here, but a 2′–5′ bond cannot be completely ruled out. **e**, HPLC analysis of compounds synthesized by the purified wild-type *B. fragilis* Cmr complex in vitro. Cmr synthesizes the signal molecule SAM-AMP from ATP and SAM (trace i). Cmr also accepts SAH and sinefungin (SFG) as substrates (traces iii and v, respectively). Traces ii, iv and vi are control reactions. **f**, TLC analysis of in vitro reaction products. SAM, SAH and sinefungin plus ATP yielded radioactive products (red stars) but ATP alone did not. cA3 generated by wild-type *V. metoecus* Cmr complex[15] is shown for comparison. Uncropped HPLC and TLC data are presented in Supplementary Data Fig. 2.

activity (Fig. 3c,d). Moreover, the double mutant displayed an enhanced pppApA synthase activity when compared with the wild-type enzyme, suggesting a partial reversion of the acceptor binding site to favour ATP over SAM (Extended Data Fig. 7). A deeper understanding of the reaction mechanism and substrate specificity of the SAM-AMP synthases will require structural data in the presence of ligands, and could also involve discrimination by the Cas5 subunit, which is in the vicinity of the ATP2 ligand.

SAM-AMP signalling and turnover

To test the suggestion that SAM-AMP is the activator of the CorA effector, we expressed *B. fragilis* CorA in *E. coli* and purified the protein to near homogeneity in the presence of detergent (Extended Data Fig. 8a). CorA was incubated with radiolabelled SAM-AMP, SAH-AMP or cA3 and then analysed by native gel electrophoresis. A clear shifted species, close to the wells of the gel, was observed to accumulate as the CorA protein was incubated at increasing concentrations with SAM-AMP or SAH-AMP (Fig. 4a). By contrast, cA3 was not shifted. These data support a model where CorA binds the SAM-AMP second messenger to provide immunity. To investigate this in more detail, we generated a model of the pentameric *B. fragilis* CorA structure (Fig. 4b and Extended Data Fig. 8) and mapped the positions of conserved residues in the CorA-1 clade identified from a multiple sequence alignment (Supplementary

Data Fig. 10). A cluster of conserved residues at the interdomain interface hinted at a putative SAM-AMP-binding site. To test this, we created two site-directed variants of CorA by mutating two pairs of conserved residues (R152–R153 and D219–D220) in this cluster to alanine. The variant proteins were expressed similarly to the wild-type CorA, but no longer provided immunity in the plasmid challenge assay (Extended Data Fig. 8). Although these observations are consistent with a role in SAM-AMP binding, we cannot rule out the possibility that these mutations alter the quaternary structure of the protein. Although the mechanism of the CorA effector has not yet been determined, it most probably functions as a SAM-AMP-activated membrane channel, analogous to the Csx28 protein associated with Cas13[35] and to a number of other predicted membrane proteins associated with type III CRISPR systems[16]

As described previously, most type III systems with a CorA effector also encode a PDE of the NrN or DEDD family[16]. We therefore incubated the *B. fragilis* NrN protein with SAM-AMP and observed that it specifically degrades SAM-AMP (Fig. 4c), but not the linear dinucleotide pApA, or cOA molecules cA2–6 (Extended Data Fig. 9a). One possibility is that specialized NrN and DEDD-family PDEs represent a type of 'off switch' to reset the system, analogous to the ring nucleases that degrade cOA molecules in canonical type III CRISPR systems[10]. In the *Clostridia*, the NrN protein is replaced with a predicted SAM lyase (Fig. 1b), suggesting an alternative means to degrade the SAM-AMP signalling molecule.
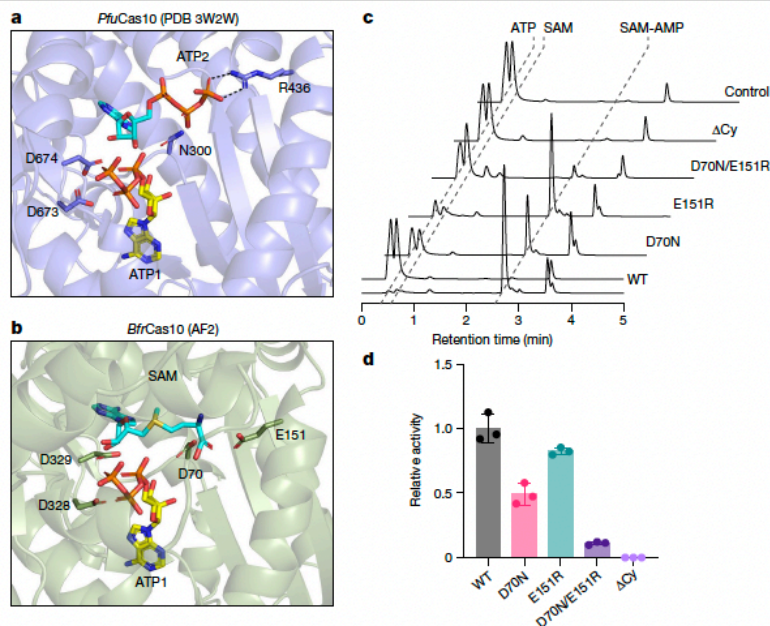
Fig. 3 | Changes in the active site of Cas10 during synthesis of SAM-AMP. **a**, The crystal structure of the *P. furiosus* (*Pfu*) Cas10 subunit with 2 bound ATP molecules[33]. Side chains for the two metal binding aspartate residues of the DD motif, together with residues N300 and R436 that interact with ATP2, are shown. **b**, Equivalent view of the AF2 model of the *B. fragilis* (*Bfr*) Cas10 structure with ATP1 from the *P. furiosus* Cas10 structure and ATP2 replaced by SAM. The precise conformation and position of SAM is unknown. The conserved acidic residues D70, E151, D328 and D329 are shown. **c**, In vitro SAM-AMP synthase activity of wild-type and variant *B. fragilis* Cmr, analysed by HPLC following incubation of 2 μM Cmr with 0.5 mM ATP and SAM for 30 min. Raw HPLC data are presented in Supplementary Data Fig. 3. **d**, Relative SAM-AMP synthase activity of Cmr variants. Three independent experiments; data are mean ± s.d., calculated using GraphPad Prism 9.

We tested this by cloning and expressing the SAM lyase from *C. botulinum* and measuring its ability to degrade SAM-AMP, observing efficient degradation of the molecule to 5′-methylthioadenosine (MTA) (Fig. 4c,d). The other product of a lyase reaction, L-homoserine lactone, is not detectable by UV. The *C. botulinum* lyase degrades SAM-AMP more efficiently than SAM (Extended Data Fig. 9b), consistent with a specialized role in defence.

Discussion

The polymerase active site of Cas10, the catalytic subunit of type III CRISPR systems, which consists of two DNA polymerase family B palm domains, is known to synthesize a range of cOA second messengers for antiviral defence. Here we have shown that some type III CRISPR systems signal via synthesis of SAM-AMP, a previously unknown molecule created by the adenylation of SAM (Fig. 5), which thus represents a novel nucleotide-based second messenger. In bacteria, the most recently discovered anti-phage signalling molecules include the cUMP and cCMP of the PYCSAR system[36], a wide range of cyclic di- and tri-nucleotides of the CBASS system[37] and the cOAs typically made by type III CRISPR systems[6,7]. Given the structural similarity between ATP and SAM, it is perhaps not surprising that SAM can substitute for ATP as an acceptor for a new 5′–3′ phosphodiester bond in the active site of nucleotide cyclases, following limited sequence divergence. Clearly this reaction reaches a natural end point as there is no possibility of cyclization or further polymerization. Judging by the distribution of CorA effectors, SAM-AMP signalling has a patchy but wide distribution in members of the *bacteroidetes, firmicutes, δ-proteobacteria, ε-proteobacteria* and *euryarchaea*. This is consistent with the high levels of defence system gain (by lateral gene transfer) and loss observed

generally, and may be a reflection of the pressures exerted by viruses, driving diversity.

CRISPR-associated CorA proteins are predicted to have a N-terminal soluble domain fused to a C-terminal transmembrane helical domain related to the CorA family of divalent cation transporters[16]. We postulate that binding of SAM-AMP to the cytoplasmic domain results in an opening of the transmembrane pore to effect immunity, but alternative mechanisms of membrane disruption have been observed for bacterial immune effectors[38], so this is a priority for future studies. The CorA effectors seem to be obligately associated with degradative enzymes such as NrN in *B. fragilis*, sometimes even being fused[16]. The observation that SAM-AMP is easily purified from extracts of *E. coli* expressing the activated *B. fragilis* Cmr system suggests that SAM-AMP is not a substrate for the generalist ribonucleases present in bacteria, necessitating the addition of a specialized PDE such as NrN. In these respects, NrN is reminiscent of the ring nucleases (Crn1–3 and Csx3) that are frequently found associated with cOA generating CRISPR systems[10]. This suggests that it is beneficial to the cell to deplete the SAM-AMP signalling molecule, perhaps to avoid unnecessary cell death when phage infection has been cleared. In this regard, it is telling that the NrN PDE is sometimes replaced by a SAM-AMP lyase—an enzyme that degrades SAM-AMP using an entirely different mechanism[23], yielding different products. SAM lyases are typically phage related genes and are thought to function by neutralizing DNA methylases in host restriction-modification systems[23,24]. In the context of CorA-family CRISPR systems, SAM-AMP lyases encoded by mobile genetic elements may also function as anti-CRISPRs, similarly to viral ring nucleases[39].

Some important open questions remain. It is difficult to explain the toxicity of CorA when no SAM-AMP is synthesized and NrN is absent,
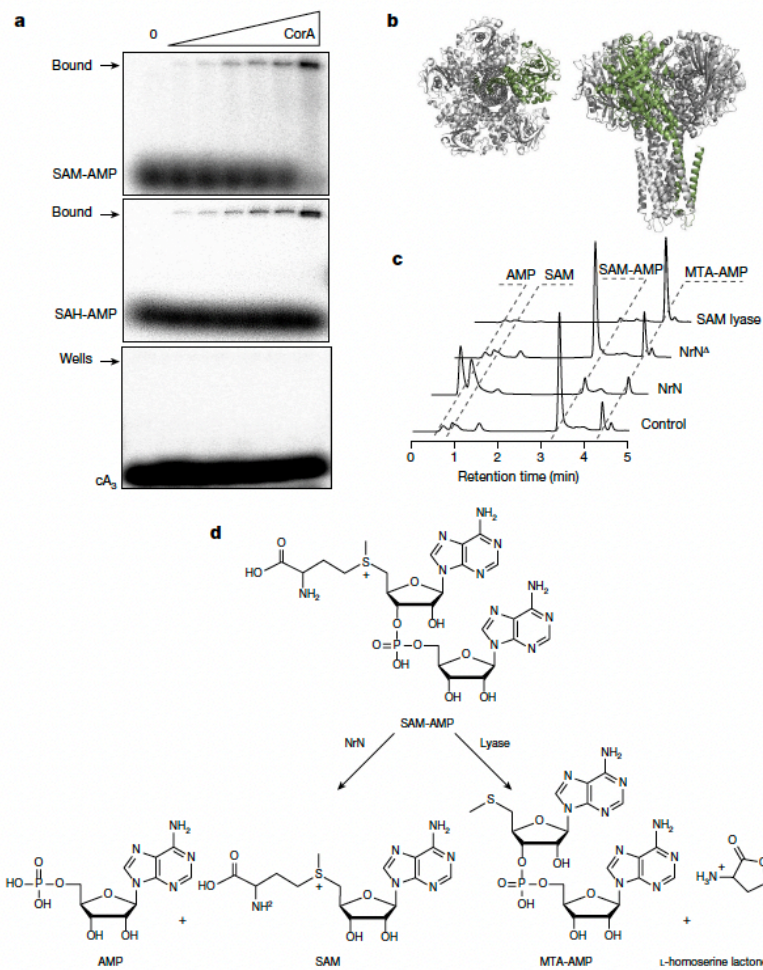
Article

Fig. 4 | Binding and degradation of SAM-AMP by ancillary CRISPR proteins.
a, CorA binds SAM-AMP and SAH-AMP, but not cA₃ (1 μM ^{32}P-labelled ligand incubated with 0, 0.0625, 0.125, 0.35, 0.75, 1.5 and 3.3 μM CorA), illustrated by acrylamide gel electrophoresis and phosphorimaging. Uncropped gels are shown in Supplementary Data Fig. 4. **b**, Two views of the predicted structure of the pentameric CorA channel, with one subunit coloured green. **c**, NrN specifically degrades SAM-AMP to SAM and AMP. HPLC analysis of samples in which purified SAM-AMP was incubated with NrN and NrN^A, an inactive variant with D85A/H86A/H87A mutations. *C. botulinum* lyase degrades SAM-AMP to generate MTA and L-homoserine lactone (not UV visible). Small amounts of MTA are present in the SAM-AMP sample purified from *E. coli*. Uncropped HPLC traces are available in Supplementary Data Fig. 4. **d**, Schematic representation of the reactions catalysed by NrN and SAM-AMP lyase.

as well as the observation that both CorA and NrN are required for immunity. These data suggest a close functional link between CorA and NrN, although we have detected no physical interaction between the two proteins in vitro. Rather than functioning in a manner analogous to ring nucleases, an alternative hypothesis is that NrN (or SAM-AMP lyase) is required to prevent de-sensitization of the CorA channel—a phenomenon observed for other pentameric ligand-gated ion channels when activator concentrations remain high[40]. Answers to these questions will probably require further analysis of the system in a cognate host at native expression levels, coupled with structure–function studies of the CorA channel.

Given the wide range of SAM and ATP analogues available, the discovery of an enzymatic route to synthesis of SAM-AMP opens the way to the generation of a new family of bioactive molecules. For example, there is considerable interest in the development of specific inhibitors of methyltransferases, a large family of enzymes (more than 300 methyltransferases are encoded in the human genome) involved in many key cellular reactions[41]. Depending on the specificity of the Cas10 enzyme, a range of SAM and ATP analogues could be provided as building blocks to make a diverse family of SAM-AMP analogues with altered properties. As we have seen, replacement of the methyl group on the sulfur atom of SAM with a proton (in SAH) or an amino group (sinefungin) still supports catalysis by *B. fragilis* Cas10. Many other modification sites are available on the parental molecules.

