The exploration & exploitation of CRISPR-Cas type III defence strategies

A. Jurre Steens

Propositions

1. CRISPR-Cas type III immune systems are extremely creative in finding ways to kill its host.
(this thesis)

2. Studying prokaryotic immune systems that have conceptual similarities to eukaryotic immune systems allows us to learn about both of them.
(this thesis)

3. Evolution is like baking bread, with basic ingredients you can achieve a huge variety of outcomes.

4. The skills required for becoming an academic largely overlap with those required for becoming an entrepreneur.

5. The current rise of online media makes televised weather forecasts obsolete.

6. Discovering life beyond our planet will profoundly transform the course of human history.

Propositions belonging to the thesis, entitled

The exploration & exploitation of CRISPR-Cas type III defence strategies

A. Jurre Steens
Wageningen, 17th May 2024

The exploration & exploitation of CRISPR-Cas type III defence strategies

A. Jurre Steens

Thesis committee

Promotor
Prof. Dr John van der Oost
Personal chair at the Laboratory of Microbiology
Wageningen University & Research

Co-promotor
Dr Raymond H. J. Staals
Associate professor at the Laboratory of Microbiology
Wageningen University & Research

Other members
Prof. Dr Gorben Pijlman, Wageningen University & Research
Prof. Dr Martin Jinek, University of Zurich, Switzerland
Prof. Dr Malcolm F. White, University of St Andrews, Scotland
Dr Wen Y. Wu, University of Groningen

This research was conducted under the auspices of VLAG Graduate School (Biobased, Biomolecular, Chemical, Food and Nutrition Sciences)

The exploration & exploitation of CRISPR-Cas type III defence strategies

A. Jurre Steens

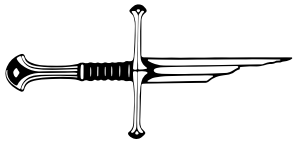
Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of Rector Magnificus,
Prof. Dr C. Kroeze
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 17 May 2024
at 4 p.m. in the Omnia Auditorium

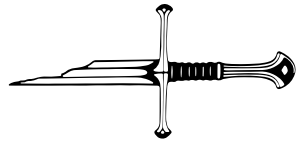
A. Jurre Steens
The exploration & exploitation of CRISPR-Cas type III defence strategies
284 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2024)
With references, with summary in English

ISBN: 978-94-6469-915-9
DOI: https://doi.org/10.18174/649997

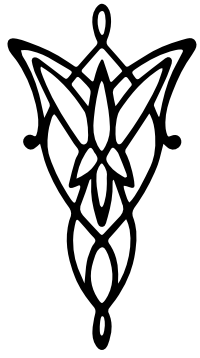
Table of contents





Chapter 1

General introduction and thesis outline



General introduction

Evolution

"Survival of the fittest" is a frequently used quote from the famous work by Charles Darwin. Unfortunately, however, it is often misinterpreted by the general public [1]. Immediately, thoughts of predators that hunt their prey, or animals fighting within their species to compete for mating partners appear. In both situations, the (physically) strongest individual (belonging to the same species or beyond) survives/wins and is deemed the fittest. While not completely wrong, this idea of the "fittest" does not convey the intended meaning. Many species that were fitter than the rest, at that time, have gone extinct due to changes in their environment and ecosystems, on large time-scales it is therefore more of a relative term. A better interpretation of the quote would be that species that are able to adjust the best to their environment over time are the most successful. Or, as Darwin carefully put it:

"Can it be thought improbable, seeing that variations useful to man have undoubtedly occurred (in our domestic productions), that other variations useful in some way to each being in the great and complex battle of life, should occur in the course of many successive generations? If such do occur, can we doubt (remembering that many more individuals are born than can possibly survive) that individuals having any advantage, however slight, over others, would have the best chance of surviving and procreating their kind? On the other hand, we may feel sure that any variation in the least degree injurious would be rigidly destroyed. This preservation of favourable individual differences and variations, and the destruction of those that are injurious, I have called Natural Selection, or the Survival of the Fittest."

Evolution by natural selection is based on variation. Variation in the (epi)genetics of organisms can lead to different (physical) features that can benefit the individual in the struggle for existence and survival. If this variation tends to increase the chances of survival, it is passed on during reproduction and selected for. Nature never comes up with the perfect solution in one go, it is very much a trial and error process that has been going on for billions of years on earth. This evolutionary process is better captured in a quote by Dr. Ian Malcolm:

Life finds a way

Bacteria and bacteriophages

Similar to humans, microbes also suffer from deadly infections by viruses. Bacterial viruses are often called bacteriophages, small packages of genetic information encapsuled by protein structures, and they exist on the border of being dead and alive. Generally speaking, bacteriophages can infect bacteria by injecting their genetic information (**Figure 1, A**) and take over the hosts' molecular machinery to use its genetic information to produce a next generation of bacteriophages (**Figure 1, B**) [2]. This eventually kills the bacteria by bursting out the infected cell in great numbers (**Figure 1, C**). Actually, the amount of bacteriophages on earth outnumber all organisms together, and they are responsive for a significant portion of bacterial death [3]. For billions of years bacteriophages and bacteria have evolved to "outsmart" each other, a never ending arms race: life finds a way through survival of the fittest. This has led to extremely diverse bacteriophages but also to many mechanisms by which bacteria attempt to defend themselves against them [4–7].

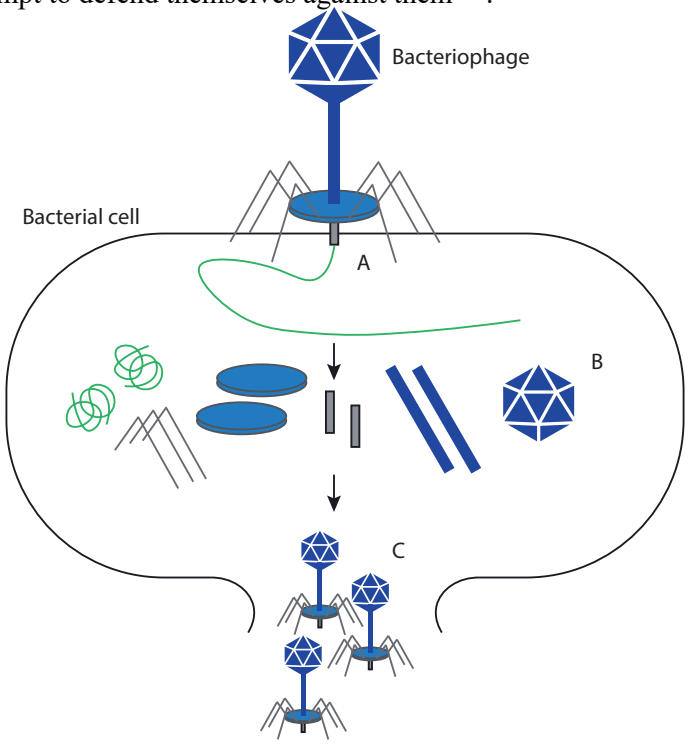


Figure 1. Schematic representation of a lytic bacteriophage infection cycle. The bacteriophage (specifically) attaches to the bacterial cell and injects its genetic information (green) into the cytoplasm (**A**). The hosts' molecular machinery is then hijacked to replicate the genetic information and various other (protein) components (**B**). After assembly of a multitude of bacteriophages they burst out of the cell, ready to infect neighbouring cells (**C**).

Phage defence systems

During the lytic bacteriophage infection cycle, the hosts' molecular machinery is hijacked to make viral offspring, after which the cell is lysed to release the new generation of phage particles. Sensing whether the integrity of the bacterial cell is affected is an indirect way to signal that a phage infection has occurred. An example of this are toxin/anti-toxin systems, which often contain stable toxin mRNA/protein components and unstable anti-toxin mRNA/protein components [8–10]. Reduced integrity of the hosts molecular machinery leads to the relative rapid depletion of the anti-toxin components, subsequently relieving the repression of the toxin and killing the cell to stop the viral replication. This process is called abortive infection.

A more direct way of sensing the presence of a bacteriophage infection can occur before or during phage assembly, where phage proteins trigger the defence system directly [11–13]. Once triggered, effector proteins get activated that can induce membrane disruption, metabolite depletion, DNA cleavage and a variety of other ways that disrupt essential cellular pathways .

An early immune response to bacteriophage infection, however, can be achieved by the sensing of invading nucleic acids. The injection and synthesis of bacteriophage nucleic acids are the first intra-cellular steps of infection and the targeting of these steps by bacterial defence systems is a widespread mechanism. Most frequent are restriction-modification (RM) systems, which modify the host genome in a specific manner [14,15]. Methylation of the nucleobases is a common modification in some of these systems, but alterations exist, such as DNA sugar-phosphate backbone modifications [14,16–20]. The invading nucleic acids are recognized by the lack of these modifications and are thereby identified as 'non-self' or 'foreign', leaving them vulnerable to restriction and thereby neutralisation.

Sequence-specific recognition by bacteriophage defence systems is the second most widespread strategy that targets the invading nucleic acids directly. The most used strategy consist of native, non-specific immune systems. The clustered regularly interspaced short palindromic repeats-CRISPR associated proteins (CRISPR-Cas) systems is the only adaptive prokaryotic immune system, present in ~40% of bacterial genomes [21].

CRISPR-Cas

The discovery of clusters of short ~30 basepair (bp) repeated palindromic sequences that were interspaced with similar lengths of variable DNA sequences puzzled researchers at the time [22,23]. Only when sequenced genomes became more abundant, this phenomenon was observed more frequently and given the name by which we known it now; CRISPR [24]. When several groups realized that the variable DNA regions of the CRISPR resembled bacteriophage DNA, it was hypothesized that these were part of an adaptive bacteriophage immune system [25–27]. The genes adjacent to the CRISPRs encoded for Cas proteins that were later proven to be essential for this defence system and furthered the mechanistic understanding of this novel bacteriophage defence system [28,29]. The biotechnological revolution that ensued has led to numerous landmark discoveries and developments, climaxing in the 2020 Nobel prize in Chemistry awarded to Emmanuelle Charpentier and Jennifer Doudna [30].

Classification of CRISPR-Cas systems

The number of CRISPR-Cas systems that have been discovered has grown immensely over the years and a systematic approach to classification has been established (**Figure 2**) [21]. The highest order in the classification has CRISPR-Cas systems divided over two classes, Class 1 and Class 2. This division is based on the composition of the effector module of the CRISPR-Cas systems, whereby the combination of the guide nucleic acid and complexed protein(s) are considered. Class 1 systems comprise of a multi-subunit effector complex whereas the Class 2 systems consist of a single multidomain effector protein complex. Within each of the classes, several types have been established. For Class 1 these are type I, III and IV while for Class 2 types II, V, VI are currently included. Various subtypes divide these further, based on CRISPR-Cas locus organisation, Cas gene composition and ancillary effector module presence.

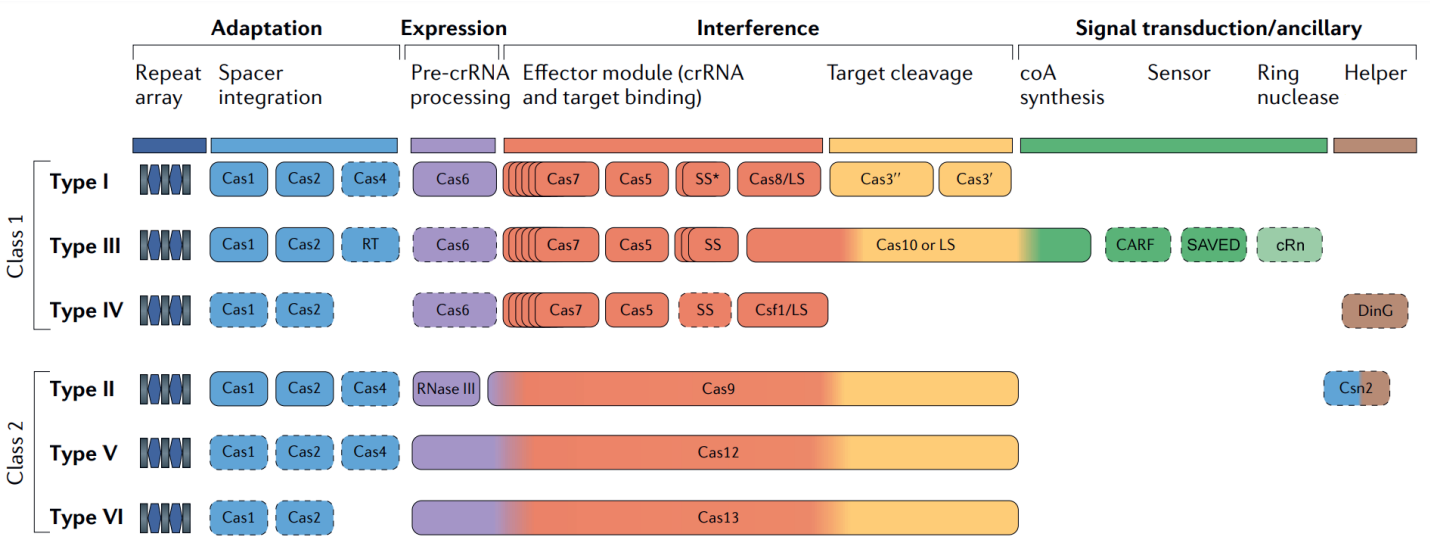


Figure 2. Schematic overview of CRISPR-Cas classification. The involved Cas genes are schematically indicated for each type, grouped by the stages of CRISPR-Cas immunity they are responsible for. Figure adapted from [21].

Stages of CRISPR-Cas immunity

CRISPR-Cas adaptive immunity consists of three general stages: Adaptation, Expression, and Interference (**Figure 3**) [31]. Upon encountering an invading mobile genetic element (MGE), a fragment of this (pre-spacer) is incorporated in the CRISPR-array between duplicated copies of the first repeat. This mechanisms of adaptation is generally mediated by a Cas1:Cas2 protein complex, but other Cas proteins have also been shown to aid in adaptation [32–34]. For example, Cas4-mediated pre-spacer length and protospacer adjacent motif (PAM) selection can also be done in this process [35–38]. The adaptation is called naïve if the MGE has not been encountered before, whereas primed adaptation occurs when a (partially matching) MGE is encountered by the system that is has already seen and leads to increased adaptation rates [39–44].

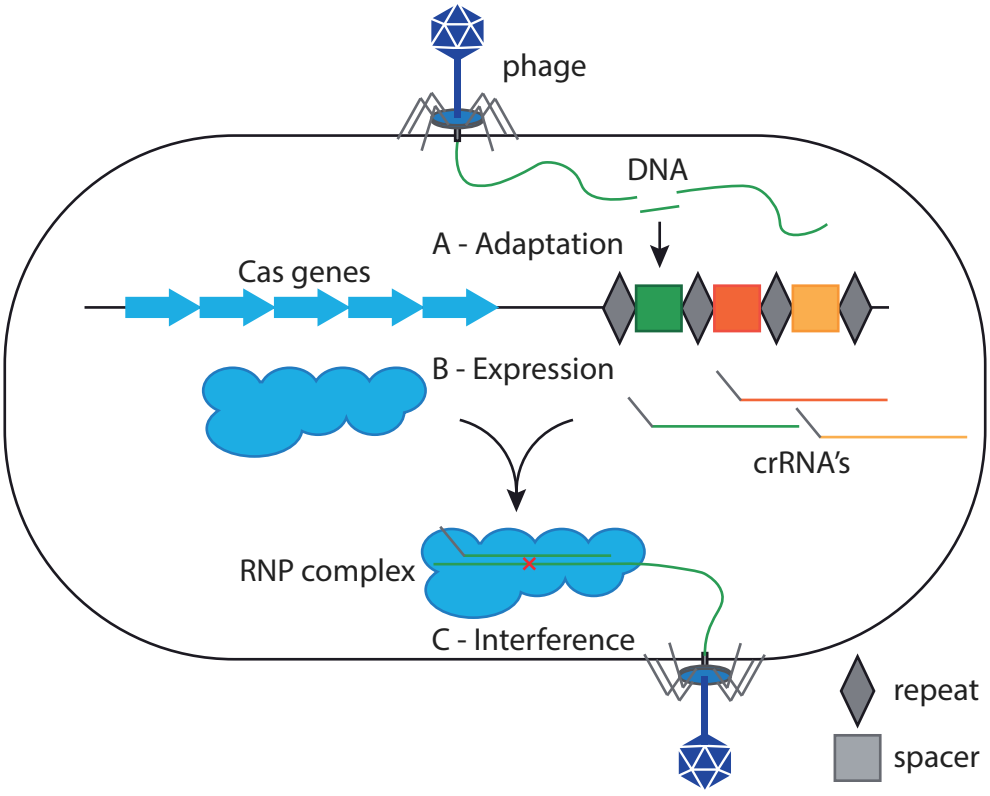


Figure 3. Schematic overview of the stages of CRISPR-Cas immunity. Adaptation occurs when a new MGE is encountered, a fragment of the MGE is incorporated into the CRISPR array (**A**). Upon expression of the associated Cas-encoding genes and CRISPR-assay and subsequent pre-crRNA maturation, they can assemble into an active RNP (**B**). When a previously encountered MGA invades again, the genomic material is specifically targeted and neutralized through (ribo)nuclease activity (**C**).

Expression of the CRISPR-array leads to the transcription of pre-CRISPR RNA (pre-crRNA), consisting of the repeats interspaced by spacers. How pre-crRNA is processed into mature crRNA depends on the type of system. For Class-1 (type I, III and IV) systems, the pre-crRNA is processed by the Cas6 ribonuclease near each 3' end of the repeat region at the base of the stem-loop structure [45–49]. Typically, the remaining 8 nt region of the repeat forms the 5' handle of the mature crRNA in the ribonucleoprotein (RNP) complex. In Class-2, all type II and some type V systems need a host ribonuclease that cleaves the duplex formed by the pre-crRNA and trans-activating RNA (tracrRNA) when bound to Cas9 [50–52]. For other type V and all type VI systems, the pre-crRNA processing is catalysed by a dedicated ribonuclease domain of the effector Cas protein itself [53–55]. Binding of the mature crRNA (or crRNA/tracrRNA), to the Cas protein(s) results in formation of the RNP complex.

In the interference stage, surveillance for the correct protospacer adjacent motif (PAM) is the first step for the DNA targeting systems: type I, II, IV and V [56–59]. The PAM-selection is a mechanism by which self and non-self targets can be distinguished by the system, mitigating the risk for accidental self-targeting activity. Upon finding the PAM, the location and sequence of which can differ substantially between CRISPR-(sub)types, base-pairing occurs between the seed region of the crRNA with the target strand [60–62]. In this process, the target strand is unwound and an R-loop structure is formed to facilitate subsequent cleavage [63]. For type I systems, a Cas3 nuclease-helicase is recruited that nicks the non-target strand and via its helicase activity further degrades the non-target strand in the 3' to 5' direction [29,64,65]. Type II and V systems perform the target DNA cleavage by themselves, which leads to a blunt-ended (Cas9) or a staggered (Cas12) double strand break, respectively [55,66–69]. The type III and VI systems are RNA targeting systems that do not have a PAM requirement but instead use 5' handle complementarity and a protospacer flanking sequence (PFS) for self/non-self distinction, respectively [53,70]. Upon base-pairing of the target RNA to the crRNA, the target RNA is cleaved in (multiple) fragments. For both DNA and RNA targeting systems the degradation of the target nucleic acids neutralizes the threat and interferes with the invading nucleic acid. However, for type III, V and VI systems, base-pairing of the target to the crRNA can also lead to non -specific collateral cleavage activity of ssRNA, ssDNA and dsDNA [53,70–74]. The biological relevance of the latter phenomenon is contributing to defence, however, not at the individual cell level but rather at the community level, by triggering dormancy or cell death, thereby preventing phage progeny [7]. The collateral activity of type III systems generally proceeds via generated secondary messenger molecules that in turn allosterically activate ancillary effector proteins that can lead to an abortive infection phenotype [75].

Type III CRISPR-Cas

The RNA targeting type III systems are thought to target MGE-derived transcripts in the context of a transcription bubble [76,77]. The seed region at the 3' region of the Cas-complex bound crRNA guide is exposed, allowing for initiation of target RNA binding before further base-pairing propagation can occur [78]. Type III effector complexes with variable backbone lengths (typically in the range of 4-6 Cas7 subunits) result in variable 3' end crRNA processing, which leads to a variable seed sequence location in crRNA guides derived from a single CRISPR spacer [78]. When compared to a fixed seed region, a variable seed implies a major hurdle for a bacteriophage to escape a certain crRNA by mutations. After base-pairing of a target RNA to the seed region, further propagation towards the Cas10-activating region (CAR) can occur. At this point, the Cas7 subunits of the type III complex generally cleave the target RNA at 6 nt intervals [79–82]. Furthermore, in the absence of 5' handle complementarity of the crRNA with the target RNA (non-self), the Cas10 subunit of the type III complex is activated. This Cas10 subunit typically contains a HD domain, activation of which results in cleavage the exposed ssDNA regions in the transcription bubble in a sequence non-specific manner. On top of this, the activation of the Palm domain leads to cyclase activity whereby ATP is used to generate cyclic oligoadenylate (cOA) messenger molecules, which can consist in different sizes ranging from two to six adenosine moieties [73,83]. Subsequently, cOAs bind and allosterically activate proteins containing CARF [CRISPR-Associated Rossmann Fold] or SAVED [SMODS-(Second Messenger Oligonucleotide or Dinucleotide Synthetase) Associated and fused to Various Effector Domains] sensory domains. These sensory domains exist as fusions to a wide range of catalytic domains [75,84]. Recently, a novel signalling molecule (SAM-AMP) was found to be produced by a Cas10 of type III complex, including a compatible allosterically activatable ancillary effector protein [85]. A commonality is that the downstream effector modules are prolific in their catalytic activity, so much so that the cellular processes are disrupted to the point of bringing the cells to a state of dormancy or even cell death [86–93]. Ultimately, this type of abortive infection mechanism sacrifices the individual cell to protect the rest of the population.

Thesis outline

The overall theme of this thesis is the exploration of the variety and mechanisms of immune responses by CRISPR-Cas type III systems, both on a fundamental level and guided by applications in the field of diagnostics.

Chapter 1 – General introduction

A general introduction about CRISPR-Cas is given that briefly touches upon that these immune systems were evolved in the never-ending arms race with bacteriophages. Furthermore, the discovery of CRISPR-Cas and the incremental advances are discussed as well as the nomenclature that is required to talk about the vast variety of systems we know of at this time. Lastly, the molecular mechanisms by which CRISPR-Cas typically functions are discussed, with a particular focus on type III.

Chapter 2 – Coevolution between bacterial CRISPR-Cas systems and their bacteriophages

The second chapter reviews the very diverse CRISPR-Cas types of heritable, adaptive immune systems in bacteria and archaea, that protect them against viruses and other parasitic genetic elements. When and why CRISPR-Cas immunity against phages evolved and how this, in turn, affects immune evasion strategies of these phages will be explored. Lastly, this interplay and subsequent co-evolution is speculated on, in terms of how the mechanisms of CRISPR-Cas immunity may affect this.

Chapter 3 – The diverse arsenal of type III CRISPR-Cas-associated CARF and SAVED effectors

The third chapter reviews and discusses the mechanisms by which cOA signals are being generated by type III CRISPR-Cas to allosterically activate downstream CARF- and SAVED-based ancillary effector proteins. An extensive overview of the experimentally characterized CARF- and SAVED-domain containing effector proteins is presented, including phenotypical outcomes if available. Lastly, examples of bioinformatically predicted CARF and SAVED proteins are discussed and the molecular mechanisms by which they may function are speculated on.

Chapter 4 – CARFish: identifying cOA-responsive effector proteins of type III CRISPR-Cas systems

The fourth chapter describes the need for a top-down approach to more extensively map the cOA signalling pathway in CRISPR-Cas type III system containing hosts and the subsequent development of such a method. The developed technique is based on immobilized cOA signalling molecules used as affinity chromatography, and is tested on bacterial and archaeal hosts that have well-characterized cOA responsive effector proteins, for validation. Furthermore, the method is tested on an archaeal host with a predicted, but not experimentally tested, cOA responsive effector protein. Lastly, candidate proteins that were identified with the method are analysed in silico, and their possible biological relevance is discussed.

Chapter 5 - Nanopore detection of single type III CRISPR-Cas produced oligo-adenylates and their stoichiometries

The fifth chapter focusses on the fact that cOAs vary in composition (3-6 adenylate monophosphate rings) and that characterization of these requires complex equipment and techniques. The chapter describes the development of a new protein nanopore assay that aims to achieve label-free detection of single cOA molecules. Together with neural-network assisted analysis of the data, the ability to identify the stoichiometry of cOA mixtures is assessed. Lastly, the composition of in vitro generated cOA mixtures from CRISPR type III-A and -B variants of Thermus thermophilus are analysed.

Chapter 6 - Compact but mighty: biology and applications of type III-E CRISPR-Cas systems

The sixth chapter reviews a multitude of papers that report on the CRISPR-Cas type III-E system, seemingly a hybrid between Class 1 and Class 2 CRISPR-Cas systems whereby some of the type III subunits are fused into a large polypeptide subunit. The current structural and biochemical understanding of this system is discussed and summarized. Furthermore, applications that have been developed that have come from fundamental understanding of this system are discussed as well.

Chapter 7 - Type III-B CRISPR-Cas signalling-based cascade of proteolytic cleavages

The seventh chapter investigates the CRISPR-Cas type III system from *Haliangium ochraceum*, with a focus on the set of associated ancillary effector proteins. Bioinformatical analysis of the origin of this set of associated genes is performed to determine whether these were co-opted from other CBASS immune systems. Furthermore, *in vitro* biochemical analysis has been done on the associated proteins to determine their function and speculate on their role in the immune defence. Lastly, the complete CRISPR-Cas type III system is transplanted and *in vivo* studies uncover the phenotypical outcomes when activated. Together, these insights reveal analogy with some eukaryotic systems which are used to create a model showing our understanding of this system.

Chapter 8 - SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation

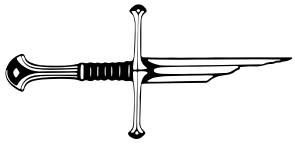
The eighth chapter reports on the investigation of targeting requirements that lead CRISPR-Cas type III-B to initiate the production of cOA signalling molecules. A multitude of *in vitro* biochemical analysis techniques will be used to gain insight in the molecular mechanism. Moreover, these findings will be reflected on with the use of previously obtained structural data. Lastly, a proof of concept is developed for a nucleic acid detection tool and demonstrated on SARS-CoV-2 clinical samples, SCOPE.

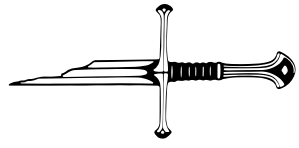
Chapter 9 – RNAPhi: a viral thermostable RNA polymerase for improved diagnostics

The ninth chapter describes the characterization of a thermostable RNA polymerase from phage origin. Structural in silico analysis combined with biochemical in vitro work will uncover how this polymerase can be used for *in vitro* transcription at elevated temperatures. Furthermore, these findings will be used in the development of SCOPEv2, a 1-step nucleic acid detection technology.

Chapter 10 – General discussion

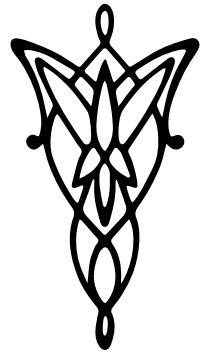
The tenth chapter contains a general summary of this thesis which is followed by a discussion of the results obtained. It also provides an outlook on future efforts that will expand and improve our understanding of the complex mechanisms by which CRISPR-Cas type III systems works.





Chapter 2

Coevolution between bacterial CRISPR-Cas systems and their bacteriophages



Adapted from:
Watson, Bridget N.J., Jurre A. Steens, Raymond H.J. Staals, Edze R. Westra, and Stineke van Houte. Coevolution between bacterial CRISPR-Cas systems and their bacteriophages. *Cell host & microbe* 29, no. 5 (2021): 715-725.

Summary

CRISPR-Cas systems provide bacteria and archaea with adaptive, heritable immunity against their viruses (bacteriophages and phages) and other parasitic genetic elements. CRISPR-Cas systems are highly diverse, and we are only beginning to understand their relative importance in phage defence. In this review, we will discuss when and why CRISPR-Cas immunity against phages evolves, and how this, in turn, selects for the evolution of immune evasion by phages. Finally, we will discuss our current understanding of if, and when, we observe coevolution between CRISPR-Cas systems and phages, and how this may be influenced by the mechanism of CRISPR-Cas immunity.

Introduction

Bacteriophages (phages) are genetically and morphologically diverse and outnumber their microbial hosts in most environments [3]. They shape microbial evolution, community structure, and their ecological functions, such as carbon and nutrient cycling 94. Bacteria use different strategies to limit infection by phages. Of these, CRISPR-Cas (clustered regularly interspaced short palindromic repeats; CRISPR-associated) immunity is the only adaptive and heritable defence system known to date [95,96]. Generally, the CRISPR-Cas immune response is thought of as a three-stage process (**Figure 1**). During adaptation, the acquisition machinery captures a fragment of the invading phage genome (known as a ‘‘protospacer’’) and incorporates it into a CRISPR array in the bacterial chromosome as a ‘‘spacer’’ sequence between repeat sequences. In the expression phase, the array is transcribed and processed into crRNAs that contain a single spacer sequence. CrRNAs associate with an effector protein or complex, which, during the interference stage, surveys the cell for genetic material complementary to the spacer. If detected, the invading nucleic acid will be cleaved or degraded. CRISPR-Cas systems have been identified in 40% of bacterial and 90% of archaeal genomes and can be grouped into two Classes containing six types 21 (**Figure 2**). In general, Class 1 is more abundant than Class 2, with types I, II, and III being the most abundant, comprising 60%, 13%, and 25% of bacterial CRISPR systems, respectively [97,98]. In archaea, type I systems makeup 64%, and type III systems 34%, while type II systems are rare [99].

In agreement, most of our knowledge of the ecology and evolution of CRISPR-Cas systems exists for type I, II, and III systems, which will therefore be the focus of this review. The weak association between the phylogenies of bacterial and archaeal hosts and the phylogenies of the CRISPR-Cas systems they carry suggests that horizontal gene transfer has mediated their spread [97]. Further, the distribution and prevalence of CRISPR-Cas systems are not uniform across bacterial and archaeal taxa or environments, and some lineages that are found across a wide range of environments appear to lack CRISPR-Cas systems altogether [100]. While the reasons for variation in distribution and prevalence are largely unknown, they are likely related to the ecology of the host, as the balance of the costs of carrying or expressing the system and the benefits of adaptive immunity will ultimately determine whether CRISPR-Cas is lost or retained. To date, several ecological factors have been correlated with CRISPR-Cas prevalence, including oxygen requirement and temperature, as more systems are found in thermophilic bacteria and those associated with an anaerobic lifestyle [98]. CRISPR-Cas systems are also more commonly found in free-living than host-associated microbes [98,100]. In addition, the presence or absence of various dsDNA repair mechanisms in the host has been linked with CRISPR-Cas prevalence, as some systems have been shown to require the activity of repair mechanisms [101]or

impair the mechanism function [102]. Explanations for the roles of these environmental and genetic factors have been proposed, but the mechanisms through which they act are in many cases not yet determined.

Diversity of CRISPR-Cas systems

CRISPR-Cas systems are diverse and can be categorized into two Classes, six types, and over 30 subtypes based on differences between the Cas proteins encoded (**Figure 2**) [21]. The adaptation machinery, Cas1 and Cas2, is conserved across almost all CRISPR-Cas variants. Class I systems (consisting of types I, III, and IV) encode multi-protein effector complexes, whereas Class II systems (types II, V, and VI) consist of single protein effectors. Across the different CRISPR-Cas types, activation of the effector components occurs when different types of nucleic acids are detected and results in different outcomes for the cells and invading genetic materials (**Figure 2**). In type I systems, the Cas protein effector complex (known as Cascade) recognizes double-strand (ds) DNA and recruits Cas3 to progressively degrade the foreign DNA [29,65]. Interference relies on recognizing a short sequence, called the protospacer adjacent motif (PAM) and a perfect spacer-protospacer match in the 8 nt adjacent to the PAM, called the ''seed'' sequence [56,103]. Clearing the phage DNA will typically result in host survival [65]. However, if phage infection progresses before the infection can be cleared, cell death may occur, which is likely due to the irreversible damage inflicted on bacterial processes in the cell for phage replication, resulting in abortive infection [104]. In both instances, phages will be removed from the population to protect vulnerable cells. Type II systems also require the recognition of a (different) PAM and seed sequence complementarity in the dsDNA target sequence [52]. The effector protein, Cas9, will generate a double-strand break in the genome of the invading genetic material and bacteria survive infection [105]. In type III systems, target RNAs resulting from transcription of the foreign DNA will bind to the effector complex and subsequently activate the Cas10 effector, initiating different pathways of defence, including target RNA and DNA cleavage [77,106–110]. Cas10 also generates cyclic oligoadenylates (cOAs), which activate CARF effector proteins to induce non-specific degradation of RNA in the cell [73,83]. While immunity mediated by type III systems can result in host survival [111], the collateral RNA damage can induce dormancy and possibly cell death. Further, no canonical PAM is recognized by type III systems, although some systems recognize motifs flanking the target RNA (called rPAMs) [112]. Type IV systems are typically found on plasmids or plasmid-like elements, where they are thought to be involved in plasmid competition [113]. As they typically lack an adaptation module, it is thought they may associate with compatible host-encoded CRISPR-Cas systems [113]. Like type II systems, type V cuts both strands of the target DNA, albeit asymmetrically [55]. Additionally, activation of the Cas12 protein can result in nontarget ssDNA degradation [72], although more work is needed to under-

stand the consequences of this collateral damage on host outcome [114]. Finally, type VI systems uniquely recognize and cleave RNA. The effector, Cas13, degrades the target RNA, as well as nontargeted RNA in the cell. Due to the high levels of RNA degradation in the cell, type VI-mediated immunity induces dormancy, but cells act as 'sinks' to remove phages from the population [115].

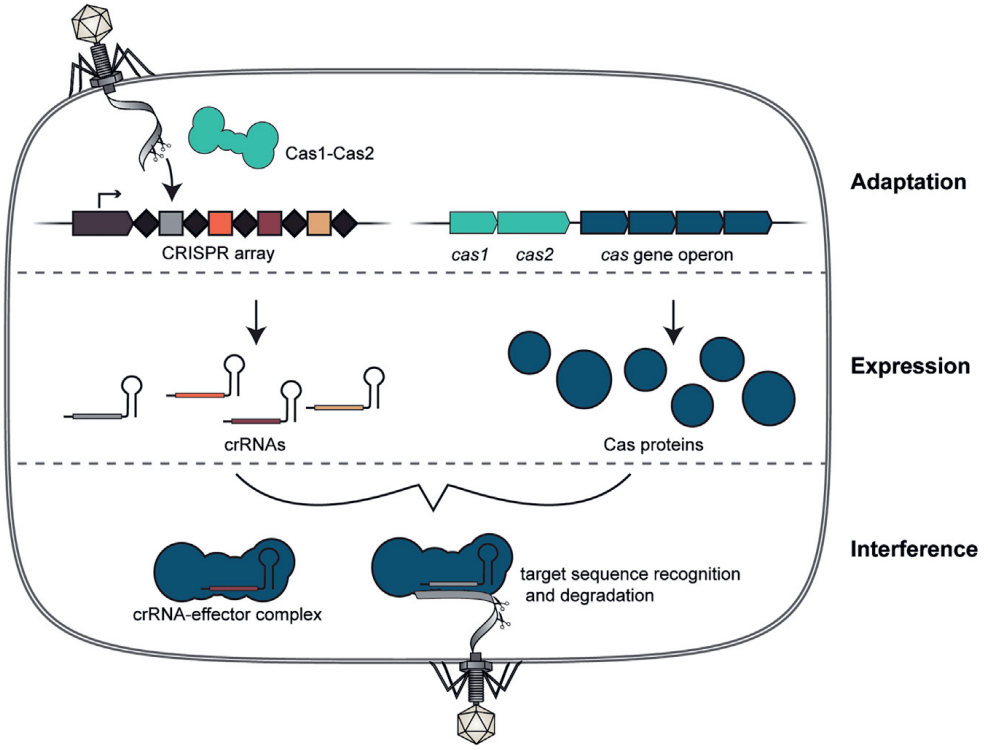


Figure 1. Overview of CRISPR-Cas immunity Adaptation: Cas1 and Cas2 capture a fragment of invading genetic material and incorporate it into the CRISPR array as a spacer (squares), between repeat sequences (black diamonds). Expression: the CRISPR array is transcribed and processed into smaller CRISPR RNAs (crRNAs) and the Cas effector protein(s) are produced. Interference: the crRNAs associate with the effector protein(s) and any sequence detected that is complementary to the crRNA is degraded.

When does CRISPR immunity evolve?

While CRISPR-based immunity can protect against phages and other parasitic mobile genetic elements, bacteria do not always rely on CRISPR-Cas immune systems for phage defence. First, CRISPR-Cas systems are not present in all prokaryotic genomes, as discussed above [21,100]. Second, even if bacteria do carry a CRISPR-Cas immune system, they may rely on other defences instead. Indeed, cells often have multiple defence options, such as surface mutation or modification (sm) and restriction-modification (R-M) systems [95]. Like CRISPR-Cas, R-M systems function following phage DNA injection and cleave unmodified DNA at certain sequence motifs. On the other hand, sm prevents phage binding and entry by altering or masking the receptor, so phages remain in the population. Recent studies have demonstrated that the defence strategies that bacteria evolve in response to phage strongly depend on the environment. When we consider an initially sensitive bacterial population that is infected with phages, the most commonly observed mechanisms to acquire resistance in the short term are mutation or loss of the receptor that is used by the phage to attach to the cell surface and CRISPR-Cas. Initially, the rates at which both types of resistance evolve (i.e., rates of spacer acquisition and rates of receptor mutation) will be one of the key determinants of the type of defence that dominates in the bacterial population, especially if different competing genotypes have similar relative fitness. In the longer term, which defence ultimately prevails will depend on whether hosts with CRISPR immunity have higher fitness than those with mutated receptors or vice versa, which in turn will depend on the environment [116]. In addition, other defence mechanisms may be acquired horizontally over these longer timescales that compete or combine with CRISPR immunity or sm to provide more robust or lower-cost defence [117]. It is becoming increasingly clear that both in the short- and long-term, the ecological context is a major determinant for the evolution of bacterial defences [118].

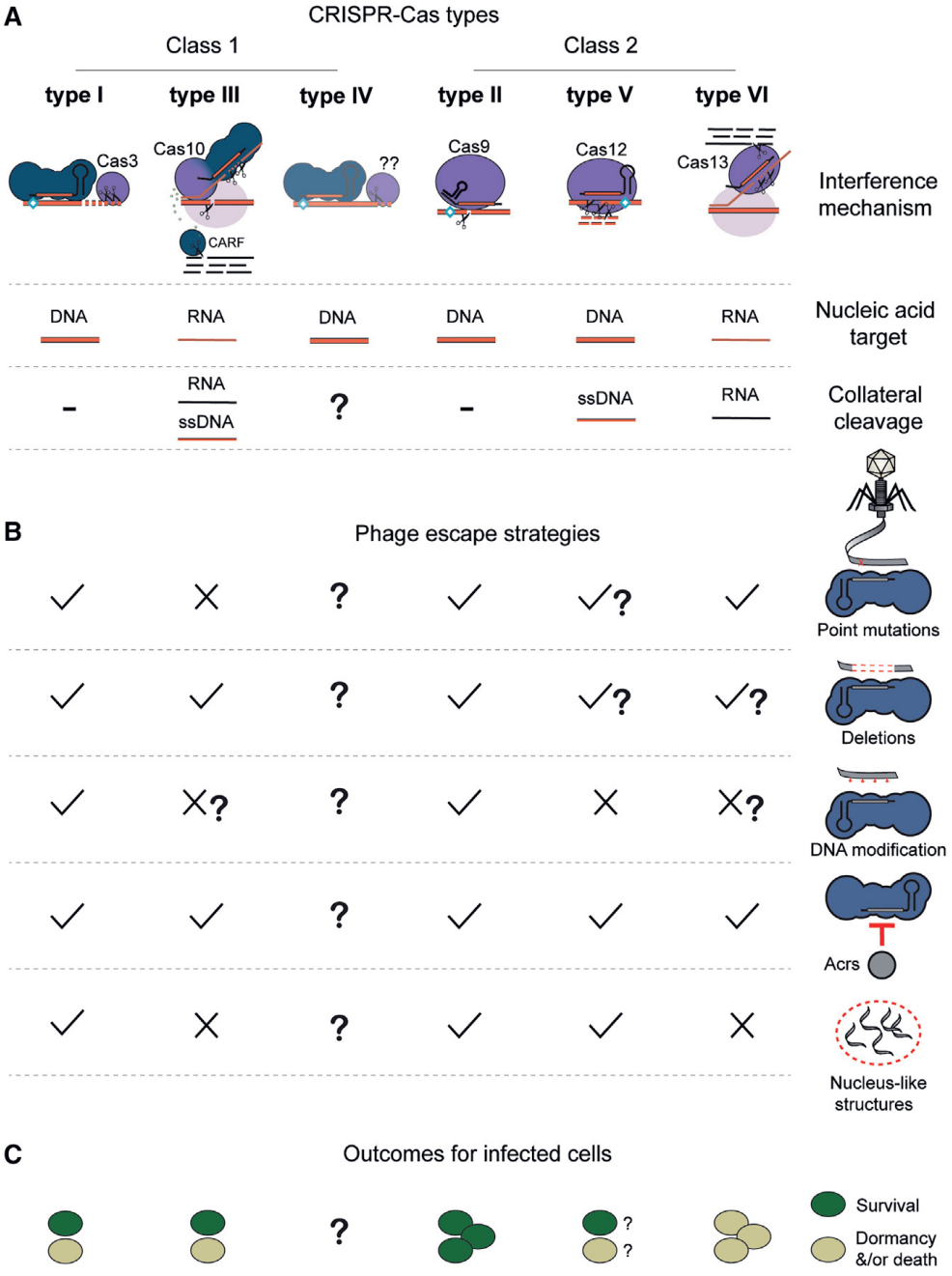


Figure 2. Summary of CRISPR-Cas types, phage escape mechanisms, and the outcomes for infected cells (**A**) The components involved in interference, the nucleic acids that are targeted and affected by collateral cleavage are shown for the different CRISPR-Cas types. The

labelled, purple proteins are the signature proteins and effector components for each type. The light-blue diamonds represent the requirement for a PAM sequence. Gray circles (types III and VI) represent replication bubbles.
(**B**) Phages can escape targeting through different strategies but these vary in their efficacy against different CRISPR-Cas types. Strategies involving alterations to the phage DNA to prevent crRNA recognition include point mutations in the PAM or seed sequence of the protospacer, deletions of the protospacer, and protospacer DNA modifications. Anti-CRISPR (Acr) proteins can inactivate CRISPR-Cas immune proteins. A nucleus-like structure in the bacterial cell that occludes CRISPR-Cas immune proteins is produced by some phages, including jumbo phages.U= phage escape, X = CRISPR-Cas is still effective (? Indicates the predicted outcome where no evidence is available).
(**C**) Type I and II systems promote survival of infected cells by cleaving invading DNA (green circle) but in some type I systems, slow phage clearance can result in cell death (light-grey circle). Type III systems can result in survival of infected cells but activation of non-specific RNA cleavage through cOA signalling induces cell dormancy. Little is known about the outcome of type V immunity, but ssDNA degradation may induce dormancy or cell death. The collateral damage induced by type VI systems through RNA degradation induces dormancy. It is not yet clear whether type IV systems provide phage resistance.

The rates of spacer acquisition

Following phage infection, the initial rise in the frequency of CRISPR-immune bacteria in a phage-sensitive population will strongly depend on both the rates of spacer acquisition and the rates at which surface-based resistance mutations are generated [119]. If the host has a higher mutation rate, then more mutations in the receptor are generated and this will result in a higher proportion of surface-based resistance in the population [120]. Three factors have been identified that determine the rates of spacer acquisition by CRISPR-Cas immune systems: defective phages, cas gene expression, and priming. The first is related to the challenge of any CRISPR-Cas system to capture a protospacer from the phage genome and carry out interference before phage replication and lysis of the bacterial cell. It turns out that high rates of spacer acquisition occur from infection with defective phages that naturally exist in phage populations [121]. This form of vaccination, whereby the CRISPR-Cas can generate immunity against the phage from an attenuated form, is analogous to the formation of defective interfering particles (DIPs) in eukaryotic viruses and the immune response they trigger [122]. DIPs are formed by error-prone viral replication and while they lack viral genetic material, they can still infect a host cell. Consequently, these attenuated viral particles can be used as a viral vaccine. Spacer adaptation is also enhanced in cells that also carry R-M systems, as these limit phage replication through phage DNA degradation, and these cleavage products can be captured by the Cas adaptation machinery 121. Generally, immunity is more readily generated when the CRISPR-Cas system is exposed to free DNA ends, through processes including

RecBCD-mediated DNA repair, degradation, or processing of DNA breaks occurring at stalled replication forks [123] or when linear phage DNA first enters the cell [124]. The rates at which spacers are being acquired, whether it be from defective or intact phages, are also determined by the expression levels of the CRISPR-Cas immune system [125,126]. Expression of CRISPR-Cas is often tightly regulated, presumably because their expression may carry costs in the absence of phage infection, for example, due to the acquisition of spacers from the host genome (reviewed in [127]). Two key factors that determine the infection risk of an individual bacterium are the density of bacterial hosts in the population, since phages spread more effectively when their density is high, and the expression levels of phage receptors on the cell surface. Consistent with this, the type I-E, I-F, and III-A CRISPR-Cas systems of *Serratia sp.* ATCC 39006 are all regulated by quorum sensing (QS) [128], and a single regulator, the Rcs stress response, regulates both CRISPR-Cas immunity and the expression of cell surface proteins [129]. In *Pseudomonas aeruginosa*, expression of both the type IV pilus, which is a key phage receptor for this species, and the CRISPR-Cas immune system are positively regulated by QS [130–132]. The alginate biosynthesis pathway represses CRISPR-Cas expression in *P. aeruginosa* cells growing on surfaces, which also ensures that the immune system is expressed at levels necessary for the risk of infection. Further, CRISPR-Cas expression levels can be induced by phage infection [133–135], as well as membrane stress [136] and metabolic status [133,137]. Priming (or primed adaptation) is a third factor that determines the rate of spacer acquisition. This mechanism of type I systems increases the rates at which new spacers are acquired from invading elements. Perfectly matched spacers or those with mismatches in the PAM or PAM-adjacent ‘‘seed’’ sequence in the targeted region can activate the incorporation of new spacers into the CRISPR array [138]. Consequently, in short-term evolution experiments, the observed levels of spacer acquisition are higher when bacteria are primed [39,138]. Another potential benefit of primed adaptation is that it introduces a bias toward spacer acquisition from parasitic DNA with sequence similarity to pre-existing spacers relative to the acquisition of spacers from the bacterial genome, which would result in autoimmunity [127]. Primed adaptation has also been observed in some type II systems, and similarly to type I systems [138], it occurs more often with perfectly matched spacers than with those carrying mismatches, since Cas9 must cleave the phage DNA to generate substrates for adaptation [43]. As a result, immunity may not be as rapidly generated in type II systems and the system is less able to quickly acquire new spacers in response to escape phages.

Fitness costs and benefits of CRISPR immunity

In the long term, the type of resistance that prevails in a population is determined by the fitness costs and benefits of each strategy, as the one with the highest net benefits will dominate [118,139]. Several factors have been identified that influence the relative fitness of bacteria with mutated surface receptors and those with CRISPR immunity. The type II-A CRISPR-Cas system of Streptococcus thermophilus has been shown to be costly to maintain and express [140], but the evolution of alternative defence such as sm is almost never observed in *S. thermophilus*, explaining why bacteria almost exclusively rely on CRISPR immunity (discussed in [116]). For *P. aeruginosa*, CRISPR-Cas immunity is also associated with a fitness cost, which was found to be strictly infection-dependent (i.e., CRISPR immune bacteria that are infected by phages have a lower fitness relative to uninfected cells) [139]. Recent mechanistic studies have shown that these phage infection-induced fitness costs are at least partially due to the production of phage-encoded proteins prior to phage removal by CRISPR-Cas [141]. Due to this cost, which is paid every time a cell gets infected, CRISPR-Cas immunity is favoured over sm at low phage densities, when infections are relatively rare, whereas high phage densities select for sm, which has a constitutive cost independent of the number of infections [139,141]. Hence, in environments where there is a constant immigration of phages [139] or where high phage densities are maintained due to the immigration of naive hosts into the population [142], bacteria with surface-based resistance ultimately dominate the population. The magnitude of the fitness costs of CRISPR immunity and surface resistance depends on the ecological context beyond phage densities. For example, it was found that the cost of surface-based resistance relative to those of CRISPR immunity are amplified in the presence of a microbial community [143]. Moreover, these amplified costs manifested in the presence of some, but not other bacterial species. Understanding the mechanistic basis of how microbial community composition determines the fitness costs and benefits of alternative phage defence strategies will be key to understanding their ecological distribution. The fitness benefits of spacer acquisition by a CRISPR-Cas immune system rely on the ability to protect against infection by genetically similar phage in the future. Since CRISPR-Cas immune systems rely on sequence identity between the phage from which a spacer is acquired and subsequent infecting phages, high genetic diversity in the phage population is predicted to reduce the benefits of acquired spacers, and therefore lead to lower levels of CRISPR immunity [144,145]. Indeed, experimental tests show that increasing phage genetic diversity promotes the evolution of more generalist defence through mutation of the phage receptor instead of CRISPR immunity [146]. Those bacteria that do evolve CRISPR immunity often carry multiple spacers, which leads to higher protection as phages are less likely to carry mutations in multiple targets [146]. The presence of genetically identical phages in the population that encode genes that block the CRISPR-Cas immune system also favours the evo-

lution of surface-based resistance [147]. CRISPR-based immunity can also have fitness costs due to the incorporation of self-targeting spacers [148]. Spacers targeting bacterial genomes are widespread, although there are often mutations in either the protospacers or the repeat sequences in the CRISPR array [149], since self-targeting spacers are cytotoxic [150,151]. While mutations might make these spacers non-functional, the mismatches can activate primed adaptation, resulting in autoimmunity [138,151]. In some type I systems, almost half of the self-targeting spacers map to prophages in the genome but targeting may be prevented by phage-encoded anti-CRISPR proteins [152]. The consequences of prophage-targeting vary between CRISPR-Cas types, as in type I systems, prophage-targeting spacers can be cytotoxic and the CRISPR-Cas system may be lost to remove the cost and maintain the prophage [153]. In contrast, in type III systems, the requirement of target transcription for targeting allows prophages to be tolerated [154]. However, fitness costs due to cytotoxicity can still occur if spacers target prophage genes that are expressed [155]. The potential for CRISPR-Cas systems to target and exclude prophages can be a disadvantage to the host, as they can provide beneficial traits [116]. Similarly, CRISPR-Cas systems can limit the uptake of novel genetic material which may restrict their evolutionary potential [156–158]. Consequently, CRISPR-Cas systems may be lost or inactivated to enable the uptake of plasmids [159], which is the case for several bacterial pathogens [160]. In accordance, in both Enterococcus faecalis and *P. aeruginosa* a negative correlation was found between CRISPR-Cas systems and horizontally acquired elements [161,162]. However, a global analysis of all sequenced genomes did not detect an interaction between CRISPR-Cas and signatures of horizontal gene transfer, suggesting that this effect may vary across taxa or ecosystems [163].

What limits the durability of CRISPR immunity?

The evolution of CRISPR immunity in bacteria can lead to rapid phage extinction. One way for phages to persist is through immune evasion, which can be achieved through various different mechanisms (extensively reviewed in [95,164]), including through mutation of target sequences, production of anti-CRISPR proteins, and physical barriers that shield phage DNA from cleavage (**Figure 2**). However, in addition to phage evolution, bacteria can also evolve to lose their CRISPR immunity, which can play an important role in the coexistence of bacteria with CRISPR immune systems and their phages.

Evolution of phage infectivity on CRISPR immune bacteria

Phages can overcome CRISPR-Cas targeting by altering the protospacer sequence through mutations, deletions, or gene rearrangements. For type I and type II systems, single point mutations in the PAM or the seed sequence can prevent CRISPR-Cas interference [165]. The probability of phages acquiring mutations to overcome type I or II CRISPR-Cas immunity depends on the composition of the host population [166]. Escape phages are most likely to evolve in a population with an intermediate level of resistant hosts, since there needs to be enough hosts on which phages can replicate, but also enough selection for acquiring infectivity. However, having hosts with multiple or different phage-targeting spacers decreases the likelihood of escape phages emerging. For escape phages that do evolve, the accumulation of point mutations that enable CRISPR escape can have negative fitness consequences [167–169], which may select against those phages in the long-term. It is more difficult for phages to overcome multiple spacers within a single host and deletion or recombination may be a more effective way to do this [169–171], although large deletions can alter the phage structure and reduce phage infectivity [169]. In contrast to type I and II, type III CRISPR-Cas systems are much more tolerant of mismatches due to a flexible seed sequence 78 and phages can only escape through deletions [111]. Phages can also evade CRISPR by providing physical barriers that shield the phage DNA from CRISPR-Cas immune complexes [164]. Examples of these include modification of the phage DNA to avoid detection by CRISPR-Cas [172], or, in the case of jumbo phages, the production of a nucleus-like structure within the host cell that surrounds the phage DNA but excludes CRISPR-Cas machinery [173,174]. Such mechanisms may have a fitness cost, although this has not been studied to date. While these physical barriers protect phages from CRISPR-Cas systems that recognize DNA, they do not protect against type III immunity as phage RNA will still be targeted in the cytoplasm [173]. Finally, phages may encode anti-CRISPR proteins (Acrs) that inhibit CRISPR-Cas activity. Different Acrs have been found to act at each phase of immunity, but most impair interference by interacting with effector proteins or complexes (See [175] from this special issue). Loss of CRISPR immunity through bacterial mutation Another mechanism that can enable the coexistence of bacteria with CRISPR immune systems and the phages they target is through the loss of spacers or mutation of cas genes [176–179]. Such spontaneous loss of immunity can provide phages with sensitive hosts that they can use to replicate, hence avoiding extinction. Both the mutation of cas genes and the loss of spacers from CRISPR arrays has been observed experimentally [159,180]. The idea that CRISPR immunity can be lost at a high rate is further supported by experiments where *Staphylococcus epidermidis* was transformed with an antibiotic resistance plasmid, resulting in estimated rates of between one in a thousand and one in ten thousand loss events per individual per generation [159]. In an observational study, spacer loss events correlated with the resurgence of a phage population that was no

longer targeted [180], and an experimental study with *S. thermophilus* DGCC7710 and its lytic phage 2972 found that the stable coexistence of bacteria and phages, which occurred despite the high frequency of CRISPR immunity in their experiments, dynamics that were best captured in a mathematical model that assumed high rates of loss of CRISPR-Cas immunity [179]. Another mechanism through which pre-existing levels of CRISPR immunity can decrease is through the acquisition of new spacers that target other genetic elements, which are inserted at the leader-proximal end of the CRISPR array. Spacers that are closer to the leader-end of the CRISPR array provide higher levels of immunity than those toward the trailer-end of the array [181], since expression of spacers decreases with distance from the leader end of the array, where transcription is initiated [182]. This may help to explain why most CRISPR arrays contain only between 10–40 spacers, even if they can in theory contain hundreds of spacers [183,184]. Indeed, theoretical studies that explore the trade-offs between effectiveness against a specific phage and coverage of as many phages as possible find that the common ranges of 10–40 spacers provide an optimal defence under a broad range of realistic parameter estimates [183,185]. The acquisition of multiple CRISPR arrays within a single host, along with sufficiently high levels of cas gene expression [104], may occur to overcome these challenges by maximizing novel spacer acquisition and memory span [186].

What does CRISPR-mediated coevolution look like?

The observations that the CRISPR-Cas system acquires a ‘‘memory’’ of past infections through spacer integration [28] and that phages can readily overcome CRISPR-Cas immunity through protospacer or PAM mutation [164,165] when using a clonal population of CRISPR-immune bacteria has led to the idea that CRISPR-immune bacteria and their phages are engaged in an ongoing coevolutionary arms race in which hosts accumulate spacers and phage accumulates point mutations [187]. This model of CRISPR-phage coevolution has been revised in recent years based on more refined models [118,179,188] and an appreciation of the mechanistic differences that exist between type I, type II, and type III systems [96].

Type I CRISPR-phage coevolution

Theoretical studies on CRISPR-phage coevolution predicted that over time, different spacers that each target the phage at different positions in its genome, appear in the population, resulting in high levels of spacer diversity at the population level. Meanwhile, phages with mutations to evade CRISPR targeting appear [144,179,187,188]. The ultimate outcome (i.e., ongoing coevolution or extinction of host and/or phage) was predicted to be influenced by the spacer acquisition rate, as well as the number of potential unique spacers (as determined by the requirement for PAM sequences)

Experimental evolution studies where a virulent mutant of phage DMS3 was used to infect *P. aeruginosa* strain PA14 found high population-level spacer diversity in the CRISPR array due to the acquisition of new spacers targeting the phage. Moreover, the majority of hosts in the population acquired a single, unique spacer [139,141]. This bacterial strain carries a type I CRISPR-Cas system that is primed against phage DMS3, which promotes rapid spacer acquisition and therefore generation of spacer diversity at the population level, as discussed above (**Figure 3**). As predicted by theory [188], the likelihood that phages are driven extinct is positively correlated with the level of CRISPR spacer diversity, and the level of spacer diversity naturally generated in a population drives phage extinct rapidly and consistently [189,190]. The immigration of susceptible hosts into the population can provide permissive hosts and enable phages to replicate [142]. However, these phages will still not evolve to be able to replicate on the CRISPR clones and will just coexist in the population, rather than coevolve with CRISPR. Hence, for CRISPR-Cas systems that generate high levels of population-level diversity, CRISPR-phage coevolution is likely only very short-lived, and in accordance, experimental studies have not found evidence to date for CRISPR-phage coevolution when bacteria carry a primed type I CRISPR-Cas system.

Type II CRISPR-phage coevolution

The most important model system for studying type II CRISPR-phage coevolution is that of the lactic acid bacterium *S. thermophilus* DGCC7710 and its phage 2972. Using this experimental system, CRISPR-Cas systems were first demonstrated to provide adaptive immunity against phages [28] and phages were shown to evolve to overcome CRISPR immunity through point mutation [165]. Co-culture studies in milk and in defined media have demonstrated that spacer acquisition occurs at a much lower rate than those observed in type I systems, despite some form of a priming mechanism in type II systems (**Figure 3**) [43]. As a result of this, bacterial populations during the early stages of a phage epidemic are virtually clonal, which is the result of a single bacterium that acquired a spacer sweeping to fixation [168,171,179,191]. Although initially, spacer abundance is determined by acquisition rates, rather than selection [192]. Several studies found that the bacteria and phage can coexist for extensive periods of time, ranging from tens to hundreds of generations [168,171]. In accordance with theory that predicts that lower rates of spacer acquisition increases the probability for CRISPR-phage coevolution [144,188], escape phages were found to emerge in these experiments, and an arms race where bacteria accumulate spacers and phage accumulate escape mutations ensues [168]. However, the arms race is asymmetrical, with the host acquiring cost-free spacers and slowly increasing the population-levels of spacer diversity, whereas the phage accumulates costly point mutations and is unable to keep up with the increase in spacers at the population level, ultimately resulting

in phage extinction [168]. Interestingly, a recent study examined how the environment shapes the coevolutionary interaction, with a focus on the role of spatial structure, which limits host and phage mobility and therefore could affect the effects of spacer diversity in the bacterial population. Data from this study suggest that escape phages emerge more readily in structured environments compared with well-mixed broth, leading to a greater number of coevolutionary cycles and, hence, a greater number of spacers that are acquired by the bacterial hosts [193].

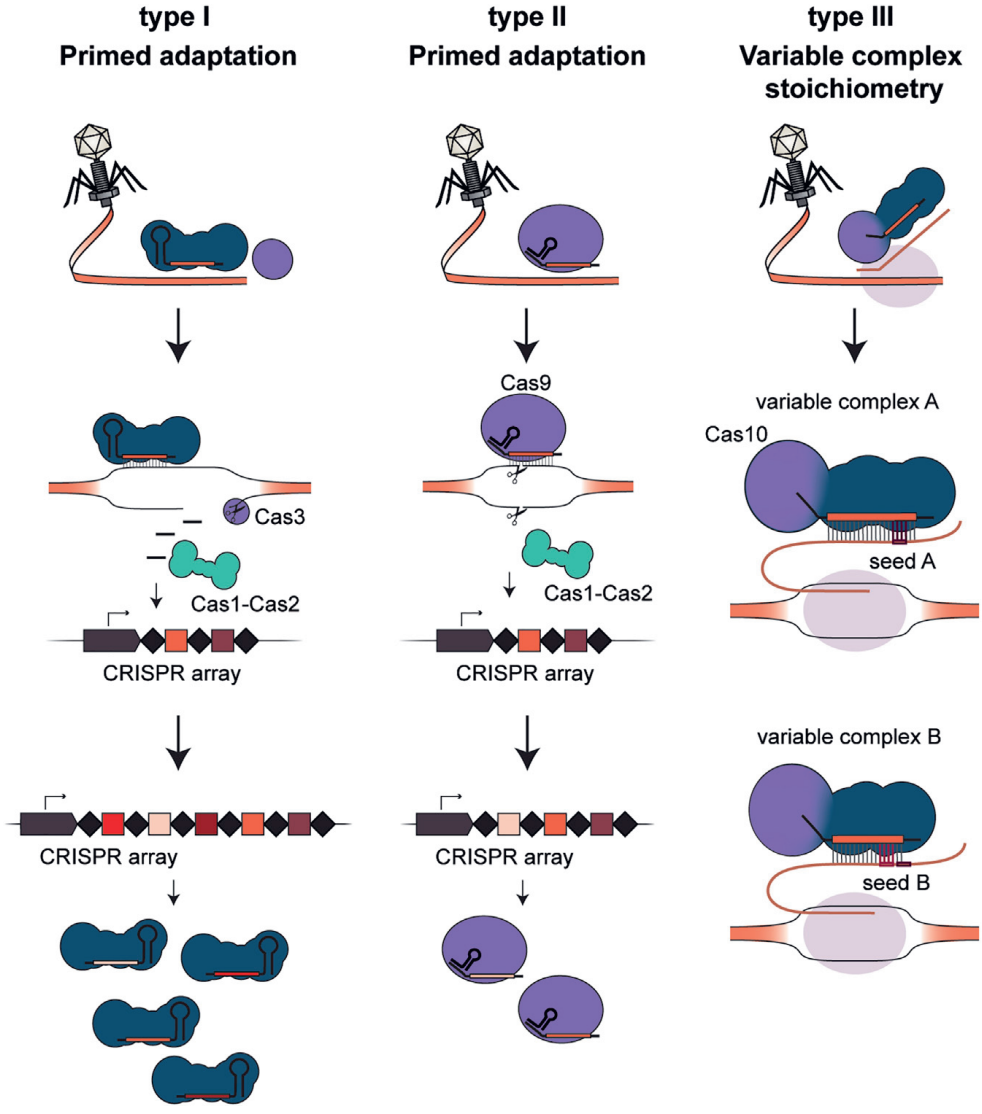


Figure 3. Diversity-generating mechanisms for the different CRISPR-Cas types
Type I systems generate population-level spacer diversity through primed adaptation when exposed to phages with perfect sequences or mutations that prevent interference. The effector complex (dark blue) will recognize and bind the target sequence. Cas3 (purple circle) is then recruited and will degrade the target sequence, generating substrates for spacer acquisition into the CRISPR array (squares represent spacers, black diamonds represent repeats) by Cas1-Cas2 (green). Type II systems can generate low levels of population-level spacer diversity, including through primed adaptation with spacers that perfectly target phages. Cleavage of the target DNA by Cas9 (purple circle) generates substrates for spacer acquisition by Cas1-Cas2 (green). Type III effector complexes vary in their size due to variability in the number of

Cas7-Cas11 (Cmr4-Cmr5) backbone segments. The variation in the length of the corresponding crRNA creates seed sequence flexibility, which then gives rise to individual-level diversity in CRISPR-Cas immune complexes. Type III effector complexes bind to target RNA (gray circles represent replication bubbles). In variable complex A, seed sequence A (purple lines) represents a sequence that must be perfectly complementary for targeting to occur. Variable complex B represents a smaller complex, with fewer backbone segments. Hence, seed B (pink lines) is in a different position to seed A.

Type III CRISPR-phage coevolution

Apart from diversity at the population level, diversity can also be generated at the individual level through the acquisition of multiple different spacers within the CRISPR locus of an individual bacterium, which makes it much harder to overcome CRISPR immunity by point mutation than single spacers. Recent studies reveal that type III systems have a distinct and unique mechanism to generate individual-level diversity, explaining why escaping from this CRISPR-Cas type is so rare, requiring deletions in targeted phage sequences [111]. Mechanistic studies have demonstrated that type III immune complexes vary in size, with smaller complexes carrying crRNA of reduced size [107,194]. The heterogeneity in the size of the crRNAs in type III systems is due to a secondary maturation step at the 3' end of the crRNAs, resulting in crRNAs with a variable 3' sequences. Crucially, a recent study showed that the type III-B system of *T. thermophilus* carries a 3' seed region that is critical for target RNA cleavage [78]. Where phages can overcome CRISPR immunity of type I and II systems through a single point mutation in the seed sequence [165,169], this is much harder in the case of this type III system, due to its variable 3' end, which defines different seed sequences [78]. By employing this unique strategy, type III complexes are able to create within-host diversity with just a single unique spacer (**Figure 3**) and are much more robust against rapidly evolving phages [111]. In addition to RNA cleavage, base pairing between the target RNA and the crRNA of type III systems activates the Cas10 subunit of the immune complex, leading to the production of cOA signalling molecules and sequence non-specific ssDNA cleavage activity 83. The activation of Cas10 is regulated by the Cas10 Activating Region (CAR) at the 5' end of the crRNA [78]. Mutations in the target at the 5' end of the crRNA affect the production of cOA but do not affect the sequence-specific target RNA cleavage [78]. The different effects of mutations in the 5' CAR or 3' seed sequence of the crRNA may help to explain why previous studies have suggested seed regions at both ends of the crRNA [195–198]. Taken together, variable 3' processing and the resulting complex composition, which defines the location of the seed sequence, along with activation of Cas10 through the 5' end of the crRNA, creates significant challenges for phage to overcome type III CRISPR immunity by mutation.

Conclusions and outlook

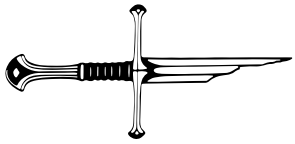
While we have made progress in understanding the fundamental concepts of CRISPR-phage coevolution using lab-based experiments, we are far from understanding the extent and mode of coevolution in natural environments. Metagenomics, comparative genomics, and experimental studies in semi-natural environments, including biofilms in a mine drainage systems [199], fish farms [200], and recolonized mouse gut [201], have provided evidence for the evolution of CRISPR immunity outside of the lab (reviewed in [116]). Moreover, these studies have identified CRISPR spacers that match co-sampled phage genomes and found that phages that persist over time often carry mutations that are predicted to enable escape from spacers found in earlier time points [180,199,202,203]. These studies suggest that in natural environments, phages and bacteria with CRISPR-Cas immune systems can coexist and coevolve. It is likely that the differences seen between natural environments and lab evolution experiments exist due to various ecological and evolutionary factors that have not been captured in lab experiments. These likely include the microbial community context in which these interactions take place, the spatial structure of the environment, as well as the levels of phage diversity. While some observational, theoretical, and experimental studies have started to explore how these and other factors impact CRISPR-phage coevolution [171,180], the generality of these observations is not yet clear, and further work with a greater number of model systems and different CRISPR-Cas types is needed to address the many open questions that remain.

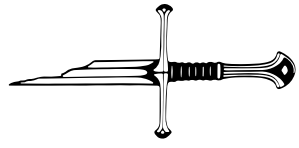
Acknowledgements

R.H.J.S. was supported by a VENI grant (016.Veni.171.047) from the Netherlands Organization for Scientific Research (NWO). E.R.W. acknowledges funding from the Natural Environment Research Council (grant ref. NE/S001921/1 and NE/M018350/1), and the European Research Council (ERC-STG-2016-714478 - EVO-IMMECH). S.v.H. acknowledges funding from the Biotechnology and Biological Sciences Research Council (BB/S017674/ 1 and BB/R010781/10).

Declaration of interest

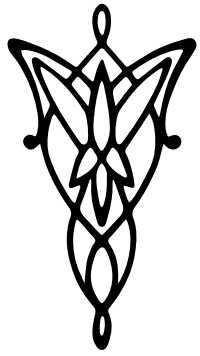
J.A.S and R.H.J.S. are shareholders of Scope Biosciences and are inventors on type III CRISPR-Cas-related patents. R.H.J.S. is a member of the scientific advisory board of Scope Biosciences.





Chapter 3

The diverse arsenal of type III CRISPR-Cas-associated CARF and SAVED effectors



Adapted from:
Steens, Jurre A., Carl Raymund P. Salazar, and Raymond H.J. Staals. The diverse arsenal of type III CRISPR–Cas-associated CARF and SAVED effectors. *Biochemical Society Transactions* 50, no. 5 (2022): 1353-1364.

Abstract

Type III CRISPR-Cas systems make use of a multi-subunit effector complex to target foreign (m)RNA transcripts complementary to the guide/CRISPR RNA (crRNA). Base pairing of the target RNA with specialized regions in the crRNA not only triggers target RNA cleavage, but also activates the characteristic Cas10 subunit and sets in motion a variety of catalytic activities that starts with the production of cyclic oligoadenylate (cOA) second messenger molecules. These messenger molecules can activate an extensive arsenal of ancillary effector proteins carrying the appropriate sensory domain. Notably, the CARF and SAVED effector proteins have been responsible for renewed interest in type III CRISPR-Cas due to the extraordinary diversity of defences against invading genetic elements. Whereas only a handful of CARF and SAVED proteins have been studied so far, many of them seem to provoke abortive infection, aimed to kill the host and provide population-wide immunity instead. A defining feature of these effector proteins is the variety of in silico-predicted catalytic domains they are fused to. In this mini-review, we discuss all currently characterized type III-associated CARF and SAVED effector proteins, highlight a few examples of predicted CARF and SAVED proteins with interesting predicted catalytic activities, and speculate how they could contribute to type III immunity.

Introduction

CRISPR-Cas is an adaptive immune system in prokaryotes that provides sequence-specific immunity against mobile genetic elements (MGEs), such as phages, transposons and (conjugative) plasmids, although other non-immune functions have been identified as well [7,28]. Well over a decade of research has highlighted the immense diversity of these systems, as reflected by their classification that currently distinguishes two main classes, six types, and many different subtypes [21]. Nevertheless, all CRISPR-Cas systems make use of an RNA-guided protein (complex) that binds and degrades complementary MGE-derived sequences. However, type III CRISPR-Cas systems seem to be equipped with an additional layer of defence that involves the production of signalling molecules and effector proteins that respond to them.

A typical type III system consists of several cas genes and a CRISPR array containing MGE-derived spacer sequences separated by repeat sequences (**Figure 1A**). Expression of the CRISPR array results in pre-crRNAs (pre-CRISPR RNAs) that are processed by the Cas6 protein into crRNAs 45. These are typically further processed at their 3' ends, resulting in mature crRNAs that start with an 8 nt repeat-derived handle at their 5' ends and with a variable 3' spacer-derived end [49]. Expression of the type III cas genes forms a complex with these crRNAs, resulting in a heterogenous population of type III ribonucleoprotein (RNP) complexes (**Figure 1A**).

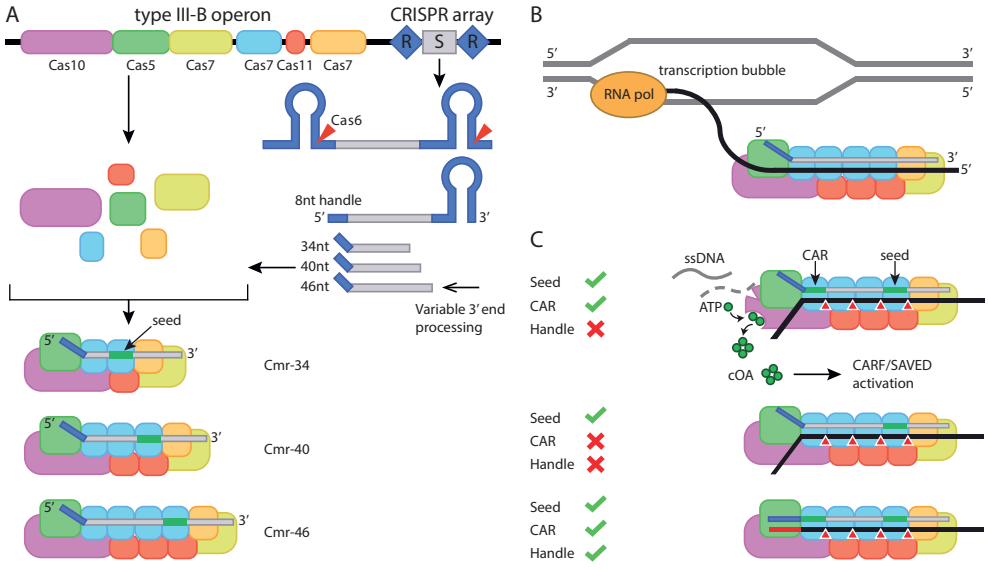


Figure 1. Mechanism of type III CRISPR-Cas immunity.
(**A**) Overview of type III effector complex formation. Expression of the cas genes and processing of the CRISPR array. Repeats and spacers are indicated by blue diamonds and grey rectangles, respectively. Following endonucleolytic cleavages of Cas6, a variable 3′ end processing step of the crRNAs leads to a heterogeneous complex size population. (**B**) Biological context of type III CRISPR–Cas systems. A transcription bubble is formed when MGE-derived dsDNA is transcribed by an RNA polymerase into (m)RNA, which is subsequently targeted by type III CRISPR–Cas. (**C**) Target RNA requirements for the various activities of type III. Note that RNA cleavage only relies on complementarity in the seed region, whereas ssDNase and cyclase (cOA production) activity requires additional strict base-pairing in the CAR and no base-pairing interactions with the 5′ handle.

Type III complexes are thought to operate in the context of a transcription bubble, where they bind MGE-derived transcripts complementary to the bound crRNA (**Figure 1B**) [76,77]. Binding is initiated at a 3’ exposed region of the crRNA called the seed sequence [78]. The variable 3’ end of the crRNA guarantees flexibility in targeting mutated RNA sequences, as the seed will be in a different location in differently sized type III complexes (**Figure 1A**). A seed-compliant target RNA will be bound and cleaved by the Cas7 subunits of the type III complex, cleaving it at 6-nt intervals [79–82]. Further base-pairing of the target RNA with the 5’ end, spacer-derived region on the crRNA will result in activation of the large Cas10 subunit of the complex; a region we designated as the CAR (Cas10-activating region). However, activation of Cas10 is prevented when binding self-RNAs (i.e. anti-sense RNA transcripts form the CRISPR array) (**Figure 1C**). This autoimmune-protection is governed by sensing base-pairing interactions between the 5’ handle and the corresponding ribonucleotides on the target RNA. Cas10 typically contains an HD domain, capable of cleav-

ing the exposed ssDNA regions in the transcription bubble) and a Palm domain. The Palm domain acts as a cyclase that generates cyclic oligoadenylate (cOA_x), where x stands for the number of adenosine residues in the ring-like structure) signalling molecules from ATP [73,83]. The number of adenosine residues can vary between different type III systems, but typically are in the range of cOA_2-cOA_6. Subsequently, cOAs bind and allosterically activate proteins containing the appropriate sensory domains: CARF (CRISPR-associated Rossmann fold) or SAVED (Second Messenger Oligonucleotide or Dinucleotide Synthetase-associated and fused to various effector domains). These sensory domains are often fused to a wide range of (predicted) catalytic domains. Over the last years, a handful of these auxiliary type III effectors have been characterized. Here, we will provide a short summary of our current understanding of these proteins (**Table 1**). Furthermore, we will highlight a couple of interesting examples of predicted, (non-nuclease) auxiliary type III effectors and speculate how they might contribute to type III immunity.

Table 1: Summary of characterized type III-associated, cOA-activatable effectors, their activating cOA species, substrate specificity, and phenotypic outcomes upon activation.

