DOUBLE KILL

the DNase and RNase activity of CRISPR-Cas type III systems

Yifan Zhu

PROPOSITIONS

1. The presence or absence of a single protein domain in two homologous multi-subunit complexes of CRISPR-Cas type III results in major functional features.
(this thesis)

2. The dual-function ability of type III-A CRISPR-Cas complexes substantially increases the robustness to neutralize genetic invaders.
(this thesis)

3. Genome editing will not only help us cure human genetic diseases but will also promote the galactic adventure.

4. Doing research and cook food are similar as both of them require creativity and patience.

5. Mastering the skills to quickly fall asleep may have a major impact on private and professional performance.

6. When drinking liquor, tasting is better than boozing, not only for Asian people.

Propositions belonging to the thesis, entitled

'Double kill - the DNase and RNase activity of CRISPR-Cas type III systems'

Yifan Zhu

Wageningen, 3 May 2019

Double kill

the DNase and RNase activity of CRISPR-Cas type III systems

Yifan Zhu

Thesis committee

Promotors

Prof. Dr Willem M. de Vos
Distinguished professor
Wageningen University & Research

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

Co-promotor

Dr Raymond H.J. Staals
Assistant professor, Laboratory of Microbiology
Wageningen University & Research

Other members

Prof. Dr Willem van Berkel, Wageningen University & Research
Dr Gorben Pijlman, Wageningen University & Research
Dr Chirlmin Joo, Delft University of Technology
Dr Peter van Baarlen, Wageningen University & Research

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Argobiotechnology, Nutrition and Health Sciences).

Double kill

the DNase and RNase activity of CRISPR-Cas type III systems

Yifan Zhu

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 3 May 2019 at
4 p.m. in the Aula.

Yifan Zhu

Double kill - the DNase and RNase activity of CRISPR-Cas type III systems,
184 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2019)
With references, with summary in English

ISBN: 978-94-6343-436-2
DOI: 10.18174/471572

Table of Contents

Chapter 1

General introduction and thesis outline

GENERAL INTRODUCTION

Host invader co-evolution

In nature there has always been a continuous arms race between predator and prey. For example, if the predator becomes faster, the prey must also become faster or the prey will eventually die out. The same principle applies to prokaryotes and (bacterio)phages that try to infect them. The prokaryotic hosts employ different strategies to defend against invading viruses [1-6]. And, to counteract, the invading viruses rapidly evolve to overcome this defense barrier [7]. This interplay between prokaryotes and viruses across every ecological niche result in endless cycles of co-evolution [8].

Overview of CRISPR-Cas

In the 20th century, a novel genetic system was discovered in prokaryotes, the CRISPR-Cas system (clustered regularly interspaced short palindromic repeats and CRISPR-associated genes). This system encodes CRISPR RNAs (crRNAs) that contain a guide fragment originating from an invader genome sequence. CrRNAs interact with CRISPR-associated (Cas) proteins forming a CRISPR-Cas ribonucleoprotein (RNP) complex that recognizes and cleaves the complementary invader genome, thus eliminating the invading virus [9-11].

A complete CRISPR-Cas system consists of two components. The first part is the CRISPR locus (also known as CRISPR array) that contains the repeat-spacer sequences preceded by a AT-rich sequence, called the 'leader'. The spacers are mainly acquired from mobile genetic elements (MGEs), but can also come from the host genome sequence at low frequencies [12]. The second component contains the variable *cas* genes that encode the Cas proteins needed for the different steps in CRISPR-Cas-mediated defense, as described below. The latest categorization distinguishes 2 classes, 6 types and around 29 subtypes, based on signature *cas* genes [13]. Although class 2 systems have become increasingly popular over the last years, due to

their use in genome editing applications, this thesis focuses on the type III subtypes within the class 1 systems (Figure 1).

The defense process of the CRISPR-Cas system can be divided into three steps (Figure 2): (I) CRISPR adaptation or acquisition, where the invading nucleic acid is scanned and recognized via at least two dedicated Cas proteins, Cas1 and Cas2, which process and integrate a short invader-derived sequence into one of the CRISPR locus on the genome of the host; (II) expression of the CRISPR array and *cas* genes, crRNA maturation and CRISPR-Cas RNP complex formation. After pre-CRISPR RNA (pre-crRNA) of the CRISPR array is transcribed by RNA polymerase, endoribonucleases, like Cas6, RNase III, Cas12, or Cas13 [14-20], will process the pre-crRNA into small mature crRNA units that harbor a spacer flanked by (parts) of the repeat sequence. This complex will associate with Cas protein(s) to form the RNP complex; (III) CRISPR interference, the RNP complex will use the crRNA as a guide to scan the incoming plasmid or phage nucleic acid or other MGEs for a possible match. Cas proteins will then cleave the nucleic acid after the crRNA matches to the target.

I: CRISPR adaptation

Adaptation (or spacer acquisition) is the step where new MGE fragments are integrated as spacers into the CRISPR array. Cas1 and Cas2 proteins, which are present in almost all CRISPR-Cas systems (Figure 1), are essential for novel spacer adaptation. In the type I system, the Cas1-Cas2 complex acts as an integrase for spacer adaptation [22]. In type II systems, however, in addition to this complex, Cas9, Csn2 and a trans activating crRNA (tracrRNA) are also involved in spacer adaptation [23-25]. Cas4, a Cas protein encoded in several type I subtypes, type II-B/C variant, and type V-B subtype, use its 5’-3’ exonuclease activity to process the protospacer by generating 3’ overhangs to facilitate spacer integration by Cas1-Cas2 [26-30]. A host protein called the integration host factor (IHF)

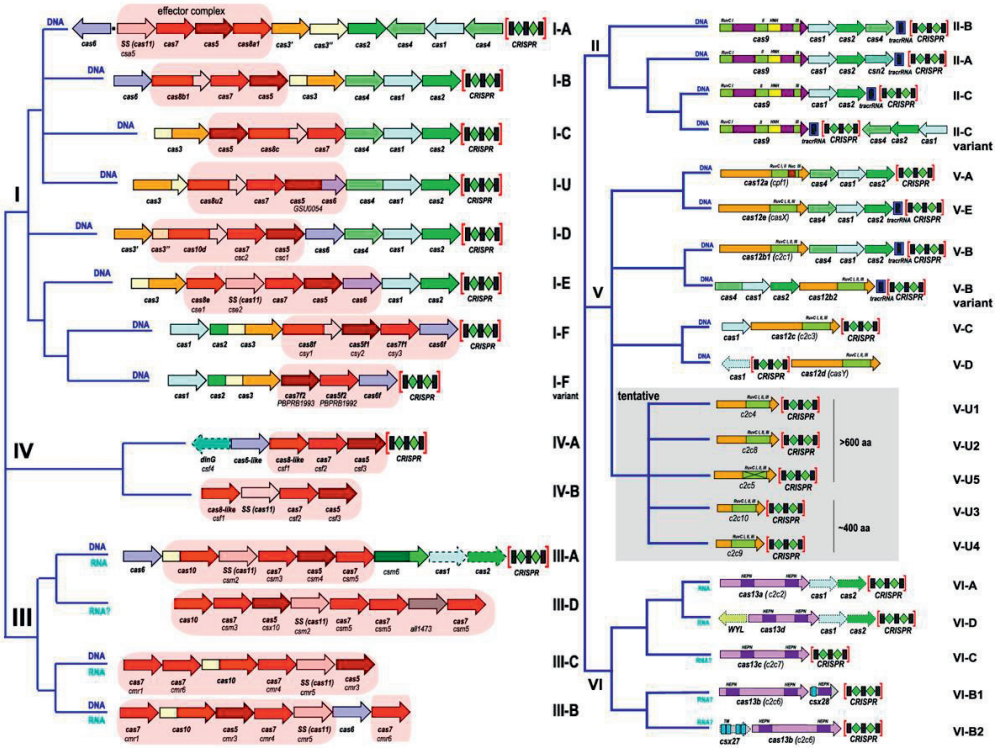


Figure 1. Classification of CRISPR-Cas systems

Schematic representation of the different types and subtypes of class 1 systems, showing the two main components that make up a CRISPR-Cas system: the *cas* genes (arrows) and the CRISPR arrays (cluster of black rectangles and green diamonds). Homologues *cas* genes are color–coded. The adaptation-specific *cas1* and *cas2* genes are almost universal among most (sub)types. Known or predicted (indicated with a question mark) targets (DNA and/or RNA) for the subtypes are indicated at the end branches of the dendrogram. The pink shaded areas represent genes encoding for Cas proteins that are part of the CRISPR RNP complex. (Image adapted from [13])

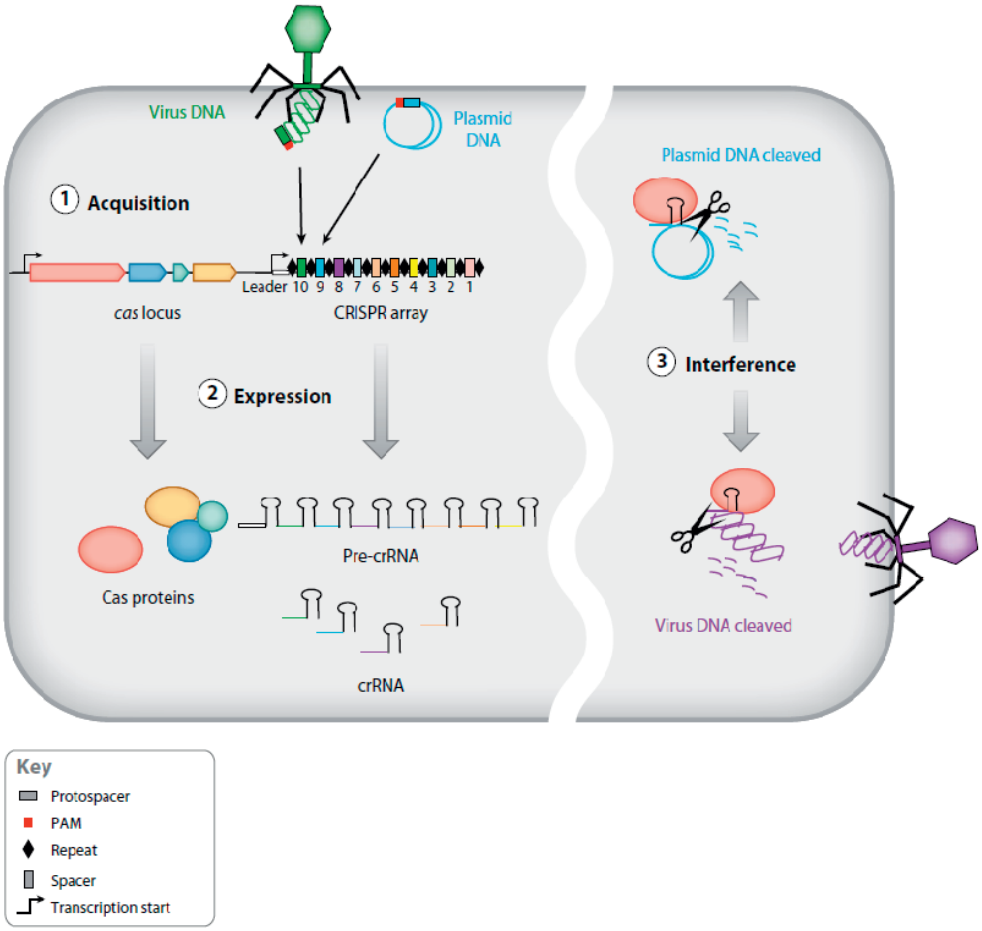


Figure 2. The three steps of CRISPR-Cas-mediated defense.

Three steps are depicted. (1) CRISPR adaptation: virus or plasmid DNA is recognized and processed as a new spacer by Cas proteins (not shown here) and integrated at the leader end of a CRISPR array on the host genome. (2) CRISPR expression: pre-crRNA is transcribed from the CRISPR array by RNA polymerase and cleaved into mature crRNA by the Cas protein(s). The mature crRNA and Cas protein(s) assemble into a CRISPR RNP complex (3) CRISPR interference: the CRISPR RNP complex (red oval) uses the crRNA as a guide to find and cleave complementary invading sequences (phages or plasmid) (Image taken from [21]).

in the type I-E system or a short sequence termed leader-anchoring site (LAS) in type II-A system is involved in the spacer integration as well [31-33]. Acquiring a spacer from the host genome is possible, albeit with low frequency, and often occurs in systems with a defective CRISPR-Cas RNP complex [25, 34]. The reason of low acquisition frequency may be because the replication frequency of MGEs is higher than that of the host genome [35]. The adaptation process mentioned above is called naïve adaptation to acquire novel spacer from a newly encountered invader. In many type I systems, however, there is a second, more rapid process of spacer acquisition, which occurs when there is a pre-existing spacer in the CRISPR array with a (partial) match the invading element, which is called primed adaptation [36-40]. All *cas* genes and the pre-existing spacer are required for primed adaptation. Primed adaptation usually results in the uptake of multiple new spacers selected in close proximity to the original protospacer location (of the pre-existing spacer) and are sometimes biased towards a particular strand depending on the CRISPR-Cas subtype [37-39, 41]. In all cases, primed adaptation creates a positive feedback loop to enhance the defense system to overcome the MGE escape mutants [42, 43].

II: CRISPR expression

Transcription of the CRISPR array is usually initiated within the leader sequence resulting in a long precursor crRNA (pre-crRNA) transcript. Subsequently, each CRISPR-Cas system processes this pre-crRNA into short mature crRNAs via a different mechanism.

Typically, Cas6 family ribonucleases are utilized in class 1 systems to generate mature crRNA by cleaving the pre-crRNA at a specific site within the repeats [44-48]. Two exceptions are the Cas5d in type I-C system [14, 15] and RNase E in subtype III-B variant (III-Bv) [49], because these systems do not encode a Cas6 homologue. In type I systems, Cas6 recognizes the repeat sequence and/or stem-loop structure and then cleaves the RNA directly downstream of the hairpin. A matured crRNA is

composed of a complete spacer with a short repeat-derived 5′ handle and 3′ stem-loop [9, 16]. The overall crRNA processing mechanism in the type III system is similar to type I, although the crRNA undergoes further trimming at the 3′ end via a non-Cas nuclease [50].

In many type II systems, a tracrRNA:pre-crRNA duplex is bound and stabilized by the Cas9 protein and then the housekeeping protein RNase III cleaves the duplex within the repeat [18]. The type V-A complex, Cas12a, has intrinsic RNase activity responsible for crRNA maturation [20]. The pre-crRNA maturation mechanism of type V-B (Cas12b) and type V-C (Cas12c) is still unknown, but it appears that these two proteins recruit tracrRNA to process pre-crRNA, akin to type II systems [51]. Similar to type V-A system, type VI-A system (Cas13a) has dual RNase activity responsible for pre-crRNA processing and target RNA cleavage [19, 52].

III: CRISPR interference

To cleave the MGEs, the RNP complex patrol in the cytoplasm to scout for target sequences (protospacers) complementary to the spacer region in the crRNA. In type I and II systems, only protospacers flanked by the correct PAM will be interrogated [9, 53-56]. Protospacers flanked by an incorrect PAM not be cleaved by the CRISPR RNP complex, even if the protospacer and spacer are completely complementary. Additionally, type I, type II, and type V systems all harbor a region in the crRNA that is sensitive to mismatches between the protospacer and the spacer region of the crRNA. This region is termed the seed sequence and is located in the PAM proximal end of the protospacer [56-64]. The seed sequence in type I system is usually 7 nucleotides (nt), in type II system it is 7 – 12 nt, and extend to around 18 nt in type V systems. Further studies demonstrated that the seed sequence region is also the starting point for initiating base pairing interaction between the protospacer and spacer part of the crRNA [65, 66].

Target interrogation by the CRISPR RNP complex starts by scanning the MGE for the presence of a PAM. After a correct PAM has been found, the

CRISPR RNP complex will check for complementarity between the complex-bound crRNA and the protospacer, starting in the abovementioned seed region. Depending on the CRISPR-Cas (sub)type, the MGE will be cleaved by the intrinsic nuclease activity of the CRISPR-Cas RNP complex (e.g. Cas9 and Cas12) or by recruiting a Cas nuclease. In type I systems for example, the Cas3 nuclease is recruited and will bind to the exposed non-target strand DNA (the strand opposite of the protospacer). Subsequently, the HD domain of Cas3 will nick this strand followed by 3' to 5' exonucleolytic cleavage to fully degrade the target DNA [43, 55, 67-69]. The Cas9 protein of type II systems contains two nuclease domains: HNH and RuvC. The HNH domain cleaves the strand containing the protospacer, whereas the RuvC domain cleaves another strand [59]. The Cas12 RNP complex from the type V system only contains one nuclease domain (RuvC) that cleaves both strands of the dsDNA target [56, 70, 71]. Although most of the mechanisms described here are about CRISPR-Cas systems that cleave DNA, there are also RNA targeting CRISPR-Cas systems (i.e. type III and type VI), which will be described in detail in chapter 2.

THESIS OUTLINE

As a consequence of coevolution, eukarya, bacteria and archaea have developed various strategies to defend themselves against the invasion of transposons, plasmids, and viruses. Unlike the highly sophisticated immune systems in eukarya, most of the defense systems found in bacteria and archaea appear to be much more rudimentary. However, in 2006, a novel defense system in these two domains was predicted and was demonstrated to have more complexity. This immune system, known as CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeat, CRISPR-associated genes), is unique in its adaptability and hereditability. Furthermore, it relies on small-noncoding RNA sequences that mediate the defense. The overall aim of this thesis is to explore the CRISPR-Cas type III system of *Thermus thermophilus* and unravel the mechanistic details of DNA and RNA interference. We are especially interested in the nucleotide substrate cleavage mechanism and how the CRISPR-Cas type III protein-RNA complex can distinguish self from non-self sequences.

Chapter 2 – Introduction to CRISPR-Cas RNA targeting systems

This chapter provides an overview of RNA cleavage mechanisms in RNA targeting CRISPR-Cas systems: the type III RNA and DNA targeting system and the type VI system that exclusively targets RNA. Other aspects of these two systems are discussed as well, such as new spacer acquisition via the reverse-transcriptase-Cas1 fusion protein, non-specific nucleic acid sequence degradation, and possible suicide or signaling route.

Chapter 3 – Structure and activity of the RNA-targeting type III-B CRISPR-Cas complex of *Thermus thermophilus*

Using different biochemical and structural approaches the cleavage mechanism of the type III-B complex from *Thermus thermophilus* (TtCmr) is addressed in chapter 3. Presented data shows that the TtCmr complex cleaves complementary RNA sequences at a fixed distance of 6 nt intervals. Furthermore, the structure of TtCmr is reconstructed by electron

microscopy (EM) data and matched with the predicted stoichiometry of the TtCmr complex. An RNA target cleavage model is proposed, revealing which subunit(s) in the TtCmr complex may be responsible for observed target RNA cleavage.

Chapter 4 – Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning

In former chapter, the stoichiometry and low resolution structure of this multi-subunit complex is presented. Here, the near-atomic resolution cryo-EM structure of the TtCmr complex is studied in the absence and presence of target RNA. Target RNA binding is shown to confer a conformational change in the complex and it is demonstrated that thumb-like β hairpins intercalate between segments of duplexed crRNA:target RNA to facilitate cleavage of the target at 6 nt intervals. Moreover, the architectural similarity between CRISPR-Cas type I and type III suggests a divergent evolution of these two immune systems from a common ancestor.

Chapter 5 – A flexible seed sequence regulates targeting by the type III CRISPR-Cmr complex

Chapter 5 focuses the mechanism of target RNA recognition and binding. We demonstrated that a short area close to the 3' terminal end of the crRNA is used to initiate binding to target RNAs, which we defined as the seed region. Base pairing of this region between crRNA and target RNA will trigger conformational changes in the complex to propagate base pairing interactions, ultimately resulting in the endoribonucleolytic degradation of the target RNA. The high resolution structural data also supports this conclusion since the seed-like sequence resides in a region in the complex that is more exposed when compared to the rest of the crRNA.

Chapter 6 – RNA Targeting by the CRISPR-Cas type III-A Csm Complex of *Thermus thermophilus*

Previous studies of type III-A Csm complexes revealed the *in vivo* DNA cleavage ability, meanwhile no *in vitro* DNA cleavage activity of the type III-A Csm complex has been reported. In this chapter, we investigated the structural and biochemical properties of the endogenous type III-A TtCsm complex. The results demonstrate that TtCsm exhibits endoribonuclease activity using a cleavage mechanism similar to the CRISPR-Cas type III-B Cmr complex. In addition, unlike type I and type III TtCmr complexes, interference by TtCsm does not proceed via initial base pairing by a seed sequence at the 5' end of the crRNA.