In conclusion, we report the discovery of SAM-AMP, which is synthesized from two of the most abundant molecules in the cell and functions as a second messenger of viral infection. This broadens the repertoire of type III CRISPR systems and may have implications for immune signalling more generally, as family B polymerases are a widespread and diverse superfamily found in all branches of life. The recent expansion of our knowledge of signalling molecules reflects the fact that Nature tends to use and repurpose such molecules in diverse cellular processes.
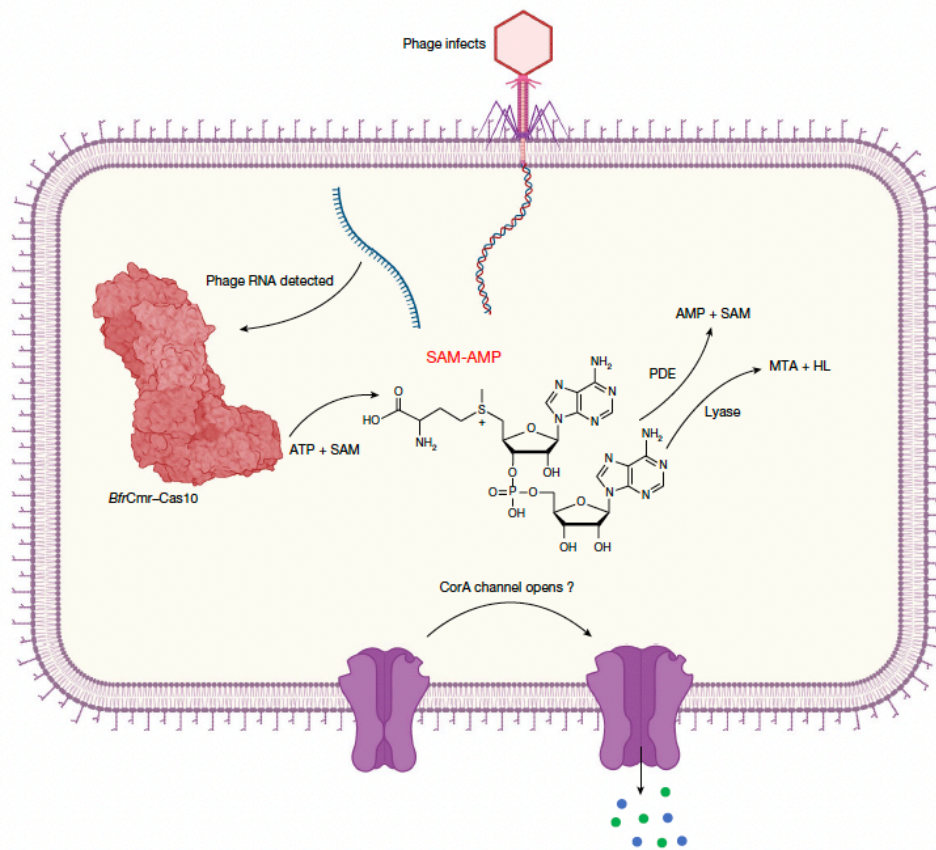
Fig. 5 | Model of the SAM-AMP immune signalling pathway. Transcription of the infecting phage genome activates the *B. fragilis* Cmr complex, resulting in synthesis of the SAM-AMP second messenger. SAM-AMP binds to the CorA membrane protein, resulting in the opening of a pore that disrupts the host membrane to combat infection. SAM-AMP is degraded by specialized PDE enzymes that hydrolyse the phosphodiester bond, generating AMP and SAM or lyases that target the methionine moiety, generating MTA and homoserine lactone (HL). These enzymes are likely to deactivate the signalling molecule to reset the system once phage have been eliminated. Created with BioRender.com.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-023-06620-5.

1. Makarova, K. S. et al. An updated evolutionary classification of CRISPR–Cas systems. *Nature Rev. Microbiol* **13**, 722–736 (2015).
2. Makarova, K. S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
3. Samai, P. et al. Co-transcriptional DNA and RNA cleavage during type III CRISPR–Cas immunity. *Cell* **161**, 1164–1174 (2015).
4. Elmore, J. R. et al. Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes Dev.* **30**, 447–459 (2016).
5. Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, C. & Siksnys, V. Spatiotemporal control of type III-A CRISPR–Cas immunity: coupling DNA degradation with the target RNA recognition. *Mol. Cell* **62**, 295–306 (2016).
6. Kazlauskiene, M., Kostiuk, G., Venclovas, C., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR–Cas systems. *Science* **357**, 605–609 (2017).
7. Niewoehner, O. et al. Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 (2017).
8. Rouillon, C., Athukoralage, J. S., Graham, S., Gruschow, S. & White, M. F. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *eLife* **7**, e36734 (2018).
9. Steens, J. A., Salazar, C. R. P. & Staals, R. H. J. The diverse arsenal of type III CRISPR–Cas-associated CARF and SAVED effectors. *Biochem. Soc. Trans.* **50**, 1353–1364 (2022).
10. Athukoralage, J. S. & White, M. F. Cyclic oligoadenylate signalling and regulation by ring nucleases during type III CRISPR defence. *RNA* **27**, 855–867 (2021).
11. Wexler, A. G. & Goodman, A. L. An insider's perspective: *Bacteroides* as a window into the microbiome. *Nat. Microbiol.* **2**, 17026 (2017).
12. Shkoporov, A. N. et al. PhiCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat. Commun.* **9**, 4781 (2018).
13. Wexler, H. M. Bacteroides: the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **20**, 593–621 (2007).
14. Tajkarimi, M. & Wexler, H. M. CRISPR–Cas systems in *Bacteroides fragilis*, an important pathobiont in the human gut microbiome. *Front. Microbiol.* **8**, 2234 (2017).
15. Grüschow, S., Adamson, C. S. & White, M. F. Specificity and sensitivity of an RNA targeting type III CRISPR complex coupled with a NucC endonuclease effector. *Nucleic Acids Res,* **49**, 13122–13134 (2021).
16. Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V. & Koonin, E. V. Systematic prediction of genes functionally linked to CRISPR–Cas systems by gene neighborhood analysis. *Proc. Natl Acad. Sci. USA* **115**, E5307–E5316 (2018).
17. Shah, S. A. et al. Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR–Cas gene cassettes reveals 39 new cas gene families. *RNA Biol.* **16**, 530–542 (2019).
18. Heo, K. et al. A pGpG-specific phosphodiesterase regulates cyclic di-GMP signaling in Vibrio cholerae. *J. Biol. Chem.* **298**, 101626 (2022).
19. Wang, F. et al. Structural and biochemical characterization of the catalytic domains of GdpP reveals a unified hydrolysis mechanism for the DHH/DHHA1 phosphodiesterase. *Biochem. J.* **475**, 191–205 (2018).
20. Hsiao, Y. Y., Duh, Y., Chen, Y. P., Wang, Y. T. & Yuan, H. S. How an exonuclease decides where to stop in trimming of nucleic acids: crystal structures of RNase T-product complexes. *Nucleic Acids Res.* **40**, 8144–8154 (2012).
21. Lee, C. W. et al. Structural basis of small RNA hydrolysis by oligoribonuclease (CpsORN) from *Colwellia psychrerythraea* strain 34H. *Sci. Rep.* **9**, 2649 (2019).

Article

22. Nicholls, T. J. et al. Dinucleotide degradation by REXO2 maintains promoter specificity in mammalian mitochondria. *Mol. Cell* **76**, 784–796.e786 (2019).
23. Guo, X. et al. Structure and mechanism of a phage-encoded SAM lyase revises catalytic function of enzyme family. *eLife* **10**, e61818 (2021).
24. Simon-Baram, H. et al. SAMase of bacteriophage T3 inactivates *Escherichia coli's* methionine S-adenosyltransferase by forming heteropolymers. *mBio* **12**, e0124221 (2021).
25. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
26. Grüschow, S., Athukoralage, J. S., Graham, S., Hoogeboom, T. & White, M. F. Cyclic oligoadenylate signalling mediates *Mycobacterium tuberculosis* CRISPR defence. *Nucl. Acids Res.* **47**, 9259–9270 (2019).
27. Shao, Y. et al. A non-stem-loop CRISPR RNA is processed by dual binding Cas6. *Structure* **24**, 547–554 (2016).
28. Tamulaitis, G., Venclovas, C. & Siksnys, V. Type III CRISPR–Cas immunity: major differences brushed aside. *Trends Microbiol.* **25**, 49–61 (2017).
29. Tamulaitis, G. et al. Programmable RNA shredding by the type III-A CRISPR–Cas system of *Streptococcus thermophilus*. *Mol. Cell* **56**, 506–517 (2014).
30. Estrella, M. A., Kuo, F. T. & Bailey, S. RNA-activated DNA cleavage by the type III-B CRISPR–Cas effector complex. *Genes Dev.* **30**, 460–470 (2016).
31. Vedel, M., Lawrence, F., Robert-Gero, M. & Lederer, E. The antifungal antibiotic sinefungin as a very active inhibitor of methyltransferases and of the transformation of chick embryo fibroblasts by Rous sarcoma virus. *Biochem. Biophys. Res. Commun.* **85**, 371–376 (1978).
32. Halliday, N. M., Hardie, K. R., Williams, P., Winzer, K. & Barrett, D. A. Quantitative liquid chromatography–tandem mass spectrometry profiling of activated methyl cycle metabolites involved in LuxS-dependent quorum sensing in *Escherichia coli*. *Anal. Biochem.* **403**, 20–29 (2010).
33. Osawa, T., Inanaga, H. & Numata, T. Crystal structure of the Cmr2–Cmr3 subcomplex in the CRISPR–Cas RNA silencing effector complex. *J. Mol. Biol.* **425**, 3811–3823 (2013).
34. Jia, N., Jones, R., Sukenick, G. & Patel, D. J. Second messenger cA₄ formation within the composite Csm1 palm pocket of type III-A CRISPR–Cas Csm complex and its release path. *Mol. Cell* **75**, 933–943 e936 (2019).
35. VanderWal, A. R. et al. Csx28 is a membrane pore that enhances CRISPR–Cas13b-dependent antiphage defense. *Science* **380**, 410–415 (2023).
36. Tal, N. et al. Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell* **184**, 5728–5739.e5716 (2021).
37. Whiteley, A. T. et al. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).
38. Duncan-Lowey, B., McNamara-Bordewick, N. K., Tal, N., Sorek, R. & Kranzusch, P. J. Effector-mediated membrane disruption controls cell death in CBASS antiphage defense. *Mol. Cell* **81**, 5039–5051.e5035 (2021).
39. Athukoralage, J. S. et al. An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. *Nature* **577**, 572–575 (2020).
40. Velisetty, P. & Chakrapani, S. Desensitization mechanism in prokaryotic ligand-gated ion channel. *J. Biol. Chem.* **287**, 18467–18477 (2012).
41. Zhang, J. & Zheng, Y. G. SAM/SAH analogs as versatile tools for SAM-dependent methyltransferases. *ACS Chem. Biol.* **11**, 583–597 (2016).
42. Russel, J., Pinilla-Redondo, R., Mayo-Munoz, D., Shah, S. A. & Sorensen, S. J. CRISPRCasTyper: automated identification, annotation, and classification of CRISPR–Cas loci. *CRISPR J* **3**, 462–469 (2020).

1 **Antiviral signaling by a cyclic nucleotide activated CRISPR protease**

2

3 Christophe Rouillon[1,2,*,#], Niels Schneberger[1,#], Haotian Chi[3], Katja Blumenstock[4], Stefano

4 Da Vela[5], Katrin Ackermann[6], Jonas Moecking[1], Martin F. Peter[1], Wolfgang Boenigk[2],

5 Reinhard Seifert[2], Bela E. Bode[6], Jonathan L. Schmid-Burgk[4], Dmitri Svergun[5], Matthias

6 Geyer[1], Malcolm F. White[3], Gregor Hagelueken[1,*]

7

8 [1] Institute of Structural Biology, University of Bonn, Venusberg-Campus 1, 53127 Bonn,

9 Germany

10 [2] Max Planck Institute for Neurobiology of Behavior – caesar, Ludwig-Erhard-Allee 2, 53175

11 Bonn, Germany

12 [3] School of Biology, University of St Andrews, North Haugh, St Andrews, KY16 9ST, UK.

13 [4] Institute of Clinical Chemistry and Clinical Pharmacology, University and University

14 Hospital Bonn, Venusberg-Campus 1, 53127 Bonn, Germany

15 [5] EMBL Hamburg Unit, c/o DESY, Notkestr. 85, 22607, Hamburg, Germany

16 [6] EaStCHEM School of Chemistry, Biomedical Sciences Research Complex, and Centre of

17 Magnetic Resonance, University of St Andrews North Haugh, St Andrews KY16 9ST, UK

18

19 [#] these authors contributed equally

20 *Corresponding author emails: back2crispr@gmail.com (C.R.), hagelueken@uni-bonn.de

21 (G.H.)

22
23

CRISPR defense systems such as the well-known DNA-targeting Cas9 and the RNA-targeting type III systems are widespread in prokaryotes[1,2]. The latter can orchestrate a complex antiviral response that is initiated by the synthesis of cyclic oligoadenylates (cOAs) upon foreign RNA recognition[3-5]. Among a large set of proteins that were linked to type III systems and predicted to bind cOAs[6,7], a CRISPR associated Lon protease (CalpL) stood out to us. The protein contains a sensor domain of the SAVED (SMODS-associated and fused to various effector domains) family[7], fused to a Lon protease effector domain. However, the mode of action of this effector was unknown. Here, we report the structure and function of CalpL and show that the soluble protein forms a stable tripartite complex with two further proteins, CalpT and CalpS, that are encoded in the same operon. Upon activation by cA_4, CalpL oligomerizes and specifically cleaves the MazF-homolog CalpT, releasing the extracytoplasmic function (ECF) sigma factor CalpS from the complex. This provides a direct connection between CRISPR-based foreign nucleic acid detection and transcriptional regulation. Furthermore, the presence of a cA_4-binding SAVED domain in a CRISPR effector reveals an unexpected link to the cyclic oligonucleotide-based antiphage signaling system (CBASS).

Main

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is a prokaryotic adaptive immune system that enables microorganisms to fend off attacks from mobile genetic elements such as phages, viruses, or plasmids[8]. The protein complex Cas1-Cas2 captures short DNAs from invaders and integrates them as "memories" into a CRISPR locus[9]. Transcripts of these "memories" are processed into small CRISPR RNAs (crRNAs) and integrated into large ribonucleoprotein (RNP) complexes, which can sense the presence of a matching foreign nucleic acid in the cell[10]. Once an invading sequence is detected, an antiviral response is triggered. Depending on the type of CRISPR system[1], this response can be markedly different, ranging from cleavage of the invading nucleic acid by the RNP as in the case of Cas9[11], to a complex multipronged defense strategy as found in type III CRISPR systems[12]. For the latter, the Cas10 subunit of the RNP has a cyclase activity that converts ATP into a recently discovered class of cyclic oligoadenylates (cOAs) upon viral RNA recognition[3–5]. The cOAs are constructed from 3 to 6, 3'-5' linked AMP units[13] and act as second messengers, typically by binding to proteins harboring a CARF (CRISPR-associated Rossmann-fold) domain[14]. There is a wide variety of CARF proteins linked to effector domains with functions ranging from RNA cleavage, supercoiled DNA nicking, dsDNA cleavage to transcription modulation[12,15–20]. The downstream effects of those cOA-activated proteins can lead to viral clearance, an abortive infection or a dormant state of the cell, enabling it to weather the phage attack[18,21].

Recently, two bioinformatic teams cataloged CARF-domain encoding genes that are likely linked to a functional type III system in bacterial and archaeal genomes[6,7]. Together, the studies revealed more than 100 such genes, including several membrane proteins and many proteins with currently unknown functions. Another study proposed that some of those type III-associated proteins contain a SAVED domain ('SMODS-associated and fused to various effectors domains'; SMODS being the acronym for 'second messenger oligonucleotide or dinucleotide synthetase'[22]) instead of CARF, reminiscent of the recently discovered CBASS system ('cyclic-oligonucleotide-based antiphage signaling systems'[23]). One of these proteins is CalpL (CRISPR associated Lon protease; initially termed Lon-CARF[6]), a 60 kDa protein with two predicted transmembrane helices, a Lon-protease domain and a SAVED4 domain[24].

Here we report the structure and function of the CalpL protein from the thermophilic bacterium *Sulfurihydrogenibium sp.* YO3AOP1. We find that CalpL is a soluble monomer and forms a 1:1:1 complex with CalpT and CalpS, encoded by adjacent genes in the locus. Once activated by cA$_4$, CalpL oligomerizes and proteolytically cleaves CalpT, releasing the CalpT$_{23}$/S complex, which has striking similarities to bacterial σ-factor/anti-σ-factor pairs.