Protein	Domain architecture	No. of subunits in the active form	Activating cOAs	Substrate	Phenotypic outcome	Ring nuclease activity
Nuclease						
Csm6	CARF-6H-HEPN	2	cOA_4 [73,204] cOA_6 [73,83,208,209]	ssRNA	Phage clearance [205] Dormancy [91]	Yes [204,206,207]
Csx1	CARF-HTH-HEPN	2 6 [215]	cOA_4 [209-214]	ssRNA	Dormancy? [91]	Yes [211,212] No [214,216]
Can1	CARF-nuclease-like-CARF-PD-D/ExK	1	cOA_4 [89] cOA_3 [90]	dsDNA [89] ssRNA [90]	Phage clearance [89,90]	No [89]
Can2/Card1	CARF-PD-D/ExK	2	cOA_3 [90] cOA_4 [92,217]	ssDNA [90,92] dsDNA [90,217] ssRNA [90,92,217]	Dormancy [92] Phage clearance [92,217]	No [92,217]
NucC	PD-D/ExK	6	cOA_3 [86,88]	dsDNA [86,88]	Cell death [86,88]	No data
Non-nucleases						
CRISPR-LON	Lon-SAVED	1	cOA_4 [93]	CRISPR-T [93]	Cell death [93]	No data
TIR-SAVED	TIR-SAVED	>3	cOA_3 [87]	NAD+ [87]	Cell death [87]	No data
Csa3	CARF-HTH	1	cOA_4 [218]	dsDNA [218]	Transcriptional regulation [218]	No [218]

CARF nucleases

The first cOA-activatable proteins to be described were CARF nucleases, in particular Csm6 and Csx1, because they are frequently encoded in type III CRISPR-Cas operons (**Figure 3A**) [219]. Csm6 and Csx1 were shown to function as RNases in type III interference, despite their lack of physical associations with type III complexes [80,108,109,205,220,221]. Investigations into how Csm6 and Csx1 are activated upon target recognition by type III complexes led to the discovery of the cOA signalling system and the function of CARF domains [73,83]. Since then, new CARF nucleases have been characterized, showing different nuclease activities aiding in defence by promiscuously degrading both self and non-self nucleic acids. Here, we will summarize the characterized CARF nucleases, their catalytic activities, and their phenotypic outcomes.

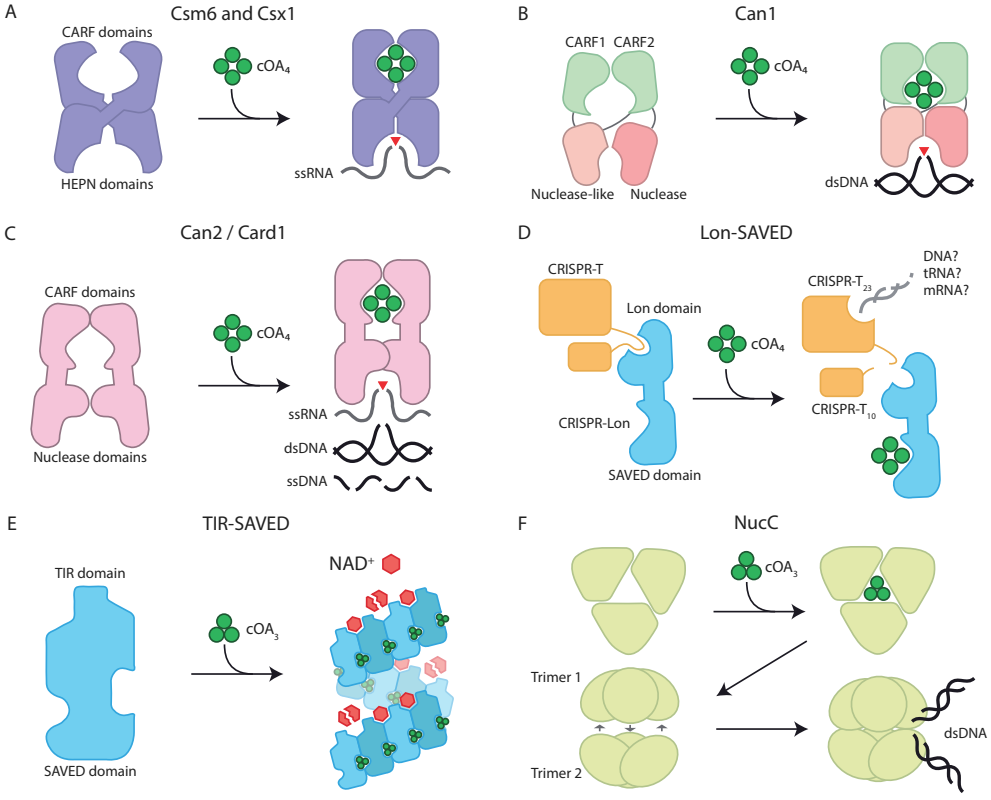


Figure 2. Schematic illustration of the activities of characterized type III-associated, cOA-activatable effectors.

(**A**) Csm6 and Csx1 homodimers bind cOA, stabilizing them into an active form where the HEPN domains catalyse ssRNA degradation. (**B**) Can1 occurs as monomers with two CARF domains, a nuclease-like and nuclease domain. Upon cOA binding, the nuclease-like and nuclease domain form a composite active site that catalyses dsDNA nicking. (**C**) Can2 homodi-

mers bind cOA and shift to an active form where the nuclease domains form a composite active site that catalyses dsDNA nicking, ssRNA and ssDNA degradation. (**D**) Lon-SAVED is initially bound to CRISPR-T, and upon cOA binding, cleaves CRISPR-T into CRISPR-T23 and CRISPR-T10. CRISPR-T23 then proceeds to degrade a yet unknown nucleic acid target. (**E**) TIR-SAVED forms superhelical structures upon cOA_3 binding, forming multiple composite NADase active sites for NAD+ degradation. (**F**) NucC homotrimers bind cOA3, causing conformational changes that promote homohexamer formation, and forming dsDNA cleavage sites across the two homotrimers.

Csm6

Among the first CARF proteins to be described was Csm6, in part due to its high prevalence in type III-A CRISPR-Cas operons [219]. These proteins have an N-terminal CARF domain and a C-terminal HEPN (higher eukaryotes and prokaryotes nucleotide-binding) domain, interspaced with an α-helical region (6H) [220]. Csm6 forms homodimers in solution to form a cOA binding pocket and a composite ribonuclease active site, in a dynamic conformational equilibrium between active and inactive forms, conferring low levels of non-specific RNase activity [205,206,220]. Upon cOA binding, Csm6 homodimers stabilize to the activated conformation formed by the histidine residues of the catalytic HEPN domain, leading to a highly-active, non-specific RNAse that cleaves single-stranded RNAs (ssRNAs) after purines (**Figure 2A**) [83,208,220]. Besides ssRNase activity, some Csm6 homologues have ring nuclease activity (conferred by either the CARF or HEPN domain) that cleaves bound cOAs, thereby autoregulating their activity [206,207]. *In vivo*, the contribution of Csm6 to type III immunity becomes important in situations with sub-optimal type III targeting, such as with late-expressed viral genes, mutated targets, or infrequently transcribed plasmid genes. Here, Csm6-mediated RNA degradation becomes indispensable when Cas10 HD-mediated DNA degradation is insufficient in halting MGE DNA accumulation [198,205]. During viral infections, Csm6 activity does not seem to impair cell growth, while for plasmid invasions, it causes temporary growth arrest until the plasmid is cleared [198,205]. This may be due to a high concentration of phage genomes and transcripts during infections, saturating the active Csm6, thereby reducing the impact on host transcripts.

Csx1

Another CARF protein that was identified early on through bioinformatic analyses of type III-B CRISPR-Cas operons was Csx1 [219]. Similar to Csm6, Csx1 also has an N-terminal CARF domain and a C-terminal HEPN domain, but are separated by a helix-turn-helix region (HTH), and forms homodimers (**Figure 2A**) [206,209,221,222]. Interestingly, *Sulfolobus islandicus* Csx1 (SisCsx1) has a unique structure where it forms a hexamer built from a trimer of homodimers [214]. Csx1 has non-specific ssRNase activity upon cOA binding, catalysed by the HEPN domains. The sequence specificity of this ssRNase activity can vary between homologues: *Pyrococcus furiosus* Csx1

(PfuCsx1) cleaves after adenosines, while SisCsx1 cleaves in between two cytosine residues [213,221]. SisCsx1 and *S. solfataricus* Csx1 do not exhibit ring nuclease activity but seem to be dependent on dedicated ring nuclease proteins to break down the cOAs [214,223]. PfuCsx1 can degrade cOA_4 through the HEPN domain, while T*hermus thermophilus* Csx1 does so through the CARF domain [211,212,214,216]. Biologically, the importance of Csx1-related ssRNase activity in type III defence varies among different species. For example, SisCsx1 is necessary for type III-B CRISPR-Cas plasmid interference in *S. islandicus* [106]. However, in *P. furiosus*, PfuCas10 and PfuCsx1 seem to serve redundant roles in plasmid interference, where a combination of PfuCas10 HD domain mutations and either PfuCsx1 or PfuCas10 Palm domain mutations are necessary to abrogate defence [212]. The impact of Csx1-mediated RNA degradation on the fitness of the host remains to be determined, but it is suggested that it would be similar to Csm6 [91].

Can1

Besides RNA, cOA-activatable nucleases can also degrade DNA, as seen with Can1 (CRISPR ancillary nuclease 1). Can1 appears to be limited to the genus Thermus, and unlike Csm6 and Csx1, Can1 operates as a monomer and contains two CARF domains separated by a nuclease-like domain and a C-terminal PD-D/ExK nuclease domain [89]. Binding of cOA to TtCan1 induces a conformational change to form a composite DNA cleavage site, formed by the nuclease-like and PD-D/ExK domains, nicking supercoiled DNA at random sites (**Figure 2B**). This nicking activity is believed to slow down viral replication by mediating the collapse of DNA replication forks and subsequently causing dsDNA breaks in rapidly replicating phage genomes. Interestingly, a recent pre-print showed that TtCan1 can change substrate specificity depending on the nature of the bound cOA, with cOA_3 binding inducing dsDNase activity, and cOA_4 binding inducing ssRNase activity in vitro [90]. The cOA_3-specific dsDNAse activity is hypothesized to be a more lethal alternative in case the cOA_4-induced ssRNase activity fails to clear the infection. Further investigations will be needed to fully understand the observed cOA-dependent activities.

Can2 / Card1

A close relative of Can1, Can2 / Card1 (cOA-activated ssRNase and ssDNase 1) has a domain architecture composed of an N-terminal CARF and a C-terminal PD-D/ExK nuclease domain, and forms homodimers similar to Csm6 and Csx1 [92,217]. Upon cOA_4 binding, two studies have shown that Can2 degrades ssRNA in vitro [92,217]. For DNase activity, one study showed ssDNase but not dsDNAse activity with the *Treponema succinifaciens* Can2, while another study showed progressive DNA nicking activity that eventually led to dsDNA degradation with *Sulfobacillus thermosulfidooxidans* and *Thioalkalivibrio sulfidiphilus* Can2 (**Figure 2C**) [92,217]. It is likely that its canonical function is DNA nicking, given two homologs are known to exhibit

this activity. In vivo, Can2 is suggested to induce dormancy in response to phage infection and plasmid transformation [92]. Can2 is thought to introduce DNA lesions in both the host and phage genome and acts in parallel with Cas10 to eliminate target DNA [92]. In contrast to this, another study found that Can2 provided phage immunity without causing any noticeable growth defects of the host, suggesting that Can2 adequately slows down phage replication similar to Can1 [217].

Non-nuclease CARF

Many of the currently characterized CARF proteins associated with type III immunity are nucleases, which aid in defence by promiscuously degrading both self and non-self nucleic acids. Although more research is needed, it appears that most of these systems operate as an altruistic mechanism to protect the population by inducing cell dormancy or cell death of the infected individual. Similar to other abortive infection mechanisms, there are multiple ways to induce dormancy or cell death and this is reflected by the many catalytic activities that are predicted to be associated with CARF proteins [84,104,115,224–230]. Here, we will discuss a few interesting examples of downstream type III effectors and speculate how they might induce dormancy or cell death.

cOA-responsive transcriptional regulator

The only experimentally characterized non-nuclease CARF effector known to date is a transcriptional regulator, Csa3, which is often found in type I-A systems (**Figure 3B**) [231]. These proteins are a fusion between a CARF domain and an HTH domain, commonly involved in DNA binding and influencing expression. The complete regulatory functions of these effectors appear to be very complex, but hints to crosstalk between type III and type I systems. It has been shown that Csa3 is involved in the regulation of type I CRISPR adaptation, as well as providing a feedback loop to type III interference [134,218,232–234]. Furthermore, Csa3-mediated activation of DNA repair genes has been demonstrated, indicating that the network of gene regulation by type III associated effectors might not be constricted to CRISPR-related genes [235].

Transmembrane CARF effectors

In silico analyses indicated that many CARF proteins are fused to a transmembrane (TM) domain [84]. One mechanism could be that these CARF-TM proteins form ion channels in the membrane upon activation by cOAs (**Figure 3C**). Subsequent depolarization of the membrane would be a means to disrupt many processes in the cell, eventually resulting in altruistic cell death, akin to some superinfection exclusion systems encoded on prophages [228]. Alternatively, CARF-TM activation could also result in complete mechanical disruption of the membrane, as seen in another abortive infection systems [226,236].

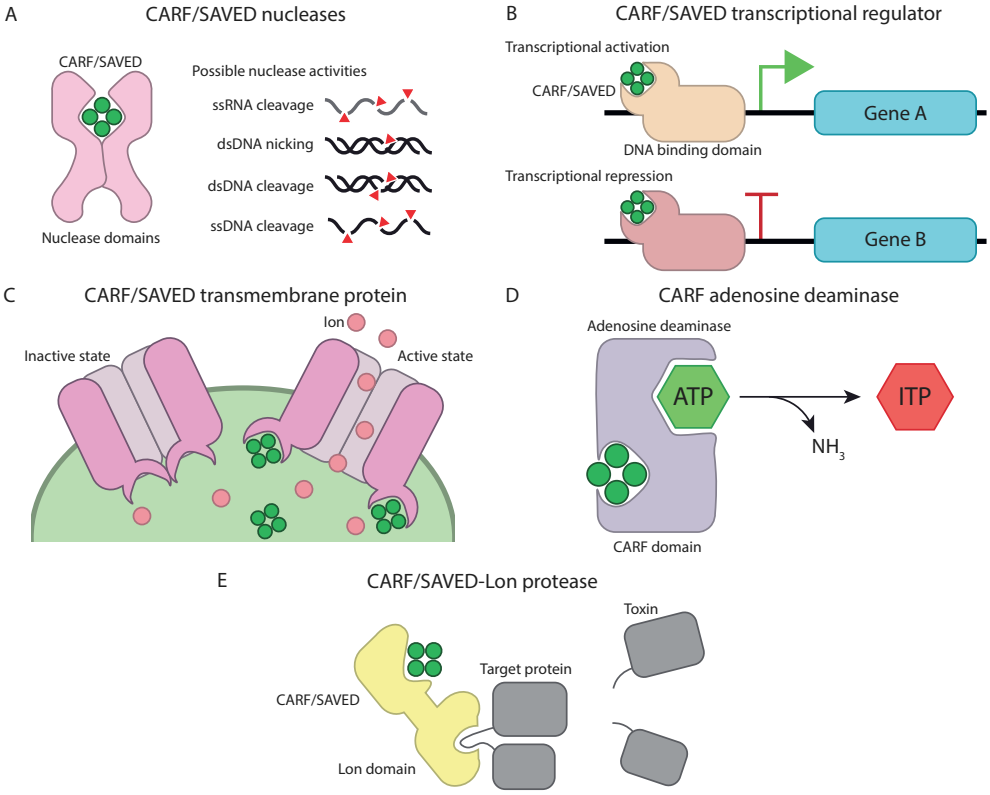


Figure 3. Schematic illustration of the anticipated activities of bioinformatically-predicted type III-associated, cOA-activatable effectors.
(**A**) CARF or SAVED proteins with promiscuous nuclease activity, cleaving both self- and non-self nucleic acids. (**B**) CARF or SAVED proteins with DNA binding domains could enhance or repress downstream effector genes. (**C**) CARF or SAVED proteins with transmembrane domains could form pores that depolarize the membrane, depriving the cell of energy. Alternative strategies to disrupt the membrane could be employed too. (**D**) CARF proteins with predicted adenosine deaminase domains converting ATP into Inosine triphosphate (ITP), depleting cellular ATP levels. (**E**) CARF or SAVED proteins with a fused Lon protease domain liberating a toxin that kills the cell.

CARF adenosine deaminase

Although not as widespread as the CARF-TM fusions, some CARF proteins have an adenosine deaminase (ADA) domain. These enzymes typically convert adenosines into inosine residues, and some non-CARF-associated ADAs have important housekeeping roles in prokaryotes, such as the editing of tRNAs [237]. Activated CARF-ADA could therefore act by dysregulating these processes to induce cell death. However, the de-amination of nucleotides could also be leveraged to deplete the cell of ATP, similar to a strategy that is employed by other phage defence systems (Figure 3D) [230].

CARF-Lon effectors

The active domain of a family of housekeeping proteases, the Lon domain, also seems to be adapted for type III-mediated phage defence, as demonstrated by the bioinformatically-predicted CARF-Lon fusions [84]. Canonical Lon proteases are involved in the degradation of misfolded and abnormal proteins as well as certain regulatory proteins [238–241]. In the context of phage defence, we speculate that the CARF-Lon protein, when activated by cOA, could either cleave essential host protein targets or acts as an aggressive promiscuous protease, both geared towards killing the host (Figure 3D).

SAVED effector proteins

In archaea and bacteria, cyclic oligonucleotide-based antiphage signalling system (CBASS) immunity systems are widespread, providing a diverse arsenal of anti-phage defence tools [242–246]. Typically, these systems encode a cGAS/DncV-like Nucleotidyltransferase (CD-NTase) protein, responsible for sensing the presence of phage and the subsequent synthesis of a second messenger molecule [247,248]. The messenger molecules resemble signalling molecules of type III defence but can contain a variety of nucleotide moieties and different linkages between them. Upon recognition by CD-NTase-associated protein (Cap) effectors, an elaborate immune response is initiated that can lead to cell death. SAVED is a common sensing domain for these Cap proteins and has long been predicted to be involved in type III immunity, but was until recently not experimentally demonstrated [84,249]. Although this sensing domain has limited sequence similarity to CARF domains, it is thought to be a highly divergent version of CARF, fused to a variety of effector domains [84,247,250].

SAVED nucleases

Similar to the abovementioned CARF-nuclease fusions, a large array of different domains predicted to confer non-specific (ribo)nucleases are commonly found in SAVED proteins (**Figure 3A**) [84]. Notably, the SAVED-HNH fusion proteins appear to be a common example. Of note, this domain, named after the catalytic residues, is also responsible for sequence-specific target cleavage in some type II CRISPR-Cas systems [21,67].

Lon-SAVED

The first example of a connection between SAVED and type III CRISPR-Cas, a Lon-SAVED protease, was recently demonstrated and revealed a new mechanism by this system to aid in defence [93]. The Lon-SAVED effector (CRISPR-Lon) contains a C-terminal SAVED sensing domain, consisting of two CARF-like domains, fused to a N-terminal Lon protease domain. Binding of a cOA_4 messenger molecule induces an allosteric change in the protein that activates this effector (**Figure 2D**). Interestingly and in contrast to canonical Lon proteases, CRISPR-Lon appears to have a specific target protein, CRISPR-T. The 32 kDa CRISPR-T protein is cleaved by activated CRISPR-Lon into two fragments (~23 and ~10 kDa). The ~23 kDa fragment bears structural similarity to MazF, which is a toxin known to cleave specific rRNA, mRNA and tRNA molecules, leading to abortive infection [251,252]. The triggering of a toxin/cell-death signal by a protease is something observed in both prokaryotes and eukaryotes and seems to be an evolutionary conserved strategy for inducing cell death [253,254]. It is therefore anticipated that similar type III CRISPR-Cas protease-mediated defence strategies will be uncovered, acting through different protease-like toxins that perturb essential cellular targets.

TIR-SAVED

The Toll/interleukin-1 receptor (TIR) domain is widely found in all domains of life [255–259]. In humans, this domain is often present in Toll-like receptors to mediate signalling for innate immunity. In response to binding their ligand(s), TIR domains of certain immune receptors in plants synthesize a signalling molecule to induce cell death [258]. In Thoeris, a bacterial anti-phage defence system, TIR domains are responsible for the production of nicotinamide adenine dinucleotide (NAD) derived signalling molecules which in turn allosterically activate a TIR domain-containing enzyme that aggressively depletes NAD to arrest cell growth [260]. The depletion of NAD is a strategy that is also employed by a prokaryotic short Argonaute immune system upon the detection of invading DNA [261].

The modularity of known CBASS systems and their interplay with CRISPR-Cas defence is highlighted by a recent study on a TIR-SAVED effector protein [87]. The CBASS system it originated from generates cOA3 messenger molecules. These are

bound by TIR-SAVED and mediate its multimerization, forming composite active sites to degrade NAD+ (**Figure 2E**). TIR-SAVED can induce cell death *in vivo* when placed in the context of a type III system by replacing the canonical Csm6 CARF ribonuclease by the TIR-SAVED effector. This demonstrates interchangeability between CBASS and CRISPR-Cas defence systems, but natural examples of type III CRISPR-Cas systems in combination with the NucC nuclease and TIR-SAVED effectors exist as well [84]. The widespread usage of TIR domains in defence systems across the domains of life can be seen as proof that several eukaryotic immune systems originated from an ancestral prokaryotic anti-phage system [259].

Transmembrane SAVED effectors

Similar to CARF effectors, many SAVED proteins containing a transmembrane domain have been predicted bioinformatically [84]. A similar mechanism as described for the CARF-TM fusions could be employed by this type of SAVED effector (**Figure 3B**). Although rather speculative, an exciting possibility arises that these transmembrane type III effectors position the sensory SAVED domain on the outside of the cell, hinting at intercellular signalling. Signalling the presence of infection to others in the population can be seen in other CRISPR-Cas systems as a means to strengthen the immune response [128,132].

NucC, a non-CARF and non-SAVED effector

cOA-activatable effectors are not limited to CARF and SAVED proteins, as exemplified by NucC. NucC (nuclease, CD-NTase associated) is a CBASS-associated protein that has also been found in [31] type III CRISPR-Cas loci [88]. In both CBASS and type III systems, NucC forms homotrimers with three active sites on the outer edge. Upon binding of cOA_3, pairs of NucC homotrimers bind to form homohexamers, juxtaposing pairs of partial active sites between the two homotrimers and forming dsDNase active sites (**Figure 2F**). This results in double-stranded breaks on dsDNA with two-base 3' overhangs [88,90]. *In vivo*, in both CBASS and type III CRISPR-Cas systems, NucC appears to act through an abortive infection mechanism whereby its activation causes the complete destruction of the host chromosome, culminating in cell death [86,88]. This can be a beneficial characteristic of this nuclease to overcome phage escape strategies, such as the ability of type III-associated NucC to overcome jumbo phage infections in Serratia [86]. These phages protect their DNA after injection using a proteinaceous nucleus inside the host [262]. Instead of targeting phage DNA, NucC degrades the host genome and thereby kills the host to prevent phage progeny (abortive infection).

Perspective

(**i**) Type III CRISPR-Cas systems are sophisticated multi-layered immune systems, primarily aiming to clear invading MGEs by cleaving target RNAs complementary to the guide RNA, but will also produce cOA signalling molecules to activate its second layer of defence, mediated by CARF and SAVED proteins.

(**ii**) CARF and SAVED proteins have a plethora of catalytic activities associated with them, most of which seem to be geared towards killing the infected host (and thereby preventing viral progeny) to provide population-wide immunity; a mechanism known as abortive infection. Collateral damage observed in other CRISPR-Cas systems indicates that this strategy is not limited to type III [7].

(**iii**) Most CARF proteins characterized to date are sequence-unspecific (ribo) nucleases that induce cell death or dormancy by cleaving both self and non-self nucleic acids. However, bioinformatic analyses have shown that many other CARF and SAVED proteins are fused to other catalytic domains (proteases, deaminases, NADases, etc). If and how these different activities contribute to type III mediated defence and what effect they will have on the fitness of the host, will be an interesting challenge for the future.

Competing Interests

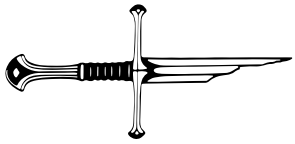
J.A.S. is a founder and shareholder of Scope Biosciences. R.H.J.S. is a shareholder and member of the scientific board of Scope Biosciences.

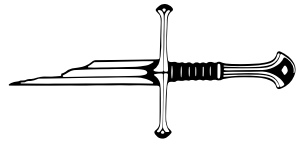
Author contributions

All authors contributed to the manuscript, read and approved the final manuscript.

Funding

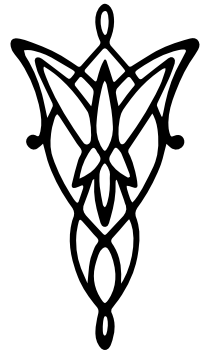
R.H.J.S is supported by a VIDI grant (VI.Vidi.203.074) from The Netherlands Organization for Scientific Research (NWO).





Chapter 4

CARFish: identifying cOA-binding effector proteins of type III CRISPR-Cas systems



Manuscript in preparation:
Jurre A. Steens, Patrick Barendse, Jasper Groen, Sjef Boeren, John van der Oost, Daan C. Swarts, Raymond H.J. Staals

Abstract

Type III CRISPR-Cas systems have the ability of using cyclic oligoadenylate (cOA) signalling molecules to defend their host against invading genetic elements. When the type III CRISPR-Cas complex detects target RNA, it produces cOA signalling molecules. These molecules bind various cOA-responsive effector proteins to induce cell dormancy or cell death. Using bioinformatics to predict the network of cOA-responsive effector proteins is not straightforward, and pose the risk of missing specific cOA effectors. We therefore developed CARFish, a bottom-up approach to experimentally identify cOA binding proteins. Here, cOA molecules are immobilized to fish cOA-binding proteins from cell lysates. We experimentally demonstrate that CARFish can be used to identify cA_4 interacting proteins in several bacteria and archaea. Not only does CARFish allow the experimentally identification of bioinformatically predicted cOA binding proteins, it also allows identification of non-canonical cOA binding proteins that may be involved in the CRISPR-Cas type III defence response.

Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins comprise adaptive CRISPR-Cas immune systems found in bacteria and archaea. In CRISPR-Cas systems, different Cas proteins and CRISPR-derived RNA (crRNA) guides form ribonucleoprotein complexes to target and neutralize complementary nucleic acids of invading bacteriophages and plasmids. The highly diverse CRISPR-Cas systems are divided into two classes, seven types and dozens of subtypes [21,263]. Class I Type III CRISPR-Cas complexes have multiple unique features. Next to sequence specific target RNA degradation, crRNA:target RNA base-pairing triggers activation of Cas10, a multidomain protein subunit that is part of the type III Cas effector complex (**Figure 1A**) [79,106,108,109]. The HD-type nuclease domain of activated Cas10 cleaves neighbouring (invader) ssDNA non-specifically [77,112,177]. In addition, the PALM-type polymerase domain of Cas10 acts as an adenylate cyclase that converts ATP into cyclic oligoadenylate (cOA) molecules. The size of cA_x (x represents number of AMP moieties) molecules can vary among different type III systems, typically containing 2-6 adenosine moieties in a ring-like structure [73,83]. These cOA molecules act as signalling molecules that activate auxiliary effector proteins [75]. The cOA signal transduction pathway has been thoroughly investigated in recent years, revealing a wealth of events downstream of target RNA recognition [75,264]. This includes cOA signalling molecules triggering catalytic activities that are geared towards killing the host or inducing dormancy to prevent the spread of the invading nucleic acids among the population [86–93].

Many of the downstream effectors in cOA signalling pathways contain cOA-sensing CARF and SAVED domains, which can be identified bioinformatically. However, more divergent CARF and SAVED domains or analogous cOA sensory domains might not be identified using a bioinformatics approach [84,265]. For example, the NucC effector protein and Csx23 are activated by cA_3 and cA_4 respectively, and do not contain a typical CARF nor SAVED sensory domain [266,267]. Another complicating factor is that not all genes encoding cAO-sensing proteins reside in (close proximity to) the type III operon, obscuring their functional association with the type III system, as many members of the CARF/SAVED family of proteins even operate outside physical context of type III complexes [84]. Together, this raises the question whether the current landscape of type III effector enzymes is complete, or that additional effectors exist. In this study we describe a new experimental approach to identify cOA-interacting proteins in a high-throughput manner to obtain better insights into the expansive cOA signalling network. The methodology, which we termed CARFish, makes use of immobilized cOA molecules to pull down cOA-interacting proteins from lysates of cells of choice, followed by identification of the enriched cOA-interacting proteins by LC-MS/MS (**Figure 1B**).

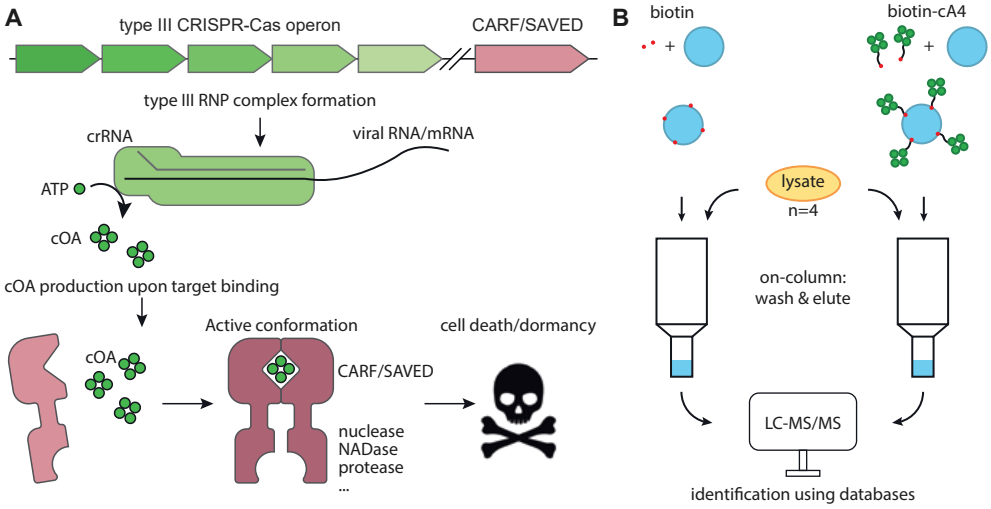


Figure 1. Schematic representation of type III CRISPR-Cas signalling pathway and CARFish. (**A**) The type III CRISPR-Cas gene operon encodes Cas proteins that together with a crRNA encoded in a CRISPR locus form a ribonucleoprotein (RNP) complex. Upon sequence-specific detection of target RNA complementary to the crRNA, the Cas10 subunit is catalytically activated. This results in target RNA cleavage by the Cas10 HD domain, and cOA production by its PALM domain. Produced cOAs act as secondary messenger molecules that allosterically activate ancillary effector proteins, eventually resulting in cell dormancy or death. (**B**) In CARFish, streptavidin beads with immobilized biotin-conjugated cA_4 molecules are used to generate a column for affinity chromatography. As control, another column prepared with streptavidin beads loaded with equimolar amounts of biotin is used. Cell lysate is applied on both columns for affinity chromatography. After washing of the column, associated proteins are eluted and analysed via LC-MS/MS. Enrichment of proteins eluted from the cA_4 beads is determined by comparing their abundance to the abundance of proteins eluted from the control biotin-covered beads.

Results

CARFish can be used to identify cOA-binding proteins

The aim of the CARFish method is to enrich cOA-interacting proteins from cell lysates. To this end, biotin-conjugated cA_4 molecules are immobilized on streptavidin beads which enables cA_4-affinity chromatography. In parallel, an experimental control is performed to correct for background protein interactions in each experiment by covering streptavidin beads with biotin alone (**Figure 1B**). The cell lysate of an organism of interest is incubated with the biotin-cA_4 beads and with the control biotin-covered beads, after which the beads are washed to remove excess proteins. Subsequentially, proteins interacting with the beads are eluted by including biotin in the elution buffer. Protein content and abundance in the elution fractions is analysed by quantitative LC-MS/MS. Enrichment of specific proteins is calculated by expressing the label free quantification (LFQ) intensity of proteins identified in the biotin-cA_4 eluate as a ratio of those identified in the control eluate.

In order to validate the CARFish methodology, we performed the method on lysates of *Thermus thermophilus* strains HB8 and HB27, which encode well-characterized cA_4 effector enzymes. Known cA_4 effectors include Csx1, Csm6 and Can1 in *T. thermophilus* HB8 and Csm6 and Can1 in *T. thermophilus* HB27 (**Figure 2A-B**) [73,84,89,211]. First, the optimal cell lysate:bead ratio and performance of the bead type (agarose streptavidin beads vs. magnetic streptavidin beads) were investigated. To this end, enrichment of Csm6 from *T. thermophilus* HB27 lysates was quantified under various conditions (**Figure S1**). These experiments demonstrated that highest Csm6 enrichment was obtained using agarose streptavidin beads with 10 mg cell pellet per µl beads, which was used for all subsequent experiments.

The optimized CARFish method was used for cA_4 pull-down experiments using lysates of *T. thermophilus* HB8 and HB27. All experiments and subsequent LC-MS analyses were performed in quadruplicates. Significance of the enrichments was determined by performing a Students t-test on the LFQ intensity of cA_4-enriched proteins compared to the LFQ intensity of the proteins present in the control eluate. The p-value in turn is used to calculate an enrichment score:

$$-\log_{10} \text{p value} \times \log_{10} \text{fold change} = \text{enrichtment score}$$

Using CARFish, we were able to identify all known cOA-interacting CARF proteins from the *T. thermophilus* HB8 lysate: Csx1, Csm6 and Can1 (**Figure 2C**). Csm6 (TTHB152) showed the highest enrichment with an enrichment score of 9.61. Csx1 (TTHB144) was also highly enriched (enrichment score 6.50), whereas Can1

showed lower but significant enrichment (enrichment score 0.9) (**Figure 2D**). For the *T. thermophilus* HB27 enrichment, a similar CARF protein enrichment was obtained (**Figure 2E**), with both Csm6 (TT_P0107) and Can1 (TT_P0110) showing strong and significant enrichment (enrichment scores 17.64 and 1.50, respectively) (**Figure 2F**). The differences in detection sensitivity might be explained by differences in affinity for the biotinylated cA_4 bait, or by different expression levels [89]. This can be corroborated by the previous findings indicating that Csm6 has a 1.88±0.09-fold higher expression compared to Can1 in *T. thermophilus* HB27 [268]. As Csm6 and Can1 might be competing for cA_4 binding, we hypothesized that Can1 detection could be enhanced by reducing the lysate to bead ratio. Indeed, when the lysate to bead ratio is reduced, the detection signal to noise ratio of Can1 is increased (**Figure S2**). This implies that varying the lysate to bead ratio can result in enhanced detection of specific proteins that might have lower affinity for the cA_4-biotinylated beads and/or are less abundant in the cell lysate.

Together, these results demonstrate that CARFish can be used to successfully identify cOA-interacting proteins in cell lysates and suggests that it can be used for the identification of novel cOA-binding proteins.

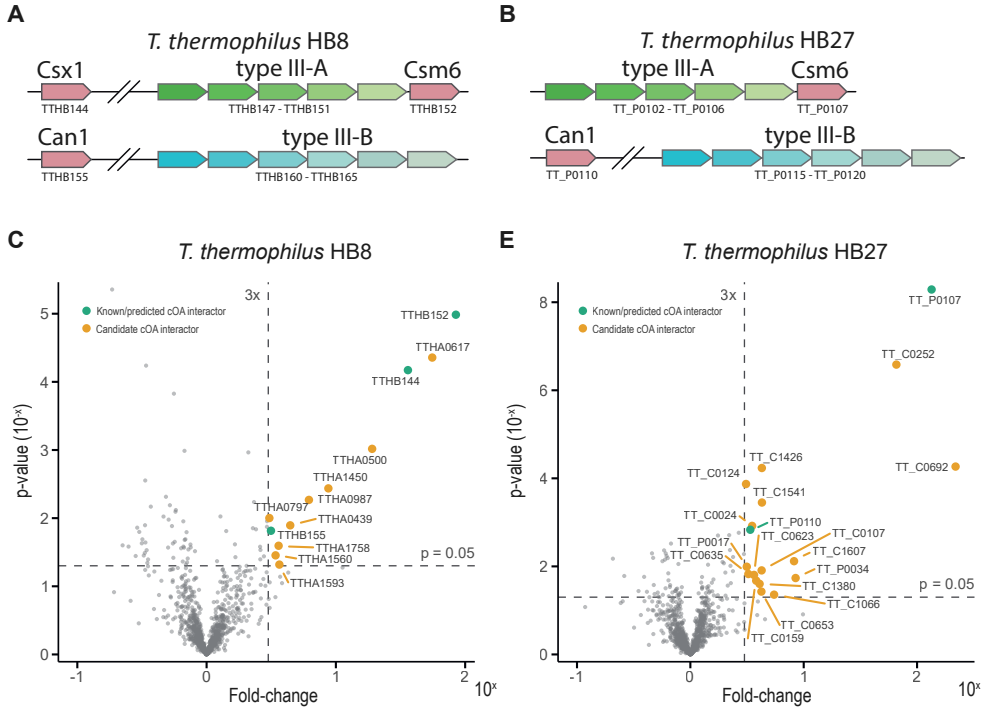
Identification of putative cOA-binding proteins in bacteria

We further investigated if CARFish can be used to identify novel cOA-binding proteins. In both T. *thermophilus* strains, CARFish resulted in the enrichment of several other putative cA_4-binding proteins (enrichment>3-fold, $p<0.05$) (**Figure 2C-F**). To determine whether these proteins have structural features that could facilitate cA_4 binding, we used AlphaFold2-predicted or experimentally determined structures (when available) to perform a structural homology search against structures from the Protein Data Bank. From the top 20 of the identified structural homologs, the associated PFAM domains were analysed for recurring or highly similar domains (**Figure 2D,F**). In addition, bound ligands and designated protein function of selected structural homologs were analysed (**Figure 2D,F**). Finally, we investigated whether the structural models of cA_4-binding candidates show structural homology to CARF domains by pairwise alignment with the structure of *T. thermophilus* HB8 Csm6 (TtCsm6, PDB: 5FSH) (**Figure 2D,F**).

For both *T. thermophilus* strains, CARFish identified a nicotinate phosphoribosyltransferase (NAPRT), with high enrichment scores of 7.60 (TTHA0617; HB8) and 11.97 (TT_C0252; HB27) respectively (**Figure 2D,F**). While NAPRT does not structurally resemble the CARF domain, it is predicted to harbour a (partial) nucleotide-binding Rossmann fold (**Figure 2D,F, S3A**). NAPRT is an ATP-dependent enzyme which converts nicotinate and phosphoribosyl pyrophosphate into nicotinate

ribonucleotide, one of the first steps in NAD biosynthesis [269]. In ligand-bound structures of NAPRT the 5P-ribose moiety of the substrate (5-phospho-α-D-ribose 1-diphosphate) is bound at the entrance of the active site [270]. We speculate that also the 5P-ribose backbone of cA_4 can bind at this site, resulting in enrichment of NAPRT during CARFish. CARFish also identified a predicted NYN/PIN domain-containing protein (TT_C0692; enrichment score: 9.98) in strain HB27 (**Figure 2F**), which contains a Rossmann fold that is structurally homologous to the CARF domain of TtCsm6 (**Figure S3B**). Furthermore, structural homology searches suggest TT_C0692 contains a NYN domain with a PIN domain-like fold, which are known to confer endoribonuclease activity and have been identified in genetic regions that also encode type III CRISPR-Cas systems (**Figure 2F**) [250,271]. Other putative cOA-binding proteins identified include a putative PLD phosphodiesterase (TTHA1450) with predicted nuclease and OB-fold domains, and a putative GGDEF adenylate cyclase (TTHA1758). Both proteins show low or no structural homology to the CARF domain of TtCsm6 (**Figure 2D**). In *T. thermophilus* HB27, the small helix-turn-helix protein TT_C0653, which shares homology with transcription factors, was enriched (**Figure 2F**).

These results show that CARFish, next to identification of known cOA-binding proteins, can be used to identify putative novel ATP- and/or cOA-binding proteins without high sequence or structural homology to known CARF sensing domains. Yet, whether any of these putative cA_4-interacting proteins can bind and/or cleave, and/or is activated by cA_4 remains to be determined experimentally.



D

Thermus thermophilus HB8

Locus	CARF	Name	Score	Fold change (10^x)	Domain analysis (AF2 structural homology to PDB25 by DALI)	CARF DALI Z-score
TTHB152	Csm6	CRISPR system endoribonuclease Csm6	9.61	1.93	Csm6, DUF1887, Card1, Can1, HEPN	35.3
TTHA0617		Nicotinate phosphoribosyltransferase	7.60	1.75	Various nicotinate proteins, ribose binding	2.9
TTHB144	Csx1	Uncharacterized protein	6.50	1.56	Csm6, DUF1887, Card1, Can1	15.5
TTHA0500		Glycerate dehydrogenase/hydroxypyruvate reductase	3.86	1.28	Glycerate kinase, binds ATP/ADP	5.9
TTHA1450		Uncharacterized protein	2.30	0.94	PLD phosphodiesterase, bilobed structure: ssDNA/RNA nuclease and OB-fold	2.8
TTHA0987		Beta-ketoadipyl CoA thiolase	1.79	0.79	Thiolase, coenzyme A binding	2.1
TTHA0439		ABC transporter, ATP-binding protein	1.23	0.65	ABC transporter, SMC_N-terminal domain (RecN)	n.h.
TTHA0797		(S)-2-hydroxy-acid oxidase subunit (GlcD)	0.97	0.49	FAD/NAD containing dehydrogenase	n.h.
TTHB155	Can1	Uncharacterized protein	0.90	0.50	DUF1887, Can1, Cad1, Csm6, Csa3	14.2
TTHA1758		GGDEF domain protein	0.89	0.56	Diguanylate/adenylate cyclase, GGDEF, DNA polymerase, GAF domain	n.h.
TTHA1560		Coenzyme A biosynthesis bifunctional protein CoaBC	0.78	0.53	Flavoprotein (FAD binding)	6.3
TTHA1593		Uncharacterized protein	0.74	0.56	Pilus assembly/response regulator domain, transcription regulator	2.6

F

Thermus thermophilus HB27

Locus	CARF	Name	Score	Fold change (10^x)	Domain analysis (AF2 structural homology to PDB25 by DALI)	CARF DALI Z-score
TT_P0107	Csm6	Uncharacterized protein	17.64	2.13	Csm6, DUF1887, Card1, Can1, ring nuclease	30.8
TT_C0252		Nicotinate phosphoribosyltransferase	11.97	1.82	Various nicotinate proteins, ribose binding	3.2
TT_C0692		Hypothetical conserved protein	9.98	2.34	NYN/PIN domain (mRNA/tRNA nuclease)	6.6
TT_C1426		RNA-splicing ligase RtcB	2.67	0.63	RNA-splicing ligase, translation initiation factor 2	n.h.
TT_C1541		[LysW]-aminoadipate kinase	2.18	0.63	α-aminoadipate carrier, intermediate in lysine biogenesis	3.3
TT_C1607		tRNA uridine 5-carboxymethylaminomethyl modification	1.94	0.91	MnmG, FAD/NAD binding	2.7
TT_C0124		Glycerate dehydrogenase/hydroxypyruvate reductase	1.90	0.49	Glycerate kinase, DUF4147, MOFRL	5.8
TT_P0034		Putative oxidoreductase	1.61	0.93	Tetrameric short-chain dehydrogenase with NAD(P)-binding Rossmann fold	6.7
TT_C0024		Probable methylmalonyl-coA epimerase	1.59	0.54	Glyoxalase 4/5, binds coenzyme A	n.h.
TT_P0110	Can1	Uncharacterized protein	1.50	0.53	Can1, DUF1887, Card1, Csm6, Csa3, ring nuclease	13.5
TT_C0107		Alpha-glucosidase/glycosyl hydrolase	1.20	0.63	Amino acid transporter, alpha amylase	2.7
TT_C0623		Beta-ketoadipyl CoA thiolase	1.01	0.56	Thiolase, binds coenzyme A	n.h.
TT_C1066		Uncharacterized protein	1.00	0.74	FAD-binding/storage/transport dodecin	n.h.
TT_P0017		Uroporphyrinogen-III C-methyltransferase	0.99	0.50	Tetrapyrrole methylase, binds SAM	3.6
TT_C1380		Orotate phosphoribosyltransferase	0.98	0.61	Orotate/adenine phosphoribosyltransferase	2.4
TT_C0159		Riboflavin biosynthesis protein	0.97	0.58	RibF, FAD synthetase, phosphopantetheine adenylyltransferase	5.4
TT_C0635		Hypothetical conserved protein	0.93	0.51	Phosphohistidine phosphatase SixA, phosphoserine phosphatase 1	2.1
TT_C0653		Uncharacterized protein	0.89	0.63	Small helix-turn-helix protein. Transcriptional regulators, sigma factors	n.h.

Figure 2. Identification of cA4-binding proteins in Thermus thermophilus strains. Schematic representations of type III CRISPR-Cas systems in *T. thermophilus* HB8 (**A**) and HB27 (**B**). Volcano plots of proteins from *T. thermophilus* HB8 (**C**) and *T. thermophilus* HB27 (**E**) lysates identified with CARFish, expressed as statistical significance (p-value, y-axis) versus magnitude of change (fold change, x-axis). Enrichment is calculated using the LFQ intensity of proteins identified with the biotin-cA_4 pull down compared to those of the biotin-only control. Enrichment is indicated by the fold change in LFQ intensity. Experiments are performed in biological quadruplicates (n=4) and significance is calculated using a Students t-test and plotted as p-value. Highlighted are proteins with known/predicted CARF-domains (green) and candidate proteins (orange) with fold change > 3 and p-value < 0.05. (**C-F**) Identified proteins with fold change > 3 and p-value < 0.05 sorted according to score ($-\log_{10}$(p-value) * $\log_{10}$(fold change)) in *T. thermophilus* HB8 (**D**) and HB27 (**F**) (full list of identified proteins in **Table S1**). Names are taken from their Uniprot annotation, and domains resulting from DALI homology search are summarized (analysis including proteins outside of top 20 by enrichment score can be found in **Table S2** and DALI results in **Supp. Data S1**). Z-score of pairwise DALI homology search to the CARF domain of *T. thermophilus* HB8 Csm6 is shown when homology was detected (n.h.: no homology).

Identification of putative cOA-binding proteins in archaea

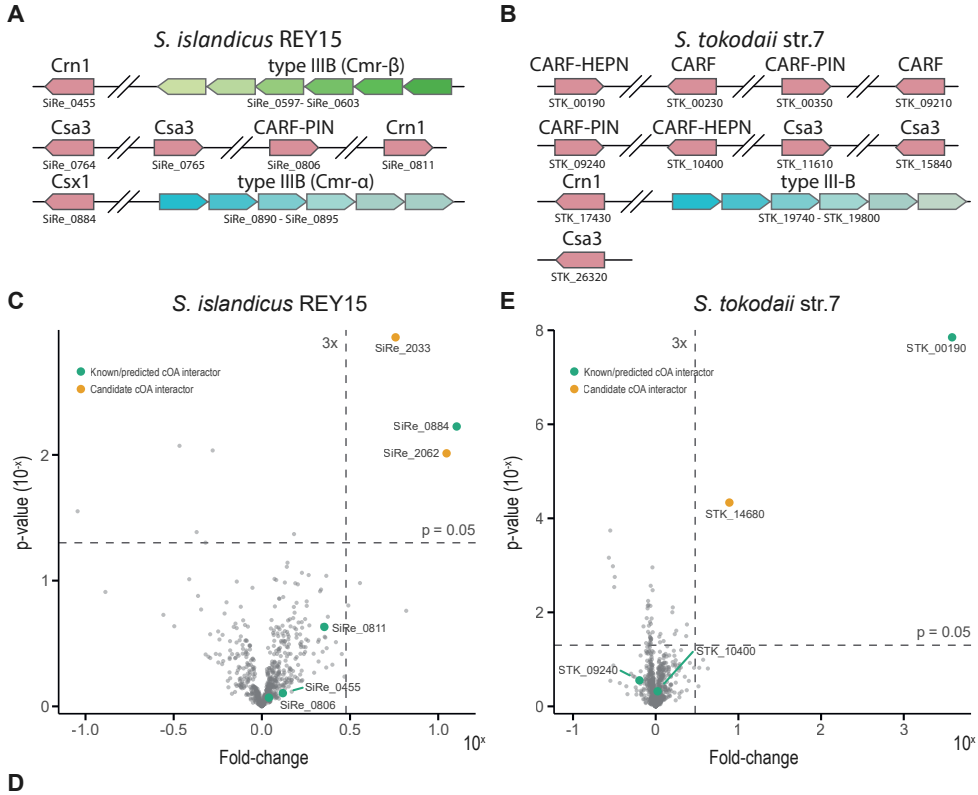
As type III CRISPR-Cas systems are more frequently encoded in archaea than in bacteria, we further explored the potential of CARFish to identify cOA-binding proteins in archaea [21,97,272]. For further validation of CARFish, we selected the thermophilic crenarchaeon *Sulfolobus islandicus* REY15A, which encodes two type III CRISPR-Cas systems including an experimentally characterized cOA-responsive CARF protein (SisCsx1) (**Figure 3A**), and *Sulfolobus tokodaii* str. 7, which encodes one type III CRISPR-Cas system with bioinformatically predicted CARF proteins (**Figure 3B**) [84,106,213,214].

Using CARFish on *S. islandicus*, the protein with the highest enrichment score is CARF domain-containing protein Csx1 (SiRe_0884) (**Figure 3C-D**), which has been demonstrated to bind to cA_4 [214]. However, other proteins that contain CARF domains, known to bind cA_4 or predicted to bind cOA of unknown size, were either not detected (such as the Csa3 transcriptional regulators SiRe_0764 and SiRe_0765) or did not exceed the enrichment cutoff values (SiRe_0811, SiRe_0455 and SiRe_0806) (**Figure 3C**) [215,234,273]. However, other putative cOA-binding proteins were identified that do meet the enrichment cut-off values. This includes a putative signal transduction protein (SiRe_2033; enrichment score: 2.22). SiRe_2033 bears no structural homology to the TtCsm6 CARF domain, and might therefore contain a previously an unknown fold (i.e. not a Rossmann fold) that could facilitate binding cOAs (**Figure 3D, S3C**). SiRe_2033 contains four cystathionine beta synthase (CBS) domains, which are known to bind adenosine derivatives (NAD, SAM, AMP etc.), metal ions

and nucleic acids [274]. Another putative cOA-binding protein that was enriched is an ATP-cone containing protein (SiRe_2062), which has no homology to CARF domains or Rossmann folds (**Figure 3D, S3D**). Other proteins in the ATP-cone family of SiRe_2062 bind ATP, and as such they allosterically regulate the functionality of associated proteins/domains i.e. by allowing oligomerization or DNA binding [275–277].

Using CARFish on lysates of *Sulfolobus tokodaii* str. 7, (**Figure 3B**) we could identify various of the predicted CARF domain-containing proteins (**Figure 3E**) [278]. This includes the Csx1 effector protein STK_00190 (enrichment score 28.14), which was bioinformatically annotated, but for which cOA-binding was not yet experimentally demonstrated (**Figure 3E-F**). A CARF-HEPN protein (STK_10400) and a CARF-PIN protein (STK_09240) were not significantly enriched (enrichment scores: 0.01 and -0.11, respectively). Other predicted cOA-binding proteins, including the ring nuclease (Crn1) and Csa3 transcriptional regulators were not detected in our experiment (**Figure 3B, E-F, Table S1**). Beyond the predicted cOA-binding proteins, a putative MIP18 family-like protein (STK_14680) was enriched (enrichment score: 3.86). STK_14680 also contains a Mrp/NBP35 ATP-binding domain, which as well as the MIP18 family-like proteins is involved in FeS cluster assembly [279,280]. Its predicted structure shares structural homology with magnetosome proteins and a type III secretion system membrane ring protein, but it bears no structural homology to the TtCsm6 CARF domain (**Figure 3F, S3E**).

Combined, these results illustrate that CARFish can be used to identify known and predict novel putative archaeal ATP- and/or cOA-binding proteins.



D

Sulfolobus islandicus REY15

Locus	CARF	Name	Score	Fold change (10^x)	Domain analysis (AF2 structural homology to PDB25 by DALI)	CARF DALI Z-score
SiRe_0884	Csx1	CRISPR-associated (Cas) DxTHG family	2.46	1.10	Csx1, DUF1887, Card1, Csa3, Csm6	8.1
SiRe_2033		Putative signal transduction protein with CBS domains	2.22	0.76	CBS domain, adenylate-binding domain, AMPK, binds AMP/ADP/ATP	n.h.
SiRe_2062		ATP-cone domain-containing protein	2.11	1.05	Small ATP-cone only protein, Cyclin, Sigma70	n.h.
SiRe_0811	Crn1	CRISPR-associated protein, APE2256 family	0.22	0.35	Crn1, SSO1393-like, DUF1887, Card1, Csa3, Csm6	10.6
SiRe_0455	Crn1	Uncharacterized protein	0.01	0.12	Crn1, DUF1887, Csm6, Card1, Csa3	13
SiRe_0806	CARF-PIN	PIN domain-containing protein	0.00	0.04	Card1, DUF1887, PIN, Can1, Crn1	9.2

F

Sulfolobus tokodaii str.7

Locus	CARF	Name	Score	Fold change (10^x)	Domain analysis (AF2 structural homology to PDB25 by DALI)	CARF DALI Z-score
STK_00190	CARF-HEPN	CRISPR-associated protein	28.14	3.58	Csx1, DUF1887, Csa3, Csm6, Card1	10.6
STK_14680		MIP18 family-like domain-containing protein	3.86	0.89	FeS-cluster, MAMB/MAMM, cation diffusion facilitator, zinc transporter	n.h.
STK_10400	CARF-HEPN	CRISPR-associated protein	0.01	0.02	Csx1, DUF1887, Crn1, Csa3, Card1	7.5
STK_09240	CARF-PIN	Uncharacterized protein	-0.11	-0.20	DUF1887, Card1, Can1, Csm6, LuxE, PIN	10

Figure 3. Identification of cA4-binding proteins in Sulfolobus species. Schematic representation of type III CRISPR systems in *S. islandicus* REY15A (**A**) or *S. tokodaii* str. 7 (**B**). Proteins from *S. islandicus* REY15A (**C**) or *S. tokodaii* str. 7 (**E**) lysate enriched in an immobilized biotin-cA_4 pull down compared to a biotin control. Enrichment is indicated by the fold change in LFQ intensity. Experiments are performed in biological quadruplicates (n=4) and significance is calculated using a Students t-test and plotted as p-value. Highlighted are proteins with known/predicted CARF-domains (green) and candidate proteins (orange) with fold change > 3 and p-value < 0.05. (C-F) Identified proteins with fold change > 3 and p-value < 0.05 sorted according to score ($-\log_{10}$(p-value) * $\log_{10}$(fold change)) in *S. islandicus*

REY15A (**D**) and *S. tokodaii* str. 7 (**F**) (full list of identified proteins in **Table S1**). Names are taken from their Uniprot annotation, and domains resulting from DALI homology search are summarized (full analysis in **Table S2** and DALI results in **Supp. Data 1**). Z-score of pairwise DALI homology search to the CARF domain of *T. thermophilus* HB8 Csm6 is shown when homology was detected (n.h.: no homology).

Discussion

CRISPR-Cas type III systems have elaborate ways to protect the population from invading nucleic acids, often through a signalling pathway mediated by cOA molecules [75]. Possibly, we have only scratched the surface of the complete set of proteins that participate in this signalling network. While *in silico* methods allow the identification of certain CARF domain- and SAVED domain-containing proteins based on homology, other cOA-binding proteins that have lower structural or sequence homology and/or rely on other domains for cOA binding, might be overlooked. Here, we present CARFish, an experimental approach for identification of cOA-interacting proteins from cell lysates in an unbiased bottom-up approach, with the goal to uncover novel cOA-responsive proteins that participate in CRISPR-Cas type III immune signalling pathways.

CARFish successfully identified previously characterized CARF domain-containing proteins in cell lysates of bacteria and archaea (**Figure 2, 3A,C-D**). Furthermore, CARFish experimentally confirmed a bioinformatically-predicted CARF effector in *S. tokodaii* (**Figure 3B, E-F**). Finally, CARFish resulted in the identification of various novel putative cA4 binding-proteins. The functional relevance of cA_4 binding by these proteins, and whether they contribute to CRISPR type III defence, remains to be established.

For both Sulfolobus strains, the ring nucleases (Crn1) was not significantly enriched using CARFish. A potential explanation could be that they degraded the immobilized cA_4 on the beads (due to RING nuclease activity associated with the CARF domain) and were subsequently washed from the column prior to elution (**Figure 3C, E**) [211,214,215]. Furthermore, the CARF-PIN and Csa3 proteins (SiRe_0764, SiRe_0765, SiRe_0806, STK_00350, STK_09240, STK_11610, STK_15840 and STK_26320) were also not enriched or detected at all (**Figure 3C, E**). It was previously shown that a Csa3 protein from *S. islandicus* (SiRe_0765) did not interact strongly with cA4, which could explain our findings [281]. In the same study it was reported that CARF-PIN from *S. tokodaii* (STK_0035) has a relatively low affinity for cA_4, and that cA_4 does not trigger RNase activity *in vitro*. A similar scenario might apply to the CARF-PIN protein in *S. islandicus*. It may be that other cOA variants are the natural ligand of these proteins or that the biotinylation of cA_4 affect the binding

significantly.

Using AlphaFold structural predictions and DALI structural homology searches, we determined that 17 out of 28 candidates identified in our CARFishing experiments are structurally homologous to CARF domains, albeit to varying degrees. This homology could be due to the abundance of Rossmann folds in all kinds of proteins that can bind nicotinate/NAD, FAD, ATP/ADP, Co-enzyme A, tRNA and nucleic acids (**Figure 2-3**) [282,283]. Whether or not these interactions are specific for (part of) cA_4, it is interesting to consider that many cA_4 molecules might bind other Rossmann fold containing proteins thereby competing with cA_4-CARF binding, which can possibly result in false negatives. Alternatively, biotin-cA_4 molecules may have been (partly) degraded by e.g. ring nucleases during the pulldown but still bound to biotin, leaving the adenosine containing molecules accessible for binding to adenosine binding proteins.

For example, the described experiments resulted in the enrichment of proteins that, given their biological function, are more likely to bind AMP or molecules that closely resemble AMP, either as substrate (5P-ribose, RNA, DNA), or as co-factor (ATP, NAD/FAD, Coenzyme A). The twelve proteins listed in **Figure 2D**, include three established CARF proteins, and nine proteins that are homologs of proteins that interact with AMP-like molecules, at least in several cases with functions that appear not related to defence strategies. It is important to confirm biochemically that candidate proteins that bind cOAs do so for a specific reason, be it to sequester or degrade cOAs, or activate auxiliary domains/proteins. To refine the list of candidate proteins with a functional cOA interaction the fraction of true positives could be improved. This could be accomplished through, for example including another control (biotin-AMP, next to biotin-only & biotin-cOA) that does bind false-positives, but binds less strongly to cA_4, and as such can be used to discard false-positives generated through AMP-like molecules. In addition, eluting with cA_4 instead of biotin could decrease biotin and carbon-linker specific interactions.

The CARFish method allows high throughput screening of a complex mixture for cA_4-binders without the need for tedious purification of individual candidates. Besides from the observation that CARFish possibly results in the identification of false negatives, another challenge is that CARFish requires an accurate proteome database for accurate identification of proteins of interest, which might not be available for all strains of interest. This could be overcome by DNA extraction and whole genome sequencing of such strains. One could envision also fishing in mixed cultures (for non-isolatable organisms), as long as a (metagenomic) database for protein identification exists or can be generated. While originally designed to identify CRISPR associated proteins, CARFish could also be used for the identification of other do-

mains that recognize cOA molecules, such as effector from CBASS immune systems [226,242,247,284,285]. The current CARFish method utilizes biotin-cA_4 to identify cA_4-binding proteins, while it is known that many sensory domains of CBASS and CRISPR systems bind cA_x of other sizes or variations thereof. We therefore envision great potential for CARFish when combined with a diverse size repertoire of biotin-cOA molecules, but also other molecules to map signalling pathways beyond CRISPR.

Competing Interests

J.A.S. is a founder and shareholder of Scope Biosciences. R.H.J.S. is a shareholder and member of the scientific board of Scope Biosciences.

Acknowledgements

This work was supported by a VIDI grant (VI.Vidi.203.074) from The Netherlands Organization for Scientific Research (NWO) to R.H.J.S and grants from the Netherlands Organization for Scientific Research (NWO, 016. Veni.192.072) and the European Research Council (ERC, ERC-2020-STG 948783) to D.C.S.

Author contributions

J.A.S., J.v.d.O. and R.H.J.S. conceived the study. J.A.S., P.B., S.B., J.v.d.O., D.C.S., and R.H.J.S. designed the experiments. J.A.S., P.B., J.G. and S.B. performed experiments and analyses. J.v.d.O., D.C.S. and R.H.J.S. supervised the project. J.A.S. and P.B. wrote the manuscript with input from J.v.d.O., D.C.S. and R.H.J.S.. All authors have read and approved the manuscript.

Materials & Methods

Lysate preparation

Thermus thermophilus cultures were grown in a medium containing 8 g/L bacto-tryptone, 4 g/L yeast extract, and 3 g/L NaCl at a final pH of 7.5, overnight at 65ºC and 200 rpm. The Sulfolobus cultures were grown in Brock medium overnight at 75ºC and 200 rpm. The cells were harvested by centrifuging for 15 minutes at 6,000xg. Cells were then lysed by sonication (50% amplitude, 1 second on, 2 seconds off) for 20 minutes in 10 mL lysis buffer per liter culture (50 mM HEPES pH 8, 100 mM NaCl, cOmplete Mini Protease Inhibitor Cocktail Tablet, 1 µL/mL DNaseI (ThermoFisher), 2 µL/mL RNaseA (ThermoFisher), 1 mg/mL lysozyme and 1 mM DTT). The lysate was centrifuged for 45 minutes at 30,000xg and subsequently passed through a 0.45 µm filter.

CARFish pull-down method

A column volume (CV) of 200 µL Streptavidin Resin (GenScript) was added to each Poly-Prep® Chromatography Column (Bio-Rad) and equilibrated with 2 CVs wash buffer (50 mM HEPES pH 8, 100 mM NaCl). Approximately 5 mL of lysate was incubated for 10 minutes at 65ºC with 3/5 CV cA_4-49-biotin 100 µM (Biolog.de) or 3/5 CV wash buffer (50 mM HEPES pH 8, 100 mM NaCl) for the sample and the negative control, respectively. The lysate was added to their respective columns with the bottom closed, a lid was added, and the columns were subsequently rotated at room temperature for one hour. The flowthrough was collected for analysis and the columns were washed with 5 CVs wash buffer. Each wash was collected for further analysis. Finally, the columns were eluted with five times 0.5 CV elution buffer (50 mM HEPES pH 8, 100 mM NaCl, 2.5 mM biotin) and collected separately. For each sample and control, the experiment was performed in biological quadruplicates.

Mass spectrometry sample preparation

The third elution fraction (usually with the most protein across replicates) was loaded on a 10% SDS-PAGE gel (Bio-Rad) and run at 200V. The gels were stained with Coomassie Brilliant Blue and destained with dH2O. To reduce cysteines, the gel was incubated in 25 mL 15 mM DTT and 50 mM ammonium bicarbonate pH 8 for 1 hour under gentle agitation. Next, it was washed with ultrapure water and subsequently incubated in 25 mL 100 mM Tris-HCl pH 8 and 20 mM acrylamide for 30 minutes in the dark for alkylation. The gel was then washed thoroughly with ultrapure water. The bands in the gel were cut out with a scalpel in pieces of ~1 mm2 and transferred to a 0.5 mL low-binding microcentrifuge tube (Eppendorf LoBind). The tubes were frozen in dry ice and thawed repeatedly. 50 µL of 5 ng/µL trypsin in 50mM ammonium bicarbonate was added to the gel and shaken overnight at room temperature. Then the pH was adjusted to 2-4 by adding 10% trifluoroacetic acid. To

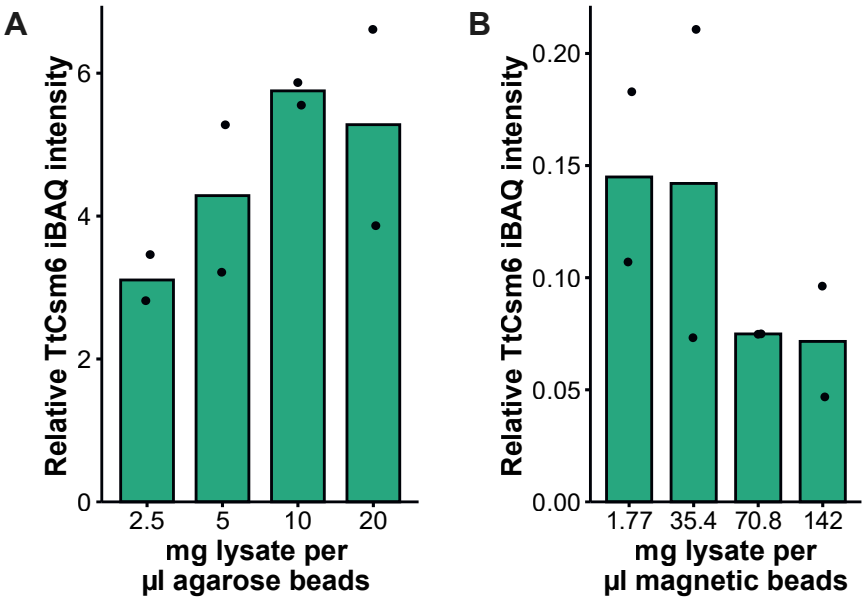
purify the peptides columns were made with 3M™ EmporeTM C18 Extraction Disk (ThermoFisher), washed with 200 µL MeOH, and 4 µL of a 50% slurry of LiChroprep® RP-18 (SigmaAldrich) was added (2µl column volume/CV). The columns were washed with 50 CVs of MeOH and equilibrated with 50 CVs 0.1% formic acid. The supernatant of the samples was added to the columns and the gel pieces were washed once more with 100 µL 0.1% formic acid which was also added to the columns. The columns were washed with 50 CVs 0.1% formic acid and subsequently eluted with 25 CVs 50% acetonitrile and 0.05% formic acid into 0.5 mL low-binding microcentrifuge tubes. The samples were then concentrated in a vacuum concentrator until the volume reached below 15 µL. The final volume was then adjusted to 50 µL with 0.1% formic acid.

nLC-MS/MS analysis

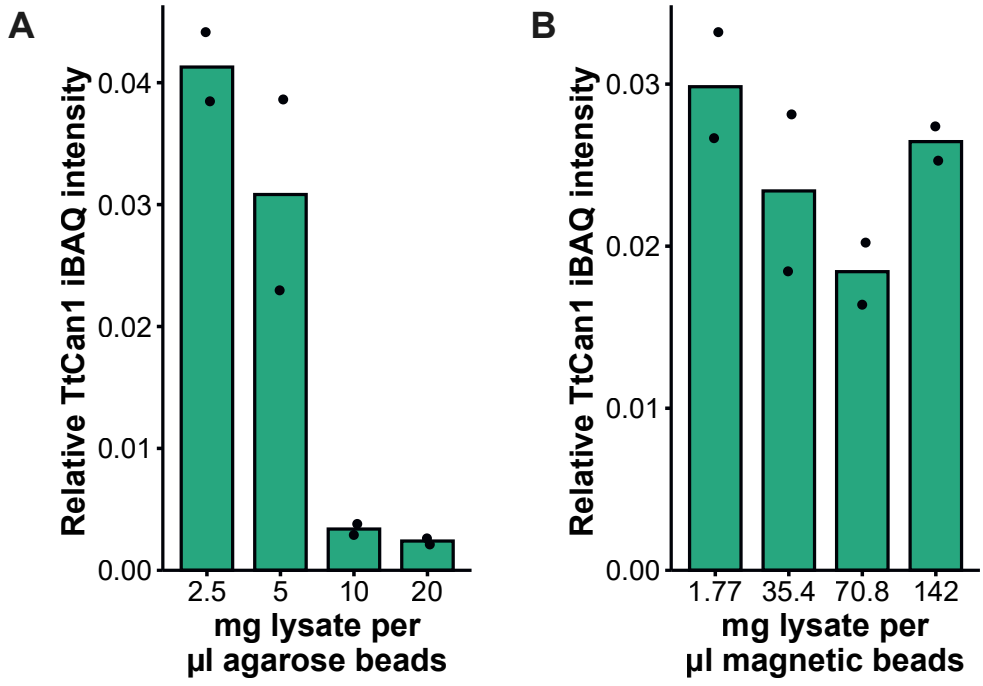
Five µL of peptide samples were loaded directly onto a 0.10 * 250 mm ReproSil-Pur 120 C18-AQ 1.9 µm beads analytical column (prepared in-house) at a constant pressure of 825 bar (flow rate of circa 600 nL/min) with 1 mL/L HCOOH in water and eluted at a flow of 0.5 µL/min with a 50 min linear gradient from 9% to 34% acetonitril in water with 1 mL/L formic acid with a Thermo EASY nanoLC1000. An electrospray potential of 3.5 kV was applied directly to the eluent via a stainless steel needle fitted into the waste line of a micro cross that was connected between the nLC and the analytical column. Full scan positive mode FTMS spectra were measured between m/z 380 and 1400 on a Exploris 480 (Thermo Fisher Scientific, Bremen, Germany) in the Orbitrap at resolution 60000. MS and MSMS AGC targets were set to 300%, 100% respectively or maximum ion injection times of 50 ms (MS) and 30 ms (MSMS) were used. HCD fragmented (Isolation width 1.2 m/z, 28% normalized collision energy) MSMS scans in a cycle time of 1.1 s of the 25 most abundant 2-5+ charged peaks in the MS scan were recorded in data dependent mode (Resolution 15000, threshold 2e4, 15 s exclusion duration for the selected m/z +/- 10 ppm) as previously described by [286]. The LC-MS data analysis (false discovery rates were set to 0.01 on peptide and protein levels) and additional result filtering (minimally 2 peptides are necessary for protein identification of which at least one is unique and at least one is unmodified) were performed as previously described [287,288]. The nLC-MS/MS system quality was checked with PTXQC using MaxQuant result files [289]. The analysis of the relative abundance of proteins was done by comparing their normalized label-free quantification (LFQ) intensities [290]. A Students t-test was performed on LFQ-values between sample and control. In addition, a combined score was calculated: $-\log_{10} p \text{ value} \times \log_{10} \text{fold change} = \text{score}$. A list of identified proteins is presented in **Table S1**.

In silico analysis of protein candidates

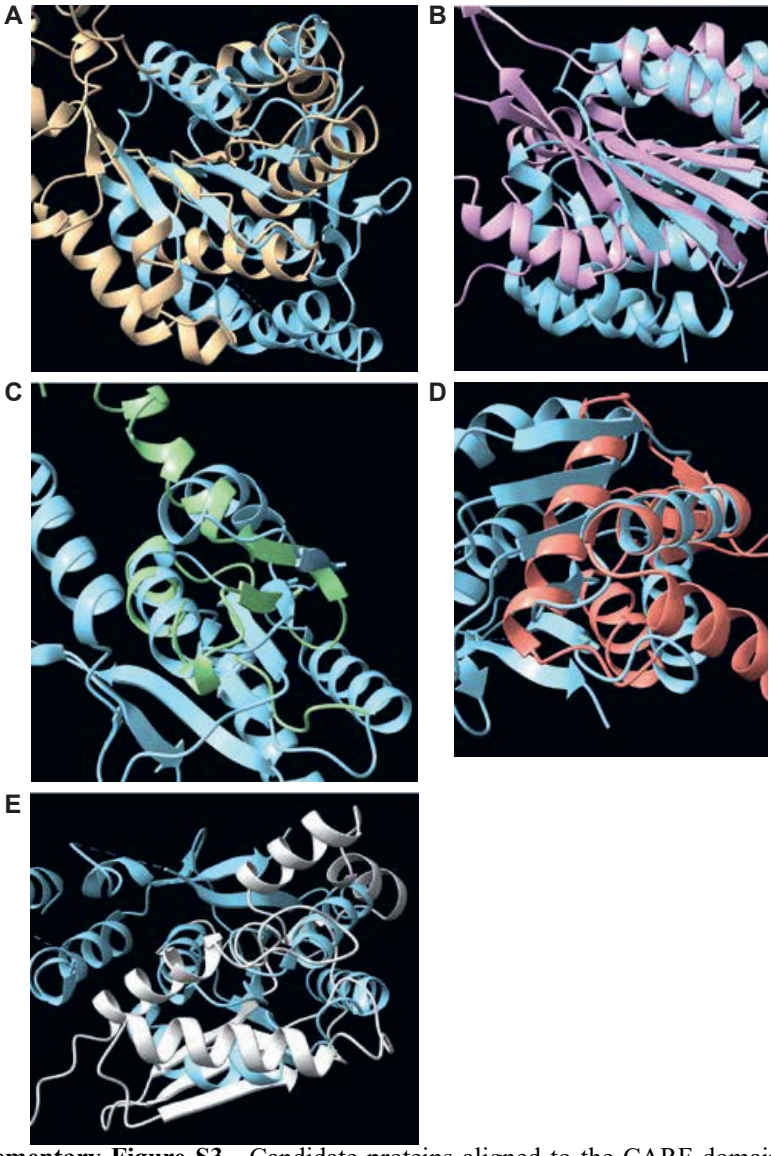
Each protein candidate with $p < 0.05$ and fold change > 3 was analysed further in silico. Proteins in the top 20 sorted by score and fold change were included in the analysis. For each candidate, a Alphafold 2 monomer was retrieved (AlphaFold Protein Structure Database), and in case an experimental structure was available this was retrieved as well (PDBe) [291]. Candidate structures were submitted to DALI structural homology search against a selection of pdb structures that are less than 25 % identical in sequence (PDB25) [292]. The PFAM domain composition of the top 20 most homologous structures were analysed if the DALI Z-score > 5.0. If certain domains of candidate proteins did not align with the identified homologous structures while other domains did, proteins were truncated to unaligned domains and resubmitted for DALI homology search. In addition, a DALI pairwise structural homology search with candidate proteins and the CARF domain of TtCsm6 (PDB: 5FSH, 1-185aa) was performed. The complete analysis is summarized in **Table S2**, and the DALI search results are in **Supp. Data S1**.



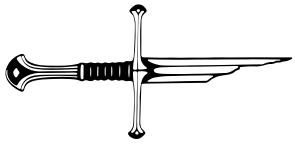
Supplementary Figure S1. Relative Csm6 iBAQ intensity in duplicate *T. thermophilus* HB27 pulldowns. iBAQ intensity of TtCsm6 relative to the average iBAQ intensity of top 50 hits sorted by average protein iBAQ intensity of all samples.

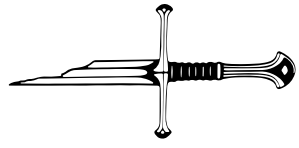


Supplementary Figure S2. Relative Can1 iBAQ intensity in duplicate *T. thermophilus* HB27 pulldowns. iBAQ intensity of TtCan1 relative to the average iBAQ intensity of top 50 hits sorted by average protein iBAQ intensity of all samples.



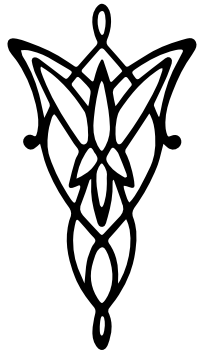
Supplementary Figure S3. Candidate proteins aligned to the CARF domain of TtCsm6 (PDB: 5FSH, 1-185aa, sky blue). (**A**) AF2 model of NAPRT of HB27 (TT_C0252, beige), RMSD is 4.1 Å over 61 α-carbon pairs. (**B**) AF2 model of NYN/PIN protein of HB27 (TT_C0692, lavender magenta), RMSD is 3.4 Å over 100 α-carbon pairs. (**C**) AF2 model of the CBS-domain containing protein SiRe_2033 of *S. islandicus* (light green), no structural homology detected. (**D**) AF2 model of the ATP-cone domain-containing protein SiRe_2062 of *S. islandicus* (salmon), no structural homology detected. (**E**) AF2 model of the MIP18 family-like protein STK_14680) of *S. tokodaii* (Light Gray), no structural homology detected. Given RMSD was determined using DALI.