Chapter 7 – Biochemical and structural analysis of DNA targeting by the *Thermus thermophilus* CRISPR-Cas type III-A system

In chapter 7 we looked deeper into how type III-A TtCsm complexes use a target RNA to stimulate DNase activity. Our results indicate that DNA will be targeted after a crRNA-target RNA duplex is formed in the TtCsm complex. We show that this DNase activity is biased towards cleaving thymidine residues in the DNA target. Furthermore, this endonucleolytic DNase activity can be controlled by base pairing interactions between the 5' repeat region of the crRNA and 3' sequences of the protospacer on the target RNA. Cryo-EM analysis reveals structural similarity between DNA nucleases from type I and III-A TtCsm effector complexes.

REFERENCES

1. Swarts, D.C., et al., *DNA-guided DNA interference by a prokaryotic Argonaute.* Nature, 2014. **507**(7491): p. 258-261.
2. Goldfarb, T., et al., *BREX is a novel phage resistance system widespread in microbial genomes.* The EMBO Journal, 2015. **34**(2): p. 169-183.
3. Aguado, L.C., et al., *RNase III nucleases from diverse kingdoms serve as antiviral effectors.* Nature, 2017. **547**(7661): p. 114-117.
4. Harms, A., et al., *Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology*. Molecular Cell, 2017. **70**(5): p. 768-784.
5. Ofir, G., et al., *DISARM is a widespread bacterial defence system with broad anti-phage activities.* Nat Microbiol, 2017. **3**(1): p. 90-98.
6. Doron, S., et al., *Systematic discovery of antiphage defense systems in the microbial pangenome.* Science, 2018. **359**(6379): p. eaar4120.
7. Labrie, S.J., J.E. Samson, and S. Moineau, *Bacteriophage resistance mechanisms.* Nat Rev Microbiol, 2010. **8**(5): p. 317-327.
8. Samson, J.E., et al., *Revenge of the phages: defeating bacterial defences.* Nat Rev Microbiol, 2013. **11**(10): p. 675-687.
9. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-964.
10. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA.* Science, 2008. **322**(5909): p. 1843-1845.
11. Mohanraju, P., et al., *Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems.* Science, 2016. **353**(6299): p. aad5147.
12. Stern, A., et al., *Self-targeting by CRISPR: gene regulation or autoimmunity?* Trends Genet, 2010. **26**(8): p. 335-340.
13. Makarova, K.S., Y.I. Wolf, and E.V. Koonin, *Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?* The CRISPR Journal, 2018. **1**(5): p. 325-336.
14. Garside, E.L., et al., *Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases.* RNA, 2012. **18**(11): p. 2020-2028.
15. Nam, K.H., et al., *Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system.* Structure, 2012. **20**(9): p. 1574-1584.
16. Carte, J., et al., *Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes.* Genes Dev, 2008. **22**(24): p. 3489-3496.
17. Sefcikova, J., et al., *Cas6 processes tight and relaxed repeat RNA via multiple mechanisms: A hypothesis.* Bioessays, 2017. **39**(6).
18. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-607.
19. East-Seletsky, A., et al., *Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection.* Nature, 2016. **538**(7624): p. 270–273.
20. Fonfara, I., et al., *The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA.* Nature, 2016. **532**(7600): p. 517-521.
21. Bhaya, D., M. Davison, and R. Barrangou, *CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation.* Annu Rev Genet, 2011. **45**: p. 273-297.
22. Nuñez, J.K., et al., *Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity.* Nat Struct Mol Biol, 2014. **21**(6): p. 528-534.
23. Heler, R., et al., *Cas9 specifies functional viral targets during CRISPR-Cas adaptation.* Nature, 2015. **519**(7542): p. 199-202.
24. Heler, R., et al., *Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR-Cas Immune Response.* Mol Cell, 2017. **65**(1): p. 168-175.
25. Wei, Y., R.M. Terns, and M.P. Terns, *Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation.* Genes Dev, 2015. **29**(4): p. 356-361.
26. Zhang, J., T. Kasciukovic, and M.F. White, *The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster.* PLoS One, 2012. **7**(10): p. e47232.
27. Krupovic, M., et al., *Evolution of an archaeal virus nucleocapsid protein from the CRISPR-associated Cas4 nuclease.* Biol Direct, 2015. **10**(1): p. 65.
28. Kieper, S.N., et al., *Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation*. Cell Rep, 2018. **22**(13): p. 3377-3384.
29. Lee, H., et al., *Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays.* Molecular Cell, 2018. **70**(1): p. 48-59.
30. Shiimori, M., et al., *Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci.* Molecular Cell, 2018. **70**(5): p. 814-824.e6.
31. Nuñez, J.K., et al., *CRISPR Immunological Memory Requires a Host Factor for Specificity.* Molecular Cell, 2016. **62**(6): p. 824-833.
32. Yoganand, K.N., et al., *Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system.* Nucleic Acids Res, 2016. **45**(1): p. 367-381.
33. Wright, A.V. and J.A. Doudna, *Protecting genome integrity during CRISPR immune adaptation.* Nat Struct Mol Biol, 2016. **23**(10): p. 876-883.

34. Yosef, I., M.G. Goren, and U. Qimron, *Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli.* Nucleic Acids Res, 2012. **40**(12): p. 5569-5576.
35. Levy, A., et al., *CRISPR adaptation biases explain preference for acquisition of foreign DNA.* Nature, 2015. **520**(7548): p. 505-510.
36. Datsenko, K.A., et al., *Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system.* Nat Commun, 2012. **3**: p. 945.
37. Li, M., et al., *Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process.* Nucleic Acids Res, 2014. **42**(4): p. 2483-2492.
38. Richter, C., et al., *Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer.* Nucleic Acids Res, 2014. **42**(13): p. 8516-8526.
39. Swarts, D.C., et al., *CRISPR interference directs strand specific spacer acquisition.* PLoS One, 2012. **7**(4): p. e35888.
40. Vorontsova, D., et al., *Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery.* Nucleic Acids Res, 2015. **43**(22): p. 10848-10860.
41. Shmakov, S., et al., *Pervasive generation of oppositely oriented spacers during CRISPR adaptation.* Nucleic Acids Res, 2014. **42**(9): p. 5907-5916.
42. Staals, R.H., et al., *Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system.* Nat Commun, 2016. **7**: p. 12853.
43. Künne, T., et al., *Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation.* Molecular Cell, 2016. **63**(5): p. 852-864.
44. Gesner, E.M., et al., *Recognition and maturation of effector RNAs in a CRISPR interference pathway.* Nat Struct Mol Biol, 2011. **18**(6): p. 688-692.
45. Haurwitz, R.E., et al., *Sequence- and structure-specific RNA processing by a CRISPR endonuclease.* Science, 2010. **329**(5997): p. 1355-1358.
46. Sashital, D.G., M. Jinek, and J.A. Doudna, *An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3.* Nat Struct Mol Biol, 2011. **18**(6): p. 680-687.
47. Sokolowski, R.D., S. Graham, and M.F. White, *Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system.* Nucleic Acids Res, 2014. **42**(10): p. 6532-6541.
48. Ozcan, A., et al., *Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum.* Nat Microbiol, 2019. **4**(1): p. 89-96.
49. Behler, J., et al., *The host-encoded RNase E endonuclease as the crRNA maturation enzyme in a CRISPR-Cas subtype III-Bv system.* Nat Microbiol, 2018. **3**(3): p. 367-377.
50. Walker, F.C., et al., *Molecular determinants for CRISPR RNA maturation in the Cas10–Csm complex and roles for non-Cas nucleases.* Nucleic Acids Res, 2016. **45**(4): p. 2112-2123.
51. Shmakov, S., et al., *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems.* Molecular Cell, 2015. **60**(3): p. 385-397.
52. Abudayyeh, O.O., et al., *C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector.* Science, 2016. **353**(6299): p. aaf5573.
53. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-1712.
54. Garneau, J.E., et al., *The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA.* Nature, 2010. **468**(7320): p. 67-71.
55. Westra, E.R., et al., *CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3.* Molecular Cell, 2012. **46**(5): p. 595-605.
56. Zetsche, B., et al., *Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System.* Cell, 2015. **163**(3): p. 759-771.
57. 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, 2011. **108**(25): p. 10098-10103.
58. Wiedenheft, B., et al., *RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions.* Proc Natl Acad Sci U S A, 2011b. **108**(25): p. 10092-10097.
59. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-821.
60. Künne, T., D.C. Swarts, and S.J.J. Brouns, *Planting the seed: target recognition of short guide RNAs.* Trends Microbiol, 2014. **22**(2): p. 74-83.
61. Kim, Y.K., Y.G. Kim, and B.H. Oh, *Crystal structure and nucleic acid-binding activity of the CRISPR-associated protein Csx1 of Pyrococcus furiosus.* Proteins, 2013. **81**(2): p. 261-270.
62. Kim, D., et al., *Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells.* Nat Biotechnol, 2016. **34**(8): p. 863-868.
63. Kleinstiver, B.P., et al., *Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells.* Nat Biotechnol, 2016. **34**(8): p. 869-874.
64. Liu, L., et al., *C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism.* Molecular Cell, 2017. **65**(2): p. 310-322.
65. Rutkauskas, M., et al., *Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection.* Cell Rep, 2015. **10**(9): p. 1534-1543.
66. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.* Nature, 2014. **507**(7490): p. 62-67.
67. Sinkunas, T., et al., *Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system.* EMBO j, 2011. **30**(7): p. 1335-1342.
68. Mulepati, S. and S. Bailey, *In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target.* J Biol Chem, 2013. **288**(31): p. 22184-22192.

69. Huo, Y., et al., *Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation.* Nat Struct Mol Biol, 2014. **21**(9): p. 771-777.
70. Yang, H., et al., *PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease.* Cell, 2016. **167**(7): p. 1814-1828.e12.
71. Swarts, D.C., J. van der Oost, and M. Jinek, *Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a.* Molecular Cell, 2017. **66**(2): p. 221-233.e4.

Chapter 2

Shooting the messenger: RNA-targeting CRISPR-Cas systems

This Chapter has been published as:

Zhu, Y., Klompe, S.E., Vlot, M., van der Oost, J., Staals, R.H.J. (2018). Shooting the messenger: RNA-targeting CRISPR-Cas systems. Biosci Rep. 38(3)

ABSTRACT

Since the discovery of CRISPR-Cas immune systems astonishing progress has been made on revealing their mechanistic foundations. Due to the immense potential as genome engineering tools, research has mainly focused on a subset of Cas nucleases that target DNA. In addition, however, distinct types of RNA-targeting CRISPR-Cas systems have been identified. The focus of this review will be on the interference mechanisms of the RNA targeting type III and type VI CRISPR-Cas systems, their biological relevance and their potential for applications.

INTRODUCTION TO CRISPR-Cas

CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats, CRISPR-associated genes) is a prokaryotic adaptive immune system that is present in about half of the currently available bacterial and archaeal genomes. It provides sequence-specific defense against mobile genetic elements (MGEs) such as bacteriophages and conjugative plasmids, but also plays a role in host gene regulation [1-6].

The mechanism of adaptive immunity by CRISPR-Cas systems can be divided into three different phases: adaptation, expression/processing, and interference. The defense is initiated by a process called 'adaptation' or 'spacer acquisition' during which short, MGE-derived DNA sequences, called spacers, are stored in repetitive loci on the host chromosome. The spacers are separated by repetitive DNA sequences called repeats, collectively forming the CRISPR array. These CRISPRs function as a genetic memory storage that allows the host to recognize future invasions by previously encountered MGEs. In the second phase a CRISPR array is transcribed into a long RNA transcript that is subsequently processed into multiple mature CRISPR RNAs (crRNAs) [7-9]. In addition, Cas proteins are encoded by *cas* genes that are generally located in close proximity to the CRISPR array. Together the crRNA and Cas protein(s) form a ribonucleoprotein (RNP) complex that will patrol the cell. In the interference phase, these RNPs use complementarity to the crRNA to detect recurring nucleic acid invaders and utilize the nuclease activity of the Cas proteins to degrade the invading nucleic acids [10].

A very broad range of CRISPR-Cas systems has been discovered to date, resulting in a classification system and a common nomenclature for the associated *cas* genes. The latest classification divides these systems into two classes [11-13], each of which are further divided into three different types and numerous subtypes based on signature *cas* genes. Class 1 systems (encompassing type I, III, and the putative type IV systems) utilize

multisubunit RNP complexes, while class 2 systems (encompassing the type II, V and VI systems) utilize single protein RNP complexes.

Even though every CRISPR-Cas system has its own characteristics, most rely on direct targeting of invading DNA. Two CRISPR-Cas types, however, have been shown to differ from this standard by using RNA as their bona-fide target (Fig. 1). Many recent reviews have focused on DNA targeting CRISPR-Cas systems. Instead, this review will discuss the state of the art of the two RNA-targeting CRISPR-Cas types, type III and VI. Their interference mechanisms and fundamental biology will be discussed.

CRISPR-Cas TYPE III SYSTEMS

Type III systems belong to the class 1 CRISPR-Cas systems, and as such, their RNP complexes are composed of 4-6 different Cas proteins (Fig. 2). Even though there are subtype-specific variations between different type III RNP complexes, the general structure is largely conserved. The general structure is comprised of two intertwined helical protein filaments, the backbone of which is formed by multiple copies of the Cas7 protein and multiple copies of the small subunit Cas11 (Csm2 or Cmr5) [14-17]. The signature protein Cas10, together with Cas5, caps one end of the RNP, whereas different variants of the Cas7 protein cap the other end of the structure. In addition, the RNP contains a mature crRNA that comprises a general repeat-derived sequence called the 5'-handle and an unique spacer-derived sequence. The 5'-handle is tightly anchored within the Cas10/Cas5 cap while the spacer-derived sequence of the crRNA spans the backbone of the RNP [18-22].

Type III CRISPR-Cas systems are divided into four subtypes (subtype III-A to III-D). Subtypes III-A and III-B (Csm and Cmr) have been described first, while the less frequently occurring variant subtypes III-C and III-D systems have subsequently been discovered. The signature subunit Cas10 typically contains a HD-type nuclease domain and a GGDD-type cyclase domain

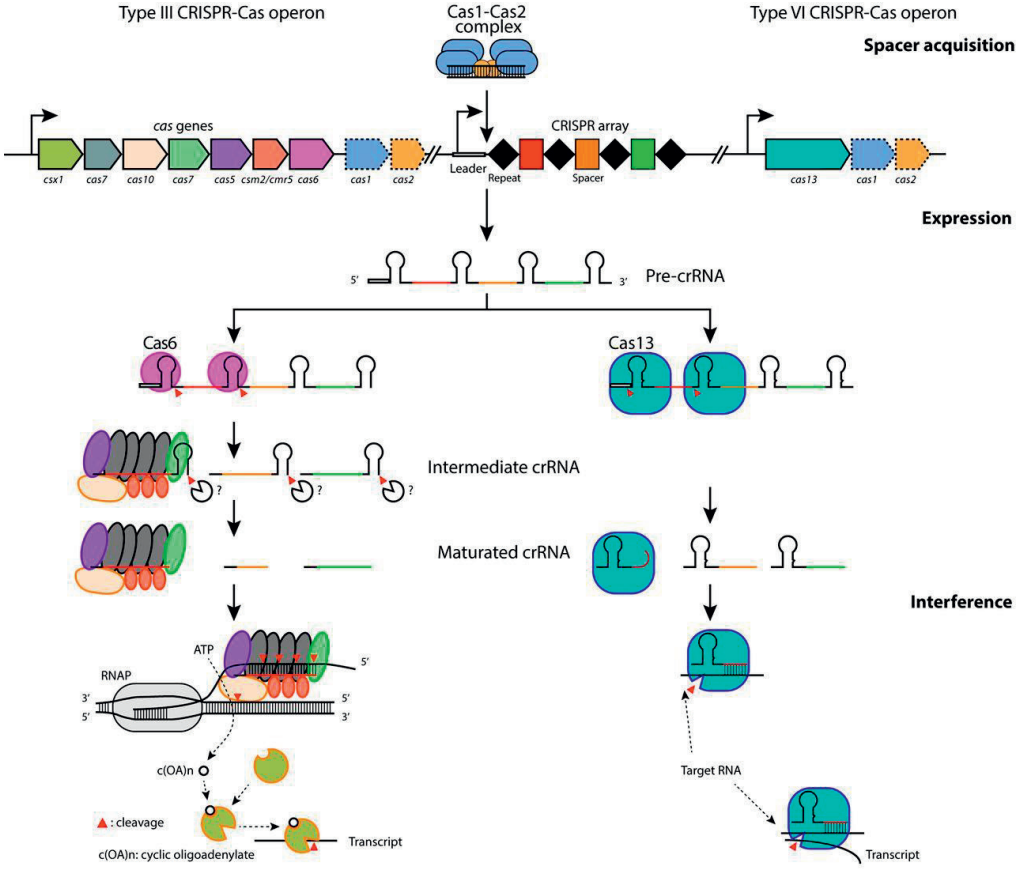


Figure. 1. Overview of the three phases of type III and type VI CRISPR-Cas systems.

Spacer acquisition, upon entry of a mobile genetic element (in this case, a bacteriophage genome), the Cas1 (teal) and Cas2 (orange) proteins select and process a region on the invading DNA to generate a spacer, which will be integrated at the leader-first repeat junction of the CRISPR array, consisting of repeats (black diamonds) and pre-existing spacers (multiple colors). Expression, the CRISPR locus is transcribed into a long pre-crRNA transcript. The Cas6 endoribonuclease (light pink) and Cas13 (light blue) proteins cleave the pre-crRNA at fixed positions down- (Cas6) or upstream (Cas13) of the stem-loop structure (formed by the palindromic nature of the repeats). In type III systems, the 3' end of the intermediate crRNA is further processed by an unknown mechanism. The mature crRNAs assemble with Cas protein(s) to form a functional RNP complex. Interference, the RNP complex scans transcripts for a complementary RNA target, after which the RNA target is degraded by the Cas7 nuclease (in type III) or Cas13 (in type VI). In type III systems, base pairing between the crRNA and the target RNA will activate the HD and palm domain of Cas10 for ssDNA cleavage and c(OA)s biosynthesis respectively. c(OA)s will play a role as a second messenger to trigger the sequence non-specific RNA cleavage activity of Csx1 (olive). In type VI systems, sequence non-specific RNA cleavage is conferred by the HEPN domain in Cas13 after target RNA binding.

(referred to as the palm domain). The amino acid sequence of the Cas10 subunit of subtype III-C differs substantially from that of the other type III subtypes, probably reflecting an inactive cyclase-like domain [12]. Subtype III-D systems usually harbor a Cas10 that lacks the HD domain.

Spacer adaptation in type III systems

Cas1 and Cas2 proteins are widely conserved among CRISPR-Cas systems and were found to be essential for spacer acquisition [23]. Surprisingly, not all type III systems harbor these genes [12]. The lack of these genes is likely complemented by other CRISPR-Cas systems present in the genome. As such, type III systems are expected to use Cas1 and Cas2 from other systems *in trans* and have been shown to utilize mature crRNAs derived from other CRISPR arrays [11, 12, 24].

A well-conserved feature of all CRISPR-Cas systems is that a complex of Cas1 and Cas2, sometimes complemented with additional proteins, is responsible for the selection of protospacers in invading genomes and the subsequent integration as new spacers in the CRISPR array in the host chromosome. A unique additional mechanism for spacer acquisition by some type III systems was recently discovered. In these specific systems, Cas1 was found to be fused to a reverse transcriptase domain (RT) [25, 26]. Since RT enzymes can generate DNA from an RNA template, an RT-Cas1 fusion suggested a possible function for spacer adaptation from RNA. Indeed, this fusion has recently been demonstrated to be capable of spacer acquisition from RNA while retaining its ability to acquire spacers from DNA [25].

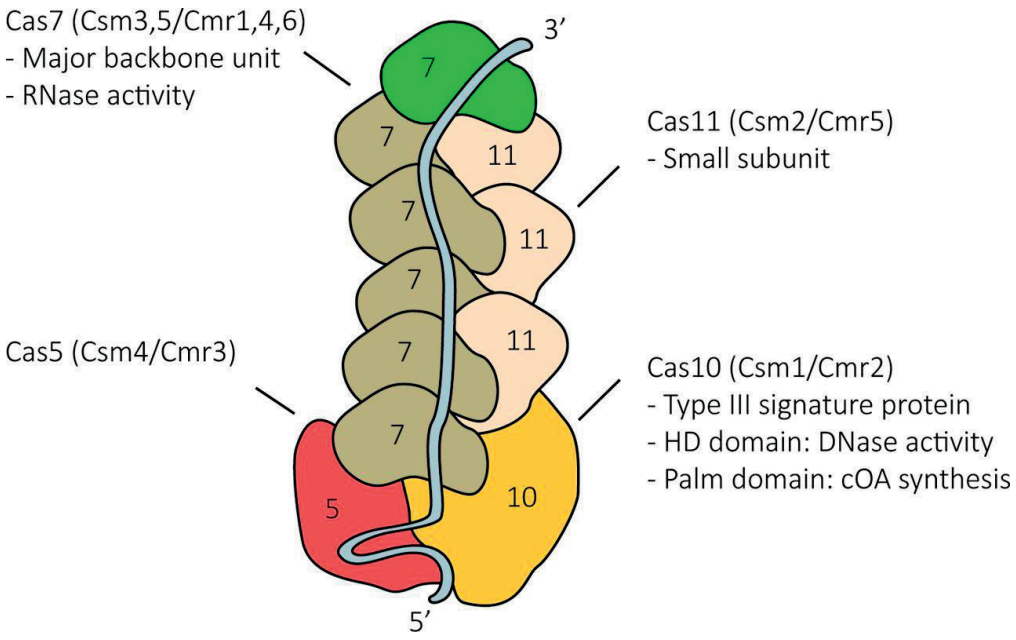


Figure. 2. Structural arrangement of a typical type III CRISPR-Cas complex.