3

Structure of CalpL

A synthetic, codon-optimized variant of the CalpL gene from *Sulfurihydrogenibium sp.* YO3AOP1 (UniProt ID B2V8L9) was expressed in *E. coli*. Although predicted to be a trans-membrane protein[6,7], the protein was found in the soluble fraction of the cell lysate and behaved as a monomer during size exclusion chromatography (Extended Data Fig. 1a,b and below). CalpL was crystallised and the structure was solved at a resolution of 2.1 Å by single-wavelength anomalous dispersion phasing and refined to R/R_{free} values of 19.3/22.5 (Fig. 1a, Extended Data Fig. 1c,d, Extended Data Table 1)[25,26].

The Lon protease domain consists of a four-stranded mixed β-sheet (β1-4), sandwiched between αD of the N-terminal domain and αG, H, J on the other side (Fig. 1a, Extended Data Fig. 1d). Various close structural homologs were identified and are listed in Extended Data Fig. 2. In CalpL, the catalytic Ser-Lys dyad, a hallmark of Lon proteases, is formed by S152 (loop β4-αH) and K193 (αJ) and lies at the end of a narrow channel that presumably binds the substrate peptide (Extended Data Fig. 2c). A superposition of the Lon-protease domain of CalpL with the acyl-enzyme intermediate state of the yellowfin ascites virus ATP-independent Lon protease[27] hints at the location of the P1 site in CalpL (Extended Data Fig. 2d). Structural modelling indicated that only amino acids with small hydrophobic side chains such as Ala or Gly can be accommodated in this site.

The C-terminal part of CalpL folds into a SAVED4 domain[24]. It consists of two pseudo-symmetric CARF-like domains with a pseudo-two-fold axis running between helices αP and αS (Fig. 1a). Interestingly, the TMHMM 2.0 server predicted that those helices and the directly preceding β-strands form transmembrane helices or at least membrane associated helices[6,28] (Extended Data Fig. 1b), which is clearly not the case. The SAVED domain has an extensive, positively charged cavity on its molecular surface, suited to bind a cOA ligand. The CARF- or SAVED-domains of the cOA-activated effector proteins Cap4, Cap5, and Can1 are structural homologs found in CBASS and CRISPR systems [17,23,29] (Extended Data Fig. 2b). Despite the low sequence identities (9-14% identical amino acids) the fold of the CARF-like domains is conserved. The position of the effector domain relative to the SAVED or CARF domains is, however, entirely different between the four structures (Extended Data Fig. 2e-g).

Surface plasmon resonance experiments showed that of the four tested cOAs CalpL selectively binds cA4 with a dissociation constant of ~1 nM (Fig. 1b). We determined a 2.2 Å crystal structure of the CalpL/cA4 complex and found the ligand bound to the SAVED domain at the expected position (Fig. 1b,c). As for the SAVED domain itself, the bound cA4 molecule adopts a pseudo two-fold symmetry (Extended Data Fig. 3a). The cyclic tetra adenylate is

4

involved in a large number of polar and hydrophobic interactions with the SAVED domain, such that three of the four phosphate groups of the ligand are solvent exposed (Extended Data Fig. 3a). A comparison between the apo- and the cA_4-complex structure reveals no major conformational changes, apart from small shifts in the loops surrounding the cA_4 ligand (Extended Data Fig. 3b). A comparison with the cA_3 bound Cap4 structure (Extended Data Fig. 3c) illustrates how the loops on top of the SAVED domains shape the binding site for the cognate ligand in each structure.

The N-terminal domain of CalpL forms a bundle of six α-helices (αA-F, Fig. 1a, Extended Data Fig. 1d) with weak structural similarity to the N-terminal protein-protein interaction domain of Katanin p60-N in the p60p80-CAMSAP complex (Extended Data Fig. 2b)[30].

CalpL specifically cleaves CalpT

We used the WebFLAGs server[31] to study the gene neighborhood of CalpL homologs and noticed a small 812 bp open reading frame (271 amino acids, 31.8 kDa, UniProt ID B2V8L8) with no annotated function upstream of the *calpL* gene (Fig. 2a). We analyzed its sequence with HHPRED[32] and found homologies to the MazF toxin in the N-terminal half of the protein and weak homologies to DUF2080, a "domain of unknown function", in the C-terminal half (Fig. 2a). We predicted the structure with AlphaFold2[33]. The software produced a model of a two-domain protein with a ~23 kDa and a ~10 kDa domain connected by an apparently flexible linker (Fig. 2b). The structural model was submitted to the DALI server[34], revealing structural similarities to MazF-like toxins (N-terminal fragment) and various immunoglobulin fold containing proteins (C-terminal fragment). Interestingly, the predicted structure appears as a structural mimic of the MazEF complex with helices αA, αD, and αE blocking the region that binds to the ssRNA target of MazF in a similar fashion to MazE (Extended Data Fig. 4a,b)[35]. We investigated whether this protein (named "CalpT" for "target") is cleaved by the CalpL protease. The gene was expressed in *E. coli* and the protein purified to near homogeneity (Extended Data Fig. 4c). CalpL, CalpT, and different cOAs (3, 4, 5, 6) were mixed at 1:1:1.5 molar ratios and incubated at 60 °C for one hour. Strikingly, we found that in the presence of cA_4, CalpT was cleaved by CalpL. SDS-PAGE analysis revealed two distinct cleavage products with molecular weights of 23 kDa and 10 kDa, respectively, suggesting a single cleavage site (Fig. 2c). The activity for the other cOAs was significantly lower. We repeated the experiment with an S152A variant of CalpL, which lacks the nucleophilic serine needed for its peptidase activity. Since this variant showed no protease activity, the CalpL protease active site is responsible for the observed proteolytic activity (Fig. 2c).

5

The peptide sequences of the two cleavage fragments were determined with peptide mass fingerprinting (Extended Data Fig. 4d). This analysis confirmed that the 23 kDa (CalpT23) fragment corresponds to the N-terminal two-thirds of the CalpT protein and the 10 kDa fragment (CalpT10) to the C-terminal one-third. Based on this result and considering the predicted structure (Fig. 2b), we mapped the location of the cleavage site to the stretch of residues between amino acids ~170-200 of CalpT. As mentioned above, our CalpL structure suggested that only peptides containing an alanine or glycine as the P1 residue will fit into the active site of CalpL. We therefore created glutamic acid mutants of all four alanine residues in the cleavage region: A172, A182, A195, and A201 (Fig. 2b, magenta spheres; the stretch of residues does not contain any glycine). Peptidase assays with all four CalpT variants were conducted and only the A195E mutation abolished the cleavage completely (Extended Data Fig. 4e). At this position, the amino acid sequence reads V_{190}LRHVA|ST, where A195 is most likely the P1 residue. Notably, A195 is conserved amongst CalpT homologs (Fig. 2d). The peptide fingerprint data in Extended Data Fig. 4d also supports this conclusion, as for CalpT23, the peptide coverage extended almost exactly to the identified cleavage site. We did not observe any non-tryptic peptides that corresponded to the identified cleavage site. Thus, the exact molecular weights of the CalpT cleavage products are 23.0 kDa (CalpT23) and 8.7 kDa (CalpT10), fitting to the sizes observed in SDS PAGE analysis (Fig. 2c).

CalpL and CalpT form a 1:1 complex

To test whether CalpL and CalpT form a stable complex, we analysed the individual proteins and their equimolar mixtures by SEC-MALS (Fig. 3a). CalpL alone eluted in a single peak at 17.1 ml and the CalpT protein eluted at 17.9 ml, both at the expected molecular weights for the monomeric proteins. The 1:1 mixture of the two proteins resulted in a single elution peak at 16.2 ml. The MW_{MALS} of the complex was 82.4 kDa, suggestive of a 1:1 complex of the CalpL and CalpT proteins (52 + 30.5 kDa). We used surface plasmon resonance (SPR) to quantify the interaction strength between CalpL and -T and found that the two proteins form a very strong complex with a sub-nanomolar K_D (Extended Data Fig. 5a). Interestingly, a similar affinity was observed for a construct where CalpT10 (including the cleavage site) was fused to a VHH domain targeting an unrelated protein (Extended Data Fig. 5b,c). The artificial construct was readily cleaved by the protease upon activation by cA4 (Extended Data Fig. 5d). Thus, the CalpT23 fragment plays no important role either in the formation of the CalpL/T complex or in the cleavage process. A second artificial construct, where the CalpT10 moiety was also replaced

176 by an unrelated VHH was not cleaved (Extended Data Fig. 5d). Hence, the CalpT$_{10}$ subunit is
177 required for cleavage.

178   To follow the fate of the complex after cleavage, we repeated the experiment in presence of
179 an excess (1:1.1) of cA$_4$ (Fig. 3a, violet). Here, we observed three peaks corresponding to
180 CalpL/T$_{10}$, CalpT$_{23}$ and cA$_4$. Correspondingly, for the inactive CalpL S152A variant, the
181 CalpL/T complex was observed but the addition of cA$_4$ did not lead to the observed split into
182 three peaks (Extended Data Fig. 5e). We also checked whether the four cleavage-site variants
183 of CalpT could still form a complex with CalpL (Extended Data Fig. 5f). Whereas A172E,
184 A182E, and the P1 site variant A195E did, the A201E variant did not. The glutamate at this
185 position apparently weakened the interaction, explaining the reduced cleavage efficiency of this
186 mutant (Extended Data Fig. 4e).

187   We isolated the CalpL/T$_{10}$ complex for crystallization and determined its structure at 3.3 Å
188 resolution by molecular replacement, using the CalpL crystal structure and the AlphaFold
189 model of CalpT$_{10}$ as search models (Fig. 3a, Extended Data Fig. 5g, Extended Data Table 1).
190 Indeed, the CalpT$_{10}$ fragment binds to the N-terminal domain of CalpL and, as indicated by the
191 mutation analysis, A201 of CalpT is part of the interface in addition to the CalpL hydrophobic
192 residues W28, L6, V14, L18, E20, E13, K8, H2 and CalpT residues K200, Y210, Y203, E222
193 (Extended Data Fig. 5g). A small-angle X-ray scattering (SAXS) experiment was performed to
194 exclude a crystal packing artefact, by measuring the SAXS profiles of CalpL and CalpL/T$_{10}$
195 complex by SEC-SAXS (Extended Data Table 2, experimental session I). The two profiles were
196 fitted simultaneously using the multi-phase *ab initio* shape reconstruction program MONSA.
197 The *ab initio* model was in excellent agreement with the SAXS data and compares well to our
198 crystal structure (Extended Data Fig. 5h), thus confirming the arrangement of the two subunits
199 in solution.

200

201 **cA$_4$ induced oligomerization of CalpL/T**
202 Intriguingly, while the C-terminal part of the CalpL cleavage site (T197 of CalpT) is visible in
203 the complex crystal structure, it is more than 35 Å away from the protease active site (Fig. 3a),
204 indicating that a cA$_4$ induced structural rearrangement of CalpL must occur to allow cleavage
205 of CalpT. Recent studies on other SAVED-domain containing CBASS effectors demonstrated
206 that cOA binding induces an oligomerization, which then activates the effector[23,29,36]. Dynamic
207 light scattering (DLS) and SAXS experiments showed such a cA$_4$- and protein concentration-
208 dependent oligomerization of CalpL (Fig. 3b, Extended Data Fig. 6a,b). A representative *ab*
209 *initio* model for monomeric CalpL in the presence of cA$_4$ was obtained by SEC-SAXS

7

210 (Extended Data Table 2), resulting in an elongated and slightly bent model featuring two main
211 lobes connected by a slightly thinner region at SAXS resolution. Using this monomeric unit,
212 the concentration series in the range 2-5 mg/mL was modeled as a dimerizing mixture by a
213 global SASREFMX fitting without imposing symmetry elements (as non-identical binding
214 interfaces are to be expected in the presence of the cA_4 ligand). SAXS modelling of the
215 dimerizing mixture produced stable solutions (normalized spatial discrepancy (NSD) ~0.95)
216 featuring elongated shapes that were large enough to accommodate two CalpL molecules
217 (Extended Data Fig. 6c).

218   We noticed a distinct positively charged patch on the face opposite of the cA_4 binding site,
219 where R361, R338 and K364 coordinate a sulfate ion in the cA_4 complex structure (Fig. 1c).
220 Due to this structural feature, binding of the negatively charged cA_4 molecule would result in a
221 charge complementarity between the top and bottom sides of the SAVED domain. This supports
222 an arrangement where two or more CalpL molecules would form stacks, with cA_4 sandwiched
223 in between, similar to the architecture observed for other SAVED domain oligomers[23,29,36,37].
224 The SAXS *ab initio* model in Extended Data Fig. 6c would best agree with a staggered
225 arrangement of the CalpL monomers. An attractive model for the activation of CalpL by cA_4
226 would thus be an in-trans cleavage reaction in the observed oligomers. To test this, we
227 performed a cleavage assay, were mixtures of preformed CalpL/T complexes, for instance
228 CalpL/T (wt/wt), CalpL/T (S152A/wt), CalpL/T (wt/A195E), were tested for cA_4 induced
229 cleavage. A 1:1 mixture of CalpL/T (S152A/wt) with CalpL/T (wt/A195E), *i.e.*, two complexes
230 that are not capable of in-cis cleavage, led to 50% cleaved CalpT (Fig. 3c). The remaining 50%
231 could not be cleaved due to the A195E mutation, but this could be titrated by changing the ratio
232 of the two complexes (Fig. 3c). Further, a mixture of CalpL/T (S152A/wt) with uncomplexed
233 CalpL led to complete cleavage of CalpT. To support the idea of oligomerization induced
234 cleavage, we introduced mutations to the backside of the SAVED domain, aiming to disturb
235 the presumed oligomerization interface (Fig. 3d and 1c). While mutant R361E had a ~50%
236 reduced activity, R338E had no cleavage activity. We also found that the R493C mutant used
237 for the SPR experiments had a ~50% reduced cleavage activity.

238   While our data show that in-trans cleavage occurs, we can currently not distinguish, whether
239 cleavage occurs inside one particular CalpL/T oligomer or between two oligomers. For the
240 latter, three CalpL/T units would have to assemble for the VA/ST sequence in CalpT to be able
241 to reach the protease active site of a CalpL molecule in the oligomer. A high-resolution structure
242 of such a CalpL/T stack will be necessary to unravel the molecular details of the activation.
243

8

CalpL/T bind the ECF σ factor CalpS

Our initial assumption that CalpT will have a MazF-like nuclease activity could not be confirmed experimentally. We could neither identify any signs of an RNase activity biochemically (Extended Data Fig. 7a-c), nor with RNase-seq of random libraries (Extended Data Fig. 7d,e). Moreover, expression of CalpT$_{23}$ constructs in *E. coli* were tolerated by the cells (not shown). We also looked at a dimerization of the CalpT$_{23}$ fragment, similar to the active dimeric MazF enzyme. To test this, we spin labelled the CalpL/CalpT complex at position 119 of CalpT (Extended Data Fig. 7f,g) and measured the interspin distance in the presence and absence of cA$_4$. According to the EPR data, the cA$_4$ induced cleavage did not lead to changes of the conformational state of the CalpT$_{23}$ (Extended Data Fig. 7h-j).