Chapter 5

Nanopores reveal the stoichiometry of single oligo-adenylates produced by type III CRISPR-Cas



Manuscript in submission:
David Fuentenebro-Navas, Jurre A. Steens, Carlos De Lannoy, Ben Noordijk, Dick de Ridder, Raymond HJ Staals, and Sonja Schmid.

Abstract

Cyclic oligoadenylates (cOAs) are small second messenger molecules produced by the type III CRISPR-Cas system as part of the prokaryotic immune response. The role of cOAs is to allosterically activate downstream effector proteins that induce dormancy or cell death, and thus abort viral spread through the population. Interestingly, different type III systems have been reported to utilize different cOA stoichiometries (with 3 to 6 adenylate monophosphates). However, so far, their characterization has only been possible in bulk and with sophisticated equipment, while a portable assay with single-molecule resolution has been lacking. Here, we demonstrate the label-free detection of single cOA molecules using a simple protein nanopore assay. It sensitively identifies the stoichiometry of individual cOA molecules and their mixtures from synthetic and enzymatic origin. To achieve this, we trained a convolutional neural network (CNN) and validated it with a series of experiments on mono- and polydisperse cOA samples. Ultimately, we determined the stoichiometric composition of cOAs produced enzymatically by the CRISPR type III-A and III-B variants of *Thermus thermophilus*. Interestingly, both variants produce cOAs of nearly identical composition, and we discuss the biological implications of this finding. The presented nanopore-CNN workflow with single-cOA resolution can be adapted to many other signalling molecules (including eukaryotic ones), and it may be integrated into portable handheld devices with potential point-of-care applications.

Introduction

Second messenger molecules and other small metabolites serve a wide variety of essential signalling, activation, and regulation purposes in the biological cell, such as spatial and temporal regulation of cellular responses, signal transduction between cell membrane and nucleus, and neuro-transmission [293]. However, due to their small size, they are particularly difficult to detect, quantify, and study. In the present work, we focus on cyclic oligoadenylate molecules (cOAs) which play a crucial role in the type III CRISPR-Cas adaptive immune system of prokaryotes [73,83]. This immune system evolved among bacteria and archaea to combat invading plasmids, bacteriophages, and other mobile genetic elements (MGEs) [25,28,294]. It works by storing short DNA sequences of the encountered MGEs in an array of clustered, regularly interspaced short palindromic repeats, the CRISPR array [295]. During subsequent infections, the CRISPR array is transcribed, processed into short CRISPR RNAs (crRNAs, or guide RNAs), and incorporated into a single CRISPR-associated (Cas) protein or into multi-subunit protein complexes, with different modes of action (classified in several classes and types) [21,295]. The crRNA-protein complexes then bind and degrade invading complementary MGEs.

The cOAs are second messengers produced by the type III CRISPR-Cas complex (**Figure 1A**). This ribonucleoprotein complex is endowed with three catalytic activities: sequence-specific RNAase activity by the Cas7 subunits, non-specific ssDNA cleavage by the HD domain of Cas10, and the ATP-cyclase activity of the palm domain of Cas10 responsible for generating the cOAs [73,210]. The stoichiometry of these cOA molecules can vary between different hosts and CRISPR type III subtypes, but they typically contain three to six adenosine monophosphates (AMP) in a ring structure (cA_3–cA_6, **Figure 1A**) [83,207,214]. The cOAs of the type III CRISPR response were found to activate particular families of downstream effector proteins containing an appropriate sensory domain, such as CARF (CRISPR-associated Rossman fold) and SAVED (second messenger oligonucleotide or dinucleotide synthetase (SMODS)-associated and fused to various effector domains). The cOA-binding domain of these proteins is often fused to other different domains with a wide variety of catalytic activities, such as (ribo-)nucleases and proteases [84,265]. Activation of these enzymes by the cOAs results in a 'secondary line of defence' by type III systems, leading to the degradation of essential host biomolecules that induce either cellular dormancy or even cell death, preventing viral (or other MGE) propagation through the population [206,247].

As the diverse catalytic functions that are activated by cOAs are only just being unveiled, several questions remain unsolved. In particular: do the Cas10 homologues all produce monodisperse cOAs, or rather a polydisperse distribution? In addition, it has been reported that several distinct type III CRISPR systems (including different CARF and SAVED proteins) can be found within one host, despite using the same CRISPR array [81,82,89]. For example, in *Thermus thermophilus* HB8, considered here, the genome encodes two different CRISPR-Cas type III systems, termed III-A and III-B. This is remarkable, and raises the question: what is the evolutionary benefit of having multiple distinct systems for a given task? One possible reason for the co-occurrence of multiple type III subtypes in one host is that they may each produce a unique subset of cOA stoichiometries, thus providing a regulatory benefit by activating distinct effector proteins in a fine-tuned, cOA-stoichiometry-dependent way. To test this hypothesis and elucidate cOA-dependent regulation mechanisms, a simple and rapid method to directly detect small amounts of enzymatically produced cOAs, and even quantify their stoichiometric composition with single-molecule resolution would be instrumental.

Here, we demonstrate the label-free detection of single cOA molecules and their stoichiometries using nanopore experiments, where a pore protein embedded in a free-standing lipid bilayer acts as a sensor for single cOA molecules (**Figure 1B**). An applied positive voltage drives the negatively charged cOA molecules through the nanopore by electrophoresis. While translocating, the cOA partially blocks the ionic through-pore current, resulting in a characteristic current blockade signal (cf. resistive pulse sensing) [296]. Nanopores are by definition single-molecule sensors due to their small size, which we chose here to be comparable to the cOA molecule. More generally, nanopore technology is best known for commercial devices offering inexpensive and portable DNA sequencing with long-reads that are revolutionizing the life sciences [297,298]. The sensitivity of protein nanopores is remarkable: even single enantiomers (chiral variants) in small-molecular racemates can be distinguished [299,300]. In nanopore signal processing, neural networks play a crucial role [301–303]. An important advantage of neural networks is their ability to implicitly extract features from the data provided during training, in contrast to other machine learning approaches that require the manual definition of informative features (e.g. hidden Markov models). Neural networks are, therefore, able to learn and combine more subtle characteristics, such as the shapes of blockade events and their current fluctuations [304,305]. For nanopore signal processing, this enabled improved quantitative analyses [302]. In this study, we use a convolutional neural network (CNN) to quantitatively infer the stoichiometric composition of cOA mixtures – including samples from enzymatic origin.

Our results present the first detection of cOAs with single-molecule resolution, in a sensitive nanopore assay. We compare a range of synthetic cOAs with known monodisperse stoichiometries of three to six adenosine monophosphate (AMP) monomers (subsequently: cA_3 to cA_6). Using this calibration data as a training set for our CNN, we turn to cOA mixtures of know polydisperse composition, and validate the capacity of the CNN to quantify the correct ratio of stoichiometries involved. We then use this new label-free cOA identification pipeline to study enzymatic cOA samples. Specifically, we identify the stoichiometric composition of cOAs produced by different type III-A and III-B complexes, and compare them with their CARF activation capability. Lastly, we discuss the implications of our results on the hypothesized evolutionary benefit of multiple type III CRISPR systems (possibly producing varied cOA stoichiometries) in one prokaryotic host. Altogether, our label-free cOA identification assay has proven its capacity to identify small second messengers with single-molecule resolution – even from enzymatic mixtures – and also their stoichiometry. This makes it a powerful tool for metabolic research of CRISPR-Cas and beyond, where fast, label-free, and quantitative readouts matter.

Results and Discussion

Detecting single cyclic oligo-adenylates with protein nanopores

Figure 1 shows the capacity of the alpha-hemolysin (α-HL) pore protein to detect single cOA molecules. In contrast to the clean α-HL baseline measured in the absence of cOA molecules (**Figure 1C**), characteristic resistive pulses appear after cOA addition to the 'cis' side of the pore (see definition in **Figure 1B**, and data in **Figure 1D**), each representing a single cOA translocation (**Figure 1E**). As expected, the cOA event rate is concentration dependent (**Figure S1**), and the event duration decreases with increasing voltage (**Figure S2, Table S1**), indicating that cis-to-trans translocations take place (rather than cis-to-cis escapes). This behaviour is expected for substantially charged molecules, like the poly-nucleic cOAs, for which the electrophoretic driving force dominates over other contributions (e.g. electro-osmosis) [306]. The current blockade ΔI and event duration Δt depend on the specific analyte and measurement conditions. For cOAs with a stoichiometry of six AMP subunits (cA6), we find mean values and standard deviations of $\Delta I = 400 \pm 30$ pA and $\Delta t = 315 \pm 26$ μs, measured at +200mV in 3M KCl (**Figure 1D, E, Table S1**).

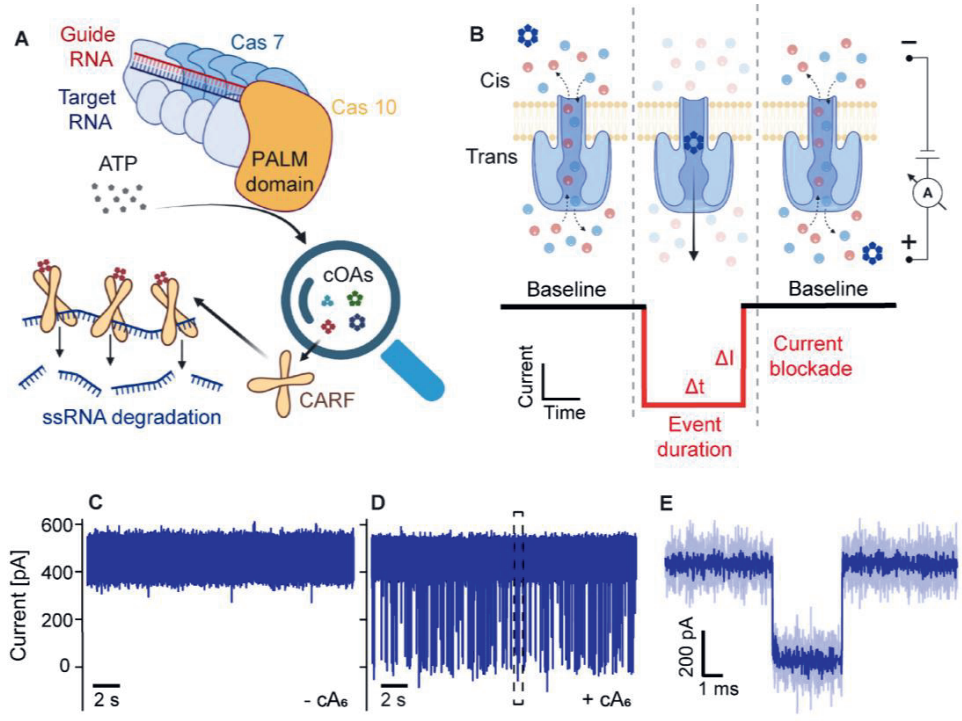


Figure 1: Nanopore detection of single CRISPR second messenger molecules. (**A**) Schematic depiction of a type III CRISPR-Cas complex. Target RNA binding activates the cyclase activity of the palm domain of Cas10 for cOA synthesis from ATP molecules. The cOA molecules activate CARF proteins, such as the non-specific ribonuclease Csx1. (**B**) Schematic arrangement of the α-hemolysin (α-HL) nanopore experiment. Upon voltage application, an ionic current flows through the pore. A cOA translocation, driven by electrophoresis, is observed as a resistive pulse with current blockade ΔI and duration Δt. We refer to the depicted pore arrangement as a 'trans-inserted' α-HL, where the vestibule is on the trans side. (**C-E**) α-HL recordings at +200 mV in a 3 M KCl buffer, obtained as illustrated in (**B**): (**C**) Baseline trace measured without an analyte. (**D**) α-HL recording after addition of 10 μM cA_6 to the cis compartment. (**E**) A single cA_6 translocation event extracted from the trace in (**D**)), as indicated by the dashed rectangle. The dark blue line represents filtered data, the light blue line represents measured raw data (see Methods).

For the single-molecule resolution of these small cOA molecules, the choice of the nanopore and precise experimental conditions are crucial, as several factors affect the sensing performance. For example, while the MspA nanopore with its much pointier constriction site provides better spatial resolution than α-HL in DNA sequencing applications [307,308], it barely resolved cOA translocations (**Figure S3**), which limits its utility for our application in small-molecule sensing. In contrast, the α-HL with its long narrow stem yields single-molecule observations that are well resolved in time, even for these small cOA molecules. In addition, we found that the translocations through a trans-inserted α-HL (illustrated in **Figure 1B**) were more uniform than the translocations with a cis-inserted pore (**Figure S4**). Similar findings have previously been attributed to less heterogeneous excursions and interactions in the wide pore lumen of α-HL. Overall, a trans-inserted α-HL in 3M KCl and +200mV bias provided an optimal signal-to-noise ratio to resolve single cA_6 molecules.

Nanopore event durations reveal the stoichiometry of cOA molecules

Encouraged by the successful label-free single-molecule detection of cA_6, we probed the sensitivity of our assay to detect even smaller cOAs: cA_3, cA_4, and cA_5 depicted in **Figure 2A-D**. Biologically, this is highly relevant, since type III CRISPR-Cas complexes have been reported to produce cOAs with varying stoichiometries [75]. Experimentally, however, it has been impossible up to now to resolve these differences among single cOAs, given their small, cyclized structure (**Figure S5** shows a 3D representation). Gratifyingly, with the presented nanopore assay, all cOAs down to cA_3 can be resolved at +120mV, as evident from the current traces and zoomed-in events shown in **Figure 2A-D**. The measured event durations scale with the cOA size, as the larger cOAs have a reduced translocation probability through α-HL's 1.4 nm constriction site (**Figure 2E, F, S6, S7, Table S1**). Interestingly, however, cA_3 to cA_6 all cause similar current blockades, i.e. the larger cOAs do not measurably block more ionic current (**Figure 2G, H, S6, S7**), likely due to local positioning inside the nanopore and 3D folding effects (cf. **Figure S5**). To distinguish cOA stoichiometries, we therefore focus on the distinct event durations. Their clear trend can serve as a proxy for cOA identification, while it is also evident that cA_3 and cA_4 cannot yet be distinguished based on the overlapping event durations alone (**Figure 2E**, bottom panel). We therefore turned to neural networks, which have the ability to recognize more subtle patterns in the nanopore translocation events and allow the identification of cOA stoichiometries at the single-event level.

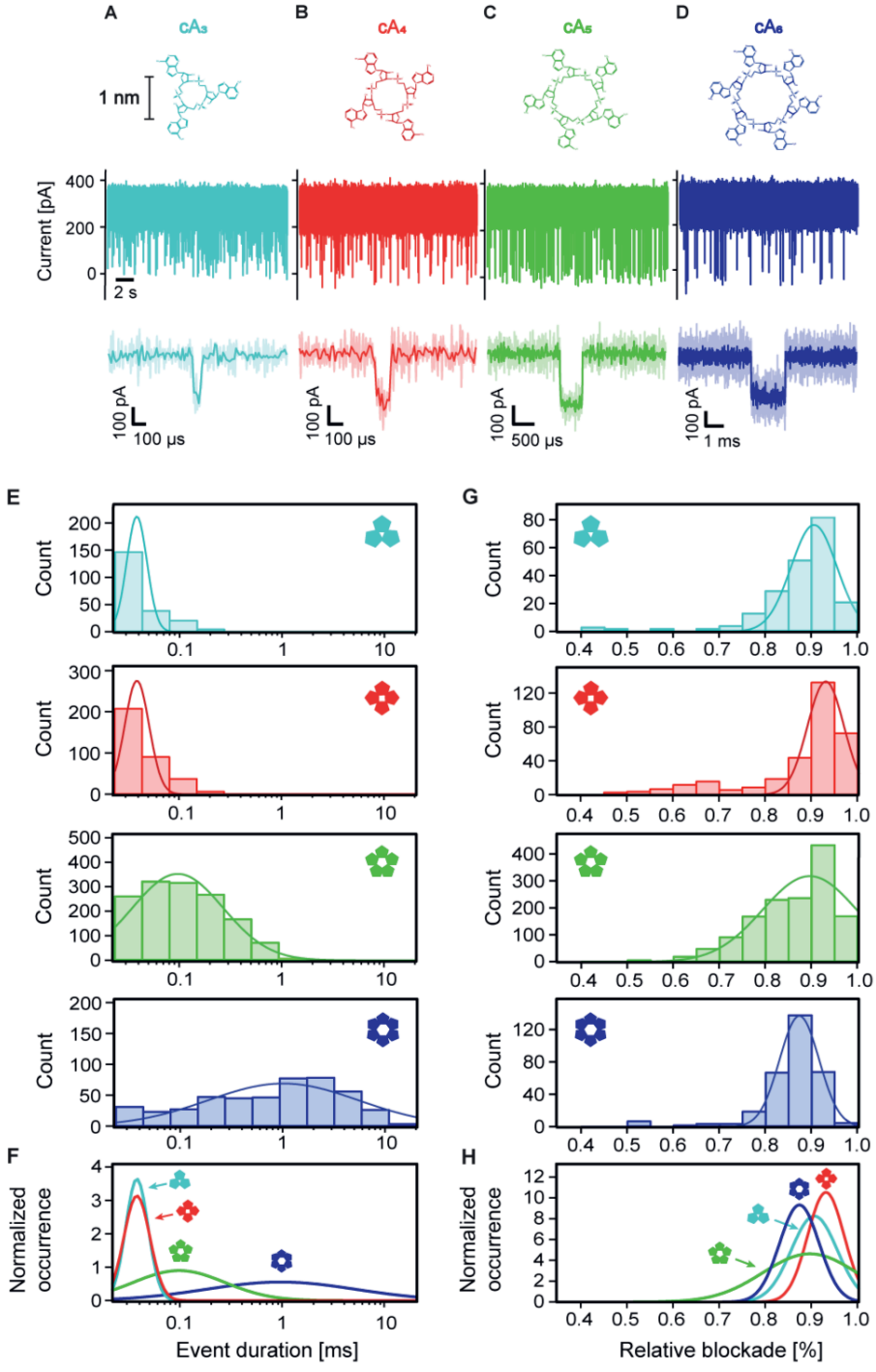

A
cA3
1 nm
B
cA4
C
cA5
D
cA6
Current [pA]
400
200
0
2 s
100 pA
100 µs
100 pA
100 µs
100 pA
500 µs
100 pA
1 ms
E
G
Count
F
H
Normalized occurrence
Event duration [ms]
Relative blockade [%]

Figure 2: Nanopore recordings reveal the stoichiometry of cyclic oligo-adenylates. (**A-D**) Nanopore recordings showing the label-free detection of single cOA molecules with different stoichiometries: cA_3, cA_4, cA_5, cA_6, respectively. Top: molecular structure of the cOAs with 3-6 AMP monomers as indicated (same scale bar for (**A-D**)). Centre: nanopore current trace with short blockade events indicating cOA translocations, obtained at +120 mV with 10 µM of the respective cOAs (same scale bar and axis for **A-D**). Bottom: zoom-in on a single translocation event of the respective cOA. (**E**) Event duration histograms for the four cOAs with lognormal fits to the data. (**F**) Overlay of all distributions in (**E**), with normalized integrals. (**G**) Relative blockade histograms for the four cOA datasets in (**E**) with normal fits to the data. (**H**) Overlay of the distributions in (**G**) with normalized integrals.

Neural networks can infer the stoichiometry of single cOA molecules

Neural networks have proven very useful in nanopore signal interpretation [309], thanks to their ability to recognize features beyond the mere event duration and current blockade discussed above. Hence, to differentiate between nanopore events caused by cOAs of different stoichiometries, we trained a convolutional neural network (CNN) for single-event classification (**Figure 3A**). The training data consisted of cOA events obtained from mono-disperse cOA samples with known stoichiometry (see Methods), and the data used for model evaluation was obtained from different experiments than the training data, to achieve valid accuracy estimates. Indeed, our CNN outperforms a more conventional machine learning approach (a k-nearest neighbour classifier considering only event duration and current blockade, **Figure S8**), which suggests that the CNN recognizes additional signal properties, such as event shape and current fluctuations. Nevertheless, the events of the two smallest cOAs (i.e. cA_3 and cA_4, with very similar duration and current blockades) are not well separated by the CNN. We note that this can likely be solved, in the future, using alternative (possibly engineered) protein nanopores, capable of distinguishing both stoichiometries. For this study however, we move on with a joint class of cA3/4. Using this approach, the CNN correctly identified 83%, 64%, and 70% of (unseen) mono-disperse $cA_{3/4}$, cA_5, cA_6 events, respectively (**Figure 3B**). Erroneous classifications occur mainly between adjacent stoichiometries (e.g. cA_6 misclassified as cA_5, or cA_5 misclassified as cA_6 or cA3/4), and mainly towards lower stoichiometries (e.g. cA_5 misclassified as $cA_{3/4}$, rather than cA_6). The latter reflects the event duration distributions and their overlap (**Figure 2**): cA_5 and cA_6 events misclassified as $cA_{3/4}$ are marked by short event durations (**Figure S9**). Altogether, 73% of all unseen events in monodisperse samples are correctly identified by the CNN.

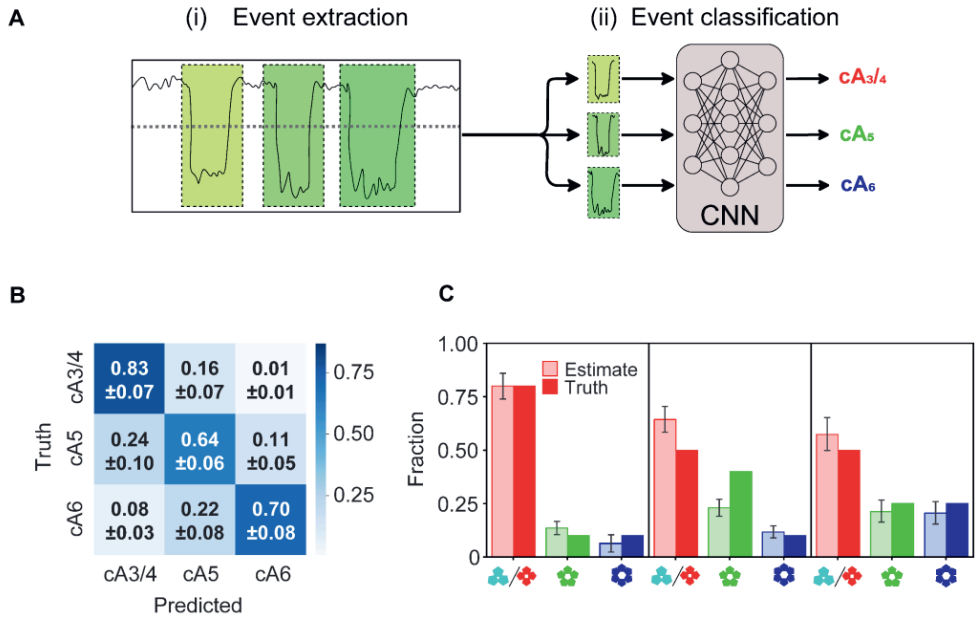


Figure 3: Neural network workflow and validation of the cOA stoichiometry inference from nanopore recordings. (**A**) The cOA classification pipeline consists of (i) extraction of the cOA translocation events, (ii) cOA-classification per-event using a convolutional neural network (CNN). See Methods for details and code. (**B**) Confusion matrix showing the performance for the single-event classification of 2,058 cOA translocation events (not included in the training set). Fractions denote relative prediction frequencies per ground truth class (mean ± standard deviation over 10 cross-validation folds). (**C**) cOA stoichiometry distribution analysis: comparison of the experimental input (dark-coloured) vs. the inferred output (light-coloured) obtained for cOA mixtures of known composition, as indicated (percentages from left to right: 80:10:10, 50:40:10, and 50:25:25). Error bars denote 95% prediction intervals obtained by 10-fold cross validation, which express uncertainty introduced by data variability given this CNN architecture and fitting procedure.

Quantifying polydisperse cOA mixtures and CNN validation

We next assessed the ability of the trained CNN to estimate the stoichiometric composition of polydisperse samples. We prepared three cOA mixtures with known composition, and acquired nanopore recordings of each one. **Figure 3C** shows the true and inferred stoichiometry fractions, with prediction intervals obtained by 10-fold cross validation (see Methods). In qualitative terms, the relative abundances (high vs. low) are correctly identified in all cases, down to the smallest tested fraction of 10%. Quantitatively, the $cA_{3/4}$ population is partially overestimated, while the cA5 population is sometimes underestimated, which is consistent with our results for the mono-disperse cOAs (previous section). For the 50:40:10 mixture (centre in **Figure 3C**), the deviation is larger than the model's prediction interval, indicating additional imperfections beyond the training and test data variability.

Importantly, all ground truth differences in population size were correctly identified by the CNN, which inferred significantly different populations in these cases (t-test, $p << 0.01$, **Table S2**). Similarly, identical ground truth populations are also inferred to be identical – with the exception of the smallest populations of 10%. This indicates the resolution limit of the CNN classification procedure for population sizes, which we (conservatively) estimate to be 15%. Individual uncertainties for the cOA identification procedure can be estimated as 17% for cA3,4, 36% for cA5, and 30% for cA_6, based on the confusion matrix (1 – diagonal value in **Figure 3B**). Equipped with this nanopore-CNN workflow with validated accuracies and uncertainty estimates, we moved on to study cOA mixtures of enzymatic origin.

The stoichiometry of cOAs produced by type III-A and III-B CRISPR-Cas complexes

We next turned to biological cOAs, produced in vitro by two different CRISPR-Cas variants – type III-A and III-B – that coexist in *T. thermophilus* HB8 (see Methods). If the benefit of having not one but two different type III subtypes in one organism is to activate different CARF proteins, then their cOA composition should differ. We directly tested this hypothesis with nanopore experiments using cOAs produced in vitro by reconstituted type III complexes [81,108]. First, we checked if other substances present in the enzymatic reaction mix, such as ATP or short RNAs (i.e. the guide RNA and the target RNA), would interfere with the experiment. However, as they do not produce detectable nanopore signals (**Figure S10**), they do not interfere with the cOA quantification. Likely, the cyclized structure of the cOAs is essential for their detection, whereas the short non-cyclic RNA molecules cannot be resolved because they translocate faster through the nanopore – too fast, in fact, for the time resolution of our experiment (100kHz, cf. **Figure S10**). Interestingly, we find that the

two CRISPR-Cas subtypes produce nearly identical cOA distributions (**Figure 4A, B**). The comparison with the monodisperse calibration data (**Figure 4C**) shows that $cA_{3/4}$ are the predominant stoichiometries in both cases. Using the pre-trained CNN workflow, we quantified the relative abundance of $cA_{3/4}$ as 89±16% and 81±14% for type III-A and III-B, respectively (inferred fraction ± identification uncertainty, see last section). Only minor amounts of cA_5 (8±3% and 12±5%), and cA_6 (3±1% and 7±3%) were detected for type III-A and III-B, respectively, which lie at the resolution limit of our technique. In absolute numbers, this converts to 870 ± 420 cOA molecules produced per type III-A complex, and 2800 ± 840 cOAs per type III-B complex under the conditions used (see Methods). In summary, the presented label-free nanopore-CNN workflow revealed that the cOA produced by the Cas10 subunit of *T. thermophilus* type III-A and III-B complexes are very similar, and they predominantly consist of $cA_{3/4}$ (80-90%).

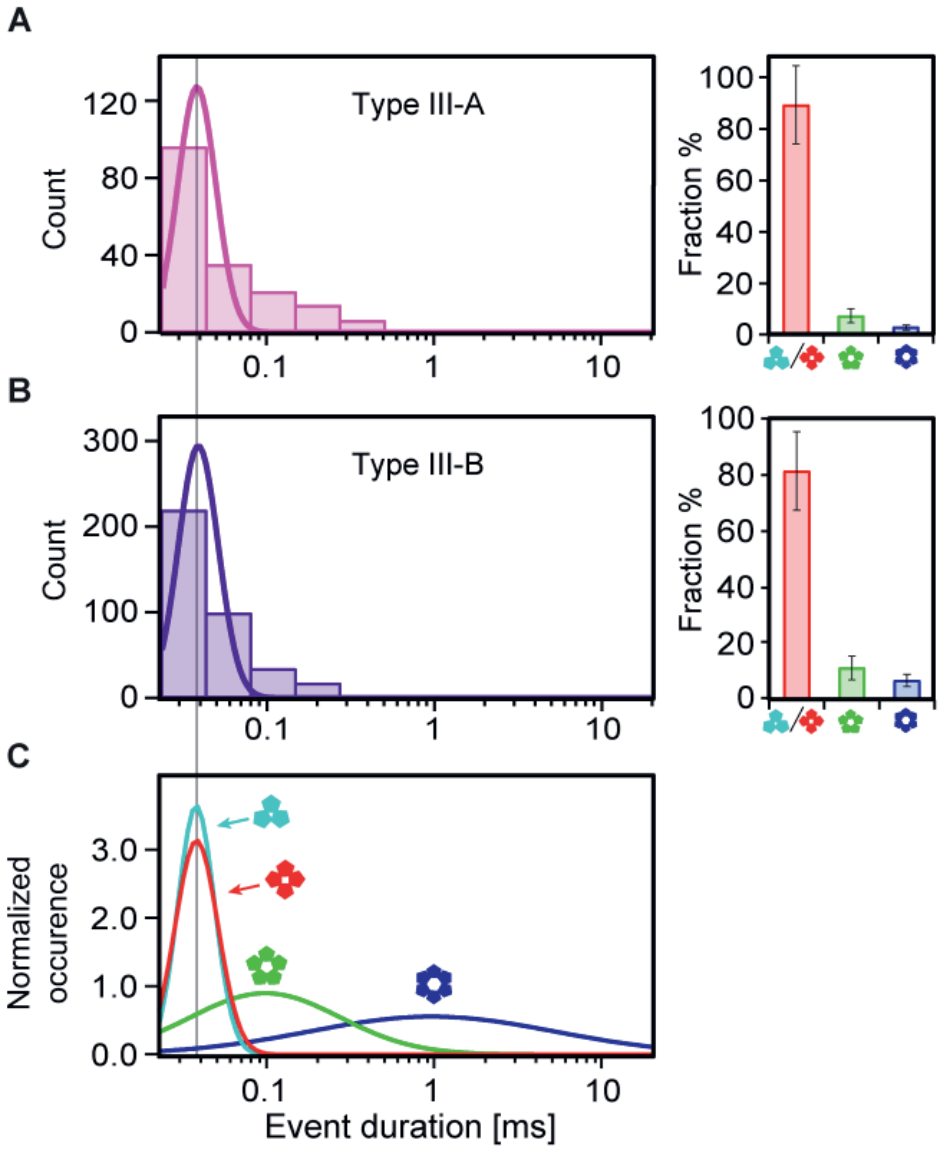


Figure 4: Identification of the cOA composition produced by CRISPR-Cas type III-A and III-B. (**A**) Histograms of the nanopore event durations measured for the cOAs produced by the type III-A variant (measurement conditions as in **Figure 2A-D**). The CNN-based quantification of the stoichiometric composition is shown on the right. Error bars denote the identification uncertainty of each cOA class, estimated as 1 – the diagonal of the confusion matrix in **Figure 3B**. (**B**) Same as A) but for the cOAs produced by the type III-B variant. (**C**) Calibration data obtained with monodisperse cOA (cf. **Figure 2E**). A vertical line through panels **A-C** provides a guide to the eye.

The cOAs produced by CRISPR-Cas type III-A and III-B activate a cA_4-specific CARF

After establishing that the type III-A and III-B complexes of *T. thermophilus* produce similar cOAs, we moved on to test their capacity for downstream activation of the CARF protein Csx1 from the same host (TtCsx1). Since this CARF protein has non-specific RNase activity, a fluorescent RNA-cleavage reporter system was used to screen the activation of CARF proteins by cOAs from synthetic and enzymatic origin (**Figure 5A**, Methods). As expected [211], among the synthetic cA_3 to cA_6, only cA_4 led to TtCsx1 activation, resulting in RNA cleavage (**Figure 5B**). For the cOAs produced by the type III-A and III-B variants (**Figure 5C**), we found that both activate the RNase activity of TtCsx1 much beyond the negative controls (no target, NT). Hence, we can conclude from **Figure 5C** that both variants (type III-A and III-B) must indeed produce cA_4, while other stoichiometries cannot be excluded based on **Figure 5C** alone. However, together with the nanopore-CNN results above, our results show that both type III variants produce a nearly identical cOA composition, including cA_4. These results argue against the hypothesis that individual Cas10 homologues produce distinct cOA compositions for a homologue-specific downstream regulation in *T. Thermophilus*. The presented nanopore-CNN workflow has thus elucidated the hitherto unknown stoichiometric composition of the cOA second messengers of these type III-A and III-B systems, and given the results, the purpose of multiple CRISPR-Cas homologues in one host remains an open question.

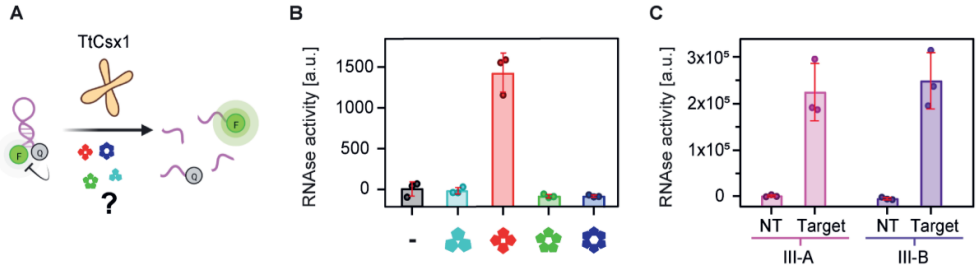


Figure 5: The CARF activation by enzymatic and synthetic cOAs reveals stoichiometric specificity. (**A**) Schematic representation of the RNA cleavage assay. Reporter RNA molecules conjugated with a fluorophore (F) and a quencher (Q) are incubated with the CARF RNase TtCsx1 and various cOA samples. If the cOAs activate TtCsx1, it cleaves the reporter RNA, leading to increased fluorescence intensities compared to the negative controls. (**B**) Activity assay of TtCsx1 after the addition of synthetic monodisperse cOAs, or no cOAs in case of the negative control (-). Assays were run in triplicates. Also, cA_2 was tested and did not activate TtCsx1 (**Figure S11**). (**C**) TtCsx1 RNase activity upon addition of cOAs produced in an in vitro reaction mixture by endogenous type III-A and III-B from *T. thermophilus*, as indicated. cOAs are only produced in the presence of complementary target RNA (Target). Reactions with non-complementary target RNA (no target, NT) serve as a negative control for the assay, run in triplicates.

Conclusion

We present a cOA identification workflow with single-molecule resolution (**Figure 6**) that combines the potential of nanopore technology, neural networks, and enzymatic assays. For the first time, we individually detected these small cOA second-messenger molecules from the type III CRISPR-Cas systems using nanopores. Based on the observed nanopore event durations, we were able to distinguish $cA_{3/4}$ from cA_5, and cA_6. Using monodisperse synthetic cOAs, we acquired calibration data for each stoichiometry, and used it to train a CNN, which we validated using cOA samples of known polydisperse composition. We then detected cOAs of enzymatic origin, produced by the type III-A and type III-B CRISPR-Cas complexes of *T. thermophilus*, and quantified the previously unknown stoichiometric composition using the validated CNN. Additional enzyme activity assays revealed the stoichiometry-specific activation of CARF proteins, and further complement the nanopore-CNN results.

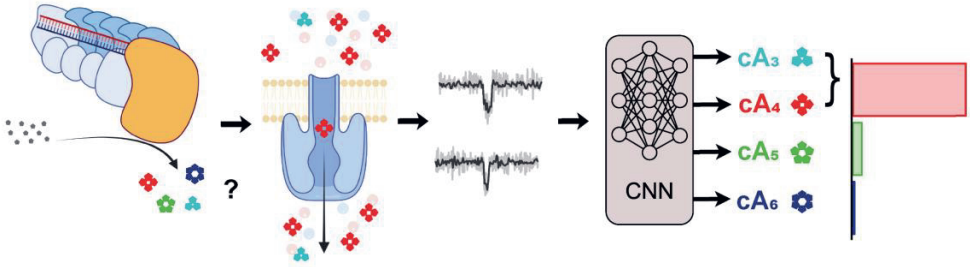


Figure 6: The nanopore-CNN pipeline for the identification of single second messengers from type III CRISPR-Cas. Biological cOAs of unknown stoichiometry, produced by CRISPR type III-A and III-B complexes, are detected one-by-one in nanopore experiments. Single-molecule events extracted from the nanopore reads are then identified by a pre-trained CNN, yielding the stoichiometric composition of the biological cOA samples.

We conclude from the nanopore experiments that (i) cyclic oligo-adenylates (cA_3 to cA_6) can be sensitively detected at the single-molecule level; (ii) the $cA_{3/4}$, cA_5, cA_6 stoichiometries cause distinguishable nanopore event durations; and (iii) the type III-A and type III-B CRISPR-Cas complexes produce cOAs of qualitatively near-identical stoichiometry: predominantly the small cA_3 or cA_4. Using a pre-trained CNN, we (iv) identify the stoichiometries of the cOA at the single-event level with an accuracy of 73%, and we (v) quantify the cOA composition produced by *T. thermophilus* CRISPR-Cas type III-A and type III-B variants as 89±16% and 81±14% cA3/4, respectively, and only trace amounts of cA_5 and cA_6 in both cases. We also estimate (vi) the absolute number of cOAs produced which amounts to 870

± 420 cOAs per type III-A complex, and 2800 ± 840 cOAs per type III-B complex, under the conditions used. Lastly, (vii) the enzymatic RNA cleavage assays verified that the combined $cA_{3,4}$ class must contain cA_4, proving that CRISPR-Cas type III-A and III-B both produce this specific second messenger. Altogether, we thus find nearly identical stoichiometric cOA compositions in both variants, which challenges the hypothesis that co-occurring type III complexes would produce distinct cOA for specific downstream regulation purposes.

In the future, the presented nanopore-CNN pipeline can be used to elucidate the cOA stoichiometries of different CRISPR-Cas type III, and it is readily adaptable to additional signalling molecules, which include – but are not limited to – signalling molecules of other antiviral immune systems, such as CBASS (cyclic oligonucleotide-based antiphage signalling system) [236]. Additional protein nanopores will be explored to detect also cA_2 and to distinguish cA_3 from cA_4, which was not possible with our current hemolysin nanopore. Promising candidates include (possibly engineered) pore proteins with a long narrow channel (hemolysin, aerolysin), or with multiple constrictions (CsgG) [311,312]. Altogether, we demonstrated the first single-molecule sensor for CRISPR second messengers. We achieved stoichiometry sensitivity using affordable nanopore technology that can be integrated into handheld devices and may, in the future, enable quantitative point-of-care diagnostics with single-molecule resolution [78].

Materials and Methods

Protein nanopore recordings were performed using the Montal-Mueller technique, using an Axopatch 200B amplifier and a Digidata 1550B digitizer (both Molecular Devices) with a sampling rate of 500kHz, and low-pass filtered at 100kHz [313]. A free-standing lipid bilayer was formed in a custom-built flowcell with two buffer reservoirs separated by a Teflon film (GoodFellow, Huntingdon, England) with a small electro-sparked aperture (ca. 100µm diameter, obtained with a spark generator, Daedalion, Colorado, United States). The Teflon film was pre-treated on each side with 10µl of a hexadecane solution (10% hexadecane, Acros Organics, Geel, Belgium) in pentane (Alfa Aesar, Massachusetts, United States), and let dry for 5 minutes. Both reservoirs were filled with 400µl of measurement buffer each (3M KCl, 100mM Tris, 0.1 mM EDTA, pH 8), connected to Ag/AgCl electrodes (silver wire with 0.5mm diameter, Advent, Oxford, England, chloridized in household bleach), and the flowcell was placed inside a Faraday cage. Bilayers were built by adding 10µl of a 1,2-diphytanoyl-sn-glycero-3-phosphocholine solution (DPhPC, 10mg/ml, Avanti Polar Lipids, Alabama, United States) in pentane onto each buffer reservoir and pipetting up and down as described previously [313]. α-HL oligomers (kindly provided by Sergey Kalachikov, Columbia University) were added to the trans reservoir (with the working electrode under positive polarity) and their spontaneous membrane insertion caused a characteristic current. The cOA solution (500µM, Biolog, Bremen, Germany) was added to the cis reservoir (with the ground electrode) to a final concentration of 10µM.

Nanopore data processing was performed in Igor Pro (v6.37, Wavemetrics, Oregon, United States) using custom code. Event detection was performed after filtering the signal with a digital 80 kHz low-pass filter and median-conserving decimation to a final sampling rate of 80kHz (25µs time resolution). Events were extracted by applying a threshold at 65% of the open-pore current. The mean current blockades were calculated from the extracted data. To calculate event durations, 1,000-fold bootstrapping with replacement was performed, where each subset was fit with a single-exponential (with X offset). The uncertainty is expressed as the standard deviation of all bootstrapped time constants.

Neural networks: per-event classification. To classify single cOA translocation events, we trained a 1D convolutional neural network (CNN) implemented using Tensorflow [314]. As training data, we used 48 traces of +120 mV translocations from synthetic monodisperse cOA samples with known stoichiometries. Events were extracted as described above, individually normalized and scaled by the standard deviation. To provide information about the relative blockade, we include 47 datapoints (5,875 ms) of the baseline signal on either side of the extracted event. Moreover, to

ensure consistent input size, we pad each to a width of 250 datapoints with zeroes on the right-hand side. The original event duration is concatenated to this signal as a final feature.

We feed each individual padded event into the CNN, which performs multi-class classification. The CNN consists of two 1D convolutional layers with 10 filters of width 25 and a ReLu activation function, each of which is followed by a batch normalization and 20% dropout operations. Next, values are maxpooled with a pool size of 10, and fed into a dense layer with 4 output nodes (one per cOA class) and a softmax activation function. The CNN is trained for 100 epochs at a learning rate of 0.001 with the Adam optimizer that minimizes cross entropy loss based on the 3 one-hot encoded cOA classes. During training, we perform oversampling to prevent class imbalance. Five restarts were performed, after which the classifier with the highest training accuracy was selected for testing. Training was done on a laptop equipped with an Intel Core i7 processor (8 cores @3GHz, Santa Clara, CA) and a T500 GPU (NVIDIA, Santa Clara, CA). Training runtime was around 20 minutes.

CNN validation was done using a 10-fold cross-validation scheme. Importantly, events of a given trace were never split between training and test sets, thus ensuring that no trace-specific characteristics were learned. The reported overall accuracy was calculated over the predictions of all folds merged into one data set. We compared the CNN performance to a traditional feature-extraction-based classifier by performing k-nearest neighbour (KNN) classification based on relative blockade and (log transformed) dwell times (k=3 after hyperparameter optimization), implemented using sci-kit learn [315]. In cOA mixture classifications, t-tests and two one-sided t-tests (TOST) procedures were used to test difference and equivalence respectively between all cOA distributions. The TOST procedure tests whether two cOA relative distributions $N(\mu_{cAx}, \sigma_{cAx})$ and $N(\mu_{cAy}, \sigma_{cAy})$ have means diverging less than a given level δ, by performing two one-sided t-tests, with null-hypotheses:

$$H_{01}: \mu_{cAx} - \mu_{cAy} < -\delta$$

$$H_{02}: \mu_{cAx} - \mu_{cAy} > \delta$$

Where rejection of both H_{01} and H_{02} means that the alternative hypothesis that the means differ less than δ must be accepted, or:

$$H_1 : -\delta > \mu_{cAx} - \mu_{cAy} > \delta$$

Both methods are implemented in python using the scipy package and available in the main code repository for this paper (see Code availability) [316].

The abundance estimation ability was evaluated by running inference using the trained model on polydisperse samples of known composition. To obtain prediction intervals reflecting variation induced by training data, prediction was repeated with all 10 classifiers obtained during cross-validation. We compensate for differences in event rate by using cOA-specific constants which convert from event count to estimated abundance in a sample. The constants are estimated by counting the number of events in the training files, and normalizing for concentration and time. The values of these correction factors (in units of events s^{-1} μMol^{-1}) were 0.17±0.05, 0.62 ±0.7 and 0.12 ±0.3 for $cA_{3/4}$, cA_5, and cA_6 respectively.

Enzymatic cOA synthesis and quantification:
Either TtCsm (type III-A) or TtCmr (type III-B) endogenous protein complexes from *Thermus thermophilus* (purified as described previously [81,108]) were incubated in 150 mM NaCl, 20 mM Tris-HCl pH 8, 10 mM DTT, 2 mM MgCl and 1 mM ATP in a total volume of 20 µl, to which 200 nM of non-target or target RNA (IDT, Coralville, Iowa) was added, as indicated. The reactions were carried out in triplicate by incubating the samples for 60 minutes at 65°C.

Target RNA sequence:

5′ GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGGCCAUCG 3′

Non-target RNA sequence:

5′ UGAUGAGGUAGUAGGUUGUAUAGUAAGCUUGGCACUGGCCGUCGUUUACG 3′

The absolute numbers of each cOA produced by the type III-A and III-B complexes were calculated from the observed nanopore event rate r_e (in units of s^{-1}), the CNN deduced fraction cAx% of each stoichiometry class 'x', and the pre-defined correction factor CF_x (see last section), the dilution factor d (between the enzyme reaction and the nanopore experiment), and the Avogadro constant N_A:

$$n(cOA) = r_e * (cA_{3/4}\% / CF_{3/4} + cA_5\% / CF_5 + cA_6\% / CF_6) * d * N_A$$

For the type III-A and III-B measurements, the respective event rates were 0.47±0.19 s^{-1} and 1.17±0.11 s^{-1}, and the dilution factors were d = 21 and d = 27.6, respectively. We report the cOA numbers normalized per type III-A or III-B complex (their concentration during cOA synthesis was 62.5 nM), and the uncertainty was calculated by error propagation of the reported individual uncertainty estimates.

CARF cOA stringency assay. The CARF protein Csx1 from *Thermus thermophilus*, 1μM TtCsx1, purified as described previously [78], was incubated with 250 nM RNaseAlertTM (ThermoFisher, Waltham, USA) in 150 mM NaCl, 20 mM Tris-HCl pH 8, 10 mM DTT, and 1 mM ATP in a total volume of 20 μl, to which 1 μM of synthetic cOAn (n=2-6, Biolog, Bremen, Germany) was added. The reactions were carried out in triplicate for 60 minutes at 65ºC in a Bio-Rad CFX384TM Real-Time System (Hercules USA), measuring FAM at 1 min intervals. Data shows the relative fluorescence at 30 min.

CARF activation assay with enzymatic cOA from type III CRISPR. Either type III-A or type III-B endogenous complexes from *Thermus Thermophilus* were incubated with 1 μM TtCsx1 (the CARF protein), 250 nM RNaseAlertTM (ThermoFisher, Waltham, USA) in 150 mM NaCl, 20 mM Tris-HCl pH 8, 10 mM DTT, 2 mM MgCl and 1 mM ATP in a total volume of 20 μl, to which 200 nM of non-target or target RNA (IDT, Coralville, Iowa) was added, as indicated. The reactions were carried out in triplicate for 60 minutes at 65ºC in a ThermoFisher QuantStudio 1 Real-Time PCR System (Waltham USA). Data shows the relative fluorescence at 30 min.

Target RNA sequence:

5′ GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGGCCAUCG 3′

Non-target RNA sequence:

5′ UGAUGAGGUAGUAGGUUGUAUAGUAAGCUUGGCACUGGCCGUCGUUUACG 3′

Code availability. CNN training and evaluation code is freely available on Gitlab at https://github.com/cvdelannoy/coa_classifier

Notes

J.A.S. is a founder and shareholder of Scope Biosciences B.V. R.H.J.S is a shareholder and members of the scientific board of Scope Biosciences B.V. J.A.S and R.H.J.S are inventors on CRISPR-Cas related patents/patent applications.

Acknowledgments

We thank Sergey Kalachikov and Jens Gundlach for kindly providing aHL and MspA, respectively. We thank Chenyu Wen for critical reading of the manuscript, and Giovanni Maglia for early discussions on instrumentation. R.H.J.S. is supported by a VIDI grant (VI.Vidi.203.074) from The Netherlands Organization for Scientific Research (NWO). SS acknowledges IPM4 funding from Wageningen University, and the NWO XL grant ProPore: OCENW.XL21.XL21.003.

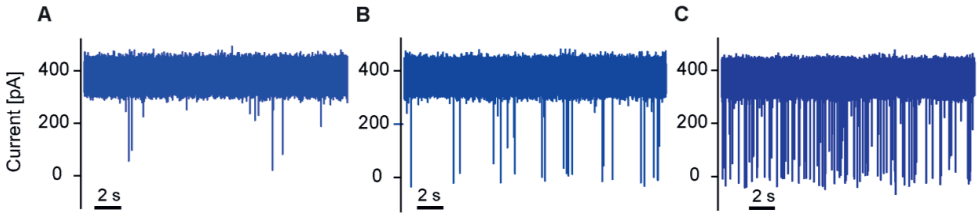


Figure S1: Concentration-dependent cOA event rates. Current traces recorded with a trans-inserted αHL pore at +160mV in a 3M KCl buffer (Methods) after the addition of cA_6 to the cis chamber: (**A**) 100nM, (**B**) 1µM, (**C**) 10µM cA_6. The event rates are: 0.2, 0.7, and 3.8 events/s, respectively. cA_6 events are also detectable below 100 nM, albeit at correspondingly slower event rates.

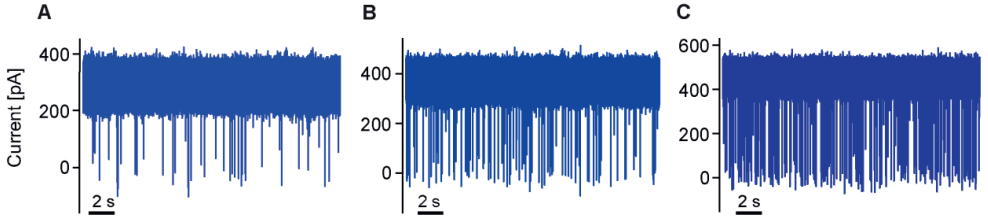


Figure S2: Voltage-dependent event rates and durations. Current traces recorded with a trans-inserted α-HL pore in a 3M KCl buffer (Methods) after the addition of 10µM cA_6 to the cis chamber, recorded at (**A**) +120mV, (**B**) +160mV, (**C**) +200mV. As expected for electrophoresis-driven translocation events, the event rate increases with voltage (1.7, 4.3, and 7 events/s, respectively) while the event duration decreases. The voltage-dependent event durations for cA_3, cA_4, cA_5, cA_6 are specified in **Supplementary table 1**.

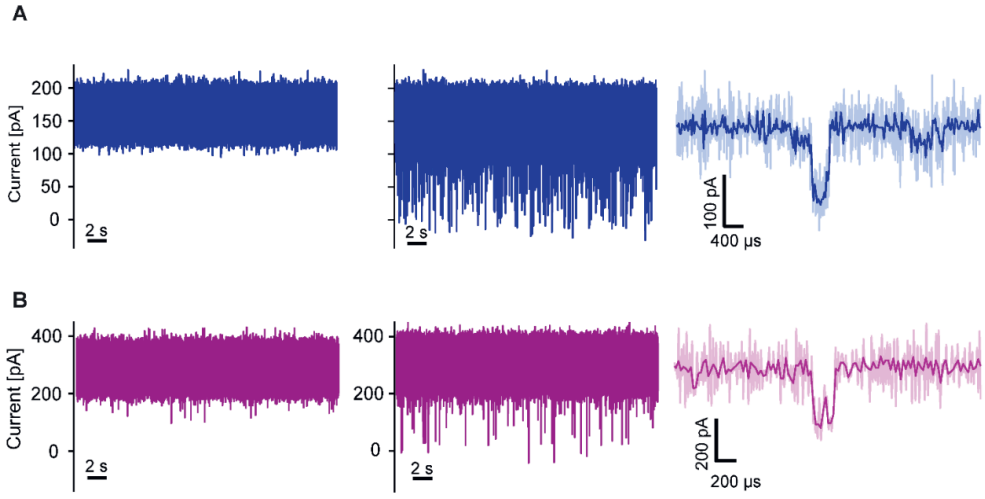


Figure S3: The MspA nanopore detects cOAs with shorter event durations than α-HL. (**A**) α-HL. (**B**) MspA. Both **A, B**) were measured under identical conditions: a cis-inserted pore at +160mV in a 1M KCl buffer (Methods). Left: baseline without analyte. Middle: current trace recorded after the addition of 10µM cA_6 to the cis chamber. Right: zoom-in on a single event. The events detected with MspA are often truncated and generally shorter than those with α-HL, making the latter a more suitable choice for cOA detection.

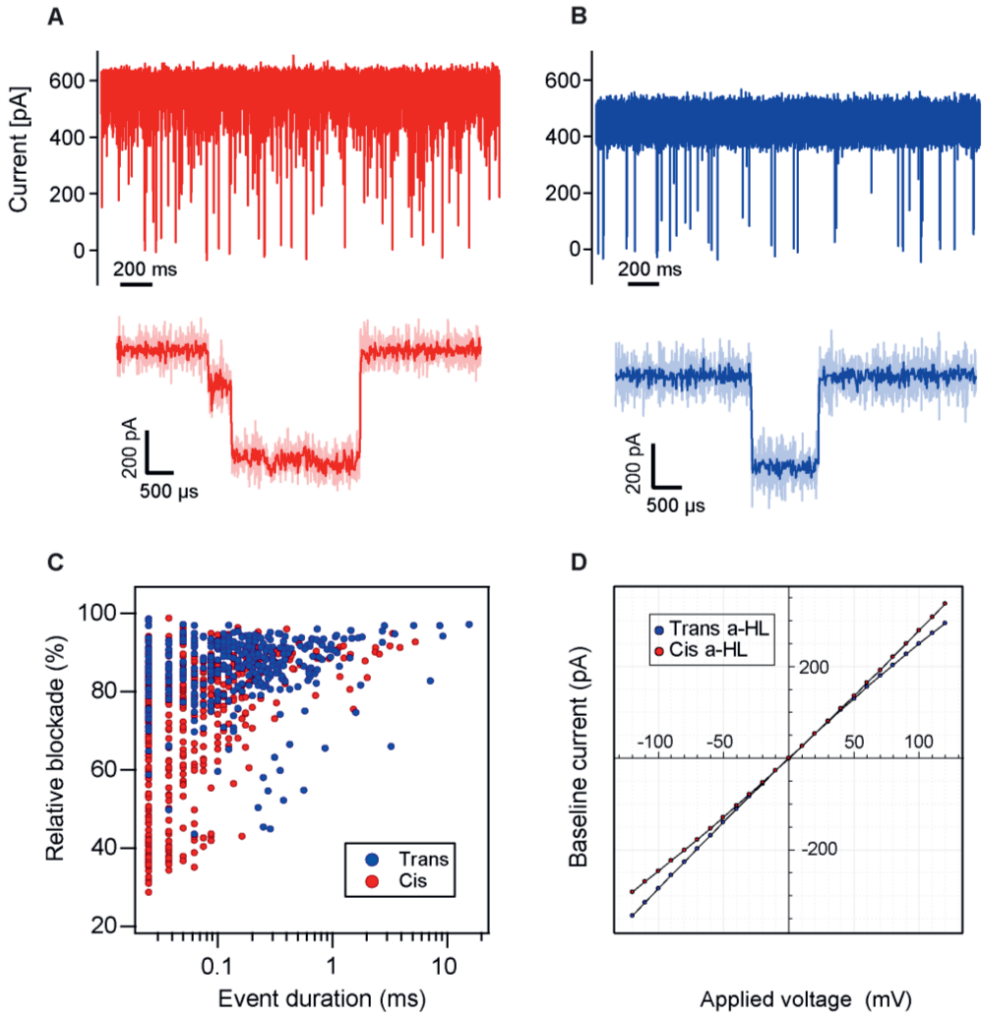


Figure S4: Comparison of the cOA detection capacity of cis vs. trans-inserted α-HL nanopores. (**A, B**) show a current trace and event zoom-in recorded with a cis- (**A**) or trans-inserted α-HL pore (**B**), respectively, both measured at +200mV, in 3M KCl, with 10µM cA_6 in cis. The events observed with a cis-inserted α-HL were less uniform. (**C**) Scatter plot of blockade vs. duration for the events cis in **A** and trans events in **B**, respectively. The events of the cis-inserted α-HL (red) are more heterogenous (spread out) than the more uniform ones of the trans-inserted pore (blue). (**D**) Current-voltage curves of cis- and a trans-inserted α-HL pores as indicated (without cOA).

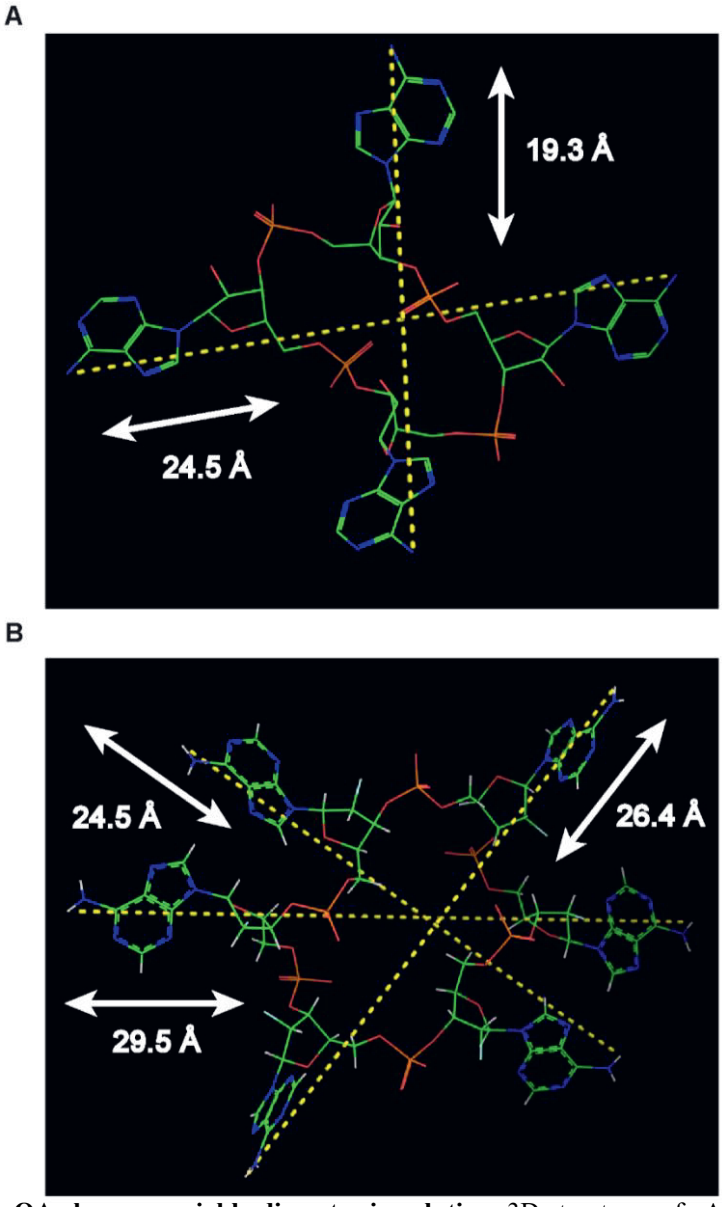


Figure S5: cOAs have a variable diameter in solution. 3D structures of cA_4 and cA_6 in complex with (**A**) SsCsa3 (PDB ID: 6W11) and (**B**) EiCsm6 (PDB ID: 6TUG) proteins, respectively. The proteins have been hidden for visualization purposes. The cOA have flexible bonds and fold in solution. This leads to a variable diameter, which allows them to squeeze through the narrower α-HL pore (1.2 nm). Note that the molecule shown at the bottom is cyclic hexa-2'-fluoro-hexa-dAMP (cFA_6), a cA_6 mimic resistant to degradation but still able to activate its substrate EiCsm6.

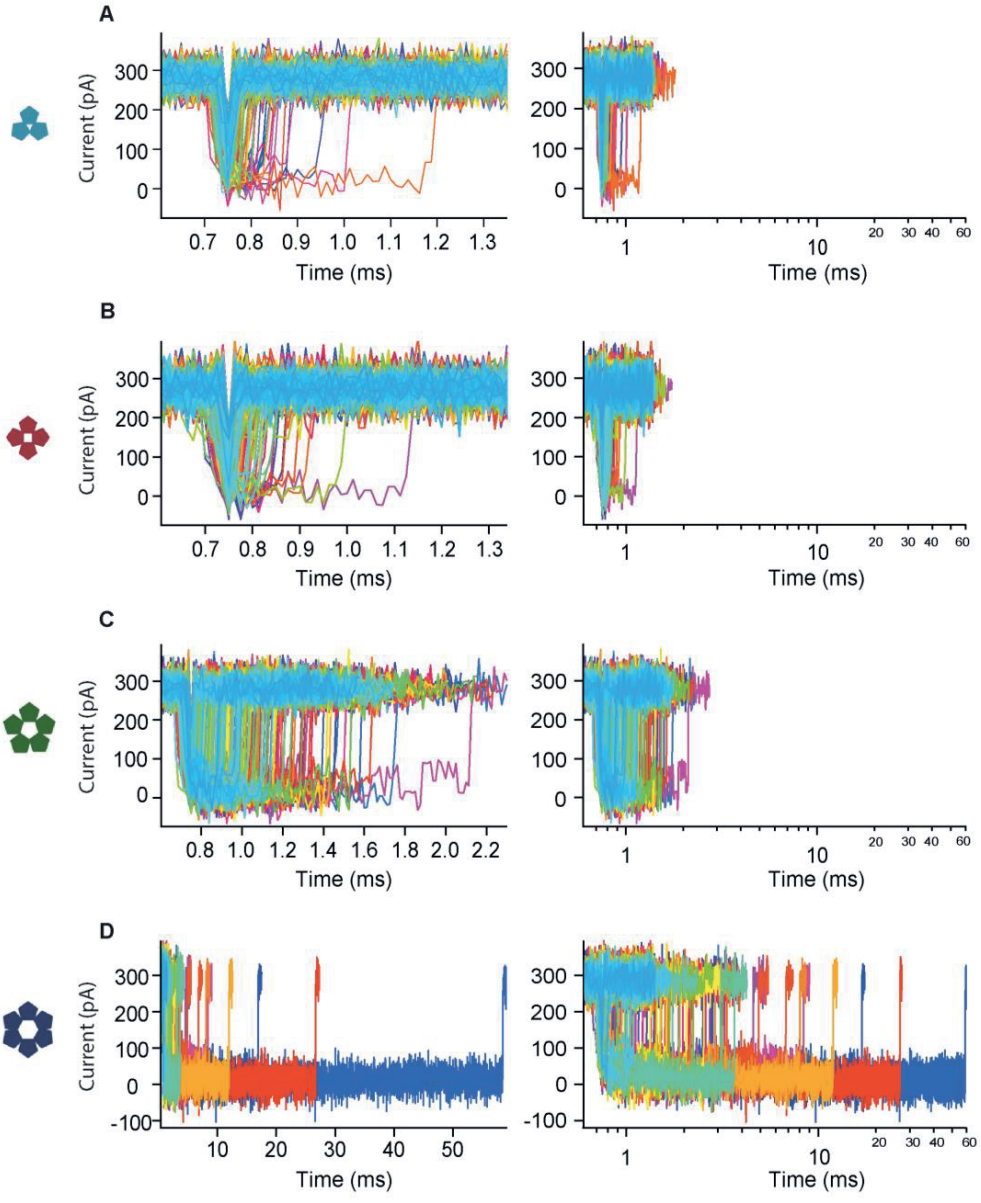


Figure S6: Aligned nanopore translocation events of cOAs: (**A**) cA_3, (**B**) cA_4, (**C**) cA_5, (**D**) cA_6. The events were obtained as in **Figure 2A** (10µM cOA in cis chamber, trans-inserted α-HL pore, +120mV, 3M KCl buffer). The events are aligned at their starting point and individually coloured. Plots with linear (right) and logarithmic (left) time axes are shown.

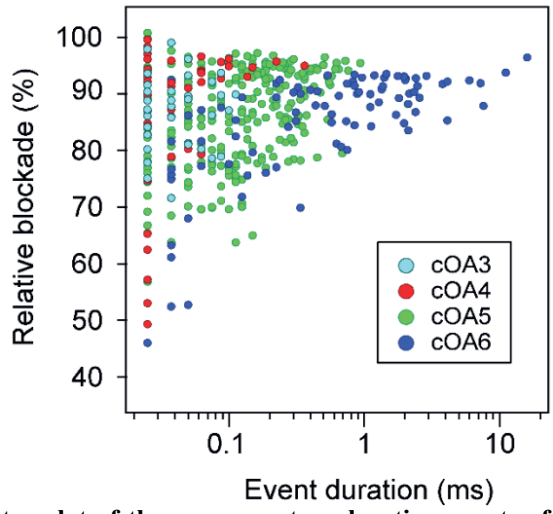


Figure S7: 2D scatter plot of the nanopore translocation events of the indicated cOA molecules. Each dot represents a single translocation event, represented by its relative current blockade vs. its duration.

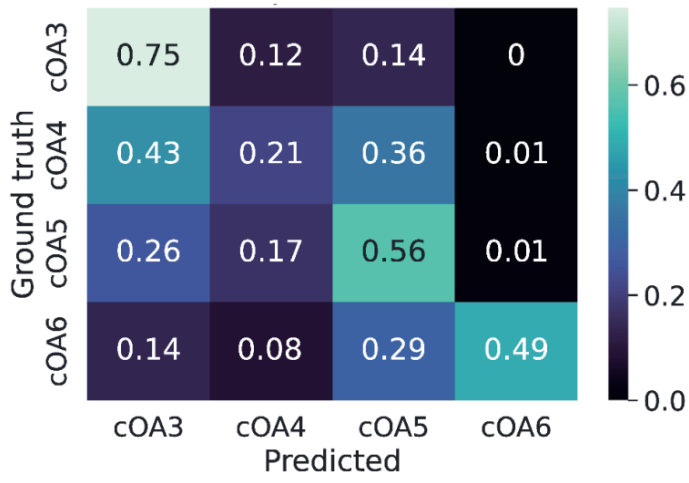


Figure S8: k-nearest neighbour performance on cOA event classification. This confusion matrix shows the performance of a k-nearest neighbour (kNN) classifier for cOA event identification, fed with the same data as the CNNs (**Figure 3B**). The kNN only uses current blockade and event duration as features for classification, and performs poorly when compared to the trained CNN. This is likely due to the fact that the CNN is able to recognize more subtle patterns present in the signal, such as event shape and current fluctuations, leading to a better classification of unseen events.

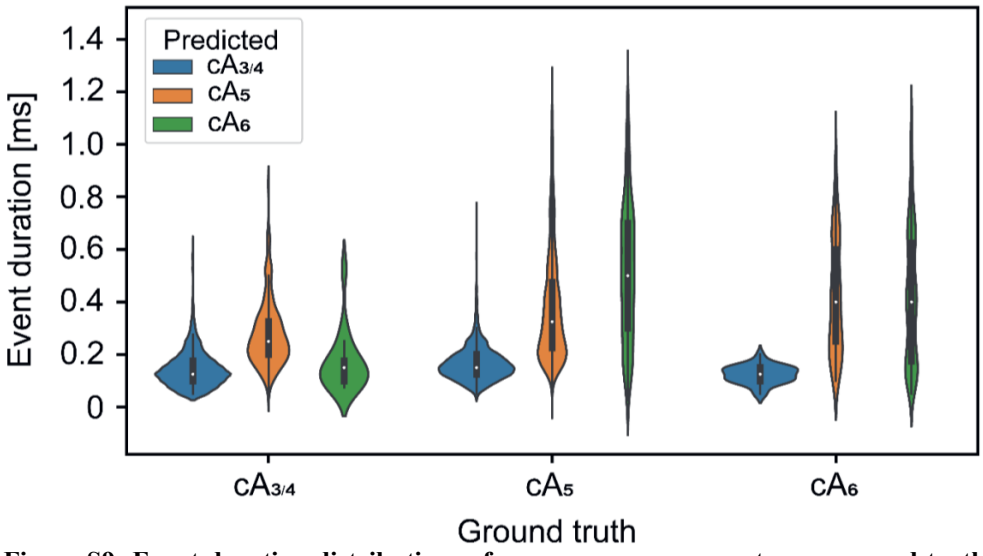


Figure S9: Event duration distributions of unseen nanopore events, per ground truth and CNN predicted stoichiometry. Test events from mono-disperse samples were classified using the trained CNN and subsequently analysed on event duration for each combination of ground truth and CNN-predicted stoichiometry. Distributions are shown as violin plots, where each violin surface is normalized to the same surface area to allow for visual distribution comparison. Events that were erroneously predicted to be of lower-stoichiometry than indicated by the ground truth showed shorter event durations, indicating that this feature played a major role in misclassifications.

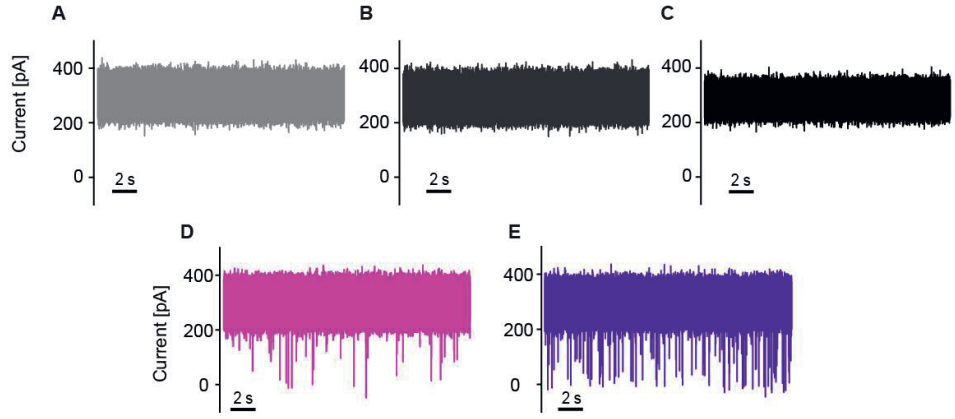


Figure S10: Negative control: neither ATP nor RNA from enzymatic reactions cause detectable nanopore events. All current traces were recorded with a trans-inserted α-HL pore at +120mV in 3M KCl buffer. (**A**) Baseline without analyte. (**B**) 1µM ATP in the cis chamber (as used in **Figure 4B,C**, cf. Methods) does not cause any detectable events. (**C**) The enzyme reaction mix (incl. enzyme, RNA, as used in **Figure 4B,C**, see Methods) does not cause any detectable events in the absence of ATP (precluding enzymatic cOA synthesis). **B**) and **C**) confirm that the events detected in **Figure 4B,C** originate from bona-fide Cas10- produced cOAs, and not from other molecules present in the reaction. (**D, E**) Current traces recorded with the active reaction mix (incl. ATP) show typical cOA events, indicating the presence of

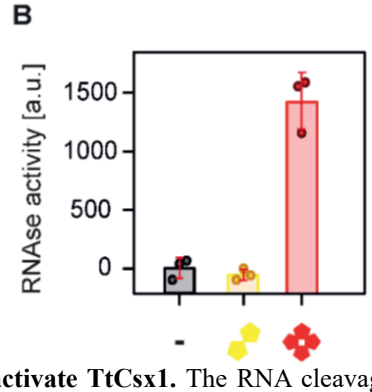


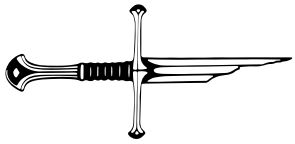
Figure S11: cA_2 does not activate TtCsx1. The RNA cleavage assay (cf. **Figure 5**) was extended to cA_2. In contrast to cA_4, it does not activate the RNAse activity TtCsx1 beyond the negative control (-).

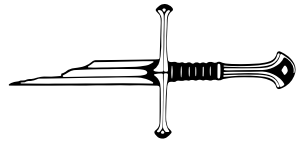
Table S1: Nanopore event durations of cA_{3-6} at different voltages. The time constants and standard deviations (SD) were obtained by 1,000-fold bootstrapping with replacement, fitting the bootstrapped event duration histograms to an exponential function. The average tau and SD values from the 1,000 bootstrapped fits are reported. The nanopore recordings were performed as described in **Figure S2**, and the values were calculated from at least n=150 events. All cOAs show the same trend: decreasing event durations with increasing voltages, indicating translocations instead of trapping events. For the smaller cA_3 and cA_4, with shorter event durations, higher voltages push the event durations below the time resolution (25µs) and the events are not resolved. Altogether, a voltage of +120mV was found to provide the optimal balance between distinct event durations, capture rate, and signal-to-noise for the detection of all four cOAs tested in this work.

	Event duration & SD (µs)			
	cA3	cA4	cA5	cA6
+80mV	32 ± 5	109 ± 22	297 ± 33	2010 ± 944
+120mV	15 ± 1	22 ± 2	96 ± 4	1213 ± 119
+160mV	Not resolved	14 ± 3	61 ± 6	396 ± 39
+200mV	Not resolved	Not resolved	55 ± 5	315 ± 26

Table S2: Pair-wise statistical difference and equivalence test p-values, for neural network classifications of cOA mixtures. t-test p-values and two one-sided t-test (TOST) procedure p-values are reported for difference and equivalence testing respectively. Nomenclature: sample '50_40_10' indicates 50% $cA_{3/4}$, 40% cA_5 and 10% cA_6. While the cA_5 and cA_6 in the 80_10_10 mixture differ significantly according to the t-test ($p << 0.01$), in pairwise equivalence tests, they deviate significantly less than 10% (TOST procedure, $p<<0.01$).

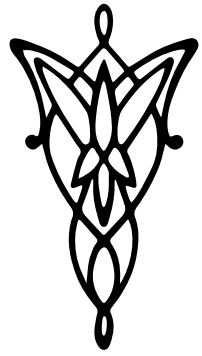
Sample	**cA_A**	**cA_B**	**p_{TOST}**	**$p_{t\text{-}test}$**
50_25_25	$cA_{3/4}$	cA_5	1	3.77×10^{-13}
50_25_25	$cA_{3/4}$	cA_6	1	2.71×10^{-13}
50_25_25	cA_5	cA_6	1.21×10^{-5}	4.88×10^{-1}
50_40_10	$cA_{3/4}$	cA_5	1	1.95×10^{-16}
50_40_10	$cA_{3/4}$	cA_6	1	2.61×10^{-19}
50_40_10	cA_5	cA_6	8.18×10^{-1}	1.25×10^{-8}
80_10_10	$cA_{3/4}$	cA_5	1	3.76×10^{-20}
80_10_10	$cA_{3/4}$	cA_6	1	8.54×10^{-21}
80_10_10	cA_5	cA_6	3.76×10^{-2}	1.33×10^{-6}
type III-A	$cA_{3/4}$	cA_5	1	1.29×10^{-305}
type III-A	$cA_{3/4}$	cA_6	1	1.44×10^{-300}
type III-A	cA_5	cA_6	3.33×10^{-91}	2.07×10^{-74}
type III-B	$cA_{3/4}$	cA_5	1	6.03×10^{-269}
type III-B	$cA_{3/4}$	cA_6	1	4.25×10^{-261}
type III-B	cA_5	cA_6	3.56×10^{-66}	2.65×10^{-50}





Chapter 6

Compact but mighty: biology and applications of type III-E CRISPR-Cas systems



Adapted from:
Steens, Jurre A., John van der Oost, and Raymond H.J. Staals. Compact but mighty: Biology and applications of type III-E CRISPR-Cas systems. *Molecular Cell* 82, no. 23 (2022): 4405-4406.

In this issue, Liu *et al.* present an in-depth study aiming to unravel the structural, biochemical and physiological aspects of how type III-E CRISPR-Cas systems trigger abortive infection by activating a protease upon target RNA recognition [317].

CRISPR-Cas defence systems are divided into two classes, the effector complex of Class 1 systems consists of multiple, single-domain Cas proteins, and Class 2 systems utilize a single, multi-domain Cas protein as effector. Both classes bind a CRISPR-derived crRNA that guides the Cas protein(s) towards their complementary target sequence for cleavage. Type III CRISPR Cas systems (subtypes A-F) belong to Class 1, with the unique feature that they use their crRNA to target complementary RNA. After binding, the target RNA is cleaved at multiple positions and a last self/non-self check (**Figure 1**) is executed. In case of a non-self RNA, activation of a DNase domain (not shown) and a cyclase domain of the Cas10 subunit (**Figure 1**) results in dormancy or death of the host cell by activating ancillary CARF or SAVED proteins. After the discovery of the unique subtype III-E, a series of studies recently provided relevant fundamental insights into this novel type III CRISPR-Cas system (**Figure 1**) [317–325]. In this issue of Molecular Cell, Liu *et al.* combined structural insights with biochemical and in vivo experiments to unravel the multifaceted immune response of this unique CRISPR-Cas system.