Homologous subunits are depicted by the same color (old nomenclature is mentioned in brackets).

Nuclease activities of type III systems

Currently, type III systems are the only CRISPR-Cas systems known to utilize three different nuclease activities for invader neutralization. For many years, various targets have been identified for different CRISPR-Cas type III subtypes. Initial *in vivo* experiments with the type III-A Csm-crRNA complex of *Staphylococcus epidermidis* indicated that immunity relied on degradation of target DNA [3]. This was quickly followed up with *in vitro* experiments with the Cmr complex of the type III-B Cmr in *Pyrococcus furiosus*, which revealed that not DNA but RNA was specifically degraded [27]. This apparent conundrum came to an end when it became clear that a general type III mechanism is based on transcription-dependent (specific RNA binding and cleavage) and subsequent (non-specific) targeting of DNA [28, 29]. Interestingly, additional non-specific RNase activity by type III systems was recently demonstrated to occur through allosteric activation of

a hyperactive Cas ribonuclease [30, 31]. In this section these three individual nuclease activities will be discussed.

Specific RNA cleavage

The only sequence-specific degradation facilitated by type III systems is that of the target RNA (containing the protospacer sequence complementary to the crRNA guide). Binding of the target RNA to the crRNA of the effector complex positions the target along the Cas7 subunits forming the RNP backbone [16, 32, 33]. Structural data revealed that a conserved β-hairpin present in Cas7 (Csm3 or Cmr4) disrupts the RNA duplex at 6 nt intervals [34]. This steric conflict results in the outward 'flipping' of nucleotides, positioning them for cleavage facilitated by a conserved aspartate residue which is also present in Cas7 [16, 34]. Since Cas7 is present in multiple copies, this results in multiple cleavage sites within the targeted RNA sequence. In accordance with the RNP conformation these cut-sites are separated by 6nt intervals and are at fixed distances from the crRNA 5' handle [16, 21, 22, 32, 35]. This 'ruler' mechanism is found in all different type III systems [14, 16, 20, 34, 36, 37].

Interestingly the binding and degradation of the target RNA has been reported to be relatively flexible when it comes to mutations within the protospacer sequence [14, 22, 35, 38]. This low stringency implies that MGE mismatch mutants are still being recognized and neutralized, and hence that escape through mutagenesis is less likely. It is possible that the host can endure the occasional off-target degradation of RNA since the effects would not be as disastrous as for off-target DNA damage. Until now, no self-discrimination mechanism has been found regarding RNA cleavage activity nor has the exact seed region been defined. Complementarity between the conserved 5' handle of the crRNA and the 3' flanking region of the target RNA (which indicates self-targeting) does not abrogate RNA cleavage [22, 36]. Although a potential risk for self-targeting could be RNA generated through transcription of the CRISPR array, it is not expected to

be much of a problem since transcription of the CRISPR array typically only occurs in the sense direction, producing RNAs that are identical but not complementary to the crRNA.

Non-specific ssDNA cleavage

In addition to specific target RNA degradation, type III systems were found to facilitate immunity through DNA cleavage [3]. A major break-through was the determination that DNA cleavage relies on transcription of the protospacer sequence (Fig. 3). On top of that, immunity was only detected when transcription occurred antisense to the protospacer [28, 36]. The directional restriction of transcription strongly suggested that a protospacer-containing RNA strand is required for DNA degradation. Several subsequent *in vitro* studies indeed showed that binding of the RNP to a target RNA promotes DNA cleavage [17, 35, 36]. What came as surprise was that RNA-induced activation resulted in degradation of both target and non-target DNA sequences [17, 35, 39]. This indicated that the DNase activity of type III systems is non-specific [35].

Mutational analysis showed that the HD domain of Cas10 is the metal-dependent DNase responsible for this activity [17, 35]. Additional *in vitro* work determined that the nuclease activity preferably targets single-stranded DNA, implicating that transcription is required for the production of target RNA as well as for presenting the DNA in a single-stranded form [17]. Upon transcription, RNA polymerase opens the DNA double-helix and uses the template strand to generate an RNA transcript. Meanwhile, the non-template strand is expected to be accessible on the outside of the transcription machinery [40]. This would be in accordance with *in vivo* data revealing that the non-template strand was cleaved and further degraded while the template strand remained intact [36].

Furthermore, it was shown that three nucleotides complementarity (rPAM) between the crRNA 5’ handle and the 3’ region flanking the RNA protospacer inhibits DNase activity, thereby generating a mechanism to avoid targeting

of the host CRISPR array [17, 39]. However, this system does not save the host from possible collateral DNA cleavage once activated by a bona-fide (invader-derived) target RNA. This issue might be resolved by the spatial-temporal restriction of the DNase activity. Since activation of Cas10 requires the RNP to be bound to the RNA target, the complex is locally restricted to the vicinity of invader RNA transcripts. Moreover, it has been shown that upon cleavage of the RNA target, the cleavage products are released, potentially returning Cas10 to a non-active state [17]. This hypothesis is supported by the detection of increased DNA cleavage upon inactivation of the RNase activity (by mutating the active site of Cas7). Inactivation of the RNase activity causes the target RNA to remain bound to the complex, thereby keeping the Cas10 subunit in an activated state [35].

Non-specific RNA degradation

In addition to the HD domain mentioned before, Cas10 also contains a palm domain (a domain that has been implicated with cyclase activity in other proteins) harboring a conserved GGDD motif. Although the function of the palm domain in Cas10 has long been enigmatic, recent work has shown that it is involved in the conversion of ATP into a range of cyclic oligoadenylates (c(OA)s) in a metal-dependent way [30, 31]. The synthesis of c(OA)s by Cas10 is not constitutive, but requires constant activation by binding of an RNA target to the RNP, and requires non-complementarity between the crRNA 5’ handle and its cognate target RNA [30, 31]. It is interesting to note that both the HD-dependent DNase activity and the GGDD-dependent (palm domain) c(OA)s synthesis of Cas10 depend on the non-complementarity of the 5'-handle. This suggests that a common conformational change upon proper target recognition may activate both active sites.

CRISPR-Cas type III systems often harbor genes that encode Csm6 or Csx1 family members [30, 41-43]. These homologous proteins contain a C-terminal higher eukaryotes and prokaryotes nucleotide-binding (HEPN)

domain and an N-terminal CRISPR-associated Rossman fold (CARF) domain [31, 41, 44]. A wide range of proteins that belong to the HEPN super family have been shown to harbor HEPN-based, metal-independent RNase activity, suggesting that Csm6 or Csx1 also exercise this function [41]. However, neither of the proteins seem to stably associate with the type III effector complex or the adaptation machinery [19, 45]. It has long been proposed that the CARF domains of Csm6/Csx1 function as specific sensors for an activation signal [41]. Intriguingly, it was shown that the CARF domain of the homodimeric Csm6 complex can bind the second messengers (c(OA)s) produced by the palm domain of Cas10. Binding of the c(OA)s to Csm6 or

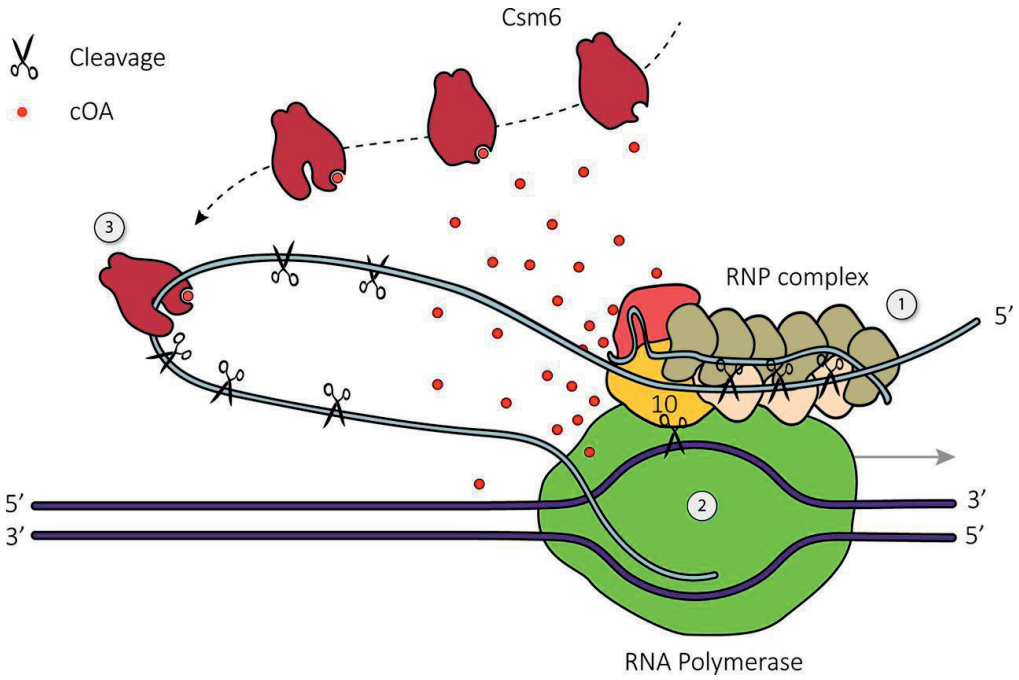


Figure. 3. Model for Type III RNA and DNA targeting.

(1) Complementarity between the crRNA and target RNA induces RNA cleavages at fixed 6 nt intervals by the multiple copies of Cas7 present in the RNP backbone. (2) Complementarity between the crRNA and target RNA in combination with non-complementarity between the 5′ crRNA handle and the sequence 3′ of the protospacer region of the target RNA result in the activation of the HD and palm domains of Cas10. The HD domain of Cas10 non-specifically degrades single-stranded DNA that is likely to be available in the transcription bubble caused by RNA polymerase (green) during transcription. The Cas10 palm domain converts ATP to cyclic oligoadenylates (c(OA)s) that function as diffusible second messengers. (3) c(OA)s bind to the CARF domains of the Csm6 homodimer. This interaction allosterically activates the sequence non-specific ssRNase activity of its HEPN domains

Csx1 has been shown to trigger the allosteric activation of its RNase activity [30, 31]. This means that recognition of an invader by the type III effector complex not only activates the non-specific nuclease activity by Cas10, but also induces c(OA)s generation by Cas10, which in turn allosterically activates the non-specific ssRNA activity of Csm6. It was determined that *Streptococcus thermophilus* Csm6 (StCsm6) preferentially cleaves at GA or AA. However, the presence of c(OA)s seemed to reduce this apparent preference [30].

Cas10 homologs can produce a range of c(OA)s. For example, StCsm6 is activated by cA_6 while *Thermus thermophilus* Csm6 (TtCsm6) is activated by cA_4 [30, 31]. This indicates that c(OA)s are second messengers common to type III CRISPR-Cas systems, even though the exact nature of the messenger molecule may differ between systems [30]. Notably, Csm6 and Csx1 proteins from different systems have been shown to exercise RNase activity in the absence of c(OA)s [43-45]. Since this RNase activity is non-specific, this is potentially detrimental for the host. However, the addition of c(OA)s increases the RNase activity dramatically. It is very well possible that the host can withstand the low background RNase activity in the absence of c(OA)s, or that conditions *in vivo* somehow quench this background activity [30, 41].

One of the questions that remain to be answered for c(OA)s-induced RNase activity is if, and how, this activity is restrained to invader transcripts. c(OA)s synthesis is regulated through Cas10 activation, and binding of an RNA target induces c(OA)s synthesis while target cleavage by Cas7 terminates c(OA)s synthesis. It is possible that Csm6/Csx1 activity is spatially restricted to the area of c(OA)s production, and that it is temporally restricted by the half-life of c(OA)s-activated Csm6/Csx1 but additional data is required to confirm this hypothesis. Another possibility is that Csm6/Csx1 is not restricted to intruder RNA at all and that it works as a suicide-switch that protects the population, rather than the individual cell, from infection.

Type III immunity

Type III systems are the only CRISPR-Cas systems that evolved three entirely different nuclease activities. The question remains whether these different nuclease activities are all required to provide efficient protection against invaders. In this section we will speculate on the origins of type III immunity.

Immunity primarily against RNA or DNA invaders?

CRISPR systems that adopt two different RNase activities and show spacer adaptation from RNA targets might suggest that RNA invaders are the primary target. The RNase activity of Csm6/Csx1 could ensure complete clearance of the invading RNA, while the Cas7 could facilitate temporal control of this activity. In this scenario, Cas7-mediated RNA cleavage is followed by the release of the RNA degradation fragments from the complex, thereby down-regulating the second messenger (c(OA)s) synthesis. The DNase activity could be accounted for by the fact that some RNA phages replicate through a DNA intermediate. The fact that type III systems can indeed convey resistance to RNA phages supports this hypothesis [22]. A counter argument to this theory is that the CRISPR arrays associated with at least some type III system seem to match DNA phages, rather than RNA phages [17, 43]. Additionally, RNA phages have a notoriously high mutation rate which would require swift spacer adaptation by the CRISPR system in order to provide immunity [43].

Distinguishing lytic from lysogenic infections

One aspect of CRISPR-Cas immunity that is often overlooked is that it can target any intruding MGE, regardless of whether the cargo is harmful or beneficial. Temperate phages can proliferate through two different cycles, a lytic cycle that includes avid expression of its genetic information and results in cell death, and a lysogenic cycle in which the viral DNA is incorporated in the host genome and relies on host proliferation for replication. While the lytic cycle is deadly for the host and should be

combated, the lysogenic cycle allows for potentially beneficial genes to be incorporated in the host genome [46, 47]. Targeting of these lysogenic infections by CRISPR-Cas could therefore be disadvantageous for the host rather than advantageous [48]. Using its three different nuclease activities, type III systems can tolerate lysogenic infections while responding quickly to lytic onsets, thereby offering the perfect compromise between defense and genetic availability [29, 48]. Type III systems that target genes expressed early in the lytic cycle were reported to neutralize the invader using the DNase activity of Cas10 alone [43]. In contrast, targeting of late-expressed genes require the RNase activity of Csm6 to prevent the accumulation of phage RNA [43, 49]. This is in accordance with data showing that type III systems can battle lytic infections regardless of whether it targets early or late expressed genes [29]. The RNase activity of Cas7 does not seem to be required for immunity but might function as a switch to lower the DNase activity of Cas10 after releasing the cleaved target RNA from the complex [19, 36].

Escape mutant targeting

Type III systems are surprisingly flexible regarding mutations in the target sequence [22, 35, 38]. DNase activity alone is enough to facilitate immunity provided there is full complementarity between the crRNA guide and its RNA target [43]. However, upon introduction of multiple mutations within the protospacer, the RNase activity of Csm6 is required to provide immunity. This suggests that the Cas10 DNase activity is less efficient when the binding affinity between the crRNA and the target RNA is reduced. Csm6/Csx1 is therefore required to prevent accumulation of viral RNA, thereby buying time for the DNase activity of Cas10. This assumption is supported by the fact that a specific type III system not only uses the crRNAs of a co-existing CRISPR-Cas type I-F system, but is also able to neutralize escape mutants from that particular system [24].

Possible suicide or signaling route

Considering that type III systems rely on two sequence-independent nucleases opens up the possibility to use it as a 'suicide-switch' or whether it is necessary to cope a high concentration of invader RNA/DNA. It does not seem likely that this is the natural purpose of the system since Cas10 activity is switched off after target RNA degradation and release. Even though it is not currently known how long synthesized c(OA)s will remain present in the cell after Cas10 deactivation, and how long c(OA)s-bound Csm6/Csx1 stays active, thus far there have been no reports on Csm6/Csx1 causing cell death.

CRISPR-Cas TYPE VI SYSTEMS

Unlike any other CRISPR-Cas system, type VI effector proteins have been demonstrated to exclusively cleave RNA targets [50-54]. In fact, type VI effector proteins are endowed with two distinct active sites, both conferring RNase activity: one for pre-crRNA processing and one involved in target RNA degradation, as described below. Type VI effectors are among the most divergent CRISPR-Cas proteins studied to date and encompass three subtypes: subtype VI-A (Cas13a/C2c2), VI-B (Cas13b1 and Cas13b2) and VI-C (Cas13c) [13, 54]. These Cas13 variants share very low sequence similarity, but can be classified as type VI Cas proteins based on the presence of two conserved HEPN-like RNase domains [51, 53, 54]. Although these two domains appear to be a conserved feature of Cas13 enzymes and are typically located close to the two terminal ends, their spacing within the protein appears to be unique for each subtype [54].

Currently, there are three published crystal structures of type VI-A Cas13a proteins: Cas13a from *Leptotrichia shahii* (LshCas13a), *Lachnospiraceae bacterium* (LbaCas13a) and *Leptotrichia buccalis* (LbuCas13a) [55-57]. Akin to other Class 2 complexes, the crRNA-Cas13a complex has a bi-lobed architecture, consisting of a nuclease (NUC) lobe and a crRNA recognition (REC) lobe. Despite this similar bi-lobed setup, the overall structure and

domains of Cas13a bear very little resemblance to other Class 2 nucleases, such as Cas9 and Cpf1 [58, 59]. Instead, the crRNA-bound form of Cas13a adopts a "clenched fist"-like structure, with the REC lobe being imperfectly stacked on top of the NUC lobe. The REC lobe has a variable N-terminal domain (NTD) followed by a helical domain (Helical-1), whereas the NUC lobe consists of the two abovementioned HEPN domains (HEPN-1 and HEPN-2) separated by a linker domain (Helical-3). Noteworthy, the HEPN-1 domain is split into two subdomains by another helical domain (Helical-2) [57]. The NTD, Helical-1 and HEPN2 domains form a narrow, positively charged cleft that anchors the 5' repeat-derived end of the bound crRNA (the 5' handle), whereas the 3' end of the crRNA is bound by the Helical-2 domain. The first few nucleotides of the guide portion of the crRNA are buried in the NUC lobe, whereas the central region of the crRNA is in a solvent-exposed state [50, 57].

In agreement with the absence of a dedicated pre-crRNA processing nuclease (i.e. Cas6 in type I and III systems), the type VI pre-crRNAs are processed by the Cas13 itself [53]. Typically, the repeats in type VI pre-crRNAs form a bulge-containing stem-looped structure (Fig. 1), which appears to be a conserved and essential feature of mature crRNAs in Cas13a effectors, as disruption of the bulge impedes target RNA degradation, whereas pre-crRNA processing remained unaffected [55-57]. Pre-crRNA processing in type VI involves metal-independent cleavages upstream of the stem-loop and does not require a trans-activating crRNA (tracrRNA) or host factors [53]. Mutagenesis of positively charged residues in the Helical-1 domain (e.g. Arg438 and Lys441 in LshCas13a) abrogated pre-crRNA processing, suggesting that this domain is responsible for the formation and binding of the 5' handle. In agreement with the lack of conservation of some of these residues, the length of the 5' handle (determined by the cleavage site) can vary between different Cas13 homologues [57, 60]. After pre-crRNA processing, the 5' and 3' ends of the mature crRNA are anchored within the complex and are ready to engage in target binding.

The presence of two HEPN RNase domains in Cas13 suggested that RNA is the bona-fide target substrate for type VI CRISPR-Cas systems. Indeed, a heterologously expressed type VI CRISPR-Cas system in *E. coli* conferred sequence-specific immunity against the lytic phage MS2 (an ssRNA phage that does not have a DNA stage within its life cycle), indicating that Cas13 specifically targets ssRNAs [50, 51].

However, it was noted that Cas13-mediated immunity against phage MS2 was accompanied by a growth suppression phenotype. Subsequent biochemical and structural studies unraveled the molecular basis for this phenomenon. It was shown that binding of its cognate target RNA (i.e. a ssRNA complementary to the bound crRNA) converts Cas13 into a sequence non-specific ribonuclease, conferring rapid degradation of neighboring RNAs, an activity often referred to as 'collateral damage'. Target RNA binding results in the formation of an A-form dsRNA helix (crRNA:target RNA) which causes substantial conformational changes in Cas13a to accommodate the propagating duplex within a positively charged channel in the NUC lobe. Concomitantly, these rearrangements bring the catalytic residues of the two HEPN domains into close proximity, thereby forming a single, composite catalytic site for RNA cleavage [56]. Strikingly, this newly formed catalytic site is formed at a considerable distance from the crRNA:target RNA duplex on the external surface of the protein. Therefore, it was concluded that this highly accessible active site would not only degrade the target RNA *in cis* (provided that the target RNA is long enough to reach the active site), but also conferred promiscuous RNase activity, causing the 'collateral damage' to other non-target RNAs *in trans*. Although most RNAs appear to be vulnerable to the promiscuous RNAse activity of Cas13, there are subfamily-specific differences in the nucleotide preferences (e.g. targeting mostly adenosine or uridine-rich RNAs) of these ribonucleases [50, 60, 61].