All this turned our attention towards CalpS, a third conserved protein encoded by the operon (224 amino acids, 26.5 kDa, UniProt ID B2V8L7). The protein has strong sequence similarities to ECF family σ factors, which tailor transcription in diverse stress conditions[32,38,39] (Extended Data Fig. 8a). Interestingly, AlphaFold2 supported a heterotrimeric complex between CalpL, -T and -S, which was consistent with our finding that the CalpT$_{23}$ domain is not involved in CalpL/T complex formation (Fig. 4a). The predicted CalpT/S interface has a combined buried surface area of ~4000 Å^2 [40], high confidence scores in the interface area and convincing sidechain interactions. In the prediction, CalpT binds to both the σ$_2$- and σ$_4$-domains of the σ-factor and blocks most of the -10-region interface (Extended Data Fig. 8bc). The same interface is targeted by so called anti-σ-factors, preventing the interaction of the σ-factor/RNA-polymerase (RNAP) complex with its cognate promotor (Extended Data Fig. 8d)[38,39]. Note that in the predicted CalpT/S complex, the σ$_2$- and σ$_4$ domains are tied together in a way that would not allow the σ-factor to bind to the RNAP (Extended Data Fig. 8e).

To put the existence of this complex to the test, we cloned the *calpS* gene and co-expressed the His-CalpT/S proteins in *E. coli*. As predicted, the two proteins formed a stable complex that could be isolated by gel filtration (Fig. 4b). Furthermore, addition of CalpL to the CalpT/S complex resulted in a ternary complex, that disintegrated into CalpL/T$_{10}$ and CalpT$_{23}$/S upon addition of cA$_4$ (Fig. 4bcd).

We noticed that expression of CalpS alone (instead of coexpression with CalpT) led to a copurification of the protein with the α- and β subunits of the DNA-directed RNA polymerase of *E. coli* (44% sequence identity between β subunits of *Sulfurihydrogenibium sp.* and *E. coli* RNAP). This corroborates the prediction that CalpS is a σ-factor and that CalpT inhibits its interaction with the RNAP (Fig. 4e). Thus, CalpT has striking functional similarities to anti-σ-

9

factor proteins and literally links the cA$_4$ sensor CalpL to the transcription machinery of the cell.

Discussion

Protease signaling cascades are a common scheme in evolution that are often employed in emergency situations. Prokaryotic type-II toxin-antitoxin (TA) systems, for instance, are activated by degradation of the antitoxin by ATP-dependent Lon proteases[41]. The innate immune system of higher organisms also employs proteases such as caspases to initiate and amplify fast responses to external threats. Another well-known example are the cascades of proteases that control the clotting of blood[42]. Our work shows that the CalpL/T/S cascade amalgamates aspects of different defense systems such as CRISPR, CBASS, toxin/antitoxin systems, and σ/anti-σ-factors into a cA$_4$-controlled "fast response" signaling cascade.

Our data are summarized in the model sketched in Fig. 5, where in its inactive state, CalpL will be present in a 1:1:1 complex with CalpT and CalpS. Upon detection of a foreign RNA by the type III effector complex, cA$_4$ will be synthesized by its Cas10 subunit and the second messenger will bind to the SAVED domain of CalpL with nanomolar affinity. This drastically changes the surface electrostatics of the SAVED domain, enabling oligomerization of CalpL, as observed by SAXS and DLS (Fig. 3b and Extended Data Fig. 6). This activation mechanism has now emerged as a common theme in the SAVED-domain based effectors of CBASS defense systems[23,29,36] but has not been observed in CRISPR systems. The CalpL oligomerization triggers an in-trans cleavage of CalpT (Fig. 3c,d), releasing the CalpT$_{23}$/S subcomplex in a strictly cA$_4$ dependent manner (Fig. 4bcd). Some CARF domain proteins are known to auto deactivate by degrading cOA species [12], but this has not been observed for SAVED domains and has not yet been investigated for CalpL.

CalpS is member of the ECF family of ECF σ factors, which play a role in the sensing of extracellular stress events, such as cell envelope- or oxidative stress[38,39]. In striking resemblance to the Calp cascade, the release of such anti-σ factors is orchestrated by a sequence of proteolytic events called regulated intermembrane proteolysis (RIP)[43]. The activation of σ^E from *E. coli*, for instance, proceeds via proteolytic cleavage of the membrane bound anti sigma factor RseA, releasing a soluble σ-factor/anti-σ-factor complex. The anti-σ-factor is subsequently degraded by ATP-dependent ClpXP proteases. [38,44]. Following the established paradigms, one might speculate that further proteolysis of CalpT$_{23}$ releases the sigma factor CalpS to allow transcriptional response.

10

310   Recently, two studies have revealed that type III-E CRISPR systems also function by
311   activating the protease Csx29 (also known as TRP-CHAT) [45,46]. There are notable differences,
312   as the protease (from the Caspase family) is completely unrelated to CalpL and is not activated
313   by cOA. Csx29, part of the Type III-E effector complex, is activated by conformational changes
314   upon foreign RNA detection and cleaves an uncharacterized protein (Csx30) encoded in the
315   operon [47]. Interestingly, Csx30 binds a σ-factor also homologous to the ECF family (termed
316   CASP-σ). Furthermore, it was shown that the Csx30 protein inhibits CASP-σ and this inhibition
317   is relieved by Csx29 mediated proteolytic cleavage. CASP-σ has a high affinity for a DNA
318   sequence that is found in the promotor of Cas1-2, proteins of CRISPR adaptation for the
319   acquisition of new viral memories [48]. Since the CRISPR effectors and the two proteases are
320   completely unrelated, this appears to be a striking example of convergent evolution.

321   Here, we have uncovered a cOA-mediated signaling cascade from viral RNA detection to the
322   proteolytic release of a σ-factor that binds RNA polymerase. Notably, the Cas10 proteins
323   associated with Calp operons lack an HD nuclease domain and auxiliary cOA activated
324   nucleases such as Csx1 are rarely found, suggesting transcriptional changes sufficient for
325   CRISPR antiviral immunity in organisms such as *Sulfurihydrogenibium*. It will be exciting to
326   find out the DNA targets of CalpS to understand better how the Calp cascade shapes the
327   antiviral response, buying the organism enough time to survive a viral attack[18,21].
328

11

Main text references

1. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Classification and nomenclature of CRISPR-Cas systems: where from here. *The CRISPR Journal* 1, 325-336 (2018).
2. Zhu, Y., Klompe, S. E., Vlot, M., van der Oost, J. & Staals, R. H. J. Shooting the messenger: RNA-targetting CRISPR-Cas systems. *Bioscience reports* 38, BSR20170788 (2018).
3. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* 357, 605-609 (2017).
4. Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A. & Jinek, M. Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* 548, 543-548 (2017).
5. Rouillon, C., Athukoralage, J. S., Graham, S., Grüschow, S. & White, M. F. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *Elife* 7, e36734 (2018).
6. Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V. & Koonin, E. V. Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc Natl Acad Sci U S A* 115, E5307-E5316 (2018).
7. Shah, S. A., Alkhnbashi, O. S., Behler, J., Han, W., She, Q., Hess, W. R., Garrett, R. A. & Backofen, R. Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new cas gene families. *RNA Biol* 16, 530-542 (2019).
8. Gasiunas, G., Sinkunas, T. & Siksnys, V. Molecular mechanisms of CRISPR-mediated microbial immunity. *Cell Mol Life Sci* 71, 449-465 (2014).
9. Sasnauskas, G. & Siksnys, V. CRISPR adaptation from a structural perspective. *Curr Opin Struct Biol* 65, 17-25 (2020).
10. Jiang, F. & Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics* 46, 505-529 (2017).
11. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012).
12. Athukoralage, J. S. & White, M. F. Cyclic oligoadenylate signalling and regulation by ring nucleases during type III CRISPR defence. *RNA* rna.078739.121 (2021).
13. Jia, N., Jones, R., Sukenick, G. & Patel, D. J. Second messenger cA4 formation within the composite Csm1 Palm pocket of type III-A CRISPR-Cas Csm complex and its release path. *Molecular cell* 75, 933-943. e6 (2019).
14. Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V. & Aravind, L. CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Frontiers in genetics* 5, 102 (2014).
15. Lau, R. K., Ye, Q., Birkholz, E. A., Berg, K. R., Patel, L., Mathews, I. T., Watrous, J. D., Ego, K., Whiteley, A. T. & Lowey, B. Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity. *Molecular cell* 77, 723-733. e6 (2020).
16. Lintner, N. G., Frankel, K. A., Tsutakawa, S. E., Alsbury, D. L., Copié, V., Young, M. J., Tainer, J. A. & Lawrence, C. M. The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. *Journal of molecular biology* 405, 939-955 (2011).
17. McMahon, S. A., Zhu, W., Graham, S., Rambo, R., White, M. F. & Gloster, T. M. Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nat Commun* 11, 500 (2020).

12

379 18. Rostøl, J. T., Xie, W., Kuryavyi, V., Maguin, P., Kao, K., Froom, R., Patel, D. J. &
380   Marraffini, L. A. The Card1 nuclease provides defence during type III CRISPR
381   immunity. *Nature* 590, 624-629 (2021).
382 19. Garcia-Doval, C., Schwede, F., Berk, C., Rostøl, J. T., Niewoehner, O., Tejero, O., Hall,
383   J., Marraffini, L. A. & Jinek, M. Activation and self-inactivation mechanisms of the
384   cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. *Nature communications*
385   11, 1-9 (2020).
386 20. Lawrence, C. M., Charbonneau, A. & Gauvin, C. Cyclic Tetra-Adenylate (cA4)
387   Activates CRISPR Associated Transcription Factor Csa3, Providing Feedback
388   Activation of Protospacer Acquisition and crRNA Expression. *The FASEB Journal* 34,
389   1-1 (2020).
390 21. Meeske, A. J., Nakandakari-Higa, S. & Marraffini, L. A. Cas13-induced cellular
391   dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* 570, 241-245
392   (2019).
393 22. Burroughs, A. M., Zhang, D., Schäffer, D. E., Iyer, L. M. & Aravind, L. Comparative
394   genomic analyses reveal a vast, novel network of nucleotide-centric systems in
395   biological conflicts, immunity and signaling. *Nucleic Acids Res* 43, 10633-10654
396   (2015).
397 23. Lowey, B., Whiteley, A. T., Keszei, A. F. A., Morehouse, B. R., Mathews, I. T., Antine,
398   S. P., Cabrera, V. J., Kashin, D., Niemann, P., Jain, M., Schwede, F., Mekalanos, J. J.,
399   Shao, S., Lee, A. S. Y. & Kranzusch, P. J. CBASS Immunity Uses CARF-Related
400   Effectors to Sense 3'-5'- and 2'-5'-Linked Cyclic Oligonucleotide Signals and Protect
401   Bacteria from Phage Infection. *Cell* 182, 38-49.e17 (2020).
402 24. Makarova, K. S., Timinskas, A., Wolf, Y. I., Gussow, A. B., Siksnys, V., Venclovas, Č.
403   & Koonin, E. V. Evolutionary and functional classification of the CARF domain
404   superfamily, key sensors in prokaryotic antivirus defense. *Nucleic acids research* 48,
405   8828-8847 (2020).
406 25. Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R.,
407   McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K.,
408   Storoni, L. C., Terwilliger, T. C. & Adams, P. D. Automated structure solution with the
409   PHENIX suite. *Methods in Molecular Biology (Clifton, N.J.)* 426, 419-435 (2008).
410 26. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G.
411   J., Murray, L. W., Richardson, J. S. & Richardson, D. C. MolProbity: all-atom structure
412   validation for macromolecular crystallography. *Acta Crystallographica Section D* 66,
413   12-21 (2010).
414 27. Chung, I. Y. & Paetzel, M. Crystal structures of yellowtail ascites virus VP4 protease:
415   trapping an internal cleavage site trans acyl-enzyme complex in a native Ser/Lys dyad
416   active site. *J Biol Chem* 288, 13068-13081 (2013).
417 28. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting
418   transmembrane protein topology with a hidden markov model: application to complete
419   genomes. *Journal of Molecular Biology* 305, 567-580 (2001).
420 29. Fatma, S., Chakravarti, A., Zeng, X. & Huang, R. H. Molecular mechanisms of the
421   CdnG-Cap5 antiphage defense system employing 3', 2'-cGAMP as the second
422   messenger. *Nature Communications* (2021).
423 30. Jiang, K., Faltova, L., Hua, S., Capitani, G., Prota, A. E., Landgraf, C., Volkmer, R.,
424   Kammerer, R. A., Steinmetz, M. O. & Akhmanova, A. Structural Basis of Formation of
425   the Microtubule Minus-End-Regulating CAMSAP-Katanin Complex. *Structure* 26, 375-
426   382.e4 (2018).
427 31. Saha, C. K., Sanches Pires, R., Brolin, H., Delannoy, M. & Atkinson, G. C. FlaGs and
428   webFlaGs: discovering novel biology through the analysis of gene neighbourhood
429   conservation. *Bioinformatics* 37, 1312-1314 (2021).

13

430   32.   Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F.,
431   Söding, J., Lupas, A. N. & Alva, V. A completely reimplemented MPI bioinformatics
432   toolkit with a new HHpred server at its core. *Journal of molecular biology* 430, 2237-
433   2243 (2018).

434   33.   Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O.,
435   Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C.,
436   Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R.,
437   Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M.,
438   Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W.,
439   Kavukcuoglu, K., Kohli, P. & Hassabis, D. Highly accurate protein structure prediction
440   with AlphaFold. *Nature* 596, 583-589 (2021).

441   34.   Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. 38, W545-9
442   (2010).

443   35.   Simanshu, D. K., Yamaguchi, Y., Park, J.-H., Inouye, M. & Patel, D. J. Structural basis
444   of mRNA recognition and cleavage by toxin MazF and its regulation by antitoxin MazE
445   in Bacillus subtilis. *Molecular cell* 52, 447-458 (2013).

446   36.   Hogrel, G., Guild, A., Graham, S., Rickman, H., Grushow, S., Bertrand, Q., Spagnolo,
447   L. & White, M. F. Cyclic nucleotide-induced superhelical structure activates a bacterial
448   TIR immune effector. *bioRxiv* (2022).

449   37.   Hogrel, G., Guild, A., Graham, S., Rickman, H., Grüschow, S., Bertrand, Q., Spagnolo,
450   L. & White, M. F. Cyclic nucleotide-induced helical structure activates a TIR immune
451   effector. *Nature* 1-5 (2022).

452   38.   Paget, M. S. Bacterial sigma factors and anti-sigma factors: structure, function and
453   distribution. *Biomolecules* 5, 1245-1265 (2015).

454   39.   Sineva, E., Savkina, M. & Ades, S. E. Themes and variations in gene regulation by
455   extracytoplasmic function (ECF) sigma factors. *Current opinion in microbiology* 36,
456   128-137 (2017).

457   40.   Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline
458   state. *Journal of Molecular Biology* 372, 774-797 (2007).

459   41.   Schuster, C. F. & Bertram, R. Toxin–antitoxin systems are ubiquitous and versatile
460   modulators of prokaryotic cell fate. *FEMS microbiology letters* 340, 73-85 (2013).

461   42.   Walsh, P. N. & Ahmad, S. S. Proteases in blood clotting. *Essays in biochemistry* 38, 95-
462   112 (2002).

463   43.   Brown, M. S., Ye, J., Rawson, R. B. & Goldstein, J. L. Regulated intramembrane
464   proteolysis: a control mechanism conserved from bacteria to humans. *Cell* 100, 391-398
465   (2000).