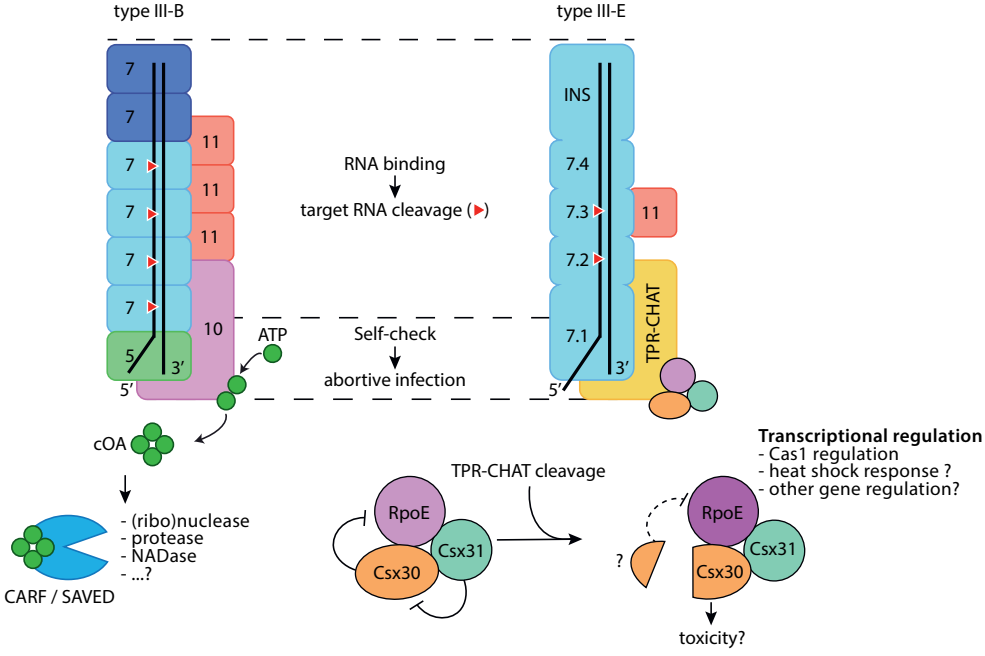


Figure 1. Schematic comparison between type III-A-D (taking III-B as an example) and type III-E CRISPR-Cas complexes and the mechanism of inducing abortive infection respectively

At present, two type III-E systems have been characterized in detail: the Cas7-11 complex of Desulfonema ishimotonii, and the gRAMP complex of *Candidatus Scalindua brodae* [321,323]. Both systems have a similar genomic organization, including (**i**) a large chimeric gene encoding the Cas7-11 polypeptide, (**ii**) a gene encoding TPR-CHAT (also called Csx29, resembling caspase-like proteases, whose eukaryotic counterparts contribute to programmed cell death), and (**iii**) a cluster of three genes that encode RpoE (an alternative sigma factor) and two proteins of unknown function (Csx31 and Csx30).

Liu et al. present an in-depth structural and biochemical analysis of the type III-E system of *Candidatus Scalindua brodae* [317]. Several high-resolution cryo-EM structures are presented of the crRNA-bound gRAMP complex and the gRAMP/TPR-CHAT (Craspase) complex. Both complexes are shown targeting RNA molecules flanked by a 3′ anti-tag (**Figure 1**). This sequence is checked for complementary (self) or non-complementary (non-self) to the crRNA 5′ tag sequence to avoid auto-immunity. The series of structures provide snapshots of the most important steps underlying type III-E immunity. Like all Class 1 CRISPR-Cas systems, the gRAMP structure contains a crRNA that runs along the Cas7 backbone. However, unlike the typical 8 nucleotide 5′ tag, the guide associated with gRAMP has an unusually long 5′ tag (~14-28 nucleotides). Binding of target RNA causes rotation of the Cas11 domain, which facilitates crRNA:target-RNA duplex formation through a positively charged surface on Cas11. The thumb-like structures of the Cas7.2 and Cas7.3 domains interrupt this RNA duplex at positions 4 (site 1) and 10 (site 2) of the spacer sequence by pushing these bases outwards, moving the target RNA into a cleavage-competent position. The authors identified two aspartic acid residues (D547 in Cas7.2, D806 in Cas7.3) involved in cleavage at sites 1 and 2 (together with D698 in Cas7.2 identified earlier [323]). Lastly, structures of Craspase with target RNAs containing self or non-self 3′ anti-tag sequences demonstrated that non-complementarity (non-self) of the first 4 protospacer-proximal nucleotides are crucial for TPR-CHAT activation, whereas prior work showed that target RNA degradation was unaffected [323]. Taken together, it appears Craspase follows a 2-step defence strategy analogous to that of other type III systems [78]: (**i**) a less stringent first line of defence where both self and non-self RNAs are cleaved, and (**ii**) a stringent second line of defence where only non-self RNA recognition results in signal transduction and activation of downstream effectors (**Figure 1**). In type III-E, binding of non-self RNAs allosterically induces rearrangement of the TPR-CHAT catalytic site into an active proteolytic state that specifically cleaves Csx30 and provokes cellular dormancy or death (**Figure 1**).

Besides structural analyses, Liu et al. performed in vivo phage challenge experiments. The complete set of system genes (gRAMP, TPR-CHAT, Csx30, Csx31 and RpoE) was introduced in *E. coli*, together with a CRISPR array containing spacers

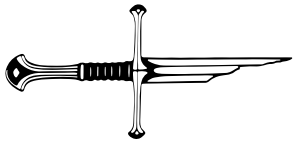
against phage lambda. When challenged with phage, cells exhibit a robust defence phenotype, which was significantly impaired by a catalytically dead TPR-CHAT mutant, or removal of Csx30, Csx31 or RpoE. Growth experiments using phage at different multiplicities of infection (MOI) showed culture collapse when using a MOI > 1, indicative of a defence system that operates through abortive infection. Introducing both the C- and the N-terminal fragments of Csx30 in the presence of Csx31 and RpoE (but without gRAMP and TPR-CHAT) leads to immediate cell death, demonstrating that the cleaved form of Cxs30 mediates cell death and requires Csx31 and RpoE.

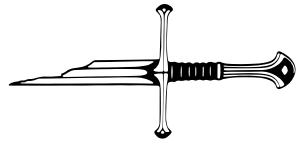
Besides the interesting biology behind these systems, the opportunities to repurpose them are plentiful. The lack of collateral damage and the streamlined nature of this type III-E system makes it a good candidate for knockdown of specific RNA targets [320,321]. Furthermore, *in vivo* and *in vitro* RNA sensing strategies have been engineered, allowing for RNA transcript sensing in live cells and a sensitive RNA-detecting diagnostic tool [320,322].

The exact killing mechanism is an unresolved question that could lead to a better understanding of how this system operates. The study from Liu *et al.* together with two recent studies have provided important pieces of the intriguing type III-E puzzle. Liu et al. showed that Csx30 and its cleaved N-terminal part (Csx30.1) stably binds RpoE and RpoE-Csx31, indicating that these proteins form a heterotrimeric complex. Overexpression of Csx30.1, but not its C-terminal counterpart, resulted in E. coli growth suppression [320]. RpoE acts as a transcriptional regulator, whose activity is inhibited by Csx30 [322]. In case of the type III-E system of *D. ishimotonii*, RpoE binds a particular motif in certain promoter sequences. Interestingly, one of these promoters was located in the type III-E locus, upstream the gene encoding Cas1, a key enzyme of CRISPR adaptation. This may suggest the exciting scenario coupling CRISPR interference and CRISPR adaptation in type III systems, akin to priming in type I systems. More studies addressing the exact role Csx30, Csx31 and RpoE are necessary to elucidate their contribution to type III-E immunity. Finally, it is currently unclear whether III-E complexes use similar target RNA requirements for activating TPR-CHAT, compared to Cas10 activation in other type III systems. For example, mutations in both the 3′ and the 5′ end of the protospacer can prevent TPR-CHAT activation. Analogously, Cas10 activation in other type III systems requires compliance in the 3′ seed as well as strict base-pairing in the 5′ end of the crRNA [78]. It would be interesting to know whether one of these regions represents “the seed region” (affecting both RNA degradation and TPR-CHAT activation) and the other “the CAR region” only affecting TPR-CHAT activation. Given the pace of research in the type III field, we will probably know the answers soon.

Declaration of interests

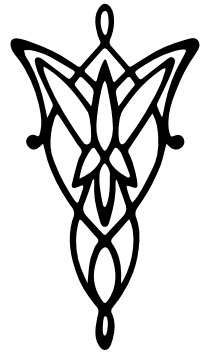
The authors declare the following competing interests. J.A.S. is a founder and shareholder of Scope Biosciences B.V.. J.v.d.O. and R.H.J.S. are shareholders and members of the scientific board of Scope Biosciences B.V.. J.v.d.O. is scientific advisor of NTrans Technologies and Hudson River Biotechnology. J.A.S., J.v.d.O. and R.H.J.S. are inventors on CRISPR-Cas related patents/patent applications.





Chapter 7

Type III-B CRISPR-Cas cascade of proteolytic cleavages



Adapted from:
Steens, Jurre A., Jack PK Bravo, Carl Raymund P. Salazar, Caglar Yildiz, Afonso M. Amieiro, Stephan Köstlbacher, Stijn HP Prinsen, Ane S. Andres, Constantinos Patinios, Arjan Barendregt, Richard A. Scheltema, Thijs J. G. Ettema, John van der Oost, David W. Taylor and Raymond H.J. Staals. Type III-B CRISPR-Cas cascade of proteolytic cleavages. *Science* 383, no. 6682 (2024): 512-519.

Abstract

The generation of cyclic oligoadenylates and subsequent allosteric activation of proteins that carry sensory domains is a distinctive feature of type III CRISPR-Cas systems. In this work, we characterize a set of associated genes of a type III-B system from Haliangium ochraceum, that contains two caspase-like proteases, SAVED-CHAT and PCaspase (prokaryotic caspase), co-opted from a cyclic oligonucleotide-based antiphage signalling system (CBASS). Cyclic tri–adenosine monophosphate (AMP)–induced oligomerization of SAVED-CHAT activates proteolytic activity of the CHAT domains, which specifically cleave and activate PCaspase. Subsequently, activated PCaspase cleaves a multitude of proteins, which results in a strong interference phenotype *in vivo* in *Escherichia coli*. Taken together, our findings reveal how a CRISPR-Cas-based detection of a target RNA triggers a cascade of caspase-associated proteolytic activities.

Introduction

Type III CRISPR-Cas systems are adaptive immune systems in prokaryotes that use CRISPR-derived RNA (crRNA) guides to target the complementary RNA of invading viruses and plasmids. Type III systems have many distinctive features, including a rapidly expanding network of signal transduction pathways to trigger dormancy or cell death [75,264,326].

A typical type III operon encodes multiple Cas proteins that form a type III effector complex with a mature crRNA. These complexes bind complementary target RNAs, a process that is initiated at an exposed seed region at the 3' end of the crRNA. Seed binding licenses complete base pairing between the target RNA and the crRNA, thereby activating Cas10, the signature multidomain subunit of the type III complex [78]. The histidine-aspartate (HD) nuclease domain of activated Cas10 degrades single-stranded DNA substrates in a non-sequence specific manner, whereas the Palm domain acts as a cyclase to convert adenosine triphosphate (ATP) into signalling molecules called cyclic oligoadenylates (cOAs), i.e. rings of three to six adenosine monophosphate (AMP) moieties [70,73,83,112,327]. These cOAs activate a set of effector proteins carrying cOA-binding domain, CARF or SAVED [73,83,242,247]. These sensory domains are generally fused to catalytic (e.g., ribonucleases (RNases), deoxyribonucleases (DNases), nicotinamide adenine dinucleotide nucleosidases (NADases), and toxins), which leads to diverse activities [84,249,265]. Over the last few years, a handful of these CARF and SAVED proteins have been characterized, revealing various activities geared towards killing the host or inducing dormancy, which stops the spread of the invading nucleic acid (e.g. phage progeny, plasmid propagation, etc.) in a process known as abortive infection. Recent work on type III systems has indicated that proteases also play a role in type III immunity, particularly TPR-CHAT (Csx29) which physically associates with the type III-E complex, and a cOA-activated Lon-like protease (CalpL) in a type III-B system [93,323].

Results

Type III-B CRISPR-Cas-CBASS hybrid

We identified a set of genes flanking an operon that encodes a type III-B CRISPR-Cas protein complex (*Cmr1* to *Cmr6*) in *Haliangium ochraceum* DSM 14365 (**Figure 1A**). We observed a *SAVED-CHAT* gene, which encodes a fusion protein of a SAVED domain and a CHAT domain (related to cysteine proteases that include caspases, which are involved in programmed cell death in eukaryotes) (**Figure 1B**). Further downstream, we observed a gene encoding a caspase-like cysteine protease (annotated as "Peptidase C14 caspase catalytic subunit p20"), which we named

PCaspase (prokaryotic caspase) (**Figure 1C**). Located in between these genes are three genes that encode (**i**) a predicted sigma factor we named *PC-σ* (prokaryotic caspase sigma factor); (**ii**) a hypothetical protein we named *PCi* (prokaryotic caspase inhibitor), owing to predicted structural homology with the CI-2 family of serine protease inhibitors; and (**iii**) a serine-threonine protein kinase we named *PCk* (prokaryotic caspase kinase). A similar operon appears to be located elsewhere on the chromosome of *H. ochraceum* (**Figure S1A**).

We explored the co-occurrence of these genes in other prokaryotic genomes. We identified six additional bacterial genomes that encode *SAVED-CHAT* and *PCaspase* genes within the same gene cluster from the phyla Myxococcota (n=2), Bacteriodota (n=2) and Cyanobacteria (n=1) (**Figure 1D**). Only the gene cluster of *Enhygromyxa salina_A* contained gene set comparable to that of *H. ochraceum*, excluding *PCi*, for which homologs could not be identified in other prokaryotes. Two of the operons encoded type III CRISPR-Cas systems in the neighbouring regions, whereas the remaining four operons encoded cGAS/DncV-like nucleotidyltransferase (CD-NTase) domains, indicative of cyclic oligonucleotide-based antiphage signalling systems (CBASS) [236,248].

Sensory domains like SAVED are often fused to different types of effector domains, which implies a separate evolutionary origin of the two domains in *SAVED-CHAT* (**Figure 1B and E**). Phylogenetic analysis showed that the CHAT domains of the two *H. ochraceum SAVED-CHAT* variants were only distantly related, indicating independent acquisition events (**Figure S1B**). The SAVED domains, however, formed a monophyletic clade (Figure 1E) and were likely acquired once and duplicated into two *SAVED* copies to which CHAT domains of different origin were subsequently added. This scenario is further supported by phylogenies of *PCaspase*, *PC-σ*, and *PCk* (**Figure S1C-E**), which all support monophyly of the two respective copies. The SAVED domains are most closely related to those of other Myxococcota and might be part of a conserved system in these bacteria. All but one gene neighbourhood of *SAVED-CHAT* relatives contained a setup reminiscent of a CBASS defense system (**Figure 1F**), with two or three putative cyclases, a protein kinase homolog of *PCk*, a sigma-70-like sigma factor, and a ubiquitin-activating *thiF* gene. Taken together, these findings suggest that cOA sensory and effector components of a CBASS system were co-opted in *H. ochraceum* to work in tandem with a type III CRISPR-Cas system. The pairing of a type III CRISPR-Cas "signal generator" with a CBASS effector is reminiscent of NucC (a cyclic tri-AMP (cA_3)-activated, promiscuous nuclease) that has been described previously [173,266].

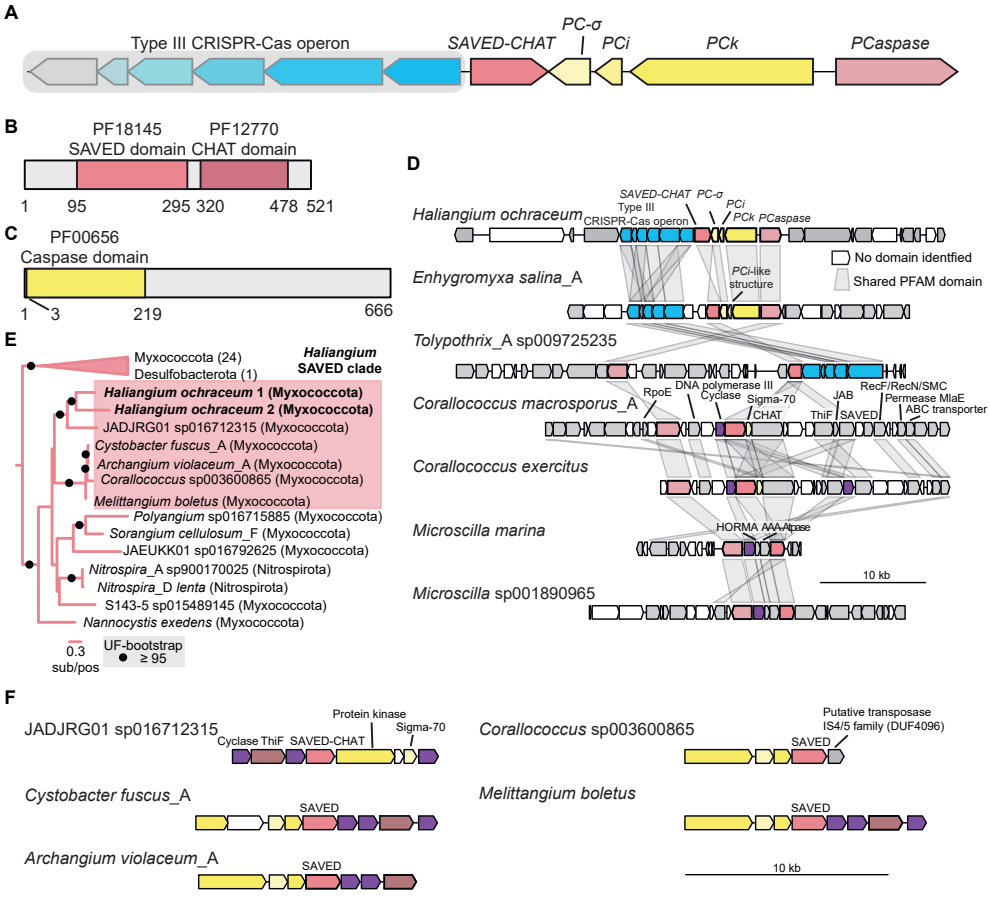


Figure 1. Type III CRISPR-Cas operon associated SAVED-CHAT-PCaspase gene cluster setup and CBASS origin. (**A**) Schematic representation of the *H. ochraceum* type III-B CRISPR-Cas operon containing the *Cmr1-6* genes encoding the crRNA-guided type III protein complex and its associated genes: *SAVED-CHAT, PC-σ, PCi, PCk* and *PCaspase* (locus tags Hoch_1313-1323). (**B**) Domain architecture of SAVED-CHAT. (**C**) Domain architecture of PCaspase. (**D**) Genomic neighbourhoods containing *SAVED-CHAT* and *PCaspase*. (**E**) SAVED domain phylogenetic midpoint-rooted tree. (F) CBASS type operons containing *Haliangium* SAVED clade domains with associated proteins containing cyclase, protein kinase, Sigma-70 and transposase domains.

SAVED-CHAT and PCaspase cleavage activity

We anticipated that SAVED-CHAT acts as a cOA-activated protease with a selective substrate repertoire similar to that of other CHAT cysteine proteases [328]. To test this hypothesis, we purified all type III associated proteins from *H. ochraceum*, excluding PCk, which could not be cloned individually or in combination with PC-σ and PCi, likely because of toxicity. We investigated SAVED-CHAT substrate specificity using in vitro cleavage assays with the other three proteins and observed that SAVED-CHAT specifically cleaves PCaspase in a cA_3-dependent and co-factor-independent manner (**Figure 2A, S2, S3**), but not PC-σ and PCi (**Figure 2B**). SAVED-CHAT cleaved PCaspase into several defined fragments in addition to a myriad of products. A catalytically dead version of SAVED-CHAT (dSAVED-CHAT, which contains substitutions H375A (His^{375}→Ala) and C422A (Cys^{422}→Ala)) showed no activity (**Figure S2**). Furthermore, we noticed that the addition of cA_3 prevented the migration of SAVED-CHAT in native polyacrylamide gel electrophoresis (PAGE), which indicates that SAVED-CHAT oligomerizes to form large complexes in the presence of cA_3 (**Figure S4**).

To elucidate how PCaspase is activated upon cleavage by SAVED-CHAT, we used AlphaFold2 (AF2) to predict the structure of PCaspase. The H79-C146 catalytic dyad of PCaspase was obstructed by an unstructured loop (residues 144 to 163) (**Figure S5A, B**), which suggests a mechanism whereby PCaspase is autoinhibited.

Because the protease active site of SAVED-CHAT lies within a narrow channel, we hypothesized that the PCaspase cleavage site would reside within an unstructured loop. To identify a putative PCaspase cleavage site, we used AF2 multimer to co-fold the CHAT domain of SAVED-CHAT with ~10 to 20 residue long unstructured peptides of PCaspase. We found that the peptide that obstructs the PCaspase catalytic dyad was predicted to form an interaction with SAVED-CHAT, with a single arginine residue at position 153 (R153) predicted to fit into the active site with a high confidence score (89.3) (F**igure S5C**). An AF2 model of the processed PCaspase fragment residues 1 to 153 appears to be catalytically competent. Indeed, activated SAVED-CHAT was unable to cleave an R153A (Arg^{153}→Ala) PCaspase mutant, which shows that R153 is essential for cleavage (**Figure 2C**). Shotgun and topdown proteomic profiling of cleaved PCaspase fragments confirmed cleavage at R153 (**Figure S6**). Notably, R153 was highly conserved in other PCaspase homologs that co-occurred with SAVED-CHAT (**Figure S7A**). The AF2 model predicted that PCaspase R153 would interact with SAVED-CHAT asparagine residue N420 and glutamic acid residue E328 (**Figure S5F**), which were fully conserved or conserved as pairs in several homologs with a co-occurring PCaspase (**Figure S7B**).

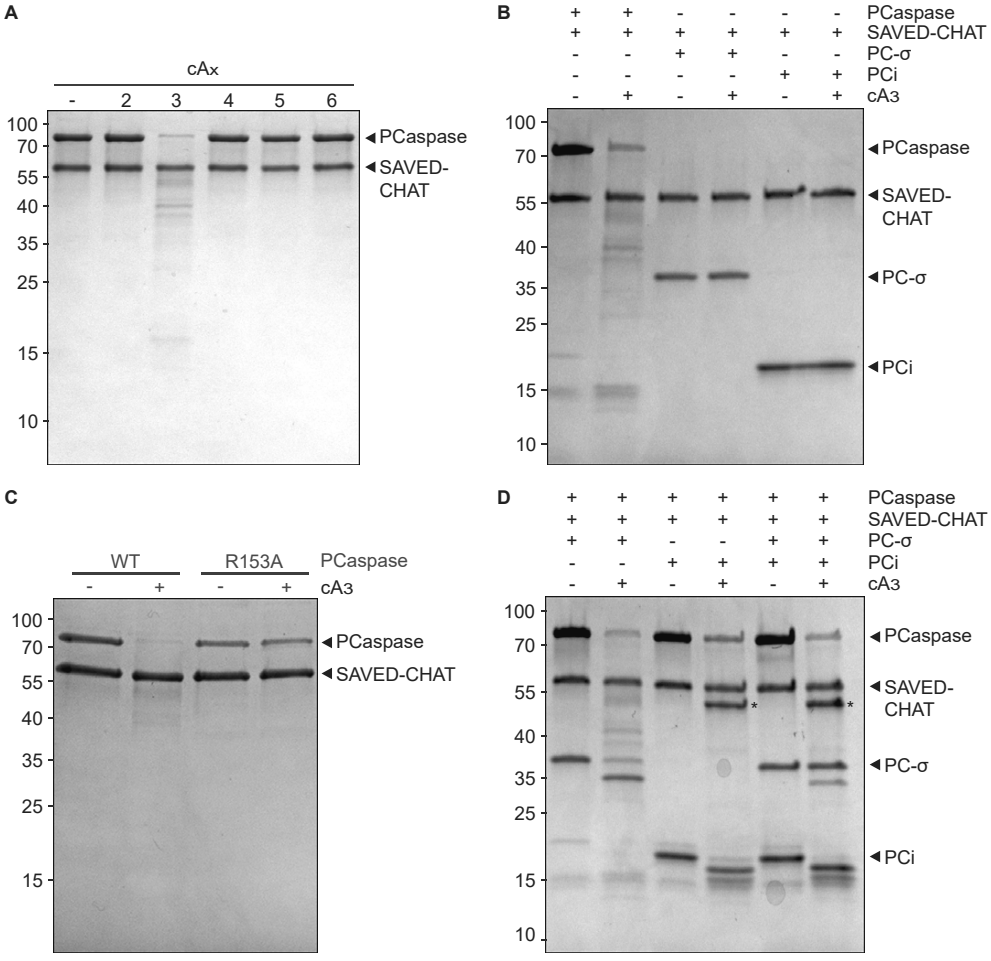


Figure 2. cA_3-induced SAVED-CHAT and subsequent PCaspase activity. (**A**) SDS-PAGE analysis of SAVED-CHAT cleavage, showing its dependency on cA_3 for cleaving PCaspase. (**B**) SDS-PAGE analysis of cleavage activity of activated SAVED-CHAT on PC-σ and PCi. (**C**) SDS-PAGE analysis of SAVED-CHAT cleavage on WT (wild-type) PCaspase and the R153A PCaspase mutant (**D**) SDS-PAGE analysis of cleavage activity of activated PCaspase on PC-σ and PCi. Activated PCaspase cleaves PC-σ and PCi individually, PC-σ is cleaved less when PCi is present. Stabilized PCaspase fragment is indicated with an asterix "*".

Next, we performed protein cleavage assays with SAVED-CHAT and PCaspase using different combinations of PC-σ and PCi (**Figure 2D**). PCaspase cleavage by activated SAVED-CHAT resulted in the subsequent cleavage of PC-σ and PCi into defined cleavage products. These results indicate that PCaspase cleavage by SAVED-CHAT activates its proteolytic activity. When PCi was present, one of the cleaved PCaspase fragments was stabilized and other degradation products of PCaspase were absent, which hints at self-degradation (**Figure 2D**). Furthermore, PCi decreased the efficiency of PCaspase-mediated PC-σ cleavage in a concentration-dependent manner, indicating that PCi acts as an inhibitor for PCaspase activity (**Figure 2D, S8**).

PCaspase-mediated cleavage of PC-σ and PCi prompted us to hypothesize that PCaspase might have a broad substrate repertoire. Indeed, activated PCaspase was able to degrade the biologically unrelated casein, whereas SAVED-CHAT by itself did not (**Figure 3A, S9**). Other unrelated proteins were also degraded by activated PCaspase but not activated by SAVED-CHAT, which further demonstrates the broad range of PCaspase substrates (**Figure S10**). Furthermore, a catalytically dead mutant of PCaspase (dPCaspase, which contains substitutions H79A (His^{79}→Ala) and (Cys^{146}→Ala) C146A) showed no activity (**Figure 3A**). PCi also inhibited PCaspase-mediated casein degradation in a concentration-dependent manner, which indicates that the PCi inhibitory effect is independent of PC-σ (**Figure S11**).

The broad substrate repertoire of PCaspase allowed us to monitor its activity in real-time by providing a small fluorophore-quencher peptide as a substrate (**Figure 3B, S12**). We showed an inverse relationship between PCaspase reaction kinetics and PCi concentration (**Figure 3C**). We probed the sensitivity and kinetics of this detection method to cA_3, and achieved near-instant signal generation in the presence of >31.2 nM of cA_3, with a limit of detection of 15.6 nM (**Figure 3D**). Lastly, we elucidated the target residue of activated PCaspase by testing six different fluorophore-quencher peptides that contained different combinations of candidate residues typically targeted by caspase-like cysteine proteases [329]. All cleaved peptides contained an arginine (**Figure 3E**), which suggests that activated PCaspase cleaves at arginines in unstructured peptides.

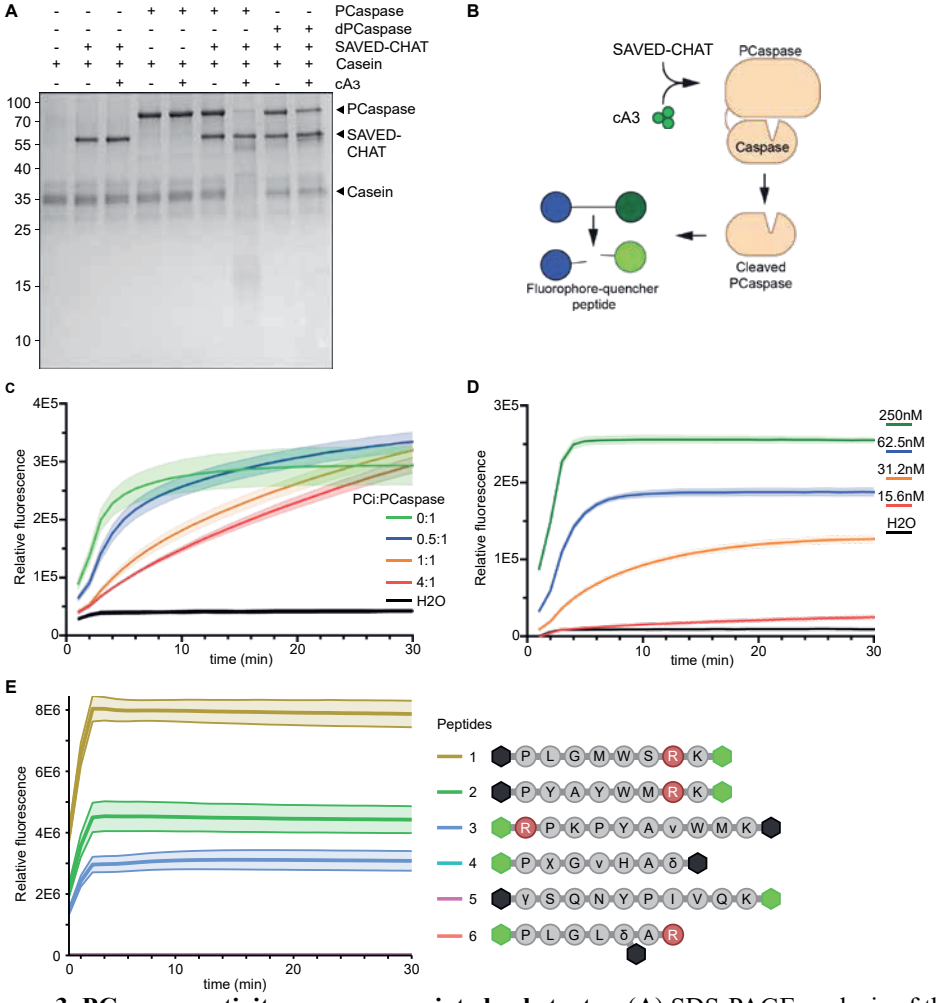


Figure 3. PCaspase activity on non-associated substrates. (**A**) SDS-PAGE analysis of the proteolytic activation of PCaspase by SAVED-CHAT, which results in degradation of the casein substrate, in contrast to activated SAVED-CHAT by itself or the dPCaspase catalytic mutant (H79A, C146A). (**B**) Schematic overview of PCaspase activity assay using a fluorescent peptide reporter. (**C**) Increasing the molar ratio (PCi:PCaspase) of PCi reduces the cleavage activity of a carboxy-fluorescein (FAM)–peptide substrate by activated PCaspase. The shaded regions represent the standard error of the mean (n = 3 technical replicates). (**D**) Sensitivity for cA_3 of the FAM-peptide visualization method is 15.6 nM. The shaded regions represent the standard error of the mean (n = 3 technical replicates). (**E**) PCaspase cleaves reporter peptides that contain arginines. Arginines are in red, fluorophores are in green, and quenchers are in black. Shaded regions represent the standard error of the mean (n = 3 technical replicates). ν, norvaline; c, cyclohexylalanine; d, diaminobutyric acid; g, GABA. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; V, Val; W, Trp; and Y, Tyr.

SAVED-CHAT and PCaspase activation leads to a strong defense mechanism phenotype

Owing to the apparent broad substrate specificity of PCaspase, we hypothesized that the sequential activation of *H. ochraceum* type III CRISPR-Cas by a target RNA, SAVED-CHAT, and eventually PCaspase could affect cell viability. We therefore constructed the pHochTypeIII plasmid, which encodes *H. ochraceum Cmr1* to *Cmr6* (forming the type III complex), csb2 (a cas6 homolog processing the pre-crRNA), and a minimal CRISPR array containing two repeats and a spacer (**Figure S13A**). The SAVED-CHAT and PCaspase genes were expressed in different combinations from the pEffector plasmid (**Figure S13B**). NucC (a cA_3-responsive nuclease that causes abortive infection) from *Escherichia. coli* MS115-1 and an empty vector were used as positive and negative controls, respectively (**Figure S13B**) [88]. Target and non-target plasmids encoded an isopropyl-β-D-thiogalactopyranoside (IPTG) inducible protospacer that was complementary and non complementary to the spacer, respectively (**Figure S13C**) [330].

The target or non-target plasmid was transformed into *E. coli* BL21-AI that harboured pHochTypeIII and one of the pEffector plasmids (**Figure 4A**). Ten-fold serial dilutions were plated on glucose- or IPTG-containing plates to repress or induce, respectively, the target or non-target RNA to calculate transformation efficiencies. First, we demonstrated that the *H. ochraceum* type III system produces cA_3 upon target RNA recognition and leads to an abortive infection phenotype in the presence of NucC (**Figure 4B**) [88]. When both SAVED-CHAT and PCaspase were present, target RNA expression reduced transformation efficiency by six orders of magnitude (**Figure 4B, S13D**). Additionally, we investigated the in vivo effect of expressing different combinations of the two expected PCaspase fragments after R153 cleavage, along with SAVED-CHAT. None yielded a strong reduction in transformation efficiency compared with expression of full-length, wildtype PCaspase, perhaps owing to changes in the structural conformations of the fragments or weaker interactions when they were expressed separately (**Figure S14**). Overall, we confirmed that *H. ochraceum* type III-B produces (at least) cA_3 and mounts a robust defence response in combination with its effector proteins SAVED-CHAT and PCaspase.

To investigate the inhibitory effect of PCi on PCaspase *in vivo*, we cloned it with and without PC-σ into the pSAVED-CHAT_PCaspase plasmid (**Figure S13B**). The results showed that upon target RNA expression, PCi improved transformation efficiency by two orders of magnitude (irrespective of the presence of PC-σ) compared with SAVED-CHAT and PCaspase alone (**Figure 4C, S13E**). This is in agreement with the observed *in vitro* PCi-mediated PCaspase inhibition, and shows that *in vivo*, PCi alleviates the PCaspase-mediated defense mechanism phenotype. Unfortumate-

ly, we were unable to produce a strain that expressed PC-σ in combination with SAVED-CHAT and PCaspase, likely because of toxicity.

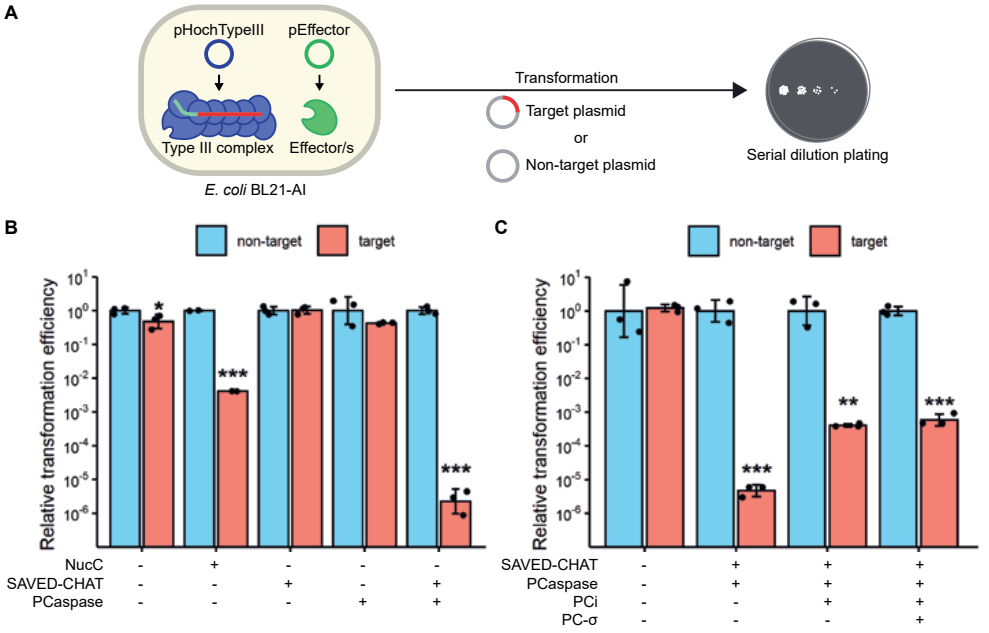


Figure 4. PCaspase activation reduces the transformation efficiency of a target plasmid. (**A**) Schematic overview of the experimental setup in *E. coli* BL21-AI. (**B**) Transformation efficiencies (relative to the non-target control) of target or non-target plasmid in E. coli co-expressing *H. ochraceum* type III CRISPR-Cas complex and different combinations of effector proteins. (**C**) Transformation efficiencies (relative to the non-target control) when SAVED-CHAT and PCaspase are co-expressed with PCi and PC-σ. Error bars represent the standard deviation of the mean. Statistical significance was calculated using one-sided unpaired Welch's t-test. * - $p < 0.05$, ** - $p < 0.005$, *** - $p < 0.0005$ ($n = 3$ biological replicates).

Structural basis for SAVED-CHAT activation

To understand the structural basis for cA_3-induced activation of SAVED-CHAT, we determined a cryogenic electron microscopy (cryo-EM) structure of the activated SAVED-CHAT in complex with cA_3. Size-exclusion chromatography and negative staining indicated that SAVED-CHAT is monomeric in solution, whereas cA_3 binding induced the formation of large, polydisperse oligomers as observed with native PAGE (**Figure S4, S15, S19**). Prolonged incubation of cA_3 with SAVED-CHAT resulted in the formation of large, polymorphous aggregates that were not amenable to structural determination (**Figure S16A**). However, by vitrifying immediately after mixing SAVED-CHAT and cA_3, we were able to obtain a structure of the filament at a global resolution of 3.1 Å (**Figure S16B-H, S17 and Table S1**).

SAVED-CHAT oligomerization results in long, curved filaments with the SAVED domain on the inside and the CHAT domains at the periphery, with a curvature of ~10º between each monomer (**Figure 5A, S16**). SAVED-CHAT monomers assemble through head-to-tail oligomerization, with a single cA_3 bound at the interface between two SAVED domains, forming 'singlet' filaments. We additionally observed partial "doublets" in which two antiparallel filaments make a cross-fiber interaction that spans up to three SAVED-CHAT monomers (**Figure 5A**).

In the singlet regions of our map, the CHAT domains were poorly resolved and often absent, which was likely due to conformational flexibility (**Figure 5B, C**). Analysis of AF2 models of monomeric SAVED-CHAT revealed a high degree of flexibility between the two globular domains, with an unstructured linker (residues 297 to 304) acting as a hinge to accommodate a maximum displacement of up to 35 Å (**Figure S18A**). It is likely that the strain induced by the filament curvature prevents more than three successive CHAT domains from forming stable inter-fibre contacts.

The intra-filament head-to-tail CHAT-CHAT interface is composed of a four-helix bundle with two helices from each protomer (**Figure 5D**). This bundle is held together through a network of predominantly electrostatic interactions, which is consistent with a SAVED-CHAT mutant in which this interface was perturbed (**Figure S19**). The CHAT-CHAT inter-filament doublet interface is mediated by an unusual π-π stacking interaction between two R349 residues (the same residue from different monomers), reinforced by additional electrostatic contacts (**Figure 5E**).

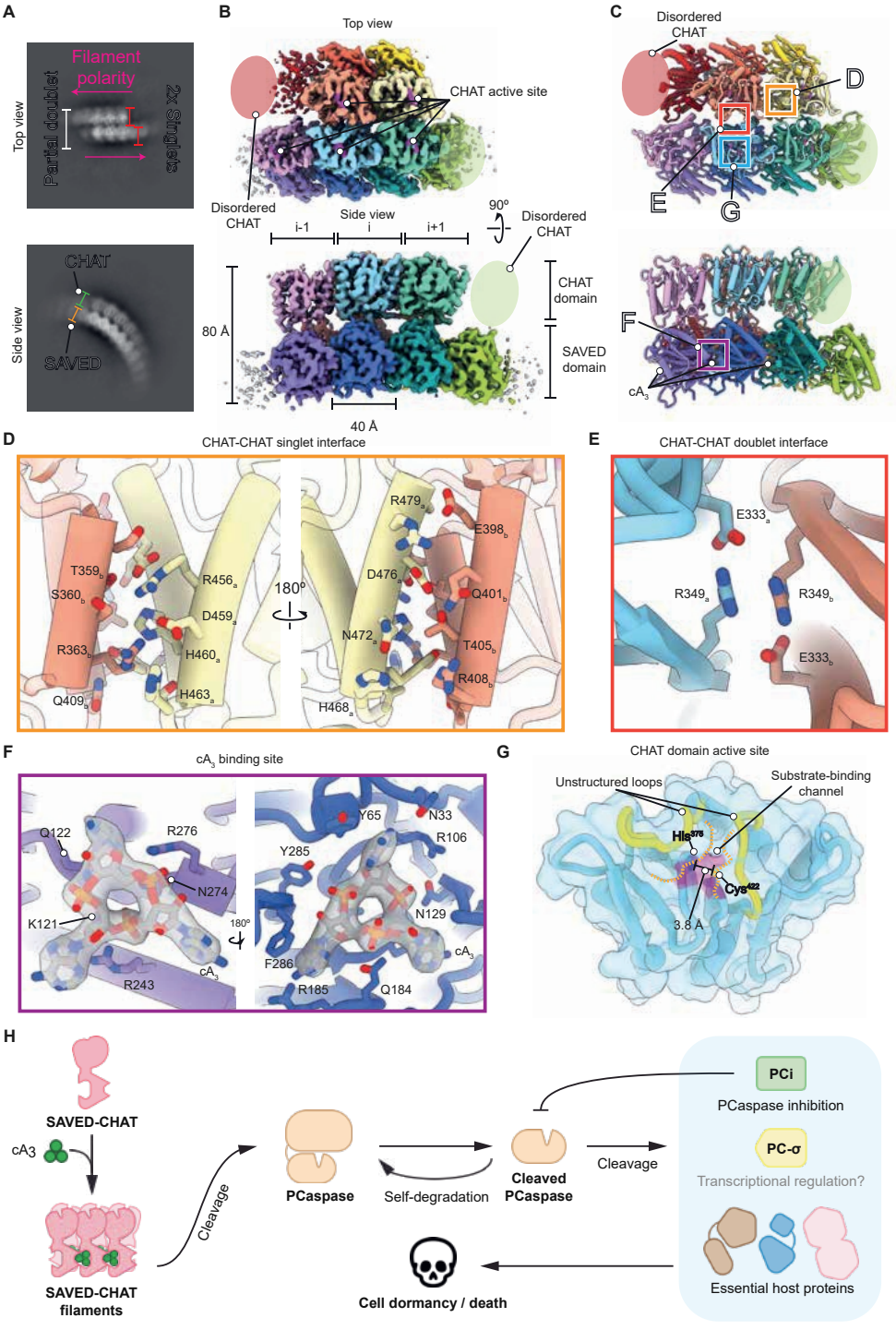

A
Filament polarity
Top view
Partial doublet
2x singlets
Side view
CHAT
SAVED
B
Top view
CHAT active site
Disordered CHAT
Side view
i-1
i
i+1
90°
Disordered CHAT
CHAT domain
SAVED domain
80 Å
40 Å
C
Disordered CHAT
D
E
G
F
cA3
D
CHAT-CHAT singlet interface
T359b
S360b
R363b
Q409a
R456a
D459a
H460a
H463a
180°
R479a
D476a
N472a
H468a
E398b
Q401b
T405b
R408b
E
CHAT-CHAT doublet interface
E333a
R349a
R349b
E333b
F
cA3 binding site
Q122
R276
N274
K121
R243
cA3
180°
Y65
N33
R106
Y285
N129
cA3
F286
R185
Q184
G
CHAT domain active site
Unstructured loops
Substrate-binding channel
His376
Cys422
3.8 Å
H
SAVED-CHAT
cA3
SAVED-CHAT filaments
Cleavage
PCaspase
Self-degradation
Cleaved PCaspase
Cleavage
PCi
PCaspase inhibition
PC-σ
Transcriptional regulation?
Essential host proteins
Cell dormancy / death

Figure 5. Structural basis for SAVED-CHAT activation by cA_3. (**A**) Representative 2D class averages of SAVED-CHAT bound to cA_3, showing side and top views. In the side view, the filament curvature is evident, with the CHAT domain at the periphery of the arch. In the top view, singlet filaments form a partial inter-filament doublet, with two singlets running in opposite polarities forming cross-fibre contacts spanning ~3 monomers. (**B, C**) 3.1 Å-resolution cryo-EM reconstruction and model of the cA_3-bound SAVED-CHAT filament. In some monomers, the CHAT domain is disordered and absent from the reconstruction (red and light green monomers). (**D**) Close-up view of the CHAT-CHAT singlet intra-filament interface, consisting of a four-helix bundle. (**E**) Close-up view of the CHAT-CHAT doublet inter-filament interface. (**F**) cA_3 binding site, at the interface between adjacent SAVED domains. Cryo-EM density for cA_3 is shown as translucent grey. (**G**) CHAT domain active site. The residues H375 and C422 that comprise catalytic dyad (magenta) are 3.8 Å apart and located within the substrate-binding channel beneath two unstructured gating loops (yellow). (**H**) Binding of cA_3 to the SAVED domain of SAVED-CHAT induces oligomerization that activates its CHAT domain. Activated SAVED-CHAT cleaves PCaspase, which subsequently mediates further downstream events by cleaving PCi (inhibitor of PCaspase), PC-σ (transcriptional response?) and/or other essential host proteins. Hypothesized events are in grey.

The cA_3 is buried within the intra filament interface between two SAVED domains, and participates extensively in hydrogen bonding, electrostatic and stacking interactions, similar to its role in other SAVED domains (**Figure 5F, S20**). This network of contacts suggests that cA_3 acts as a molecular glue to bridge SAVED domain intra-filament interactions (**Figure S19**), which subsequently provides a platform for CHAT-CHAT inter-domain interactions that allosterically activates protease activity. This is a distinct mechanism of activation from the TIR-SAVED filaments which assemble a composite active site across two adjacent TIR domains within a filament upon cA_3 binding, or the CRISPR-associated Lon protease (CalpL) activation that appears to have a concentration-dependent effect, reflecting the mechanistic diversity of cOA-responsive effector proteins [87,93].

Within the resolved doublet CHAT domains, the active site is positioned underneath two unstructured loops, which creates a narrow substrate-binding channel (**Figure 5G**). This channel is wide enough to accommodate unstructured peptides but would likely be inaccessible to highly structured peptides. The residues of the H375-C422 catalytic dyad is aligned and positioned ~4Å apart, confirming that the cA_3-bound filament represents the active conformation of SAVED-CHAT. Comparison with the CHAT domain of the activated type III-E protein TPR-CHAT bound to proteolytic substrate Csx30, revealed a near identical active-site configuration in which the catalytic dyad is aligned and the substrate-binding channel is exposed (**Figure S18B**). We propose that SAVED-CHAT activation may occur through a conformational change in the CHAT active site, where autoinhibition is allosterically alleviated through interdomain CHAT-CHAT contacts that occur after cA_3-induced oligomeri-

zation. This is consistent with the well-established activation mechanisms of eukaryotic caspases, which are inhibited as monomers and become allosterically activated upon dimerization [331].

Discussion

Caspases are a family of cysteine proteases that generally use a catalytic cysteine residue to target proteins by cleaving a peptide bond. Caspases in animals typically cut after an aspartate, whereas caspases in fungi and plants (metacaspases) cut after arginine or lysine [332–337]. In eukaryotes, caspases are key players in signal transduction pathways that often trigger programmed cell death (apoptosis) [337,338]. These apoptotic pathways consist of two classes of caspases: initiators and executioners, with the former proteolytically activating the latter to trigger further downstream events [331]. Our work highlights a conceptual similarity on this theme in prokaryotes: a cOA-sensory SAVED domain that activates a caspase-like CHAT domain (analogous to an initiator), which in turn activates a second caspase-like protease PCaspase (analogous to an executioner) that acts as the central node causing a cascade of events. PCaspase appears to modify downstream targets that promote dormancy and/or cell death *in vivo*.

The model in **Figure 5H** summarizes our present understanding of the *H. ochraceum* type III-B CRISPR-Cas system and its effector components, in which detection of target RNA results in cA_3 generation (not shown in the schematic) [73,83]. The synthesized cA_3 binds to the sensory domain of SAVED-CHAT and acts as a molecular glue to stabilize SAVED:SAVED dimerization (**Figure 5**). The subsequent multimerization of the unusual antiparallel SAVED-CHAT doublet filament allosterically activates the CHAT domains, an unprecedented mechanism for CARF and SAVED effector proteins (**Figure 5**). The activated CHAT domains in turn cleave PCaspase, which becomes an active, promiscuous protease that cleaves a wide range of substrates. In contrast to the substrate specificity of SAVED-CHAT, PCaspase has a broader substrate repertoire, as reflected by the cleavage of PC-σ, PCi, and other unrelated proteins and fluorophore-quencher peptides in vitro (**Figure 2D, 3, S10**). This is further supported by the strong abortive infection like phenotype that is observed when PCaspase is activated in vivo (**Figure 4**), where cleavage of host factors most likely leads to the significantly reduced transformation efficiency.

We found that PCi is a potent inhibitor of PCaspase both in vitro and *in vivo*, which hints at an auto regulated feedback loop that specifically controls PCaspase activity (**Figure 2**). Furthermore, the predicted sigma factor PC-σ might be involved in coordinating additional defence strategies through transcriptional regulation (**Figure 5H**), as has been suggested for CASP-σ in type III-E systems [322]. In that study, it was

shown that CASP-σ is part of a sub-complex together with Csx30 (a hypothesized toxin) and Csx31, which both lack any similarity to either PCi or PCk (**Figure S21**). Upon target RNA binding by the type III-E complex, the associated protease TPR-CHAT cleaves Csx30 and releases CASP-σ for transcriptional regulation of target genes that may include Cas1, which hints at a connection between interference and spacer adaptation. Another recent work on a type III-B system showed that CalpL (a Lon-like protease fused to a SAVED domain) forms a complex with CalpT (a potential anti-sigma factor) and CalpS (a sigma factor) [93]. Upon cOA signalling, CalpL oligomerizes and subsequently cleaves CalpT, releasing a subcomplex consisting of a CalpT fragment and CalpS that is thought to regulate gene expression by RNA polymerase.

Probably owing to toxicity, all attempts to (co)-purify PCk or include it in the in vivo experiments failed. We hypothesize that activated PCaspase might proteolytically modify and activate PCk, thereby activating another signal transduction pathway. Bacterial serine-threonine kinases act in stress response and anti-phage defence by phosphorylating several target proteins, and in some cases hampering growth [339,340]. Hence, it might be that the anticipated cytotoxicity of PCk reflects its serine-threonine kinase activity, which may instigate abortive infection, either directly in response to a PCaspase-mediated cleavage, or indirectly through transcriptional activation by PC-σ. Another remaining question is how PCaspase activity is suppressed by PCi. We speculate that this negative-feedback mechanism may allow for downregulating the stress response to a reversible dormancy state, rather than going irreversibly to cell death.

In recent years, an unprecedented diversity of type III CRISPR-Cas cOA-based signal transduction pathways have been described, but many details remain to be uncovered [75,264]. Most notably, the full impact of further downstream events after activation of CRISPR-associated proteases is not fully understood and is very likely to vary between different CRISPR-Cas systems. The variant described in this work is an intriguing example of evolutionary tinkering of prokaryotic immune systems that transduce stress signals through a set of caspase-like proteases to trigger a strong defence response.

Acknowledgments

We would like to thank Rodrigo Fregoso Ocampo for assistance with negative stain EM imaging.

Funding

Dutch Research Council (NWO) VIDI grant VI.Vidi.203.074 (RHJS)
Dutch Research Council (NWO) Spinoza grant SPI 93-537 (JvdO)
European Research Council (ERC) Advanced grant ERC-AdG-834279 (JvdO)
European Research Council (ERC) CoG grant 817834 (TJGE)
Dutch Research Council (NWO) VICI grant VI.C.192.016 (TJGE)
Volkswagen Foundation grant 96725 (TJGE)
National Institute of General Medical Sciences of the National Institutes of Health R35GM138348 (DWT)
Welch Foundation Research Grant F-1938 (DWT)
Robert J. Kleberg, Jr. and Helen C. Kleberg Foundation Medical Research Grant (DWT)
American Cancer Society Research Scholar RSG-21-050-01-DMC (DWT)

Author contributions

Conceptualization and design: JAS, JvdO, RHJS
Gene cluster and phylogenetic analyses: SK, TJGE
Experiments and analyses: JAS, JPKB, CRPS, CY, AMA, SHPP, ASA, CP, AnB, AB, RAS
Writing: JAS, JPKB, JvdO, SK, TJGE, DWT, RHJS

Competing interests

JAS and SHPP are founders and shareholders of Scope Biosciences B.V.. JvdO and RHJS are shareholders and members of the scientific board of Scope Biosciences B.V.. JvdO is a scientific advisor of NTrans Technologies and Hudson River Biotechnology. JAS, SHPP, JvdO, and RHJS are inventors on CRISPR-Cas related patents/patent applications. JPKB, CRPS, CY, AMA, SK, ASA, CP, AnB, AB, RAS, TJGE and DWT declare that they have no competing interests.

Data and materials availability
All data is available in the manuscript. The raw mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD046897.

Supplementary Materials

Materials and Methods

Computational identification of SAVED-CHAT/PCaspase gene clusters

We downloaded predicted protein coding sequences for all species representatives of archaea and bacteria from the genome taxonomy database (GTDB) (PMID: 30148503) release 207(https://data.gtdb.ecogenomic.org/releases/release207/207.0/genomic_files_reps/gtdb_proteins_aa_reps_r207.tar.gz). According to annotated Pfam (PMID: 33125078) domains in PCaspase (locus tag Hoch_1323) and SAVED-CHAT (locus tag Hoch_1319), we searched with Pfam hidden Markov models (HMM) of Caspase (PF00656), SAVED (PF18145) and CHAT (PF12770) domains against the species representative proteomes with the "hmmsearch" program in HMMER v3.1b2 (PMID: 22039361) with default parameters. The identified candidate proteins were reannotated using InterProScan v5.50-84 (PMID: 24451626) with Pfam domains (-appl Pfam) to remove false positive hits. We screened 62,291 bacterial and 3,412 archaeal species representative genomes for domain fusions of SAVED and CHAT, and a caspase domain within 10 genes distance. For each of the three Pfam domains, we defined candidate gene clusters. Final gene clusters were formed by merging overlapping candidate gene clusters and annotated with Pfam domains. Gene clusters were visualized with GenoPlotR v0.8.11 (PMID: 20624783).

Phylogenetic analysis of the SAVED-CHAT/PCaspase gene cluster proteins

For *SAVED-CHAT* and *PCaspase*, we performed domain specific phylogenetic analysis. We extracted all regions with these domain annotations from the GTDB prokaryotic species representative genome dataset. We additionally performed Pfam domain annotation as described above for the eukaryotic protein sequences in the eggNOG v5 (PMID: 30418610) database (http://eggnog5.embl.de/download/eggnog_5.0/e5.proteomes.faa) and extracted the respective protein domains. We then merged prokaryotic and eukaryotic protein domains and performed multiple sequence alignment (MSA) with FAMSA v1.6.2 (PMID: 27670777). After gentle trimming with trimAl (PMID: 19505945) v1.4.rev15 (-gt 0.02), we calculated approximate maximum likelihood (ML) trees with FastTree v2.1.11 (PMID: 20224823). We extracted focal clades for respective *H. ochraceum* proteins in the two gene clusters. The respective focal clade sequences were aligned with MAFFT L-INS-i (PMID: 23329690) v7.427 and we inferred ML trees with IQ-TREE (PMID: 32011700) v2.0.3 (-m MFP –bb 1000). For *PC-σ* and *PCk* (locus tags: Hoch_1320 and Hoch_1322, respectively) we searched the GTDB representative species protein dataset with MMseqs2 (PMID: 29035372) v14.7e284 (--max-seqs 2000) to identify

the highest scoring homologs. We aligned sequences with FAMSA, trimmed MSAs with trimAL (-gt 0.02) and calculated trees with FastTree. Trees were visualized with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

PCaspase and CHAT domain alignments

We extracted PCaspases that co-occur in a genome neighbourhood with SAVED-CHAT and performed full length protein multiple sequence alignment using MAFFT L-INS-I as described above.

We likewise extracted the predicted CHAT domains from proteins containing a SAVED and CHAT domain in the same genomic neighbourhood as a PCaspase. We extracted only the CHAT domains for analysis, as the domain order of SAVED and CHAT domain in these proteins varies and makes full protein multiple sequence alignment infeasible (i.e., CHAT can be the N- or C-terminal domain). We performed structure guided multiple sequence alignment on ESMfold (PMID: 36927031) protein structure models of the respective CHAT domains with PROMALS3D (PMID: 18287115). Multiple sequence alignments were visualized with Jalview v2 (PMID: 19151095).

Plasmids and strains for protein purification

Electrocompetent *E. coli* DH5α or DH10B (Invitrogen) were used for general cloning and plasmid propagation. Full-length genes of SAVED-CHAT, PCaspase, PC-σ and PCi from *H. ochraceum* DSM 14365 (NCBI reference sequence NC_013440.1) were codon-harmonized for expression in *E. coli* BL21(DE3) or BL21-AI and synthesized with an N-terminal Strep-tag II by either Integrated DNA Technologies (IDT) or Twist Bioscience. The synthetic genes were then cloned into the expression vector pJS-BCD under an IPTG-inducible T7 promoter. SAVED-CHAT and PCaspase mutants were generated by PCR mutagenesis. All expression plasmids (**Table S2**) were transformed into electrocompetent *E. coli* BL21(DE3) or BL21-AI for protein expression and purification.

Protein purification

For recombinant expression and purification of SAVED-CHAT, PCaspase (and mutants thereof), PC-σ and PCi, the expression vector pJS-BCD was used (**Table S2**). *E.coli* BL21(DE3) or BL21-AI cells transformed with the expression vector were inoculated in 1.5 L lysogeny broth (LB) and grown at 37°C with 140 rpm shaking. After an optical density (O.D. 600 nm) of 0.6-0.7 was reached, a cold shock on ice for 30-60 min was performed. Protein expression was induced by adding 0.5 mM

IPTG (isopropyl β-D-1-thiogalactopyranoside) in the case of BL21(DE3) or 0.2% arabinose in the case of BL21-AI. Subsequently, cells were grown at 16 °C for 18-24 hours and harvested via centrifugation. Pelleted cells were resuspended in 20-30 mL Wash Buffer (150 mM NaCl, 100 mM Tris, pH 8.0) and a cOmplete™ protease inhibitor tablet (Roche) was added. Cells were lysed by sonication and lysate was clarified by centrifugation and subsequent filtration (0.45 μM). The supernatant was applied to a StrepTrap XT 5 mL column (Cytiva) and the bound protein was eluted with Elution Buffer (Wash Buffer + 50 mM biotin). Finally, a Superdex 200 (Cytiva) size exclusion chromatography column was used with SEC Buffer (500 mM NaCl, 100 mM Tris, pH 8.0, 10% glycerol) as eluate. After confirmation with SDS-PAGE analysis, protein fractions were pooled and concentrated to a working stock of 5 μM.

Protein cleavage activity assay

All cleavage assays were performed in a buffer containing 125 mM NaCl, 10 mM Tris-HCl (pH 8.0), and 1 mM DTT. For SAVED-CHAT, PCaspase (and mutants thereof), PC-σ, and PCi, a final protein concentration of 0.5 μM was added, unless otherwise specified. Depending on the assay, a final concentration of 1 μM cA_3 (Biolog.de), 5 mM EDTA, 5 mM EGTA, 2 mM $MgCl_2$, 2 mM CaCl, 1 mM ATP and 0.1 mg/mL casein was added. After incubation for 1h at 35°C, SDS loading dye was added and incubated for a further 5 min at 95°C. SDS-PAGE analysis using Coomassie blue staining was utilized to visualize the results (Biorad Gel Doc XR).

SAVED-CHAT complex formation on native PAGE

For native PAGE analysis of SAVED-CHAT (and mutants thereof) complex formation, 2.5 μM SAVED-CHAT was incubated with or without 1 μM cA_3 for 1h at 35°C in a buffer containing 125 mM NaCl, 10 mM Tris-HCl (pH 8.0), and 1 mM DTT. Afterwards, the reaction was run on a native 4-20% polyacrylamide gel, stained with Coomassie blue, and visualized (Biorad Gel Doc XR). To better visualize intermediate complex formation, the incubation time was shortened from 1h to 15 min.

Real-time PCaspase activity assay

SAVED-CHAT/PCaspase activity assays were conducted in vitro in activity buffer (125 mM NaCl, 10 mM Tris, 1 mM DTT, pH 8.0) to which different combinations of component were added: SAVED-CHAT (0.5 μM), PCaspase (0.5 μM), cA_3 (15.6 nM to 1 μM), and/or FAM-peptide substrate (5 μM) (Eurogentec AS-60579-01). Assays were incubated for one hour at 37C with a FAM channel measurement at 1 min intervals in a Thermo Scientific Quantstudio 1 RT-qPCR instrument running Quantstudio Design & Analysis software (v1.5.2). Data was visualized using either Graphpad

Prism 9 (n=3) or R version 4.3.0 with standard error depiction.

Bottom-Up Proteomics / In-Gel Digestion

SAVED-CHAT, PCaspase, and PCi were combined in a cleavage reaction in the same manner as the protein cleavage activity assay and ran on SDS-PAGE. Bands from the gel were excised, washed with water, and then treated with Acetonitrile for shrinking. Subsequently, the bands were reduced using 6.5 mM DTT at 55 °C for 1 hour, followed by alkylation in 54 mM IAA at room temperature for 30 minutes in the dark. Both DTT and IAA were dissolved in 50 mM Ammonium bicarbonate pH 8.5. After further shrinking the gel pieces with acetonitrile, trypsin was added, and the mixture was kept on ice for 30 minutes. Excess trypsin was removed, and the gel pieces were covered with cold ammonium carbonate pH 8.5 and incubated overnight at 37 °C. The resulting supernatant was transferred to new Eppendorf tubes. Peptides were extracted by adding 100% acetonitrile to the gel pieces. The two peptide extracts were combined and dried using a vacuum centrifuge. The recuperated peptides were dissolved in 40 µl of 2% formic acid before undergoing

LC-MS/MS analysis

The digested peptides were separated through online reversed-phase chromatography using a Dionex UltiMate 3000 system (Thermo Fisher Scientific). Initially, peptides were trapped and washed for 1 minute on an Acclaim Pepmap 100 C18 cartridge (5 mm × 0.3 mm, 5 µm) in solvent A (0.1% v/v formic acid in water) at a flow rate of 30 µl/min. Subsequently, the peptides were separated on an analytical column (packed with Poroshell 120 EC C18; dimensions 50 cm × 75 µm, 2.7 µm, Agilent Technologies) with a flow rate of 300 nl/min. The column was housed in a compartment maintained at 40 °C and was coupled to a Thermo Scientific Exploris 480 mass spectrometer (Thermo Scientific, Bremen, Germany). Sample elution occurred over a 50-minute gradient. Initially, the gradient was kept at 9% B (0.1% v/v formic acid in 80% acetonitrile) for 1 minute. Then, the percentage of B was ramped from 13% to 44% over 37 minutes, followed by an increase from 44% to 55% over 5 minutes, and finally from 55% to 99% in 1 minute, which was maintained for 5 minutes. Peptide analysis was conducted with a resolution setting of 60,000 in MS1. MS1 scans were obtained with standard automatic gain control (AGC) target and injection time settings, within a scan range of 375–1,600 Th. Precursors were selected with a 1.4 Th window and fragmented with a normalized collision energy of 28%. MS2 scans were acquired at a resolution of 15,000, with standard AGC target and maximum injection time settings, starting from 120 Th.

The RAW data was processed using MaxQuant with standard settings except for a few [341]. Specifically, a FASTA file containing the proteins of interest was employed, automatically supplemented by contaminant proteins during analysis. Fixed modifications was set to Carbamidomethyl on Cysteine, while variable modifications was set to Acetylation on the Protein N-term and Oxidation on Methionine. To enable absolute quantitation, Match Between Runs, Label-Free Quantification and iBAQ was activated. The resulting protein groups table was imported into Excel, and bar charts were generated using the iBAQ columns. Additionally, the sequence coverage plot was generated using FragmentLab (freely available at https://scheltemalab.com/software/software-fragmentlab/).

Top down LC-MS/MS

SAVED-CHAT, PCaspase, and PCi were combined in a cleavage reaction in the same manner as the protein cleavage activity assay. The reaction was halted by heating to 95 °C for 5 minutes, followed by acidification of the solution to 10% formic acid. Reversed-phase liquid chromatography was performed using a Thermo Scientific Vanquish Neo UHPLC instrument, featuring a 0.15 mm x 150 mm MAbPac RP analytical column, directly linked to an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Scientific, San Jose, CA, USA). To enhance chromatographic separation, the analytical column was heated to 60 °C in a column compartment. Post-reaction, samples were separated over 22 minutes at a flow rate of 1 µL/min. Gradient elution was achieved using two mobile phases: A (0.1% formic acid in Milli-Q) and B (0.1% formic acid in acetonitrile). Mobile phase B was held constant at 20% for 2 minutes, then increased from 20% to 45% over 15 minutes, followed by a step to 95% over 1 minute, maintained for 4 minutes. Mass spectrometric data were acquired with the instrument operating in Intact Protein and Low-Pressure mode. The spray voltage was set to 2.0 kV, with a capillary temperature of 300 °C, and source-induced dissociation set at 15 V. Two MS methods were utilized for charge state assignment and accuracy, with resolutions set at either 7,500 or 120,000 (both at m/z = 200 Th). Individual proteins or reaction mixtures were recorded at a resolution of 7,500 (@ m/z 200) in MS1 for better detection of charge distributions of large proteins (> 30 kDa). Reacted Caspase proteins were analysed at a resolution of 120,000 (@ m/z 200) in MS1 for more precise mass detection of smaller proteins (< 30 kDa). Low-resolution MS1 scans were recorded with a m/z range of 500-3,000 Th, with a 250% AGC target and maximum injection time set to either 100 ms for the 7,500 resolution or 50 ms for the 120,000 resolution, with 5 µscans recorded for the 7,500 resolution and 1 µscans for the 120,000 resolution per scan. MS/MS scans were acquired at a resolution of 120,000, with a maximum injection time of 250 ms, a 200% AGC target, and 5 µscans averaged per scan. The ions of interest were mass-selected by a quadrupole in a 2 Th isolation window, or targeted mass was selected when nec-

essary, and accumulated to the AGC target prior to fragmentation. Electron-transfer dissociation (ETD) was performed with the following settings: 8 ms reaction time, a maximum injection time of 200 ms, and an AGC target of 1e6 for the ETD reagent. For the data-dependent MS/MS acquisition strategy, the intensity threshold was set to 2e4 of minimum precursor intensity. MS/MS scans were recorded in the range of m/z = 350-5,000 Th using high mass range quadrupole isolation.

Fragmentation spectra were extracted using FlashDeconv version 3.0.0 with standard settings, except for the merging method parameter, which was set to 1 to enable moving Gaussian averaging. The obtained MS2 data files were processed using an in-house developed script that generated a protein sequence for every position along the protein backbone, allowing a maximum precursor tolerance of 50 Da. The script then generated theoretical fragments from each sequence and compared them with the recorded spectra. The highest-scoring matches were subsequently analysed using ProSight Lite [342].

Plasmid interference assay

To assess SAVED-CHAT/PCaspase-mediated in vivo response, electrocompetent *E. coli* BL21-AI (Invitrogen) carrying pHochTypeIII and one of the pEffectors (**Figure S9A, B**) were prepared for target/non-target plasmid transformation (**Table S1**). For pHochTypeIII, *H. orchaceum* type III-B *cmr1-6, csb2*, and a minimal CRISPR array carrying one spacer sequence were placed individually under the control of T7 promoters. For the pEffectors, the various effector genes (and mutants or fragments thereof) were cloned into one operon and placed under the control of a constitutively expressing lacUV5 promoter. For the target/non-target plasmids, a non-coding RNA was placed under the control of a trc promoter and lacO (**Figure S9C**). This non-coding RNA carries the protospacer targeted by the spacer of the crRNA guide encoded by the CRISPR array on pHochTypeIII.

The transformations were carried out in biological triplicates, using 100 ng of target or non-target plasmid, by electroporation (BTX electroporation system). After transformation, all cells were recovered in 1 mL of LB at 37°C for 1h. Ten-fold dilutions were plated on LB agar medium containing 0.2% arabinose, 34 µg/mL chloramphenicol, 50 µg/mL carbenicillin, 50 µg/mL kanamycin, and either 1 mM IPTG or 0.2% glucose to induce or repress target RNA transcription, respectively. Finally, the plates were incubated overnight at 30°C, and transformation efficiencies were quantified.

To assess the activity of cleaved PCaspase fragments *in vivo*, genes coding for either PCaspase residues 1-153, 154-666, and their combination, were cloned into a pEffector coding for SAVED-CHAT as well (**Table S2**). The resulting pSAVED-CHAT_PCaspase fragments plasmids were transformed into BL21-AI in biological triplicates, using 5.25 fmol for each plasmid, and recovered in 1 mL of LB at 37°C for 1h. Ten-fold dilutions were plated on LB agar medium containing 50 µg/mL carbenicillin. Finally, the plates were incubated overnight at 37°C, and transformation efficiencies were quantified.

Data was analyzed and visualized using R version 4.3.0, with statistical significance calculated by one-sided unpaired Welch's t-test.

Structural analyses

10 µM SAVED-CHAT was mixed with 125 µM cA_3. 2.5 µl of complex was immediately applied to C-flat grids (1.2/1.3, 300 mesh) which had been plasma-cleaned for 30 seconds in a Solarus 950 plasma cleaner (Gatan) with a 4:1 ratio of O_2/H_2. Grids were blotted with Vitrobot Mark IV (Thermo Fisher) for 6 seconds, blot force 0 at 4ºC & 100% humidity, and plunge-frozen in liquid ethane. Data were collected on a FEI Glacios cryo-TEM equipped with a Falcon 4 detector. Data was collected in SerialEM, with a pixel size of 0.94 Å, a defocus range of -1.5 - -2.5 µm, and a total exposure time of 15s resulting in a total accumulated dose of 40 e/Å^2 which was split into 60 EER fractions. Motion correction, CTF estimation and particle picking was performed on-the-fly using cryoSPARC Live v4.0.0-privatebeta. Data were collected on a FEI Glacios cryo-TEM equipped with a Falcon 4 detector, as described for the binary complex. In total, 2,550 movies were collected. All subsequent data processing was performed in cryoSPARC v3.2. 1,229,738 particle co-ordinates were picked, selected and subjected to 2D classification. Multiple rounds of *ab-initio* modelling and heterogeneous refinement resulted in a subset of 39,001 particles, which yielded a 3.1 Å-resolution which was used for modelling. For modelling, multiple AlphaFold2 models of the individual SAVED and CHAT domains were rigid-body fitted into the map. cA_3 was built in Coot, and other adjustments were performed in Isolde, before running real-space refinement in Phenix. All structural figures and movies were generated using ChimeraX.

For negative stain EM experiments, grids were prepared by incubating 250 nM wild-type or mutant SAVED-CHAT with 5 µM cA_3 for 15 minutes on ice, before grid preparation, which was prepared as previously described, and imaged using a JEOL NEOARM operated at 200 kV, at a nominal magnification of 50,000x [343].

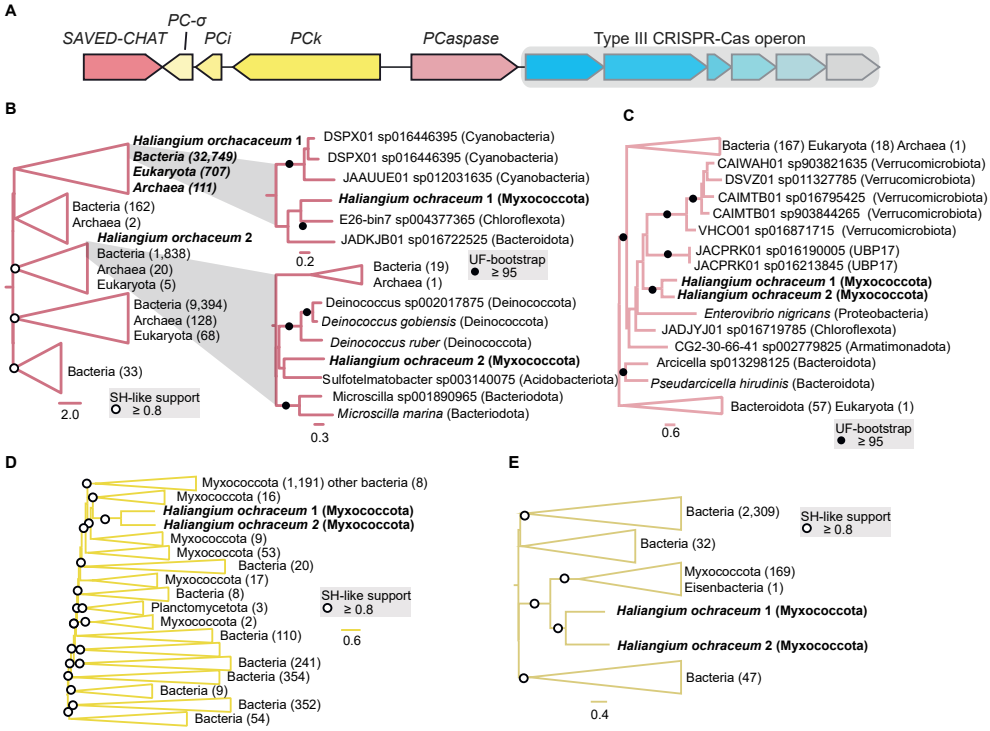


Figure S1. Second SAVED-CHAT/PCaspase like gene cluster and protein phylogenies. (**A**) Schematic representation of the second *H. ochraceum* type III CRISPR-Cas operon and its associated genes: *SAVED-CHAT, PC-σ, PCi, PCk*, and *PCaspase* (locus tags Hoch_5578-5588). (**B**) CHAT domain general and individual focal clade phylogenetic trees. (**C**) Phylogenetic tree of caspase domain focal clade. (**D, E)** Phylogenetic trees of *PCk* and *PC-σ*, respectively. All trees are midpoint-rooted, scale bars indicate substitutions per position in the alignment.

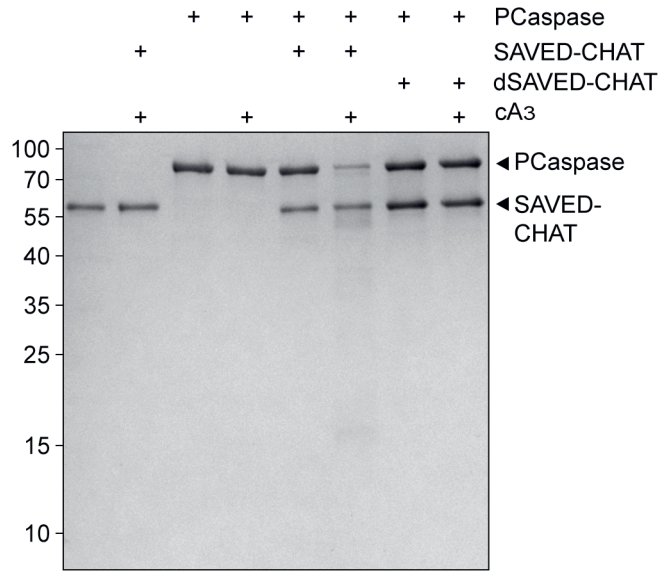


Figure S2. SDS-PAGE analysis of wildtype and catalytically dead SAVED-CHAT activity. PCaspase is specifically cleaved by cA_3-activated SAVED-CHAT, opposed to the SAVED-CHAT H375A/C422A catalytic mutant (dSAVED-CHAT).