These observations gave rise to several questions regarding outcome on the individual cell undergoing type VI immunity, similar to those arising from the observed collateral damage activity by CRISPR-Cas type III

immunity described above. For example: is the collateral damage caused by Cas13 aimed at cleaning up local, high concentrations of foreign RNA after binding its target RNA, or does it contribute to thwarting the phage infection by inducing programmed cell death or dormancy [62]? Regardless of the answer, it is evident that this activity must be tightly controlled to prevent unwanted spontaneous cellular toxicity. In some type VI-B systems for example, the type VI-associated genes *csx27* and *csx28* appear to have a regulatory role by either repressing or stimulating the collateral activity of Cas13b respectively [51]. Although the molecular basis for the modulation of Cas13b by Csx27 and Csx28 is currently unknown, it seems to be specific for type VI-B systems, as no clear homologues appear to be associated with type VI-A and VI-C systems. Another layer of control is the requirements of target RNA recognition by Cas13. More specifically, the central, solvent exposed region of the crRNA is thought to initiate nucleation with its cognate target RNA, indicating that Cas13a has a central seed region, opposed to the apically located seed region in type I and type II systems [56]. Mutagenesis of the target RNA complementary to the central seed region of the crRNA demonstrated that two mismatches are enough to prevent Cas13 activation both *in vitro* and *in vivo* and this result was further supported by crystal structure study of target-bound LbuCas13a [50, 56]. In addition, target RNA recognition requires the presence protospacer flanking site (PFS) 3′ of the protospacer, although some Cas13b homologues require both 3′ and 5′ PFS, while others appear to have none [50, 51, 63]. In the case of LshCas13a for example, a non-G PFS motif 3′ of the protospacer is essential for robust RNA interference. Speculatively, base pairing of the last nucleotide 5′ handle (cytosine) of the crRNA with a G as the 3′ PFS of the target RNA could disrupt the proper positioning of the 5′ handle within LshCas13 (described above) and thereby hamper target recognition/degradation [50, 56].

These strict requirements dictating the specificity of Cas13 are perhaps counterbalanced by its unprecedented sensitivity to recognize specific

target RNAs that meet these criteria within a heterogeneous population of non-target RNAs. It has been reported that Cas13 can detect target RNAs with femtomolar sensitivity (even down to the attomolar range, when combined with a pre-amplification step), which opened up opportunities to adopt Cas13 for nucleic acid detection applications, for example in disease diagnostics [64, 65]. In addition, Cas13 has recently been repurposed to provide sequence-specific immunity towards RNA viruses in plants and knockdown and tracking of specific mRNAs in mammalian cells [66, 67]. By fusing a catalytic-dead mutant of Cas13 fused to RNA modifying enzymes, it is now possible to perform highly-specific *in vivo* RNA editing [67]. At last, a nucleic acid detection platform has been developed which combines nucleases from both the type VI and the type III systems (Cas13 and Csm6 respectively) to increase the signal sensitivity, which illustrates the potential of these RNA-guided RNA cleavage CRISPR-Cas systems to be developed into new applications [65].

Next to the questions surrounding the biological role and promiscuous ribonuclease activity of Cas13, it is also still open for debate how type VI systems acquire spacers. PSI-BLAST and HHpred analyses suggested that some of the subtype VI-A loci have adaptation modules, indicating that these type VI systems can acquire new spacers from RNA bacteriophages directly [53]. However, acquisition from RNA generally requires the involvement of a reverse-transcriptase enzyme, as was shown for some type III systems [25, 26]. However, with the exception of *Lachnospiraceae bacterium* MA2020, most type VI systems do not seem to encode a separate reverse transcriptase [53]. Another possibility is that that type VI systems rely on the adaptation modules and CRISPR arrays from other CRISPR-Cas systems residing in the same host, as was recently demonstrated for the type I and III systems in *Marinomonas mediterranea [24]*. Nevertheless, the details underlying spacer acquisition in type VI systems are still waiting to be explored.

CONCLUDING REMARKS

Although type III CRISPR-Cas systems were among the first to be characterized, only recently are we starting to develop a thorough understanding of how these systems operate. Type III systems are endowed with many different sequence-specific and sequence non-specific (deoxy)ribonuclease activities, which are coordinated in a highly organized fashion. The cue for all these activities is the detection and binding of the target RNA, and in this respect, shows conceptual resemblance to the evolutionary distant type VI system (paradoxically, the latest identified CRISPR-Cas system). Both systems seem to confer 'collateral damage' towards cellular RNAs, and as outlined in the review, it remains elusive whether this is controlled in a spatial and/or temporal fashion (i.e. to combat phage infection), or whether it sacrifices the infected cell for the sake of the population by inducing programmed cell death. The recently discovered involvement of second messenger signaling in conferring this behavior in type III interference was unexpected and intriguing, as it is reminiscent of the usage such molecules in mammalian innate immunity. It also illustrates that there might still be many unexpected aspects in CRISPR-Cas immunity that remain to be discovered.

AUTHOR CONTRIBUTIONS

Y.Z., R.H.J.S. and J.v.d.O. contributed to the initial design of this study. All authors contributed to the writing of the manuscript.

ACKNOWLEDGEMENTS

Y.Z. is supported by the Dutch governments Gravitation grant (SIAM Grant 024.002.002 to W.d.V.). M.V. is supported by the Netherlands Organization for Scientific Research (NWO) VIDI Grant (864.11.005) and European Research Council Stg grant (638707). R.H.J.S. is supported by an NWO/Veni grant (016.Veni.171.047), and J.v.d.O by an NWO/TOP grant (714.015.001).

REFERENCES

1. Makarova, K.S., et al., *A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action.* Biol Direct, 2006. **1**(7).
2. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-1712.
3. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA.* Science, 2008. **322**(5909): p. 1843-1845.
4. Louwen, R., et al., *The role of CRISPR-Cas systems in virulence of pathogenic bacteria.* Microbiol Mol Biol Rev, 2014. **78**(1): p. 74-88.
5. Barrangou, R., *The roles of CRISPR-Cas systems in adaptive immunity and beyond.* Curr Opin Immunol, 2015. **32**: p. 36-41.
6. Westra, E.R., A. Buckling, and P.C. Fineran, *CRISPR-Cas systems: beyond adaptive immunity.* Nat Rev Microbiol, 2014. **12**(5): p. 317-326.
7. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-607.
8. Carte, J., et al., *Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes.* Genes Dev, 2008. **22**(24): p. 3489-3496.
9. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-964.
10. Mohanraju, P., et al., *Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems.* Science, 2016. **353**(6299): p. aad5147.
11. Makarova, K.S., et al., *Evolution and classification of the CRISPR-Cas systems.* Nat Rev Microbiol, 2011. **9**(6): p. 467-477.
12. Makarova, K.S., et al., *An updated evolutionary classification of CRISPR-Cas systems.* Nat Rev Microbiol, 2015. **13**(11): p. 722–736.
13. Koonin, E.V., K.S. Makarova, and F. Zhang, *Diversity, classification and evolution of CRISPR-Cas systems.* Curr Opin Microbiol, 2017. **37**: p. 67-78.
14. Staals, R.H.J., et al., *RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus.* Molecular Cell, 2014. **56**(4): p. 518-530.
15. Spilman, M., et al., *Structure of an RNA silencing complex of the CRISPR-Cas immune system.* Molecular Cell, 2013. **52**(1): p. 146-152.
16. Benda, C., et al., *Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4.* Molecular Cell, 2014. **56**(1): p. 43-54.
17. Kazlauskiene, M., et al., *Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition.* Molecular Cell, 2016. **62**(2): p. 295-306.
18. Zhang, J., et al., *Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity.* Molecular Cell, 2012. **45**(3): p. 303-313.
19. Hatoum-Aslan, A., et al., *A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs.* J Biol Chem, 2013. **288**(39): p. 27888-27897.
20. Staals, R.H., et al., *Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus.* Molecular Cell, 2013. **52**(1): p. 135-145.
21. Hale, C.R., et al., *Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex.* Genes Dev, 2014. **28**(21): p. 2432-2443.
22. Tamulaitis, G., et al., *Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus.* Molecular Cell, 2014. **56**(4): p. 506-517.
23. Yosef, I., M.G. Goren, and U. Qimron, *Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli.* Nucleic Acids Res, 2012. **40**(12): p. 5569-5576.
24. Silas, S., et al., *Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems.* Elife, 2017. **6**: p. e27601.
25. Silas, S., et al., *Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein.* Science, 2016. **351**(6276): p. aad4234.
26. Toro, N., F. Martinez-Abarca, and A. Gonzalez-Delgado, *The Reverse Transcriptases Associated with CRISPR-Cas Systems.* Sci Rep, 2017. **7**(1): p. 7089.
27. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex.* Cell, 2009. **139**(5): p. 945-956.
28. Deng, L., et al., *A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus.* Mol Microbiol, 2013. **87**(5): p. 1088-1099.
29. Goldberg, G.W., et al., *Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting.* Nature, 2014. **514**(7524): p. 633-637.
30. Kazlauskiene, M., et al., *A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems.* Science, 2017. **357**(6351): p. 605-609.
31. Niewoehner, O., et al., *Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers.* Nature, 2017. **548**(7669): p. 543–548.
32. Zhu, X. and K. Ye, *Cmr4 is the slicer in the RNA-targeting Cmr CRISPR complex.* Nucleic Acids Res, 2015. **43**(2): p. 1257-1267.
33. Taylor, D.W., et al., *Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning.* Science, 2015. **348**(6234): p. 581-585.

34. Osawa, T., et al., *Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog.* Molecular Cell, 2015. **58**(3): p. 418-430.
35. Estrella, M.A., F.T. Kuo, and S. Bailey, *RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex.* Genes Dev, 2016. **30**(4): p. 460-470.
36. Samai, P., et al., *Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity.* Cell, 2015. **161**(5): p. 1164-1174.
37. Zhang, J., et al., *Multiple nucleic acid cleavage modes in divergent type III CRISPR systems.* Nucleic Acids Res, 2016. **44**(4): p. 1789-1799.
38. Maniv, I., et al., *Impact of different target sequences on type III CRISPR-Cas immunity.* J Bacteriol, 2016. **198**(6): p. 941-950.
39. Elmore, J.R., et al., *Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system.* Genes Dev, 2016. **30**(4): p. 447-459.
40. Wang, D. and R. Landick, *Nuclease cleavage of the upstream half of the nontemplate strand DNA in an Escherichia coli transcription elongation complex causes upstream translocation and transcriptional arrest.* J Biol Chem, 1997. **272**(9): p. 5989-5994.
41. Anantharaman, V., et al., *Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing.* Biol Direct, 2013. **8**(15).
42. Makarova, K.S., et al., *CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems.* Front Genet, 2014. **5**: p. 102.
43. Jiang, W., P. Samai, and L.A. Marraffini, *Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity.* Cell, 2016. **164**(4): p. 710-721.
44. Niewoehner, O. and M. Jinek, *Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6.* RNA, 2016. **22**(3): p. 318-329.
45. Sheppard, N.F., et al., *The CRISPR-associated Csx1 protein of Pyrococcus furiosus is an adenosine-specific endoribonuclease.* RNA, 2016. **22**(2): p. 216-224.
46. Cumby, N., A.R. Davidson, and K.L. Maxwell, *The moron comes of age.* Bacteriophage, 2012. **2**(4): p. 225-228.
47. Brussow, H., C. Canchaya, and W.D. Hardt, *Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion.* Microbiol Mol Biol Rev, 2004. **68**(3): p. 560-602.
48. Goldberg, G.W., et al., *Incomplete prophage tolerance by type III-A CRISPR-Cas systems reduces the fitness of lysogenic hosts.* Nat Commun, 2018. **9**(1).
49. Hatoum-Aslan, A., et al., *Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system.* J Bacteriol, 2014. **196**(2): p. 310-317.
50. Abudayyeh, O.O., et al., *C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector.* Science, 2016. **353**(6299): p. aaf5573.
51. Smargon, A.A., et al., *Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28.* Molecular Cell, 2017. **65**(4): p. 618-630.
52. Tamulaitis, G., Č. Venclovas, and V. Siksnys, *Type III CRISPR-Cas Immunity: Major Differences Brushed Aside.* Trends Microbiol, 2017. **25**(1): p. 49-61.
53. Shmakov, S., et al., *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems.* Molecular Cell, 2015. **60**(3): p. 385-397.
54. Shmakov, S., et al., *Diversity and evolution of class 2 CRISPR-Cas systems.* Nat Rev Microbiol, 2017. **15**(3): p. 169-182.
55. Knott, G.J., et al., *Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme.* Nat Struct Mol Biol, 2017. **24**: p. 825-833.
56. Liu, L., et al., *The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a.* Cell, 2017. **170**(4): p. 714-726.
57. Liu L, L.X., et al., *Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities.* Cell, 2017. **168**(1-2): p. 121-134.
58. Nishimasu, H., et al., *Crystal structure of Cas9 in complex with guide RNA and target DNA.* Cell, 2014. **156**(5): p. 935-949.
59. Yamano, T., et al., *Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA.* Cell, 2016. **165**(4): p. 949-962.
60. East-Seletsky, A., et al., *Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection.* Nature, 2016. **538**(7624): p. 270–273.
61. East-Seletsky, A., et al., *RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes.* Molecular Cell, 2017. **66**(3): p. 373-383.e3.
62. Makarova, K.S., et al., *Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes.* Biol Direct, 2012. **7**(40).
63. Abudayyeh, O.O., et al., *RNA targeting with CRISPR-Cas13.* Nature, 2017. **550**(7675): p. 280-284.
64. Gootenberg, J.S., et al., *Nucleic acid detection with CRISPR-Cas13a/C2c2.* Science, 2017. **356**(6336): p. 438-442.
65. Gootenberg, J.S., et al., *Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6.* Science, 2018. **360**(6387): p. 439-444.
66. Aman, R., et al., *RNA virus interference via CRISPR/Cas13a system in plants.* Genome Biol, 2018. **19**(1): p. 1-9.
67. Cox, D.B.T., et al., *RNA editing with CRISPR-Cas13.* Science, 2017. **358**(6366): p. 1019-1027.

Chapter 3

Structure and activity of the RNA-targeting type III-B CRISPR-Cas complex of *Thermus thermophilus*

This Chapter has been published as:

Staals, R.H.J.*, Agari, Y.*, Maki-Yonekura, S.*, Zhu, Y., Taylor, D.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K., Masuda, A., Dohmae, N., Schaap, P.J., Doudna, J.A., Heck, A.J.R., Yonekura, K., van der Oost, J., Shinkai, A. (2013) Structure and activity of the RNA-targeting type III-B CRISPR-Cas complex of *Thermus thermophilus*. Mol Cell. 52(1):135-145

*contributed equally

ABSTRACT

The CRISPR-Cas system is a prokaryotic host defense system against genetic elements. The type III-B CRISPR-Cas system of the bacterium *Thermus thermophilus*, the TtCmr complex, is composed of six different protein subunits (Cmr1-6) and one crRNA with a stoichiometry of $Cmr1_1 2_1 3_1 4_4 5_3 6_1$:$crRNA_1$. The TtCmr complex copurifies with crRNA species of 40 and 46 nt, originating from a distinct subset of CRISPR loci and spacers. The TtCmr complex cleaves the target RNA at multiple sites with 6 nt intervals via a 5′ ruler mechanism. Electron microscopy revealed that the structure of TtCmr resembles a ‘‘sea worm’’ and is composed of a Cmr2-3 heterodimer ‘‘tail,’’ a helical backbone of Cmr4 subunits capped by Cmr5 subunits, and a curled ‘‘head’’ containing Cmr1 and Cmr6. Despite having a backbone of only four Cmr4 subunits and being both longer and narrower, the overall architecture of TtCmr resembles that of type I Cascade complexes.

INTRODUCTION

The CRISPR-Cas (clustered regularly interspaced short palindromic repeat CRISPR-associated proteins) system is a recently discovered, prokaryotic host defense system (found in most archaea and many bacteria) targeting mobile genetic elements such as phages and plasmids [1-4]. Typically, CRISPR loci are composed of 25-38 bp direct DNA repeats, separated by unique spacer sequences of a similar length [5], which are derived from previous encounters with invading elements. The cas genes are often located in close proximity to these CRISPR loci and are used to categorize the system into one of the three major CRISPR-Cas (sub)types [6]. The exact mechanisms of CRISPR-mediated defense are distinct between the different CRISPR-Cas types and subtypes. In general, the CRISPR loci are transcribed and processed into small RNAs called CRISPR RNAs (crRNAs), each composed of a separate spacer sequence flanked by part(s) of the repeat sequence. The mature crRNAs are then loaded on either a protein complex (for type I and type III CRISPR-Cas systems), composed of several Cas proteins [2, 7], or a single Cas protein, Cas9 [8, 9], in the case of a type II CRISPR-Cas system. The complex is then guided toward its complementary target, resulting in base-pairing interactions between the crRNA and the invading genetic element and generally, followed by degradation of the invader. New spacers can be acquired in a still poorly understood process called ‘‘adaptation,’’ in which a new spacer (derived from the invader) will be integrated in a directional fashion into the chromosomal CRISPR spacer array.

In type III systems, crRNA synthesis starts with the transcription of the CRISPR array. The resulting pre-crRNA is most likely first cleaved in the repeat sequence by one or more Cas6 proteins followed by further processing at the 3’ end, potentially involving other Cas proteins to generate the mature crRNA [10, 11]. Consistent with other CRISPR systems, subtype III-A seems to target DNA invaders [12]. However, the subtype III-B is unique, in that it has been shown to target RNA rather than DNA. The

effector complex of the subtype III-B system, the Cmr complex, binds crRNA, and cleaves a target RNA complementary to the bound crRNA[7]. Comparison of different Cmr modules revealed that subtype III-B contains either six Cmr genes (Cmr-α module) and an associated csx1 gene, such as the Cmr complex from *Pyrococcus furiosus* [7], or seven Cmr genes (Cmr-b module), as has been shown in *Sulfolobus* species [13, 14]. Surprisingly, the modus operandi of the Cmr complexes encoded by these Cmr modules appears to be markedly different. The Cmr-α complex of *P*. *furiosus* cuts target RNA molecules according to a ‘‘ruler mechanism,’’ cleaving the target 14 nucleotides (nt) upstream of the 3’ end of the basepaired crRNA [7]. The Cmr-β complex of *S*. *solfataricus* on the other hand, has been shown to cleave substrate RNAs at UA dinucleotides [14]. The ribonuclease activities of a bacterial Cmr complex have not been addressed so far.

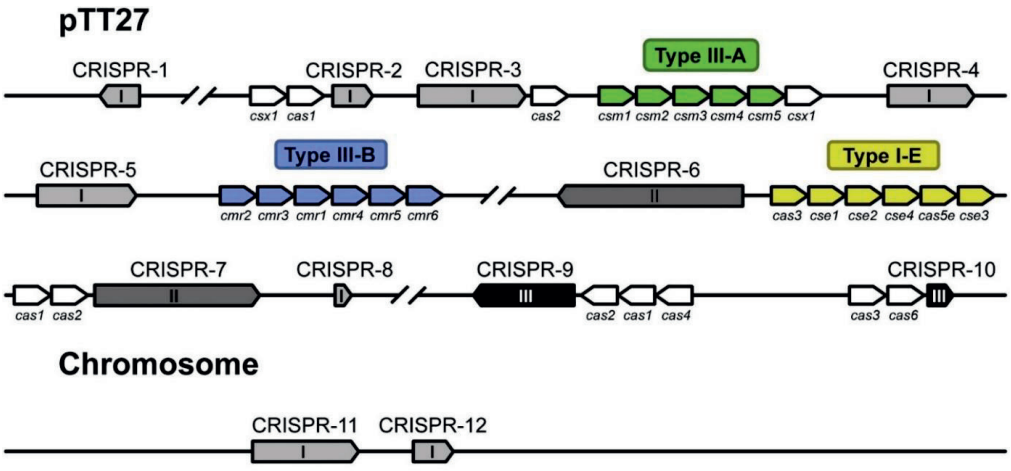
The thermophilic bacterium *Thermus thermophilus* HB8 has 11 CRISPR loci in total, two of which are located on the chromosome, while the other nine loci are found on a 260 kbp plasmid pTT27 [15] (Figure 1). Here, the CRISPR locus 8 is not considered a genuine CRISPR because this locus consists of only one (type I) repeat sequence and does not contain any spacers [15]. These loci are classified into three types depending on the nucleotide sequence of the repeat [15]. All 11 CRISPRs are unidirectionally expressed and processed to mature crRNAs, although the stability or processing can be different between crRNAs [16]. The 5’ ends of all the crRNAs predominantly retain eight nucleotides derived from repeat sequences (the 5’ handle), while their 3’ ends are variable depending on the repeat sequence type, suggesting that this strain has multiple crRNA processing systems [16]. More than 30 Cas proteins are encoded by pTT27, including subtypes I-E, III-A, and III-B [15, 16] (Figure 1). Expression of CRISPR loci and most cas genes are upregulated by infection with the myophage ΦYS40 [15]. Furthermore, an operon containing the subtype I-E genes and one containing the subtype III-A genes are positively regulated by cAMP receptor protein (CRP) in a cAMP-dependent manner [15, 17].