466   44.   Fei, X., Bell, T. A., Barkow, S. R., Baker, T. A. & Sauer, R. T. Structural basis of
467   ClpXP recognition and unfolding of ssrA-tagged substrates. *Elife* 9, e61496 (2020).

468   45.   Hu, C., van Beljouw, S. P. B., Nam, K. H., Schuler, G., Ding, F., Cui, Y., Rodríguez-
469   Molina, A., Haagsma, A. C., Valk, M. & Pabst, M. Craspase is a CRISPR RNA-guided,
470   RNA-activated protease. *Science* eadd5064 (2022).

471   46.   van Beljouw, S. P. B., Haagsma, A. C., Rodríguez-Molina, A., van den Berg, D. F.,
472   Vink, J. N. A. & Brouns, S. J. J. The gRAMP CRISPR-Cas effector is an RNA
473   endonuclease complexed with a caspase-like peptidase. *Science* 373, 1349-1353 (2021).

474   47.   Kato, K., Okazaki, S., Schmitt-Ulms, C., Jiang, K., Zhou, W., Ishikawa, J., Isayama, Y.,
475   Adachi, S., Nishizawa, T. & Makarova, K. S. RNA-triggered protein cleavage and cell
476   growth arrest by the type III-E CRISPR nuclease-protease. *Science* eadd7347 (2022).

477   48.   Strecker, J., Demircioglu, F. E., Li, D., Faure, G., Wilkinson, M. E., Gootenberg, J. S.,
478   Abudayyeh, O. O., Nishimasu, H., Macrae, R. K. & Zhang, F. RNA-activated protein
479   cleavage with a CRISPR-associated endopeptidase. *Science* eadd7450 (2022).

480

14

Figure legends

Fig. 1 | Structure of apo and cA$_4$ bound CalpL. a, Overall structure of CalpL in the apo state. The structure is shown as a cartoon model and the individual domains are labelled and color-coded. The N- and C-termini, as well as the protease active site residues, are marked by spheres. The positions of key structural elements are indicated. **b**, Single cycle kinetics SPR measurements of different cOAs binding to immobilized CalpL. The experiment was performed multiple times (n=3 technical replicates) for cA$_4$ and once for the other cOAs. **c**, The CalpL/cA$_4$ complex structure. CalpL is shown as a surface model and the electrostatic potential is mapped onto the structure (blue - positive, red – negative). The bound cA$_4$ molecule is shown as spheres. Turning the structure by 180° reveals a positively charged patch opposite of the SAVED domain.

Fig. 2 | CalpL is activated by cA$_4$ and cleaves CalpT. a, The WebFLAGs server [31] was used to investigate the genomic neighborhood of CalpL (green). The primary structure of CalpT (red) is shown on top. Regions with homologies found by HHPRED [32] are marked. **b**, A structural prediction (AlphaFold2, [33]) of CalpT. The protein is shown as cartoon and colored according to the prediction confidence (pLDDT [49], predicted local distance difference test). **c**, SDS-PAGE analysis of CalpL induced cleavage of CalpT. The experiment was repeated multiple times (n =2 biological replicates and n>>3 technical replicates) **d**, Sequence alignment [50] showing that the identified P1 site at A195 is conserved among CalpT homologs. For gel source data, see Supplementary Figure 1.

Fig. 3 | CalpL and -T form a stable complex and cA$_4$-induced oligomerization of CalpL. a, SEC-MALS traces (solid lines: UV$_{280}$, dashed lines: MW$_{MALS}$) of proteolysis reactions with different combinations of CalpL wt, CalpT wt, and cOA. The SEC-MALS experiment was performed multiple times with slight variations of buffer and concentrations (n = 3 technical replicates). The schematic indicates the molecular species behind the individual peaks. Inset: Crystal structure of the CalpL/T$_{10}$ complex. The distance of 36 Å between the P-1 position (S196) and the protease active site is indicated. **b**, Top: Concentration-normalized small angle X-ray scattering curves recorded at four different concentrations of CalpL. For each experiment, thirty sample intensity frames and sixty buffer intensity frames were collected and averaged. For each data set and angular point the errors were computed following the Poisson statistics. The data points represent the average intensity difference (sample-buffer) and the error bars represent the standard deviation. The experiment was performed once for each concentration. Bottom: Molecular weights from forward scattering I(0) calculated from the SAXS curves plotted vs the concentration. The apparent molecular weight of the protein in the presence of cA$_4$ increases with the concentration. **c**, Protease assays with preformed CalpL/T complexes as indicated in the figure. The experiment was performed three times (n = 3 technical replicates). **d**, Protease assays with CalpL mutants in the positive patch on the backside of the SAVED domain (Fig. 1). The experiment was performed twice (n = 2 technical replicates). For gel source data, see Supplementary Figure 1.

Fig. 4 | CalpT links cA$_4$ detection to the transcription machinery of the cell. a, Structural information about the CalpL/T/S complex as obtained by crystallography, SAXS and AlphaFold2. **b-d**, CalpT and S form a complex that is stable during gelfiltration and disintegrates into CalpL/T$_{10}$ and CalpT$_{23}$/S. The experiments were performed twice (n=2 technical replicates). **e**, His-CalpS and CalpT can be copurified from *E. coli*. Overexpression of CalpS alone leads to copurification of the α- and β-subunits of the *E. coli* RNAP. The complex formations were replicated multiple times in two independent laboratories (n>3 biological replicates). For gel source data, see Supplementary Figure 1.

15

533 **Fig. 5 | Model for CalpL/T/S mediated antiviral defense. a**, The *calpL*, *calpT* and *calpS*
534 genes are located in close proximity to the type III-B CRISPR genes of *Sulfurihydrogenibium*
535 *sp*. YO3AOP1 (modified from [6]). **b**, Once activated, the Cas10 subunit of the RNP synthesizes
536 cA_4 from ATP. The second messenger binds to preformed CalpL/T/S complexes.
537 Oligomerization leads to proteolytic cleavage of CalpT, releasing the $CalpT_{23}/S$ fragment.
538 $CalpT_{23}$ is likely degraded by proteases, allowing CalpS to bind to the RNA polymerase.
539

540   **Methods**

541   **Expression and purification of CalpL**

542   The codon-optimized gene for CRISPR–Lon was cloned into a pET11a vector with an N-

543   terminal 10xHis-TEV tag. Site-directed mutagenesis was performed according to a protocol by

544   Liu et al.[51]. All CalpL constructs were expressed in lysogeny broth (LB) medium. *E. coli*

545   BL21(DE3) cells were grown at 37 °C until an OD_{600} of 0.6-0.8 was reached. Then, protein

546   expression was started by induction with 0.4 mM IPTG, and the cell suspension was incubated

547   at 30 °C for 4.5 h with shaking. Cells were harvested by centrifugation at 4,000*rcf for 25 min.

548   at 20 °C and resuspended in lysis buffer (20 mM Tris, 50 mM NaCl, pH 8.0). The cells were

549   lysed with a sonicator and cell debris was removed by centrifugation at 48,000*rcf for 45 min.

550   at 4 °C. For protein purification, Ni^{2+}-affinity chromatography (20 mM Tris, 50 mM NaCl,

551   pH 8.0; 500 mM imidazole was included for elution) was followed by size-exclusion

552   chromatography (20 mM Tris, 50 mM NaCl, pH 8.0) using a Superdex 200 16/600 column.

553   After that, the His-tag was cleaved off by overnight incubation at 4 °C with a 1:50 molar ratio

554   of protein to TEV protease (20 mM Tris, 50 mM NaCl, pH 8.0). A second Ni^{2+}-affinity

555   chromatography was used to remove the TEV protease and uncleaved protein. The purity of the

556   protein was checked by SDS-PAGE after each purification step. After successful purification,

557   the proteins were concentrated, flash-frozen in liquid nitrogen, and stored at -80 °C in 20 mM

558   Tris, 50 mM NaCl, pH 8.0. The selenomethionine derivative of CalpL was prepared using *E.*

559   *coli* B834 cells and the "SelenoMethionine Medium Complete" kit from Molecular Dimensions

560   according to the instructions. Protein expression and purification were done in the same way as

561   for the native protein.

562

563   **Expression and purification of CalpT**

564   The codon-optimized synthetic gene (BioCat) for CalpT (UNIPROT-ID: B2V8L8), including

565   an N-terminal 10x His-TEV tag was cloned into a pET11a vector. Protein expression was done

566   using the same expression strain and the same conditions as for CalpL. Cells were harvested by

567   centrifugation at 4,000*rcf for 25 min. at 20 °C and resuspended in lysis buffer (25 mM Tris,

568   500 mM NaCl, 10% glycerol, 1 mM DTT, pH 8.0). The cells were lysed with a sonicator and

569   cell debris was removed by centrifugation at 48,000*rcf for 45 min. at 20 °C. For protein

570   purification, Ni^{2+}-affinity chromatography (25 mM Tris, 500 mM NaCl, 1 mM DTT, 10%

571   glycerol, pH 8.0; 1 M imidazole was included for elution) was followed by size-exclusion

572   chromatography (25 mM Tris, 500 mM NaCl, 1 mM DTT, 10% glycerol, pH 8.0) using a

573   Superdex 75 16/600 column. After that, the His-tag was cleaved off by overnight incubation at

17

574   4 °C with a 20:1 ratio (m/m) of protein to TEV protease (25 mM Tris, 500 mM NaCl, 1 mM
575   DTT, 10% glycerol, pH 8.0). A second Ni^{2+}-affinity chromatography was used to separate the
576   TEV protease and uncleaved protein. The purity of the protein was checked by SDS-PAGE
577   after each purification step. After successful purification, the proteins were concentrated, flash-
578   frozen in liquid nitrogen, and stored at -80 °C in 25 mM Tris, 500 mM NaCl, 1 mM DTT, 10%
579   glycerol, pH 8.0.
580

Expression and purification of CalpS

582   The codon-optimized gene of CalpS was purchased from Integrated DNA Technologies (IDT,
583   Coralville, Iowa, USA) as a G-Bock with flanking restriction sites for cloning. *SF* was cloned
584   into NcoI and BamHI restriction sites of vector pEV5HisTEV[52], allowing expressed proteins
585   with an N-terminal 8x His-TEV tag. For expression, *E. coli* C43(DE3) cells with sequencing-
586   verified construct were incubated at 37 °C with shaking at 180 rpm until OD_{600} of the cells was
587   between 0.6 and 0.8. Then, the cell culture was grown at 16 °C overnight after inducing with
588   0.2 mM IPTG. The cell pellet was collected by centrifugation at 4000 rpm (Beckman Coulter
589   Avanti JXN-26; JLA8.1 rotor) at 4 °C for 15 min. For purification, cell pellet was resuspended
590   into buffer A (50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 20 mM imidazole, and 10% glycerol) and
591   lysed by sonication. The cleared cell lysate was loaded onto a 5 mL HisTrap FF column (GE
592   Healthcare) equilibrated with buffer A. The His-tagged SF was eluted in a linear gradient with
593   buffer B (50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 0.5 M imidazole, and 10% glycerol). The his-
594   tag was then removed by incubating with TVE protease at room temperature overnight before
595   recovering TEV-cleaved SF through a HisTrap column again. Size-exclusion chromatography
596   was finally used to purify the SF in SEC buffer (20 mM Tris–HCl, 0.25 M NaCl, 1 mM DTT,
597   10% glycerol, pH 7.5). The purity of SF was evaluated on the SDS-PAGE at each purification
598   step. Concentrated SF was flash-frozen in liquid nitrogen and stored at -70 °C.
599

Co-expression and co-purification of sigma factor (SF) and CalpT

601   For expression of His-tagged CalpS with CalpT, the fragment of CalpT flanking NdeI and XhoI
602   sites was cloned into MCS-2 of vector pCDFDuet™-1 (Novagen, Merck Millipore). The
603   constructs pEV5HisTEV-SF and pCDFDuet-CalpT were co-transformed into *E. coli* C43(DE3)
604   cells. The cell was induced by 0.2 mM IPTG after reaching OD600 of 0.6-0.8 and grown
605   overnight at 16 °C with shaking at 180 rpm.
606   For expression of His-tagged CalpT with CalpS, a G-Block of SF was constructed into MCS-1
607   (NcoI and BamHI) of vector pCDFDuet™-1. *E. coli* C43(DE3) cells were transformed with

18

608   constructs pET11a-CalpT and pCDFDuet-SF and grown at 37°C with shaking at 180 rpm. Cell

609   was induced by 0.2 mM IPTG once OD600 of 0.6-0.8 and cultivated at 16 °C overnight, then

610   purified as described above.

611

Protease assay

613   For protease activity assays CalpL and CalpT were used at a final concentration of c = 4.64 μM

614   each. The different cOAs were used at a final concentration of c = 5.11 μM. The protein

615   solutions were prepared in 20 mM Tris, 50 mM NaCl, pH 8.0 and incubated for 1 hr at 60 °C.

616   Subsequently, the cOA was added and the mixture was incubated for another 1 hr at 60 °C. For

617   SDS-PAGE 3 μl of 4x SDS-loading buffer was added to 9 μl of the sample, the mixture was

618   heated for 5 min at 94 °C and 10 μl were loaded to a 15% polyacrylamide gel, which was run

619   at 250 V for 40 min.

620

Size exclusion chromatography analysis (SEC)

622   To determine the interaction of the complex of CalpS and CalpT with CalpL, the SEC runs

623   were carried out on a Superose6 increase 10/300 chromatography column (GE Healthcare)

624   equilibrated with SEC buffer (20 mM Tris, 0.25 M NaCl, 1 mM DTT, 10% glycerol, pH 8.0).

625   The injected volume of tested sample solution was 200 ul at flow rate of 0.5 ml min^{-1}. The final

626   concentrations were set to c(CalpL) = 63.3 μmol*l^{-1}, c(CalpT/S) = 115.8 μmol*l^{-1}, and c(cA$_4$)

627   = 60 μmol*l^{-1} diluted by using SEC buffer. All samples were incubated at 60 °C for 60 min

628   before cooling down to room temperature and loading onto column.

629

Pull-down assay

631   The magnetic nickel beads-based immobilized metal affinity chromatography (IMAC) was

632   performed to detect releasing of CalpS from CalpL/T/S complex. The complex of His-tagged

633   CalpS and CalpT was incubated with CalpL in binding buffer (20 mM Tris-HCl, pH 7.5, 60

634   mM NaCl, 0.01% Tween^{TM}-20) at 60 °C for 1 hr in presence or absence of cA$_4$. After cooling

635   down to room temperature, the sample solution was mixed with pre-equilibrated beads (Magne,

636   His Ni particle, Promega) with binding buffer on a roller for 20 min at 4 °C. The beads were

637   washed tree times with 300 ul wash buffer (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 mM

638   imidazole, 0.01% Tween^{TM}-20) before eluted twice using 25 ul elution buffer (20 mM Tris-

639   HCl, pH 8.0, 120 mM NaCl, 300 mM imidazole, 0.01% Tween^{TM}-20). The samples from first

640   elution and 20% input were analyzed on the SDS-PAGE. The final concentrations were set to

19

641   c(CalpT/S) = 0.208mg/ml, c(CalpL) = 0.127 mg/ml, and c(cA$_4$) = 2.5 µmol*l^{-1} diluted by

642   binding buffer.