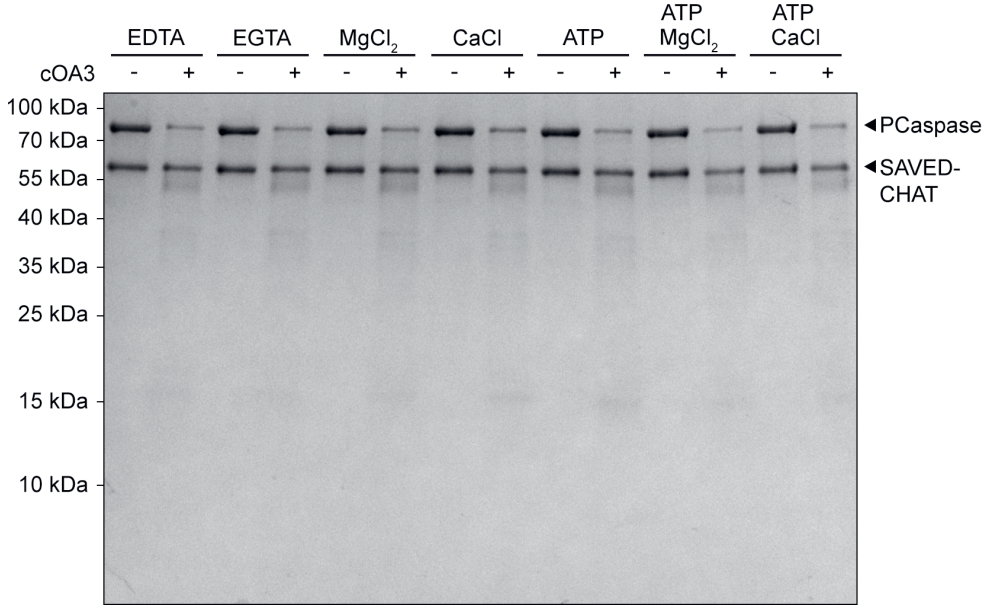


Figure S3. SDS-PAGE analysis of co-factor requirements of PCaspase cleavage by SAVED-CHAT. There is no apparent dependence on a co-factor or the catalytic activity of SAVED-CHAT on PCaspase when activated by cA_3.

Figure S4. Native PAGE analysis of SAVED-CHAT after incubation with cA_3. Oligomerization of SAVED-CHAT monomers hampers migration into a native PAGE gel in a cA_3-dependent manner.

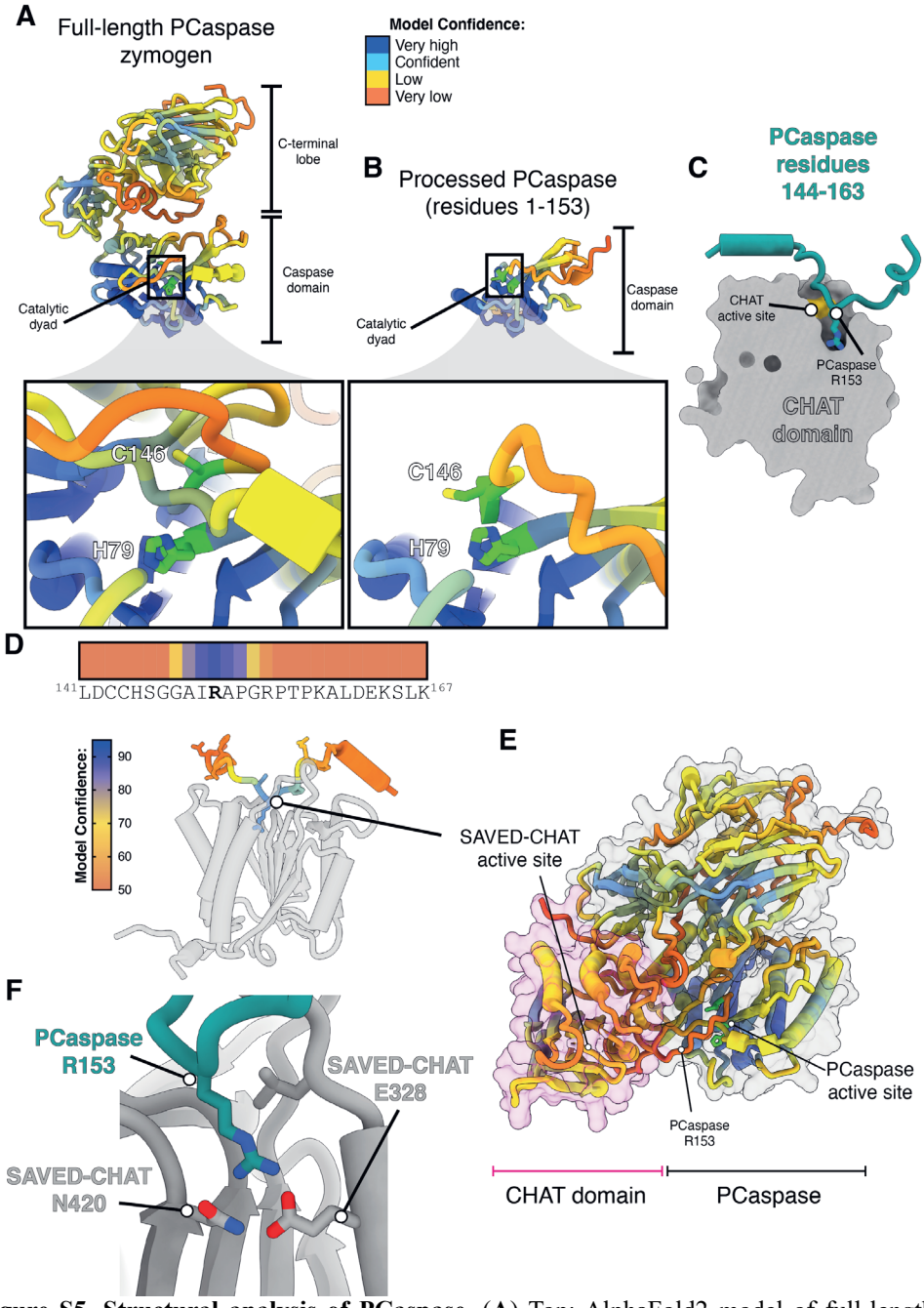
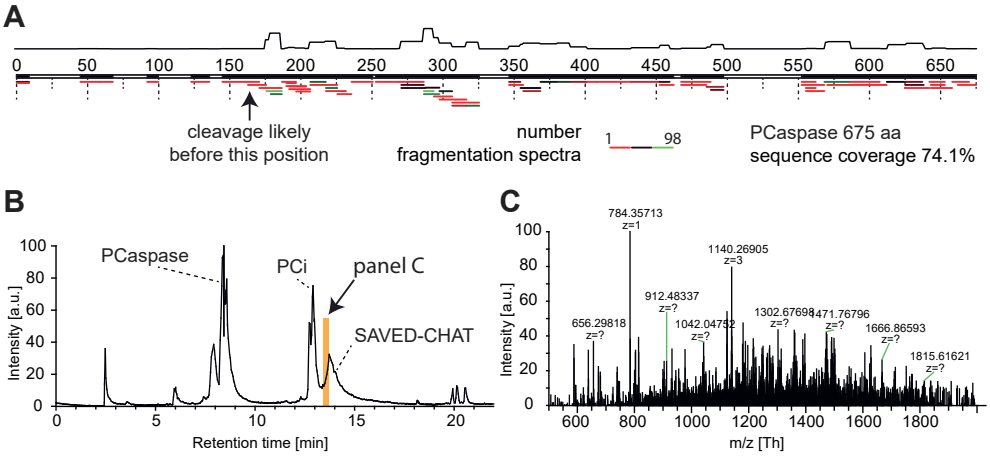


Figure S5. Structural analysis of PCaspase. (**A**) Top: AlphaFold2 model of full-length PCaspase, coloured by AF2 pLDDT score. Below: zoom-in of catalytic dyad, which is occluded by a flexible loop (residues 144-163). (**B**) Top: AlphaFold2 model of processed

PCaspase (residues 1-153). Below: zoom-in of catalytic dyad, which is now in a catalytically competent conformation. (**C**) AlphaFold2 multimer co-folding of the CHAT domain of SAVED-CHAT with PCaspase auto-inhibitory loop (residues 144-163). R153 is bound in a cleft, positioning the peptide within the CHAT domain active site. (**D**) AF2 model of PCaspase fragment bound to CHAT domain of SAVED-CHAT, with the PCaspase fragment coloured by model confidence. Only R153 and adjacent residues are predicted with high confidence (up to 89.3 for R153), indicating that the flanking regions of the peptide are likely disordered. (**E**) AF2 model of full-length PCaspase and SAVED-CHAT CHAT domain. The CHAT domain is predicted with low confidence, and is incongruent with the wedging of R153 in the CHAT domain active site. The low confidence score indicates that this model does not represent a state that allows SAVED-CHAT to cleave and activate PCaspase. (**F**) Interactions between PCaspase R153 and the CHAT domain from AF2 model in panel **C** and **D**.



D

```
  1 M⌊W S H P⌉Q⌉F⌉E⌉K⌉S⌉A⌉M⌉K⌉R⌉A⌉L⌉L I G A P⌉L A H⌉S  25
 26 A T H⌉A⌉L A G V A⌊N⌉D I D⌉R A G⌉E I L R⌉L Y⌉G F H  50
 51 C D D V L F G E R A⌊T⌉R⌉D⌉G I V A R L E A L I A E  75
 76 T Q A D D A V V I Y F S G H G G R V V N T G I⌊V⌊K 100
101⌊D⌉L P G G⌊D Y R P G⌊A H⌊Q F L V P⌊E⌊D⌊Y⌊Q P⌊K⌊A⌊Q 125
126⌊T⌊F⌊T⌊G⌊I⌊L⌊L⌊D⌊F⌊E L R Y⌊L V A⌊R L A A⌊R T E⌊N⌊V S 150
151 V I L⌉D C⌊C H S G G A I R 163
```

Figure S6. Shotgun and top-down proteomics profiling to determine the PCaspase cleavage site (by SAVED-CHAT). (**A**) Gel band annotated with "*" in **Figure 2D** was excised and subjected to independent shotgun proteomics analysis. A prominent indication of PCaspase cleavage was observed, based on detected C-terminal PCaspase peptides. The sequence coverage plot suggests an approximate position around amino acid 160 as the probable cleavage site. Although peptides were identified before this position, they are likely false positives, given the scarcity of fragmentation scans compared to the peptides identified after the cleavage point (see colour coding Number fragmentation spectra). (**B**) Topdown proteomics profiling of reaction mixture identical to the one used in panel **A**. Chromatography profile with the different intact proteins are indicated (peak splitting was observed for intact proteins analysed in isolation as well). Even though PCaspase is cleaved, the intact protein ionizes well and a sizable peak remains. The identified cleavage products co-elutes with SAVED-CHAT, compromising the analysis. Orange bar indicates elution area used for analysis in panel **C**. (**C**) By targeting the m/z values of PCaspase specifically, we were able to collect informative fragmentation spectra. (**D**) Annotation of the spectra provides high sequence coverage in both the N-terminal as well as the C-terminal direction (red brackets) for the cleavage product at position 163 (including the ten residues of the Strep-tag II).

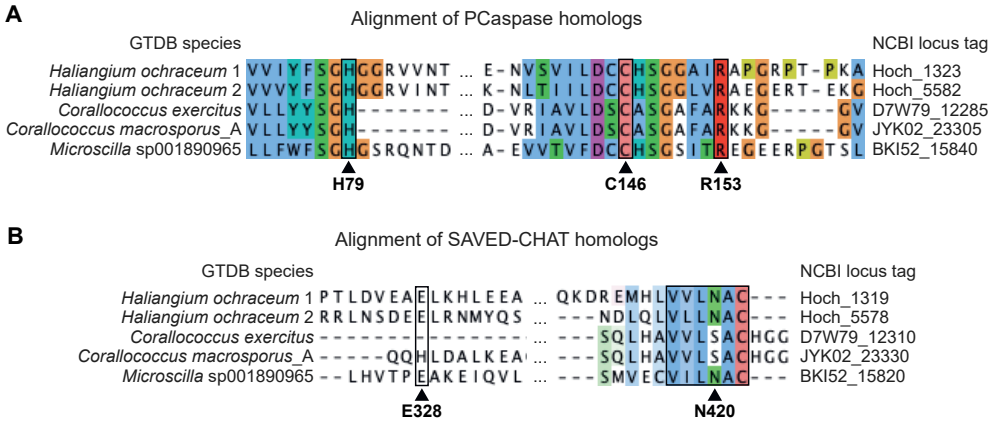


Figure S7. Alignment of PCaspase and SAVED-CHAT homologs co-occurring in other species. (**A**) Active site residues H79, C146, as well as the cleavage site residue, R153, are highly conserved in PCaspase homologs across species with a co-occurring SAVED-CHAT homolog. (**B**) E328 and N420 are conserved as pairs in SAVED-CHAT homologs in some species with a co-occurring PCaspase, whereas residues 417-422 were highly conserved.

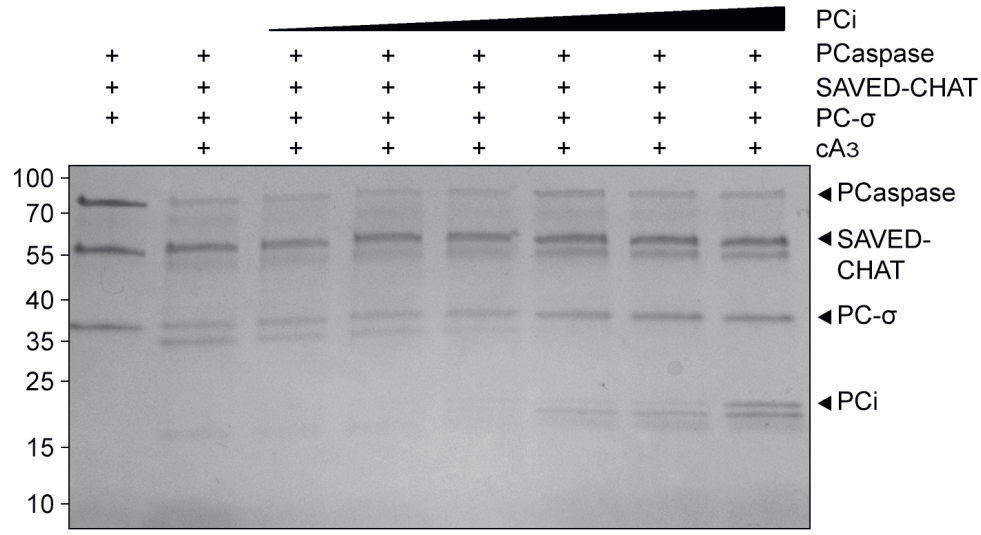


Figure S8. SDS-PAGE analysis of the inhibitory effects of PCi on PCaspase cleavage activity on PC-σ. An increasing concentration of PCi reduces the ability of PCaspase to cleave PC-σ, cleavage of PCi is also incomplete. Molar ratios of PCi:PCaspase range from 1:20 – 2:1.

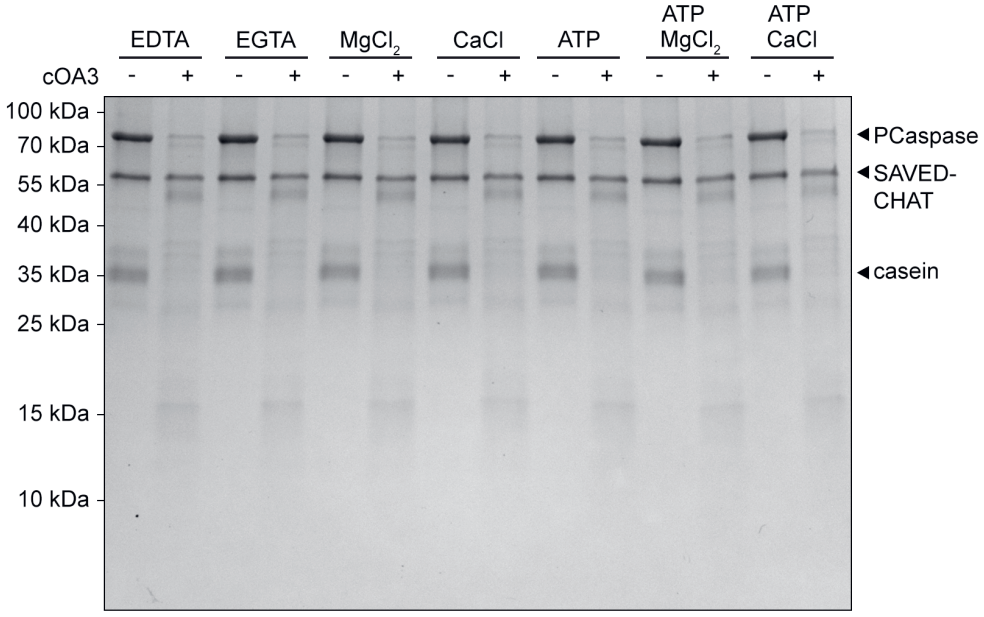


Figure S9. SDS-PAGE analysis of co-factor requirements of PCaspase cleavage activity of casein. SDS-PAGE protein cleavage assays with SAVED-CHAT, PCaspase and casein, demonstrating that the cleavage activity on casein of activated PCaspase does not require co-factors.

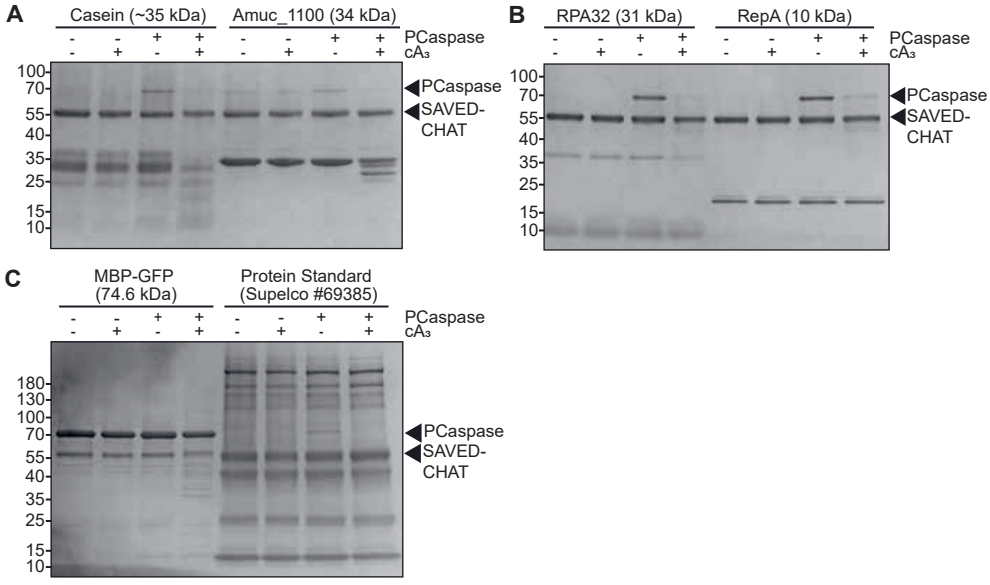


Figure S10. SDS-PAGE analyses of protease activity of SAVED-CHAT or PCaspase on various unrelated protein substrates. SAVED-CHAT was present in all reactions. (**A**) Casein and Amuc_1100 were cleaved when PCaspase is activated. (**B**) RepA was cleaved when PCaspase is activated, while RPA2 remained intact. (**C**) MBP-GFP was cleaved when PCaspase is activated, while a protein marker standard remained intact.

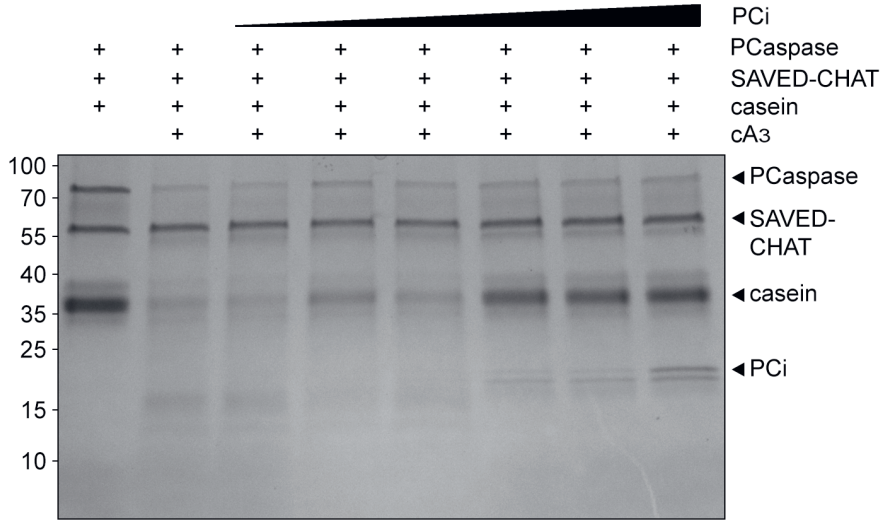


Figure S11. SDS-PAGE analysis of the inhibitory effects of PCi on casein cleavage by activated PCaspase. An increasing concentration of PCi reduces the ability of activated PCaspase to cleave casein and complete cleavage of PCi. Molar ratios of PCi:PCaspase range from 1:20 – 2:1.

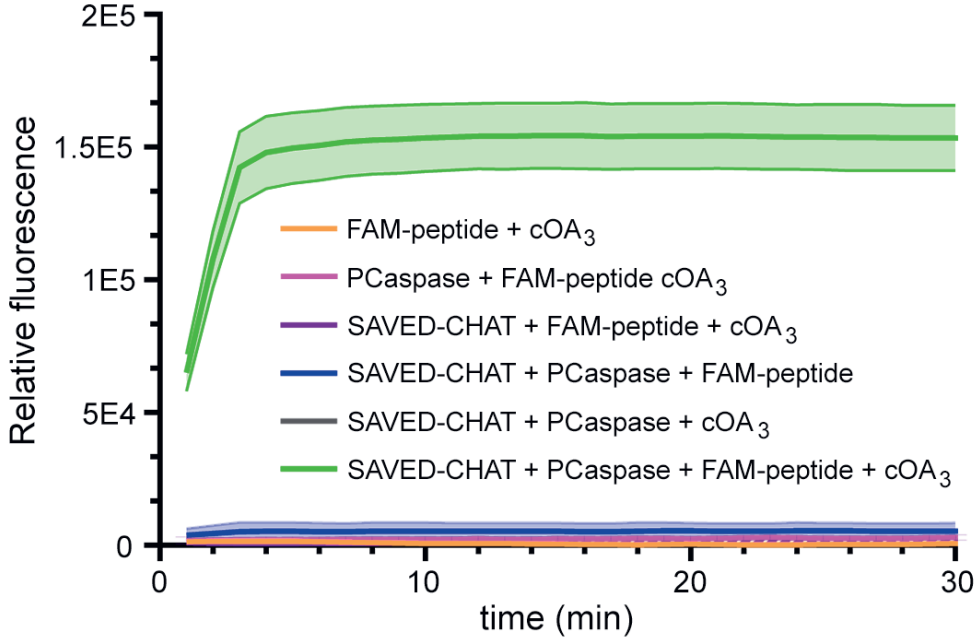


Figure S12. FAM-peptide is specifically cleaved by activated PCaspase. Various combinations of proteins in graph are represented by coloured lines. The transparent bands represent the standard error of the mean (technical replicates, n=3).

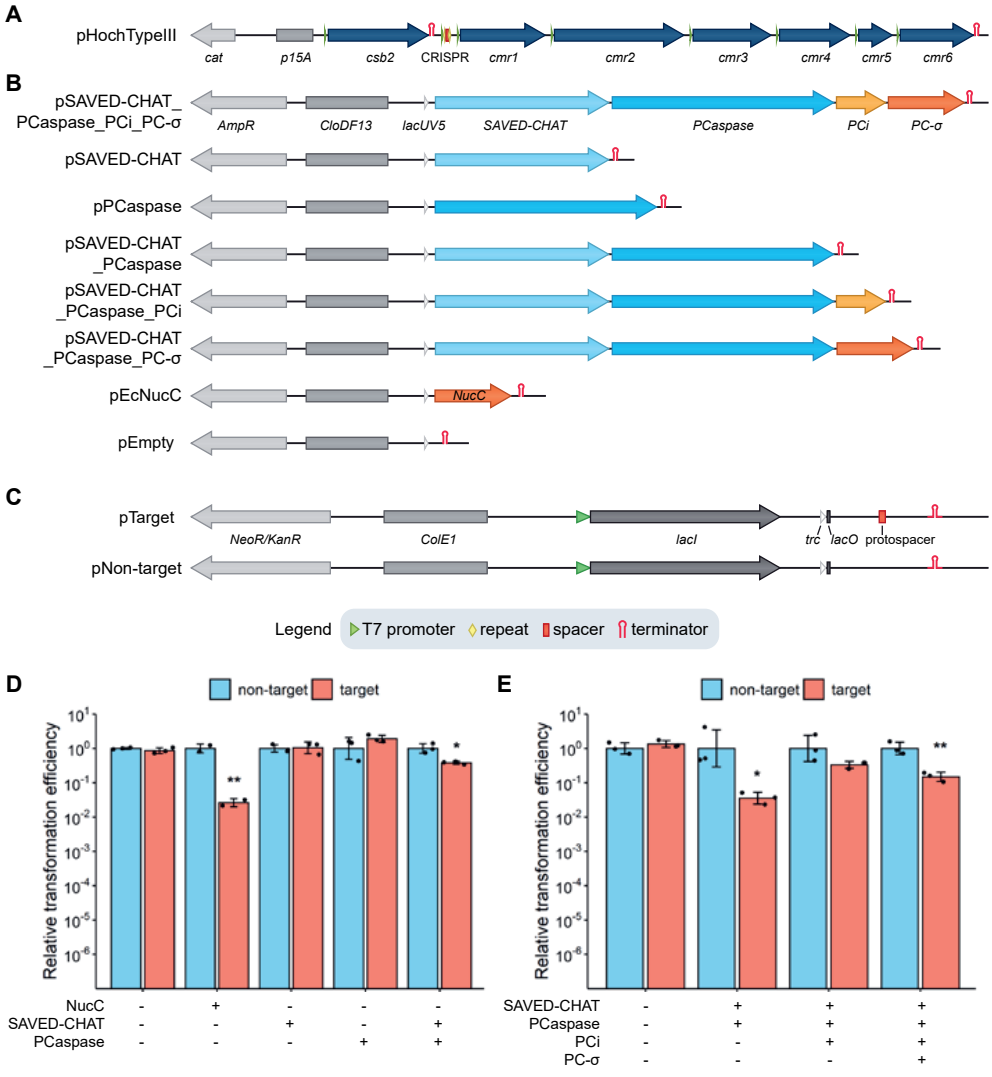


Figure S13. Linearized plasmid maps and transformation efficiencies with repressed protein expression for the plasmid challenge assay. (**A**) Map of the *H. ochraceum* type III CRISPR-Cas expression plasmid pHochTypeIII, with *Cmr1-Cmr6* from the *SAVED-CHAT* genomic neighbourhood, *csb2* from a co-occurring type I-G system, and the associated CRISPR array with a single spacer sequence targeting a protospacer on pTarget. (**B**) Maps of the effector expression plasmids, showing the different combinations of effectors used in the study. pEcNucC and pEmpty were used as positive and negative controls respectively. (**C**) Maps of the pTarget and pNon-target, having identical backbones except for a protospacer on pTarget. (**D**) Transformation efficiencies (relative to the non-target) of pTarget or pNon-target in *E. coli* with repressed expression of *H. ochraceum* type III CRISPR-Cas complex and different combinations of SAVED-CHAT and PCaspase. (**E**) Transformation efficiencies (rela-

tive to the non-target) with repressed expression of different combinations of SAVED-CHAT, PCaspase, PCi, and PC-σ. Statistical significance was calculated using one-sided unpaired Welch's t-test. * - $p < 0.05$, ** - $p < 0.005$, *** - $p < 0.0005$ (n=3).

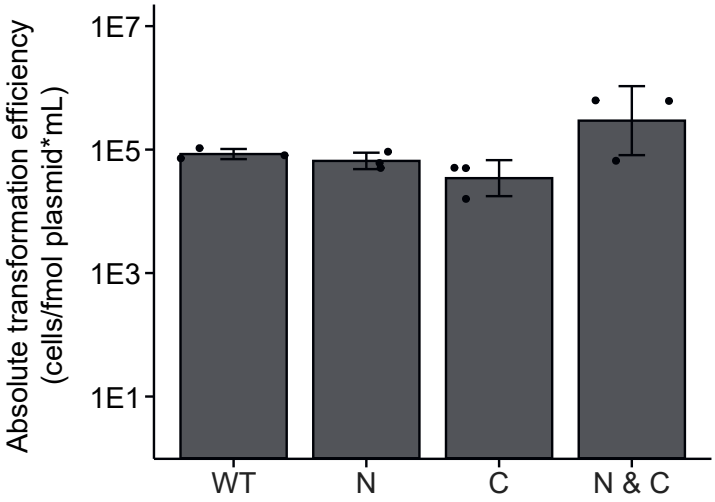


Figure S14. Expression of the post-cleaved PCaspase fragments in *E. coli*. Transformation efficiencies of plasmids constitutively expressing PCaspase ("WT"), or the N- (aa 1-153) and/or C-terminal (aa 154-666) PCaspase fragment(s).

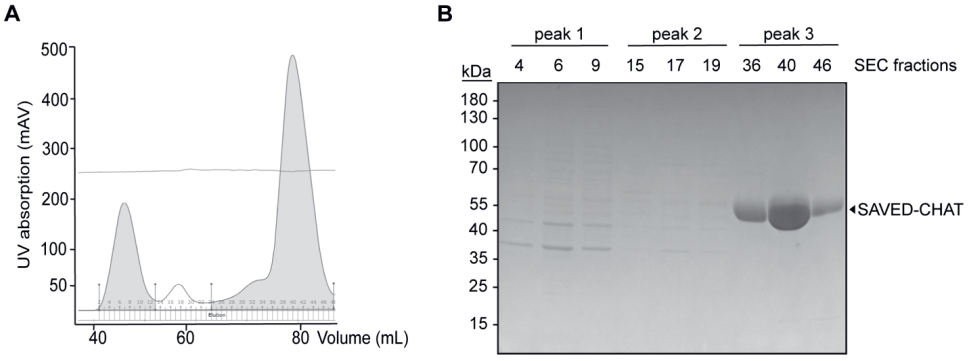


Figure S15. Size exclusion purification of SAVED-CHAT. (**A**) Chromatogram obtained from samples after affinity chromatography purification of SAVED-CHAT on a Superdex pg200 column. (**B**) SDS-PAGE analysis of the eluted fractions.

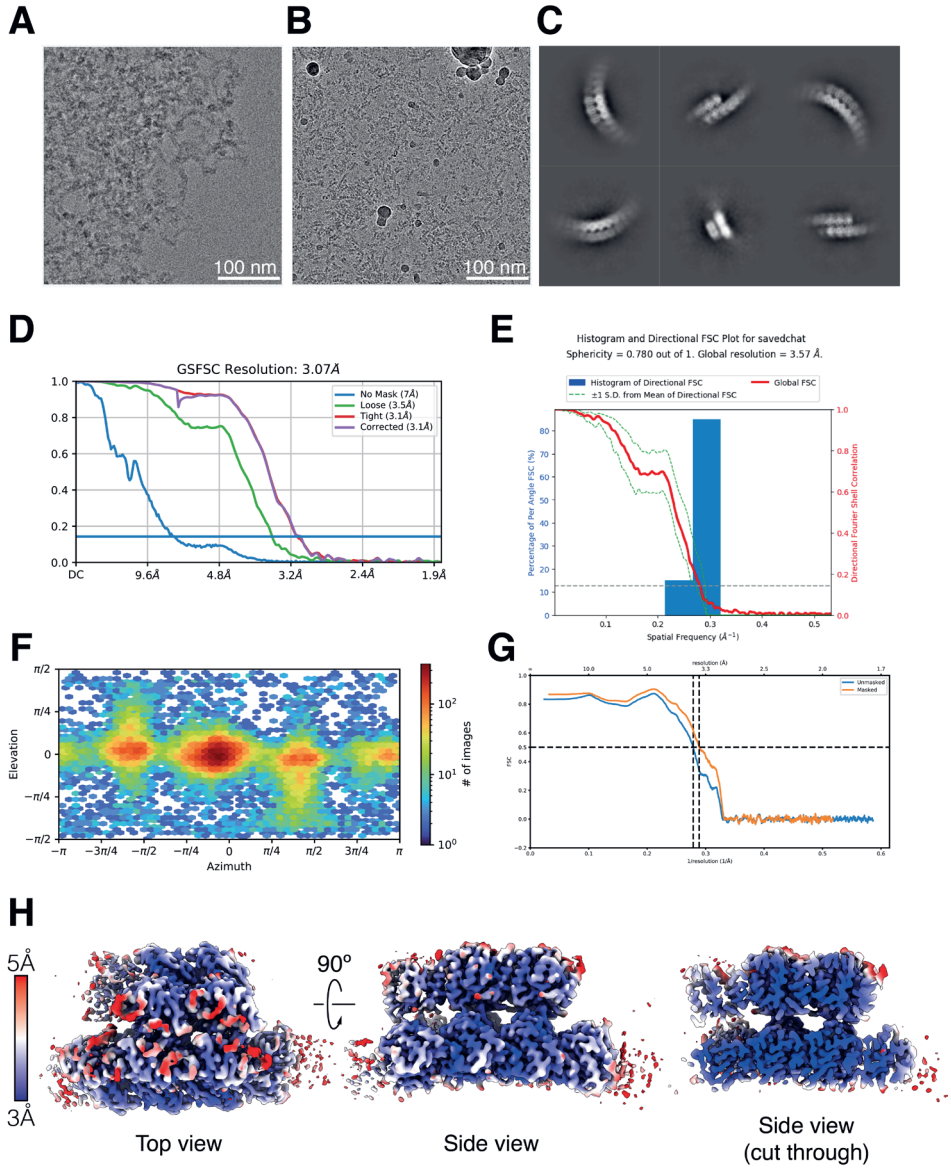


Figure S16. Cryo-EM data collection and validation. (**A**) Cryo-EM micrograph of SAVED-CHAT incubated for 1 hour at 25ºC with cA_3, resulting in large aggregates. (**B**) Cryo-EM micrograph of SAVED-CHAT incubated with cA_3 and vitrified immediately (within 10 seconds), preventing aggregation. (**C**) Cryo-EM 2D classes of SAVED-CHAT-cA_3 filaments. (**D**) Gold-standard Fourier Shell Correlation (FSC) of SAVED-CHAT-cA3 3D reconstruction. (**E**) Directional FSC of reconstruction, corresponding to **D**. (**F**) Euler orientation distribution plot. (**G**) Map-to-model FSC, with a resolution of ~3.4 Å at the 0.5 threshold. (**H**) SAVED-CHAT-cA_3 filament reconstruction coloured by local resolution.

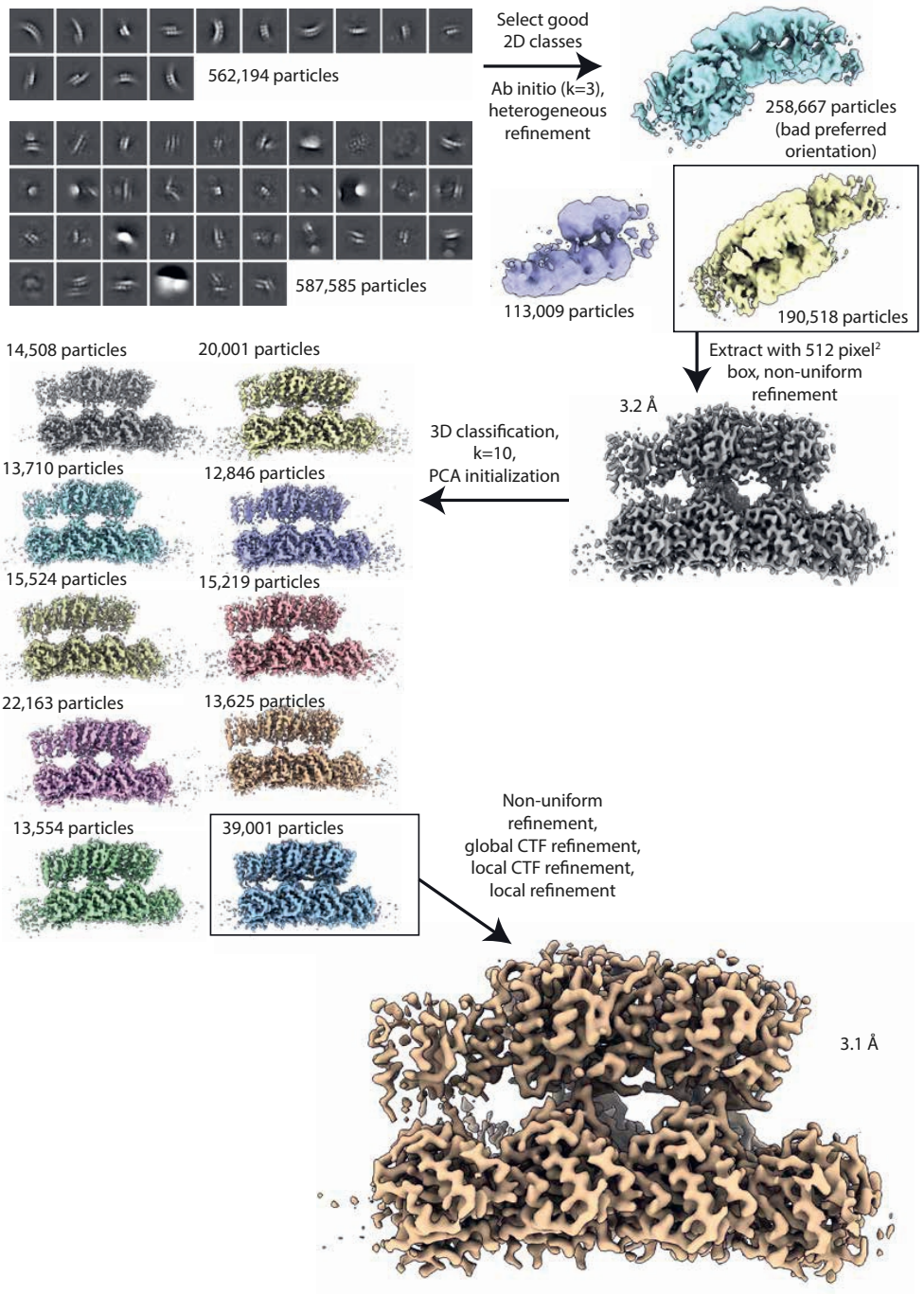


Figure S17. Cryo-EM data processing workflow. Flowchart outlining cryo-EM processing performed to obtain the structure of the active SAVED-CHAT filament.

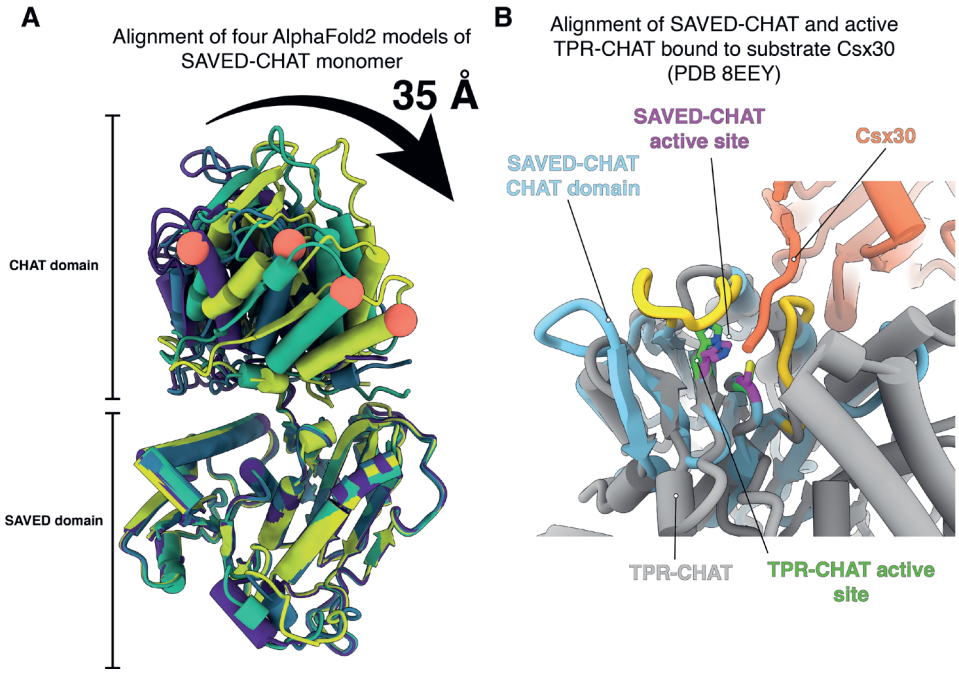


Figure S18. Structural analysis of SAVED-CHAT. (**A**) Flexibility of SAVED-CHAT monomer. Four AlphaFold2 models (coloured purple, dark cyan, green and chartreuse) were aligned based on the SAVED domain. Red sphere corresponds to the same residue in all four models, highlighting the flexibility. (**B**) Alignment of a single SAVED-CHAT proteolytic active site with TPR-CHAT (Csx29) bound to type III-E CRISPR effector in complex with activating non-self target RNA and substrate peptide Csx30. TPR-CHAT and SAVED-CHAT His-Cys active site catalytic dyad aligned in both structures, poised for cleavage of Csx30, confirming that the structure of SAVED-CHAT is in an active conformation.

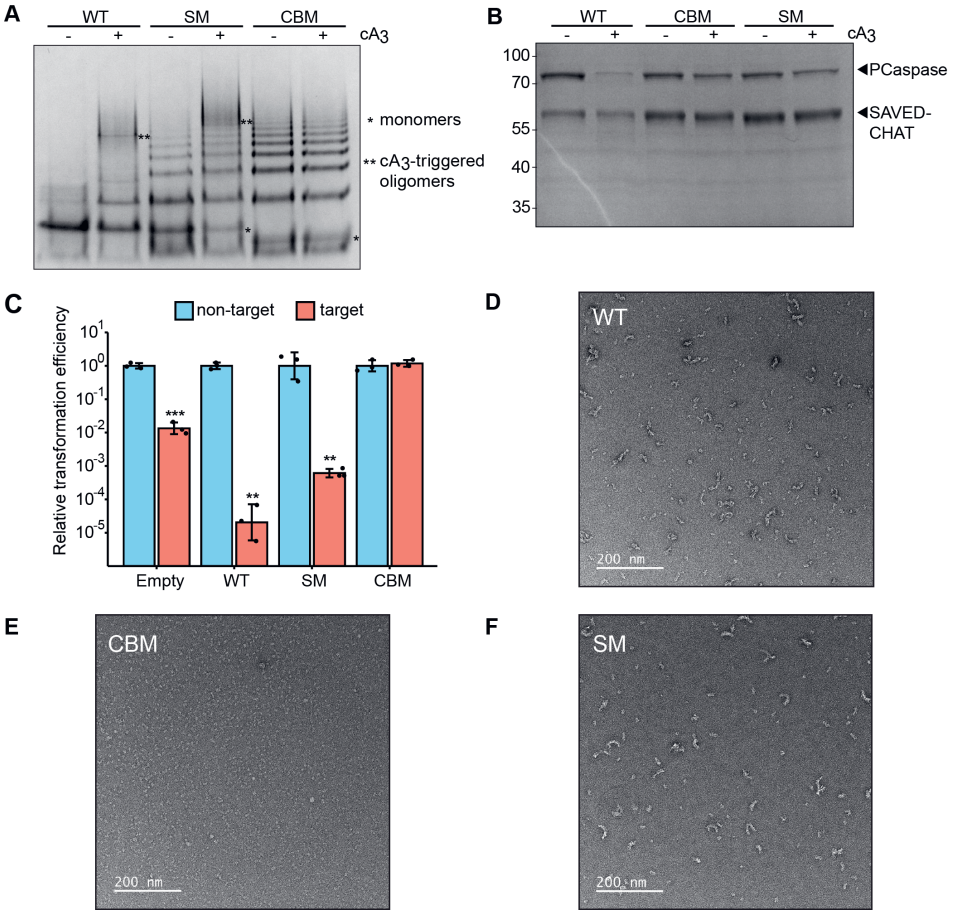


Figure S19. Impact of SAVED-CHAT mutations in the cA_3-binding site ("CBM", K121E, Q122A, N274A, R276E, Y285A, F286A) or residues involved in singlet formation ("SM", R456E, D459K, D476K, R479E) on oligomerization and protease activity. (**A**) Native PAGE analysis showing different SAVED-CHAT oligomers (compared to the wild-type) for the singlet mutant, and no induced oligomerization for the cA_3-binding mutant. (**B**) SDS-PAGE analysis showing the absence of protease activity by the two SAVED-CHAT mutants. (**C**) In vivo assay showing that the SAVED-CHAT defence response is negatively impacted by the singlet interface mutations and completely prevented by cA_3 binding site mutations. Negative staining of (**D**) wildtype SAVED-CHAT "WT", (**E**) the cA_3-binding mutant "CBM" and (**F**) the singlet mutants "SM", showing oligomerization of wildtype and the singlet mutant but not of the cA_3-binding mutant.

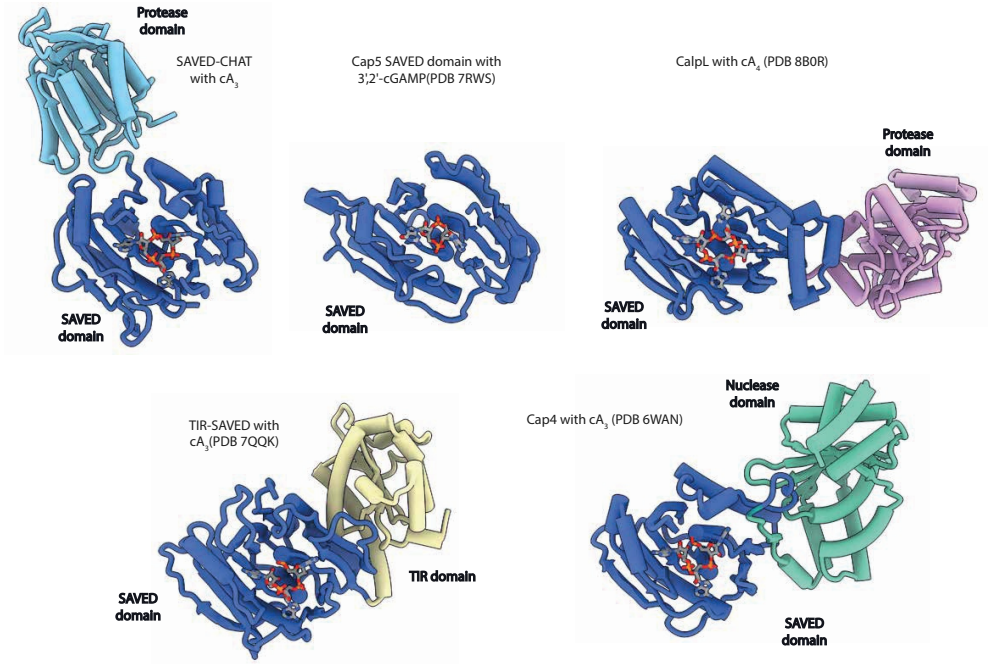


Figure S20. Comparison of SAVED domains in complex with different small molecule activators. SAVED-CHAT (blue, this study) bound to cA_3. Cap5 SAVED recognizes 3'2'-cGAMP, to activate a fused HNH domain [284]. CalpL recognizes cA_4 to activate a Lon protease domain [93], TIR-SAVED binds cA_3 to activate the TIR and degrade NAD+ [87], and Cap4 recognizes cA_3 to activate a nuclease domain [247].

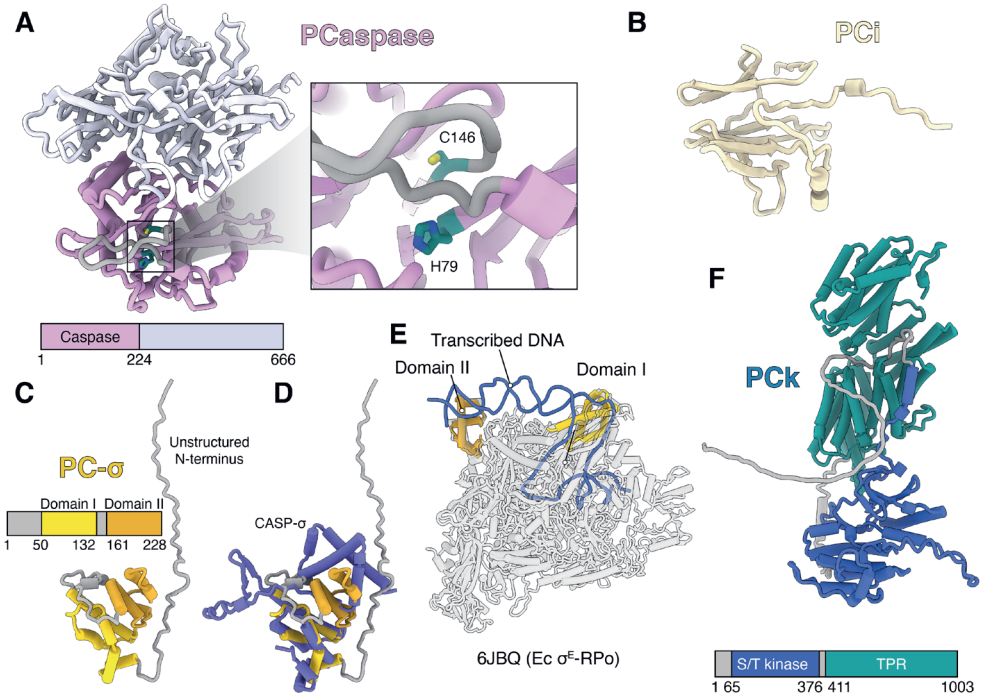


Figure S21. AlphaFold2 analysis of other factors encoded within the type III-associated genes. (**A**) Predicted model of PCaspase. The proteolytic CHAT domain is shown in pink, and the C-terminal putative regulatory domain is shown in light grey. Right: close-up view of catalytic dyad H79 and C146 in an autoinhibited conformation, with a peptide occluding the active site. (**B**) Precited model of PCi. (**C**) Predicted model of PC-σ. Two sigma factor domains (Domain I and Domain II) are shown in yellow and brown. These domains are attached by a flexible linker. (**D**) alignment of PC-σ with CASP-σ [322]. (**E**) Alignment of two PC-σ domains with *E. coli* RNA polymerase bound to σE (PDB ID 6JBQ) [344]. (**F**) Predicted model of PCk. S/T kinase domain is shown in blue, and TPR is in dark cyan.

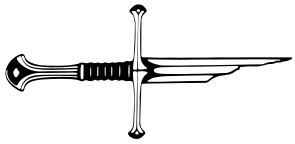
Table S1. Cryo-EM data collection, refinement, and validation statistics.

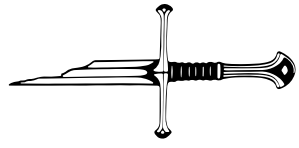
cA_3-bound SAVED-CHAT (EMDB-41358) (PDB 8TL0)	
Data collection and processing	
Voltage (kV)	200
Electron exposure (e–/Å^2)	49
Defocus range (μm)	-1.5 - -2.5
Pixel size (Å)	0.94
Symmetry imposed	C1
Initial particle images (no.)	3,824,645
Final particle images (no.)	39,001
Map resolution (Å)	3.1
FSC threshold	0.143
Map resolution range (Å)	2.5 - >5
Refinement	
Initial model used (PDB code)	N/A
Model resolution (Å)	3.4
FSC threshold	0.5
Model resolution range (Å)	2.3 - >5
Map sharpening B factor (Å^2)	-31.1
B factors (Å^2)	
Protein	23148
Nucleotides	2818
Ligand	21
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.678
Validation	
MolProbity score	1.67
Clashscore	6.18
Poor rotamers (%)	0.21
Ramachandran plot	
Favored (%)	95.27
Allowed (%)	4.73
Disallowed (%)	0

Table S2. Plasmids used in this study.

Plasmid name	Main components	Addgene ref.
pJS-BCD-Strep-SAVED-CHAT	N-terminal Strep-tagged SAVED-CHAT	214030
pJS-BCD-Strep-dSAVED-CHAT	N-terminal Strep-tagged dSAVED-CHAT (H375A / C422A)	214031
pJS-BCD-Strep-SAVED-CHAT-cbm	N-terminal Strep-tagged SAVED-CHAT cA_3 binding mutant (K121E, Q122A, N274A, R276E, Y285A, F286A)	214032
pJS-BCD-Strep-SAVED-CHAT-sm	N-terminal Strep-tagged SAVED-CHAT singlet mutant (R456E, D459K, D476K, R479E)	214033
pJS-BCD-Strep-PCaspase	N-terminal Strep-tagged PCaspase	214034
pJS-BCD-Strep-dPCaspase	N-terminal Strep-tagged dPCaspase (H79A / C146A)	214035
pJS-BCD-Strep-PCaspase-R153A	N-terminal Strep-tagged PCaspase R153A mutant	214036
pJS-BCD-Strep-PC-σ	N-terminal Strep-tagged PC-σ	214037
pJS-BCD-Strep-PCi	N-terminal Strep-tagged PCi	214038
pHochTypeIII	csb2, cmr1-cmr6, minimal CRISPR array with one spacer	214039
pSAVED-CHAT_PCaspase_PCi_PC-σ	SAVED-CHAT, PCaspase, PCi & PC-σ	214040
pSAVED-CHAT	SAVED-CHAT	214041
pPCaspase	PCaspase	214042
pSAVED-CHAT_PCaspase	SAVED-CHAT & PCaspase	214043
pSAVED-CHAT_PCaspase-N-term	SAVED-CHAT & PCaspase residues 1-153	214044
pSAVED-CHAT_PCaspase-C-term	SAVED-CHAT & PCaspase residues 154-666	214045
pSAVED-CHAT_PCaspase-N-C-term	SAVED-CHAT, PCaspase residues 1-153 & PCaspase residues 154-666	214046
pSAVED-CHAT-cbm_PCaspase	SAVED-CHAT cA3 binding mutant (K121E, Q122A, N274A, R276E, Y285A, F286A) & PCaspase	214047

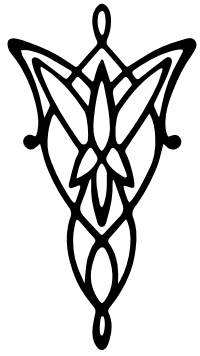
pSAVED-CHAT-sm_ PCaspase	SAVED-CHAT singlet mutant (R456E, D459K, D476K, R479E) & PCaspase	214048
pSAVED-CHAT_PCaspase_ PCi	SAVED-CHAT, PCaspase & PCi	214049
pEcNucC	E. coli NucC	214050
pEmpty	N/A	214051
pTarget	Target sequence	214052
pNon-target	Non-target sequence	214052





Chapter 8

SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation



Adapted from:
Steens, Jurre A., Yifan Zhu, David W. Taylor, Jack PK Bravo, Stijn HP Prinsen, Cor D. Schoen, Bart JF Keijser, Michel Ossendrijver, L. Marije Hofstra, Stan J.J. Brouns, Akeo Shinkai, John van der oost, Raymond H.J. Staals. SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation. *Nature communications* 12, no. 1 (2021): 5033.

Abstract

Characteristic properties of type III CRISPR-Cas systems include recognition of target RNA and the subsequent induction of a multifaceted immune response. This involves sequence-specific cleavage of the target RNA and production of cyclic oligoadenylate (cOA) molecules. In this study, we discovered that an exposed seed region at the 3' end of the crRNA is essential for target RNA binding and cleavage, whereas base pairing at the 5' end of the crRNA is required to trigger cOA production. Moreover, we uncovered that the variation in the size and composition of type III complexes within a single host results in variable seed regions. This may prevent escape by invading genetic elements, while controlling cOA production tightly to prevent unnecessary damage to the host. Lastly, we used these findings to develop a new diagnostic tool, SCOPE, for the specific detection of SARS-CoV-2 from human nasal swab samples, revealing sensitivities in the atto-molar range.

Introduction

As a widespread prokaryotic adaptive immune system, CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated) systems target and cleave genetic material of viruses and other mobile genetic elements (MGEs) [41,96,97,345]. A CRISPR array is composed of alternating repeat and spacer sequences, typically with the repeats consisting of identical sequences and the spacers consisting of variable sequence fragments acquired from invading MGEs [25–27]. The CRISPR array is generally located adjacent to a set of CRISPR-associated (cas) genes encoding the Cas proteins.

In type III CRISPR-Cas systems, the CRISPR array is expressed and processed into mature CRISPR RNA (crRNA) by the Cas6 ribonuclease [45]. A second maturation event occurs where the crRNA is trimmed at the 3' end [346,347]. The mature crRNAs form a ribonucleoprotein complex together with a set of Cas proteins: the type III effector complex (**Figure 1A**). In the interference stage, type III CRISPR-Cas systems are unique in that they attack nucleic acids in three distinct ways.

Unlike other CRISPR-Cas systems that exclusively target either DNA (type I, II and V) or RNA (type VI), many type III systems have the capacity to target both RNA and DNA [70,77,106,109,112,327,348]. Structural and biochemical analyses of the multi-subunit type III complexes have revealed the subunits responsible for these activities and showed that they are induced in a step-wise manner, starting with the binding of a complementary target RNA [79,81,82,327,349–351]. After binding, the target RNA is cleaved by the Cas7 ribonuclease, which is present in multiple copies and constitutes the backbone of type III complexes. As such, type III interference complexes have multiple (2-4) active sites, cleaving the target RNA at 6 nt intervals (**Figure 1A**) [74,108,109,327,352]. Simultaneously, target RNA binding activates two distinct catalytic domains of the Cas10 protein, the large subunit of type III complexes. Activation of the HD domain of Cas10 confers sequence-nonspecific DNase activity [70,73,74], while activation of its Palm domain triggers oligoadenylate cyclase activity, producing cyclic oligoadenylate (cOA) second messenger molecules that allosterically activate CARF (CRISPR-associated Rossmann fold) proteins [73,83]. Most of the CARF proteins that have been characterized so far appear to be promiscuous RNases (fusion of CARF and HEPN domains), cleaving both viral and host RNAs, thereby potentially inducing cell dormancy or cell death [211,223,265].

Previously, we characterized the structural and enzymatic features of the endogenous type III-B Cmr complex from *T. thermophilus* HB8 (TtCmr) [82,353]. We showed that TtCmr adopts a structure similar to type I (Cascade) complexes: a backbone consisting of several Cas7 (Cmr4) subunits, associated with multiple copies of the

small subunit Cas11 (Cmr5). The complex is capped at one end by a heterodimer of the large subunit Cas10 (Cmr2) and Cas5 (Cmr3), and at the other end by a heterodimer of Cas7-like subunits (Cmr1 and Cmr6) [82]. It is important to note that the Cas10 subunit of the TtCmr complex lacks the HD domain, and hence does not have DNase activity [81]. The mature crRNA runs along the Cas7 backbone of the complex, with its 5' repeat-derived end (the 5' handle) anchored by Cas10/Cas5, and its 3' end located at the Cmr1/Cmr6 end [79,82].

Interestingly, the 3' end of the mature crRNA is variable in type III systems, due to an uncharacterized 3' processing event following the endonucleolytic cleavages of Cas6 [80,194,354]. Although the details of this 3' processing event are not known, it is hypothesized that the heterogenous nature of the TtCmr complex might be responsible for this. Indeed, analysis of the crRNA-content of the endogenous TtCmr complex showed that it indeed co-purifies with mature crRNAs of different sizes (with variable 3' ends), with a distinct 6-nt pattern: 34, 40 & 46 nt [81]. In addition, our previously obtained cryo-EM structures revealed that the native population of TtCmr complexes consisted of larger (with a stoichiometry of $Cmr1_1 2_1 3_1 4_4 5_3 6_1$) and smaller complexes (i.e. $Cmr1_1 2_1 3_1 4_3 5_2 6_1$ and $Cmr1_1 2_1 3_1 4_2 5_1 6_1$), with the smaller complexes lacking one or two Cas7–Cas11 (Cmr4-Cmr5) backbone segment(s) (**Figure 1A**). Taken together, these data indicate that the 3' end of the mature crRNA is determined by the stoichiometry of the TtCmr complex. In this scenario, it is likely that the 3' end is generated by a (non-Cas) host ribonuclease, that shortens the unprotected, protruding 3' end of the bound crRNA [354].

In this study, we set out to understand the biological significance of these differently-sized Cmr complexes. This revealed yet another unique feature of type III: a flexible seed region at the 3' end of the crRNA guides, that appears to be important for these systems to prevent phage escapees. Additionally, we identified another key feature at the 5' end of the type III crRNA that, upon binding a perfectly match target sequence, triggers the catalytic activities of Cas10, thereby ensuring tight control over CARF protein activation. These characteristics formed the basis for the development of a highly sensitive novel type III diagnostics platform called SCOPE (**S**creening using **C**RISPR **O**ligoadenylate-**P**erceptive **E**ffectors).

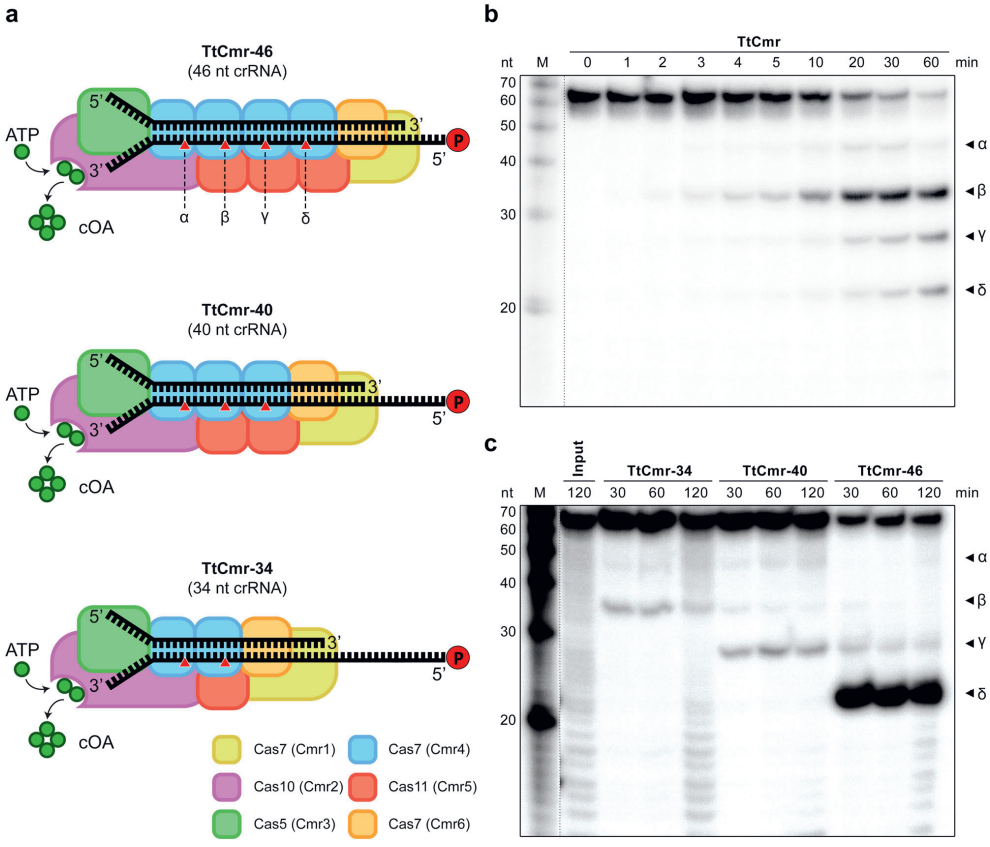


Figure 1. In vitro RNase activity assays with the endogenous and reconstituted TtCmr complexes. (**A**) Schematic illustration of the different reconstituted TtCmr complexes used in the activity assays shown in panel **C**, pre-loaded with either the 46 (TtCmr-46), 40 (TtCmr-40) or 34 nt (TtCmr-34) crRNA (top strand). Red triangles indicate the anticipated cleavage sites (α, β, γ and δ) in the 4.5 target RNA (bottom strand, **Table S1**) by the endoribonuclease activity of the Cas7 subunits. The target RNA was radiolabelled at the 5' end with ^{32}P γ-ATP ("P" in the red circle) (**B**) Denaturing PAGE analysis of the activity assay using a 5' labelled target RNA complementary to the crRNA incubated with the endogenous TtCmr complex. A single stranded RNA marker ("M") was used as size standards as indicated on the left. (**C**) Activity assays similar to panel **B** but using the reconstituted complexes. Discontinuous gel lanes are indicated by a dashed line. The results of the cleavage assays are representative results of three (1B) or two (1C) replicates.

Results

Size variation of TtCmr complexes

We previously demonstrated that endogenous type III-B Cmr complexes purified from *T. thermophilus* HB8 (TtCmr) are loaded with mature crRNA guides of different lengths (34-40-46 nt), and that the crRNA-4.5 (CRISPR array 4, spacer 5) is most abundant (**Figure 1A**) [81]. This endogenous Cmr complex specifically cleaves complementary target RNAs (4.5 target RNA) at 6 nt intervals, corresponding to the Cas7 subunits in the backbone of the complex. Consequently, this results in 5' labelled degradation products of 39 (α), 33 (β), 27 (γ) and 21 (δ) nucleotides (**Figure 1B**). However, the heterogeneous nature of the crRNA-content of the endogenous Cmr complexes [81,82], complicates the interpretation of these results. Therefore, to further reveal the mechanism of target RNA cleavage, we used *E.coli*-produced subunits to reconstitute three different Cmr complexes bound to a single crRNA (crRNA-4.5) of a defined length. Based on their abundance in their native host [81], we chose to include crRNA lengths of either 34 (TtCmr-34), 40 (TtCmr-40) or 46 nt (TtCmr-46). Opposed to all four (α-δ) 5'-labeled degradation products observed with the endogenous complex, each of the reconstituted complexes produced defined degradation products decreasing in size with a longer crRNA (**Figure 1C**). This is consistent with the idea that the composition of the complex corresponds to the length of the crRNA, with larger complexes (e.g. TtCmr-46) harbouring more cleavage sites, hence cleaving more closely to the labelled 5' end of the target RNA. Smaller complexes, such as TtCmr-40 and TtCmr-34, lack one or two Cas7–Cas11 (Cmr4-Cmr5) backbone segment(s), respectively, and therefore cleave the target RNA at less and more distal locations (further away from the 5' label), resulting in larger degradation products. These results show that the population of endogenous TtCmr complexes is a heterogenous mixture of bigger and smaller complexes, cleaving their cognate target RNAs at different positions.

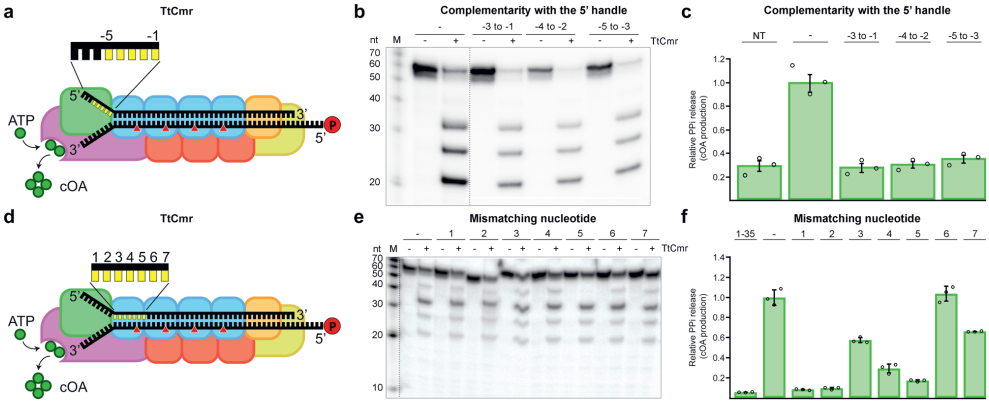


Figure 2. Impact of complementarity in the 5' handle and mismatches in the first spacer region on RNA targeting and cOA production. (**A**) Schematic illustration of the TtCmr complex bound to a target RNA (4.5 target RNA, **Table S1**), showing the different subunits in different colours, crRNA (top strand) and target RNA (bottom strand). The target RNA was labelled at the 5' end with ^{32}P ("P" in the red circle). Red triangles indicate the cleavage sites within TtCmr. Highlighted in yellow are the first 5 nucleotides in the 5' handle of the crRNA (nucleotides -1 to -5). (**B**) Different target RNAs (**Table S1**) with matches to the 5' handle were used in an activity assay and analysed on denaturing PAGE. (**C**) Impact of 5' handle complementarity on PPi release as a measure of COA production. (**D**) Similar schematic illustration as panel A with the first 7 nucleotides of the spacer region of the crRNA highlighted in yellow. (**E**) Similar activity assay as in panel **B** but with RNA targets (**Table S1**) with single mismatches in the first 7 nucleotides of the spacer region of the crRNA. (**F**) Impact of mismatches in the target RNA with the first 7 nucleotides of the spacer region of the crRNA on PPi release, as a measure for cOA production. Discontinuous gel lanes are indicated by a dashed line. The results of the cleavage assays and cOA production assays displayed in this figure are representative results of three replicates. Error bars represent the standard deviation.

Flexible 3' seed region

To investigate the significance of these type III complexes with different stoichiometries, we performed activity assays to probe for differences in seed requirements for RNA targeting as well as for the production of cyclic oligoadenylate (cOA) second messengers. In the structurally-related type I effector complexes (i.e. the Cascade complex), DNA targeting is governed by two factors: the PAM (protospacer adjacent motif) and the seed [103,355,356]. In the RNA targeting type III systems, however, self/non-self discrimination is conferred by an rPAM (RNA protospacer-adjacent motif) [112]. This motif checks for complementarity between the 5' handle of the crRNA (8 nucleotides, referred to as nucleotides -8 to -1) and the corresponding 3' region flanking the protospacer (**Figure 2A**) [70,198,327,357]. Since TtCmr is devoid of DNase activity we tested whether RNase activity and production of cOA are affected by target

RNAs with complementarity to the 5' handle of the crRNA (**Figure 2A**) [81,82]. The cleavage activity assays with the endogenous Cmr complex showed that these 'self-like' substrates had no substantial effect on RNA cleavage activity (**Figure 2B**). To probe for their impact on cOA production, we developed a pyrophosphatase-based assay that directly reflects the oligoadenylate oligomerization, as a measure for cOA production. A by-product of cOA production is the formation of pyrophosphate (PPi) [73], which can be converted to free phosphate (Pi) by a thermostable pyrophosphatase enzyme and be subsequently visualized by the Malachite Green colorimetric technique. Using this assay, we show that target RNAs with complementarity to the 5' handle reduced the production of cOA to background levels, comparable to using a non-target (NT) target RNA (**Figure 2C**). Similar results were obtained for the reconstituted TtCmr-46 and TtCmr-40 complexes (**Figure S1**). We conclude that cOA production, but not target RNA cleavage, is affected by complementarity between target RNA and the 5' handle of the crRNA, thereby mitigating detrimental consequences that cOA production might cause (e.g. cell death/dormancy) when binding antisense transcripts from the CRISPR array [77,109].

Next, we tested whether TtCmr utilizes a seed similar as described for the Cascade complexes of type I systems. Activity assays were performed by incubating the TtCmr complex with target RNAs containing single mismatch mutations in the first 7 nt of the spacer region of the crRNA, which is the region on the crRNA base pairing with the protospacer (**Figure 2D**). The results showed that RNA targeting was not affected by these mutations, although a mismatch at position 5 abolished cleavage at the adjacent site, as demonstrated by the missing 39 nt degradation product (**Figure 2E**). Interestingly however, cOA production was greatly affected by these mismatches, in particular at positions 1 and 2 (**Figure 2F**). Similar results were obtained with the reconstituted TtCmr-46 and TtCmr-40 complexes (**Figure S2**). These results indicate that the seed is either lacking or located in a different region of the crRNA. However, since base pairing at the most 5' region does appear to be critical for cOA production, we designated this segment as the **C**as10-**a**ctivating **r**egion (CAR).

Since the seed is defined as the region on the crRNA that initiates base pairing with its target, we performed EMSA binding assays with the endogenous TtCmr complex (**Figure S3**). To probe for regions crucial for initiating base paring, we used RNA targets with different mismatching segments (**Figure 3A**). We observed that targets with mismatches in the first three segments (nucleotides 1-5, 7-11, and 13-17) did not influence the binding of the target RNA by the TtCmr complex as the migration was similar to that of the fully complementary RNA target control (WT). However, mismatches in the fourth and fifth segments (nucleotides 19-23 and 25-29) substantially affected the electrophoretic mobility of the TtCmr/crRNA-target RNA tertiary complex, suggesting a seed region at the 3' end of the crRNA.

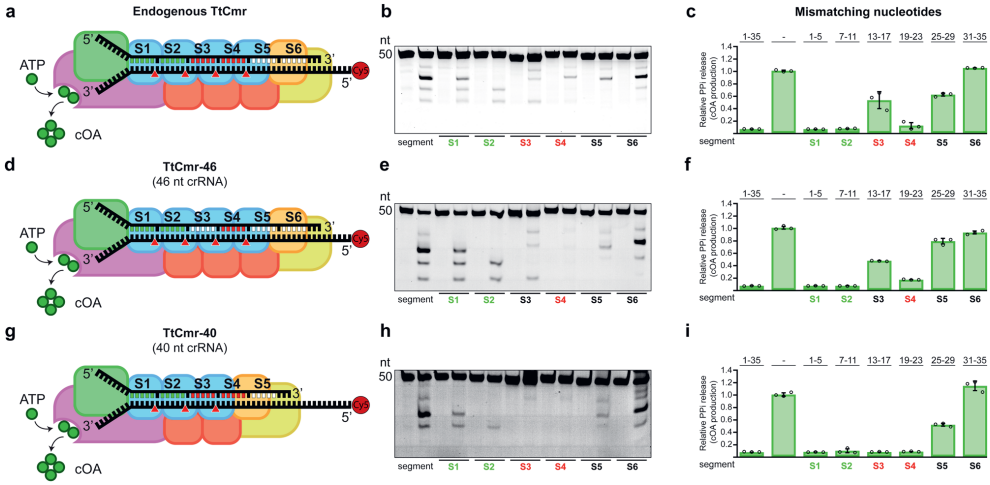


Figure 3. A flexible seed region at the 3' end of the crRNA. (**A**) Schematic illustration of the endogenous TtCmr complex. (**B**) Different target RNAs with segments mismatches were used in an activity assay and analysed on denaturing PAGE. (**C**) Impact of target RNAs with mismatches in the indicated segments on the production of cOA. (**D**) Schematic overview of the 46 nt crRNA complex (TtCmr-46). (**E**) Similar to panel **B**, using the 46 nt crRNA complex (TtCmr-46). (**F**) Similar to panel **C**, using 46 nt crRNA complex (TtCmr-46). (**G**) Schematic overview of the 40 nt crRNA complex (TtCmr-40). (**H**) Similar to panel B, using the 40 nt crRNA complex (TtCmr-40). (**I**) Similar to panel **C**, using the 40 nt crRNA complex (TtCmr-40). Target RNA contain a 5' end Cy5 label (red circle); mismatched segments are indicated with S1-S6. Red triangles indicate the cleavage sites within TtCmr. CAR segments are indicated in green, seed segments are indicated in red. The results of the cleavage assays and cOA production assays displayed in this figure are representative results of three independent experiments. Error bars represent the standard deviation.

To further investigate the impact of such a uniquely located seed on RNA degradation and cOA production, we performed activity assays with the endogenous TtCmr complexes using RNA targets containing different mismatching segments (**Figure 3A-C**). In agreement with our previous findings, RNA targets with mismatches in the first segment (S1, nucleotides 1-5) did not interfere with target degradation, despite skipping one cleavage site downstream of the mismatched segment. Similarly, mismatching of the segments S2-S5 (nucleotides 7-12, 13-17, 19-23 and 25-29) resulted in skipping both the adjacent (up- and downstream) cleavage sites, whereas cleavage at the other sites was unaffected. Mismatches in segment S6 (nucleotides 31-35) had no effect on RNA degradation, other than skipping the upstream cleavage site (**Figure 3B**). In contrast, some of the mismatching segments substantially affected the cOA production (**Figure 3C**). In agreement with results in **Figure 2F**, segment mismatches in the CAR (S1 and S2 region) completely abolished the production of cOA. Mismatches in segments S3, S5 and S6 had a minor impact on the production

of cOA, whereas a major effect on cOA production was observed with mismatches in segment S4.