Elucidation of the functional and structural mechanisms and roles of all CRISPR-Cas machineries in this strain will provide a systematic understanding of the host defense systems. In this study, we investigated the structure and function of a bacterial Cmr-α effector complex of the subtype III-B CRISPR-Cas system: the *T. thermophilus* Cmr complex (TtCmr).

RESULTS

Preparation and Initial Characterization of the Cmr Complex from *T. thermophilus*

We constructed a recombinant *T. thermophilus* strain that produces the Cmr6 protein fused with a (His)6 tag at its C-terminus. The Cmr complex was purified to homogeneity by three different column chromatography steps, as described in the Supplemental Experimental Procedures. The purified protein complex was composed of six proteins (Figure 2A). We confirmed by mass spectrometry based peptide sequencing that the proteins corresponded to TTHB162 (Cmr1), TTHB160 (Cmr2), TTHB161 (Cmr3), TTHB163 (Cmr4), TTHB164 (Cmr5), and TTHB165 (Cmr6) (data not shown), indicating that the *T. thermophilus* Cmr complex is composed of six subunits. According to the N-terminal amino acid sequence analysis, the N-terminal methionine residues of the Cmr4 and Cmr6 proteins in the complex were deleted. A blue native polyacrylamide gel electrophoresis (BN-PAGE) analysis of the Cmr complex revealed two bands: an 350 kDa major band and an 310 kDa minor band (Figure 2B). These bands suggest multiple stoichiometries of the Cmr complex may co-occur (addressed below). The complex was subjected to the gel permeation chromatography (GPC)-high-performance liquid chromatography (HPLC) in the presence of 6 M guanidine-HCl. We identified the peaks using individually expressed recombinant Cmr proteins as standards (Figures S1, S2A, and S2B). The UV spectrum of the peak at 44 min detected from the gel-filtration chromatogram was similar to that of a synthetic RNA (Figure S2C),



Locus	Genomic location	Representative repeat sequence (5' to 3')	Repeat type	# Spacers
CRISPR-1	18303-18044	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	3
CRISPR-2	133766-133954	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	2
CRISPR-3	135156-136099	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	12
CRISPR-4	144129-144842	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	9
CRISPR-5	146042-146983	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	12
CRISPR-6	190947-189507	GTAGTCCCCACGCACGTGGGGATGGACCG	II	23
CRISPR-7	200811-202078	GTAGTCCCCACGCGTGTGGGGATGGACCG	II	20
CRISPR-8	210807-210842	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGATAC	I	0
CRISPR-9	228237-227324	GTTGCAAACCCCGTCAGCCTCGTAGAGGATTGAAAC	III	12
CRISPR-10	239108-239214	GTTGCAAACCTCGTTAGCCTCGTAGAGGATTGAAAC	III	1
CRISPR-11	872101-873199	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	14
CRISPR-12	874397-874734	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	4

Figure 1. Schematic Representation of CRISPR Arrays and cas Genes on the Genome and Plasmid pTT27 of *T. thermophilus* HB8

CRISPR arrays (1–12) are indicated in different grayscales, depending on the repeat type (I, II, or III). *Cas*(-related) genes belonging to a particular CRISPR-Cas subtype are colored in green (subtype III-A), blue (subtype III-B) or yellow (subtype I-E). Additional *cas* genes are indicated in white. For each of these CRISPR arrays, the bottom panel summarized the genomic location, the representative repeat sequence, repeat type and the number of spacers. The 5' handles, as found by our deep sequencing analysis, are underlined.

suggesting that this fraction is derived from crRNA. Transcription of *T. thermophilus cmr* genes increases after infection of ΦYS40 phage [15]; however, yield and subunit composition of the Cmr complex from phage-infected cells were not significantly different from that of noninfected cells (data not shown).

extracted from the purified *T. thermophilus* Cmr complex and analyzed by denaturing gel electrophoresis, which showed that the complex binds two major crRNA species of 40 and 46 nt, although low-abundant levels of crRNA species of other sizes could be observed, as well (Figure S3A). No substantial differences in abundance or size distribution could be discerned when comparing the Cmr-bound crRNAs from either the uninfected or the phage-infected cells (data not shown). Successful polyadenylation of their 3' ends by *E. coli* poly(A) polymerase indicates that these crRNA have accessible, unphosphorylated 3' ends (Figure S3A). Furthermore, the 5'-phosphate-dependent ribonuclease Xrn-1 was not able to degrade the purified crRNAs prior to polynucleotide kinase (PNK) treatment, suggesting the presence of a 5' hydroxyl group (Figure S3B). In order to establish the exact nature of the Cmr-bound crRNAs, as well as to determine any changes in the crRNA population upon phage infection, the crRNAs were prepared for high-throughput sequencing using 2374 paired-end reads.

The sequencing reads were mapped to the genome of *T. thermophilus*, which revealed a considerable bias in the abundance of different crRNAs incorporated into the complex (Figure 3A). This variation in abundance is evident when individual spacers within one CRISPR locus are considered and some CRISPR loci seem to be favored in contributing mature crRNAs for the Cmr complex when compared to others. A detailed overview of all the reads is given in Figure S4. Out of the 11 total CRISPR loci, crRNAs that are destined for incorporation into the Cmr complex originated predominantly from only four of these loci (CRISPR-1,2,4, and 11). Although CRISPR-2 was slightly underrepresented in our data set, this bias is in good agreement with the relatively high expression of these loci [16]. In accordance with the observed crRNA sizes on denaturing gels (Figure S3A), the size-distribution resulting from the crRNA sequencing showed that the major population consists of lengths of 40 (24.3%) and 46 nt (16.7%) (Figure 3B). This size-distribution is observed for the total crRNA population, and also when individual cases (such as CRISPR-4.5, spacer 5 from locus 4,

the most abundant crRNAs in the data set) are considered (Figure 3C). The presence of crRNAs of other sizes within the complex are reminiscent of processing intermediates, suggesting that the final steps in crRNA maturation can occur when bound to the complex. crRNAs bound to the Cmr complex retain the last 8 nt of the Type I repeat sequence at their 5′ end [15]. The observation that the majority of reads (81.2%) started with the sequence ''ATTGCGAC'' could suggest that this 5′ handle promotes binding to the complex. While reads originating from Type II repeats were almost completely absent in our data set, a small number of reads originated from CRISPR-9 and -10 (containing Type III repeats). The last 8 nt of the repeats in these latter two CRISPR loci are very similar to those of Type I repeats (''ATTGAAAC''). These data indicate that the Cmr complex favors 40 and 46 nt crRNAs with Type I (-like) repeats.

Stoichiometry of the Cmr Complex

To further investigate the structural basis of the interaction between Cmr subunits and the stoichiometry of the complex, we determined the composition of the Cmr protein assembly using native mass spectrometry [18, 19]. Denaturing and tandem MS analyses provided accurate mass measurements for each subunit of Cmr. The measured masses of the individual Cmr subunits were consistent with their theoretical values (Table S1). Analysis of the intact assembly by native MS revealed the presence of two major components. From their well-resolved charge state distributions, we could determine masses of 364,610 ± 46 Da and 319,060 ± 40 Da (Figure 4A), in good agreement with the data from native gel electrophoresis. Although we could measure the masses quite accurately, this did not provide direct information on the stoichiometry of the assembly. However, complexes composed of only one copy of each Cmr component ($Cmr1_1 2_1 3_1 4_1 5_1 6_1$:$crRNA_1$) would have a mass of only 246 kDa, indicating that some subunits are present in multiple copies.

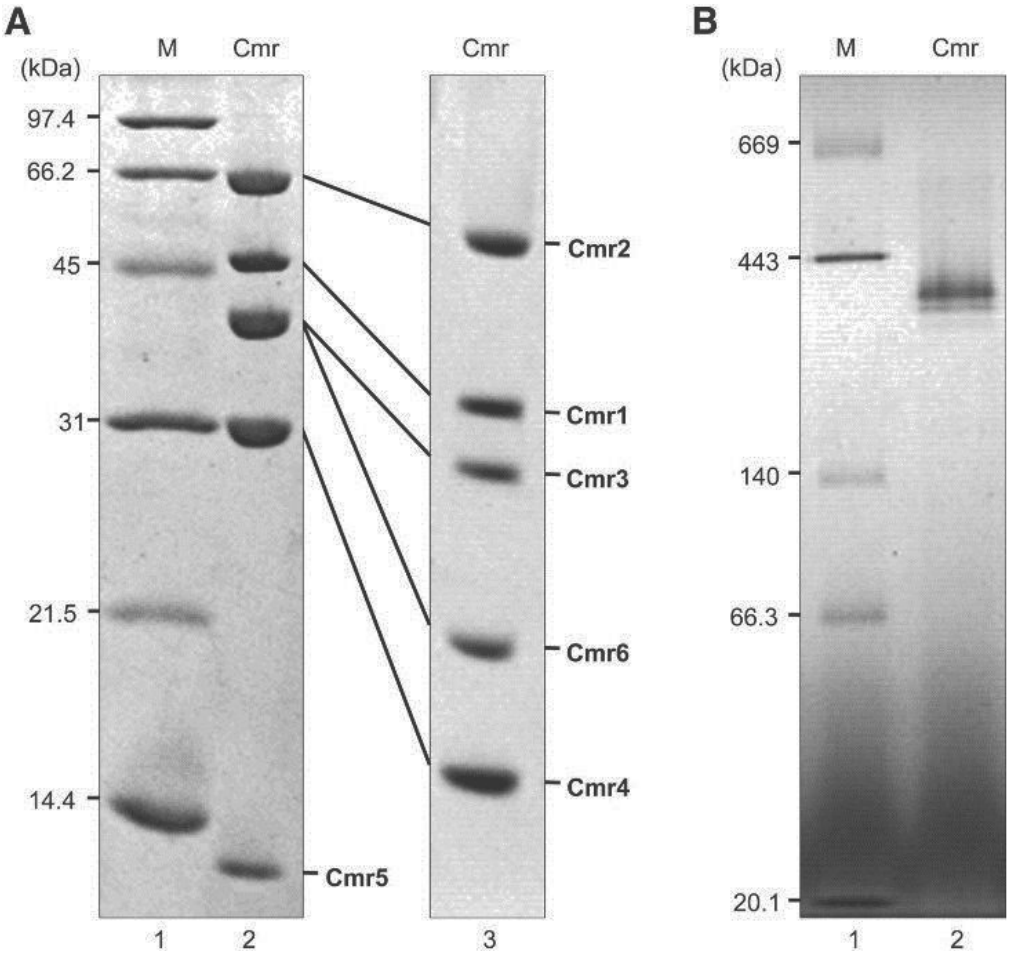


Figure 2. Isolation and Characterization of the Native TtCmr Complex

(A) SDS-PAGE analysis of the *T. thermophilus* Cmr complex. The purified protein (2 mg) was analyzed on a 15% polyacrylamide gel (lane 2), or an 8% polyacrylamide gel containing 8 M urea (lane 3), and the gels were stained with Coomassie Brilliant Blue R-250. Each subunit is indicated. The Cmr6 has a $(His)_6$-tag at its C terminus. Lane 1, molecular-weight markers (Bio-Rad Lab- oratories, Inc.): phosphorylase b (97.4 kDa), serum albumin (66 kDa), oval- bumin (45 kDa), carbonic anhydrase (31 kDa), trypsin inhibitor (21.5 kDa), lysozyme (14.4 kDa). (B) BN-PAGE analysis of the *T. thermophilus* Cmr complex. Two μg of the Cmr complex was analyzed on a 4%–16% linear polyacrylamide gradient gel in the presence of 0.02% Coomassie Brilliant Blue G-250 (lane 2). Lane 1, molecular weight markers (Sekisui Medical, Co, Ltd.): thyroglobulin (669 kDa), ferritin (443 kDa), lactate dehydrogenase (140 kDa), bovine serum albumin (66.3 kDa), and trypsin inhibitor (20.1 kDa). Protein concentration used was determined by the Bradford Method (Bradford, 1976), using bovine serum albumin as a standard.

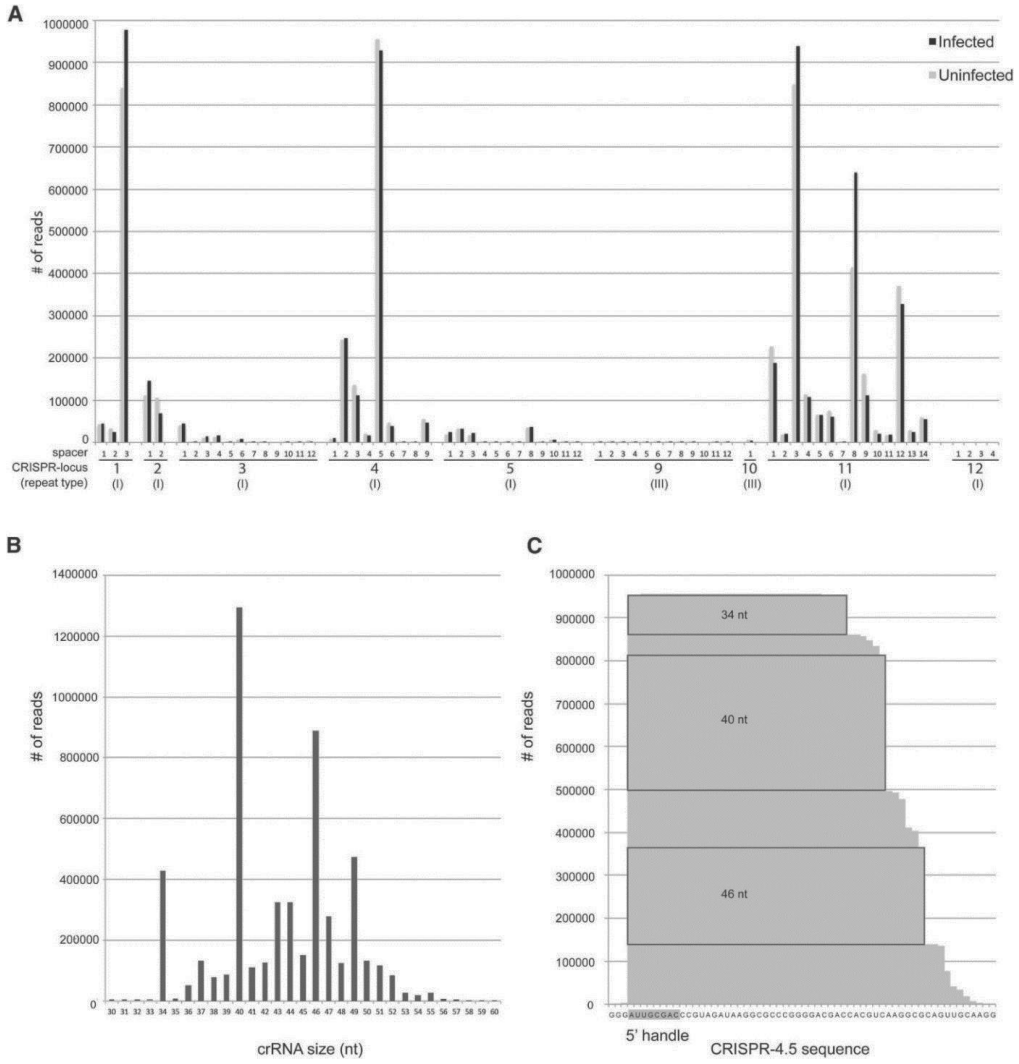


Figure 3. Deep-Sequencing Analysis of the crRNA Content of the TtCmr Complex

(A) Distribution of crRNA reads over the 12 CRISPR loci present in the genome of *T. thermophilus* HB8. (B) Histogram illustrating the size distribution of the mapped crRNA reads. (C) Deep-sequencing profile of the crRNA reads mapping to CRISPR-4.5. The 5' handle, which are the first eight nucleotides, derived from the upstream repeat sequence are indicated in gray. Abundant crRNA species of 34, 40, and 46 nt are indicated in boxes. An overview of all the mapped reads is given in Figure S4.

The two major complexes of 365 and 319 kDa, most likely represent the intact Cmr and the Cmr subcomplex lacking Cmr1, respectively. The measured mass difference between the two assemblies is 45.5 kDa (mass of Cmr1 monomer is 44,479 Da). Collision-induced dissociation experiments on mass-selected ions of the intact Cmr confirmed the facile dissociation of

Cmr1, hinting at a more loosely attachment and a peripheral location of this subunit. Interestingly, besides the elimination of Cmr1, no other subunits were expelled from the intact complex under collision induced dissociation conditions. Selection and activation of the 319 kDa Cmr subcomplex did not result in substantial fragmentations, indicating a high gasphase stability of the Cmr complex.

As the Cmr assembly could not be disrupted by tandem MS, we sought to further explore the Cmr structure using a combination of in-solution dissociation of the intact Cmr complex and reconstitution of Cmr complexes lacking single Cmr proteins. From all these mass spectrometric data (Figure 4B and Table S2) the stoichiometry of the complex was determined to be $Cmr1_1 2_1 3_1 4_4 5_3 6_1$:$crRNA_1$, i.e., containing multiple copies of Cmr4 and Cmr5. With this stoichiometry the expected mass would be 364,036 kDa, in good agreement with a measured mass of 364,610 kDa. As shown before [18, 20] these types of mass spectrometric experiments can reveal a wide range of other structural details. For instance, we observed a stable heterodimer between Cmr2 and Cmr3, in agreement with earlier observations [21, 22]. Having multiple copies present, the backbone of the Cmr complex is likely composed of Cmr4 and Cmr5, wrapped around the crRNA molecule.

Activity of the Cmr Complex

To evaluate the nucleolytic capabilities of the Cmr complex, an RNA substrate was generated by in vitro transcription of a hybridized oligonucleotide pair in the presence of α-^{32}P UTP. This resulted in a 5’ nt internally labeled RNA substrate, complementary to one of the most abundant crRNAs found incorporated into the complex: CRISPR-4.5 (Figure 3A). Ribonuclease activity was assayed by incubating the RNA substrate with the Cmr complex in buffers containing different concentrations of Mg^{2+} (Figure 5A). The results showed that the complex harbors Mg^{2+}-dependent endoribonuclease activity, although activity was also observed in the presence of Mn^{2+} (data not shown). The presence of multiple, distinct

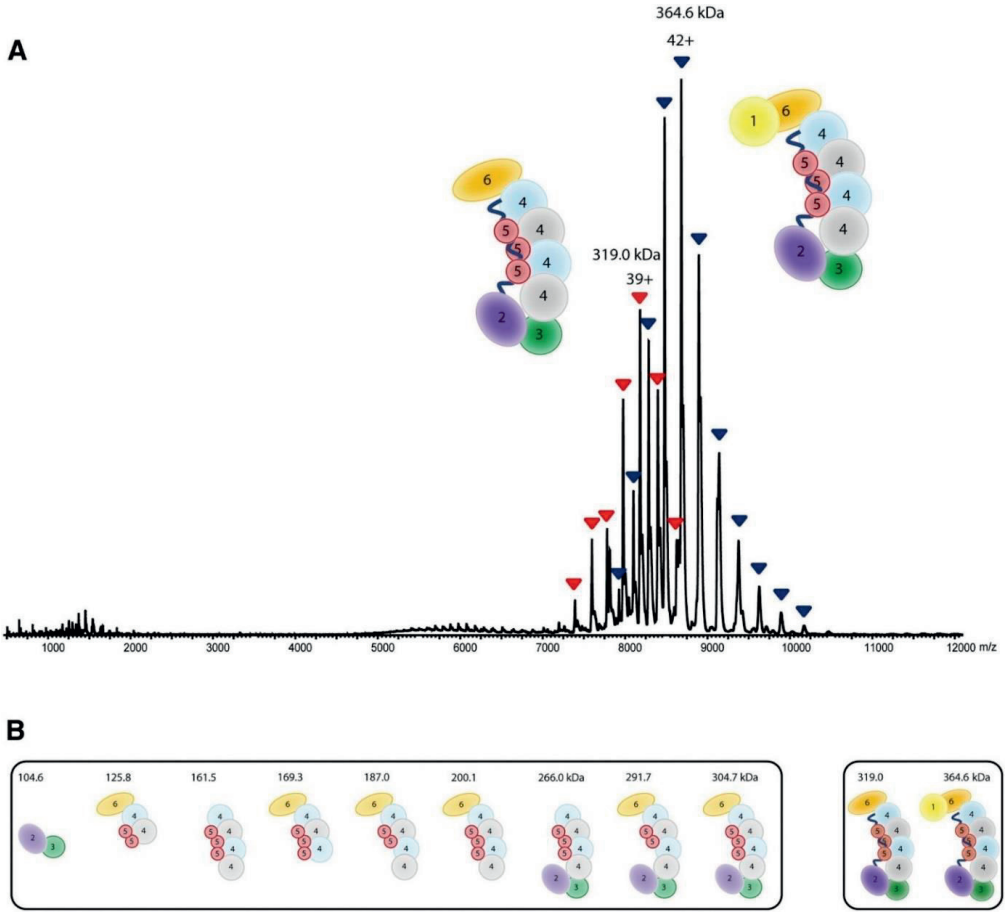


Figure 4. Subunit Composition of the TtCmr Complex

(A) Native nano-electrospray ionization mass spectrum of Cmr. Two major, well-resolved charge state distributions are present at high m/z values, corresponding to complexes of 365 kDa (blue) and 319 kDa (red). More exact masses are provided in Table S1. (B) Cmr (sub)complexes detected by native mass spectrometry. These (sub)complexes were detected and mass analyzed by reconstituting the Cmr complex lacking individual Cmr subunits. More accurate masses and assignments are given in Table S2.

cleavage products indicates that the target RNA is cleaved at different sites. No activity was observed when complementary ssDNA or dsDNA substrates were used (Figure S5).