643

Analytical gel filtration and SEC-MALS analysis

645   To investigate the complex formation of CalpS with CalpT and CalpL, analytical gel filtration

646   was carried out on a SD 200 increase 10/300 column. When purifying CalpS, the complex

647   consisting of CalpS and DNA-directed RNA polymerase subunits alpha and beta, eluted in a

648   defined 50 mAU peak on a SD 200 16/600 gel filtration column. This peak was pooled and

649   concentrated to V ≈ 1.5 ml. Thereafter, 440 µl of the complex were incubated with 60 µl of

650   CalpT (350 µM), 60 µl of CalpT-CalpL S152A complex (195 µM) and 60 µl of CalpS buffer

651   (20 mM Tris–HCl, 0.25 M NaCl, 1 mM DTT, 10% glycerol, pH 7.5), respectively. After

652   centrifugation at 15000 rcf and 10 °C for 10 min, each sample was loaded onto a SD 200

653   increase 10/300 column for size-exclusion chromatography in CalpS buffer.

654   For determination of interactions between CalpL and CalpT, SEC-MALS runs were performed

655   at room temperature on an Agilent 1260 Infinity II Prime Bio LC System coupled with a Wyatt

656   miniDAWN® MALS detector, a Optilab rEX refractive index detector and a Superose6 increase

657   10/300 chromatography column (GE Healthcare) equilibrated with 25 mM Tris, 500 mM NaCl,

658   1 mM DTT, 10% glycerol, pH 8.0. Data acquisition and evaluation were carried out using

659   ASTRA 8 software (Wyatt Technologies). The flow rate was set to 0.5 ml min^{-1} and an injection

660   volume of 50 µl was used for the experiments. Final concentrations were set to

661   c(CalpL) = 51 µmol l^{-1}, c(CalpT) = 51 µmol l^{-1}, and c(cA4) = 60 µmol l^{-1} by dilution with

662   25 mM Tris, 500 mM NaCl, 1 mM DTT, 10% glycerol, pH 8.0. The proteins were incubated

663   for 40 min at 60 °C, cA$_4$ was added followed by an additional 20 min incubation at 60 °C. The

664   samples were centrifuged at 15.000*g for 10 min. before injection.

665

Mass spectrometry

667   The gel bands were excised and cut into 1 mm^3 cubes. The samples were destained with 2x

668   rinses each of ethanol, acetonitrile and 25mM ammonium bicarbonate, then subjected to

669   reduction with 10 mM dithiorethritol, followed by alkylation with 20 mM iodoacetamide. The

670   gel pieces with shrunk with acetonitrile and then soaked in 25 mM AmBic with 2 ng/µl trypsin

671   and left to digest overnight at 37 °C. The peptides were soaked from the gel with 1% formic

672   acid and concentrated to 20 µl in a speedvac. Between 1-7 µl of the sample, dependent on

673   original gel coomassie staining, was loaded onto a Eksigent 2D ultra nano HPLC with Sceix

674   5600+ mass spectrometer. The Thermoscientific Acclaim Pepmap 100 trap (20 mm x 75 µm)

20

675   and column (150 mm x 75 μm) were in trap elute configuration with a flow of 5 μl/min and 300

676   nl/min respectively. The peptides were loaded onto the trap and washed for 5 minutes at 100%

677   loading buffer (100% water, 0.05% TFA) before the trap was switched in line with the column

678   and the peptides eluted with a linear gradient over 20 minutes of 98% A to 98% B where A is

679   100% water with 0.1% formic acid and B is 80% acetonitrile, 20% water, 0.1% formic

680   acid. The eluent was sprayed directly into the nanosource of the mass spectrometer. MS data

681   was collected from 400-1250 m/z in positive ionisation for 150 msec. Data dependant

682   acquisition mode was utilized to collect MSMS data from 100-2000 m/z on the 20 strongest

683   peptides with 2-5+ charge states. The peak list was extracted from the .wiff file using

684   MSconvert and the .mgf file searched against an inhouse database of 7000 protein sequences to

685   which the sequences of the proteins of interest were added. The following settings were used

686   in the mascot search, trypsin, and semi trypsin as digest enzymes, fixed modification of

687   carbamidomethyl (c) and variable modifcation of oxidation (M). MS tolerance was set at

688   20 ppm and MSMS at 0.1 Da.

689

690   **Surface plasmon resonance of cOA and CalpT binding to CalpL**

691   All surface plasmon resonance experiments were run on a Biacore[TM] 8K instrument (GE

692   healthcare life sciences), using a streptavidin-functionalized sensor chip (Serie S Sensor Chip

693   SA, GE healthcare life sciences). Data was recorded at a rate of 10 Hz and 25°C flow cell

694   temperature. The running buffer contained 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 5%

695   Glycerol, 0.05% TWEEN20. After three initial injections of 1 M NaCl in 50 mM NaOH (10

696   μL/min, 60 s), the biotinylated CalpL construct, R493C-biotin, was immobilized on the chip

697   (86 nM, 5 μL/min, 180 s). Binding of CoA and CalpT was measured as single cycle kinetics.

698   For the cOAs, a series of seven different concentrations (0.086, 0.26, 0.78, 2.33, 7, 21, 63 nM)

699   were injected at a flow rate of 30 μL/min (contact time: 120s, dissociation time: 600 s). For

700   CalpT and NIS038, a series of seven different concentrations (0.0625, 0.25, 1, 4, 16, 64, 256

701   nM) were injected applying the same parameters as above. The recorded data were double

702   referenced by reference flow cell and blank cycle subtraction and data was analysed and fitted

703   using the Biacore Insight Evaluation Software.

704

705   **X-ray crystallography**

706   Pure CalpL protein was concentrated to 20 mg/ml and crystallized at 20 °C using a Gryphon

707   pipetting robot (Art Robbins) and commercial crystallization screens (Molecular Dimensions)

708   using sitting drop plates. Hexagonal crystals appeared after one day in condition D7 of the

21

709   JCSG+ screen. Several rounds of optimization in sitting- and hanging drop plates were
710   performed to achieve well-diffracting crystals. The final crystallization condition was 0.1 M
711   Tris-Cl pH 8.0, 38.8% PEG400, 0.29 M Li_2SO_4. The SeMet derivative (see above) was
712   crystallized under similar conditions and yielded identical crystals. The crystals were harvested
713   without further cryo-protection and a diffraction dataset was recorded at beamline P13
714   ($\lambda = 0.9795$) operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg,
715   Germany)[53]. The diffraction data were automatically processed with XDS[54]. The structure was
716   solved using phenix.autosol and refined with phenix.refine[55]. Further model building was
717   performed in Coot[56] and figures were prepared with PyMOL (www.pymol.org). The geometry
718   of the model was checked with MolProbity[57]. The molprobity score was 1.43, the clashscore
719   4.53 and the Ramachandran statistics (outliers/favored) were 0.0/97.0%.

720   The same crystallization condition was used to obtain CalpL crystals for soaking with cA_4.
721   After harvesting the crystals, they were incubated for approximately 3 minutes in a solution of
722   mother liquor supplemented with 5% PEG 400 and 5 mM cA_4. Diffraction experiments were
723   done at beamline P13 ($\lambda = 0.\,0.97626$) at the PETRA III storage ring (DESY, Hamburg,
724   Germany)[53]. A full dataset was recorded and automatically processed with XDS[54]. Molecular
725   replacement was run with PHASER[58] using CalpL as search model. Further refinement of the
726   structure was done with phenix.refine[55]. Using Coot[56], cyclic tetraadenylate could be fitted
727   perfectly into a defined difference electron density inside the cA_4 binding pocket. The
728   molprobity score was 1.32, the clashscore 2.85 and the Ramachandran statistics
729   (outliers/favored) were 0.4/96.6%.

730   To obtain the crystal structure of the CalpL/T_{10} complex, CalpL and CalpT protein solutions
731   were mixed at 1:1 molar ratio (155 μM each) and incubated at 50 °C for 40 min. After
732   incubation, cA_4 was added to a final concentration of 175 μM followed by 20 min of incubation
733   at 50 °C. A total volume of 400 μl was loaded to a SD 200 increase 10/300 column and size-
734   exclusion chromatography was conducted using CalpL buffer. The complex eluted in one single
735   peak which was pooled and concentrated to approximately 30 mg/ml. Sitting drop
736   crystallization plates were set up as described above. Crystals were obtained after several days
737   in condition E2 of the JCSG+ screen (2 M ammonium sulfate, 0.1 M ammonium cacodylate,
738   0.2 M NaCl at pH 6.5). The crystals were harvested with 35% glycerol for cryo-protection and
739   a diffraction dataset was recorded at beamline P13 ($\lambda = 0.\,0.9762$) operated by EMBL Hamburg
740   at the PETRA III storage ring (DESY, Hamburg, Germany)[53]. Automatic data processing was
741   achieved using XDS[54]. CalpL was used as search model for molecular replacement with
742   PHASER [58]. Structure refinement and model building was done with phenix.refine[55] and

22

743 Coot[56], respectively. The molprobity score was 2.36, the clashscore 11.93 and the
744 Ramachandran statistics (outliers/favored) were 0.9/92.3%.

745 Geometric parameters of all described structures were checked with MolProbity[57]. All figures
746 were prepared with PyMOL (www.pymol.org).

747

Small-angle X-ray scattering

749 Small-angle X-ray scattering (SAXS) experiments were conducted at the P12[59] beamline of the
750 Petra III synchrotron (Hamburg, Germany), in two separate experimental sessions. The
751 parameters for the data collections are reported in Extended Data Table 2. The scattering
752 intensity, $I(s)$ was collected as a function of the momentum transfer defined as $s=(4\pi\sin\theta)/\lambda$,
753 where 2θ is the scattering angle and λ the X-ray wavelength employed.

754 The buffer employed for all SAXS experiments, also employed for SEC and for background
755 subtraction, contained 20 mM Tris pH=8.0, 50 mM NaCl. CalpL, CalpL/T$_{10}$ and CalpL
756 monomer SAXS curves were collected by SEC-SAXS at room temperature[60], employing a
757 SD200 5/150 increase SEC column (GE Healthcare) online to the SAXS flow capillary. For
758 each SEC-SAXS run, 15 min elutions at 0.3 mL/min flow rate were performed, collecting 900x
759 1 s exposures on the eluate.

760 Concentration series of CalpL in the range 1-5 mg/mL with and without a 1.2-fold molar excess
761 of cA$_4$ were collected in batch mode after centrifugation 30 min at 30000 xg (5 °C), and 30
762 exposures of 0.1 s were collected while flowing 35 μL of solution through the 1 mm quartz
763 capillary. The apparent molecular weights for these measurements are obtained from the
764 forward scattering, $I(0)$, using the $I(0)$ of a SAXS curve from bovine serum albumin at 1.9
765 mg/mL in a HEPES buffer as secondary standard.

766 The primary data reduction was performed with the program SASFLOW[61], including automatic
767 selection of the exposures to monitor for radiation damage prior to data averaging. The reduced
768 data inspected and processed to obtain the overall protein parameters using PRIMUS[62] and the
769 programs of the ATSAS suite[63]. Comparison of experimental SAXS curves with the crystal
770 structures was performed with CRYSOL[64] *Ab initio* modeling of low-resolution protein
771 structures was performed with DAMMIF[65], and MONSA[66] for multi-phase modeling, repeating
772 10 modeling runs. The resulting models were compared and averaged using DAMAVER[67],
773 providing a normalized spatial discrepancy (NSD) value[68] that reflects the stability of the
774 structural reconstructions (stable reconstructions have NSD<1). SAXS curves of CalpL at 2, 3,
775 5 mg/mL in the presence of cA$_4$ were globally fitted as a mixture of monomer and dimer, using
776 the program SASREFMX[69]. The mixture results in an overall scattering intensity which is the

23

777 linear combination of the scattering intensity of monomer and dimer species at varying volume
778 fractions along the concentration series.
779

780 **Pulsed EPR experiments**

781 For site-specific spin labelling, a single cysteine mutant E119C of CalpT of was expressed and
782 purified as described for wild-type CalpT. After purification, 250 μl of a 315 μM CalpT E119C
783 solution was bound to Ni^{2+}-NTA beads. These were washed with 10 ml reducing buffer (25 mM
784 Tris, pH = 8, 250 mM NaCl, 1 mM TCEP, 10% glycerol) and 20 ml wash buffer (25 mM Tris,
785 pH = 8, 250 mM NaCl, 10% glycerol). Thereafter, the protein was eluted in buffer containing
786 MTSSL (25 mM Tris, pH = 8, 250 mM NaCl, 1 M imidazole, 10% glycerol, 0.6 mM MTSSL).
787 A PD10 desalting column was used to remove imidazole and free spin label, as well as for
788 buffer exchange of CalpT E119R1, wild-type CalpL and CalpL S152A. All final buffers
789 contained D_2O instead of H_2O and no reducing agent. An activity assay was done as described
790 using the prepared samples with a final protein concentration of 55 μM and deuterated buffer
791 (25 mM Tris, pH = 8, 250 mM NaCl, 5% glycerol). The success was checked on SDS-PAGE.
792 All samples were flash frozen to $N_{2\,(l)}$ and sent on dry ice for the measurement. The labelling
793 efficiency was determined to be 103% by cw-EPR spectroscopy (average of two
794 measurements).
795 For pulse EPR measurements, samples of spin-labeled CalpL/T E119R1 in presence or absence
796 of 1 molar equivalent cA4 were mixed with 45% (v/v) deuterated ethylene glycol to yield
797 27.5 μM CalpL/T E119R1 in 65 μl final volume. Samples were transferred to 3 mm EPR quartz
798 tubes, flash-frozen and stored in liquid nitrogen until use.
799 Pulsed electron-electron double resonance (PELDOR)[70] distance measurements were obtained
800 at Q-band frequency (34 GHz) on a Bruker ELEXSYS E580 spectrometer with 3 mm
801 cylindrical resonator (ER 5106QT2-2w, TE012 mode) using a pulse travelling wave tube
802 (TWT) amplifier (Applied Systems Engineering) with nominal output of 150 W and an arbitrary
803 waveform generator for rectangular pulses.
804 PELDOR experiments were performed with the 4-pulse DEER[71,72] pulse sequence $(\pi/2(\nu_A) - $
805 $\tau_1 - \pi(\nu_A) - (\tau_1 + t) - \pi(\nu_B) - (\tau_2 - t) - \pi(\nu_A) - \tau_2 -$ echo) at 50 K, with a frequency offset (pump
806 – detection frequency) of +80 MHz (~3 mT). Shot repetition time (SRT) was set to 2.5 ms; τ_1
807 was set to 380 ns, and τ_2 was set to 5000 ns. Pulse lengths were 16 and 32 ns for $\pi/2$ and π
808 detection, and 12 ns for the inversion π pump pulse. Unwanted echoes were suppressed with a
809 16-step phase cycle and nuclear modulation was averaged by adding 16 traces with τ_1

24

810   incremented by 8 ns. The pump pulse was placed on the resonance frequency of the resonator

811   and applied to the maximum of the nitroxide field-swept spectrum.