Since the endogenous complex is a mixture of longer and shorter complexes, we switched to using the TtCmr-46 or TtCmr-40 reconstituted complexes in order to pinpoint this crucial region more precisely (**Figure 3D-I**). The TtCmr-46 complex almost completely mirrored the results obtained with the endogenous complex, with the exception that mismatches in segment S4 seem to abolish the RNA targeting activity (**Figure 3E**). Similarly, mismatches in the CAR (S1 and S2) as well as in S4 diminish cOA production. However, this essential region appeared to have shifted one segment in the TtCmr-40 complex, with strict base pairing requirements for RNA targeting in the third and fourth segments (**Figure 3H**). Again, effects on cOA production mirrored these results (**Figure 3I**). Taken together, these results demonstrate the existence of a 3' located seed region in TtCmr that shifts towards the 5' end of the crRNA in case of smaller guides (in smaller TtCmr complexes). We propose that together, these regions act as flexible seed sequences in TtCmr.

Seed of TtCmr structure allows for structural rearrangements

To determine the structural basis for the 3' seed in RNA targeting, we interrogated our previously determined cryo-electron microscopy structures of TtCmr with both a 46 and 40-nt crRNA (EMD-2898 and EMD-2899, respectively) [82]. In both structures, the 3' end of the crRNA is largely exposed, as it is only cradled by the Cmr1/6 heterodimer along one side of the RNA strand (**Figure 4A, B**). In contrast, the seed region of the crRNA immediately upstream of the 3' end (23 to 38 nt) is partially buried and sandwiched between the Cmr4 and Cmr5 subunits and are less exposed (**Figure 4C, D**). The 3' end of the crRNA is thus primed for transmitting conformational changes and repositioning Cmr5 subunits along the complex to facilitate complete target binding. This strongly suggests that complementarity between the crRNA and target in this region is critical for propagation of base-pairing along the length of the complex, and that it is sensitive to mismatches. Importantly, because the crRNA is shortened by one segment (6 nucleotides) in TtCmr40 compared to TtCmr46, the seed shifts towards the 5' end of the crRNA (17 to 32 nt) in the smaller complex. Interestingly, recent structures of the *S. islandicus* Cmr complex also reveal an exposed 3' end of the crRNA [358]. This indicates that the exposed 3' end is a common feature among Cmr complexes from different organisms.

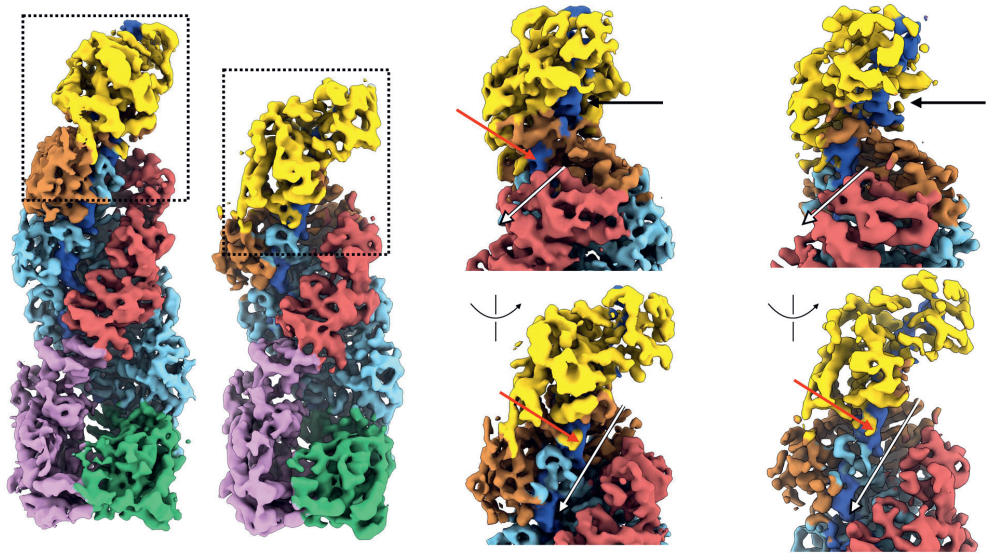


Figure 4. Structural basis for flexible 3' seed region in TtCmr complexes. (**A**) Overall structure of TtCmr complex with 46 nt crRNA (EMD-2898). (**B**) Overall structure of TtCmr complex with a 40 nt crRNA (EMD-2899). Complexes are coloured as in Figure 1. Boxed regions refer to close-up views shown in panels **C** and **D**. (**C**) Close-up view of the top of TtCmr46. (**D**) Close-up view of the top of TtCmr34. The 'open' seed region located at the 3' end of the crRNA is more accessible than the upstream regions located towards the 5' end, which are protected by Cmr5 subunits. The 3' end of the crRNA is thus primed for propagating conformational changes and repositioning Cmr5 subunits along the complex upon target binding.

SCOPE - a TtCmr-based nucleic acid detection tool

Based on these stringent target RNA requirements (i.e. the need for high target complementarity at the seed and the CAR to initiate cOA production), we concluded that type III CRISPR-Cas systems have a high potential for being repurposed as a novel, highly sensitive, robust nucleic acid detection tool. To investigate this possibility, we opted to couple the production of the second messenger to an easy read-out. In nature, these cOAs specifically bind to proteins with a CARF domain, causing an allosteric activation of fused enzyme domains. A well-characterized example of such an CARF-associated enzyme is a cOA-dependent non-specific RNase (TTHB144) of *T. thermophilus* HB8 [211]. We selected this enzyme to establish a synthetic signal transduction route consisting of an RNA-targeting TtCmr/crRNA complex that generates cOA molecules, which in turn trigger the cleavage of a reporter RNA by TTHB144 thereby generating a detectable fluorescence signal.

We first performed in vitro activity assays, using the TtCmr-46 complex, to which we added purified TTHB144 and a 5' Cy5-labelled reporter RNA. We observed defined degradation products of the reporter RNA only when both the 4.5 target RNA (T) and TTHB144 are present (**Figure 5A**), whereas a non-target RNA (NT) did not induce this activity. A guide/target mismatch at the seed region (segment 4 (S4) of TtCmr-46) greatly diminished the intensity of reporter RNA degradation products. In agreement with earlier results, guide/target mismatches in the CAR (1-5 mismatches in segment 1) completely abolished TTHB144 activation, as seen by the lack of reporter RNA degradation products (**Figure 5A**). These results show that the target RNA sequence requirements to activate TTHB144 perfectly match with those of cOA production, and that our setup can discriminate single nucleotide differences.

To generate an easy read-out for our tool, we performed a similar assay, with a fluorophore-quencher reporter RNA and measured fluorescence in real-time. For this assay, we used a crRNA that targets the E-gene of the SARS-CoV-2, one of the genes used in various RT-qPCR tests for the corona virus, validated by FIND (**Table S1**) [359]. Using this setup, a minimal detectable concentration of 1 nM target RNA concentration was achieved with a fluorescence signal detectable within seconds after starting the incubation (**Figure 5B**).

To enable DNA detection and to enhance the sensitivity of our tool even further, we included a pre-amplification step akin to those used by other CRISPR-Cas based diagnostic tools (Cas12a, Cas13) [360]. For this proof of principle, we designed a (RT-) LAMP pre-amplification step that specifically amplifies the SARS-CoV-2 E-gene, simultaneously adding a T7 promotor to the amplicon that allowed for a subsequent in vitro transcription step (**Table S1, Figure 5C**) [361]. To determine the limit of detection (LOD) of this 2-step approach, ten-fold dilutions of the complete synthetic RNA genome of SARS-CoV-2 were used. The inclusion of the LAMP pre-amplification indeed enhanced the sensitivity of our tool, reaching sensitivities in the attomolar (10-18 M) range within a timespan of ~35 minutes (30 min pre-amplification (1) + 5 min CRISPR detection (2)) (**Figure 5D**).

To validate our 2-step setup in a more complex reaction environment, we continued by testing human nasal swab samples, which were collected at an on-site SARS-CoV-2 testing facility. Out of the 80 samples tested, our tool scored 62 of them positively. We validated these results by comparing them to a PCR-based test (current standard for SARS-CoV-2 testing) that was performed on the same samples in parallel, which were in excellent agreement up to a relevant Ct-value of ~37. In addition, 20 samples that were considered negative by PCR (Ct value not determined, ND), were scored negative by SCOPE as well (**Figure 5E, Table S2** and **Figure S5**).

Lastly, the thermophilic nature of the type III proteins (Thermus-derived Cmr complex and CARF-RNase) of our tool offers an attractive opportunity for a one-pot reaction by combining the LAMP pre-amplification step with CRISPR detection, as the optimum temperature of the reactions is in the same range. Due to the maximum temperature tolerance of an appropriate, commercially available RNA polymerase (Hi-T7 RNAP, NEB), we performed a one-pot LAMP-CRISPR assay at a temperature of 55°C. The limit of detection was determined at 800 aM, using a synthetic version of the SARS-CoV-2 E-gene, demonstrating the feasibility of this approach (**Figure 5F**).

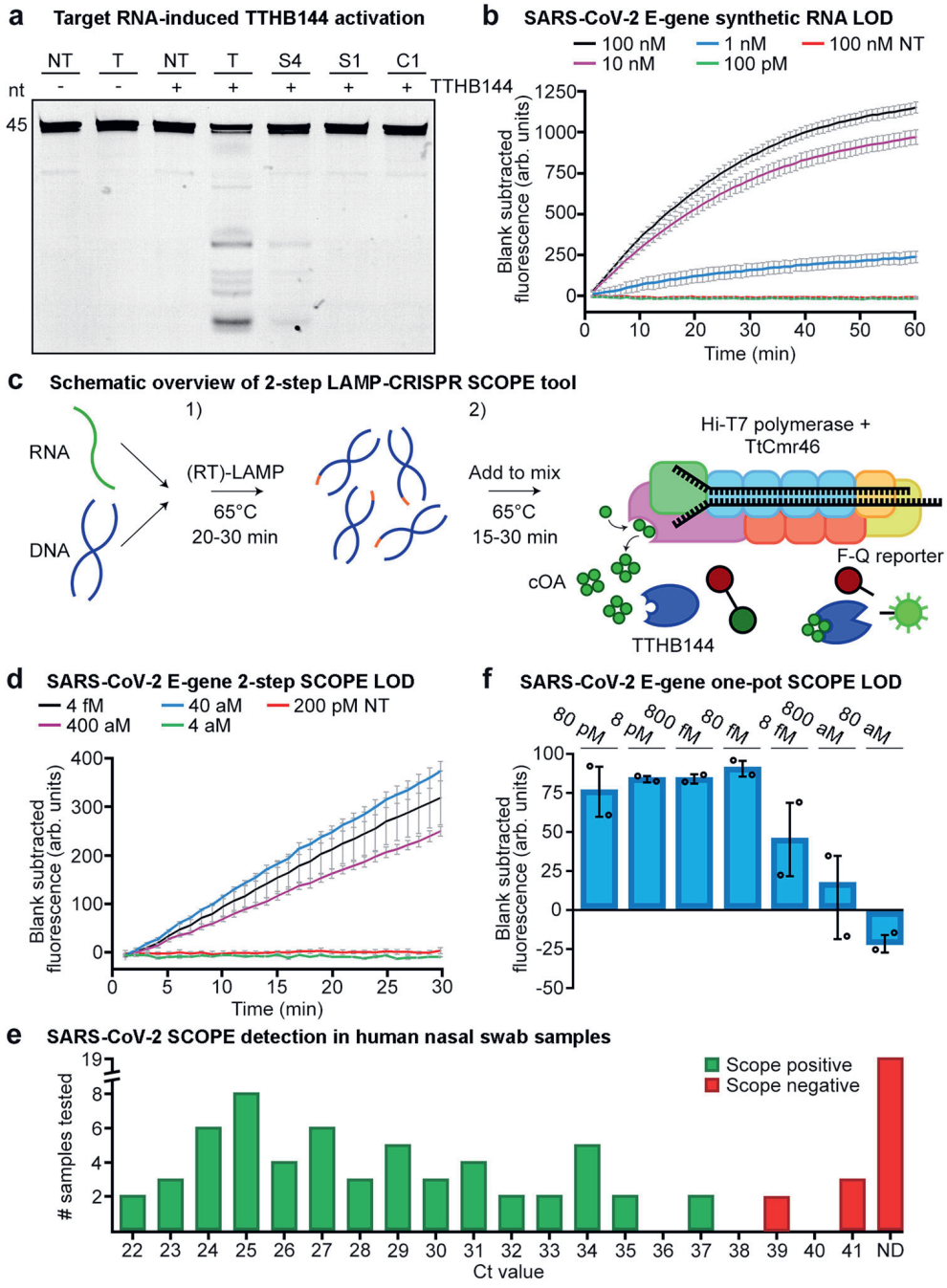


Figure 5. A novel type III CRISPR-Cas tool for the sensitive detection of nucleic acids. (**A**) Denaturing PAGE resulting from activity assays performed with the CARF protein TTHB144 and a 5' Cy5 labelled reporter RNA. Activation of TTHB144 due to cOAs pro-

duced by the reconstituted TtCmr-46 complex was monitored by offering fully complementary (T) target RNAs, or target RNA with mismatches in segments one (S1), four (S4) or by a single mismatch in the CAR (C1). A fully non-target RNA (NT) was used as a control. Results of the TTHB144 cleavage assay is representative of results obtained from three replicates. (**B**) Limit of detection (LOD) assay using reconstituted Cmr-46 on a synthetic SARS-CoV-2 E-gene and a fluorophore-quencher reporter RNA masking construct measured over time. A non-target RNA (NT) was used as a control(n=3). (**C**) Schematic overview of the 2-step reaction setup consisting out of (1) a (RT)-LAMP based pre-amplification step and (2) a T7-based *in vitro* transcription and type III CRISPR detection step. (**D**) Limit of detection assay using reconstituted Cmr-46 in the 2-step setup (depicted in panel **C**), with a SARS-CoV-2 synthetic full RNA genome as target (n=3). (**E**) Detection of SARS-CoV-2 in human swab samples. Ct-values of qPCR analysis (Orf1ab gene) of 81 samples are depicted on the X-axis with the true negative samples displayed as not determined (ND). See **Figure S5** and **Table S2** for Ct-values of qPCR analysis of SARS-CoV-2 samples and respective Scope tool score. (**F**) One-pot LAMP-CRISPR limit of detection assay, using reconstituted Cmr-46, on a synthetic SARS-CoV-2 E-gene (n = 3). All error bars in this figure represent the standard deviation.

Discussion

Recent advancements in our understanding of type III CRISPR-Cas systems have highlighted that they have unique mechanistic features compared to other CRISPR-Cas systems. Examples of this include the requirement for reverse-transcriptase activity for some type III systems during the adaptation [362–364] phase and the potentially large signalling network mediated by cOA molecules in the interference phase [73,83,210,216,265,281]. Furthermore, type III systems are exceptional in the sense that they are the only CRISPR-Cas system characterized to date capable of targeting both RNA (guide dependent) and DNA (collateral). However, the latest classification in type III CRISPR-Cas systems suggested that not all type III systems are endowed with DNase activity, due to an inactivated or missing HD domain in Cas10 [21]. This suggests that RNA is the bona fide target of these systems, as is the case for the type III-B system (Cmr-β) of *S. islandicus* [106], and of *T. thermophilus* presented here.

Yet another unique feature of type III systems is the variable crRNA length with a typical 6 nt periodicity, which, in turn, corresponds to the variable number of Cas7 subunits that constitute the backbone of type III-A and type III-B complexes [45,79–81,107,109,110]. Consequently, the cellular population of type III complexes are a heterogenous mixture of complexes with at least three different sizes [80,109,353]. The biological significance of these observations has long remained elusive (see below). Lastly, studies addressing the seed in type III systems are somewhat conflicting, with one report proposing a complete absence of a seed [155], whereas other studies report a seed region in either the 5' [195,198] or 3' [196,197] end of the crRNA guide. These discrepancies

can be explained by the different methods used to pinpoint certain regions of importance on the crRNA. For instance, read-outs that either directly or indirectly look at cOA production, such as phage challenge experiments or conformational changes in Cas10, will point towards a bigger importance of the 5' region. However, studies on the relationship of seed requirements and cOA production have paved the way for a more thorough and complete investigation [210]. Here, we dissected the importance of different size-variants of the TtCmr complex and the different regions on the crRNA by looking at each step from target RNA binding to CARF protein activation individually. Together, our data resulted in a model presented in **Figure 6**.

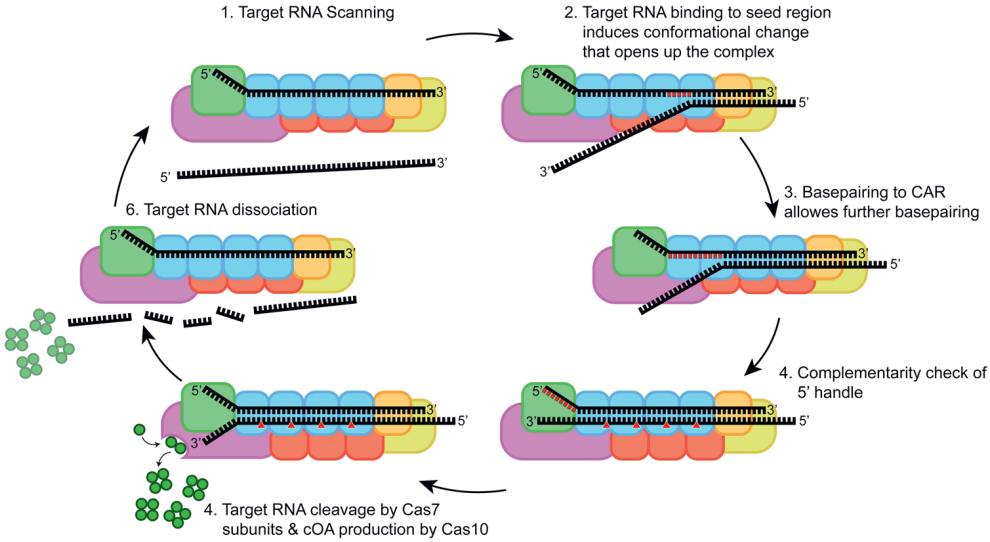


Figure 6. Schematic model of TtCmr target binding and subsequent activities. (**1**) TtCmr complex with bound crRNA is scanning for complimentary target RNA. (**2**) Target RNA binding is initiated at seed region, which induces a conformational change that allows further base pairing. (**3**) Full base pairing of target RNA to the crRNA, activating Cas10. (**4**) Target RNA is cleaved by Cas7 subunits and cOA is produced by Cas10. (**5**) Cleaved target RNA dissociates from TtCmr.

Firstly, we demonstrated that target RNA binding is initiated by a largely exposed 3' region on the crRNA, which we propose should be designated as the seed region (**Figure 3, S3**). This region coincides with the first region of the crRNA leaving the Cas7/Cas11 backbone that extends to Cmr1 and Cmr6 on that side of the complex. In agreement with our observations, previous studies showed that target RNA binding in the type III-B system requires Cmr1 and Cmr6, and that base pairing in a confined region within the 3' end of the crRNA is crucial for interference [110,352]. We therefore propose that Cmr1 and Cmr6 are involved in the proper positioning of the seed region to initiate base pairing with its cognate target RNA. The position of Cmr1 and

Cmr6 on the crRNA is determined by the variable 3' end of the crRNA. Indeed, we observed that shorter complexes (with the 40 nt crRNA) had a seed region, that was shifted one segment towards the 5' end (**Figure 3H-I**). We conclude that target RNA binding (and cleavage) is governed by a flexible 3' located seed region. This is an important difference with the structurally-related type I complexes, which harbor a fixed 5' seed region [103].

Secondly, base pairing at the seed region promotes a conformational change within the TtCmr complex, opening a channel along the Cas7/Cas11 backbone wide enough to accommodate further base pairing interactions of the crRNA with its target RNA (**Figures 4, 6**) [82]. Our results show that, in contrast to target RNA degradation, cOA production is dependent on base pairing with the first seven 5' nucleotides of the spacer part of the crRNA (with the exception of the nucleotide at position six), a region that we designated as the CAR (Cas10 Activating Region). In agreement with previous work on type I and type III complexes, the nucleotide on the crRNA after each segment (i.e. every sixth nucleotide) is excluded from base pairing with the target RNA, due to the thumb-like extension of Cas7 [82,103].

Lastly, even after fulfilling RNA targeting requirements (starting at the seed, followed by the CAR) we showed that a final checkpoint ensures that no cOA is produced when targeting self-RNAs (i.e. anti-sense transcripts from the CRISPR array, which has been reported to occur occasionally) [365]. Although self-RNAs are bound and cleaved by TtCmr, the complementarity between the 5' handle and the corresponding region on the self-RNA does completely prevent cOA production.

Although the biological significance of our findings still awaits further investigation in vivo, we anticipate that the flexibility of the seed region in type III systems will lower the chances that MGEs escape CRISPR-Cas targeting by introducing mutations in the protospacer. In type I CRISPR-Cas systems, MGE escapees promote the rapid acquisition of additional spacers, in a process called primed adaption [39,44], which ensure within-host spacer diversity. Our findings suggest that type III systems can create this diversity with just a single spacer. On one hand, the ability to recognize heavily mutated MGEs might be beneficial for the host to provide a robust interference response towards MGEs [155,366]. In *Marinomonas mediterranea* for example, a horizontally acquired type III-B system was recently shown to effectively bolster the immune response of its native type I-F systems to cope with phage escapees [367]. On the other hand, this same flexibility might come at the cost of a higher risk of self-targeting. While incidental binding and degradation of self-RNAs by itself might not represent a large fitness cost to the host, the subsequent activation of the Cas10 Palm domain (resulting in the production of the second messenger that allosterically activates for example promiscuous, sequence non-specific RNase

activity) might have more detrimental consequences [70,327]. This scenario is in good agreement with our results, showing that activation of the Cas10 Palm domain is indeed only induced when very specific conditions are met (matching seed, matching CAR and non-matching 5' handle).

The stringent control of cOA production has motivated us to repurpose Type III CRISPR-Cas systems for the specific detection of nucleic acids. We first developed a novel pyrophosphatase-based colorimetric assay, which allows for easy quantification of oligoadenylate production. When combined with appropriate Pi calibration curves, this method could further be developed for the absolute quantification cOAs or even to quantify the virus titers in a sample. For signal amplification purposes and to achieve an easy read-out (RNase activity), we selected TTHB144 - one of the three native cOA-activatable (CARF) proteins present in the genome of T. thermophilus [73,89,211]. While TTHB152 is also characterized as a CARF-RNase, it is encoded in the type III-A operon so for consistency we opted to use TTHB144. We observed highly similar trends in target RNA sequence requirements (CAR, and segments mutants) for activation of the RNase activity of TTHB144 compared to those governing cOA production (**Figures 2, 3**). This shows that the stringent control of TTHB144 activation could be utilized to make the highest possible distinctions between target RNAs, i.e. monitoring even a single nucleotide difference (**Figure 5A**, C1 mutant). Next to the specificity aspect, due to the combination of an intrinsic signal amplification step (Cas10 producing a multitude of cOA molecules) and the high ribonuclease activity of TTHB144 [211], our tool was able to very quickly generate a fluorescence signal: detecting 1nM of target RNA within seconds (**Figure 5B**). For most diagnostic applications however, a sensitivity of 1nM is not high enough [360]. Therefore, we adapted a previously established SARS-CoV-2 LAMP pre-amplification reaction to boost the sensitivity of our diagnostic tool even further 361. Indeed, after testing this 2-step protocol (**Figure 5C**) on a SARS-CoV-2 RNA reference genome, we determined its limit of detection at 40 aM (~40 copies/µL). Validating our test on human nasal swab samples indicated our test to be 100% accurate up to a Ct value of ~37, which is higher than most accepted cut-offs for diagnosis 359. Equally important, we did not detect any false positives out of the 20 negative samples tested.

Ideally, a true one-pot reaction is preferred to minimize steps and to reduce the risk of cross-contamination. We therefore investigated the performance of our system in a one-pot setup, at a temperature of 55°C (to comply with the upper temperature limit of Hi-T7 polymerase). Despite using sub-optimal temperature for the CRISPR detection step, we were able to achieve a limit of detection of 800 aM using a synthetic gene template (**Figure 5F**). We showed the feasibility of the one-pot LAMP CRISPR detection approach with some room for optimization regarding the amount of signal generated and reducing the incubation time. Although similar reaction component

concentrations compared to the 2-step protocol were used, both the LOD (800 aM) and reaction time (180 min) were affected in the one-pot approach. This is likely due to the sub-optimal incubation temperature of 55°C combined with potential interfering reaction components to the LAMP reaction. Optimization consisting of changing TTHB144 and reporter RNA concentrations, for example, could improve this issue. Furthermore, the LAMP reaction by itself could be designed and optimized to work more efficiently at a temperature of 55°C by choosing a different DNA polymerase or change the primer design to reduce overall reaction times.

A couple of CRISPR-based nucleic acids detection platforms have been developed over the last years, such as the Class2-based DETECTR and SHERLOCK platforms [72,368]. However, SCOPE is the first Class 1-based CRISPR-Cas nucleic acid detection tool, with some very useful characteristics: highly sensitive (requiring very little sample input, see methods) and specific (flexible seed, stringent CAR), quick (detection within seconds), flexible (PAM-independent guide design), highly robust (long shelf-life). Furthermore, the potential for a 1-pot system in combination with RT-LAMP offers opportunities to increase the throughput of our tool. The thermophilic nature of the system potentially means that the system is less affected by, enzyme mediated, inhibitory factors in crudely extracted samples. While the 2-step approach currently limits efficient high-throughput testing due to the extra (manual) step that is required, an efficient 1-pot system would mitigate this. The use of a standard fluorophore and no proprietary plastics further facilitates integration in current high throughput testing facilities. All these favourable features make the SCOPE approach a highly attractive alternative over the currently available detection tools.

Acknowledgments

We thank the Nogales and Doudna labs for their support on the structural analyses (University of California, Berkeley, CA, USA). R.H.J.S was supported by a VENI grant (016.Veni.171.047), J.V.D.O. by a TOP grant (714.015.001), and S.J.J.B. by a VICI grant (VI.C.182.027), all from The Netherlands Organization for Scientific Research (NWO). This work was supported in part by Welch Foundation grant F-1938 (to D.W.T.), National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) R35GM138348 (to D.W.T.), and a Robert J. Kleberg, Jr. and Helen C. Kleberg Foundation Medical Research Award (to D.W.T.). D.W.T is a CPRIT Scholar supported by the Cancer Prevention and Research Institute of Texas (RR160088) and an Army Young Investigator supported by the Army Research Office (W911NF-19-1-0021). This work was also supported by the David Taylor Excellence Fund in Structural Biology made possible with support from Judy and Henry Sauer (to D.W.T.).

Author Contributions

R.H.J.S., Y.Z., J.v.d.O. and J.A.S. conceived of and designed the study. Y.Z., J.A.S., D.W. T., J.P.B, S.H.P.P. and A.S. performed experiments and analyses. R.H.J.S., S.J.J.B., J.v.d.O. provided experimental guidance. R.H.J.S. and J.A.S. wrote the manuscript with significant input from other authors.

Declaration of Interests

J.A.S, S.H.P.P are founders and shareholders of Scope Biosciences. J.v.d.O, R.H.J.S are shareholders and members of the scientific board of Scope Biosciences. J.A.S, J.v.d.O, R.H.J.S, S.H.P.P are inventors on type III CRISPR-Cas related patents.

Materials and Methods

Purification of the Cmr complex and individual subunits

A *T. thermophilus* HB8 strain was created expressing a genomic His-tagged Cm6 subunit. The (His)6-tagged TtCmr complex purification was achieved via several chromatography steps as detailed previously [81]. The purified TtCmr sample was further concentrated using Vivaspin 20 centrifugal concentrator. For reconstituted complex, each subunit was individually expressed and purified. All subunits were cloned in a bicistronic design elements containing expression plasmid under an Isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible T7 promoter and N-terminal Streptavidin tag. Expression plasmid containing *E. coli* BL21(DE3) were grown at 37°C until ~OD_{600}=0.6, after which the culture was cold-shocked on ice for 1h. IPTG was added to a final concentration of 0.5-1 mM and the culture was incubated at 20°C for 16h. Cells were harvested and resuspended in Wash Buffer (150mM NaCl, 100mM Tris-HCl, pH 8) and a Complete protease inhibitor tablet was added (Roche). Cells were lysed by sonication (25% amplitude 1 sec on, 2 sec off, Bandelin Sonopuls) and spun down at 30.000g for 45 min, subsequent lysate was filter (0.45µm) clarified. A StrepTrap HP (GE) column was equilibrated using Wash Buffer and the lysate was run over it. Wash Buffer with d-Desthiobiotin added to a final concentration of 2.5 mM was used to elute. The elution fractions were pooled, concentrated and run over a HiLoad® 16/600 Superdex® 75 pg size exclusion chromatography column for further purification. Purification of TTHB144 was performed in similar fashion.

Reconstitution of TtCmr complexes

For TtCmr46, 3.5 µL crRNA (700 ng) was added to 3.5 µL 1X Cmr buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl). Subsequently, the subunits were added to the reaction mixture in a specific order (Cmr3, Cmr2, Cmr4, Cmr5, Cmr6, Cmr1) to a final concentration of 2.5 µM, 2.5 µM, 10 µM, 7.5 µM, 2.5 µM and 2.5 µM respectively, to make up a total reaction volume of 20 µL. For TtCmr40 final concentration of subunits was adjusted to 2.5 µM, 2.5 µM, 7.5 µM, 5 µM, 2.5 µM and 2.5 µM respectively. The reaction mixture was incubated at 65˚C for 30 minutes.

In vitro cleavage activity assays

RNA substrates (listed in **Table S1**) were either 5’ labelled by T4 polynucleotide kinase (NEB) and 5’ ^{32}P-γ -ATP, after which they were purified from a denaturing PAGE using RNA gel elution buffer (0.5M Sodium acetate, 10mM $MgCl_2$, 1mM EDTA and 0.1% SDS) or ordered with a 5’ Cy5 fluorescent label. In vitro cleavage activity assays were conducted in TtCmr activity assay buffer (20mM Tris-HCl

pH 8.0, 150mM NaCl, 10mM DTT, 1 mM ATP, and 2 mM $MgCl_2$) using the RNA substrate and 62.5 nM TtCmr. Unless stated otherwise, the reaction was incubated at 65°C for 1 hour. RNA loading dye (containing 95% formamide, dyes left out in case of Cy5 substrates) was added to the samples after incubation boiled at 95° for 5 minutes. The samples were run on a 20% denaturing polyacrylamide gel (containing 7M urea) for about 1-4 hours at 15mA or overnight at a constant of 4mA. The image was visualized using phosphor imaging or fluorescent gel scanning (GE Amersham Typhoon).

cOA detection assay

The in vitro cOA detection assays were conducted in TtCmr activity assay buffer (20mM Tris-HCl pH 8.0, 150mM NaCl, 10mM DTT, 1 mM ATP, and 1 mM $MgCl_2$) to which Cmr-complex (62.5 nM final concentration) as well as the RNA substrates (200 nM, listed in **Table S1**) were added. The reaction was incubated at 65˚C for 1 hour after which 0.05 units of pyrophosphatase (ThermoFisher EF0221) was added, followed by an incubation at 25˚C for 30 minutes (**Figure 2**). Alternatively, thermostable pyrophosphatase (NEB #M0296) was added during the 1h incubation at 65˚C (**Figure 3**). cOA quantification was achieved by, using the Malachite Green Phosphate Assay Kit (Sigma-Aldrich MAK307). This resulted in an OD_{650} signal, which was measured on a BioTek Synergy Mx plate reader.

The unitless relative PPi levels (cOA-production) presented in **Figures 2** and **3** were calculated by expressing the OD_{650} signal from the individual RNA targets (containing the indicated mismatches) as a ratio of the OD_{650} signal obtained from the fully complementary target RNA.

EMSA

EMSAs were performed by incubating 62.5 nM TtCmr complex with 13.3 nM 5' Cy5 labelled target RNAs (**Table S1**) in Cmr binding buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1 mM DTT, 1 mM EDTA). All reactions were incubated for 20 min at 65°C before electrophoresis on a native 5% (w/v) polyacrylamide gel (PAGE), running at 15mA. The image was visualized via fluorescent gel scanning (GE Amersham Typhoon).

Structural modelling

In order to model crRNA within our previously determined TtCmr46 and TtCmr40 maps (EMD-2898 and -2899, respectively), the model of *S. islandicus* Cmr complex (PD 6S6B) was fitted as a single rigid body into the TtCmr46 map, and supervised flexible fitting was performed using Isolde 369. Maps were visualized using ChimeraX [370].

Human swab sample collection, extraction and qPCR

Combined human throat/nasopharyngeal swab samples were obtained at a community testing centre (Utrecht, the Netherlands) from adults with either SARS-CoV2-related symptoms or with contacts of infected persons. As samples were collected within the framework of the national COVID measures and testing efforts, samples were routinely collected at interval timepoints. PCR was conducted in a certified clinical laboratory and all procedures were validated according to the ISO 15189 standard. Nasopharyngeal swabs were transferred into 3 ml Universal transport medium. RNA was isolated and purified using the MagC extraction kit (Seegene) on an automatic nucleic acid extractor Hamilton MicroLAB StartLET (Bonaduz). SARS-CoV-2 qPCR was performed using the Allplex 19-nCoV multiplex Real-time PCR assay (Seegene), as described earlier [371]. Results were interpreted with Seegene Viewer data analysis software. A positive result was defined as amplification of the Orf1ab SARS-CoV-2 gene.

Nucleic acid detection tool

The target RNA induced activation of TTHB144 assays were performed similarly to the earlier described in vitro cleavage assays using reconstituted TtCmr-46 complex. However, non-labelled (non)target RNA was used and either a 5' Cy5 reporter RNA (**Figure 5A**) or commercial RNaseAlert (**Figure 5B, D, F and E**) was added, as well as 1 μM purified TTHB144.

The LAMP reaction in the 2-step LAMP-CRISPR detection setup (**Figure 5C, 5D**) was performed using the WarmStart® LAMP Kit (NEB #E1700), using previously published primers concentration described in literature (**Table S1**), designed for amplifying SARS-CoV-2 RNA [361]. Final primer concentrations of 0.2 μM, 1.6 μM and 0.8 μM for the F3/B3, FIP/BIP and LoopF/LoopB primers were used respectively. A T7 promotor sequence was added to the loop primer to allow for the subsequent in vitro transcription reaction, which affected its performance only marginally when compared to the standard Loop primer (data not shown). After pre-amplification by LAMP, subsequent T7 polymerase transcription will enrich the target RNA, partly

consisting of a template derived region which lies between the Loop primer and the B2 primer (Figure S4). The 2-step assay was performed by a 30 min LAMP pre-amplification step, using 5 µL of sample RNA extract, and a 15 min CRISPR detection step. The CRISPR detection step was performed by adding 8.75 µL CRISPR Mix (final concentrations after addition: 1X TtCmr Activity Assay Buffer, ~62.5 nM reconstituted TtCmr-46 complex, 500 nM TTHB144, 2 µL NTP Buffer Mix (NEB #E2050), 25 U Hi-T7 RNA polymerase (NEB #M0658), 250 nM RNaseAlert™ QC System v2 (ThermoFisher) and 8 mM MgCl). Measurements were made at 1- or 2-minute intervals at 65°C on a Bio-Rad CFX96 qPCR machine (**Figure 5D**) and Applied Biosystems™ 7500 (Figure 5E) (FAM channel). The SARS-CoV-2 synthetic genome was ordered from Twist Biosciences.

The one-pot LAMP-CRISPR reaction was set up as a total reaction volume of 20 µL, according to the LAMP method described earlier with additions to the earlier described CRISPR Mix (Modification: 1 µL NTP Buffer Mix (From NEB #E2050, no additional $MgCl_2$). Total incubation-time for the one-pot reaction was 180 mins.

In all cases, the limit of detection (LOD) was determined by the lowest concentration of target RNA in the reaction mixture in step 1, which led to an significant increase in fluorescence compared to the non-target RNA control. The delta-signal SCOPE value was calculated by taking the, blank subtracted, increase of fluorescent signal over time during the 15 mins of the CRISPR detection step in the 2-step protocol. Similarly, the same calculation was used for the one-pot assay, by using the signal increase over 180 min of incubation. For the human swap samples presented in **Figure 5E**, a sample was considered positive if the average delta-SCOPE value was within 2 standard deviations of all qPCR-scored positive samples.

Data availability

The structures of the different TtCmr complexes (**Figure 4**) have previously been deposited into the EMDataBank with accession codes EMD-2898 and EMD-2899 [82]. Other data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplemental Information

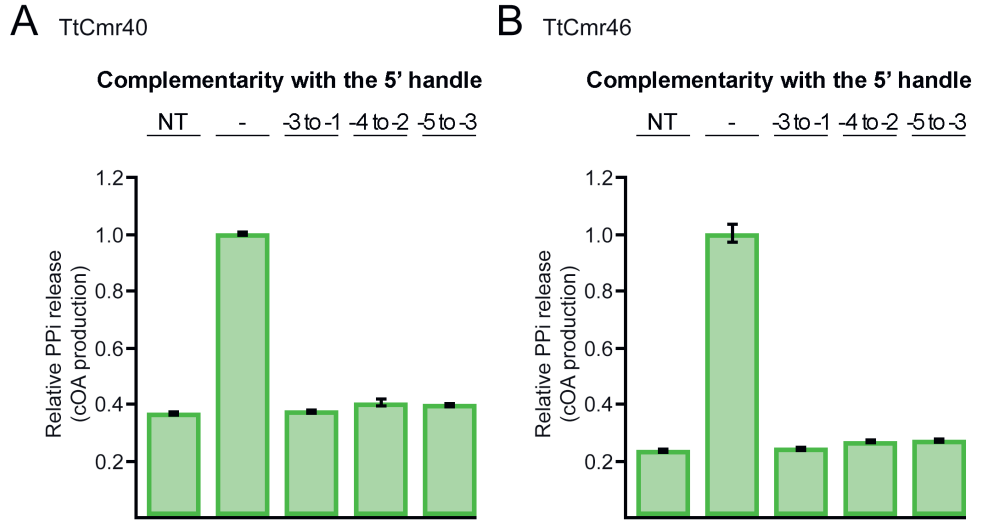


Figure S1. 5' Handle complementarity affects cOA production. (**A**) Effects of 5' handle complementarity on cOA production in TtCmr40 complex. (**B**) Similar to panel **A**, using TtCmr46 complex. NT= Non-target WT= Wild-type target. Error bars represent the standard deviation.

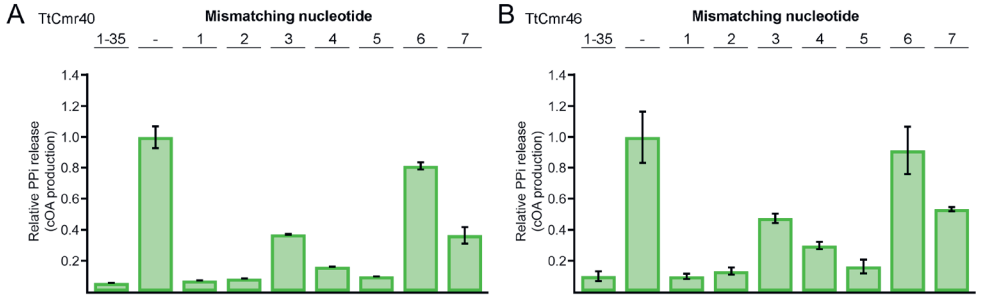


Figure S2. Complementarity in 5' region regulates cOA production. (**A**) Effect of mismatches in 5' region on production of cOA in TtCmr40 complex. (**B**) Same assay as panel **A**, using TtCmr46 complex instead. Error bars represent the standard deviation.

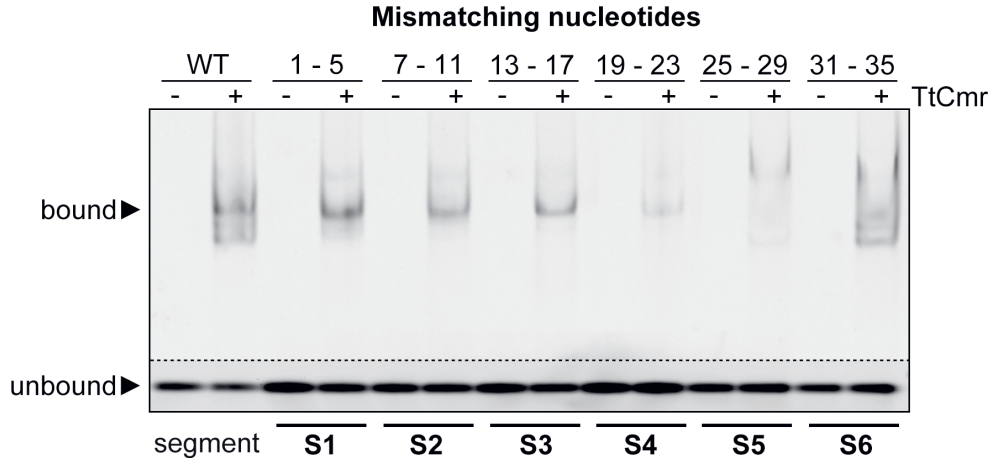


Figure S3. Base pairing of the target RNA with the TtCmr-bound crRNA is initiated at the 3' end of the crRNA. EMSA analysis of the endogenous TtCmr complex incubated with different target RNAs (**Table S1**) each containing a stretch of 5 nt mismatching with the TtCmr-bound crRNA. The results obtained are representative of results of 3 replicates.

Figure S4. Schematic overview of SARS-CoV-2 E-gene LAMP amplification. Design of SARS-CoV-2 E-gene LAMP amplification, adapted from [361]. Modification to the Loop primer is the addition of a T7 RNA polymerase promoter (**Table S1**).

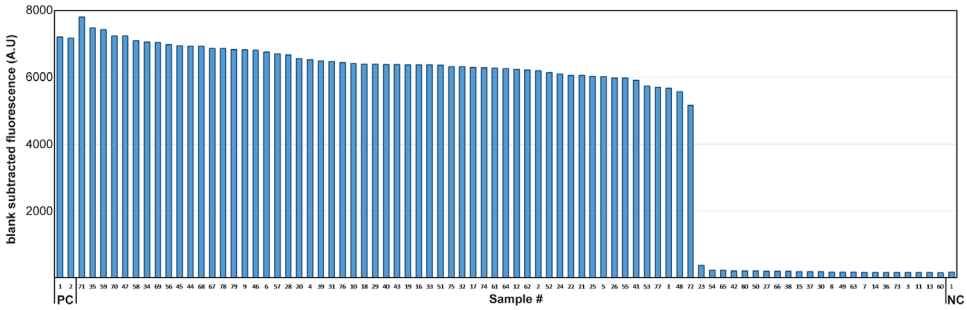


Figure S5. Graph depicting delta SCOPE values (blank subtracted increase in fluorescence) per sample, sorted by signal intensity. The delta SCOPE values used for creating **Figure 5E** are depicted in a bar graph, sorted by category (PC= positive control, NC=negative control, sample=human swab sample) and signal intensity to show the binary nature of a positive/negative SCOPE signal.

Table S1. Overview of used oligonucleotides.

	Sequence (5' to 3')	Comments
P1	AUUGCGACCCGUAGAUAAGGCGCCCGGG-GACGACCACGUCAAGGCGCAGUUGC	crRNA 4.5, 46nts (spacer region labelled as red)
P2	AUUGCGACCCGUAGAUAAGGCGCCCGGG-GACGAC	crRNA 4.5, 34nts (spacer region labelled as red)
P3	AUUGCGACCCGUAGAUAAGGCGCCCGGG-GACGACCACGUC	crRNA 4.5, 40nts (spacer region labelled as red)
P4	AUUGCGACCCGUAGAUAAGGCGCCCGGG-GACGACCACGUCAAGGCG	crRNA 4.5, 46nts (spacer region labelled as red)
P5	GUUGUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAUCUACGGCAGCGU	Target RNA 4.5 (red part complementary to crRNA 4.5)
P6	GUUGUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAUCUACGGGUCCGU	Matched -3/-1 target RNA (green part)
P7	GUUGUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAUCUACGGCUCGGU	Matched -4/-2 target RNA (green part)
P8	GUUGUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAUCUACGGCACGCU	Matched -5/-3 target RNA (green part)
P9	GAACUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAUCUACGCCCAUCG	Mismatched +1 RNA target (blue part)
P10	GAACUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAUCUACCGCCAUCG	Mismatched +2 RNA target (blue part)
P11	GAACUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAUCUAGGGCCAUCG	Mismatched +3 RNA target (blue part)

P12	GAACUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAUCUUCGGCCAUCG	Mismatched +4 RNA target (blue part)
P13	GAACUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAUCAACGGCCAUCG	Mismatched +5 RNA target (blue part)
P14	GAACUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAUGUACGGCCAUCG	Mismatched +6 RNA target (blue part)
P15	GAACUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAACUACGGCCAUCG	Mismatched +7 RNA target (blue part)
P16	GAACUGCGCCUUGACGUGGUCGUCCCCGG-GCGCCUUAUCAUGCCCCAUCG	Mismatched +1-5 RNA target (blue part), also used as 5' Cy5 variant
P17	GAACUGCGCCUUGACGUGGUCGUCCCCGG-GCGCGAAUACUACGGCCAUCG	Mismatched +7-11 RNA target (blue part), also used as 5' Cy5 variant
P18	GAACUGCGCCUUGACGUGGUCGUC-CCCCCCGCCCUUAUCUACGGCCAUCG	Mismatched +13-17 RNA target (blue part), also used as 5' Cy5 variant
P19	GAACUGCGCCUUGACGUGGUCCAGGGCGG-GCGCCUUAUCUACGGCCAUCG	Mismatched +19-23 RNA target (blue part), also used as 5' Cy5 variant
P20	GAACUGCGCCUUGACCACCACGUCCCCGG-GCGCCUUAUCUACGGCCAUCG	Mismatched +25-29 RNA target (blue part), also used as 5' Cy5 variant
P21	GAACUGCGCGAACUCGUGGUCGUCCCCGG-GCGCCUUAUCUACGGCCAUCG	Mismatched +31-35 RNA target (blue part), also used as 5' Cy5 variant
P22	AAACGACGGCCAGUGCCAAGCUUACUAUA-CAACCUACUACCUCAU	TTHB144 5' Cy5 reporter RNA
P23	UGAUGAGGUAGUAGGUUGUAUAGUAAGC-UUGGCACUGGCCGUCGUUUACG	Non-target RNA 4.5
P24	AUUGCGACCACACAAUCGAAGCG-CAGUAAGGAUGGCUAGUGUAACU	crRNA SARS-CoV-2 E-gene, 46nts (spacer region labelled as red)
P25	UAGUUACACUAGCCAUCCUUACUGCGCU-UCGAUUGUGUGCGUACUGCUGCAAUAUU-GUUA	Target RNA SARS-CoV-2 E-gene (red part complementary to P24)
P26	AUUGCGACACAAUAUUGCAGCAGUACGCA-CACAAUCGAAGCGCAGA	crRNA SARS-CoV-2 E-gene LAMP (spacer region labelled as red)
P27	AGCTGATGAGTACGAACTT	LAMP E-gene F3

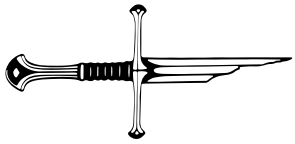
P28	TTCAGATTTTTAACACGAGAGT	LAMP E-gene B3
P29	ACCACGAAAGCAAGAAAAAGAAGTATTC-GTTTCGGAAGAGACAG	LAMP E-gene FIP
P30	TTGCTAGTTACACTAGCCATCCTTAG-GTTTTACAAGACTCACGT	LAMP E-gene BIP
P31	TAATACGACTCACTATAGCTGCGCTTC-GATTGTGTGCGT	LAMP E-gene T7 Loop

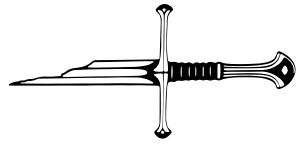
Table S2. Ct values of qPCR analysis and delta SCOPE signal values of human SARS-CoV-2 samples

Sample Name	Ct qPCR Orf1ab	qPCR result	Blank subtracted fluorescence	Scope result
Negative control	ND	Neg	164	Neg
UMCU sample 01	29.6	Pos	5664	Pos
UMCU sample 02	36.0	Pos	6199	Pos
UMCU sample 03	39.4	Pos	156	Neg
UMCU sample 04	26.2	Pos	6526	Pos
UMCU sample 05	32.6	Pos	6019	Pos
UMCU sample 06	24.9	Pos	6756	Pos
UMCU sample 07	ND	Neg	161	Neg
UMCU sample 08	41.6	Pos	170	Neg
UMCU sample 09	26.6	Pos	6828	Pos
UMCU sample 10	27.4	Pos	6417	Pos
UMCU sample 11	ND	Neg	155	Neg
UMCU sample 12	25.0	Pos	6244	Pos
UMCU sample 13	ND	Neg	152	Neg
UMCU sample 14	ND	Neg	160	Neg
UMCU sample 15	41.8	Pos	176	Neg
UMCU sample 16	25.9	Pos	6371	Pos
UMCU sample 17	30.9	Pos	6290	Pos
UMCU sample 18	26.0	Pos	6394	Pos
UMCU sample 19	31.6	Pos	6372	Pos
UMCU sample 20	25.0	Pos	6563	Pos
UMCU sample 21	25.6	Pos	6057	Pos
UMCU sample 22	28.4	Pos	6058	Pos

UMCU sample 23	ND	Neg	370	Neg
UMCU sample 24	29.2	Pos	6095	Pos
UMCU sample 25	27.7	Pos	6032	Pos
UMCU sample 26	26.6	Pos	5979	Pos
UMCU sample 27	ND	Neg	190	Neg
UMCU sample 28	34.2	Pos	6671	Pos
UMCU sample 29	34.7	Pos	6390	Pos
UMCU sample 30	ND	Neg	174	Neg
UMCU sample 31	28.6	Pos	6468	Pos
UMCU sample 32	29.3	Pos	6312	Pos
UMCU sample 33	31.3	Pos	6367	Pos
UMCU sample 34	28.5	Pos	7056	Pos
UMCU sample 35	30.2	Pos	7488	Pos
UMCU sample 36	ND	Neg	160	Neg
UMCU sample 37	ND	Neg	175	Neg
UMCU sample 38	ND	Neg	184	Neg
UMCU sample 39	25.3	Pos	6493	Pos
UMCU sample 40	24.7	Pos	6388	Pos
UMCU sample 41	34.9	Pos	5912	Pos
UMCU sample 42	ND	Neg	200	Neg
UMCU sample 43	30.9	Pos	6386	Pos
UMCU sample 44	27.3	Pos	6938	Pos
UMCU sample 45	26.8	Pos	6940	Pos
UMCU sample 46	32.3	Pos	6809	Pos
UMCU sample 47	35.5	Pos	7240	Pos
UMCU sample 48	24.7	Pos	5565	Pos
UMCU sample 49	ND	Neg	167	Neg
UMCU sample 50	41.8	Pos	195	Neg
UMCU sample 51	29.1	Pos	6360	Pos
UMCU sample 52	23.0	Pos	6146	Pos
UMCU sample 53	25.5	Pos	5731	Pos
UMCU sample 54	ND	Neg	221	Neg
UMCU sample 55	37.0	Pos	5973	Pos
UMCU sample 56	25.5	Pos	6972	Pos
UMCU sample 57	31.8	Pos	6698	Pos

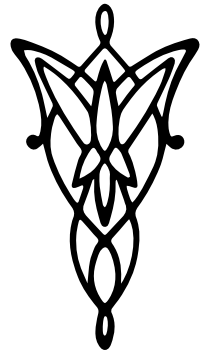
UMCU sample 58	37.6	Pos	7093	Pos
UMCU sample 59	24.1	Pos	7426	Pos
UMCU sample 60	39.4	Pos	150	Neg
UMCU sample 61	24.5	Pos	6269	Pos
UMCU sample 62	33.9	Pos	6213	Pos
UMCU sample 63	ND	Neg	165	Neg
UMCU sample 64	27.911	Pos	6262	Pos
UMCU sample 65	ND	Neg	221	Neg
UMCU sample 66	ND	Neg	185	Neg
UMCU sample 67	23.6	Pos	6865	Pos
UMCU sample 68	22.9	Pos	6938	Pos
UMCU sample 69	27.2	Pos	7046	Pos
UMCU sample 70	29.6	Pos	7241	Pos
UMCU sample 71	23.9	Pos	7802	Pos
UMCU sample 72	27.5	Pos	5164	Pos
UMCU sample 73	ND	Neg	157	Neg
UMCU sample 74	23.8	Pos	6287	Pos
UMCU sample 75	34.2	Pos	6314	Pos
UMCU sample 76	31.3	Pos	6438	Pos
UMCU sample 77	34.1	Pos	5700	Pos
UMCU sample 78	24.9	Pos	6865	Pos
UMCU sample 79	33.4	Pos	6832	Pos
UMCU sample 80	ND	Neg	200	Neg
Positive control	31.1	Pos	7205	Pos
Positive control 2	35.5	Pos	7172	Pos





Chapter 9

RNAPhi: a viral thermostable RNA polymerase for improved diagnostics



Manuscript in preparation:
Jurre A. Steens, Ane S. Andres, Bart Scholten, John van der Oost and Raymond H.J. Staals

Abstract

The use of CRISPR-Cas technology in diagnostics has been demonstrated in various studies with a variety of CRISPR-Cas systems. We previously reported on the development of SCOPE, a type III CRISPR-Cas-based diagnostic tool. A bottleneck in the initially established technology is the multi-step protocol that is required to combine the LAMP-based pre-amplification step with the CRISPR-Cas based detection and signal generation step. The solution for this issue would be the use of a thermostable RNA polymerase that would allow to convert the dsDNA amplicon into RNA that can be detected by the CRISPR-Cas system. Since such a polymerase is not commercially available, we set out to find such an enzyme. Here we describe the identification and characterization of a novel RNA polymerase from a thermophilic bacteriophage (phiFa). We showed robust RNA synthesis *in vitro* at elevated temperatures and a variety of reaction conditions. Using the potential of this polymerase to catalyse DNA-dependent RNA synthesis, we adapted RNAPhi for the DNA-to-RNA conversion step in SCOPE. As a proof of concept, we demonstrate the specific detection of the pathogenic bacterium *Pseudomonas aeruginosa*. With this, SCOPEv2 has been developed, a 1-step, one-pot CRISPR-Cas based diagnostic tool that successfully combines the LAMP pre-amplification step with the type III based CRISPR-Cas detection step.

Introduction

CRISPR-Cas is a well-characterized adaptive immune system of bacteria and archaea, providing sequence-specific neutralization of invasions by mobile genetic elements [21]. In addition, many examples have recently been reported of non-specific, collateral activity of CRISPR-Cas systems. Rather than protecting the infected cell by killing the virus, these systems can induce dormancy or even suicide of the infected host cell to protect the microbial community [7]. Hence, next to repurposing the specific targeting of nucleic acids for genome editing, the potential of the collateral activity of CRISPR-Cas for use in diagnostic applications was quickly recognized and several technologies have been developed to demonstrate this. Some examples of CRISPR-Cas diagnostics platforms are SHERLOCK and DETECTR. SHERLOCK can detect both RNA and DNA via a pre-amplification step with (reverse transcriptase) Recombinant Polymerase amplification ((RT-)RPA), followed by a Cas13-mediated detection step. In this detection step, the ssRNase collateral cleavage activity of Cas13 is used to cleave a reporter RNA construct (e.g. a fluorophore quencher pair connected by a small length of ssRNA) [72,368,372,373]. After the initial (reverse transcriptase) Loop-Mediated Isothermal Amplification ((RT-)LAMP) step, DETECTR used the Cas12a nuclease to specifically detect RNA and DNA target sequences from samples, after which the ssDNA collateral cleavage activity of the nuclease allows for cleavage of ssDNA reporter molecules. Whereas, using performing analyses using qPCR technology required complicated hardware and often can only be performed in a laboratory setting, the combination of CRISPR-Cas with isothermal amplification methods (RPA, LAMP) can provide very accurate and sensitive molecular diagnostics tools that can be used for point-of-care testing. The combination of these two different technologies is understandable, because isothermal amplification strategies by themselves have their limitations. One of these is the non-specific amplification when used with, for example, an intercalating dye as the readout, and as such often lead to misinterpretations of the results [374–377]. Verification of the amplicon using CRISPR-Cas can therefore mitigate these issues and can provide additional specificity when a template-derived region of the amplicon is targeted.

Previously, we demonstrated a proof-of-concept technology that uses LAMP in combination with T7 RNA polymerase and a CRISPR-Cas type III system, called SCOPE 78. The demonstrated proof-of-concept (SCOPEv1) showed great potential but is limited in terms of applicability due to the 2-step incubation protocol that is required, increasing hands-on time and risk of cross-contamination [78]. A key bottleneck is the maximum temperature of commercially available T7 RNA polymerase of 52°C, while both the LAMP reaction and the CRISPR-Cas type III system from Thermus thermophilus (TtCmr) work optimally at 65°C (NEB #M0658S). This prompted us

to look for an alternative DNA-dependent RNA polymerase that is active at 65°C. In contrast to the typical multi-subunit RNA polymerase complexes of prokaryotes and eukaryotes, bacteriophages often have a single-subunit RNA polymerase [378–382]. Since producing a single-subunit RNA polymerase is often simpler, the latter would be the preferred enzyme for our SCOPEv2 application.

In this study, we investigated the catalytic and structural properties of RNAPhi, the thermostable RNA polymerase from a thermophilic bacteriophage (phiFa), and tested it for its use in a streamlined one-step scopeDx assay, called SCOPEv2.

Results

PhiFa-44 acts as a DNA-dependent RNA polymerase

We set out to look for bacteriophages that can infect thermophilic hosts. The genome sequence of the *Thermus thermophilus*-infecting bacteriophage phiFa (MH673672.2) is reported by Artamonova *et al.*. This sequence revealed that the phiFa_44 gene encodes a putative RNA polymerase (QKE11339.1) [362]. This gene is located in the 5' long terminal repeat (LTR) of the phage genome, which has been suggested to be introduced first in the bacterial cells upon infection (**Figure 1A**) [362]. Furthermore, in Thermus-infected cells, it's type III CRISPR-Cas system showed a propensity for targeting genes transcribed in 5'LTR, including phiFa-44 [362]. Hence, phiFa RNA polymerase most likely plays an essential role, at least in the early stage of bacterial infection by transcribing early-phage genes.

A structural prediction of the phifa-44 encoded protein (PhiFa-44) was made using Alphafold2, after which homology with other known polymerases was investigated using Phyre2 and HHPred (**Figure 1B, S1**). One of the top results showed high structural similarity with the BH subdomain of QDE-1, a DNA/RNA-dependent RNA polymerase (DdRP/RdRP) from *Neurospora crassa*.

A structural prediction of the phifa-44 encoded protein (PhiFa-44) was made using Alphafold2, after which homology with other known polymerases was investigated using Phyre2 and HHPred (**Figure 1B, S1**). One of the top results showed high structural similarity with the BH subdomain of QDE-1 RdRP and DdRP from *Neurospora crassa* [383,384]. The structural similarity was confirmed by a multiple sequence alignment of the characteristic RNA and DNA catalytic core, the double-psi β-barrel (DPBB), with DdRPs from all domains of life, as well as viruses [385,386]. In particular, the catalytic DbDGD signature motif responsible for nucleotidyl transferase metal ion coordination appeared to be well conserved (**Figure 1C-E**). The combination of the structural analysis with the partial sequence alignment of the DPBB domain

allowed us to identify three aspartic acid residues (D366/D368/D370) belonging to the catalytic signature motif DbDGD in PhiFa-44 (**Figure 1C-E**) [384]. With this information, the wild-type phifa-44 gene and the D366A/D368A/D370A phifa-44 (dPhiFa-44) gene were cloned into an expression vector, recombinantly expressed in *E. coli* and affinity-tag purified.

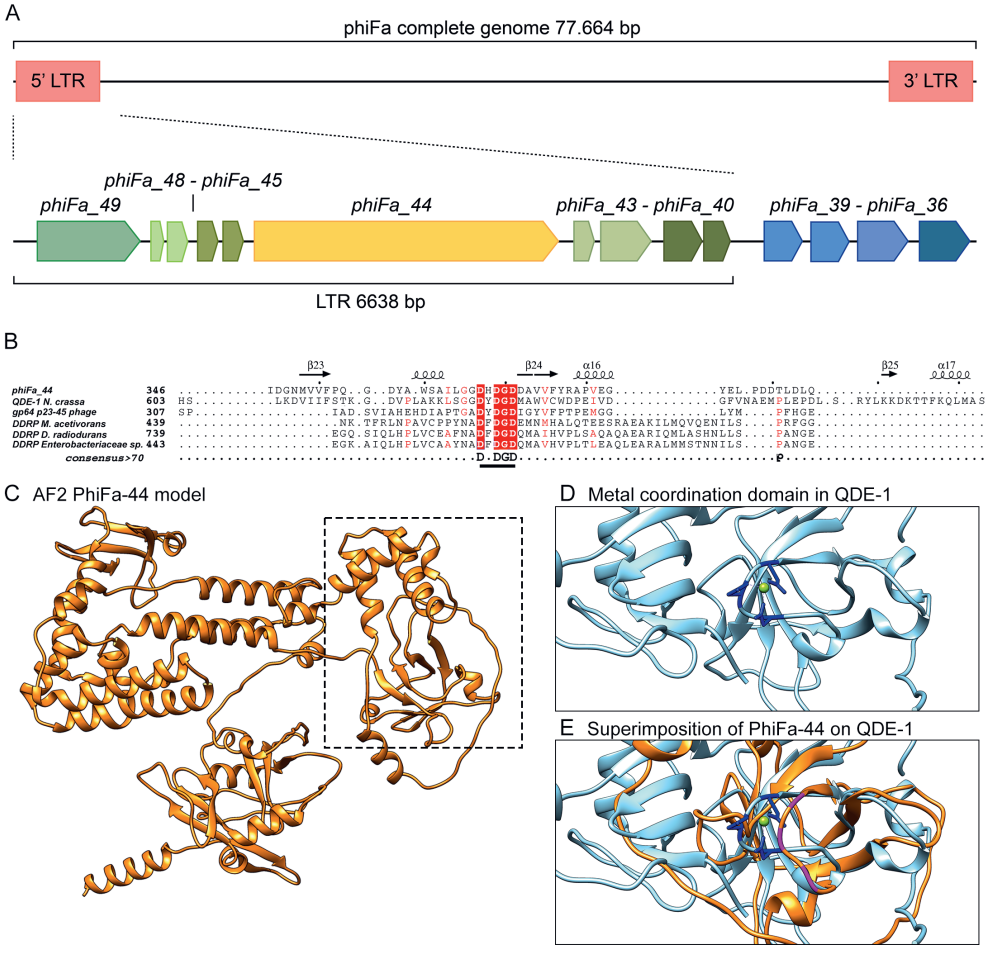


Figure 1. Analysis of phiFa phage gene phiFa_44. (**A**) Schematic overview of the genomic 5'LTR region of phiFa and neighbouring genes. (**B**) Multiple sequence alignment of a region of the double-psi β barrel (DPBB) domain of RNA polymerases from all domains of life. The DPBB catalytic motif, DbDGD, which is highly conserved in all RNA polymerases, is underlined. The residues are highlighted based on conservation: conserved in all sequences (bright red, highlighted), only in most sequences (light red). (**C**) AlphaFold2 structural prediction of PhiFa-44 protein. Dotted box highlights zoomed in region depicted in panels **D, E**. (**D**) Magnesium (green ball) coordinating aspartic acid residues (D1007, D1009, D1011) in QDE-1 polymerase (PDB: 2J7O). (**E**) Overlay of PhiFa-44 (orange) on the structurally simi-

lar region of QDE-1 (light blue). Three PhiFa-44 aspartic acid residues (D366, D368, D370) highlighted in pink are predicted to coordinate magnesium.

We anticipated that PhiFA-44 might be involved in transcribing genes from the phage's genome. The purified proteins were used to test the ability of PhiFa-44 to synthesize RNA *in vitro* from a linear, dsDNA template corresponding to the 237bp gene phiFa_43, located directly downstream the phiFa_44 gene. Since the phiFa phage can infect *Thermus thermophilus*, a thermophilic bacterium with an optimal growth temperature between 65 and 72°C, the initial assays were performed at 65°C [387]. The results show the generation of an RNA transcript by PhiFa-44, when a dsDNA template is supplied (**Figure 2A**). The addition of EDTA prevented the synthesis of RNA, demonstrating the dependency on divalent metal ions to catalyse the reaction (**Figure 2A**).

To probe which divalent metals stimulate synthesis of RNA in vitro by PhiFa-44, we conducted similar in vitro transcription (IVT) reactions in the presence of magnesium, manganese, zinc and cobalt. Both DNA and RNA polymerases, like PhiFa-44, are known to depend on divalent metal ions as cofactors for catalysis, and often show a preference for one (or more) metal ions. For PhiFa-44, most divalent ions tested had no impact on RNA synthesis, whereas both Mn2+ and Co2+ substantially stimulated transcription activity (**Figure 2B**). In agreement, mutation of the metal-coordinating residues in PhiFa-44 (dPhiFa-44) abrogated RNA synthesis completely (**Figure 2B**). These results show that PhiFa-44 is thermostable, metal-dependent RNA polymerase using at least linear dsDNA molecules as a template.

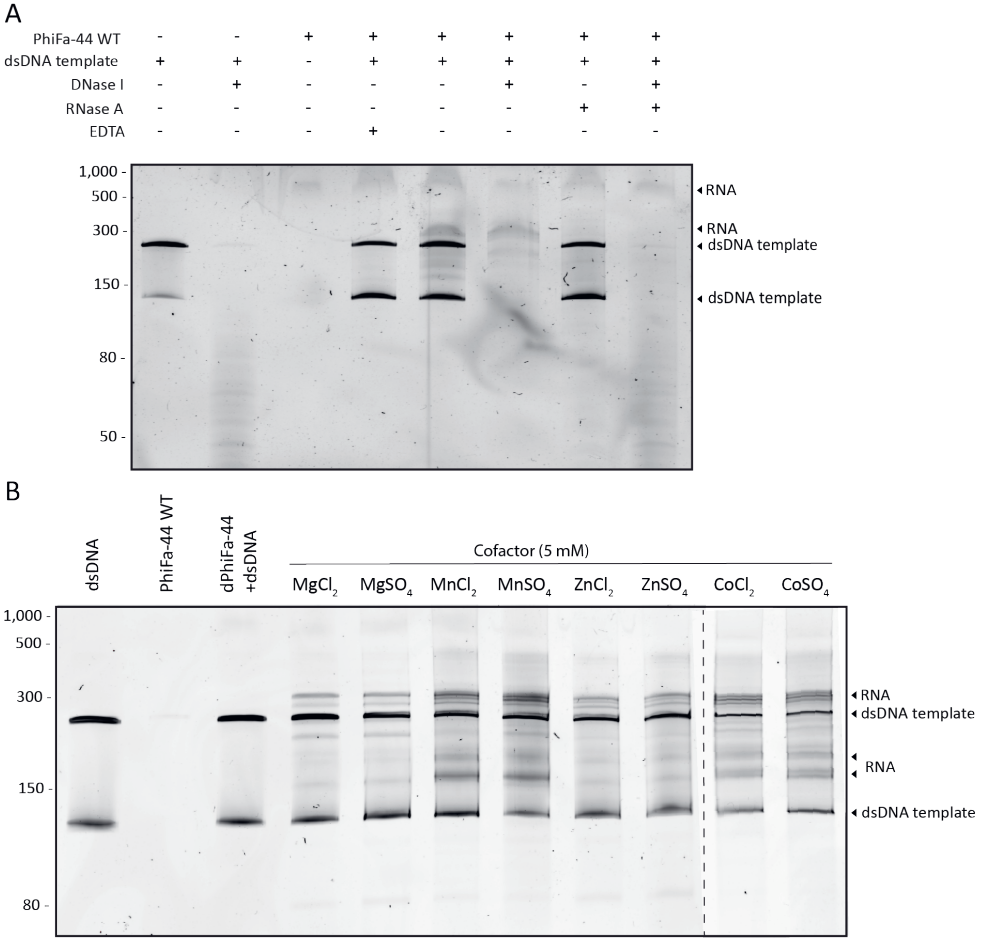


Figure 2. PhiFA-44 is a thermostable, metal-dependent RNA polymerase. Denaturing gel electrophoresis analyses of in vitro transcription reactions with PhiFa-44 using (**A**) dsDNA template at high temperatures (65C°). Addition of EDTA (5 mM) was used as a negative control (**B**) a range of divalent metal cations. A catalytically inactive PhiFa-44 (dPhiFa-44) was used as negative control. The reactions were visualized in a 15% Urea-PAGE gel after SYBR Gold staining. Low Range ssRNA ladder (NEB # N0364S) was used as marker.

Finally, we investigated the optimal incubation temperature and pH for PhiFa-44-catalyzed IVT reactions. We observed that PhiFa-44 transcribed the template DNA in all temperatures (30-80°C) tested (**Figure 3A**). Surprisingly, the highest activity was observed in the 35-40°C temperature range, which deviates from the optimal growth temperature of its native host *T. thermophilus* (65°C). Additionally, to test whether pH affects the stability of PhiFa-44 and/or influences its transcription activity, we tested different pH values within the buffering capacity of HEPES (pH 6.7-8.6) in 0.5 increments (**Figure 3B**). We observed the production of RNA peaked at pH levels between 6.7 and 7.2. Together, these results confirm the ability of PhiFa-44 to synthesize RNA from a linear double-stranded DNA template in a (seemingly) promotor-independent manner, with an optimal temperature range between 35-40°C and a pH between 6.7 and 7.2.

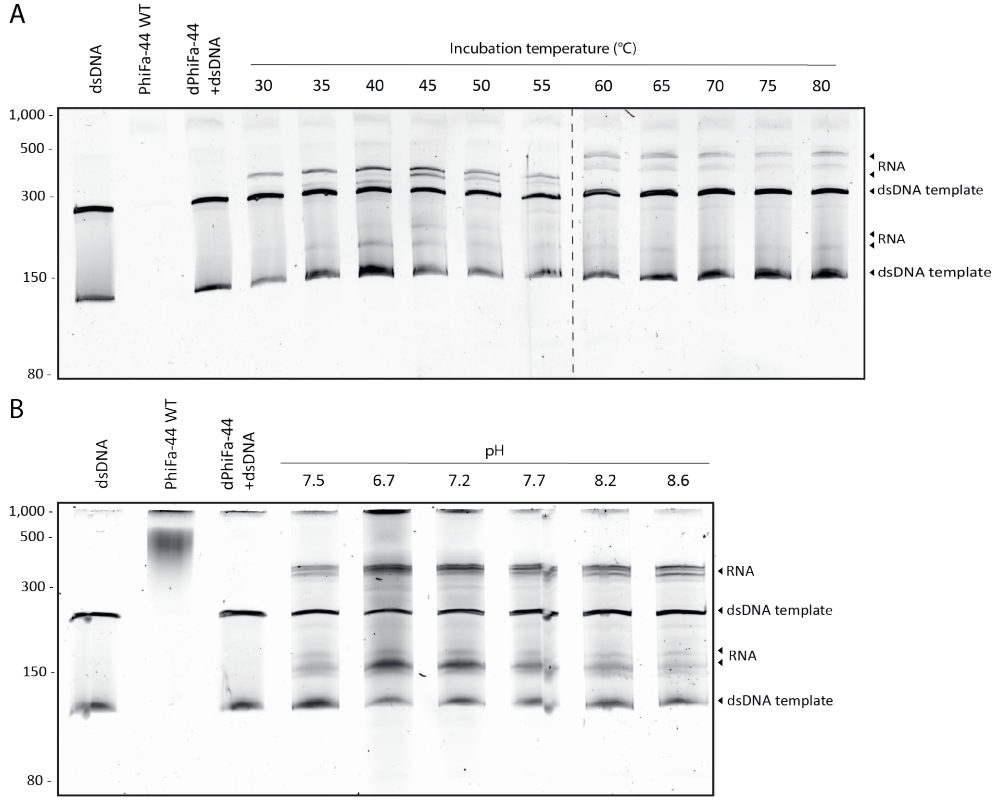


Figure 3. Temperature and pH optima for RNA synthesis by PhiFa-44. (**A**) PhiFa-44 RNA polymerase activity in a wide range of temperatures (30-80°C). A catalytically inactive PhiFa-44 (dPhiFa-44) was used as negative control. (**B**) PhiFa-44 RNA polymerase activity in a range of HEPES buffer pH. A catalytically inactive PhiFa-44 (dPhiFa-44) was used as negative control. The reactions were visualized in a 15% Urea-PAGE gel after SYBR Gold staining. Low Range ssRNA ladder (NEB # N0364S) was used as marker.

Next, we investigated the range of substrates that can be used by PhiFa-44 as a template. The homology of PhiFa-44 with QDE-1 RNA polymerase (**Figure 1B, C**), prompted us to try a wide variety of substrates, as QDE-1 polymerase has been reported to utilize both single-stranded RNA (ssRNA), and single-stranded DNA (ssDNA) but not double-stranded DNA (dsDNA), *in vitro* as well as *in vivo* [384,388]. To investigate whether ssDNA also serves as a substrate for PhiFa-44, we conducted IVT reactions with PhiFa-44 and a Cy3-labelled ssDNA substrate and analysed them by native PAGE gel electrophoresis (**Figure 4A**). When the template was incubated with PhiFa-44 for 60 minutes, the migration of the band representing the ssDNA template was severely reduced, indicating a strong interaction of the template with PhiFA-44 and/or the formation of a DNA:RNA duplex, due to transcription of the complementary RNA. To differentiate between these possibilities, we performed similar reactions for 60 minutes, after which samples were supplied with either DNase I (DNA-specific nuclease), RNase A (RNA-specific nuclease) or RNase H (nuclease specifically cleaving RNA in DNA:RNA duplexes). In agreement with the formation of a DNA:RNA duplex, addition of RNase A partially restored the migration of the band representing the ssDNA template, whereas addition of and RNase H completely restored it. To demonstrate this more clearly, we conducted a PhiFA-44 IVT reaction with a ssDNA template and monitored production of the DNA:RNA duplex over time (**Figure 4B**). The results strongly suggest that the band representing the DNA:RNA duplex quickly appeared (within a few minutes) and then increased in intensity over time, further confirming the rapid RNA synthesis by PhiFA-44.

Surprisingly, analysis of a similar reaction mixture on UREA-PAGE showed that the same shifting of the Cy3-ssDNA is observed with a higher proportion shifting in the presence of MnCl2 when compared to $MgCl_2$ (**Figure 5**). Performing the same experiment with a Cy5-ssRNA as template, a similar shift in the band and effect of $MnCl_2$ is seen (**Figure 5**). These results indicate that the apparent substrate specificity of PhiFa may consist of at least dsDNA, ssDNA and ssRNA.