In order to determine the specificity of the complex, we tested several RNA substrates complementary to Cmr-associated crRNAs that were found in either high (CRISPR-4.5), moderate (CRISPR-4.2), or low (CRISPR-2.1) levels, as well as an RNA substrate complementary to a crRNA for which

almost no reads were found (CRISPR-3.2) in our deep sequencing data set (Figure 3A). The results showed that the activity is specific for RNAs that are complementary to the Cmr-bound crRNAs and that this particular activity cannot be attributed to any potential contaminants present in the purified Cmr-complex sample (Figure 5B). Regardless of the RNA substrate, sizes of the cleavage products were found to be identical (21, 27, and 33 nt), arguing against a sequence-dependent cleavage mechanism like the Cmr complex in *S*. *solfataricus* [14]. To study the activity in more detail, a similar experiment was performed using a 3′- or 5′-labeled RNA substrate (CRISPR-4.5). While the 5′-labeled substrate was degraded to 39, 33, 27, 21, and 15 nt products, a single 12 nt degradation product was observed with the 3′-labeled substrate (please note that the 3′-labeling reaction added one additional nucleotide at the 3′ end of the RNA substrate). These results suggest that the substrate is initially cleaved at its 3′ end, followed by additional cleavages processing toward its 5′ end (Figure 5C). These findings are more in line with the ‘‘ruler mechanism’’ of *P*. *furiosus* [7]. However, the heterogeneity in the sizes of the Cmr-bound crRNAs (Figure 3B) makes it difficult to draw firm conclusions about this possibility. Therefore, the Cmr complexes were reconstituted (using individually expressed and purified Cmr subunits from *E*. *coli*) and loaded with either a 43 or a 46 nt crRNA, which are 3′-truncated versions of the same crRNA (CRISPR-4.5). These reconstituted complexes (‘‘Rec. Cmr’’) were subsequently tested in an activity assay using a synthetic, 5′-labeled RNA substrate, and the activity of these complexes was compared to the endogenous Cmr complex (‘‘Cmr’’) (Figure 5D). Surprisingly, the results showed that regardless of the 3′ end of the Cmr-bound crRNA(s), the RNA substrate was still cleaved at similar positions (Figure 5D), excluding a ruler mechanism based on the 3′ end of the crRNA. Instead, these results imply either that the Cmr complex harbors multiple active sites, or that the RNA substrate is threaded through the complex toward a single active site. By analogy to type I-E Cascade [20, 23], the 5′ handle of the crRNA is most likely recognized by one of the Cmr subunits, which in turn fixes the position

of the RNA substrate within the complex. Therefore, we favor a scenario where the Cmr complex contains multiple active sites, with Cmr4 being the most likely candidate for providing the endoribonuclease activity. Based on this model, the substrate will be cleaved at fixed distances (6 nt intervals) measured from the 5′ end of the crRNA: a 5′ ruler mechanism (Figure 6). Since the Cmr complex harbors just four Cmr4 subunits and we observed five differently sized degradation products, another Cmr subunit (possibly Cmr1 or Cmr6) might provide the remaining cleavage activity.

Three-Dimensional Structure of the Cmr Complex

We gained insight into the three-dimensional structure of the Cmr complex containing crRNA using electron microscopy (EM) and single particle analysis of negatively stained particles (Figure S6A). We reconstructed a three-dimensional (3D) map of TtCmr at ~26 Å resolution (using the 0.5 FSC criterion) from two-dimensional (2D) class averages using the common-line method followed by projection matching refinement of 9,000 individual particles images in EMAN and EMAN2 (Figures 7A and S6B). Reprojections of the map matched with high fidelity to reference-free 2D class averages, indicating good agreement between our reconstruction and experimental data (Figure 7B). The reconstruction of Cmr bound to crRNA resembled a ‘‘sea worm’’ with overall dimensions of 90X100X220 Å , compatible with the expected molecular mass of the fully intact complex (~365 kDa). NTA-nanogold particle labeling of a His_6 C-terminally tagged Cmr6 within the reconstituted complex, showed Cmr6 is located at either the head or the tail of the structure (Figure 7C). Then, we collected a larger, higher-resolution data set using the automated molecular microscopy system LEGINON (Figures S7A–S7C). Using the first reconstruction as a starting model, we obtained a reconstruction of the complex at ~22 Å resolution (using the 0.5 FSC criterion) from ~45,000 particles, showing clear additional structural features that could be segmented easily and automatically (Figure 7D). The sea worm is composed of a repeating helical backbone of four identical subunits with the size and shape of Cmr4 that

are capped by three smaller, globular densities with the size and shape of Cmr5 juxtaposed with a ~20 Å channel or groove running the length of the complex. This channel could easily accommodate an A-form RNA duplex in a hypothetical target-bound state. The distance between adjacent Cmr4 subunits within this backbone is ~25 Å consistent with a length of ~6–7 bp of A-form RNA, closely mirroring the size of our observed cleavage products and suggesting these subunits as the catalytic centers. The Cmr2-Cmr3 subcomplex crystal structure from *P. furiosus* fits into the ''tail'' with high fidelity (Figure 7E). A curled ''head'' containing Cmr1 and Cmr6 adjoins the other side of the proposed RNA-binding groove and positions these subunits on the opposite side of Cmr2-Cmr3.

DISCUSSION

In this study, we have analyzed the structure and activity of a bacterial type III-B CRISPR-Cas effector complex: the Cmr complex from *T. thermophilus*. Both in terms of structure and activity, the TtCmr complex shows interesting differences compared to its previously studied archaeal homologs [7, 14, 24].

crRNAs of different sizes were found to bind the TtCmr complex, although the major population consisted of 40 and 46 nt lengths, 1 nt smaller than previously reported for the *P. furiosis* Cmr complex [7]. Interestingly, with the exception of the abundant 49 nt crRNA species, the size distribution showed a trend for crRNAs that differ 6 nt in size: 34 nt, 40 nt, and 46 nt (Figure 3B). The maturation of these crRNAs is presumably initiated by Cas6-mediated single-site cleavage in the repeat sequence of the pre-crRNA, generating an ''1x intermediate'' crRNA that contains the last 8 nt of the repeat sequence at its 5' end. The 3' end is subsequently shortened by an unknown ribonuclease (other than Cas6) to produce the mature crRNA [7, 10, 25]. We confirmed that TTHB231 protein (NCBI accession number YP_145470) from *T. thermophilus* HB8 has Cas6-like activity, as incubation of a pre-cRNA with TTHB231 yielded cleavage products of ~70–80 nt, which

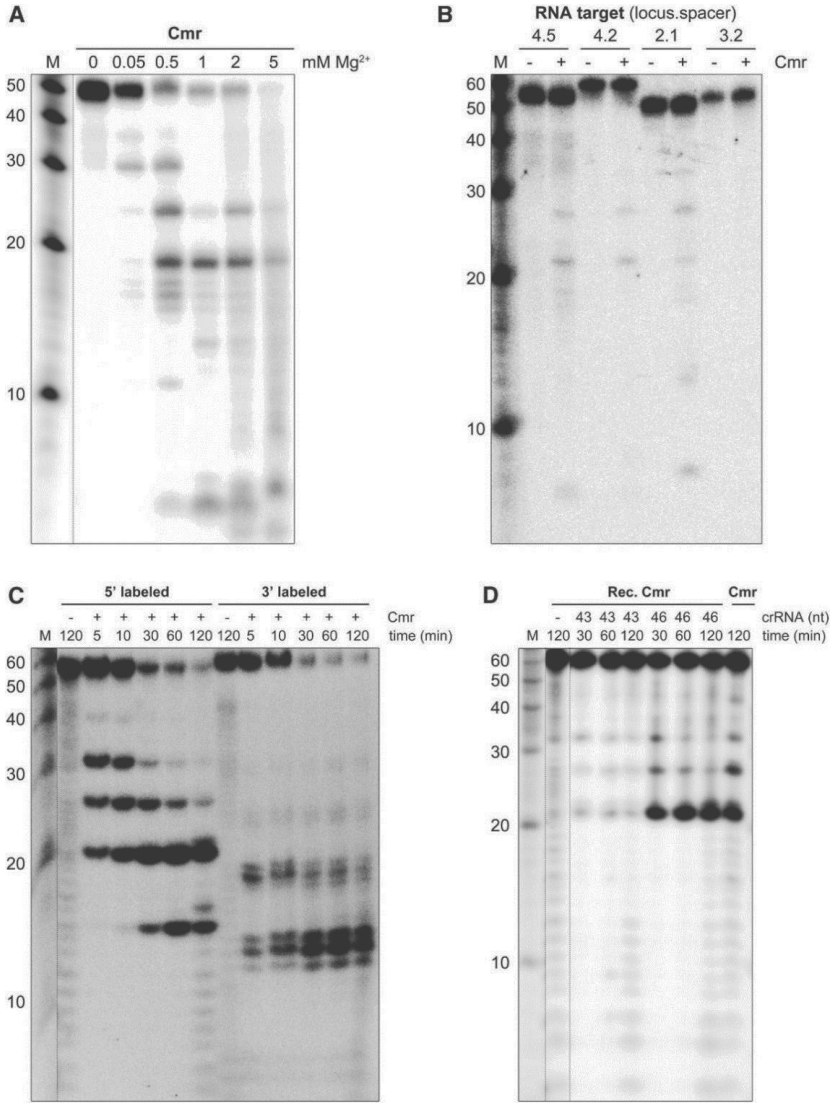


Figure 5. RNA-Cleavage Activity of the TtCmr Complex

(A) An internally labeled RNA substrate complementary to CRISPR-4.5 was incubated with the endogenous Cmr complex for 1 hr in the presence of different concentrations of Mg^{2+}. Samples were analyzed by denaturing PAGE, followed by phosphorimaging. (B) Different internally labeled RNA targets (complementary to CRISPR-4.5, 4.2, 2.1, and 3.2) were tested in a similar assay (including 2 mM Mg^{2+}), demonstrating the specificity of the Cmr complex. (C) Activity assay with the endogenous Cmr complex using a 5′- or 3′-labeled RNA substrate (CRISPR-4.5). (D) The 5′-labeled RNA substrate (CRISPR-4.5) was incubated with either the endogenous (''Cmr'') or the reconstituted Cmr complex (''Rec. Cmr'') for the indicated amount of time. The reconstituted complex was preloaded with the 43 or the 46 nt crRNA. Reconstituted Cmr complexes without crRNA (''-'') were included as a control. Noncontiguous lanes from the same gel are indicated with dotted lines. The results of a Cmr activity assay with complementary ssDNA and dsDNA substrates are depicted in Figure S5.

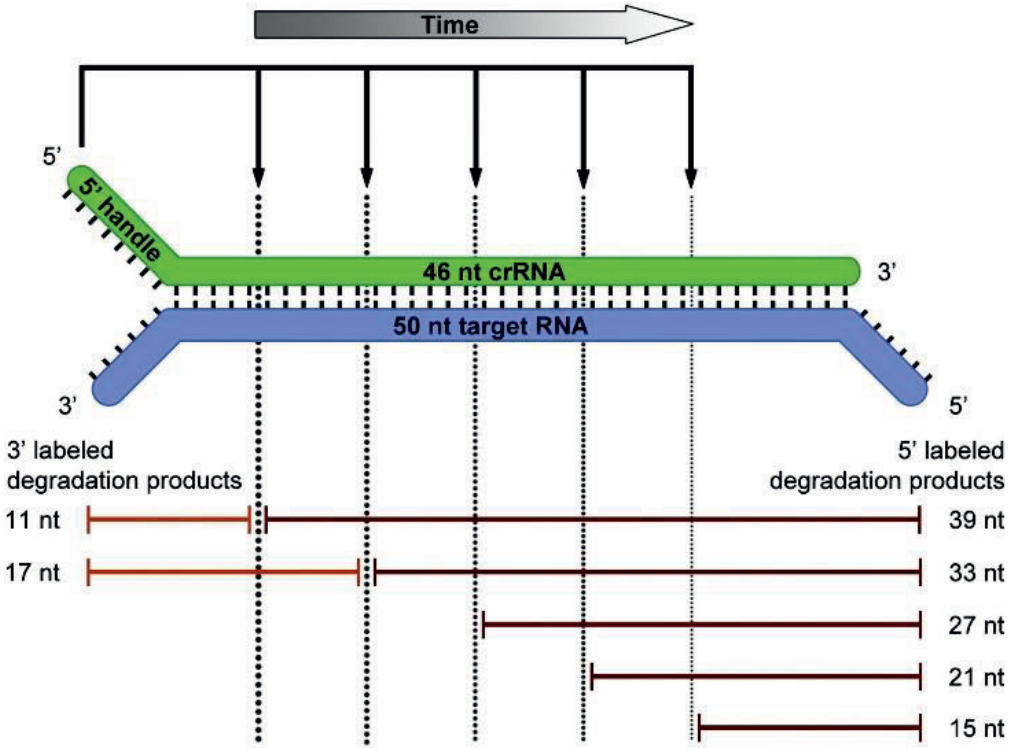


Figure 6. Model for the Cleavage Activity of the TtCmr Complex

Schematic representation of the cleavage activity deduced from the activity assays presented in Figure 5. Sizes of the different 3'- and 5'-labeled cleavage products that were observed in these assays are indicated. Note that over time, the Cmr complex cleaves the RNA substrates from the 3' end, cutting every 6 nt progressing toward the 5' end. Cmr4 is the proposed subunit responsible for this catalytic activity, while another Cmr subunit might be responsible for the remaining cleavage.

are the expected sizes of the 1x intermediate crRNAs (data not shown). The occurrence of Cmr-bound crRNAs of other sizes supports the possibility that this step in crRNA maturation occurs when bound to the complex. In addition, a recent study [26] has revealed that the 3' trimming of crRNAs in a type III-A (Csm) complex is catalyzed by a trans-acting factor not present in the complex; most likely, the same is true for type III-B (Cmr) complexes.

The population of crRNAs that was found to bind the Cmr complex shows a remarkable bias for particular spacers, the majority of which consisted of

CRISPR-1.3, CRISPR-4.5, and CRISPR-11.3, and to a lesser extent CRISPR-11.8 and CRISPR-11.12. As opposed to the upregulation of the *cmr* genes [15], the overall crRNA abundance profile did not change upon ΦYS40 phage infection (Figure 3A), indicating that this observed bias is not regulated by infection of this particular phage. The abundance profile of Cmr-bound crRNAs corresponds well with the cellular expression patterns found for these RNAs [16]. Since crRNA maturation starts with the transcription of a long pre-crRNA transcript [2, 27], the differences in abundance of the individual crRNAs can have different explanations. First, Cas6 (and/or enzymes responsible for 3' end trimming) could favor the maturation of particular crRNAs over others. Second, particular crRNAs might bind the Cmr complex more efficiently than others, potentially promoting their stability. In addition, differences in the stability of the crRNA itself might influence its abundance [14]. Lastly, internal promoters in repeat and/or spacer sequences could contribute to differences in crRNA abundance [28]. Many of these explanations are related to the secondary structure and/or sequence of the repeat and spacer sequences. However, we did not observe a clear correlation between the thermodynamic stability of in silico predicted crRNA structures and their abundance within the Cmr complex (data not shown). We found that protospacers corresponding to the CRISPR-11.1, -11.3, and -11.4 were located in the intergenic regions of the chromosomal DNA of this strain. Further studies are necessary to address the possibility for the Cmr complex to target self RNAs.

Recently, the structures of several CRISPR-Cas complexes have been elucidated. Cascade, the type I-E complex of *E. coli* is composed of Cse1, Cse2, Cas7, Cas5e (Cas5), and Cas6e proteins and forms a 405 kDa complex with a stoichiometry of 1:2:6:1:1 that binds a 61 nt crRNA [20]. The negative stain EM and small angle X-ray scattering structures of this complex show an unusual seahorse shape [20]. A subsequent cryo-EM analysis resolved the location of the individual subunits. Moreover, structures with and without target have revealed a major conformational

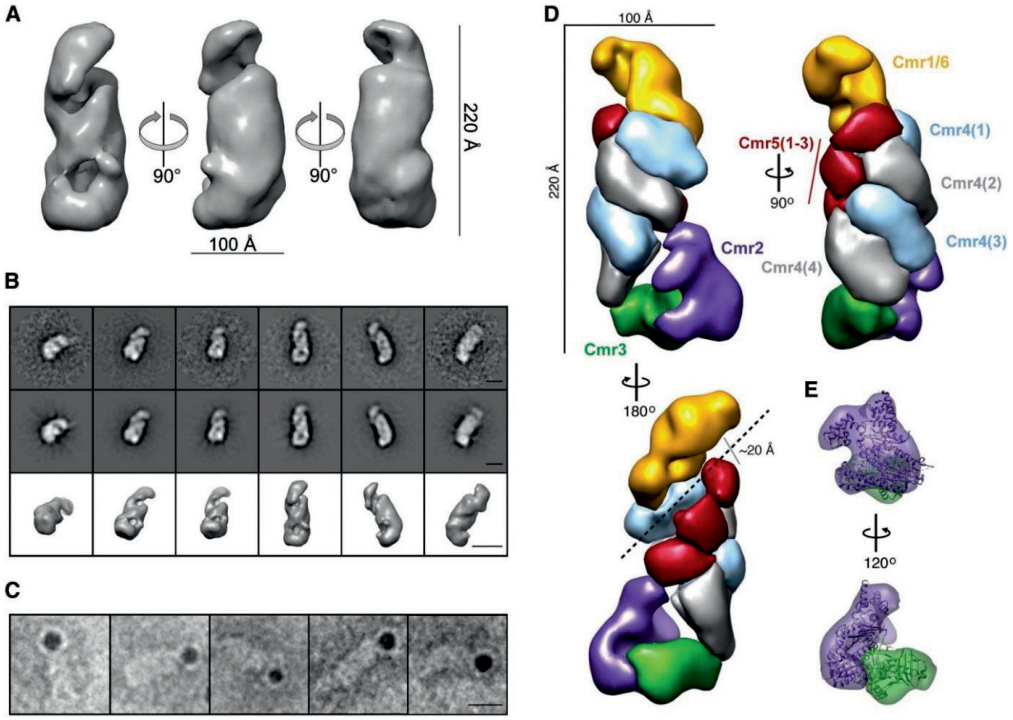


Figure 7. Molecular Architecture of the T. thermophilus Cmr Complex

(A) Surface representation of the 3D reconstruction at 26 Å resolution from negatively stained Cmr complexes. The contour level is 1.5σ. (B) Representative reference-free 2D class averages (upper row), corresponding reprojections (middle row), and surface representation (lower row) of Cmr complexes. The scale bars represent 100 Å. (C) Representative raw particle images of Cmr complexes containing Cmr6-His_6 and labeled with NTA-nanogold particles. The scale bars represent 100 Å. (D) Segmentation of the TtCmr complex reconstruction at 22 Å resolution highlighting the ''sea worm'' architecture. Segmented regions are colored and labeled as follows: Cmr1/Cmr6 (orange), Cmr2 (purple), Cmr3 (green), Cmr4 (alternating light blue and gray), and Cmr5 (red). (E) The crystal structure of the homologous Cmr2-Cmr3 subcomplex from *P. furiosus* has been docked into the tail and is color coded and labeled as in (D).

change, proposed to create a docking site for the Cas3 nuclease [29]. The negative stain EM structure of the *Pseudomonas* type I-F Csy complex, a functional homolog of the *E. coli* Cascade, composed of Csy1, -2, -3, and -4 proteins and a stoichiometry of 1:1:6:1 with a 60 nt crRNA has revealed a crescent-shaped particle, the overall morphology of which is similar to that of the *E. coli* Cascade[30]. The *Sulfolobus* type I-A Cascade complex is composed of Csa2, Cas5a, Cas6, and Csa5 proteins and crRNA, forms a helical filament, resembling the *E. coli* Cascade-like arch-shaped particle

[31]. The *Bacillus halodurans* type I-C complex, also reveals a Cascade-like architecture, with a helical backbone of Cas7 proteins [32].