812   PELDOR data were analyzed using the ComparativeDeerAnalyzer version 2.0[73,74] within

813   DeerAnalysis2022[75]; shown are the respective consensus fits and distance distributions.

814

Ribonuclease assay

815

816   Ribonuclease activity of cleaved CalpT (23-kD fragment) was assayed by incubating full-length

817   CalpT with CalpL and five different fluorescent-labelled RNA substrates, which were

818   synthesised with the fluorescent dye (6-FAM) attached at 5' end or at 3' end (purchased from

819   Integrated DNA Technologies (IDT), Extended Data Fig. 7c). The mixture of CalpL (5.5 μM)

820   and CalpT (5.5 μM) was incubated at 60 °C in 20 mM Tris-HCl, pH 8.0, 50 mM NaCl and 1

821   mM EDTA for 15 min, cA_4 (10 μM) was then added and the mixture was incubated for another

822   15 min at 60 °C, followed by adding one of the above RNA substrates into the mixture,

823   incubating for an additional 30 min at 60°C. Finally, 6 μl of the sample was analyzed on SDS-

824   PAGE (NuPAGE Bis-Tris Gel, Thermo Fisher Scientific) by heating at 95 °C for 5 min with 2

825   μL of SDS-PAGE loading dye (Thermo Fisher Scientific; NuPAGE Sample Reducing Agent

826   and LDS Sample Buffer). The remaining 14 μl of the sample were loaded to 20% acrylamide,

827   7 M urea, 1×TBE denaturing gel, which was run at 30W, 45 °C for 2 hr. The gel was finally

828   imaged by Typhoon FLA 7000 imager (GE Healthcare) at a wavelength of 532 nm (pmt

829   600~700).

830

Ribonuclease target motif profiling (RNase-Seq)

831

832   To investigate MazF activity, 20 U mRNA Interferase™ -MazF (TaKaRa, Cat# 2415A) were

833   incubated with 400 ng of a single-stranded RNA library containing 10 random bases in 1X

834   MazF Buffer (200 mM sodium phosphate, pH 7.5, and 0.05% Tween-20) at 37 °C. After 10

835   min or 2 hr, reactions were stopped by placing the samples on ice. Ribonuclease activity of

836   CalpT (23 kDa fragment) was investigated by incubating a solution containing 4.64 μM CalpL

837   and 5.57 μM CalpT in 20 mM Tris pH 8.0, 50 mM NaCl at 37 °C. After 30 min, 400 ng of a

838   ssRNA library was added together with cOA_4 to a final concentration of 5.57 μM. The reaction

839   was incubated for 10 min or 2 hr at 37 °C. To prepare next generation sequencing libraries, RT-

840   PCR was performed on 1 μl of each sample in a 10 μl reaction containing 1x KAPA HiFi

841   HotStart ReadyMix (pre-heated for 5 min at 98 °C; Roche, Cat# KK2602), 15 U WarmStart

842   RTx Reverse Transcriptase (NEB, Cat# M0380) and 0.5 μM staggered MiSeq gRNA primer

843   mix using the following temperature conditions: 15 min 65 °C, 3 min 72 °C, 30 sec 98 °C, 20

25

844   cycles of: Denaturation (10 sec at 98 °C), annealing (20 sec at 65 °C) and extension (1 min at

845   72 °C); Final extension: 5 min at 72 °C. Barcodes and Illumina-compatible constant handles

846   were added using a secondary NEBNext (NEB) PCR. Samples were pooled and column-

847   purified using QIAprep Spin columns. The final library was quantified using a NanoDrop

848   photospectrometer and sequenced on an Illumina MiSeq using the v2 chemistry.

849   The primers used for ribonuclease target motif profiling are listed in Extended Data Fig. 7e.

850

851   **Structural predictions with AlphaFold2**

852   The source code of the AlphaFold2 algorithm was downloaded from

853   https://github.com/AlphaFold and installed as described https://github.com/AlphaFold. The

854   algorithm was run locally using the CASP14 preset or via ColabFold[76].

855

856   **Statistics and reproducibility**

857   Information concerning statistics and reproducibility for the experiments shown in this study

858   are given in the figure legends of the corresponding experiments. The key findings of this study

859   (enzymatic activities and formation of macromolecular complexes) have been reproduced in

860   two laboratories (G.H. at the Institute of Structural Biology of the University of Bonn, Germany

861   and M.F.W. at the Biomedical Sciences Research Complex of the University of St Andrews,

862   Scotland, UK).

863

Additional references

49. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C. & Laydon, A. Highly accurate protein structure prediction for the human proteome. *Nature* 596, 590-596 (2021).

50. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* 42, W320-W324 (2014).

51. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology* 8, 91 (2008).

52. Rouillon, C., Athukoralage, J. S., Graham, S., Grüschow, S. & White, M. F. Investigation of the cyclic oligoadenylate signaling pathway of type III CRISPR systems. *Methods Enzymol* 616, 191-218 (2019).

53. Cianci, M., Bourenkov, G., Pompidor, G., Karpics, I., Kallio, J., Bento, I., Roessle, M., Cipriani, F., Fiedler, S. & Schneider, T. R. P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high-and low-energy phasing with variable beam focusing. *Journal of synchrotron radiation* 24, 323-332 (2017).

54. Kabsch, W. Automatic-Indexing of Rotation Diffraction Patterns. 21, 67-71 (1988).

55. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S. & McCoy, A. J. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology* 75, 861-877 (2019).

56. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* 60, 2126-2132 (2004).

57. Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J. & Chen, V. B. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* 27, 293-315 (2018).

58. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. Phasercrystallographic software. *Journal of Applied Crystallography* 40, 658-674 (2007).

59. Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Jeffries, C. M., Franke, D., Mark, D., Zengerle, R., Cipriani, F., Fiedler, S., Roessle, M., Svergun, D. I. & IUCr. Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *Journal of Applied Crystallography* 48, 431-443 (2015).

60. Graewert, M. A., Da Vela, S., Gräwert, T. W., Molodenskiy, D. S., Blanchet, C. E., Svergun, D. I. & Jeffries, C. M. Adding size exclusion chromatography (SEC) and light scattering (LS) devices to obtain high-quality small angle X-ray scattering (SAXS) data. *Crystals* 10, 975 (2020).

61. Franke, D., Kikhney, A. G. & Svergun, D. I. Automated acquisition and analysis of small angle X-ray scattering data. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 689, 52-59 (2012).

62. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. 36, 1277-1282 (2003).

63. Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A. & Borges,

914   C. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data
915   analysis. *Journal of Applied Crystallography* 54, 343-355 (2021).
916 64. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL– a Program to Evaluate X-ray
917   Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of*
918   *Applied Crystallography* 28, 768-773 (1995).
919 65. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape
920   determination in small-angle scattering. 42, 342-346 (2009).
921 66. Svergun, D. I. Restoring low resolution structure of biological macromolecules from
922   solution scattering using simulated annealing. *Biophysj* 76, 2879-2886 (1999).
923 67. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-
924   angle scattering. 36, 860-864 (2003).
925 68. Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution
926   structural models. 34, 33-41 (2001).
927 69. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M.,
928   Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. New developments in the
929   ATSAS program package for small-angle scattering data analysis. 45, 342-350 (2012).
930 70. Milov, A., Salikohov, K. & Shirov, M. Application of Endor in Electron-Spin Echo for
931   Paramagnetic Center Space Distribution in Solids. *Fizika Tverdogo Tela* 23, 975-982
932   (1981).
933 71. Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H. W. Dead-time free
934   measurement of dipole-dipole interactions between electron spins. *Journal of Magnetic*
935   *Resonance (San Diego, Calif.: 1997)* 142, 331-340 (2000).
936 72. Larsen, R. G. & Singel, D. J. Double electron–electron resonance spin–echo
937   modulation: Spectroscopic measurement of electron spin pair separations in
938   orientationally disordered solids. *The Journal of chemical physics* 98, 5134-5146
939   (1993).
940 73. Worswick, S. G., Spencer, J. A., Jeschke, G. & Kuprov, I. Deep neural network
941   processing of DEER data. *Science advances* 4, eaat5218 (2018).
942 74. Fábregas Ibáñez, L., Jeschke, G. & Stoll, S. DeerLab: a comprehensive software
943   package for analyzing dipolar electron paramagnetic resonance spectroscopy data.
944   *Magnetic Resonance* 1, 209-224 (2020).
945 75. Jeschke, G., Chechik, V., Ionita, P. & Godt, A. DeerAnalysis2006—a comprehensive
946   software package for analyzing pulsed ELDOR data. *Applied Magnetic Resonance* 30,
947   473-498 (2006).
948 76. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. & Steinegger, M.
949   ColabFold: making protein folding accessible to all. *Nature Methods* 1-4 (2022).
950 77. Cha, S. S., An, Y. J., Lee, C. R., Lee, H. S., Kim, Y. G., Kim, S. J., Kwon, K. K., De
951   Donatis, G. M., Lee, J. H., Maurizi, M. R. & Kang, S. G. Crystal structure of Lon
952   protease: molecular architecture of gated entry to a sequestered degradation chamber.
953   *EMBO J* 29, 3520-3530 (2010).
954 78. Zorzini, V., Mernik, A., Lah, J., Sterckx, Y. G. J., De Jonge, N., Garcia-Pino, A., De
955   Greve, H., Versées, W. & Loris, R. Substrate Recognition and Activity Regulation of
956   the Escherichia coli mRNA Endonuclease MazF. *Journal of Biological Chemistry* 291,
957   10950-10960 (2016).
958 79. Simanshu, D. K., Yamaguchi, Y., Park, J.-H., Inouye, M. & Patel, D. J. Structural basis
959   of mRNA recognition and cleavage by toxin MazF and its regulation by antitoxin MazE
960   in Bacillus subtilis. *Molecular cell* 52, 447-458 (2013).
961 80. Hagelueken, G., Ward, R., Naismith, J. H. & Schiemann, O. MtsslWizard: In Silico
962   Spin-Labeling and Generation of Distance Distributions in PyMOL. *Applied Magnetic*
963   *Resonance* 42, 377-391 (2012).

28

81. Campagne, S., Marsh, M. E., Capitani, G., Vorholt, J. A. & Allain, F. H. T. Structural basis for− 10 promoter element melting by environmentally induced sigma factors. *Nature structural & molecular biology* 21, 269-276 (2014).
82. Lane, W. J. & Darst, S. A. The structural basis for promoter− 35 element recognition by the group IV σ factors. *PLoS biology* 4, e269 (2006).
83. Li, L., Fang, C., Zhuang, N., Wang, T. & Zhang, Y. Structural basis for transcription initiation by bacterial ECF σ factors. *Nat Commun* 10, 1153 (2019).

Acknowledgements

Author contributions

C.R. and G.H. conceived and supervised the study and performed initial protein expression and crystallization experiments on CalpL. R.S. and W.B. cloned the initial CalpL construct. N.S., C.R., M.F.W. and G.H. designed experiments. N.S. optimized the purification of CalpL and CalpT, crystallized CalpL, CalpL/cA$_4$ and CalpL/T$_{10}$ and established and performed the cleavage assays, SEC-MALS and DLS experiments. N.S and M.F.P. cloned all mutants. G.H. and N.S. solved and refined the CalpL, CalpL/cA$_4$ and CalpL/T$_{10}$ structures. J.M. and M.G. designed and performed the SPR experiments. H.C. and M.F.W. planned and performed the ribonuclease assay. H.C., N.S. and M.F.W. cloned and purified CalpS and performed binding and coexpression experiments involving CalpS. S.D.V. performed SAXS experiments. S.D.V. and D.S. performed the SAXS data analysis and interpretation and wrote the corresponding sections. B.E.B. and K.A. performed the pulsed EPR experiments, analyzed the data, prepared figures and wrote the corresponding sections. K.B. and J.L.S.B. designed and performed the RNase-Seq assay. C.R., M.F.W., H.C., N.S., and G.H. analyzed the data and wrote the paper. All authors discussed the results and commented on the manuscript at all stages.

29

1001 **Competing interest statement**

1002 The authors declare no competing interests.

1003

1004 **Additional information statement**

1005 A previous version of this work has been published as preprint:
1006 https://doi.org/10.1101/2021.12.06.471393. Requests for materials should be addressed to G.H.
1007 (hagelueken@uni-bonn.de).

1008

1009 **Data availability statement**

1010 The crystal structures have been deposited in the PDB with the accession codes 7QDA
1011 [https://doi.org/10.2210/pdb7QDA/pdb], 8B0R [https://doi.org/10.2210/pdb8B0R/pdb], 8B0U
1012 [https://doi.org/10.2210/pdb8B0U/pdb]. The SAXS data and models have been deposited in the
1013 SASBDB with the accession codes: SASDQM4 [https://www.sasbdb.org/data/SASDQM4],
1014 SASDQN4 [https://www.sasbdb.org/data/SASDQN4], SASDQP4
1015 [https://www.sasbdb.org/data/SASDQP4], SASDQQ4
1016 [https://www.sasbdb.org/data/SASDQQ4]. The following PDB entries have been used in this
1017 study: 2H27 [https://doi.org/10.2210/pdb2H27/pdb], 3K1J
1018 [https://doi.org/10.2210/pdb3K1J/pdb], 4ME7 [https://doi.org/10.2210/pdb4ME7/pdb], 4IZJ
1019 [https://doi.org/10.2210/pdb4IZJ/pdb], 4LUP [https://10.2210/pdb4LUP/pdb], 5ZX2
1020 [https://doi.org/10.2210/pdb5ZX2/pdb], 5CR2 [https://doi.org/10.2210/pdb5CR2/pdb], 6VM6
1021 [https://doi.org/10.2210/pdb6VM6/pdb], 6SCE [https://doi.org/10.2210/pdb6SCE/pdb],
1022 7RWK [https://doi.org/10.2210/pdb7RWK/pdb]

1023

1024 **Code availability statement**

1025 No custom code was used in this work.

1026

Extended Data Figure legends

Extended Data Fig. 1 | Purification and Structure of CalpL. a, Gelfiltration chromatography (Superdex 200 16/60) of CalpL. Inset: SDS-PAGE analysis of the fractions indicated by the black bar in the chromatogram. The experiment was performed multiple times (n > 3 biological replicates). **b,** TM-prediction by the TMHMM 2.0 server[28] vs experimental structure. **c,** Representative electron density of the SeMet CalpL crystal structure. The structural model is drawn in ball-and-stick representation. Selected residues are labeled. The black mesh is a $2mF_o$-DF_c electron density map contoured at 1.0 σ. **d,** Topology diagram of CalpL. For gel source data, see Supplementary Figure 1.

Extended Data Fig. 2 | CalpL in comparison to structurally related proteins. a, CalpL is drawn as a cartoon model color-coded as in Fig. 1. The Lon protease from *T. onnorineus* (PDB-ID: 3K1J, DALI Z-score: 12.8[77]) is shown as a white cartoon model. **b,** Table listing proteins with similar domain structures. **c,** Surface electrostatics of the Lon protease active site region. The catalytic dyad is marked. The grey line marks the likely substrate binding site. **d,** Superposition of CalpL active site with the acyl-enzyme intermediate of yellowfin asciitis virus protease. CalpL is in sticks representation and color-coded as in Fig. 1. Chain D of structure 4IZJ[27] (residues 630-640) was superimposed on the corresponding residues of CalpL (150-160) leading to an r. m. s. d. of 0.314 Å. Of 4IZJ, only the acyl-enzyme intermediate is shown in sticks mode. Selected residues and the positions of the P1-P3 sites are indicated. **e,** Superposition of CalpL (color scheme as in Fig. 1) with the Cap4 protein (white, PDB-ID: 6VM6[23]). **f,** Superposition of the CalpL SAVED domain (color scheme as in Fig. 1) with the Cap5 protein (white, PDB-ID: 7RWK[29]). **g,** Superposition of the CalpL SAVED domain (color scheme as in Fig. 1) with the CARF domains of the Can1 protein (white, PDB-ID: 6SCE[17]).

Extended Data Fig. 3 | The CalpL/cA4 complex a, Close-up of cA4 (green) bound to the SAVED domain of CalpL. The blue mesh is a $2mF_o$-DF_c electron density map contoured at 1.0 σ. **b,** Superposition of CalpL apo (white) onto the cA4 complex structure (color coded as in Fig. 1). **c,** Structural alignment of the SAVED domains of CALP/cA4 and Cap4/cA3 (white).