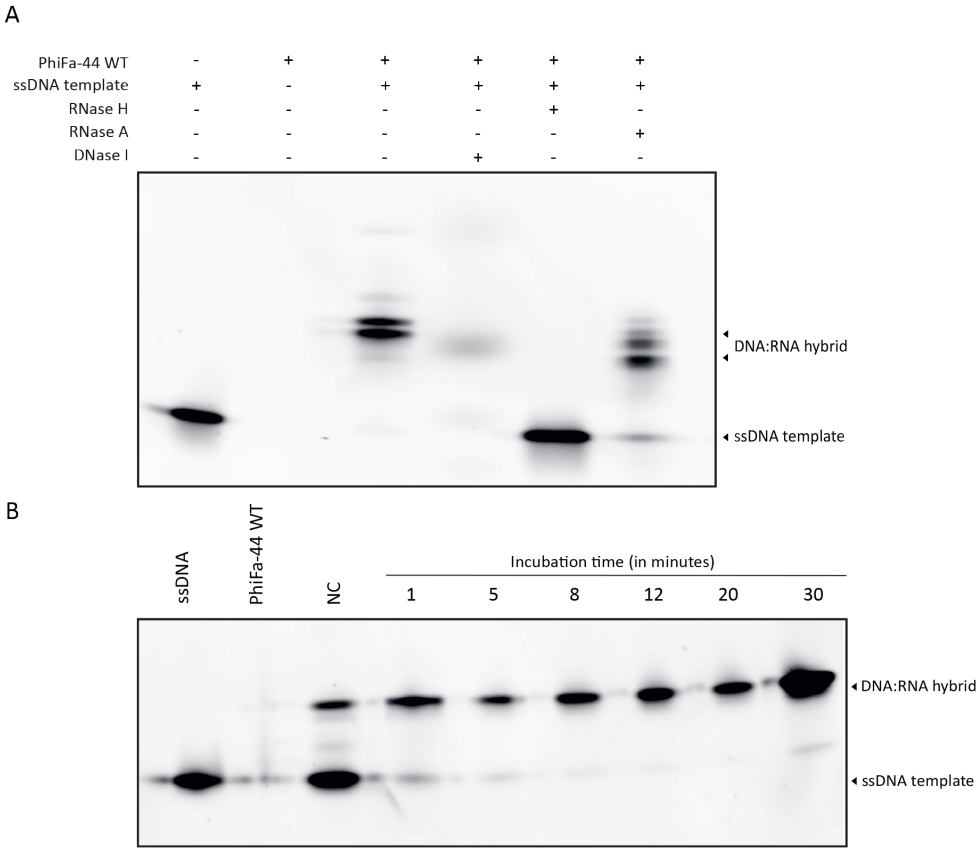


Figure 4. ssDNA template usage by PhiFa-44. (**A**) PhiFa-44 RNA polymerase activity in the presence of a ssDNA template. (**B**) Time-series of ssDNA template usage by PhiFa-44 INC: no incubation, no added $MnCl_2$. Cy3-labelled ssDNA (50 nt) was visualized using a 4-20% Native-PAGE and fluorescence measurement at EX/EM of 548/552 nm

Development of SCOPEv2

Chronic Obstructive Pulmonary Disease (COPD) is a chronic lung disease causing long-term respiratory problems in humans, which affects millions of individuals in Europe alone [389]. Typical symptoms include shortness of breath and chronic coughing, but symptoms can flare up during exacerbations caused by acute lung infections [390,391]. In these patients, acute lung infections (occurring 1-2 times per year) are a major cause of morbidity [390]. These infections are associated with accelerated loss of lung function and impairment of quality of life, meaning timely detection is important to minimize the impact of this disease. Current identification methods are based on sputum sample culturing to identify the pathogen, which takes 2-3 days. Other options (such as chest X-Ray and CRP blood testing) rapidly reveal symptomatic information about the infection but do not identify the causative agent, leading to inaccurate treatment [391]. **Pseudomonas aeruginosa**, an opportunistic pathogenic bacterium which mainly infects the upper airways is associated with acute lung infections [392,393]. *P. aeruginosa* is well-equipped with genes to survive in the airways and often contains resistance genes against commonly used antibiotics, making it difficult to cure/treat [394].

Building forward on the SCOPEv1 diagnostic platform, we opted to build a rapid and accurate diagnostic assay to specifically detect *P. aeruginosa* [78]. To show the proof of concept, the assay was first designed and validated on synthetic target material. For the target region the toxA gene was selected which encodes exotoxin A, an important virulence factor of *P. aeruginosa*, and has been used in multiple P. aeruginosa PCR-based detection methods [395,396]. The scope*Dx* assay reaction starts with a pre-amplification step to increase the sensitivity of the reaction, which is done using LAMP. For this step, a design is made to specifically amplify the toxA gene, consisting of a LF, LB, FIP, BIP, LoopF and LoopB primer (**Table S1**).

Firstly, LAMP-only reactions were performed to assess their performance on a synthetic target, Pae1, which contains a partial Tox A gene (of reference genome *P. aeruginosa* PAO1). Reactions were performed at 65 °C for 1hr and target amplification was measured using an intercalating LAMP dye. A dilution range of the synthetic target material was tested in duplo for each LAMP reaction to assess time to detection (TtD, the time to reach 20% of maximal fluorescence) and limit of detection (LoD, the lowest target dilution which is still amplified). The TtD for the duplicate reactions with 10.000, 100, 10, 1 copies/µL is 20-20, 28-30, 29-35 and 38 respectively with an LoD of 10-1 copies/µL (**Figure 6A**). Surprisingly, reactions with a negative control target (without a LAMP priming sequence) also resulted in a signal, although significantly later at 44 and 52 minutes.

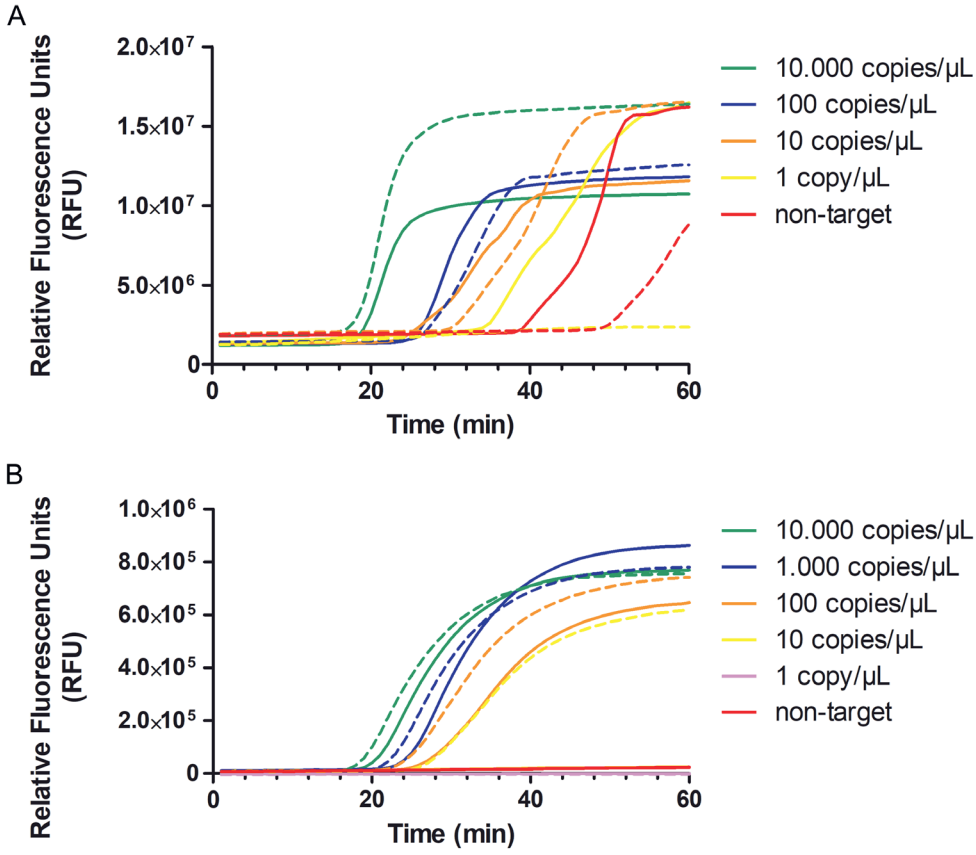


Figure 6. ***P. aeruginosa*** **synthetic gene detection.** (**A**) LAMP reactions on Pae1 synthetic target serial dilution. Dashed lines of the same colour indicate duplicate reaction. (**B**) SCOPEv2 reactions on Pae1 synthetic template serial dilution. No signal is observed in the non-target reactions. FAM fluorescent measurements are done at 1 min intervals. Dashed lines indicate technical replicates.

Despite the unspecific amplification of the non-target DNA at later timepoints, the designed LAMP primers were deemed good enough for subsequent reactions. For this next step, a CRISPR RNA (crRNA) was designed for specific detection by the *Thermus thermophilus* type III-B CRISPR-Cas complex following LAMP amplification of the synthetic Pae1 target. From our previous work, we determined that the CRISPR-Cas type III-B system tolerates mismatches between the crRNA and the target RNA in most regions [78]. There is one region however, the CAR region, that is very sensitive to single base mismatches, making it ideal to distinguish between very similar target RNAs. Taking this into account, the CAR-region of the crRNA was placed between Lb and B2. Using the designed crRNA, the CRISPR-Cas complex was reconstituted and used in a SCOPEv2 assay. The one-pot SCOPEv2 assays were

incubated at 65°C for 1 hour, with FAM fluorescence measurements at one-minute intervals. Again, the synthetic Pae1 target is detected in duplo at decreasing concentrations to assess TtD and LoD of the assay. The TtD for the duplicate reactions with 10.000, 1.000, 100, 10 copies/µL is 23-22, 28-25, 31-28 and 31 respectively with an LoD of 10-100 copies/µL (**Figure 6B**). Notably, none of the non-target reactions gave any signal. Together, these results show that RNAPhi (RNA polymerase PhiFa-44) can be used in combination with the LAMP and the CRISPR-Cas system in a one-pot SCOPEv2 approach, which greatly increases the usability of the SCOPE technology.

Discussion

Bacteriophages exhibit a plethora of highly diverse mechanisms to successfully use their host's cellular machinery for their own benefit. One of the main phage infection targets is the transcriptional machinery of the host, which is often redirected to predominantly transcribe phage genes, promoting viral infection. For instance, the P23-45 phage infects the thermophile *Thermus thermophilus* by switching the promoter specificity of the host's RNA polymerase (RNAP) to those more abundant on the phage genome, while also directly inhibiting T. thermophilus RNAP activity [397,398]. Alternatively, some phages are not completely dependent on hijacking the host's RNAP, like *E. coli* phage N4. Instead, this phage encodes two distinct RNA polymerases to transcribe early- and middle-phage genes, while redirecting E. coli's RNAP activity for the transcription of late viral RNA [399,400].

All the above-mentioned phage mechanisms depend on the activity of DNA-dependent RNA polymerases (DdRP), which synthesize RNA from a DNA template. The most studied, phage-derived DdRP family is represented by T7 RNAP. All T7-like RNAPs (e.g., T3, SP6) are single-subunit RNAPs that show high specificity for their native promoters, as well as transcribing long strands of RNA without releasing their template. Their single-subunit nature, together with their high specificity and processivity has made T7-like RNAPs attractive for various biotechnological applications. For instance, T7-like RNAPs are routinely used for *in vitro* synthesis and labelling of RNA, as well as for in vivo gene expression systems. However, these polymerases have the drawback of not working at elevated temperatures, making them unsuitable for other applications, such as their use in the SCOPE diagnostic platform. In this study, we therefore explored potential other single-subunit RNAP candidates, eventually leading to the discovery and characterization of the phage-derived PhiFa-44 polymerase.

Surprisingly, the predicted structure of PhiFa-44 did not share much structural homology with known (RNA) polymerases from other phages. This could be because viral proteins are typically highly diverse and underrepresented in most databases. When trying to find more remote PhiFa-44 homologs (using the HHpred tool from the Max Planck Institute for Biology), we found significant hits with three plant RdRPs, and the QDE-1 RNA polymerase from *N. crassa*, a filamentous fungus. The QDE-1 polymerase has been demonstrated to use ssDNA as template to generate ssRNA, as well as to transcribe dsRNA from ssRNA [384,388]. High structural similarity was found in the NTP and metal cation coordinating parts of the PhiFA-44, and conservation of several residues in these parts allowed us to construct a catalytically dead version of PhiFa-44, dPhiFa-44 (**Figure 1C-E**). Indeed, dPhiFa-44 does not show any activity in the tested conditions, indicating that indeed the three identified aspartic acid residues, which belong to the characteristic DbDGD motif, are essential for the catalytic activity of PhiFa-44 (**Figure 2B**). Further testing of its optimum reaction conditions revealed that PhiFa-44 (**i**) has a rather broad temperature optimum curve with a peak around of ~45°C, (**ii**) has a rather broad pH optimum with a preferred pH of ~6.7, and seems to prefer $MnCl_2$ as divalent cation (**Figure 2, 3**). However, the RNA synthesis activity of PhiFa-44 appears to be possible in a wide range of conditions. Unexpectedly, we demonstrate that PhiFa-44 can generate RNA from DNA templates in a seemingly, promoter-independent manner, which is not an issue for its use in the envisioned application (**Figure 2A**). Despite several attempts, no preferred promoter sequence has yet been revealed (not shown). Additionally, it was found that on top of dsDNA, both ssDNA and ssRNA are suitable templates for generation RNA products by PhiFa-44 (**Figure 4**). Surprisingly, when analyzing these products on denaturing UREA-PAGE gels, the supposed DNA-RNA and RNA-RNA duplexes seems to remain intact (**Figure 5**). This could be due to incomplete strand-separation, despite the strong denaturing conditions during electrophoresis and boiling of the sample before loading it on the gel, which would mean they are very stable. Alternatively, phiFa-44 might be using the newly synthesized RNA as a template and initiating transcription from the 3' end, thereby elongating the already existing RNA transcript (and creating a covalent bond between RNA strands). This type of transcription is called back-priming and it is common for viral RNA polymerases [401,402]. This is observed for the homologue QDE-1 polymerase in vitro as well [388]. These hypotheses should be further investigated in a follow-up study and should include these polymerases as controls.

In summary, it could be that the promoter-independent activity we are observing *in vitro* is a background activity of PhiFa-44, meaning its activity might be much higher in its native biological context *in vivo*. Whether a promoter or other co-factors (e.g. analogs of sigma factors) play a role is yet to be determined and will need further study. In vivo phage challenge experiments using (for example) *T. thermophilus* as

a host, followed by RNA-seq or Chip-seq, could be performed to map highly transcribed regions of the phiFa (or T. thermophilus) genome, facilitating the identification of its preferred promoter.

With the obtained insights into some of the activities of PNAPhi, we were able to start adapting its RNA polymerase activity for use in the SCOPE diagnostic platform, resulting in SCOPEv2. As a proof on concept, the detection of a gene from *P. aeruginosa* was chosen. First, a LAMP primer set was successfully designed and tested to function as pre-amplification step. Notably, the non-template control reaction resulted in a positive signal, albeit significantly later than the expected true positive reactions (**Figure 6A**). This observation of non-specific amplification highlights one of the problems typically associated with these kind of isothermal amplification strategies. The tested LAMP primer set was used in the new one-pot SCOPEv2 assay and was able to detect a synthetic template concentration of 10.000 copies/µL in 22.5 minutes with a detection limit of only 100 copies/µL (**Figure 6B**). Interestingly, none of the non-template control reactions gave any signal, emphasizing the potential of the SCOPEv2 strategy.

Further steps could include testing of patient samples, which would provide multiple new challenges. Firstly, the sample that would be obtained from the patient has a complex composition, e.g. including a plethora of commensal micro-organisms which should not be detected. Given that ToxA is specific to *P. aeruginosa* this is not expected to cause off-target issues. Next, *P. aeruginosa* is an opportunistic pathogen, which will likely be present in non-exacerbated patients although in far lower copy numbers. It will be of interest to assess whether quantitative testing with SCOPEv2 can be realized in the future.

As the developed SCOPEv2 assay design requires only a single incubation step, as compared to the two-step protocol in earlier work, the applicability of the technology is greatly improved. Furthermore, since the SCOPEv2 is incubated at a single temperature and requires simple read-out hardware (single temperature, fluorescence measurement), it is suitable for on-site application, which could further minimize the time to diagnosis and accurate treatment for a variety of use-cases.

Materials and Methods

Identification of phiFa-44 RNA polymerase in silico

The predicted phiFa-44 RNA polymerase aminoacidic sequence was retrieved from Uniprot (A0A6M8E0E6). First, we used Phyre2 to detect potential homologues of full-length phiFa-44 at an amino acid sequence level. Within the top results, which all turned out to be RNA polymerases, was the DNA-dependent RNA polymerase (QDE-1) from *Neurospora crassa*, with a confidence score of 97.1/100 and 21% sequence identity. Next, the 3D structure of phiFa-44 was predicted using AlphaFold2 and compared to the 3D protein structure QDE-1 (PDB: 2J7O) using the Dali server. The resulting model was visualized using Chimera X.

Purification of PhiFa_44 WT and mutant versions

E. coli BL21(D3) cells expressing Strep-MBP-TEV-PhiFa-44-His6 or Strep-MBP-TEV-dPhiFa-44-His6 constructs were grown in LB media supplemented with 1% glucose at 37°C and 120 RPM . After reaching an OD_{600}= 0.6, cells were cold-shocked on ice for 45-60 minutes, and subsequently induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20°C and 120 RPM for 16-18h. Cells were harvested at 8,000g for 10-minutes at 4°C. Subsequent, resuspension of the cell pellet with Equilibration buffer (50 mM, 100 mM Tris-HCl, pH 8.0) and a cOmplete protease inhibitor cocktail tablet (Roche) was followed by cell lysis by sonication (30% amplitude, 1s ON 1s OFF for 20 min). The soluble fraction of the lysate was separated by centrifugation (35,000g for 45 min at 4°C), filtered (0.22 μm) and loaded in the HisTrap (Cytiva) column, which had been pre-equilibrated with Equilibration buffer. After washing with Equilibration buffer supplemented with 20 mM imidazole, elution was performed using Equilibration buffer containing 500 mM imidazole. The removal of imidazole from the collected elution fractions was performed by buffer exchange using Amicon® Ultra filter columns (30,000 Da MW cut-off). To remove the Strep-MBP-tag from the PhiFa_44-His6 construct, TEV protease was added at a 1:50 ratio and the mix was incubated at 4°C overnight. The Strep-MBP-tag and PhiFa_44-His6 were separated by reverse capture chromatography using StrepTrapXT (Cytiva) column using Equilibration buffer. Next, the sample was loaded into HiTrap Heparin HP (Cytiva) column, pre-equilibrated with Equilibration buffer, and eluted using a 20-100% 1M NaCl gradient. After concentrating the sample using Amicon® Ultra filter columns (30,000 Da MW cut-off), a final size-exclusion chromatography step was performed (HiLoad Superdex 200pg 16/600 (Cytiva)) using 50 mM KCL, and 20 mM HEPES (pH 7.5) as eluent. After SDS-PAGE analysis, samples were quantified and stored at -80°C.

In vitro RNA transcription assays

The nucleic acid templates used in these assays (supplementary **Table S1**) were purified from a denaturing PAGE using the ZR small-RNA PAGE Recovery Kit (ZymoResearch, R1070), following manufacturer's instructions. Unless indicated otherwise, the RNA transcription assay was conducted in phiFa-44 activity buffer (50 mM KCl 20 mM HEPES pH 7.5, 25 mM rNTPs, 20 mM $MgCl_2$), with 55 pM of nucleic acid template and 500 mM of phifa-44 wild-type or mutant. The reaction was performed at 65°C for 2 hours. In some reactions, nucleases (DNase I, RNase A, RNase H) were added after sample denaturing (95°C for 10 minutes) and incubated at 37°C for 30 minutes. Afterwards, colourless RNA loading dye was added (47.5% formamide, 0.01% SDS, 0.5 mM EDTA), boiled at 95°C for 10 minutes and loaded into a 15% denaturing PAGE (8 M urea). The denaturing gel was run at 15-20 mA for 1-2 hours and visualized by adding SYBR Gold dye and using a gel scanning imager (GE Amersham Typhoon).
Unless indicated otherwise, all the *in vitro* RNA transcription assays were carried out following the steps and reaction components mentioned above.

Loop mediated isothermal amplification (LAMP) reactions

LAMP primer sets including loop primers were designed using Primerexplorer v5 (Fujitsu Limited) (**Table S1**). LAMP reactions were performed using WarmStart NEB kit (E1700L) in 20 µL reactions. Final primer concentrations were 0.2 µM, 1.6 µM and 0.4 µM for the F3/B3, FIP/BIP and LoopB/LoopF primers respectively. Reactions were incubated at 65°C for 60 min, with FAM fluorescence being measured at 1 min intervals in thermocycler QuantStudio 1 (ThermoFisher, A40427).
Limit of detection is taken as the lowest template dilution which is detected in all replicates. Time to detection is take as the time-point where 20% of the maximum fluorescence (baseline fluorescence subtracted) of the individual reaction is reached.

SCOPEv2 assays

Pae1-specific CRISPR-Cas complex was prepared as TtCmr46 complex described by Steens *et al.* (2021). Pae1 specific crRNA was designed in-house and can be found in **Table S1**. One-pot SCOPEv2 assays were combined by adding 10 µL 2x NEB Warmstart mix (E1700L), 1 µL Set 1 primermix (100 µM), 0.5 µL Pae1 TtCmr complex (2.5 µM), 0.6 µL TtHB144 (40 µM), 0.5 µL phiFA (25 µM), 0.2 µL rNTPs (NEB N0466L, 25 mM), 0.25 µL RNAse Alert probe (Invitrogen RNAse Alert QC System v2 (4479769), 20 µM) and 1.95 µL uLtrapure water per assay. 15 µL of this mixture is combined with 5 µL of template solution. Incubation, read-out and detection time and limit method are identical to LAMP-only reactions.

Supplementary information

A

```
>PhiFa_44
MNKTELVKKLIEEKTAELAVRPVYTGRIDEEGQWTPTSLHERYQDKVLVVPAPE
EGEPVQTIGEYEFVGGNSNRIVYIHRSIIKQWLEVAGGDSHWGAAVVLRPGKLA
SHNAVFSWEGYQLEESDEGLLATMDERDWKVLLEAYDLDPEVRLVQVTVFDLEQ
GVSAKGTLRKVLPGERPVVFPNSWKKFGNLRTQFMSILTTDVHYNPRASVNLQE
LTYYVQEAPAQWVFDALDEEIERIINLKTSPWLTTGYLTRLGFSPAWNSTLTAL
KESLCGELKKLLRRIKLSGKYGLRAKSTSSHLLRDGELLLPAYAKKHVKVGDFV
TVSRNPALPSQGWAKYRVTGFIDGNMVVFPQGDYAWSAILGGDHDGDDAVVFYR
APVEGYELPDDTLDLQAIKPVARKLDANTVEARIQRWQNEVSVNIGQFDLAARR
LLAANKLHREEAKLLSVAIQTAISLKKRVARLEDQSWWPRVAELLEESKKLAGT
TWVDMIRDGREPVGAPDWVLTIYQKVHEAIKKVQKLEPRLWLTEAKVREFAKGA
TVPASCEALIQYREKLLKAKAIAVQNEDGDTVARLNKELKEFIHVEAPKHLLGI
DEEDWKQFSRWAIGNTHISEWVFWCHPAILEELYDRFGKQTIKAIALGGTPNLA
ESEVIELEDQELIDRVFEVEGVKYALLDETGYLKAGRLKVTKVRKSVVEFEYA
```

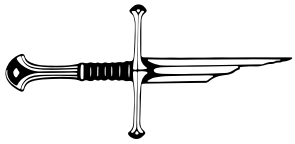
B

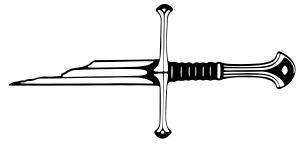
#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c7w88A	Alignment		96.5	19	**PDB header:** plant protein **Chain:** A: **PDB Molecule:** rna-dependent rna polymerase; **PDBTitle:** cryoem structure of open form zmrdr2 at 3.5 angstroms resolution **PDB Entry:** PDBe RCSB PDBj Phyre2 Run Investigator
2	c7w84A	Alignment		97.3	19	**PDB header:** plant protein **Chain:** A: **PDB Molecule:** rna-dependent rna polymerase; **PDBTitle:** cryoem structure of apo form zmrdr2 at 3.4 angstroms resolution **PDB Entry:** PDBe RCSB PDBj Phyre2 Run Investigator
3	c7rqzA	Alignment		96.9	18	**PDB header:** transcription **Chain:** A: **PDB Molecule:** rna-dependent rna polymerase 2; **PDBTitle:** structure of rna-dependent rna polymerase 2 (rdr2) from arabidopsis2 thaliana **PDB Entry:** PDBe RCSB PDBj Phyre2 Run Investigator
4	c2l7nA	Alignment		97.1	21	**PDB header:** hydrolase **Chain:** A: **PDB Molecule:** rna-dependent rna polymerase; **PDBTitle:** structure of the rnai polymerase from neurospora crassa **PDB Entry:** PDBe RCSB PDBj Phyre2 Run Investigator
5	c5fswC	Alignment		96.1	21	**PDB header:** transferase **Chain:** C: **PDB Molecule:** rna dependent rna polymerase qde-1; **PDBTitle:** rna dependent rna polymerase qde-1 from thielavia terrestris **PDB Entry:** PDBe RCSB PDBj Phyre2 Run Investigator

Figure S1. PhiFa_44 sequence and Phyre2 structural analysis. (**A**) Amino acid sequence of phiFa_44 gene. (**B**) Results of Phyre2 structural homology analysis.

Table S1. Used oligo's and synthetic gene fragments.

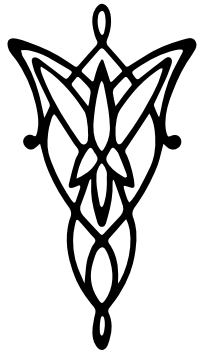
Primer name	Sequence (5'>3')
Set1 F3	GCGATGCCACCTTCTTCG
Set1 B3	TGCCTTCCCAGGTATCGT
Set1 FIP	TGACCCCGGCATGGCTGATG-TCAGGGCGCACGA-GAG
Set1 BIP	GAAAAGCGCTGGAGCGAATGGGG-GGCGAGG-TAGTTGTAGACC
Set1 LoopF	TCGGCTGCATCTCGTTG
Set1 LoopB	AGCGGCAAGGTGTTGTGC
Pae1 crRNA	AUUGCGACGCUCGACCCGCUGGACGGGGUCUA-CAACUACCUCGCCC
Pae1 synthetic target gene	TTACTAATAGCGCGATGCCACCTTCTTCGTCAGGG-CGCACGAGAGCAACGAGATGCAGCCGACGCTCG-CCATCAGCCATGCCGGGGTCAGCGTGGTCATGGC-CCAGGCCCAGCCGCGcCGGGAAAAGCGCTGGAG-CGAATGGGCCAGCGGCAAGGTGTTGTGCCTGCTC-GACCCGCTGGACGGGGTCTACAACTACCTCGC-CCAGCAGCGCTGCAACCTCGACGATACCTGGGAAG-GCAAGATCTACCGGGTGCTCGCCGGCAACCCGG-CGAAGCATGACCTGGACATCAAGCCCACGGTCAT-CAGTCATCGCCTGCACTTCCCCGAAG
dsDNA template phifa-43 (5' -> 3')	cttgcggagggttctcctccaaaaaggaggtgtacgaacatgctgaagcttgt-gcgggctactgtctacggggagttcagcgactacagaggctaccgggtagct-gaaatcgaagccgaaccgctgagcagggagttcatccgcaggctccgggattc-cgagttcttcttcgacggtaccgctcgactgcagtttactgacggtaccgtcaagtat-gtgcccattcaggagaccgatgtcgcgctcttgaacaacctgctggaggtcttg-gaagcttaa
ssDNA template (5' -> 3')	Cy3-CGATGGCCGTAGATAAGGCGCCCGGGGACGAC-CACGTCAAGGCGCAGTTC
ssRNA template (5' -> 3')	Cy5-GCUACCGGCAUCUAUUCCGCGGGCCCCUGCUG-GUGCAGUUCCGCGUCAAG





Chapter 10

Summary and general discussion



Thesis summary

CRISPR-Cas type III systems have emerged as an interesting area of research within the field of CRISPR-Cas because of their unique and multifaceted attributes. The ways by which CRISPR-Cas systems may have evolved due to constant evolutionary pressure from invading bacteriophages is discussed in this thesis and put into context of a co-evolutionary framework to explain the extreme variety of approaches that can be found in CRISPR-Cas in general. Of main interest in this research has been the messenger molecule-mediated signalling pathways that type III CRISPR-Cas are known for. Attempts have been made to increase our understanding of the fundamental knowledge on the control of these signal transduction pathways, both at structural and functional level. Methods for characterizing and mapping the signalling pathways have been developed that may be useful for future research. Furthermore, a novel type III CRISPR-Cas system with associated effector proteins has been studied, combining structural, biochemical and *in vivo* approaches. Lastly, fundamental research into these type III CRISPR-Cas systems has led to the development of a technology that can be utilized for the detection of nucleic acids. In summary, this thesis combines research prompted by questions regarding fundamental understanding with research that was conceptualized with applications in mind.

Chapter 1 discusses evolution through natural selection that relies on genetic variation that can lead to beneficial traits for survival. This evolutionary process is at the core of the interaction between bacteriophages (viruses) and bacteria, explaining how they coevolve in an ongoing arms race. A multitude of phage defence systems have been discovered to protect the host from falling prey to the phages. Sequence-specific recognition of the genomic information of phages is the second most wide-spread mechanism, with CRISPR-Cas being the focus of this thesis. The discovery of CRISPR-Cas systems, characterized by repeated palindromic DNA sequences (repeats) interspaced with variable sequences (spacers), led to the recognition of the only known adaptive anti-bacteriophage immune system. Genes coding for proteins (Cas) adjacent to CRISPR arrays were identified as part of this defence mechanism. The classification of CRISPR-Cas systems involves two main classes, further divided into types and subtypes based on the composition of effector modules and other characteristics [97]. Generally, all systems function through a three-step mechanism: adaptation, expression and interference [31]. Especially at the interference step, type III CRISPR-Cas systems are rather unique in that they use a multifaceted immune response including specific degradation of (target) RNA and the collateral cleavage of DNA, but also the production of signalling molecules. Altogether, this immune response can be employed as a way to sacrifice the infected individual to protect the population.

Chapter 2 reviews the general concept of CRISPR-Cas immunity but focusses on the coevolution of this immune system with phages. CRISPR-based immunity is not the only weapon bacteria have to defend against phages, as indicated by the absence of CRISPR-Cas systems in approximately half of the prokaryotic genomes and the presence of >100 alternative defence mechanisms, like surface mutation (sm) and restriction-modification (R-M) systems. Apparently required for survival, bacteria often possess multiple defence options, and their choice of defence strategy depends on their environment and on their history of viral infections – including exposure to anti-CRISPR systems. Short-term resistance against phages is typically achieved through mutation or loss of cell surface receptors to stop attachment of the phages, and the long-term prevalence of CRISPR-Cas is influenced by the fitness advantage conferred and the ecological context. Key in gaining CRISPR-Cas-based immunity is the rate of spacer acquisition, which is thought to be mainly affected by infection by defective phages, Cas gene expression, and priming. Phages can evade CRISPR-Cas immunity in a variety of ways, for example by altering the protospacer sequence through mutations, deletions and gene rearrangements. Furthermore, physical shielding of the genomic DNA of the phage or dedicated anti-CRISPR systems have also been described. Overall, the research in the many works that are reviewed suggests that in natural environments, phages and bacteria with CRISPR-Cas immune systems can coexist and coevolve.

Chapter 3 reviews CRISPR-Cas type III activated, downstream effector proteins that have the appropriate sensory domains for cyclic oligoadenylate (cOA) molecules: CARF and SAVED domains. The CARF and SAVED effector proteins have triggered renewed interest in type III CRISPR-Cas due to the extreme diversity of their mode of action. CARF nucleases have been studied the most, both in terms of their biochemical and structural characteristics. Furthermore, two examples of characterized SAVED effector proteins are discussed, with NADase and protease activity upon cOA binding respectively. Lastly, a NucC effector with no typical CARF or SAVED domain is described that has DNase activity upon cOA-induced multimerization. Of all described effector proteins, biochemical findings as well as phenotypical outcomes are summarized in this chapter. Lastly, a few examples of bioinformatically predicted CARF and SAVED proteins are proposed to have interesting catalytic activities, which are speculated to contribute to type III immunity. In conclusion, while not all have been experimentally characterized, many of the CARF and SAVED proteins seem to provoke abortive infection, aimed to kill the host and provide population-wide immunity.

Chapter 4 describes the development of CARFish, a new technique that aims to help in unravelling the complete CRISPR-Cas type III cOA signalling network. As described in **chapter 3**, CARF and SAVED containing proteins can be predicted bioinformatically with relative ease. However, there might be cOA-responsive effector proteins that are not readily identified by current in silico prediction methods, which prompted us to develop CARFish. A method was developed to create immobilized cA_4 signalling molecules on agarose beads, which can subsequently be used to "fish" for cOA binders from wild-type cell lysates. Enriched binders can then be identified using LC-MS/MS analysis. To validate the technology, lysates of two *Thermus thermophilus* strains were used, with known and characterized cA_4-responsive effector proteins. Furthermore, an archaeon (*Sulfolobus islandicus*) was also used for the same validation purposes. For the most part, the characterized cA_4-responsive effector proteins were greatly enriched and easily identified with CARFish. This prompted us to try the method on *Sulfolobus tokodaii*, another archaeon, that has the uncharacterized but predicted cOA-responsive effector protein CARF-HEPN. With CARFish, the most enriched protein was this CARF-HEPN protein, confirming that it is indeed a cOA-responsive effector with high affinity for cA_4 specifically. Furthermore, several candidate proteins that were similarly enriched in each of the hosts were further analysed in silico, which are interesting leads to investigate biological relevance of their apparent affinity for cA_4.

Chapter 5 describes research into a novel nanopore-based technique for the determination of the type of cOA and stoichiometries. CRISPR-Cas type III systems from different organisms can produce different cOA signalling molecules when activated, where variation can be found in the amount of AMP moieties in the ring structures. Current methods to gain insight in the cOA-type these type III systems produce in vitro has only been possible in bulk and with sophisticated equipment. In this study, the possibility of label-free detection of single cOA molecules using protein nanopore assays is explored. First, the sensitivity and signal generation profiles are established using synthetic cOA mixtures. Using this data, a convolutional neural network was successfully trained and validated with a series of experiments on mono- and polydisperse synthetic cOA mixtures. Ultimately, the neural network-enabled nanopore technology was used to successfully determine the composition of the cOAs produced *in vitro* by the CRISPR-Cas type III-A and III-B systems from *Thermus thermophilus* without the need for further sample preparation. In agreement with literature, the cOA mixture from this bacterium was determined to consist of mainly cA_4.

Chapter 6 reviews several recent works on CRISPR-Cas type III-E systems and their ancillary proteins. Type III systems are Class 1 CRISPR-Cas systems with protein complexes that typically consist of 5-6 Cas protein subunits that together with

the crRNA guide make up the active RNP effector. Type III-E is unique in that it consists of a single protein that appears to be a fusion of several Cas7 and Cas11 subunits. Furthermore, the characteristic Cas10 subunit of type III-A,-B,-C,-D complexes is missing in type III-E, where it is replaced by TPR-CHAT, a caspase-like protease. Like the other type III systems, type III-E uses an RNA guide to bind a complementary target RNA that will be cleaved at the regular intervals. Self/non-self recognition is also similarly facilitated by a 5' crRNA handle complementarity mechanism. Upon base pairing with its cognate target RNA, the TPR-CHAT subunit is allosterically activated, after which it specifically cleaves Csx30, an ancillary protein that after cleavage forms a complex with Csx31 and RpoE (resembles a sigma factor of RNA polymerase). It was shown that when challenged with phages, this system is able to launch a strong abortive infection phenotype. The exact mechanism of killing has not been determined, but it has been proposed that transcriptional regulation of (an) unknown gene(s) plays an important role. Lastly, with the knowledge gained about this system, the reviewed works describe a method for sensing RNA transcripts in vivo and for a new in vitro RNA-detection diagnostic tool.

Chapter 7 describes the investigation of a CRISPR-Cas type III-B system from *Haliangium ochraceum*, including a set of ancillary genes. Through phylogenetical analyses, it was determined that that the ancillary genes were likely co-opted from another CBASS (cyclic oligonucleotide-based antiphage signalling system) defence system. SAVED-CHAT, containing a cOA sensory domain (SAVED) and a caspase-like protease domain (CHAT), was characterized in vitro and shown to multimerize in an anti-parallel filament upon addition of cA_3 signalling molecules. Cryo-EM structural analysis shows a mode of activation of the CHAT domains upon multimerization. It was shown *in vitro* that the activated SAVED-CHAT caspase-like protease specifically cleaves PCaspase, yet another caspase-like protease, which in turn gets activated. Activated PCaspase is much less specific, as it cleaves a multitude of proteins and was shown *in vitro* to be inhibited by ancillary protein PCi. Additionally, a fluorescent peptide probe was used to be able to track PCaspase cleavage activity in real-time. *In vivo* experiments show a strong abortive infection phenotype when target RNA is presented to the type III-B system in the presence of SAVED-CHAT and PCaspase. Importantly, the inhibition of PCaspase by PCi could counteract this process, indicating some control function of the signal transduction pathway. Altogether this appears to be an effective strategy to defend a bacterial community against mobile genetic elements, probably by inducing a state of dormancy or even suicide of the infected host cell. Moreover, our findings show a cascade of proteolytic activities, with conceptual similarities to eukaryotic caspase pairs and their role in stress-dependent apoptosis.

Chapter 8 describes the investigation into the targeting requirements of CRISPR-Cas type III-B from *Thermus thermophilus* (TtCmr). The endogenous composition of sizes of TtCmr RNP complexes is heterogenous, determined by the length of crRNA and number of Cas5 and Cas7 backbone subunits. First, it was shown that the different sizes of complexes cleave the target RNA at different number of places, determined by the number of Cas7 backbone subunits. Next, the 5' crRNA handle mediated self/non-self mechanism was investigated and it was shown that complementarity did not affect target RNA degradation, but negatively affected cOA production. Subsequently, similar analyses were performed on mismatches in the 5' spacer-derived part of the crRNA, which we termed the Cas10-Activating Region (CAR). Here, no effect on target RNA degradation was observed but, depending on the exact position of the mismatch, a strong negative effect was observed on the production of cOA. Lastly, the seed region, where target RNA base pairing is initiated, was determined for differently-sized TtCmr complexes and determined both by target RNA cleavage and cOA production. This showed a flexible seed region that is dependent on the size of complex, which was further explained by previously obtained structural data and stems from an exposed region of the crRNA near the 3'end. These insights into a flexible seed region could support a model wherein escape by mutation of protospacer regions on the phages is made more difficult, since a variety of seed region requirements are generated with a single crRNA. Finally, insights into the stringent activation of cOA production has led to the development of SCOPE, a nucleic acid detection tool. SCOPE was used in a proof of concept that showed the ability to detect SARS-CoV-2 in human nasal swab samples with high speed, accuracy and sensitivity.

Chapter 9 describes the characterization of a thermostable DNA dependent RNA polymerase (RNAPhi), originating from a bacteriophage. In silico structural analysis using AlphaFold2 did not provide many homologues when searching the databases. However, the NTP and metal cation coordination domain did show predicted similarities with the BH subdomain of the QDE-1 RNA polymerase and allowed for the design of a catalytically dead version (dPhiFa-44). After heterologous expression in and purification from *E. coli*, biochemical *in vitro* experiments showed that incubation of RNAPhi with a dsDNA template and rNTPs leads to the generation of a RNA transcripts in a promoter independent manner. Furthermore, optimal in vitro reaction conditions were determined to be ~40°C, at a pH of 6.5-7.5, with a preference for Mn_2+ as a stimulating metal co-factor. However, the RNA synthesis activity of PhiFa-44 was observed in a broad range of reaction conditions. Lastly, PhiFa-44 was used in the development of SCOPEv2, going from a two-step incubation protocol (SCOPE) to a one-step protocol. This greatly increases the potential for applicability and was demonstrated as a proof-of-concept SCOPEv2 assay for the detection of Pseudomonas aeruginosa. This assay showed a sensitivity of 10 copies/µL of dsDNA

template material with a time to detection of only 31 minutes. Together, SCOPEv2 is an important improvement for the usability and applicability of the SCOPE technology.

General discussion

The increasing interest in type III CRISPR-Cas immune systems have sparked many new avenues of research and the development of tools in recent years. In this thesis, various aspects of type III CRISPR-Cas have been investigated, new systems have been characterised, novel methods have been developed and applications have been developed.

Diversity of cOA-responsive type III CRISPR-Cas effectors

The discovery of cOA signal molecule production by type III CRISPR-Cas systems has led to the rapidly growing understanding of the protein effectors that are activated by them [73,83]. As outlined in chapter 3, we currently know of CARF and SAVED domains as being common sensory domains of these effector proteins, and can be bioinformatically identified as such [84,265]. First studied and characterized are the Csm6/Csx1 family of CARF ribonuclease effector proteins [73,83,204,208,211]. Upon binding to cOA, dimerization is initiated or stabilized in such a way that the fused catalytic domain, HEPN, get activated due to a conformational change that creates an active catalytic site [75]. Rapid and indiscriminate ribonuclease activity is subsequently unleashed and will wreak havoc in the host cell, leading to cell dormancy or death of the infected cell. As such, it will thwart the proliferation of invading bacteriophages and contribute to protect the community. The mechanism of dimerization by using cOA as a "molecular" glue turns out to be a common approach for type III effectors [75]. Alternatively, the Csx1 effector protein in some hosts takes on a rather large hexameric configuration upon binding of cOA [214]. One could argue that this could be beneficial, as it will lead to localized nodes of active effector proteins instead of more distributed dimers. Furthermore, filament formation for some SAVED effector proteins has also been described, both in this thesis (**chapter 7**) and in the literature, where cOA triggers many protein subunits to homo-multimerize [87,93]. The biological relevance of this variety of mechanism of activation and formation of large complexes, if there is any, remains elusive.

Since the characterization of the Csm6/Csx1 CARF ribonuclease effector proteins, a plethora of others have been identified, including ones that have a SAVED cOA-sensing domain. An example of the latter is investigated and discussed in **chapter 7**. Also found in other CBASS systems, it appears that type III-associated proteins with a SAVED domain could have been co-opted from one or more innate immune systems [84,265]. An equally strong connection to other CBASS systems has not been found for CARF domain containing effector proteins, which is odd since it is thought of as a divergent version of SAVED. Evolutionary origins of these defence systems, as discussed in **chapter 2**, can be complex and difficult to prove with a high degree

of certainty. However, it could be rationalized that the ability of Cas10 to produce signalling molecules could only have evolved and selected for if some cOA responsive effector was already present and provided some kind of fitness advantage to the host species. Furthermore, this aspect also prompts the question why there has not been any diversity found in the type of nucleotide (i.e. UTP, GTP, CTP) that is used to create the signalling molecules. Biological relevance for the preference for ATP has not yet been investigated thoroughly. However, an exciting alternative approach by nature has been found, where Cas10 of a type III system produces a SAM-AMP signalling molecule and subsequently leads to the multimerization of membrane pore-forming effector proteins 85. Future research is expected to result in more variations on these themes.

Chimeric cOA-responsive effector proteins

The extreme diversity of catalytic domains fused to the cOA responsive domains is staggering and incredibly interesting, almost as if a deck of cards containing catalytic domains is shuffled and drawn at random and combined with a sensory domain. A commonality is conversion and/or depletion of an essential biomolecule (e.g. NAD, NADP, RNA, DNA, protein) [75]. Rather than some signal transduction or subtle allosteric confirmational change that occurs when cOA is bound, often the cOA just acts as a glue that brings two or more inactive subunits together. This can lead to the formation of composite active sites or allosteric activation of active sites which induces the, generally disruptive, catalytic activity of the effector protein.

Based on this, it would be an interesting endeavour to create synthetic chimera of cOA responsive proteins, for example by replacing the natively fused catalytic sites with different ones. Transplanting natural effector proteins could be an approach, but substitution by completely novel (to type III CRISPR-Cas) catalytic domains would be even more interesting to demonstrate the supposed endless modularity. For example, established split versions of enzymes that are already used in life-sciences can be utilized, including split-luciferase, split-horse-radish peroxidase, split-TEV protease, split-T7 polymerase and even split-Cas9 [403–409]. Besides the interesting synthetic biology aspect of this, there could be some practical uses for these chimeric, cOA-inducible CARF/SAVED effectors.

For in vitro applications like type III CRISPR-Cas nucleic acid detection systems, an alternative read-out to fluorescence measurement can be realized. In short, the detection of target nucleic acids by the type III CRISPR-Cas complex leads to the production of cOA, which in turn activates the chimeric CARF effector protein. In the case of a split luciferase approach, visible light will be generated when both parts are brought together by the cOA-bound CARF domains and in the presence of

luciferin substrate. This signal could be picked up by a relatively simple and cheap sensor, when compared to detection of fluorescence [410]. Horse-radish peroxidase can similarly be used to generate an intense colorimetric signal which can easily be seen by the naked eye, either in solution or on a lateral-flow strip [411–413]. For similar purposes, a cOA-activatable TEV protease could be used for a variety of read-outs. The cleavage of a single protein, or the separation of two proteins can be used for the de-quenching of fluorescent reporter peptides, activation of inhibited enzymes and much more [414]. Lastly, split T7 polymerase activation by cOA could be used to initiate the transcription of a dsDNA template containing a specific T7 promoter, provided in the same reaction mixture [408]. This template can encode an RNA transcript that is targeted by type III CRISPR-Cas complexes in the same mix, leading to the rapid exponential production of cOA via a positive feedback mechanism. If a signal-creating CARF/SAVED effector protein (like described above) is added as well, the sensitivity of the whole system can dramatically be increased.

In addition to this, *in vivo* applications also come to mind, such as transcript monitoring with similar chimeric cOA-responsive CARF/SAVED effectors. In this application, a measurable signal would be created when a certain transcript (RNA targeted by the type III CRISPR-Cas system) is present. Genetic circuits can also be constructed with similar approaches. For example, the transcript-specific activation of a TEV protease in vivo can be used to function as a protein-based cellular regulator. Along the same lines, the presence of a specific transcript can be used to initiate transcription of any gene of choice when using the cOA-responsive T7 polymerase, a novel method of type III CRISPR-Cas controlled gene regulation.

However, development of both *in vitro* and *in vivo* approaches with chimeric CARF/SAVED effectors needs to take into account the spontaneous dimerization that tends to occur even in the absence of cOA. This effect should be studied further and can potentially be mitigated by tuning the concentration of effector and/or by normalizing for the background level of catalytic activity. Next to this, protein engineering strategies, helped with for example laboratory evolution, could be used as well.

In summary, there is much more to explore, learn and develop from the native diversity of cOA responsive effector proteins and creative ways for hosts to kill themselves. An exciting outlook is to utilize these learnings to create synthetic chimeric effector proteins for a variety of applications.

New tools for the investigation of the type III CRISPR-Cas signalling network

A great deal of work has been described in the literature that reflects our increased understanding of the signalling network of type III CRISPR-Cas, and its ancillary effector proteins. As discussed in **chapter 3**, CARF and SAVED domains can be predicted bioinformatically with relative ease but are not always encoded close to the type III CRISPR-Cas operon [84,265]. Furthermore, it is not unthinkable that the cOA signalling molecules have interactions with other host proteins or ancillary effector proteins that have a still to-be-characterized sensory domain. This open-ended question sparked the development of a new approach to investigate the cOA interactome, CARFish, as described in **chapter 4**. By using beads containing immobilized cA_4 messenger molecules, we were able to perform protein pull-down assays on wild-type lysates to capture (novel) cOA-interacting proteins. The proof-of-concept study showed that we were able to capture and identify known cA_4-binding effector proteins from wild-type lysates of several hosts, and to confirm cOA affinity for a predicted, but not experimentally characterized CARF effector protein. In these experiments, several proteins were identified that were highly enriched but had no obvious association to the type III CRISPR-Cas defence system. In silico analysis of these findings did not result in strong leads regarding their biological relevance in an immune response. However, a trend was identified in the types of binding domains that these proteins generally contain, small molecules with similar characteristics as cOA; i.e. adenosine-like molecules. Whether these are false positives, an epiphenomenon, is still to be determined in follow-up studies. Further, optimization of the washing and elution steps could be improved to promote the binding stringency and to more strongly exclude false positive cOA-binding proteins. For example, in the current method, the elution is done by addition of biotin, releasing the bound cA_4 molecule, including the bound proteins. An alternative approach to this is elution with a high concentration of cA4, leading to the release of cA_4-specific binders. Potentially, this could exclude proteins that are bound to the beads via (partially) degraded biotin-cA_4 molecules, thereby reducing false positive hits.

However, a prerequisite for the developed technology is a sequenced and, preferably, annotated genome to allow for the LC-MS/MS analysis. If using this methodology on mixtures of lysates, for example from a sample obtained in nature, metagenomics may still provide enough fidelity to achieve interesting hits. Currently, the method has only been performed using cA4, limiting the ability to identify the broad effector protein variety that has been described, in terms of cOA molecule ring-sizes. For example, preparing binding beads on which $cA_{2\text{-}6}$ are immobilized, would allow one to be agnostic about the type of cOA you know/think the type III CRISPR-Cas system in the host of interest makes. Doing this is relatively easy if a collaborating group or specialized company is able to (chemically) make more varieties of the biotin-con-

jugated cOA molecules since the overall methodology can remain identical. On top of this, attempting a similar strategy with messenger molecules of other defence systems would be a very promising outlook as well.

A first step with promising results in the development of this methodology has been made, with plenty of improvements to be considered in follow-up studies. The top-down approach proposed in chapter 4 can work alongside common biochemical and *in vitro* techniques to get more insight into mechanisms by which type III CRISPR-Cas works *in vivo*.

Identification of the cOA production repertoire of type III CRISPR-Cas

Apart from the effector proteins that bind cOA messenger molecules, it would be interesting to investigate different type III CRISPR-Cas complexes that produce cOA variants with different AMP-ring sizes. The parameters that need to be met before cOA production is initiated is explored in chapter 8, the development of a methodology for determining the cOA repertoire in vitro is elaborated on in **chapter 5**.

Current methods for determining the sizes of cOA molecules that are being produced include HPLC-MS, gel analysis using radioactive labelling and thin layer chromatography to name a few [73,83]. While powerful, they can be cumbersome and do not deliver single molecule fidelity. At this moment, in silico analysis of different Cas10 subunits to determine the corresponding cOA ring-size is not straightforward. Therefore, we aimed to utilize nanopore technology, which is typically used for inexpensive and long-read sequencing of DNA. Generally speaking, a protein pore is embedded in a free-standing lipid bilayer and acts as a sensor for single molecules. An applied voltage drives negatively charged molecules through the pore, which while translocating the pore, partially blocks the ionic through-pore current [296–298]. The resulting current blockade signal can be characteristic for the target molecule of interest. The Oxford Nanopore Technologies platform has successfully been developed to allow sequence analysis of DNA as well as RNA [297,298].

Using synthetic cOAs, and stoichiometric mixtures thereof, we were able to train a neural network on the signals obtained from the nanopore device. After training, we could use the platform to correctly identify the stoichiometries of mixtures of synthetic cOAs. Although we could specifically identify cA_4-cA_6, the current platform does not yet allow for discriminating cA_3 and cA_4. The workflow was used to identify the stoichiometry of cOAs that are produced in vitro by purified *T. thermophilus* Csm and Cmr type III CRISPR-Cas complexes, without the need for any preparation of the sample. Although an accurate direct quantitative comparisons between Csm and Cmr cannot be made at this time since not all parameters were equal, a rough

quantification of the amount of cOA molecules suggests that Cmr produces more cOA molecules per type III complex than Csm in the conditions tested. The more interesting underlying question was whether any differences between the cOA production repertoire of Csm and Cmr from *T. thermophilus* could be found, providing a possible explanation for their co-occurrence in the same host. Hence, future optimization of the approach is required to allow for discriminating cA_3 and cA_4.

The developed method is already a great tool for determining what type of cOA molecules are produced by type III CRISPR-Cas complexes, *in vitro*. A worthwhile follow-up would be trying to develop a similar workflow that could work on lysates, with minimal sample preparation steps. This will omit the need for purification of the type III complexes, either pull-downs or recombinant expression, and characterization of the crRNA population for in vitro activation with target RNA. In this way, for example after challenging a host of interest with phage to activate the type III immune response, lysates could be analysed for the cOA molecules that are produced. This could already provide a piece of the puzzle in the further exploration of the complete signalling network including characterization of potential effector proteins. Moreover, this could direct CARFish efforts (**chapter 4**) towards using a specific cOA in the top-down characterization efforts.

Non cOA-mediated immune responses by type III CRISPR-Cas

The multifaceted, Swiss army knife-like capabilities of type III CRISPR-Cas have recently been expanded by the discovery and characterization of type III-E systems. In **chapter 6** of this thesis, we reviewed some recent studies on this system.

Besides the sequence-specific RNA targeting that is typically found in Class 1 type III CRISPR-Cas systems, ~36% of Cas10 subunit proteins contain HD-nuclease domains and ~85% of them contain conserved polymerase active-site motifs [415] responsible for sequence non-specific ssDNA cleavage and cOA production respectively. Type III-E systems depart from this general architecture of type III-A,-B,-C,-D systems in several ways. First, instead of consisting of multiple individual subunits (up to 7), a large polypeptide functions as core of the effector complex. Another unique feature of type III-E systems is that no Cas10 homologue is present. Instead, the major core forms a complex with two additional proteins: Csx29 and TRP-CHAT. The latter protein contains a caspase-like protease domain. In analogy to the activation of Cas10 in type III-A,-B,-C,-D systems, it has been shown that target RNA binding by type III-E systems, when not-self, leads to the activation of the CHAT protease domain [319,323]. Together with several ancillary proteins, RpoE, Csx30 and Csx31 it can mount an effective immune response when challenged with phages [317]. Analysis has shown that RpoE, Csx30 and Csx31 can from a stable complex, of which Csx30

is cleaved by the active TPR-CHAT subunit . While some details remain unclear, cleavage of Csx30 relieves repression of RpoE, leading to transcriptional activation of several genes of the host [322]. In analogy to the immune strategy of other type III systems, it is tempting to speculate that the latter induced genes may somehow lead to a dormant state, or even to suicide.

The type III-E systems investigated so far are a very interesting example of a Class 1 system that has characteristics of typical Class 2 systems. Furthermore, the absence of Cas10 and cOA signalling shows an novel approach by this type III CRISPR-Cas. The involvement of a caspase-like CHAT protease domain and several ancillary proteins to achieve robust immunity seems less efficient. Why an aggressive caspase-like protease would not be utilized by these type III-E systems for direct killing of the host is not clear. However, one can imagine that there are more varieties on this theme, TPR-fusion proteins that have varied catalytic domains instead of CHAT. Perhaps similar to the extreme diversity seen in CARF/SAVED effector proteins (**chapter 3**).

A cOA-independent immune response strategy is an interesting trend in type III CRISPR-Cas research. Recent work showed that some type III systems contain Cas10 fusion proteins, where for example NucC is fused to it [415]. NucC is a known cOA-responsive effector protein, which in the discovered system is thought to be activated by the cOA produced by the Cas10 fusion protein. This cOA would mediate dimerization with a separately encoded NucC flanking the operon. This is most likely still cOA mediated but other Cas10 fusion proteins might be cOA-independent. For example, the same study found a Cas10-AbiEii fusion protein, AbiEii is known for being part of a toxin/anti-toxin system that performs its function by polymerization of uncharged tRNAs [415–418]. It is feasible that the activation of the AbiEii part of the fusion protein is not cOA mediated, an exciting outlook for future work. Moreover, there is the possibility of a diverse set of Cas10 fusion proteins to be found in nature and this could be the starting point for a whole range of interesting studies in the future.

Promiscuous proteolytic activity as main instigator of abortive infection

Only recently, studies have emerged that describe type III CRISPR-Cas immune mechanisms do not solely target nucleic acids with their downstream immune response mechanisms. Meaning that also proteins and other essential molecules are targeted through the activation of CARF/SAVED effector proteins (**chapter 3**) .

Work on a type III CRISPR-Cas system and associated effector proteins from *Haliangium ochraceum* is described in **chapter 7**. Next to a typical type III-B system,

several ancillary effector proteins can be found in the operon. This system attracted our attention because of the apparent lack of nuclease effector proteins. Instead, the canonical type III genes clustered with two genes that encode potential proteases: a caspase-like protein with a SAVED domain (SAVED-CHAT) and a second caspase-like protein (prokaryotic caspase, PCaspase). Phylogenetic analyses showed a high likelihood that both effector proteins were co-opted from another CBASS immune system to use the Cas10-produced cOA messenger molecules. In fact, we demonstrated that SAVED-CHAT oligomerizes upon addition of cA_3, forming anti-parallel double filament structures. This leads to the activation of the CHAT caspase-like domains. It turns out that the cA_3-activated SAVED-CHAT, in turn, specifically cleaves and activates PCaspase, unleashing non-specific protease activity on a variety of protein substrates. The exact cleavage site where PCaspase is cleaved by activated SAVED-CHAT was determined to be the R153 residue, as shown by mass spectrometry of the cleavage product observed in the in vitro experiments (**chapter 7**). Activated SAVED-CHAT in combination with PCaspase (when heterologous expressed in *E. coli*) lead to a very strong abortive infection phenotype, as measured by the >6 orders of magnitude decrease in transformation efficiency. However, it was demonstrated that at least two of the other associated genes could also be involved in the immune response or regulation thereof. PCi (PCaspase inhibitor) showed ability to inhibit the activity of activated PCaspase, both *in vitro* and *in vivo*. This hints to an regulatory function of PCi on the overall system, potentially reducing low levels of unintentional activation, or reverting the initiated process towards a dormant state, in case of successful neutralization of the viral infection. Moreover, PC-σ (Prokaryotic Caspase-controlled sigma factor) is also cleaved *in vitro* by activated PCaspase. No biological relevance of this finding has yet been demonstrated in the current study but undoubtedly will be investigated in future studies; in analogy with the aforementioned type III-E system, the activity of the potential transcription factor may somehow contribute to cell death

Overall, the type III system in this study is a very interesting example of an immune response that is largely mediated via proteolytic activities and novel is this regard. Furthermore, striking conceptual similarities to eukaryotic caspase apoptosis pathways can be found in the way that a central protease node (SAVED-CHAT) leads to the activation of an executioner protease (PCaspase) that does the eventual killing [419]. Next to this, the utilization of a CHAT protease domain by type III to launch an effective immune response has also been seen in type III-E systems (**chapter 6**) [317,319–323]. In these systems, the proteolytic degradation does not directly lead to killing of the host, whereas in the type III system in *Haliangium ochraceum* this does seem to be the case.

Although the initial investigation into this rather complex system revealed several unique features, several aspects still require further investigation. The formation of double filaments of SAVED-CHAT after addition of cA_3 seems rather odd. It is unclear if there is any biological relevance to the formation of these doublet filaments, as compared to for example the formation of "simple" homodimers instead. This filament-forming behaviour has also been seen in a TIR-SAVED effector protein that was recently described in the literature (chapter 3). We cannot rule out that this filament formation is a phenomenon that only occurs *in vitro*, in condition with most likely higher concentration of SAVED-CHAT than in *in vivo* conditions. Future studies could focus on elucidating this with (fluorescently) labelling or tagging the SAVED-CHAT proteins *in vivo* and monitoring filament formation upon activation of the type III CRISPR-Cas system. We did not observe a clear phenotype in the absence of PCaspase, so a PCaspase knockout of *Haliangium ochraceum* should be created for efficient monitoring of this filament formation *in vivo*. Still, strategies used must employ very stringent control over the activation cOA production to avoid killing or burdening the cells unintendedly. This could be achieved by, for example, using a SiBR-Cas based control system for the transcription of target RNA [420]. On top of this, a catalytically dead variant of SAVED-CHAT could also be made to mitigate the cellular burden even more, this catalytically dead version was already demonstrated in **chapter 7**. Obviously, this requires the development of genetic tools for this strain of *Haliangium ochraceum*.

The observed abortive infection phenotype is very strong *in vivo* in *E. coli*. Demonstrating this in the native host is something that can expand our understanding of the immune response further. It would be interesting to know if activated PCaspase either specifically affects certain metabolic pathways or other key processes, or rather targets proteins non-specifically. Complete proteomics on *Haliangium ochraceum* lysate that is cleaved by activated PCaspase might provide information to answer this question. A more elegant approach is to take lysates of, for example, phage challenged cells (in the presence or absence of a phage-targeting spacer) and analysing its proteome. This could be repeated and compared to in a similar setup with cells encoding a catalytically inactive version of the type III immune system. Alternatively, a tightly regulated *in vivo* approach, similar to the one described for investigating filament formation, could be employed to remove experimental variables that using phages would introduce.

Next to this, the function of PC-σ in the immune response, if any, is completely unknown at this moment. While annotated as a sigma factor, based on bioinformatical predictions, any indication of its involvement in the regulation of genes is currently lacking. Transcriptomics on cells in which the complete type III CRISPR-Cas system of **Haliangium ochraceum** is activated could shed light on this question. At

present, it is unclear whether PC-σ forms a complex with PCi and/or PCk (Prokaryotic Caspase-controlled kinase) to perform its function. An affinity-tagged version of PC-σ could be used for a pull-down experiment to determine complex formation in the native host *in vivo*. Alternatively, SEC-MALS studies on the complex formation *in vitro* may also lead to valuable insights. It is likely that the sigma factor interacts with the host's transcriptional machinery, which may also be an outcome arising from the pull-down experiments. Shortly mentioned, the role of PCk remains completely unclear since all efforts to produce this protein *in vitro* or introduce it *in vivo* have not been successful. It is exciting to postulate that this protein, when activated by an unknown mechanism, starts phosphorylating essential enzymes to modulate their activity, for example. An approach could be to try and produce this protein using TXTL technology, mitigating supposed toxic effects that are observed when trying to recombinantly produce it in *E. coli*.

Lastly, the very strong abortive infection phenotype observed with just SAVED-CHAT and PCaspase begs the question why PCi, PC-σ and PCk are still required. The role of PCi could be understood as a means of regulation of the system to prevent unintended activation. Whether potential gene regulation by PC-σ aids in the abortive infection aspect of the immune response remains a mystery. As hypothesized for RpoE in the type III-E systems (**chapter 6**), genes that are involved in adaptation could be upregulated to increase spacer generation and incorporation in the CRISPR array [322]. As for PCk, experimentally demonstrating the exact mechanism by which it functions *in vivo* may be rather difficult. Powerful mass spectrometry techniques need to be employed to determine phosphorylation sites that could potentially be on a myriad of host cellular proteins. All of this does not yet consider the fact that in *Haliangium ochraceum* there is a complete second type III operon with almost identical ancillary effector proteins. Assuming both operons are expressed simultaneously and based on high sequence similarities of both gene clusters, effector proteins from the one operon can affect effector proteins from the other operon and vice versa. However, this needs to be experimentally demonstrated. In relation to this, the crRNA content of each type III complex should be investigated and sequenced to determine which crRNAs end up in which complex, as it is difficult to think of a reason why both (metabolically expensive) immune systems should co-occur in the same host. This is especially of interest since Haliangium ochraceum contains an abnormally large amount of spacers. This co-occurrence is seen more often for type III CRISPR-Cas systems and nobody has yet to come forward with a compelling theory as to why.

In summary, a great deal fundamental aspects about the type III CRISPR-Cas system discussed in **chapter 7** have been uncovered, while a variety of open questions remain to be tackled in future exciting studies.

Flexible seed region of type III CRISPR-Cas

In literature it has been demonstrated that type III CRISPR-Cas is relatively flexible in its targeting requirements in vivo, generously allowing mismatches while still being effective in the immune response it will mount [155,196]. This may seem counter-intuitive, why be so "sloppy" when the ultimate outcome could be death of the host cell. By studying this phenomenon in more detail *in vitro*, as described in **chapter 8**, this conundrum is at least partially explained.

It has been known for a relatively long time that the endogenous population of type III CRISPR-Cas complexes are heterogenous in their stoichiometry, with differing lengths of crRNAs determining the length of the complex. The study in this thesis took this concept and replicated this in a controlled manner, by reconstituting type III-B CRISPR-Cas complexes from *T. thermophilus in vitro*, using crRNAs of a set length. By employing a variety of (biochemical) techniques, we investigated the presence of a seed region. This region is where base paring of the target RNA to the crRNA is initiated, forming the first checkpoint before further base pairing propagation can occur. A seed region was found at the 3' region of the crRNA for the complex size variants tested. However, when using smaller complexes, e.g. 40nt crRNA instead of 46nt crRNA, the seed region shifts towards the 5' end of the crRNA (corresponding to the truncated 3' end). This not only has implications for sequence specific degradation of the target RNA, the broader and more extreme cOA-mediated immune response can also not be initiated if full base pairing can't be started at the seed.

In the context of the ongoing arms' race with bacteriophages (**chapter 1, 2**), the phage may escape immunity by mutating the seed region and thereby preventing the initiation of the immune response. However, one of the hypotheses coming from **chapter 8** is that by having a heterogenous endogenous population of type III CRISPR-Cas complexes, a variety of seed regions exist for each spacer in the CRISPR array. This may make it much harder for the phage to escape immunity by mutating this region, a considerably larger region needs to be deleted or mutated. In conclusion, the flexible seed region observed for the type III CRISPR-Cas system that was studied might allow for a more robust defence against bacteriophages (**chapter 2**).

Whether this theory is actually true remains to be investigated in a proper *in vivo* follow-up study. Ideally an *in vivo* system is set up which can be challenged with bacteriophages to screen for the frequency of escapee bacteriophages. Subsequent sequencing of these bacteriophages will provide insight into which mutations and/or deletion allow escaping type III targeting. The type III CRISPR-Cas system needs to have a set size, i.e. to limit the flexibility of the seed in the crRNA. Controlling

the process of crRNA maturation and trimming to set lengths might be too hard to achieve *in vivo*. Therefore, an approach could be to shorten the spacer in the CRISPR assay to for example 34nt (including 8nt handle). This way, at least the upper limit of crRNA length is set. One could then challenge the host containing this system and compare it against the native system, scoring the bacteriophage escapee frequency and mechanism. If the hypothesis is true, much more escapees should occur in the system with limited heterogeneity of complex sizes (i.e. with crRNA length of 34nt). Expansion of this work could be to also include, for example, spacers of length 40nt and 46nt and performing the same experiments. This increases the heterogeneity in a stepwise manner, coming closer to the native conditions. It could be expected that the reduction of bacteriophage escapees is not linear as the crRNA lengths increase, but rather exponential. Lastly, is might be interesting to perform some fitness analysis on the bacteriophages of each experiment (i.e. 34nt vs 40nt vs 46nt vs endogenous). In the biological context, it could be that escapees found in these experiments are impeded in such a manner that they would rapidly be outcompeted in the larger bacteriophage population.

In summary, an unexpected flexible seed mechanism of type III CRISPR-Cas systems might aid in being highly efficient in the robustness of the immune response against bacteriophages. It achieves this with a limited amount of spacers. Future studies into this hypothesis are needed to test this and expand our knowledge on the intricacies of type III CRISPR-Cas.

Type III CRISPR-Cas nucleic acid detection tools

Soon after the discovery of collateral cleavage activity of Cas12 and Cas13 systems, the potential for their use in nucleic acid detection applications was noted. This has led to innovative nucleic acid detection technologies like DETECTR and SHERLOCK as early as 2018, further developed by spin-off companies from UC Berkeley and MIT Broad respectively [72,368,372]. These technologies promise quick, sensitive and accurate diagnostics for point-of-care applications. In **chapter 8**, the potential for type III CRISPR-Cas systems in a diagnostic context is investigated.

In the developed SCOPE technology, target RNA recognition by the type III CRISPR-Cas complex leads to the rapid production of cOA molecules. Subsequent activation of a Csx1 CARF effector protein (**chapter 3**) leads to rapid ssRNA cleavage activity, visualized by the presence of a fluorophore-quencher reporter RNA. The sensitivity of this technology was at the time ~1 nM of target RNA, a concentration at which sufficient signal was created. While being very rapid (2-3 minutes), however, the sensitivity is not high enough for most diagnostic applications. To enhance this, a LAMP isothermal pre-amplification step was introduced. One of the primers

has a 5' overhang containing a T7 RNA polymerase promoter sequence, which ends up in the amplicon generated by the LAMP reaction. Subsequent transcription of the amplicon generates RNA that is targeted by the type III CRISPR-Cas complex using a pre-specific crRNA. When targeting a region of the amplicon that is not introduced by the primers used for the LAMP amplification, verification of the amplicon is performed which leads to the generation of a (fluorescent) signal. This increased the sensitivity of the overall assay (scope*Dx*®) to attomolar levels, improving the usability of the technology in cases such as infectious diseases. In most infectious disease user cases, low levels of, for example, viral copies need to be detected to be a useful diagnostic tool. A proof-of-concept study was performed on human swab samples containing the SARS-CoV-2 virus, which showed very good performance when compared to the golden standard, qPCR.

The elevated temperature (65°C) at which the (RT-)LAMP pre-amplification works can help in reducing sample preparation since it helps the release of genetic information from cells or viral particles. In addition, LAMP is inherently not very sensitive to inhibitory effects of compounds when compared to for example qPCR, helping in elevating the amounts of target genetic information for subsequent detection by type III CRISPR-Cas in more crude samples. Lastly, when targeting RNA based viruses for example, secondary structures of this target RNA are less stable and more prone to amplification and detection at the higher temperature like 65°C. Unfortunately, LAMP amplification is prone to false positives when used in non-ideal conditions like crude sample matrices. Furthermore, it can lack specificity in the detection of small genetic differences, which are both mitigated by using type III CRISPR-Cas to verify the generated amplicon and compounding the specificity of both technologies.

While very promising, the scopeDx® technology in **chapter 8**, requires a two-step incubation protocol due to the temperature limitations on the T7 RNA polymerase that was used. No other commercially available thermostable RNA polymerase was available at the time. The maximum temperature of the used T7 RNA polymerase lies around 52°C, significantly lower than the 65°C that is optimal for LAMP and the type III CRISPR-Cas system from *T. thermophilus*. The need for a two-step incubation protocol increases hands-on time and introduces a large risk for cross-contamination when opening the reaction tubes. This bottleneck greatly diminishes the applicability of scope*Dx*® in a real diagnostic setting.

To overcome the two-step incubation protocol and work towards a one-step protocol, the search for a novel thermostable RNA polymerase was started, which is described in *chapter 9*. Due to their, typically, single subunit polymerases, bacteriophages that infect thermophilic bacteria were prime targets for our genome mining efforts. Characterized in literature, efforts focused further on the *T. thermophilus* infecting

bacteriophage PhiFa [362]. Its genome contains two copies of a predicted RNA-dependent RNA polymerase (phiFa-44), as per automatic database annotation. Since the genome of phiFa is dsDNA, efforts continued with this gene in particular, as it was more likely to be a DNA-dependent RNA polymerase. As demonstrated in *chapter 9*, purified PhiFa-44 showed DNA-dependent RNA polymerase activity *in vitro*. Analysis of co-factor dependence was also performed and consistent RNA polymerase activity could be achieved. Surprisingly, this was observed for dsDNA templates in a promoter independent manner. Furthermore, some RNA synthesis activity was also observed when using ssRNA and ssDNA templates (**chapter 9**).

In the biological context, it is unexpected that a bacteriophage polymerase would function in the absence of a specific promoter sequence. It would be way more efficient for the bacteriophage if the polymerase would function in a sequence and strand specific manner, instead of seemingly transcribing every DNA sequence it encounters. The currently observed indiscriminate polymerase activity could be a side-effect of the *in vitro* conditions used. The current study did not investigate this aspect any further, but several follow-up experiments could be performed. Having the phiFa genome would allow for *in vitro* transcription by PhiFa-44, subsequent sequencing of the obtained products might reveal genes that are highly transcribed. Using this information, a consensus promoter sequence might be obtained. Alternatively, during infection by phiFa, certain host genes might be regulated strongly in favor of bacteriophage proliferation. Transcriptome analysis on phiFa infected hosts compared to non-infected hosts might reveal these, which again may lead to a consensus promoter sequence. A CHIP-seq approach by using a pull-down of PhiFa-44 from an infected host may also reveal the same information. Alternatively, PhiFa-44 might require one or more protein co-factors. A pull-down experiment on PhiFa-44 from an infected host and subsequent mass spectrometry analysis would be able to reveal this.

Many questions concerning the mechanistic details of the PhiFa-44 polymerase remain unanswered at this moment. Still, this did not get in the way of attempting to use the PhiFa-44 polymerase in a scope*Dx*® reaction, using an assay designed for the detection of *Pseudomonas aeruginosa* as a proof-of-concept. Remarkably, the concept worked very well, as showcased by a one-step incubation protocol that achieved a sensitivity of 10 copies/µL in 30 minutes, scope*Dx*®v2. This even removed the need for the introduction of a primer in the LAMP pre-amplification that contains a 5' promoter sequence overhang. The development of the one-step incubation protocol greatly improves applicability in practise, and it is very exciting to see where this technology will be used in the future.

Next in the development of the scopeDx™ technology

The current scope*Dx*®v2 technology is very useful for the easy detection of nucleic acids in a wide range of applications. However, further usability could be added to the technology if minute genetic differences (e.g. SNPs) could be detected in a similarly easy fashion.

In **chapter 8**, investigation into the Cas10-Activating-Region (CAR) responsible for cOA production regulation may provide the solution. There it was shown that in vitro, single base pair mismatches at key locations (especially nt 1 and 2) resulted in the almost complete diminishing of cOA production. This quality can be leveraged in the context of SNP detection, if the crRNA is designed in such a way that the mismatch(es) occur at the most sensitive locations in the CAR. The majority of the scope*Dx*®v2 technology can remain the same, only clever crRNA design can be leveraged to achieve single-nucleotide resolution. Additional tuning may be required as a true binary result, signal or no signal, is desired to reduce the need for interpretation of the results. This tuning can be achieved by introducing artificial mismatches in the CAR, potentially increasing the SNP differentiating ability of the system.

Another desirable trait for nucleic acid detection technologies is the ability to multiplex, specific detection of multiple sequences. This is a more difficult hurdle to overcome since signal is created by a Csx1 ribonuclease that is activated by cOA molecules. If a mixture of type III CRISPR-Cas complexes is used that contain different crRNAs, detection of any of them leads to an indistinguishable signal. This would be an "OR-gate", which may still be useful if the diagnostic problem just requires the user to know if any or more of the targeted sequences are present. More ideal would be the ability to create different signals (e.g. different fluorescence per target), something that might be achieved by using type III CRISPR-Cas complexes and CARF proteins from different hosts together in one reaction. For example, one of the type III CRISPR-Cas systems produces cA_3 molecules, while the other produces cA_6 molecules. The extreme diversity of CARF/SAVED effectors (chapter 3) most likely will allow for using effector proteins that specifically get activated by each type of cOA molecule respectively. If cA3 leads to the activation of a ssRNase and cA_6 leads to the activation of a ssDNase, reporter RNA and DNA molecules containing different fluorophores will lead to a different signals, respectively. Next to nuclease based read-outs, other catalytic activities like proteases (**chapter 7**) can also be utilized, or combinations thereof, to achieve the same outcome.

Another approach is using a type III CRISPR-Cas system with a completely different signalling molecule, as described in literature recently [85]. Then, two type III CRISPR-Cas variants can be used in the same reaction that are coupled to different

effector proteins, due to the difference of signalling molecules. A variation on this theme is artificially changing the signalling molecules produced natively. As far as we know, type III CRISPR-Cas has only been found to perform cyclization of ATP to produce cyclic signalling molecules (cA_{2-6}). It would be very interesting, yet rather risky, to try and use laboratory evolution to change this nucleotide specificity to any of the others, UTP, GTP, CTP.

Accordingly, efforts will need to be undertaken to change the signal molecule preference of the effector protein, for example a SAVED-CHAT protease (**chapter 7**). The coordination and initiation of cOA production by the Cas10 subunit is complicated and therefore might impose difficulties to start with. A prerequisite of this approach is the ability to use synthetic variants of, for example, cyclic oligoguanosine (cOG). Synthetic versions of cOAs are commercially available so it is conceivable that cOG can also be chemically synthesized.

With the obtained structural insights of the SAVED-CHAT effector protein, random mutations could be introduced in the regions that are responsible for cA_3 binding (**chapter 7**). The subsequent creation of a gene plasmid library can be used for a Transcription Translation (TXTL)-based screening method. The gene library should be added to TXTL mixture, in such a way that (less than) one copy of plasmid ends up in each aliquot. Next, incubation of the TXTL mixture, in e.g. 384-well format, leads to the production of the mutant SAVED-CHAT effector proteins. Alternatively, if the library size is too large to be feasible with large scale screening of 384-well plates, having multiple variants (plasmids) per aliquot will still help narrow down to working effector proteins. Funnelling, positive aliquots towards a more focused single plasmid per aliquot protocol may be time efficient. Getting a positive readout can be achieved as previously established in **chapter 7**, a fluorescent readout of SAVED-CHAT activity. Addition of the required SAVED-CHAT activity assay components, including cG_3, and subsequent incubation will lead to a fluorescent signal in wells where a cOG-sensitive SAVED-CHAT protein is present. If found, the next step will be confirming this cOG sensitivity and loss of cA_3 sensitivity in vitro.

With a cG_3-sensitive SAVED-CHAT effector, laboratory evolution of the type III Cas10 subunit can begin. A similar approach can be taken with the laboratory evolution of Cas10, the complete screening for the production of cG_3 has been established in the previous step. A prerequisite is that the type III CRISPR-Cas complex forms in the TXTL environment, including crRNA maturation. Assuming this to be true, screening of the Cas10 library would mean a similar setup to the previous step, addition of screening components (cG_3 sensitive SAVED-CHAT, PCaspase, fluorescent peptide) and a target RNA will lead to a fluorescent signal if successful. In this case, that would mean that the type III CRISPR-Cas complex produces cG_3 molecules

upon binding the provided target RNA.