On the other hand, the *Sulfolobus* Cmr complex is composed of Cmr1, -2, -3, -4, -5, -6, and -7 proteins with a proposed stoichiometry of 1:1:1:1:1:1:6, part of which forms a ‘‘crab-claw’’ structure unlike the Cascade complex [14]. The overall structure of the *T*. *thermophilus* Cmr complex reveals no obvious similarity to the crab-claw structure of the *Sulfolobus* Cmr complex. Interestingly, it does have a repeating backbone that resembles the helical Cas7 backbone and architecture of the *E*. *coli* Cascade complex (Figures S7G–S7H). It might be that the crRNA is covered by the Cmr4 backbone of the *T*. *thermophilus* Cmr complex like the *E*. *coli* Cascade-crRNA complex [29]. The Cascade and Cmr complexes are classified into different types of the CRISPR-Cas systems, i.e., types I and III, respectively [6]. However, their architectural similarities suggest common ancestry of at least some of their core subunits [33]. Despite these similarities, Cascade(-like) complexes should be regarded as surveillance complexes, as nuclease activity (Cas3) is only recruited upon successful recognition of the protospacer, while type II and type III-B complexes have intrinsic nuclease activities, and should therefore be regarded as effector complexes.

In this study, we have demonstrated that the TtCmr complex harbors Mg^{2+}-dependent endoribonuclease activity (Figure 5A), which is specific for RNA targets complementary to the crRNAs that reside in the complex (Figure 5B). The Cmr2 subunit (NCBI accession number YP_145399) of the *T*. *thermophilus* Cmr complex does not have the N-terminal HD nuclease domain present in the Cmr2 proteins of the *Pyrococcus* and *Sulfolobus* Cmr complexes. This is consistent with the observation that the N-terminal HD domain is not involved in cleavage of target RNA [34]. Instead of a single cleavage site 14 nt upstream of the 3’ end of the crRNA observed in *P*. *furiosus* [7], the TtCmr complex has been demonstrated to cleave the target at multiple positions, spaced 6 nt apart (Figures 5 and 6). Our experiments

with the endogenous and reconstituted Cmr complexes with variable 3' crRNA ends showed that the degradation products are not a result of a 3' ruler mechanism (Figure 5D). Rather, our results demonstrated that cleavage is initiated at the 3' end of the RNA substrate, followed by sequential endonucleolytic activity every six nucleotides toward the 5' end (Figure 5C), in other words, a 5' ruler mechanism (Figure 6). Cmr may use multiple cleavage sites to increase the cleavage efficiency and catalytic turnover, and possibly to improve the chance of interference in the case of targets with partial mismatches, secondary structure, and/or associated proteins. Thus, this mechanism might render the Cmr complex more effective.

Although this study does not directly identify the ribonuclease of the complex, our data strongly suggest that Cmr4 fulfills this role. The main evidence for this is the observation that the Cmr complex probably harbors multiple active sites. If it is assumed that these sites are harbored in the same subunit, only Cmr4 and Cmr5 remain likely candidates. Cmr5 has previously been shown to be dispensable for activity, while Cmr4 was not [7]. In addition, the sequential 6 nt cleavages correspond well with the observed distance between two adjacent Cmr4 subunits. Interestingly, the 39, 33, 27, and 21 nt degradation products seem to be generated rather quickly when compared to the 15 nt degradation product (Figure 5C). We propose that these early products are generated by Cmr4, while another Cmr subunit would be responsible for the 15 nt product. Generating and testing catalytically dead mutants of Cmr4 (and other subunits) will therefore be an interesting challenge for future experiments.

EXPERIMENTAL PROCEDURES

Purification of the Cmr Complex

T. *thermophilus* HB8 cells producing the $(His)_6$-tagged Cmr complex were constructed and cultivated as described in the Supplemental Experimental Procedures. The Cmr complex was isolated using a nickel resin column, followed by anion exchange and gel-filtration chromatography steps as described in detail in the Supplemental Experimental Procedures.

Expression and Purification of Recombinant Cmr Proteins

The *cmr* genes were each amplified by genomic PCR and cloned under the control of the pET vector (Merck). The Cmr proteins were each expressed in the *E*. *coli* BL21(DE3) strain or its derivative with no tag, except for Cmr6, which has a $(His)_6$ tag at its N terminal followed by a tobacco etch virus (TEV) protease-recognition site [35]. We confirmed the N-terminal amino acid sequences of the purified proteins. The N-terminal methionine of the Cmr4 protein was deleted. The Cmr6 protein has G-H residues at its N terminal, which are derived from the TEV protease recognition site. The detailed experimental procedures are described in the Supplemental Experimental Procedures, and the SDS-PAGE analysis of the purified proteins is shown in Figure S1.

Native Mass Spectrometry

Cmr was buffer-exchanged to 0.175 M ammonium acetate (pH 8.0) at 4°C, using five sequential steps on a centrifugal filter with a cut-off of 10 kDa (Sartorius). The complex was sprayed at a concentration of 1 mM from borosilicate glass capillaries. An LCT electrospray time-of-flight or modified quadrupole time-of-flight instruments (both from Waters, UK) adjusted for optimal performance in high mass detection were used [36, 37]. Exact mass measurements of the individual Cmr proteins were acquired under denaturing conditions (1% and 10% formic acid). Reconstituted subcomplexes were heated at 65°C prior to buffer exchange (performed at

40°C). Instrument settings were as follows: needle voltage 1.3 kV, cone voltage 175 V, source pressure 9 mbar. Xenon was used as the collision gas for tandem mass spectrometric analysis at a pressure of 2 X 10^{-2} mbar. The collision voltage was varied between 10 and 200 V.

Deep Sequencing of the Bioinformatic Analysis of the Cmr-Bound crRNAs

Cmr-bound crRNAs were purified by phenol-chloroform isoamyl alcohol (PCI) extraction, as detailed in the Supplemental Experimental Procedures. The strand-specific RNA-seq library was prepared according to the directional mRNA-seq library preparation protocol provided by Illumina. This protocol involves utilizing the small RNA sample prep kit and the mRNA-seq library prep kit. crRNAs were treated with a phosphatase and kinase prior to ligation of RNA adapters to the 5′ and 3′ ends. The sequential ligation use of different adapters allowed for subsequent orientation of the sequencing reads. The ligated RNAs were then reverse transcribed and amplified by PCR. The library was sequenced with an Illumina Genome Analyzer IIx (Plateforme de Plateforme de Séquençage à Haut Débit Imagif, Gif-sur-Yvette, France). A total of 6,479,050 overlapping forward and reversed 74 nt Illumina reads were obtained and aligned against its mate using blast. Nonoverlapping pairs were removed (774,457). Mapping of the adaptor-stripped reads to the genome *T*. *thermophilus* HB8 was performed with Bowtie2 using the default settings [38]. Reads containing any insertions, deletions, or mismatches regarding to the reference genome were discarded. As a result, 5,704,593 (92.74%) of the remaining reads could be mapped against the reference genome. Visualization was performed using Excel and Matplotlib [39].

Reconstitution of the Cmr Complex

Synthetic 43 and 46 nt crRNAs corresponding to CRISPR4.5 were purchased from Integrated DNA Technologies (IDT), the sequences of which are listed in Table S3. Cmr complexes were reconstituted by preheating the individual

Cmr subunits (Cmr1-6) to 65°C, after which they were added together in a $Cmr1_1 2_1 3_1 4_4 5_3 6_1$-$crRNA_1$ molar ratio, followed by incubation for 10 min in at 65°C under gentle agitation.

RNA Activity Assays

RNA substrates for the activity assays were prepared by hybridizing two complementary DNA oligos, containing a SP6 promoter in front of the desired RNA sequence. A full description of all the oligos is given in Table S3. Internally, radiolabeled RNA was obtained by in vitro transcription of the hybrized oligos with SP6 RNA polymerase in the presence of α-^{32}P UTP, followed by denaturing gel purification. The 5′ nt target RNA sequence complementary to CRISPR-4.5 was purchased from IDT. This particular RNA was either 5′ or 3′ labeled using T4 polynucleotide kinase and ^{32}P γ-ATP or T4 RNA ligase 1 and ^{32}P pCp respectively, followed by denaturing gel purification. Target RNA substrates were incubated at 65°C with 5 nM endogenous Cmr complex or 63 nM reconstituted Cmr complex in RNA degradation buffer (20 mM Tris-HCl, pH 8, 150 mM NaCl, 10 mM DTT, 1 mM ATP, and 2 mM $MgCl_2$). Reactions were stopped by transferring the tubes to ice followed by the addition of formamide RNA loading buffer. Samples and a 5′ ^{32}P labeled RNA marker (Decade Markers, Ambion) were separated on a 20% polyacrylamide denaturing gel, containing 7 M urea and visualized by phosphorimaging.

Single-Particle Electron Microscopy and Analysis

A few microliters of 0.05 to 0.1 mg ml^{-1} purified protein solution was applied onto a carbon-coated grid and negatively stained with 2% uranyl acetate. We examined the sample grids with a JEM-2100 electron microscope (JEOL) with a LaB_6 gun operated at an accelerating voltage of 200 kV. Images were recorded on a slow-scan charge-coupled device (SSCCD) camera (MegaScan; Gatan, Pleasanton, CA) at a final magnification of 65,000 and at defocus settings of 8,900–27,000 Å. The magnification was calibrated from catalase crystals. For single-particle analysis of the negatively stained

particles, the EMAN (http://blake.bcm.edu/emanwiki/EMAN1) and EMAN2 (http://blake.bcm.edu/emanwiki/EMAN2) software suites were used for the following analysis. The contrast transfer function (CTF) was estimated using ''e2ctf.py.'' The phase reversal due to the CTF was corrected with ''applyctf.'' We manually picked particles from 105 EM images by using ''e2boxer.py.'' The samples might contain complexes lacking the Cmr1 subunit, but these complexes appeared to be less abundant (Figures 2 and 4). We selected particles with consistent dimensions, most of which represent the full Cmr complex. After a low-pass filter was applied, the set of particles were iteratively aligned with ''cenalignint.'' Three-dimensional maps were constructed from the aligned images by using ''startnrclasses'' and ''startAny.'' The structure was refined by ''refine.'' The 3D map obtained was used as a starting model for higher-resolution refinement, with no filter applied. The total number of molecular images included in the 3D reconstruction was 8,976, after a few dozen iterations starting from 9,105 molecules. The Fourier shell correlation (FSC) was calculated between two volumes, each generated from half of the data set. The resolution was taken to be the spatial frequency at which the FSC drops below 0.5. Full CTF correction produced a map with the identical structural features although the map appeared to be at lower resolution. We also examined TtCmr complexes with automated electron microscopy and carried out segmentation of the 3D reconstruction (Figures 7D and S7), the details of which are described in the Supplemental Experimental Procedures.

Gold Labeling

The Cmr complex with a C-terminally His_6-tagged Cmr6 was incubated at 4°C overnight with 250 nM nickel-nitrilotriacetic acid (Ni^{2+}-NTA) nanogold particles (5 nm; Nanoprobes) [40]. We examined the samples with the JEOL electron microscope as described above.

ACCESSION NUMBERS

The EM-derived density maps for the first structure and from the larger data set have been deposited in the EMDB under accession numbers EMD-5719 and EMD-2418, respectively. The deep sequencing data set of the Cmr-bound crRNAs are deposited in the SRA database under accession number SRP029745.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, three tables, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2013.09.013.

ACKNOWLEDGEMENTS

We thank Aimi Osaki for construction of the recombinant *T. thermophilus* strain, Kayoko Matsumoto and Toshi Arima for purification of the Cmr proteins, Drs. Kwang Kim and Seiki Kuramitsu for identification of the Cmr subunits by the MS/MS analysis, and Aaron Jansen for his help with the activity assays. E. Nogales, H.-W. Wang, P. Grob, and T. Houweling are acknowledged for assistance with EM and image-processing. This work was supported by a Grant-in-Aid for Scientific Research (C), 25440013, from the Ministry of Education, Culture, Sports, Science and Technology, Japan (to A.S.), and by an ALW grant (820.02.003 to J.v.d.O.) from the Netherlands Organization for Scientific Research (N.W.O.). E.v.D., A.B., and A.J.R.H. were supported by the Netherlands Proteomics Centre. D.W.T. is a National Science Foundation Graduate Research Fellow, and J.A.D. is a Howard Hughes Medical Institute Investigator.

REFERENCES

1. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-1712.
2. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-964.
3. Horvath, P. and R. Barrangou, *CRISPR/Cas, the immune system of bacteria and archaea.* Science, 2010. **327**(5962): p. 167-170.
4. Karginov, F.V. and G.J. Hannon, *The CRISPR system: small RNA-guided defense in bacteria and archaea.* Molecular Cell, 2010. **37**(1): p. 7-19.
5. Grissa, I., G. Vergnaud, and C. Pourcel, *The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats.* BMC Bioinformatics, 2007. **8**: p. 172.
6. Makarova, K.S., et al., *Evolution and classification of the CRISPR-Cas systems.* Nat Rev Microbiol, 2011. **9**(6): p. 467-477.
7. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex.* Cell, 2009. **139**(5): p. 945-956.
8. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.* Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-E2586.
9. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-821.
10. Carte, J., et al., *Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes.* Genes Dev, 2008. **22**(24): p. 3489-3496.
11. Scholz, I., et al., *CRISPR-Cas systems in the cyanobacterium Synechocystis sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein.* PLoS One, 2013. **8**(2): p. e56470.
12. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA.* Science, 2008. **322**(5909): p. 1843-1845.
13. Deng, L., et al., *A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus.* Mol Microbiol, 2013. **87**(5): p. 1088-1099.
14. Zhang, J., et al., *Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity.* Molecular Cell, 2012. **45**(3): p. 303-313.
15. Agari, Y., et al., *Transcription profile of Thermus thermophilus CRISPR systems after phage infection.* J Mol Biol, 2010. **395**(2): p. 270-281.
16. Juranek, S., et al., *A genome-wide view of the expression and processing patterns of Thermus thermophilus HB8 CRISPR RNAs.* RNA, 2012. **18**(4): p. 7837-94.
17. Shinkai, A., et al., *Transcription activation mediated by a cyclic AMP receptor protein from Thermus thermophilus HB8.* J Bacteriol, 2007. **189**(10): p. 3891-901.
18. van Duijn, E., et al., *Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced shot-palindromic-repeats (CRISPR)-associated protein complexes from Escherichia coli and Pseudomonas aeruginosa.* Mol Cell Proteomics, 2012. **11**(11): p. 1430-1441.
19. Heck, A.J., *Native mass spectrometry: a bridge between interactomics and structural biology.* Nat Methods, 2008. **5**(11): p. 927-933.
20. Jore, M.M., et al., *Structural basis for CRISPR RNA-guided DNA recognition by Cascade.* Nat Struct Mol Biol, 2011. **18**(5): p. 529-536.
21. Osawa, T., H. Inanaga, and T. Numata, *Crystal structure of the Cmr2-Cmr3 subcomplex in the CRISPR-Cas RNA silencing effector complex.* J Mol Biol, 2013. **425**(20): p. 3811-3823.
22. Shao, Y., et al., *Structure of the Cmr2-Cmr3 subcomplex of the Cmr RNA silencing complex.* Structure, 2013. **21**(3): p. 376-384.
23. Westra, E.R., et al., *The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity.* Annu Rev Genet, 2012. **46**: p. 311-339.
24. Hale, C.R., et al., *Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs.* Molecular Cell, 2012. **45**(3): p. 292-302.
25. Wang, R., et al., *Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage.* Structure, 2011. **19**(2): p. 257-264.
26. Hatoum-Aslan, A., et al., *A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs.* J Biol Chem, 2013. **288**(39): p. 27888-27897.
27. Pougach, K., et al., *Transcription, processing and function of CRISPR cassettes in Escherichia coli.* Mol Microbiol, 2010. **77**(6): p. 1367-1379.
28. Deng, L., et al., *Modulation of CRISPR locus transcription by the repeat-binding protein Cbp1 in Sulfolobus.* Nucleic Acids Res, 2012. **40**(6): p. 2470-2480.
29. Wiedenheft, B., et al., *Structures of the RNA-guided surveillance complex from a bacterial immune system.* Nature, 2011a. **477**(7365): p. 486-489.
30. Wiedenheft, B., et al., *RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions.* Proc Natl Acad Sci U S A, 2011b. **108**(25): p. 10092-10097.
31. Lintner, N.G., et al., *Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE).* J Biol Chem, 2011. **286**(24): p. 21643-21656.
32. Nam, K.H., et al., *Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system.* Structure, 2012. **20**(9): p. 1574-1584.

33. Reeks, J., J.H. Naismith, and M.F. White, *CRISPR interference: a structural perspective.* Biochem J, 2013. **453**(2): p. 155-166.
34. Cocozaki, A.I., et al., *Structure of the Cmr2 subunit of the CRISPR-Cas RNA silencing complex.* Structure, 2012. **20**(3): p. 545-553.
35. Fang, L., et al., *An improved strategy for high-level production of TEV protease in Escherichia coli and its purification and characterization.* Protein Expr Purif, 2007. **51**(1): p. 102-109.
36. Tahallah, N., et al., *The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument.* Rapid Commun Mass Spectrom, 2001. **15**(8): p. 596-601.
37. van den Heuvel, R.H., et al., *Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry.* Anal Chem, 2006. **78**(21): p. 7473-7483.
38. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-359.
39. Hunter, J.D., *Matplotlib: A 2D Graphics Environment.* Computing in Science & Engineering, 2007. **9**(3): p. 90-95.
40. Hainfeld, J.F., et al., *Ni-NTA-gold clusters target His-tagged proteins.* J Struct Biol, 1999. **127**(2): p. 185-198.

Chapter 4

Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning

This chapter has been published as:

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

*contributed equally

ABSTRACT

Adaptive immunity in bacteria involves RNA-guided surveillance complexes that use CRISPR (clustered regularly interspaced short palindromic repeats)-associated (Cas) proteins together with CRISPR RNAs (crRNAs) to target invasive nucleic acids for degradation. Whereas type I and type II CRISPR-Cas surveillance complexes target double-stranded DNA, type III complexes target single-stranded RNA. Near-atomic resolution cryo–electron microscopy reconstructions of native type III Cmr (CRISPR RAMP module) complexes in the absence and presence of target RNA reveal a helical protein arrangement that positions the crRNA for substrate binding. Thumblike β hairpins intercalate between segments of duplexed crRNA:target RNA to facilitate cleavage of the target at 6-nucleotide intervals. The Cmr complex is architecturally similar to the type I CRISPR-Cascade complex, suggesting divergent evolution of these immune systems from a common ancestor.

RESULTS

Bacteria and archaea defend themselves against infection using adaptive immune systems comprising CRISPR (clustered regularly interspaced short palindromic repeats) arrays and CRISPR-associated (Cas) genes [1]. A defining feature of CRISPR-Cas systems is the use of Cas proteins in complex with small CRISPR RNAs (crRNAs) to identify and cleave complementary target sequences in foreign DNA [2, 3]. Whereas type I and type II CRISPR–Cas systems recognize target sequences in double-helical DNA that is locally unwound to enable DNA target strand cleavage [4, 5], type III systems bind and cleave single-stranded RNA target sequences [6, 7].

The effector complex of the type III system from *Thermus thermophilus* (Cmr) is a 12-subunit assembly composed of six Cmr subunits (Cmr1–6) and a crRNA with a stoichiometry of $Cmr1_1 2_1 3_1 4_4 5_3 6_1$: $crRNA_1$ [7]. The Cmr complex binds to target RNA that is complementary to the bound 40- or 46-nucleotide (nt) crRNA and cleaves the target at 6-nt intervals measured from the 5′ end of the crRNA sequence [7, 8]. Although structural studies revealed an overall capsule-like architecture of the Cmr complex [7], the molecular basis of subunit assembly, crRNA binding, and single-stranded RNA (ssRNA) target recognition and cleavage by the intact surveillance complex remains unknown.

We performed cryo–electron microscopy (cryo-EM) of the intact ~350-kD Cmr complex in the absence and presence of target ssRNA. We purified endogenous apo-Cmr (containing a crRNA) and used this sample for stepwise assembly with a 56-nt biotinylated ssRNA target followed by purification using streptavidin affinity chromatography. Cryo-EM micrographs of frozen-hydrated samples of both apo-Cmr and target-bound Cmr showed monodisperse particles with wormlike features (fig. S1). Using Leginon [9], we acquired ~7000 and ~4000 micrographs and automatically picked ~700,000 and ~300,000 apo- and target-bound Cmr particles,

respectively [10]. After three-dimensional (3D) classification and single-particle reconstruction (supplementary materials and methods) [11], we obtained structures of intact apo-Cmr and target-bound Cmr at ~4.1 and 4.4 Å resolution (figs. S1 and S2) from a final set of 250,000 and 175,000 particles, respectively. Additionally, we obtained the structure of a smaller apo-Cmr species revealed during our 3D classification at ~4.4 Å resolution from a second class of ~100,000 particles.