Extended Data Fig. 4 | CalpT is a MazF homolog and the target of the CalpL protease. a, b, A superposition of the predicted CalpT structure (compare Fig. 2B) with one monomer of the MazF/ssRNA complex (purple/orange) (PDB-IDs: 5CR2[78]). The AlphaFold2[33] prediction confidence is mapped onto the CalpT structure (pLDDT[49], predicted local distance difference test). **b,** A superposition of the predicted CalpT structure (compare Fig. 2b) with one monomer of the MazE/F complex (PDB-IDs: 4ME7[79]). The AlphaFold2[33] prediction confidence is mapped onto the CalpT structure (pLDDT[49], predicted local distance difference test). **c,** Gel filtration chromatography (Superdex 75 16/60) of CalpT The experiment was performed multiple times (n > 3 biological replicates). According to the MALS data in Fig. 3, isolated CalpT behaves as a monomer. **d,** Peptide fingerprints of cleavage bands. The indicated gel-bands were cut from the gel and submitted for identification at the Mass spectrometry and proteomics facility at the University of St Andrews (Fife, UK, https://mass-spec.wp.st-andrews.ac.uk). Red letters indicate peptides that were identified in the respective sample. The experiment was performed once. **e,** Mutational analysis of potential CalpL cleavage sites in CalpT. The positions of the mutants are indicated as magenta spheres on the right. (pLDDT[49], predicted local distance difference test) The experiment was performed twice (n = 2 technical replicates). **e)** SDS-PAGE analysis of the fractions indicated by the black bar in a) The experiment was performed multiple times (n > 3 technical replicates). For gel source data, see Supplementary Figure 1.

Extended Data Fig. 5 | Characterization of the CalpL/T complex. a, Single cycle kinetics SPR data of the CalpL/T interaction. The interaction is very strong but cannot be satisfyingly fitted with a 1:1 binding model. The experiment was performed twice (n = 2 technical replicates). **b,** As **a**), but an artificial construct of an unspecific VHH fused to CalpT was used as analyte in this experiment. The interaction is very similar to the CalpL/T interaction. The experiment was performed twice (n = 2 technical replicates). **c,** Schematics of two artificial constructs containing the CalpL cleavage site. **d,** CalpL cleaves an artificial construct of an unspecific VHH fused to CalpT$_{10}$ but not a construct of two VHHs fused by the CalpL cleavage site. The experiment was performed once. **e,** SEC-MALS traces (solid lines: UV$_{280}$, dashed lines: MW$_{MALS}$) of proteolysis reactions with different combinations of CalpL S152A, CalpT, and cOA. The schematic indicates the molecular species behind the individual peaks. The experiments were performed twice with slightly different buffer conditions (n=2 technical replicates). **f,** Binding of CalpL wt to the indicated CalpT mutants in the absence of cA4. The schematic indicates the position of the mutant in the CalpL/T complex. The experiments were performed once. **g,** Representative electron density of the CalpL/T$_{10}$ crystal structure. Selected residues are labeled. The black mesh is a 2mF$_o$-DF$_c$ electron density map contoured at 1.0 σ. **h,** SEC-SAXS experiment of the CalpL/T$_{10}$ complex. The experiment was performed once. Thirty sample intensity frames and sixty buffer intensity frames were collected and averaged. For each data set and angular point the errors were computed following the Poisson statistics. The data points represent the average intensity difference (sample-buffer) and the error bars represent the standard deviation. For gel source data, see Supplementary Figure 1.

Extended Data Fig. 6 | cA$_4$ induced oligomerization of CalpL studied by DLS and SAXS. a, Dynamic light scattering experiments (six timeseries, each series marked by a dashed circle, single data points are shown) at different protein concentrations and in the absence (t=0: light grey to t= 60 min: dark grey) and presence (t=0: cyan to t= 60 min: violet) of cA4 reveal a cA4-dependent oligomerization of CalpL. The experiment was performed twice (n=2 technical replicates) **b,** SAXS experiments at different concentrations. The experiments were performed once. For each experiment, thirty sample intensity frames and sixty buffer intensity frames were collected and averaged. For each data set and angular point the errors were computed following the Poisson statistics. The data points represent the average intensity difference (sample-buffer) and the error bars represent the standard deviation. **c,** *Ab initio/rigid-body* model of a CalpL dimer created with DAMMIF and SASREFMX by a global fit of a monomer-dimer mixture to the different concentrations (red lines). The crystal structure of the CalpL monomer is shown on the same scale.

Extended Data Fig. 7 | Probing the RNase activity of the activated toxin and checking for cA4 induced dimerization of CalpT with pulsed EPR. a, Fluorescence image of the denaturing PAGE to determine ribonuclease activity of the reactions in **b**) against six fluorescently labelled RNA substrates (listed in **c**)). No cleavage was observed after 30 min incubation with RNAs at 60 °C. The experiment was performed three times (n=3 biological replicates) **b,** SDS-PAGE analysis of cA4-induced cleavage of CalpT (33 kDa) by CalpL. Cleavage is complete after 60 min at 60 °C. The experiment was performed three times (n=3 biological replicates) **c,** Sequences of the RNA substrates **d,** left: MazF was incubated with a single stranded RNA library containing 10 random bases. Illumina sequencing was used to check for sequences that were cleaved by MazF. Compared to a control reaction without MazF, sequences containing the known MazF target site (ACA) were depleted. right: same experiment but with CalpL/T ± cA4 instead of MazF. No off-diagonal sequences and hence no ssRNase activity were observed. The experiment was performed two times (n=2 biological replicates). **e,** Oligonucleotides for the experiments in **d**) **f,** AlphaFold2 dimer models of CalpT$_{23}$. **g,** Best model (pLDDT[49], predicted local distance difference test) including MTSSL spin label[80]. **h,** X-

band cw-EPR spectrum of of CalpL/T E119R1. The amount of free label (sharp spikes) is ~10%. The labelling efficiency determined as ~100%. **i,** PELDOR time traces of CalpL/T E119R1 in the presence (red) and absence (black) of cA4. **j,** Consensus distributions and corresponding uncertainty bands. Colored bars indicate reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable; red: no quantification possible). Predicted distance calculated with mtsslWizard[80]. The EPR experiment was performed twice (n=2 technical replicates). For gel source data, see Supplementary Figure 1.

Extended Data Fig. 8 | AlphaFold2 predictions of CalpS. a, Prediction of CalpS alone. The protein is shown as cartoon and colored according to the prediction confidence (pLDDT[49], predicted local distance difference test) **b,** Prediction of the CalpT/S complex. **c,** Superposition of CalpS with 4LUP[81] and 2H27[82] identify the DNA binding regions of CalpS. **d,** Model of CalpS in the context of a RNAP/ECF σ-factor/promotor complex (PDB: 5ZX2[83], grey, yellow, green) from *M. tuberculosis*. Note that the linker region between the σ2 and σ4 subunits of CalpS has been cut to allow the superposition of the σ2 and σ4 domains onto those of 5ZX2. The linker is long enough to bind to the RNAP in a similar way as the σ-factor in the 5ZX2 structure (yellow).

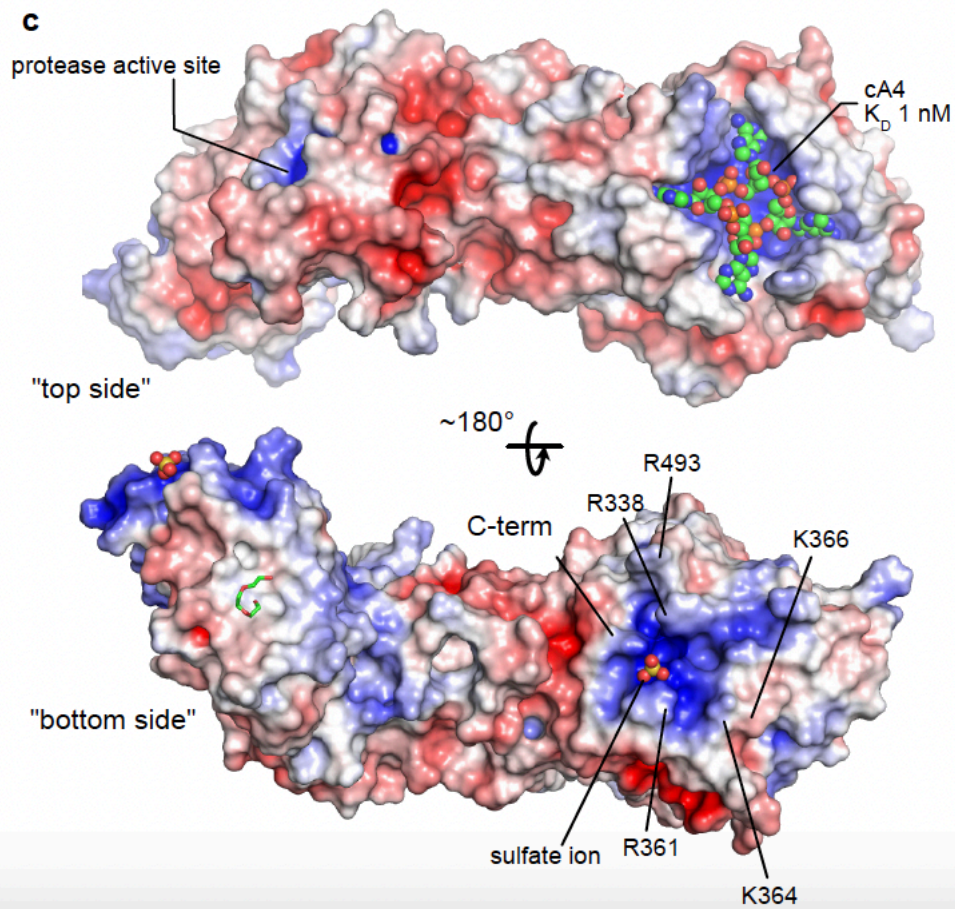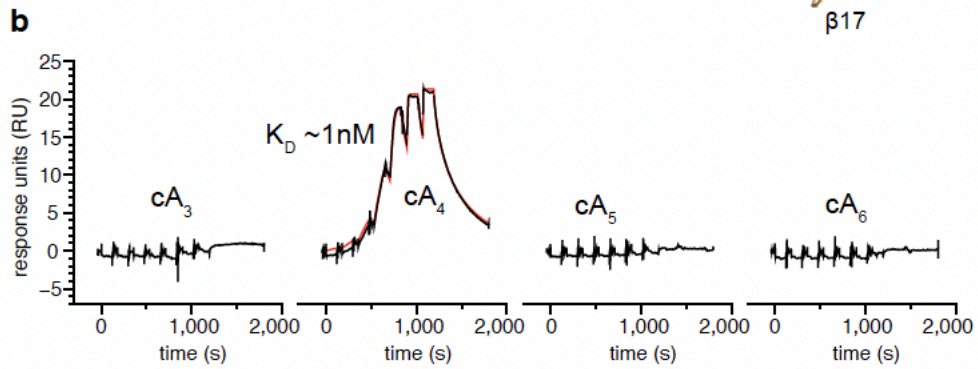**Extended Data Table 1 | Data collection and refinement statistics for crystallographic structures.**
*Values in parenthesis are for the highest-resolution shell.
One crystal was used for each data collection.

Extended Data Table 2 | SAXS data collection and parameters.
*Rg (radius of gyration) from Guinier approximation, †Rg from real-space pair distance distribution function, ‡Largest intramolecular distance, Dmax, §Bayesian molecular mass (Mr) estimate and credibility interval (>90% probability), ‖Porod volume from regularized curve, ¶Mr from VP, #,**Volume and Mr of the *ab initio* models, **Normalized Spatial Discrepancy of the *ab initio* reconstructions, ††The χ^2 value is given for the most representative DAMMIF *ab initio* reconstruction.

a N-terminal domain; Lon protease domain; SAVED domain; protease active site; S152; K193; αG; αA; αC; αD; β2; β1; N; αJ; "hinge"; αK; αM; αL; αP; αS; αO; C; αQ; β19; β17; cOA binding site

b response units (RU); $K_D \sim 1nM$; cA_3; cA_4; cA_5; cA_6; time (s)

c protease active site; $cA4$ K_D 1 nM; "top side"; ~180°; C-term; R493; R338; K366; "bottom side"; sulfate ion; R361; K364

a

MazF-like   DUF 2080

CalpT

CalpS   CalpL

Sulfurihydrogenibium sp. YO3AOP1

b

CalpT model
pLDDT

<50 low confidence   >90 high confidence

DUF2080-like   A201   A195   αG   αF   C

A172

αB

αE

αC   αD

MazF-like   RNA binding site in MazF

αA   N

~23 kDa   ~10 kDa

c

15% SDS-PAGE, Coomassie

d

cleavage site

170	180	190	200	
RIFEILAEPEE...QVIETAPE.RISEFVLR..HVASTSQKATY				CalpT
RIFNIIDELEKEEYNIIPLSPE.REENYRYT..AVASTEKNTAL				C1DV82
RIFDILEEED....FMNSIYNSLVEREIPLPEVAAAGEKDTFAE				WP_097000650.1
RVFDLIEEAD....NIIYLAPT.KL..EKQP..LSAGKENTVYT.				NPA52460.1
FIFEEDENWD....NVISVDTT.KHDDIKFA..LAASVEKTTAR				HEK25337.1
RVFEE...EE....NVIQISPE.REKEILIP..LVAGREKTVFS				WP_200674064.1
RIFEELLE.E....NVIQIPPE.REEQIKLK..LVAGKEKLTSR				A0A432RDI1
LPFEQLPELE....NTIKLPPK.RE..ILLE..LVAGKEKTTAF				HHG75456.1

225

a

CalpL/T
CalpL/T+cA$_4$
CalpL
CalpT
cA$_4$

CalpT
23 kDa
10 kDa

CalpL
58 kDa

CalpL

K193 S196 A201
36 Å
S152   C-term
CalpT$_{10}$

cA$_4$

b

CalpL w/o cA$_4$
—1 —2 —3 —5 mg/ml
CalpL +cA$_4$
—1 —2 —3 —5 mg/ml
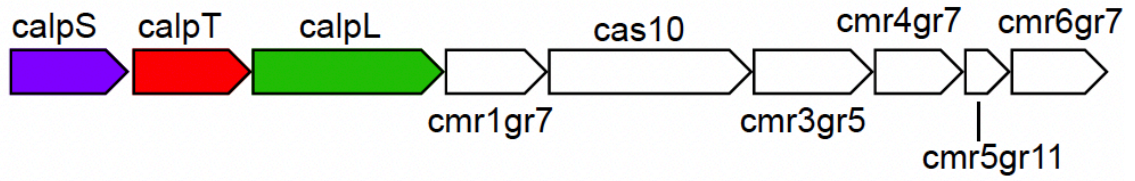
c

15% SDS-PAGE, coomassie

d

15% SDS-PAGE, coomassie

a *Sulfurihydrogenibium sp.* YO3AOP1

calpS calpT calpL cmr1gr7 cas10 cmr3gr5 cmr4gr7 cmr5gr11 cmr6gr7

b

type-III complex

DNA transcription RNA Cas10 ATP

CalpS CalpT CalpL cA₄ synthesis cA₄

in-trans cleavage.
conformational change?
cleavage between oligomers?

RNAP

CalpT₁₀ CalpT₂₃/S adaption to phage attack

degradation of CalpT₂₃

228