If the TXTL setup is not ideal for the reconstitution of type III CRISPR-Cas complexes and required crRNA maturation, an alternative *in vivo* approach might be taken. For this, an up and running *in vivo* type III CRISPR-Cas system is needed, such as the one demonstrated in **chapter 7**. However, several features will need to be changed to this system. First, Cas10 should be introduced on a separate plasmid, as this would be supplied as the gene library. Secondly, proper activation of type III CRISPR-Cas should not lead to cell death so PCaspase should not be included. Additionally, activation of SAVED-CHAT should lead to the proteolytic degradation of a toxin system, for example. Thereby, activation of type III CRISPR-Cas that leads to the production of cG_3 in turn activates the cG_3 sensitive SAVED-CHAT effector protein. Subsequent repression or degradation of the toxin system leads to survivors that can be screened for upon plating on selective plates. As with all *in vivo* laboratory evolution strategies, the pressure for mutations may lead to unwanted changes in other parts of the system.

With a cG_3-producing type III CRISPR-Cas system, a second read-out can be created when used in combination with a cOA-producing system. While providing learnings on a fundamental level, the above proposed strategies may actually improve key characteristics of the scope*Dx*® technology further.

Concluding words

This thesis covers a relatively broad range of topics, albeit within the scope of type III CRISPR-Cas systems. I am very pleased to have been able to deepen our knowledge on these systems, albeit only by a small amount. On this topic alone there is much more fundamental research to be done to get a good grasp on the ever-surprising type III systems. Furthermore, the potential applications that could be derived from this increased understanding are very exciting as well. The amount of variety of CRISPR-Cas systems is simply astounding, I am very much looking forward to seeing what this field of research will bring in terms of fundamental insights and applications that aim to better the world.

References

1. Darwin, C. , K. L. *On the Origin of Species by Means of Natural Selection, or, The Preservation of Favoured Races in the Struggle for Life*. (1859).

2. Young, R. Y. Bacteriophage lysis: mechanism and regulation. *Microbiological reviews*, **56**(3), 430-481 (1992).

3. Suttle, C. A. Marine viruses - Major players in the global ecosystem. *Nat Rev Microbiol* **5**, 801–812 (2007).

4. Prigent, M., Leroy, M., Confalonieri, F., Dutertre, M. & DuBow, M. S. A diversity of bacteriophage forms and genomes can be isolated from the surface sands of the Sahara Desert. *Extremophiles* **9**, 289–296 (2005).

5. Grose, J. H. & Casjens, S. R. Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family Enterobacteriaceae. *Virology* **468**, 421–443 (2014).

6. Jacobs-Sera, D. *et al.* Genomic diversity of bacteriophages infecting Microbacterium spp. *PLoS One* **15**, (2020).

7. Mohanraju, Prarthana, et al. Alternative functions of CRISPR–Cas systems in the evolutionary arms race. *Nature Reviews Microbiology* **20**(6): 351-364 (2022).

8. Guegler, C. K. & Laub, M. T. Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection. *Mol Cell* **81**, 2361-2373.e9 (2021).

9. Hsueh, B. Y. *et al.* Phage defence by deaminase-mediated depletion of deoxynucleotides in bacteria. *Nat Microbiol* **7**, 1210–1220 (2022).

10. Koga, M., Otsuka, Y., Lemire, S. & Yonesaki, T. Escherichia coli rnlA and rnlB compose a novel toxin-antitoxin system. *Genetics* **187**, 123–130 (2011).

11. Gao, L. A. *et al.* Prokaryotic innate immunity through pattern recognition of conserved viral proteins. *Science* **377**(6607) (2022).

12. Garb, J. *et al.* Multiple phage resistance systems inhibit infection via SIR2-dependent NAD+ depletion. *Nat Microbiol* **7**, 1849–1856 (2022).

13. Zhang, T. *et al.* Direct activation of a bacterial innate immune system by a viral capsid protein. *Nature* **612**, 132–140 (2022).

14. Goldfarb, T. *et al.* BREX is a novel phage resistance system widespread in microbial genomes . *EMBO J* **34**, 169–183 (2015).

15. Tock, M. R. & Dryden, D. T. F. The biology of restriction and anti-restriction. *Current Opinion in Microbiology* vol. **8** 466–472 (2005).

16. Gordeeva, J. *et al.* BREX system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Res* **47**, 253–265 (2019).

17. Wang, L., Jiang, S., Deng, Z., Dedon, P. C. & Chen, S. DNA phosphorothioate modification - A new multi-functional epigenetic system in bacteria. *FEMS Microbiology Reviews* vol. **43** 109–122 (2019).

18. Wang, S. *et al.* SSPABCD-SSPFGH constitutes a new type of DNA phosphorothioate-based bacterial defense system. *mBio* **12**, (2021).

19. Xiong, L. *et al.* A new type of DNA phosphorothioation-based antiviral system in archaea. *Nat Commun* **10**, (2019).

20. Xiong, X. *et al.* SspABCD–SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities. *Nat Microbiol* **5**, 917–928 (2020).

21. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* vol. **18** 67–83 (2020).

22. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. *Nucleotide Sequence of the Iap Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in Escherichia Coli, and Identification of the Gene Product*. *JOURNAL OF BACTERIOLOGY* vol. **169** (1987).

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

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

25. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (N Y)* **151**, 2551–2561 (2005).

26. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174–182 (2005).

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

28. Rodolphe Barrangou *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **31**(5819), 1709–1712 (2007).

29. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**(5891), 960–964 (2008).

30. van der Oost, J. & Patinios, C. The genome editing revolution. *Trends in Biotechnology* vol. **41** 396–409 (2023).

31. van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M. & Brouns, S. J. J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends in Biochemical Sciences* vol. **34** 401–407 (2009).

32. Nuñez, J. K. *et al.* Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* **21**, 528–534 (2014).

33. Wiedenheft, B. *et al.* Structural Basis for DNase Activity of a Conserved Protein Implicated in CRISPR-Mediated Genome Defense. *Structure* **17**, 904–912 (2009).

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

35. Kieper, S. N. *et al.* Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Rep* **22**, 3377–3384 (2018).

36. Lee, H., Zhou, Y., Taylor, D. W. & Sashital, D. G. Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. *Mol Cell* **70**, 48-59.e5 (2018).

37. Rollie, C., Graham, S., Rouillon, C. & White, M. F. NAR breakthrough article: Pre-spacer processing and specific integration in a type I-A CRISPR system. *Nucleic Acids Res* **46**, 1007–1020 (2018).

38. Shiimori, M., Garrett, S. C., Graveley, B. R. & Terns, M. P. Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci. *Mol Cell* **70**, 814-824.e6 (2018).

39. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* **3**, (2012).

40. Fineran, P. C. *et al.* Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc Natl Acad Sci U S A* **111**, (2014).

41. Jackson, S. A. *et al.* CRISPR-Cas: Adapting to change. *Science* **356**(6333), (2017).

42. Künne, T. *et al.* Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Mol Cell* **63**, 852–864 (2016).

43. Nussenzweig, P. M., McGinn, J. & Marraffini, L. A. Cas9 Cleavage of Viral Genomes Primes the Acquisition of New Immunological Memories. *Cell Host Microbe* **26**, 515-526.e6 (2019).

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

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

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

47. Wang, R., Preamplume, G., Terns, M. P., Terns, R. M. & Li, H. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: Recognition and cleavage. *Structure* **19**, 257–264 (2011).

48. Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science* **329**(5997), 1355–1358 (2010).

49. Hale, C., Kleppe, K., Terns, R. M. & Terns, M. P. Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA* **14**, 2572–2579 (2008).

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

51. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).

52. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**(6096), 816-821 (2012).

53. Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**(6299), (2016).

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

55. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759–771 (2015).

56. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **18**, 529–536 (2011).

57. Redding, S. *et al.* Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell* **163**, 854–865 (2015).

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

59. Vink, J. N. A. *et al.* Direct Visualization of Native CRISPR Target Search in Live Bacteria Reveals Cascade DNA Surveillance Mechanism. *Mol Cell* **77**, 39-50.e10 (2020).

60. Leenay, R. T. *et al.* Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol Cell* **62**, 137–147 (2016).

61. Swarts, D. C., van der Oost, J. & Jinek, M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol Cell* **66**, 221-233.e4 (2017).

62. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32**, 670–676 (2014).

63. Rutkauskas, M. *et al.* Directional R-loop formation by the CRISPR-cas surveillance complex cascade provides efficient off-target site rejection. *Cell Rep* **10**, 1534–1543 (2015).

64. Loeff, L., Brouns, S. J. J. & Joo, C. Repetitive DNA Reeling by the Cascade-Cas3 Complex in Nucleotide Unwinding Steps. *Mol Cell* **70**, 385-394.e3 (2018).

65. Westra, E. R. *et al.* CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Mol Cell* **46**, 595–605 (2012).

66. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, (2012).

67. Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell* **60**, 385–397 (2015).

68. Wu, D., Guan, X., Zhu, Y., Ren, K. & Huang, Z. Structural basis of stringent PAM recognition by CRISPR-C2c1 in complex with sgRNA. *Cell Research* vol. 27 705–708 (2017).

69. Yan, W. X. *et al.* Functionally diverse type V CRISPR-Cas systems. *Science* **363**(6422), 88–91 (2019).

70. Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č. & 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).

71. Bravo, J. P. K. *et al.* RNA targeting unleashes indiscriminate nuclease activity of CRISPR–Cas12a2. *Nature* **613**, 582–587 (2023).

72. Chen, J. S. *et al.* CRISPR-Cas12a Target Binding Unleashes Indiscriminate Single-Stranded DNase Activity. *Science* **360**(6387), 436-439 (2018).

73. Niewoehner, O. *et al.* Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 (2017).

74. You, L. *et al.* Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. *Cell* **176**, 239-253.e16 (2019).

75. Steens, J. A., Salazar, C. R. P. & Staals, R. H. J. The diverse arsenal of type III CRISPR–Cas associated CARF and SAVED effectors. *Biochemical Society Transactions* vol. **50** 1353–1364 (2022).

76. Lin, J., Shen, Y., Ni, J. & She, Q. A type III-A CRISPR-Cas system mediates co-Transcriptional DNA cleavage at the transcriptional bubbles in close proximity to active effectors. *Nucleic Acids Res* **49**, 7628–7643 (2021).

77. Samai, P. *et al.* Co-transcriptional DNA and RNA cleavage during type III CRISPR-cas immunity. *Cell* **161**, 1164–1174 (2015).

78. Steens, J. A. *et al.* SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation. *Nat Commun* **12**, (2021).

79. Benda, C. *et al.* Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4. *Mol Cell* **56**, 43–54 (2014).

80. Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W. & Marraffini, L. A. A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *Journal of Biological Chemistry* **288**, 27888–27897 (2013).

81. Staals, R. H. J. *et al.* Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. *Mol Cell* **52**, 135–145 (2013).

82. David W. Taylor *et al.* Structures of the CRISPR-Cmr complex reveal mode ofRNA target positioning. *Science (1979)* **348**, 578–581 (2015).

83. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A Cyclic Oligonucleotide Signaling Pathway in Type III CRISPR-Cas Systems. *Science* **357**(6351) 605-609 (2017).

84. Makarova, K. S. *et al.* Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic Acids Research* vol. **48** 8828–8847 (2020).

85. Chi, H. *et al.* Antiviral type III CRISPR signalling via conjugation of ATP and SAM. *Nature* **622**, 826–833 (2023).

86. Mayo-Muñoz, D. *et al.* Type III CRISPR-Cas provides resistance against nucleus-forming jumbo phages via 1 abortive infection 2. *Mol Cell* **82**, 4471–4486 (2022).

87. Hogrel, G. *et al.* Cyclic nucleotide-induced helical structure activates a TIR immune effector. *Nature* **608**, 808–812 (2022).

88. Lau, R. K. *et al.* Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. *Mol Cell* **77**, 723-733.e6 (2020).

89. McMahon, S. A. *et al.* Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nat Commun* **11**, (2020).

90. Nemudraia, A. *et al.* Sequence-speciic capture and concentration of viral RNA by type III CRISPR system enhances diagnostic. *Nature* **13**, 7762 (2022).

91. Rostøl, J. T. & Marraffini, L. A. Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR–Cas immunity. *Nat Microbiol* **4**, 656–662 (2019).

92. Rostøl, J. T. *et al.* The Card1 nuclease provides defence during type III CRISPR immunity. *Nature* **590**, 624–629 (2021).

93. Rouillon, C. *et al.* Antiviral signalling by a cyclic nucleotide activated CRISPR protease. *Nature* **614**, 168–174 (2023).

94. Koskella, B. & Brockhurst, M. A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev* **38**, 916–931 (2014).

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

96. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* vol. **172** 1239–1259 (2018).

97. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722–736 (2015).

98. Weissman, J. L., Laljani, R. M. R., Fagan, W. F. & Johnson, P. L. F. Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy. *ISME Journal* **13**, 2589–2602 (2019).

99. Burstein, D. *et al.* New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).

100. Burstein, D. *et al.* Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat Commun* **7**, (2016).

101. Bernheim, A., Bikard, D., Touchon, M. & Rocha, E. P. C. A matter of background: DNA repair pathways as a possible cause for the sparse distribution of CRISPR-Cas systems in bacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, (2019).

102. Bernheim, A. *et al.* Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria. *Nat Commun* **8**, (2017).

103. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **108**, 10098–10103 (2011).

104. Watson, B. N. J. *et al.* Type I-F CRISPR-Cas resistance against virulent phages results in abortive infection and provides population-level immunity. *Nat Commun* **10**, (2019).

105. Garneau, J. E. *et al.* The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).

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

107. Hale, C. R. *et al.* RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* **139**, 945–956 (2009).

108. Staals, R. H. J. *et al.* RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of Thermus thermophilus. *Mol Cell* **56**, 518–530 (2014).

109. Tamulaitis, G. *et al.* Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Mol Cell* **56**, 506–517 (2014).

110. Zhang, J. *et al.* Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity. *Mol Cell* **45**, 303–313 (2012).

111. Pyenson, N. C., Gayvert, K., Varble, A., Elemento, O. & Marraffini, L. A. Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. *Cell Host Microbe* **22**, 343-353.e3 (2017).

112. 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).

113. Pinilla-Redondo, R. *et al.* Type IV CRISPR-Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Res* **48**, 2000–2012 (2020).

114. Varble, A. & Marraffini, L. A. Three New Cs for CRISPR: Collateral, Communicate, Cooperate. *Trends in Genetics* vol. **35** 446–456 (2019).

115. Meeske, A. J., Nakandakari-Higa, S. & Marraffini, L. A. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* **570**, 241–245 (2019).

116. Westra, E. R. & Levin, B. R. It is unclear how important CRISPR-Cas systems are for protecting natural populations of bacteria against infections by mobile genetic elements. *Proceedings of the National Academy of Sciences* **117**, 27777–27785 (2020).

117. Dimitriu, T., Szczelkun, M. D. & Westra, E. R. Evolutionary Ecology and Interplay of Prokaryotic Innate and Adaptive Immune Systems. *Current Biology* vol. **30** R1189–R1202 (2020).

118. van Houte, S., Buckling, A. & Westra, E. R. Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiology and Molecular Biology Reviews* **80**, 745–763 (2016).

119. Gurney, J., Pleška, M. & Levin, B. R. Why put up with immunity when there is resistance: An excursion into the population and evolutionary dynamics of restriction-modification and CRISPR-Cas. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, (2019).

120. Chevallereau, A., Meaden, S., Van Houte, S., Westra, E. R. & Rollie, C. The effect of bacterial mutation rate on the evolution of CRISPR-Cas adaptive immunity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, (2019).

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

122. Yang, Y. *et al.* The antiviral and antitumor effects of defective interfering particles/genomes and their mechanisms. *Front Microbiol* **10**, (2019).

123. Levy, A. *et al.* CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505–510 (2015).

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

125. Patterson, A. G., Yevstigneyeva, M. S. & Fineran, P. C. Regulation of CRISPR–Cas adaptive immune systems. *Current Opinion in Microbiology* vol. **37** 1–7 (2017).

126. Shivram, H., Cress, B. F., Knott, G. J. & Doudna, J. A. Controlling and enhancing CRISPR systems. *Nature Chemical Biology* vol. **17** 10–19 (2021).

127. Weissman, J. L., Stoltzfus, A., Westra, E. R. & Johnson, P. L. F. Avoidance of Self during CRISPR Immunization. *Trends in Microbiology* vol. **28** 543–553 (2020).

128. Patterson, A. G. *et al.* Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems. *Mol Cell* **64**, 1102–1108 (2016).

129. Smith, L. M. *et al.* The Rcs stress response inversely controls surface and CRISPR–Cas adaptive immunity to discriminate plasmids and phages. *Nat Microbiol* **6**, 162–172 (2021).

130. Borges, A. L. *et al.* Bacterial alginate regulators and phage homologs repress CRISPR–Cas immunity. *Nat Microbiol* **5**, 679–687 (2020).

131. Broniewski, J. M., Chisnall, M. A. W., Høyland-Kroghsbo, N. M., Buckling, A. & Westra, E. R. The effect of Quorum sensing inhibitors on the evolution of CRISPR-based phage immunity in Pseudomonas aeruginosa. *ISME Journal* **15**, 2465–2473 (2021).

132. Høyland-Kroghsbo, N. M. *et al.* Quorum sensing controls the pseudomonas aeruginosa CRISPR-Cas adaptive immune system. *Proc Natl Acad Sci U S A* **114**, 131–135 (2017).

133. Agari, Y. *et al.* Transcription Profile of Thermus thermophilus CRISPR Systems after Phage Infection. *J Mol Biol* **395**, 270–281 (2010).

134. He, F., Vestergaard, G., Peng, W., She, Q. & Peng, X. CRISPR-Cas type I-A Cascade complex couples viral infection surveillance to host transcriptional regulation in the dependence of Csa3b. *Nucleic Acids Res* **45**, 1902–1913 (2017).

135. Quax, T. E. F. *et al.* Massive Activation of Archaeal Defense Genes during Viral Infection. *J Virol* **87**, 8419–8428 (2013).

136. Ratner, H. K., Sampson, T. R. & Weiss, D. S. I can see CRISPR now, even when phage are gone: A view on alternative CRISPR-Cas functions from the prokaryotic envelope. *Current Opinion in Infectious Diseases* vol. **28** 267–274 (2015).

137. Patterson, A. G., Chang, J. T., Taylor, C. & Fineran, P. C. Regulation of the type I-F CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. *Nucleic Acids Res* **43**, 6038–6048 (2015).

138. Staals, R. H. J. *et al.* Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat Commun* **7**, (2016).

139. Westra, E. R. *et al.* Parasite exposure drives selective evolution of constitutive versus inducible defense. *Current Biology* **25**, 1043–1049 (2015).

140. Vale, P. F. *et al.* Costs of CRISPR-Cas-mediated resistance in Streptococcus thermophilus. *Proceedings of the Royal Society B: Biological Sciences* **282**, (2015).

141. Meaden, S. *et al.* Phage gene expression and host responses lead to infection-dependent costs of CRISPR immunity. *ISME Journal* **15**, 534–544 (2021).

142. Chabas, H., Van Houte, S., Høyland-Kroghsbo, N. M., Buckling, A. & Westra, E. R. Immigration of susceptible hosts triggers the evolution of alternative parasite defence strategies. *Proceedings of the Royal Society B: Biological Sciences* **283**, (2016).

143. Alseth, E. O. *et al.* Bacterial biodiversity drives the evolution of CRISPR-based phage resistance. *Nature* **574**, 549–552 (2019).

144. Iranzo, J., Lobkovsky, A. E., Wolf, Y. I. & Koonin, E. V. Evolutionary dynamics of the prokaryotic adaptive immunity system CRISPR-Cas in an explicit ecological context. *J Bacteriol* **195**, 3834–3844 (2013).

145. Weinberger, A. D., Wolf, Y. I., Lobkovsky, A. E., Gilmore, M. S. & Koonin, E. V. Viral diversity threshold for adaptive immunity in prokaryotes. *mBio* **3**, (2012).

146. Broniewski, J. M., Meaden, S., Paterson, S., Buckling, A. & Westra, E. R. The effect of phage genetic diversity on bacterial resistance evolution. *ISME Journal* **14**, 828–836 (2020).

147. Chevallereau, A. *et al.* Exploitation of the Cooperative Behaviors of Anti-CRISPR Phages. *Cell Host Microbe* **27**, 189-198.e6 (2020).

148. Wimmer, F. & Beisel, C. L. CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers. *Frontiers in Microbiology* vol. **10** (2020).

149. Stern, A., Keren, L., Wurtzel, O., Amitai, G. & Sorek, R. Self-targeting by CRISPR: Gene regulation or autoimmunity? *Trends in Genetics* vol. **26** 335–340 (2010).

150. Gomaa, A. A. *et al.* Programmable removal of bacterial strains by use of genome- targeting CRISPR-cas systems. *mBio* **5**, (2014).

151. Vercoe, R. B. *et al.* Cytotoxic Chromosomal Targeting by CRISPR/Cas Systems Can Reshape Bacterial Genomes and Expel or Remodel Pathogenicity Islands. *PLoS Genet* **9**, (2013).

152. Nobrega, F. L., Walinga, H., Dutilh, B. E. & Brouns, S. J. J. Prophages are associated with extensive CRISPR-Cas auto-immunity. *Nucleic Acids Res* **48**, 12074–12084 (2020).

153. Rollie, C. *et al.* Targeting of temperate phages drives loss of type I CRISPR–Cas systems. *Nature* **578**, 149–153 (2020).

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

155. Goldberg, G. W. *et al.* Incomplete prophage tolerance by type III-A CRISPR-Cas systems reduces the fitness of lysogenic hosts. *Nat Commun* **9**, (2018).

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

157. Marraffini, L. A. & Sontheimer, E. J. CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science* **322**(5909), 1843–1845 (2008).

158. Watson, B. N. J., Staals, R. H. J. & Fineran, P. C. CRISPR-cas-mediated phage resistance enhances horizontal gene transfer by transduction. *mBio* **9**, (2018).

159. Jiang, W. *et al.* Dealing with the Evolutionary Downside of CRISPR Immunity: Bacteria and Beneficial Plasmids. *PLoS Genet* **9**, (2013).

160. Hatoum-Aslan, A. & Marraffini, L. A. Impact of CRISPR immunity on the emergence and virulence of bacterial pathogens. *Current Opinion in Microbiology* vol. **17** 82–90 (2014).

161. Palmer, K. L. & Gilmore, M. S. Multidrug-resistant enterococci lack CRISPR-cas. *mBio* **1**, (2010).

162. Wheatley, R. M. & MacLean, R. C. CRISPR-Cas systems restrict horizontal gene transfer in Pseudomonas aeruginosa. *ISME Journal* **15**, 1420–1433 (2021).

163. Gophna, U. *et al.* No evidence of inhibition of horizontal gene transfer by CRISPR-Cas on evolutionary timescales. *ISME Journal* **9**, 2021–2027 (2015).

164. Malone, L. M., Birkholz, N. & Fineran, P. C. Conquering CRISPR: how phages overcome bacterial adaptive immunity. *Current Opinion in Biotechnology* vol. **68** 30–36 (2021).

165. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J Bacteriol* **190**, 1390–1400 (2008).

166. Chabas, H. *et al.* Evolutionary emergence of infectious diseases in heterogeneous host populations. *PLoS Biol* **16**, (2018).

167. Chabas, H. *et al.* Variability in the durability of CRISPR-Cas immunity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, (2019).

168. Common, J., Morley, D., Westra, E. R. & Van Houte, S. CRISPR-Cas immunity leads to a coevolutionary arms race between Streptococcus thermophilus and lytic phage. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, (2019).

169. Watson, B. N. J. *et al.* Different genetic and morphological outcomes for phages targeted by single or multiple CRISPR-Cas spacers. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, (2019).

170. Han, P., Niestemski, L. R., Barrick, J. E. & Deem, M. W. Physical model of the immune response of bacteria against bacteriophage through the adaptive CRISPR-Cas immune system. *Phys Biol* **10**, (2013).

171. Paez-Espino, D. *et al.* CRISPR immunity drives rapid phage genome evolution in streptococcus thermophilus. *mBio* **6**, 1–9 (2015).

172. Vlot, M. *et al.* 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**, 873–885 (2018).

173. Malone, L. M. *et al.* A jumbo phage that forms a nucleus-like structure evades CRISPR–Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nature Microbiology* vol. **5** 48–55 (2020).

174. Mendoza, S. D. *et al.* A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. *Nature* **577**, 244–248 (2020).

175. Li, Y. & Bondy-Denomy, J. Anti-CRISPRs go viral: The infection biology of CRISPR-Cas inhibitors. *Cell Host and Microbe* vol. **29** 704–714 (2021).

176. Bradde, S., Vucelja, M., Teşileanu, T. & Balasubramanian, V. Dynamics of adaptive immunity against phage in bacterial populations. *PLoS Comput Biol* **13**, (2017).

177. Han, P. & Deem, M. W. Non-classical phase diagram for virus bacterial coevolution mediated by clustered regularly interspaced short palindromic repeats. *J R Soc Interface* **14**, (2017).

178. Levin, B. R., Moineau, S., Bushman, M. & Barrangou, R. The Population and Evolutionary Dynamics of Phage and Bacteria with CRISPR-Mediated Immunity. *PLoS Genet* **9**, (2013).

179. Weissman, J. L. *et al.* Immune loss as a driver of coexistence during host-phage coevolution. *ISME Journal* **12**, 585–597 (2018).

180. Weinberger, A. D. *et al.* Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput Biol* **8**, (2012).

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

182. Zoephel, J. & Randau, L. RNA-Seq analyses reveal CRISPR RNA processing and regulation patterns. in *Biochemical Society Transactions* vol. 41 1459–1463 (2013).

183. Bradde, S., Nourmohammad, A., Goyal, S. & Balasubramanian, V. The size of the immune repertoire of bacteria. *roceedings of the National Academy of Sciences* **117**, 5144–5151 (2020).

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

185. Martynov, A., Severinov, K. & Ispolatov, I. Optimal number of spacers in CRISPR arrays. *PLoS Comput Biol* **13**, (2017).

186. Weissman, JL., Fagan, W. F. & Johnson, P. L. F. Selective Maintenance of Multiple CRISPR Arrays Across Prokaryotes. *CRISPR J* **1**, 405–413 (2018).

187. Levin, B. R. Nasty viruses, costly plasmids, population dynamics, and the conditions for establishing and maintaining CRISPR-mediated adaptive immunity in bacteria. *PLoS Genet* **6**, 1–12 (2010).

188. Childs, L. M., England, W. E., Young, M. J., Weitz, J. S. & Whitaker, R. J. CRISPR-induced distributed immunity in microbial populations. *PLoS One* **9**, (2014).

189. Morley, D., Broniewski, J. M., Westra, E. R., Buckling, A. & van Houte, S. Host diversity limits the evolution of parasite local adaptation. *Mol Ecol* **26**, 1756–1763 (2017).

190. Van Houte, S. *et al.* The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532**, 385–388 (2016).

191. Paez-Espino, D. *et al.* Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* **4**, (2013).

192. Heler, R., Wright, A. V., Vucelja, M., Doudna, J. A. & Marraffini, L. A. Spacer Acquisition Rates Determine the Immunological Diversity of the Type II CRISPR-Cas Immune Response. *Cell Host Microbe* **25**, 242-249.e3 (2019).

193. Pyenson, N. C. & Marraffini, L. A. Co-evolution within structured bacterial communities results in multiple expansion of CRISPR loci and enhanced immunity. *Elife* **9**, (2020).

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

195. Cao, L. *et al.* Identification and functional study of type III-A CRISPR-Cas systems in clinical isolates of Staphylococcus aureus. *International Journal of Medical Microbiology* **306**, 686–696 (2016).

196. Manica, A., Zebec, Z., Steinkellner, J. & Schleper, C. Unexpectedly broad target recognition of the CRISPR-mediated virus defence system in the archaeon sulfolobus solfataricus. *Nucleic Acids Res* **41**, 10509–10517 (2013).

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

198. Wang, L. *et al.* Dynamics of Cas10 Govern Discrimination between Self and Non-self in Type III CRISPR-Cas Immunity. *Mol Cell* **73**, 278-290.e4 (2019).

199. Andersson, A. F. & Banfield, J. F. Virus Population Dynamics and Acquired Virus Resistance in Natural Microbial Communities. *Science* **320**(5879), 1047–1050 (2008).

200. Laanto, E., Hoikkala, V., Ravantti, J. & Sundberg, L. R. Long-term genomic coevolution of host-parasite interaction in the natural environment. *Nat Commun* **8**, (2017).

201. Cornuault, J. K. *et al.* The enemy from within: a prophage of Roseburia intestinalis systematically turns lytic in the mouse gut, driving bacterial adaptation by CRISPR spacer acquisition. *ISME Journal* **14**, 771–787 (2020).

202. Martínez Arbas, S. *et al.* Roles of bacteriophages, plasmids and CRISPR immunity in microbial community dynamics revealed using time-series integrated meta-omics. *Nat Microbiol* **6**, 123–135 (2021).

203. Sun, C. L., Thomas, B. C., Barrangou, R. & Banfield, J. F. Metagenomic reconstructions of bacterial CRISPR loci constrain population histories. *ISME Journal* **10**, 858–870 (2016).

204. Jia, N., Jones, R., Yang, G., Ouerfelli, O. & Patel, D. J. CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise cA4 Cleavage with ApA>p Formation Terminating RNase Activity. *Mol Cell* **75**, 944-956.e6 (2019).

205. Jiang, W., Samai, P. & Marraffini, L. A. Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity. *Cell* **164**, 710–721 (2016).

206. Garcia-Doval, C. *et al.* Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. *Nat Commun* **11**, (2020).

207. Smalakyte, D. *et al.* Type III-A CRISPR-associated protein Csm6 degrades cyclic hexa-adenylate activator using both CARF and HEPN domains. *Nucleic Acids Res* **48**, 9204–9217 (2020).

208. Foster, K., Kalter, J., Woodside, W., Terns, R. M. & Terns, M. P. The ribonuclease activity of Csm6 is required for anti-plasmid immunity by Type III-A CRISPR-Cas systems. *RNA Biol* **16**, 449–460 (2019).

209. Grüschow, S., Athukoralage, J. S., Graham, S., Hoogeboom, T. & White, M. F. Cyclic oligoadenylate signalling mediates Mycobacterium tuberculosis CRISPR defence. *Nucleic Acids Res* **47**, 9259–9270 (2019).

210. Rouillon, C., Athukoralage, J. S., Graham, S., Grü, S. & White, M. F. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *Elife* **7** (2018).

211. Athukoralage, J. S., Graham, S., Grüschow, S., Rouillon, C. & White, M. F. A Type III CRISPR Ancillary Ribonuclease Degrades Its Cyclic Oligoadenylate Activator. *J Mol Biol* **431**, 2894–2899 (2019).

212. Foster, K., Grüschow, S., Bailey, S., White, M. F. & Terns, M. P. Regulation of the RNA and DNA nuclease activities required for Pyrococcus furiosus Type III-B CRISPR-Cas immunity. *Nucleic Acids Res* **48**, 4418–4434 (2020).

213. Han, W., Pan, S., López-Méndez, B., Montoya, G. & She, Q. Allosteric regulation of Csx1, a type IIIB-associated CARF domain ribonuclease by RNas carrying a tetraadenylate tail. *Nucleic Acids Res* **45**, 10740–10750 (2017).

214. Molina, R. *et al.* Structure of Csx1-cOA4 complex reveals the basis of RNA decay in Type III-B CRISPR-Cas. *Nat Commun* **10**, (2019).

215. Molina, R. *et al.* Structural basis of cyclic oligoadenylate degradation by ancillary Type III CRISPR-Cas ring nucleases. *Nucleic Acids Res* **49**, 12577–12590 (2021).

216. Athukoralage, J. S., Rouillon, C., Graham, S., Grüschow, S. & White, M. F. Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. *Nature* **562**, 277–280 (2018).

217. Zhu, W. *et al.* The CRISPR ancillary effector Can2 is a dual-specificity nuclease potentiating type III CRISPR defence. *Nucleic Acids Res* **49**, 2777–2789 (2021).

218. Charbonneau, A. A., Eckert, D. M., Gauvin, C. C., Lintner, N. G. & Lawrence, C. M. Cyclic tetra-adenylate (Ca4) recognition by csa3; implications for an integrated class 1 crispr-cas immune response in saccharolobus solfataricus. *Biomolecules* **11**, (2021).

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

220. Niewoehner, O. & Jinek, M. Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *RNA* **22**, 318–329 (2016).

221. Sheppard, N. F., Glover, C. V. C., Terns, R. M. & Terns, M. P. The CRISPR-associated Csx1 protein of Pyrococcus furiosus is an adenosine-specific endoribonuclease. *RNA* **22**, 216–224 (2016).

222. Kim, Y. K., Kim, Y. G. & Oh, B. H. Crystal structure and nucleic acid-binding activity of the CRISPR-Associated protein Csx1 of Pyrococcus furiosus. *Proteins: Structure, Function and Bioinformatics* **81**, 261–270 (2013).

223. Athukoralage, J. S. *et al.* The dynamic interplay of host and viral enzymes in type iii crispr-mediated cyclic nucleotide signalling. *Elife* **9**, (2020).

224. Aizenman, E., Engelberg-Kulkat, H. & Glaser, G. An Escherichia Coli Chromosomal 'Addiction Module' Regulated by 3',5'-Bispyrophosphate: A Model for Programmed Bacterial Cell Death. *Genetics* vol. **93** (1996).

225. Bingham, R., Ekunwe, S. I. N., Falk, S., Snyder, L. & Kleanthous, C. The major head protein of bacteriophage T4 binds specifically to elongation factor Tu. *Journal of Biological Chemistry* **275**, 23219–23226 (2000).

226. 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.e5 (2021).

227. Lopatina, A., Tal, N. & Sorek, R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annu. Rev. Virol* **7**, 371–384 (2020).

228. Parma, D. H. *et al. The Rex System of Bacteriophage k: Tolerance and Altruistic Cell Death*. (1992).

229. Tal, N. *et al.* Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell* **184**, 5728-5739.e16 (2021).

230. Tal, N. *et al.* Bacteria deplete deoxynucleotides to defend against bacteriophage infection. *Nat Microbiol* **7**, 1200–1209 (2022).

231. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1**, 0474–0483 (2005).

232. Lintner, N. G. *et al.* The structure of the CRISPR-associated protein csa3 provides insight into the regulation of the CRISPR/Cas system. *J Mol Biol* **405**, 939–955 (2011).

233. Xia, P., Dutta, A., Gupta, K., Batish, M. & Parashar, V. Structural basis of cyclic oligoadenylate binding to the transcription factor Csa3 outlines cross talk between type III and type I CRISPR systems. *Journal of Biological Chemistry* **298**, (2022).

234. Ye, Q. *et al.* CRISPR-Associated Factor Csa3b Regulates CRISPR Adaptation and Cmr-Mediated RNA Interference in Sulfolobus islandicus. *Front Microbiol* **11**, (2020).

235. Liu, T. *et al.* Coupling transcriptional activation of CRISPR-Cas system and DNA repair genes by Csa3a in Sulfolobus islandicus. *Nucleic Acids Res* **45**, 8978–8992 (2017).

236. Cohen, D. *et al.* Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574**, 691–695 (2019).

237. Wolf, J. *et al.* tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. *The EMBO journal* **21**(14), 3841-3851 (2002).

238. Lee, I. & Suzuki, C. K. Functional mechanics of the ATP-dependent Lon protease- lessons from endogenous protein and synthetic peptide substrates. *Biochimica et Biophysica Acta - Proteins and Proteomics* vol. **1784** 727–735 (2008).

239. Pinti, M. *et al.* Emerging role of Lon protease as a master regulator of mitochondrial functions. *Biochim Biophys Acta Bioenerg* **1857**, 1300–1306 (2016).

240. Wlodawer, A., Sekula, B., Gustchina, A. & Rotanova, T. V. Structure and the Mode of Activity of Lon Proteases from Diverse Organisms. *Journal of Molecular Biology* vol. **434** (2022).

241. Rotanova, T. V. *et al.* Slicing a protease: Structural features of the ATP-dependent Lon proteases gleaned from investigations of isolated domains. *Protein Science* **15**, 1815–1828 (2006).

242. Lowey, B. & Kranzusch, P. J. CD-NTases and nucleotide second messenger signaling. *Current Biology* **30**, 1106–1108 (2020).

243. Duncan-Lowey, B. & Kranzusch, P. J. CBASS phage defense and evolution of antiviral nucleotide signaling. *Current Opinion in Immunology* vol. **74** 156–163 (2022).

244. Millman, A., Melamed, S., Amitai, G. & Sorek, R. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. *Nat Microbiol* **5**, 1608–1615 (2020).

245. Morehouse, B. R. *et al.* STING cyclic dinucleotide sensing originated in bacteria. *Nature* **586**, 429–433 (2020).

246. White, M. F. Bacteria SAVED from Viruses. *Cell* vol. **182** 5–6 (2020).

247. Lowey, B. *et al.* 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**, 38-49.e17 (2020).

248. Whiteley, A. T. *et al.* Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).

249. Burroughs, A. M., Zhang, D., Schäffer, D. E., Iyer, L. M. & Aravind, L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res* **43**, 10633–10654 (2015).

250. 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).

251. Nariya, H. & Inouye, M. MazF, an mRNA Interferase, Mediates Programmed Cell Death during Multicellular Myxococcus Development. *Cell* **132**, 55–66 (2008).

252. Zhang, Y. *et al.* MazF Cleaves Cellular mRNAs Specifically at ACA to Block Protein Synthesis in Escherichia coli. *Mol Cell* **12**, 913–923 (2003).

253. Fink, S. L. & Cookson, B. T. Apoptosis, pyroptosis, and necrosis: Mechanistic description of dead and dying eukaryotic cells. *Infection and Immunity* vol. **73** 1907–1916 (2005).

254. Koonin, E. V & Aravind, L. Origin and evolution of eukaryotic apoptosis: the bacterial connection. *Cell Death Differ* **9**, 394–404 (2002).

255. Essuman, K., Milbrandt, J., Dangl, J. L. & Nishimura, M. T. Shared TIR enzymatic functions regulate cell death and immunity across the tree of life. *Science* vol. **377**(6605) (2022).

256. Essuman, K. *et al.* TIR Domain Proteins Are an Ancient Family of NAD+-Consuming Enzymes. *Current Biology* **28**, 421-430.e4 (2018).

257. O'Neill, L. A. J. & Bowie, A. G. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nature Reviews Immunology* vol. **7** 353–364 (2007).

258. Wan, L. *et al.* TIR Domains of Plant Immune Receptors Are NAD +-Cleaving Enzymes That Promote Cell Death. *Science* **365**(6455) (2022).

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

260. Ofir, G. *et al.* Antiviral activity of bacterial TIR domains via immune signalling molecules. *Nature* **600**, 116–120 (2021).

261. Koopal, B. *et al.* Short prokaryotic Argonaute systems trigger cell death upon detection of invading DNA. *Cell* **185**, 1471-1486.e19 (2022).

262. Chaikeeratisak, V. *et al.* Assembly of a nucleus-like structure during viral replication in bacteria. *Science* **355**, 194–197 (2017).

263. Altae-Tran, H. *et al.* Uncovering the functional diversity of rare CRISPR-Cas systems with deep terascale clustering. *Science* **382**(6321), eadi1910 (2023).

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

265. Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V. & Aravind, L. CARF and WYL domains: Ligand-binding regulators of prokaryotic defense systems. *Front Genet* **5**, (2014).

266. 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).

267. Grüschow, S. *et al.* CRISPR antiphage defence mediated by the cyclic nucleotide-binding membrane protein Csx23 2 3. *bioRxiv* (2023) doi:10.1101/2023.11.24.568546.

268. Swarts, D. C. *et al.* DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507**, 258–261 (2014).

269. Audrito, V., Messana, V. G. & Deaglio, S. NAMPT and NAPRT: Two Metabolic Enzymes With Key Roles in Inflammation. *Frontiers in Oncology* vol. **10** (2020).

270. Shin, H. W. *et al.* An enzymatic cascade of Rab5 effectors regulates phosphoinositide turnover in the endocytic pathway. *Journal of Cell Biology* **170**, 607–618 (2005).

271. Nishimura, T. *et al.* Human MARF1 is an endoribonuclease that interacts with the DCP1:2 decapping complex and degrades target mRNAs. *Nucleic Acids Res* **46**, 12008–12021 (2018).

272. Staals, R. H. J. & Brouns, S. J. J. Distribution and mechanism of the type I CRISPR-Cas systems. in *CRISPR-Cas Systems: RNA-Mediated Adaptive Immunity in Bacteria and Archaea* 145–169, (2012).

273. Zhao, R. *et al.* A Membrane-Associated DHH-DHHA1 Nuclease Degrades Type III CRISPR Second Messenger. *Cell Rep* **32**, (2020).

274. Ereño-Orbea, J., Oyenarte, I. & Martínez-Cruz, L. A. CBS domains: Ligand binding sites and conformational variability. *Archives of Biochemistry and Biophysics* vol. **540** 70–81 (2013).

275. Aravind, L., Wolf, Y. I. & Koonin, E. V. The ATP-Cone: An Evolutionarily Mobile, ATP-Binding Regulatory Domain. *J. Mol. Microbiol. Biotechnol* vol. **2**(2) 191-194 (2000).

276. Grinberg, I. *et al.* Functional analysis of the Streptomyces coelicolor NrdR ATP-cone domain: Role in nucleotide binding, oligomerization, and DNA interactions. *J Bacteriol* **191**, 1169–1179 (2009).

277. Johansson, R. *et al.* Structural Mechanism of Allosteric Activity Regulation in a Ribonucleotide Reductase with Double ATP Cones. *Structure* **24**, 906–917 (2016).

278. Zink, I. A., Wimmer, E. & Schleper, C. Heavily armed ancestors: CRISPR immunity and applications in archaea with a comparative analysis of CRISPR types in sulfolobales. *Biomolecules* vol. **10** 1–41 (2020).

279. Boyd, J. M., Drevland, R. M., Downs, D. M. & Graham, D. E. Archaeal ApbC/Nbp35 homologs function as iron-sulfur cluster carrier proteins. *J Bacteriol* **191**, 1490–1497 (2009).

280. Van Wietmarschen, N., Moradian, A., Morin, G. B., Lansdorp, P. M. & Uringa, E. J. The mammalian proteins MMS19, MIP18, and ANT2 are involved in cytoplasmic iron-sulfur cluster protein assembly. *Journal of Biological Chemistry* **287**, 43351–43358 (2012).

281. 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).

282. Medvedevid, K. E., Kinch, L. N., Schaeffer, R. D. & Grishin, N. V. Functional analysis of rossmann-like domains reveals convergent evolution of topology and reaction pathways. *PLoS Comput Biol* **15**, (2019).

283. Medvedev, K. E., Kinch, L. N., Dustin Schaeffer, R., Pei, J. & Grishin, N. V. A Fifth of the Protein World: Rossmann-like Proteins as an Evolutionarily Successful Structural unit. *J Mol Biol* **433**, (2021).

284. Fatma, S., Chakravarti, A., Zeng, X. & Huang, R. H. Molecular mechanisms of the CdnG-Cap5 antiphage defense system employing 3′,2′-cGAMP as the second messenger. *Nat Commun* **12**, (2021).

285. Slavik, K. M. & Kranzusch, P. J. Annual Review of Virology CBASS to cGAS-STING: The Origins and Mechanisms of Nucleotide Second Messenger Immune Signaling. *Annu. Rev. Virol. 2023* **10**, 423–453 (2023).

286. Feng, Y. *et al.* Comparative genomics and proteomics of Eubacterium maltosivorans: functional identification of trimethylamine methyltransferases and bacterial microcompartments in a human intestinal bacterium with a versatile lifestyle. *Environ Microbiol* **24**, 517–534 (2022).

287. Smaczniak, C. *et al.* Proteomics-based identification of low-abundance signaling and regulatory protein complexes in native plant tissues. *Nat Protoc* **7**, 2144–2158 (2012).

288. Wendrich, J. R., Boeren, S., Möller, B. K., Weijers, D. & De Rybel, B. *Plant Hormones Methods and Protocols*. http://www.springer.com/series/7651 (2017).

289. Bielow, C., Mastrobuoni, G. & Kempa, S. Proteomics Quality Control: Quality Control Software for MaxQuant Results. *J Proteome Res* **15**, 777–787 (2016).

290. Rgen Cox, J. *et al.* Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular and Cellular Proteomics* **13**, 2513–2526 (2014).

291. Varadi, M. *et al.* AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439–D444 (2022).

292. Holm, L. DALI and the persistence of protein shape. *Protein Science* **29**, 128–140 (2020).

293. Newton, A. C., Bootman, M. D. & Scott, J. Second messengers. *Cold Spring Harb Perspect Biol* **8**, (2016).

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

295. Newsom, S., Parameshwaran, H. P., Martin, L. & Rajan, R. The CRISPR-Cas Mechanism for Adaptive Immunity and Alternate Bacterial Functions Fuels Diverse Biotechnologies. *Frontiers in Cellular and Infection Microbiology* vol. **10** (2021).

296. Pan, R., Hu, K., Jiang, D., Samuni, U. & Mirkin, M. V. Electrochemical Resistive-Pulse Sensing. *J Am Chem Soc* **141**, 19555–19559 (2019).

297. Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics and Bioinformatics* vol. **14** 265–279 (2016).

298. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* **17**, (2016).

299. Jia, W. *et al.* Identification of Single-Molecule Catecholamine Enantiomers Using a Programmable Nanopore. *ACS Nano* **16**, 6615–6624 (2022).

300. Qie, F. *et al.* β-Cyclodextrin Functionalized Nanoporous Graphene Oxides for Efficient Resolution of Asparagine Enantiomers. *Chem Asian J* **13**, 2812–2817 (2018).

301. Noordijk, B. *et al.* baseLess: lightweight detection of sequences in raw MinION data. *Bioinformatics Advances* **3** (2023).

302. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* **20**, (2019).

303. Wick, R. R., Judd, L. M. & Holt, K. E. Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLoS Comput Biol* **14**, (2018).

304. Indolia, S., Goswami, A. K., Mishra, S. P. & Asopa, P. Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. in *Procedia Computer Science* vol. **132** 679–688 (2018).

305. Mrazova, I. & Kukacka, M. Can deep neural networks discover meaningful pattern features? *Procedia Computer Science* vol. **12** 194–199 (Elsevier B.V., 2012).

306. Grosberg, A. Y. & Rabin, Y. DNA capture into a nanopore: Interplay of diffusion and electrohydrodynamics. *Journal of Chemical Physics* **133**, (2010).

307. Derrington, I. M. *et al.* Nanopore DNA sequencing with MspA. *Proceedings of the National Academy of Sciences* **107**, 16060–16065 (2010).

308. Purnell, R. F. & Schmidt, J. J. Discrimination of single base substitutions in a DNA strand immobilized in a biological nanopore. *ACS Nano* **3**, 2533–2538 (2009).

309. Rang, F. J., Kloosterman, W. P. & de Ridder, J. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biology* vol. **19** (2018).

310. Cao, C. *et al.* Discrimination of oligonucleotides of different lengths with a wild-type aerolysin nanopore. *Nat Nanotechnol* **11**, 713–718 (2016).

311. Van der Verren, S. E. *et al.* A dual-constriction biological nanopore resolves homonucleotide sequences with high fidelity. *Nat Biotechnol* **38**, 1415–1420 (2020).

312. Zhou, W., Qiu, H., Guo, Y. & Guo, W. Molecular Insights into Distinct Detection Properties of α-Hemolysin, MspA, CsgG, and Aerolysin Nanopore Sensors. *Journal of Physical Chemistry B* **124**, 1611–1618 (2020).

313. Maglia, G., Heron, A. J., Stoddart, D., Japrung, D. & Bayley, H. Analysis of Single Nucleic Acid Molecules with Protein Nanopores. in *Methods in Enzymology* vol. **475** 591–623 (2010).

314. Martín Abadi *et al.* TensorFlow: A System for Large-Scale Learning. (USENIX Association, 2005).

315. Pedregosa, F. *et al.* Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* vol. **12** (2011).

316. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261–272 (2020).

317. Liu, X. *et al.* Target RNA activates the protease activity of Craspase to confer antiviral defense. *Mol Cell* **82**, 4503-4518.e8 (2022).

318. Goswami, H. N., Rai, J., Das, A. & Li, H. Molecular mechanism of active Cas7-11 in processing CRISPR RNA and interfering target RNA. *Elife* **11**, (2022).

319. Hu, C. *et al.* Craspase is a CRISPR RNA-guided, RNA-activated protease. *Science* **377**(6612), 1278–1285 (2022).

320. Kato, K. *et al.* Structure and engineering of the type III-E CRISPR-Cas7-11 effector complex. *Cell* **185**, 2324-2337.e16 (2022).

321. Özcan, A. *et al.* Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. *Nature* **597**, 720–725 (2021).

322. Strecker, J. *et al.* RNA-Activated Protein Cleavage with a CRISPR-Associated Endopeptidase. *Science* **378**(6622), 874-881 (2023).

323. Van Beljouw, S. P. B. *et al.* The gRAMP CRISPR-Cas effector is an RNA endonuclease complexed with a caspase-like peptidase. *Science* **373**(6561), 1349–1353 (2021).

324. Wang, S., Guo, M., Zhu, Y., Lin, Z. & Huang, Z. Cryo-EM structure of the type III-E CRISPR-Cas effector gRAMP in complex with TPR-CHAT. *Cell Research* vol. **32** 1128–1131 (2022).

325. Yu, G. *et al.* Structure and function of a bacterial type III-E CRISPR–Cas7-11 complex. *Nat Microbiol* **7**, 2078–2088 (2022).

326. Steens, J. A., van der Oost, J. & Staals, R. H. J. Compact but mighty: Biology and applications of type III-E CRISPR-Cas systems. *Molecular Cell* vol. **82** 4405–4406 (2022).

327. 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).

328. Aravind, L. & Koonin, E. V. Classification of the caspase-hemoglobinase fold: Detection of new families and implications for the origin of the eukaryotic separins. *Proteins: Structure, Function and Genetics* **46**, 355–367 (2002).

329. McLuskey, K. & Mottram, J. C. Comparative structural analysis of the caspase family with other clan CD cysteine peptidases. *Biochemical Journal* vol. **466** 219–232 (2015).

330. Ichikawa, H. T. *et al.* Programmable type III-A CRISPR-Cas DNA targeting modules. *PLoS One* **12**, (2017).

331. Boatright, K. M. & Salvesen, G. S. Mechanisms of caspase activation. *Current Opinion in Cell Biology* vol. **15** 725–731 (2003).

332. Alnemri, E. S. *et al.* Human ICE/CED-3 Protease. *Cell* vol. **87**, 171 (1996).

333. Carmona-Gutierrez, D., Fröhlich, K. U., Kroemer, G. & Madeo, F. Editorial: Metacaspases are caspases. Doubt no more. *Cell Death and Differentiation* vol. **17** 377–378 (2010).

334. Talanian, R. V. *et al.* Substrate specificities of caspase family proteases. *Journal of Biological Chemistry* **272**, 9677–9682 (1997).

335. Vercammen, D. *et al.* Type II metacaspases Atmc4 and Atmc9 of Arabidopsis thaliana cleave substrates after arginine and lysine. *Journal of Biological Chemistry* **279**, 45329–45336 (2004).

336. Watanabe, N. & Lam, E. Two Arabidopsis metacaspases AtMCP1b and AtMCP2b are arginine/lysine- specific cysteine proteases and activate apoptosis-like cell death in yeast. *Journal of Biological Chemistry* **280**, 14691–14699 (2005).

337. Cohen, G. M. Caspases : the executioners of apoptosis. *Biochem. J* **326**, 1–16 (1997).

338. Fan, T. J., Han, L. H., Cong, R. S. & Liang, J. Caspase family proteases and apoptosis. *Acta Biochim Biophys Sin (Shanghai)* **37**, 719–727 (2005).

339. Depardieu, F. *et al.* A Eukaryotic-like Serine/Threonine Kinase Protects Staphylococci against Phages. *Cell Host Microbe* **20**, 471–481 (2016).

340. Gratani, F. L. *et al.* E. coli Toxin YjjJ (HipH) Is a Ser/Thr Protein Kinase That Impacts Cell Division, Carbon Metabolism, and Ribosome Assembly. *mSystems* **8**, (2023).

341. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372 (2008).

342. Fellers, R. T. *et al.* ProSight Lite: Graphical software to analyze top-down mass spectrometry data. *Proteomics* **15**, 1235–1238 (2015).

343. Bravo, J. P. K. *et al.* Structural basis of rotavirus RNA chaperone displacement and RNA annealing. **118**, (2021).

344. Fang, C. *et al.* Structures and mechanism of transcription initiation by bacterial ECF factors. *Nucleic Acids Res* **47**, 7094–7104 (2019).

345. Mohanraju, P. *et al.* Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* vol. **353**(6299) (2016).

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

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

348. Han, W. *et al.* A type III-B CRISPR-Cas effector complex mediating massive target DNA destruction. *Nucleic Acids Res* **45**, 1983–1993 (2017).

349. Jung, T. Y. *et al.* Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity. *Structure* **23**, 782–790 (2015).

350. Osawa, T., Inanaga, H., Sato, C. & Numata, T. Crystal structure of the crispr-cas RNA silencing cmr complex bound to a target analog. *Mol Cell* **58**, 418–430 (2015).

351. Ramia, N. F., Tang, L., Cocozaki, A. I. & Li, H. Staphylococcus epidermidis Csm1 is a 3′-5′ exonuclease. *Nucleic Acids Res* **42**, 1129–1138 (2014).

352. Hale, C. R., Cocozaki, A., Li, H., Terns, R. M. & Terns, M. P. Target RNA capture and cleavage by the Cmr type III-B CRISPR–cas effector complex. *Genes Dev* **28**, 2432–2443 (2014).

353. Spilman, M. *et al.* Structure of an RNA Silencing Complex of the CRISPR-Cas Immune System. *Mol Cell* **52**, 146–152 (2013).

354. Walker, F. C., Chou-Zheng, L., Dunkle, J. A. & Hatoum-Aslan, A. Molecular determinants for CRISPR RNA maturation in the Cas10-Csm complex and roles for non-Cas nucleases. *Nucleic Acids Res* **45**, 2112–2123 (2017).

355. Wiedenheft, B. *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings of the National Academy of Sciences*, **108**(25), 10092-10097 (2011).

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

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

358. Sofos, N. *et al.* Structures of the Cmr-β Complex Reveal the Regulation of the Immunity Mechanism of Type III-B CRISPR-Cas. *Mol Cell* **79**, 741-757.e7 (2020).

359. FIND evaluation update: SARS-CoV-2 molecular diagnostics.

360. van Dongen, J. E. *et al.* Point-of-care CRISPR/Cas nucleic acid detection: Recent advances, challenges and opportunities. *Biosens Bioelectron* **166**, (2020).

361. Rapid Detection of SARS-CoV-2 Using Reverse transcription RT-LAMP method. *Health Evaluation and Promotion* **47**, 248–250 (2020).

362. Artamonova, D. *et al.* Spacer acquisition by Type III CRISPR–Cas system during bacteriophage infection of Thermus thermophilus. *Nucleic Acids Res* **48**, 9787–9803 (2020).

363. Silas, S. *et al.* Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* **351**(6276) (2016).

364. Toro, N., Martínez-Abarca, F. & González-Delgado, A. The Reverse Transcriptases Associated with CRISPR-Cas Systems. *Sci Rep* **7**, (2017).

365. Lillestøl, R. K. *et al.* CRISPR families of the crenarchaeal genus Sulfolobus: Bidirectional transcription and dynamic properties. *Mol Microbiol* **72**, 259–272 (2009).

366. Pyenson, N. C. & Marraffini, L. A. Type III CRISPR-Cas systems: when DNA cleavage just isn't enough. *Current Opinion in Microbiology* vol. **37** 150–154 (2017).

367. Silas, S. *et al.* Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. *Elife* **6** (2017).

368. Gootenberg, J. S. *et al.* Multiplexed and Portable Nucleic Acid Detection Platform with Cas13, Cas12a, and Csm6. *Science* **360**(6387), 439-444 (2018).

369. Croll, T. I. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).

370. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* **27**, 14–25 (2018).

371. Gremmels, H. *et al.* Real-life validation of the Panbio™ COVID-19 antigen rapid test (Abbott) in community-dwelling subjects with symptoms of potential SARS-CoV-2 infection. *EClinicalMedicine* **31**, (2021).

372. Myhrvold, C. *et al.* Field-Deployable Viral Diagnostics Using CRISPR-Cas13. *Science* **360**(6387), 444-448 (2018).

373. Petri, K. & Pattanayak, V. SHERLOCK and DETECTR Open a New Frontier in Molecular Diagnostics. *CRISPR J* **1**, 209–211 (2018).

374. Hardinge, P. & Murray, J. A. H. Reduced False Positives and Improved Reporting of Loop-Mediated Isothermal Amplification using Quenched Fluorescent Primers. *Sci Rep* **9**, (2019).

375. Schneider, L., Blakely, H. & Tripathi, A. Mathematical model to reduce loop mediated isothermal amplification (LAMP) false-positive diagnosis. *Electrophoresis* **40**, 2706–2717 (2019).

376. Suleman, E., Mtshali, M. S. & Lane, E. Investigation of false positives associated with loop-mediated isothermal amplification assays for detection of Toxoplasma gondii in archived tissue samples of captive felids. *Journal of Veterinary Diagnostic Investigation* **28**, 536–542 (2016).

377. Trung, N. T. *et al.* CRISPR-Cas12a combination to alleviate the false-positive in loop-mediated isothermal amplification-based diagnosis of Neisseria meningitidis. *BMC Infect Dis* **22**, (2022).

378. Cramer, P. *et al.* Structure of eukaryotic RNA polymerases. *Annual Review of Biophysics* vol. **37** 337–352 (2008).

379. Jain, R., Aggarwal, A. K. & Rechkoblit, O. Eukaryotic DNA polymerases. *Current Opinion in Structural Biology* vol. **53** 77–87 (2018).

380. Seth A Darst. Bacterial RNA polymerase. *Curr Opin Struct Biol* **11**, 155–162 (2001).

381. Sousa, R., Chung, Y. J., Rose, J. P. & Wang, B.-C. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. *Nature* **364**, 593–599 (1993).

382. Young, M. C., Reddy, M. K. & Von Hippel, P. H. Perspectives in Biochemistry Structure and Function of the Bacteriophage T4 DNA Polymerase Holoenzyme. *Biochemistry* **31**(37), 8675-8690 (1992).

383. Cui, R. *et al.* Structural insights into the dual activities of the two-barrel RNA polymerase QDE-1. *Nucleic Acids Res* **50**, 10169–10186 (2022).

384. Liu, Y. *et al.* The DNA/RNA-dependent RNA polymerase QDE-1 generates aberrant RNA and dsRNA for RNAi in a process requiring replication protein a and a DNA helicase. *PLoS Biol* **8**, (2010).

385. Iyer, L. M., Koonin, E. V & Aravind, L. Evolutionary Connection between the Catalytic Subunits of DNA-Dependent RNA Polymerases and Eukaryotic RNA-Dependent RNA Polymerases and the Origin of RNA Polymerases. *BMC structural biology* **3**(1), 1-23 (2003).

386. Koonin, E. V., Krupovic, M., Ishino, S. & Ishino, Y. The replication machinery of LUCA: Common origin of DNA replication and transcription. *BMC Biol* **18**, (2020).

387. Oshima', T. & Imahori, K. Description of Thermus Thermophilus, a Nonsporulating Thermophilic Bacterium from a Japanese Thermal Spa. I*nternational Journal of Systematic and Evolutionary Microbiology* **24**(1)*,* 102-112 (1974)*.*

388. Aalto, A. P., Poranen, M. M., Grimes, J. M., Stuart, D. I. & Bamford, D. H. In vitro activities of the multifunctional RNA silencing polymerase QDE-1 of Neurospora crassa. *Journal of Biological Chemistry* **285**, 29367–29374 (2010).

389. Anees ur Rehman *et al.* The economic burden of chronic obstructive pulmonary disease (COPD) in the USA, Europe, and Asia: results from a systematic review of the literature. *Expert Rev Pharmacoecon Outcomes Res* **20**, 661–672 (2020).

390. Murphy, T. F. *et al.* Pseudomonas aeruginosa in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **177**, 853–860 (2008).

391. MacLeod, M. *et al.* Chronic obstructive pulmonary disease exacerbation fundamentals: Diagnosis, treatment, prevention and disease impact. *Respirology* vol. **26** 532–551 (2021).

392. Eklöf, J. *et al.* Pseudomonas aeruginosa and risk of death and exacerbations in patients with chronic obstructive pulmonary disease: an observational cohort study of 22 053 patients. *Clinical Microbiology and Infection* **26**, 227–234 (2020).

393. Martinez-García, M. A. *et al.* Long-Term Risk of Mortality Associated with Isolation of Pseudomonas aeruginosa in COPD: A Systematic Review and Meta-Analysis. *International Journal of COPD* **17**, 371–382 (2022).

394. Finch, S., McDonnell, M. J., Abo-Leyah, H., Aliberti, S. & Chalmers, J. D. A comprehensive analysis of the impact of pseudomonas aeruginosa colonization on prognosis in adult bronchiectasis. *Ann Am Thorac Soc* **12**, 1602–1611 (2015).

395. Dong, D. *et al.* Rapid detection of Pseudomonas aeruginosa targeting the toxA gene in intensive care unit patients from Beijing, China. *Front Microbiol* **6**, (2015).

396. Song, K. P., Chan, T. K., Ji, Z. L. & Wong, S. W. Rapid identification of Pseudomonas aeruginosa from ocular isolates by PCR using exotoxin A-specific primers. *Mol Cell Probes* **14**, 199–204 (2000).

397. Ooi, W. Y. *et al.* A Thermus phage protein inhibits host RNA polymerase by preventing template DNA strand loading during open promoter complex formation. *Nucleic Acids Res* **46**, 431–441 (2018).

398. Tagami, S. *et al.* Structural basis for promoter specificity switching of RNA polymerase by a phage factor. *Genes Dev* **28**, 521–531 (2014).

399. Carter, R. H., Demidenko, A. A., Hattingh-Willis, S. & Rothman-Denes, L. B. Phage N4 RNA polymerase II recruitment to DNA by a single-stranded DNA-binding protein. *Genes Dev* **17**, 2334–2345 (2003).

400. Cho, N.-Y., Choi, M. & Rothman-Denes, L. B. The Bacteriophage N4-Coded Single-Stranded DNA-Binding Protein (N4SSB) Is the Transcriptional Activator of Escherichia Coli RNA Polymerase at N4 Late Promoters. *J. Mol. Biol* vol. **246** (1995).

401. Laurila, M. R. L., Salgado, P. S., Stuart, D. I., Grimes, J. M. & Bamford, D. H. Back-priming mode of φ6 RNA-dependent RNA polymerase. *Journal of General Virology* **86**, 521–526 (2005).

402. Noton, S. L., Aljabr, W., Hiscox, J. A., Matthews, D. A. & Fearns, R. Factors affecting de novo RNA synthesis and back-priming by the respiratory syncytial virus polymerase. *Virology* **462–463**, 318–327 (2014).

403. Gray, D. C., Mahrus, S. & Wells, J. A. Activation of specific apoptotic caspases with an engineered small-molecule-activated protease. *Cell* **142**, 637–646 (2010).

404. Martell, J. D. *et al.* A split horseradish peroxidase for the detection of intercellular protein-protein interactions and sensitive visualization of synapses. *Nat Biotechnol* **34**, 774–780 (2016).

405. Ozawa, T., Kaihara, A., Sato, M., Tachihara, K. & Umezawa, Y. Split luciferase as an optical probe for detecting protein - Protein interactions in mammalian cells based on protein splicing. *Anal Chem* **73**, 2516–2521 (2001).

406. Paulmurugan, R. & Gambhir, S. S. Firefly luciferase enzyme fragment complementation for imaging in cells and living animals. *Anal Chem* **77**, 1295–1302 (2005).

407. Remy, I. & Michnick, S. W. A highly sensitive protein-protein interaction assay based on Gaussia luciferase. *Nat Methods* **3**, 977–979 (2006).

408. Shis, D. L. & Bennett, M. R. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc Natl Acad Sci U S A* **110**, 5028–5033 (2013).

409. Wehr, M. C. *et al.* Monitoring regulated protein-protein interactions using split TEV. *Nat Methods* **3**, 985–993 (2006).

410. van der Veer, H. J. *et al.* Glow-in-the-Dark Infectious Disease Diagnostics Using CRISPR-Cas9-Based Split Luciferase Complementation. *ACS Cent Sci* **9**, 657–667 (2023).

411. He, Y. *et al.* Ultrasensitive nucleic acid biosensor based on enzyme-gold nanoparticle dual label and lateral flow strip biosensor. *Biosens Bioelectron* **26**, 2018–2024 (2011).

412. Mao, X. *et al.* Disposable nucleic acid biosensors based on gold nanoparticle probes and lateral flow strip. *Anal Chem* **81**, 1660–1668 (2009).

413. Park, J. M. *et al.* Chemiluminescence lateral flow immunoassay based on Pt nanoparticle with peroxidase activity. *Anal Chim Acta* **853**, 360–367 (2015).

414. Callahan, B. P., Stanger, M. J. & Belfort, M. Protease Activation of Split Green Fluorescent Protein. *ChemBioChem* **11**, 2259–2263 (2010).

415. Wiegand, T. *et al.* Functional and Phylogenetic Diversity of Cas10 Proteins. *CRISPR Journal* **6**, 152–162 (2023).

416. Cai, Y. *et al.* A Nucleotidyltransferase Toxin Inhibits Growth of Mycobacterium Tuberculosis through Inactivation of TRNA Acceptor Stems. *Sci. Adv* vol. **6** (2020).

417. Dy, R. L., Przybilski, R., Semeijn, K., Salmond, G. P. C. & Fineran, P. C. A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic Acids Res* **42**, 4590–4605 (2014).

418. Hampton, H. G. *et al.* Functional genomics reveals the toxin–antitoxin repertoire and AbiE activity in serratia. *Microb Genom* **6**, 1–15 (2020).

419. Van Opdenbosch, N. & Lamkanfi, M. Caspases in Cell Death, Inflammation, and Disease. *Immunity* vol. **50** 1352–1364 (2019).

420. Patinios, C. *et al.* Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas. *Nucleic Acids Res* **49**, 11392–11404 (2021).

Author affiliations

Amieiro, A.M.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Andres, A.S.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Bardis, A.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Barendregt, A.
Biomolecular Mass Spectrometry and Proteomics, University of Utrecht, 3584 CH Utrecht, The Netherlands
Barendse, P.
Laboratory of Biochemistry, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Boeren, S.
Laboratory of Biochemistry, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Bravo, J.P.K.
Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, USA
Brouns, S.J.J.
Department of Bionanoscience, Delft University of Technology, 2629HZ Delft, The Netherlands.
de Lannoy, C.
Bioinformatics Group, Wageningen University and Research, 6708PB, Wageningen, The Netherlands
Department of Bionanoscience, Delft University of Technology, 2629HZ, Delft, The Netherlands
de Ridder, D.
Bioinformatics Group, Wageningen University and Research, 6708PB, Wageningen, The Netherlands
Ettema, T.J.G.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Fuentenebro-Navas, D.
Laboratory of Biophysics, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Groen, J.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands

Hofstra, L.M.
Department of Medical Microbiology, University Medical Center Utrecht, 3584CX Utrecht, The Netherlands
Keijser, B.J.F.
TNO, 3704HE Zeist, The Netherlands
Köstlbacher, S.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Noordijk, B.
Bioinformatics Group, Wageningen University and Research, 6708PB, Wageningen, The Netherlands
Ossendrijver, M.
TNO, 3704HE Zeist, The Netherlands
Patinios, C.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Prinsen, S.H.P.
Scope Biosciences B.V., 6708WH Wageningen, The Netherlands
Salazar, C.R.P.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Scheltema, R.A.
Biomolecular Mass Spectrometry and Proteomics, University of Utrecht, 3584 CH Utrecht, The Netherlands
Schmid, S.
Laboratory of Biophysics, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Schoen, C.D.
BioInteractions and Plant Health, Wageningen Plant Research, 6708WE Wageningen, The Netherlands
Scholten, B.
Scope Biosciences B.V., 6708WH Wageningen, The Netherlands
Shinkai, A.
RIKEN Cluster for Pioneering Research, Wako, Saitama, Japan
Staals, R.H.J.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Steens, J.A.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Scope Biosciences B.V., 6708WH Wageningen, The Netherlands
Swarts, D.C.
Laboratory of Biochemistry, Wageningen University and Research, 6708WE Wageningen, The Netherlands

Taylor, D.W.
Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, USA
van der Oost, J.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands
van Houte, S.
ESI, Biosciences, University of Exeter, Cornwall Campus, Penryn TR109FE, UK
Watson, B.N.J.
ESI, Biosciences, University of Exeter, Cornwall Campus, Penryn TR109FE, UK
Westra, E.R.
ESI, Biosciences, University of Exeter, Cornwall Campus, Penryn TR109FE, UK
Yildiz, C.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands
Zhu, Y.
Laboratory of Microbiology, Wageningen University and Research, 6708WE Wageningen, The Netherlands

List of publications and patents

Publications

Watson B. N. J., **Steens J. A.**, Staals R. H. J., Westra E. R., van Houte S. Coevolution between bacterial CRISPR-Cas systems and their bacteriophages. ***Cell Host Microbe***. 2021 May 12;29(5):715-725.

Zhou, Y., Bravo, J. P., Taylor, H. N., **Steens, J. A.**, Jackson, R. N., Staals, R. H. J., & Taylor, D.W. (2021). Structure of a type IV CRISPR-Cas ribonucleoprotein complex. ***Iscience***, 24(3), 102201.

Steens, J. A., Zhu, Y., Taylor, D. W., Bravo, J. P., Prinsen, S. H. P., Schoen, C. D., Keijser, B. F., Ossendrijver, M., Hofstra, L. M., Brouns, S. J. J., Shinkai, A., van der Oost, J., Staals, R. H. J. (2021). SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation. ***Nature communications***, 12(1), 5033.

Steens, J. A., Salazar, C. R. P., Staals, R. H. J. (2022). The diverse arsenal of type III CRISPR–Cas-associated CARF and SAVED effectors. ***Biochemical Society Transactions***, 50(5), 1353-1364.

Steens, J. A., van der Oost, J., Staals, R. H. J. (2022). Compact but mighty: Biology and applications of type III-E CRISPR-Cas systems. ***Molecular Cell***, 82(23), 4405-4406.

Steens, J. A., Bravo, J. P. K., Salazar, C. R. P., Yildiz, C., Amieiro, A. M., Köstlbacher, S., Prinsen, S. H. P., Andres, A. S., Patinios, C., Bardis, A., Barendregt, A., Scheltema, R. A., Ettema, T. J. G., van der Oost, J., Taylor, D. W., Staals, R. H. J. (2024) Type III-B CRISPR-Cas cascade of proteolytic cleavages. ***Science*** 383,512-519

Fuentenebro-Navas, D., **Steens, J. A.**, de Lannoy, C, Noordijk, B., de Ridder, D., Staals, R. H. J., Schmid, S. Nanopores reveal the stoichiometry of single oligo-adenylates produced by type III CRISPR-Cas. ***Manuscript in submission***

Steens J. A., Barendse P., Groen J., Boeren, S., van der Oost, J., Swarts, D. C., Staals, R. H. J. CARFish: identifying cOA-binding effector proteins of type III CRISPR-Cas systems. ***Manuscript in preparation***

Steens, J. A., Andres, A. S., Scholten, B., van der Oost, J., Staals, R. H. J. RNAPhi: a viral thermostable RNA polymerase for improved diagnostics. ***Manuscript in preparation***

Patents

Steens, J. A., Prinsen, S. H. P., van der Oost, J., Staals, R. H. J. (Filed 2020) Type III CRISPR/Cas-based diagnostics - WO2020256553

Steens, J. A., Staals, R. H. J., van der Oost, J. (Filed 2023) Thermostable RNA polymerase - WO2023140731

Steens, J. A., van der Oost, J., Staals, R. H. J. (Filed 2023) Method and system of nucleic acid sequence detection

Overview of completed training activities

Discipline specific activities

CRISPR con 2019, Wageningen University, Wageningen (NL)
2019
NWO CHAINS 2020, NWO, (Online)
2020 - oral presentation
Rapid Methods conference 2021, Rapid Methods (Online)
2021 - oral presentation
CRISPR meeting 2021, Pasteur Institute (Online)
2021 - oral presentation
Slush 2019 Startup conference, Slush Helsinki (FI)
2019
WUR Dies Natalis 2020, WUR Wageningen (NL)
2020 - oral presentation
CRISPR meeting Boston 2022, MIT, Boston (US)
2022 - poster presentation
Plant and Animal genome conference, PAG, San Diego (US)
2023 - poster presentation
Secondment, Tor Vergata University, Rome (IT)
2023
Seminar CRISPR-Cas, NAKtuinbouw, Roelofarendsveen (NL)
2023 - oral presentation
CRISPR meeting Germany 2023, CRISPR2023, Wurzburg (DE)
2023 - poster presentation

General courses

VLAG PhD week, VLAG, Baarlo (NL)
2019
NWO GameChanger Venture challenge, NOW (NL)
2020
Techleap Pole Position programme, Techleap (NL)
2022-2023

Assisting in teaching and supervision activities

Applied Molecular Microbiology course
2019-2023
Supervising BSc, MSc and internship students
2019-2023

Other activities

Preparation of Research Proposal, MIB, Wageningen (NL)
2019
iGEM supervisor, WUR iGEM team, Wageningen (NL)
2019-2020
Assisting in peer reviews, MIB, Wageningen (NL)
2019-2023
PhD project CHIC consortium meetings (Online)
2019-2021
Internal MIB meetings, WUR, Wageningen (NL)
2019-2023
PhD trip visiting labs/institutes, WUR, California (US)
2022

About the author

Aiko Jurre Steens was born on the 19th of April 1994 in Amersfoort, The Netherlands. He studied BSc Biotechnology at Wageningen University & Research in Wageningen. He did his BSc thesis in the departement of virology, after which he continued his MSc studies in Medical Biotechnology at Wageningen University & Research. His MSc thesis was about developing affibody phage-display as part of the Wageningen 2017 iGEM project. In this multidisciplinary synthethic biology competition, together with a team of 10 in total, he worked on the creation of a novel cell-based diagnostic application for rare infectious diseases in Africa.

For his intership he went to University of Otago in New Zealand in the McLellan lab to work on novel cancer biomarker antibody developement for use in CAR T-cell therapy. Upon returning to The Netherlands he started a PhD at the Laboratory of Microbiology in the van der Oost group, Bacterial Genetics. During the course of the PhD, the core focus of the research was on type III CRISPR-Cas.

Around the same time as starting a PhD, he founded a company together with 3 friends, Scope Biosciences. There, he still focusses on developing fundamental CRISPR-Cas research towards applications in rapid on-site diagnostics. At the moment of writing, this Wageningen University spin-off company has a total of eight employees.

Acknowledgements

Dear promoters, mentors, supervisors, opponents, paranymphs, colleagues, friends, family and those who happen to stumble across this thesis, I am happy you made it all the way to the end of this thesis book.

I wholeheartedly hope I have expressed my gratitude during our interactions for your contributions, both small and large, to this thesis and my personal (doctoral) journey in a manner that is appropriate. Hence, instead of trying to sum up all the (in)tangible things that I appreciated from our interactions and make you read another couple pages, I encourage you to allow me to tell you in person.

All's well as ends Better

About the cover

Type III CRISPR-Cas systems generally produce cyclic signalling molecules (rings of AMP moieties) to initiate a range of downstream activities in the cell upon target recognitions. Eventually, the effect on the cell is generally death or the induction of a state of dormancy. These ring-shaped signal molecules yield enormous power over the faith of the cell. Hence, the cover is heavily inspired by the cover of the first print of the book Lord of the Rings by J.R.R. Tolkien. Here, the one ring is replaced by a cA_3-ring, as a reference to the type III CRISPR-Cas system characterized in chapter 7. Furthermore, the eye of Sauron in the middle is replaced by a schematical representation of the type III CRISPR-Cas complex, ever watching for invading genetic material in the cell. Graphical elements inspired by the works of Tolkien have been carried throughout the thesis.

The research described in this thesis was financially supported by European Research Council (ERC) Advanced grant ERC-AdG-834279, Horizon-2020-CHIC-760891 and from the Dutch Research Council (NWO) VIDI grant VI.Vidi.203.074.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Printed by ProefschiftMaken.nl