The structure of intact apo-Cmr resembles a capsule in which a central, double-helical core of four Cmr4 subunits and three Cmr5 subunits is capped by a Cmr2-Cmr3 heterodimer at one end and Cmr1-Cmr6 at the other (Fig. 1A). The 5′-handle of the crRNA, an invariant sequence derived from the repeat, is held in the Cmr2-Cmr3 heterodimer. An α-helical bundle within Cmr2 makes extensive contacts with the bottom Cmr5 subunit, while the body of Cmr2 engages Cmr3. The architecture of the smaller apo-Cmr species is similar to that of intact apo-Cmr and maintains the same intersubunit interactions, despite lacking one Cmr4–Cmr5 subcomplex and being shorter by ~25 Å in the longest direction (Fig. 1B). This apo-Cmr complex likely represents the shorter (40-nt) crRNA-bound species. There are several segments of additional rod-shaped density along the helical backbone of the complex engaged with target ssRNA that are absent from the apo-Cmr structure (Fig. 1, C and D).

To analyze whether the helical geometry in the smaller apo-Cmr structure is perturbed, we aligned the Cmr2-Cmr3 base of both apo-Cmr structures and found that the Cmr4 backbone subunits from this smaller species fit perfectly into their respective subunits in the intact complex (Fig. 1E). Furthermore, when we aligned the two apo-Cmr structures based on the Cmr1-Cmr6 head, the equivalent Cmr4 backbone subunits were again superimposable (Fig. 1E), suggesting that the overall geometry of the complex and the nature of the subunit interactions are preserved in the smaller apo-Cmr structure. To study potential conformational changes that result from ssRNA target binding by the Cmr complex, we aligned the Cmr4

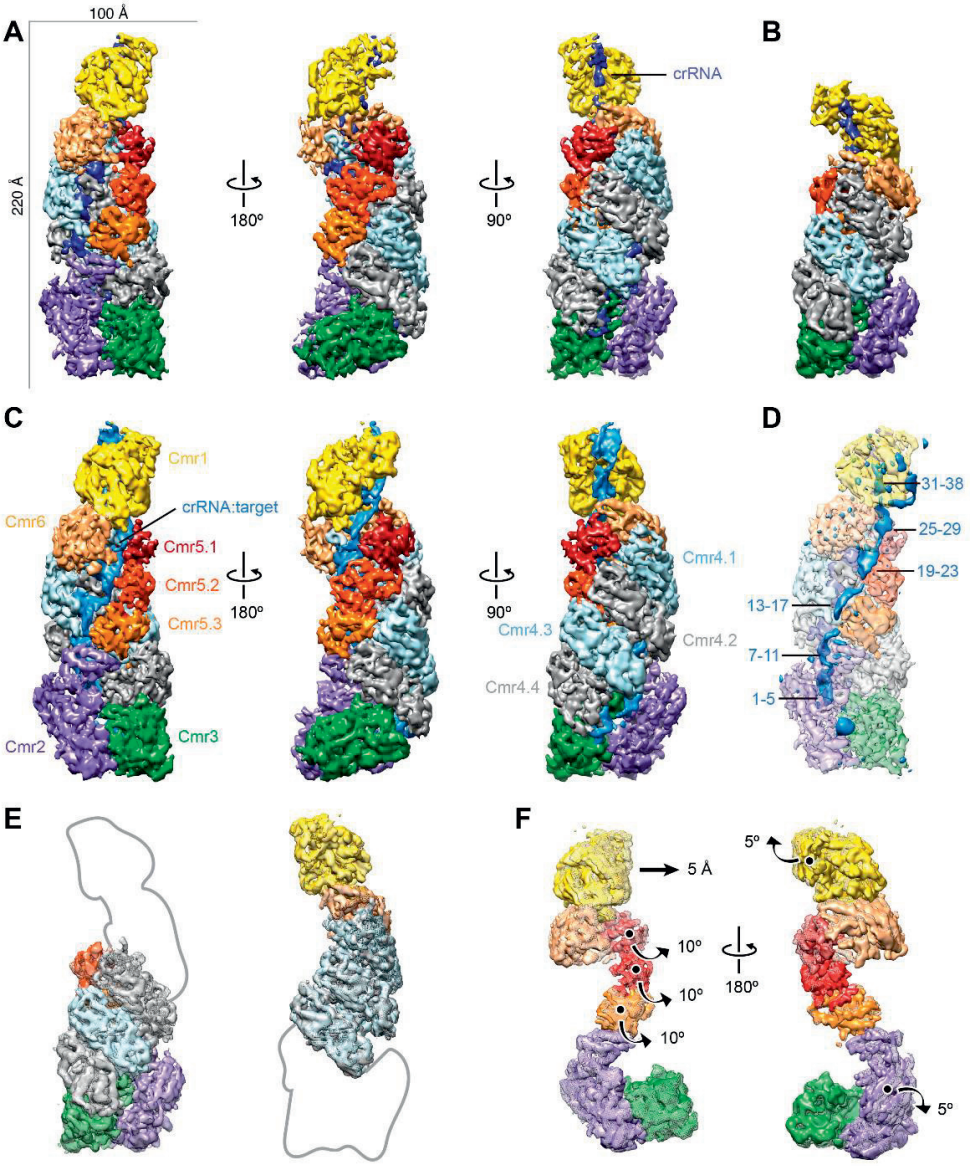


Figure 1. Architecture of the native crRNA-bound (apo) and ssRNA target-bound Cmr.

(A to C) Cryo-EM reconstructions of intact apo-Cmr (crRNA-bound) (A), a smaller apo-Cmr (B), and ssRNA target-bound Cmr (C) at 4.1, 4.5, and 4.4 Å resolution (using the 0.143 "gold standard" Fourier shell correlation criterion), respectively. Subunits are segmented and colored as indicated. (D) Difference map between intact apo-Cmr and target- bound Cmr at 10s (solid, blue) superimposed on the apo-Cmr structure (transparent). (E) Aligning the smaller (surface) and intact (mesh) apo-Cmr complexes based on Cmr2–Cmr3 (left) or Cmr1– Cmr6 (right) shows that the helical geometry is pre-served. (F) Alignment of apo-Cmr (surface) and target-bound Cmr complexes (transparent mesh) based on the Cmr4 backbone (removed for clarity). (See also movies S1 and S2.)

backbone from both apo-Cmr and target-bound Cmr structures, whose position remains relatively unchanged (cross-correlation coefficient of 0.92). This superposition shows that the remaining subunits undergo a concerted rearrangement. Upon substrate binding, the Cmr1 and Cmr2 subunits at either end rotate by ~5° in opposite directions (in addition to a 5 Å translation in the head) along axes perpendicular to the long axis of the complex (Fig. 1F). The three-subunit Cmr5 filament opens (~10° rotation) away from the center of the complex by this ratchet-like motion at the ends (Fig. 1F), exposing the crRNA and forming an elongated channel to accommodate the crRNA:target ssRNA duplex.

We observed long, thumblike β-strand extensions emerging from the palm of each Cmr4 subunit and engaged with the adjacent subunit (Fig. 2A), a feature that is reminiscent of the interactions in the *Escherichia coli* (Ec) type I CRISPR-Cascade DNA targeting complex [12, 13]. Docking of the crystal structure of *Pyrococcus furiosus* (Pf) Cmr4 [Protein Data Bank (PDB) accession code 4W8W] [14] into the density for each individual Cmr4 subunit of the Cmr4 filament shows that nearly the entire density for the thumb is unaccounted for by this crystal structure. When we aligned the structure of PfCmr4 with the core (RNA recognition motif) of the structure of Cascade EcCas7 [12] (PDB 1VY9) (root mean square deviation 2.087 Å) (fig. S3), the segments of the structure immediately preceding the unresolved stretch of residues in PfCmr4 (residues 206 to 227) align well with the segments where the thumb extends from the Cas7 structure (residues 198 to 217) and fit with the β-strand feature in our cryo-EM density for this subunit (fig. S3). We combined these two sets of atomic coordinates to create a complete, chimeric homology model of Cmr4 and used Rosetta [15] to relax the resulting atomic model into the cryo-EM density (Fig. 2B). This model of an individual Cmr4 subunit can be docked unambiguously into each of the Cmr4 backbone subunits to show a thumb-to-palm interaction network (Fig. 2B) mediated by association of the β

hairpin of the lower Cmr4 with the a1 helix of the neighboring Cmr4 subunit (Fig. 2C).

The Cmr4 backbone in the context of the ssRNA target-bound complex shows segments of ~20 Å-wide additional density anchored rigidly by the Cmr4 backbone and the other helical array of Cmr5 subunits (Fig. 3A). The thumblike β-hairpin domains of the Cmr4 filament intercalate between segments of duplexed crRNA:target RNA, distorting the crRNA:target RNA duplex after every 5-base pair segment and disrupting the formation of an extended A-form double helix (Fig. 3B). This arrangement places regions of the distorted or kinked ssRNA target in proximity to an adjacent loop density containing several of the catalytic residues (H16 and D27) [14] from the Cmr4 subunit, positioning it for productive cleavage (Fig. 3C). This rearranging of target nucleic acid is reminiscent of that occurring in Cascade. Indeed, the atomic coordinates of the nucleic acid from the ssDNA target-bound Cascade (PDB 4QYZ) [12, 13] can be accommodated easily within this additional density, placing the flipped-out base of the target strand near the catalytic loop of Cmr4 (Fig. 3, D and E).

The Cmr effector complex cleaves target ssRNAs at five sites *in vitro* [7], despite containing only four Cmr4 subunits. Reanalyzing our structures, we noticed a thumblike extension, nearly identical to those observed in individual Cmr4 subunits, originating in Cmr6 (Fig. 3F). In the context of the target-bound Cmr structure, this thumb places the target strand in a position for cleavage of the 5′-most site on the target RNA (Fig. 3G). Similarly, a thumblike domain in Cmr3 stretches into the palm of the bottom Cmr4 subunit (Fig. 3H), which stacks on top of the 5′-handle and scaffolds the 3′-most discontinuous segment of crRNA:target (Fig. 3I).

Our previous work revealed that the type I CRISPR-Cascade complex undergoes a concerted rearrangement upon target binding [16, 17]. Here, we show that the type III CRISPR-Cmr complex undergoes an analogous

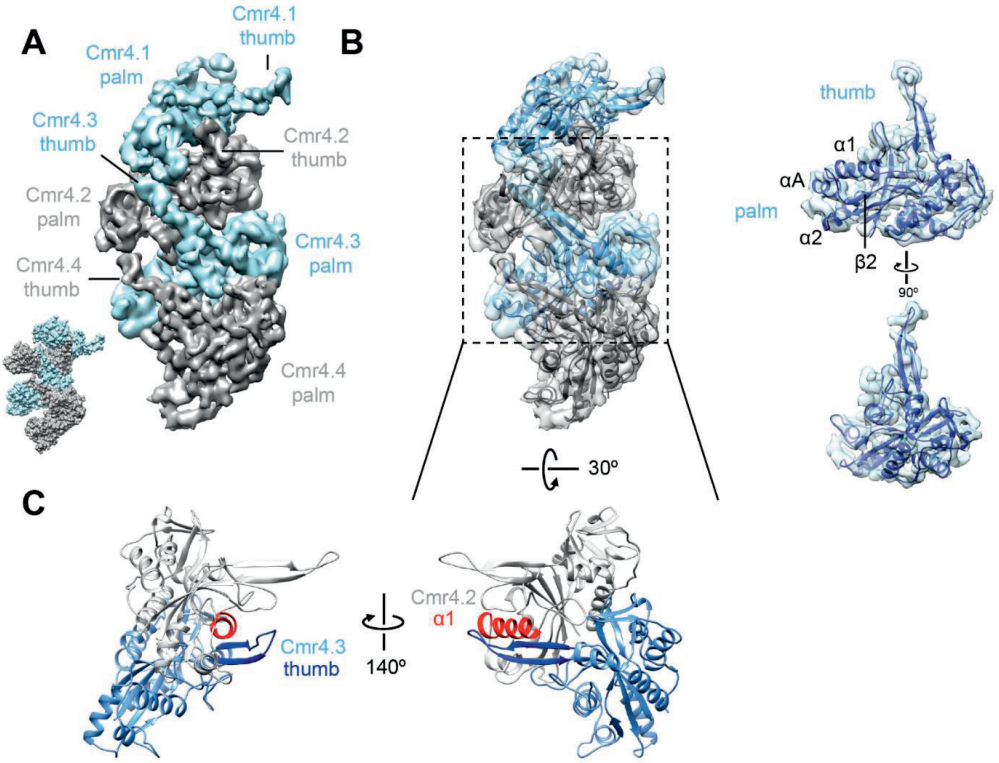


Figure 2. Thumb-to-palm interactions between adjacent Cmr4 subunits form the Cmr backbone.

(A) The Cmr4 backbone subunits (other subunits have been removed for clarity), form a helical arrangement by interaction of a thumblike density from one subunit with the palm of the subunit above. This layout is similar to the Cas7 subunits (PDB 1VY9) (12) within *E. coli* Cascade (inset). (B) Homology model of Cmr4 (right) and the Cmr4 helical oligomer (left) shows that the Cmr4 thumbs accommodate the β-hairpin extension. (C) Close-up view of Cmr4.2 and 4.3 showing that the β hairpin of Cmr4 (dark blue) associates with the a1 helix (red) of the neighboring Cmr4 subunit.

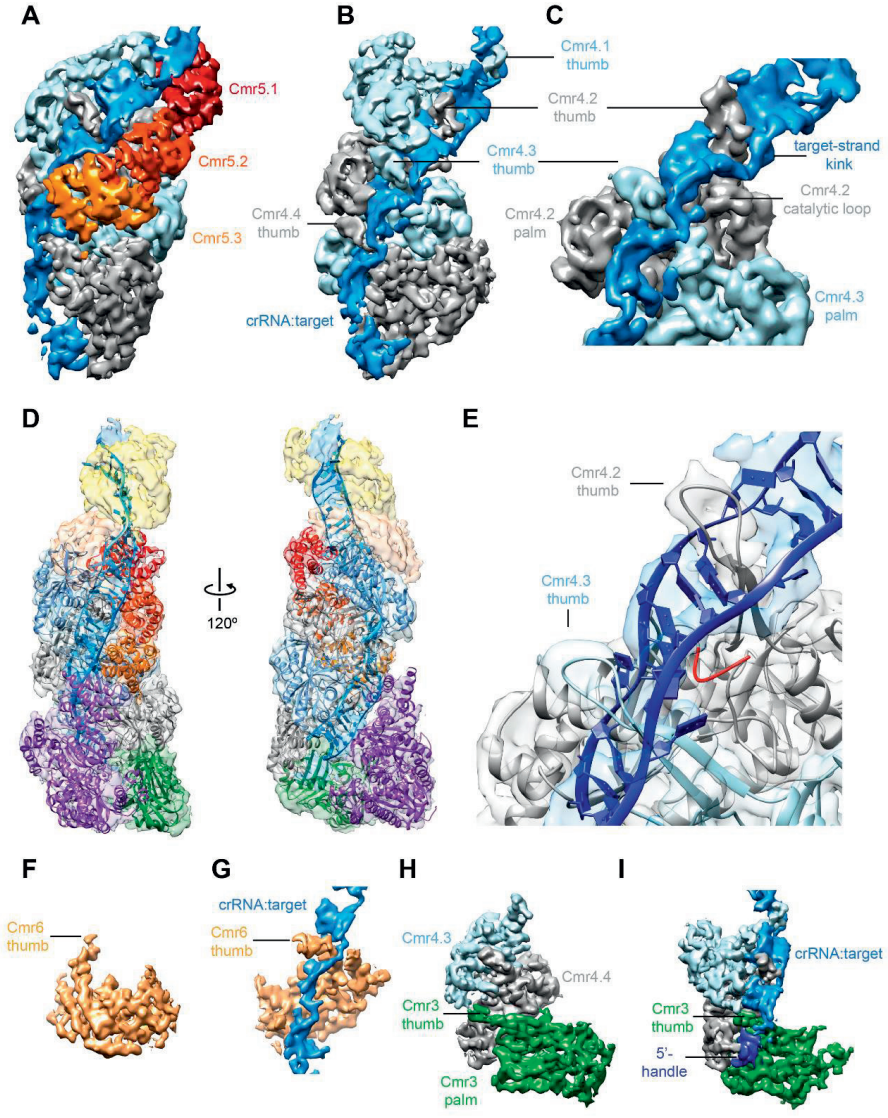


Figure 3. The Cmr4 backbone and Cmr5 subunits position target ssRNA for segmented cleavage.

(A) The channel formed between Cmr4 and Cmr5 backbones creates the binding cleft for target RNA. (B) Cmr4 β hairpins (thumbs) intercalate after every fifth base pair of duplex crRNA:target RNA, disrupting the helix. (C) Expanded view of (B) shows how the thumblike domain of the lower Cmr4 positions the kinked target near the catalytic loop. (D) Pseudo-atomic model created by docking available crystal structures of Cmr subunits and our homology model of Cmr4 into the target-bound Cmr structure. (E) Model of target recognition and cleavage shows that the loop (red) of the homology model (presented here) containing catalytic residues H16 and D27 previously identified (14) is positioned near the target. (F and G) The Cmr6 subunit (gold) also contains a long thumblike extension (F), which disrupts base-pairing between the crRNA and target RNA (G) for the 5′-most cleavage event. (H and I) The thumb of the Cmr3 subunit (H) engages the crRNA 5′-handle and bottom target:crRNA segment (I). (See also movie S3 and supplementary materials and methods.)

conformational change upon target recognition. Whereas the rearrangement in Cascade most likely permits the docking of the trans-acting Cas3 nuclease, in type III complexes this rearrangement likely regulates propagation of crRNA:target base-pairing into duplex segments and substrate recognition by the thumbs of Cmr4 subunits. Because the amino acid sequence identity between *E. coli* Cas7 and *T. thermophilus* Cmr4 is ~22% and they share conserved structural features (palm and thumb), at least the core of type I and type III CRISPR-Cas surveillance complexes likely diverged from a common ancestor. Whereas type I evolved thumblike domains for recognition of the nontarget strand, type III repurposed these thumbs for distorting a ssRNA substrate for cleavage. We also show here that apo-Cmr complexes of different sizes (corresponding to the presence or absence of one Cmr4–Cmr5) preserve their overall geometry, suggesting that this architecture is used for productive substrate recognition by crRNAs of different sizes.

These type I and type III complexes both use thumb-mediated local disruption of duplex geometry in their interactions with substrate sequences, leading to a lack of continuous double-helix formation between guide RNA and target strands. That RecA employs similar discontiguous DNA-DNA interactions for homology searches [18] hints at a common mode of substrate recognition among genome surveillance complexes. Although type II CRISPR-Cas9-RNA-ssDNA crystal structures contain a canonical crRNA-DNA helix [19, 20], crRNA could form a discontinuous helix with one strand of dsDNA targets during sequence interrogation, consistent with tolerance of large 5’ end extensions on the crRNA [21]. In the related type III CRISPR-Csm complex, discontinuous helix formation might occur during association with topologically constrained R loops formed during transcription [22].

ACKNOWLEDGEMENTS

The structures of intact apo-Cmr, smaller apo-Cmr, and target-bound Cmr have been deposited into the EMDataBank with accession codes EMD-2898, EMD-2899, and EMD-2900, respectively. We thank R. Louder, S. Howes, E. Kellogg, R. Zhang, P. Grob, Y. He, T. Houweling, Z. Yu, and M. J. de la Cruz for expert electron microscopy assistance. D.W.T. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2218-15). R.H.J.S. was supported by the University of Otago's Division of Health Sciences Career Development postdoctoral fellowship. Y.Z. and J.v.d.O. received financial support from the Netherlands Organisation for Scientific Research (NWO), via a Gravitation grant to the Soehngen Institute for Anaerobic Microbiology (024.002.002 to W.d.V.) and an ALW-TOP project (854.10.003), respectively. This work was supported in part by Japan Society for the Promotion of Science KAKENHI grant 25440013 (to A.S.). J.A.D. and E.N. are Howard Hughes Medical Institute Investigators. The *T. thermophilus* Cmr complex and *T. thermophilus* strain–producing Cmr6-His protein are available from A.S. under a material transfer agreement with RIKEN.

SUPPLEMENTARY MATERIALS

Supplemental Information includes Materials and methods, three figures, one tables, three movies, and Supplemental References can be found with this article online at www.sciencemag.org/content/348/6234/581/suppl/DC1

REFERENCES

1. Wiedenheft, B., S.H. Sternberg, and J.A. Doudna, *RNA-guided genetic silencing systems in bacteria and archaea.* Nature, 2012. **482**(7385): p. 331-338.
2. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-1712.
3. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-964.
4. Jore, M.M., et al., *Structural basis for CRISPR RNA-guided DNA recognition by Cascade.* Nat Struct Mol Biol, 2011. **18**(5): p. 529-536.
5. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-821.
6. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex.* Cell, 2009. **139**(5): p. 945-956.
7. Staals, R.H., et al., *Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus.* Molecular Cell, 2013. **52**(1): p. 135-145.
8. Hale, C.R., et al., *Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex.* Genes Dev, 2014. **28**(21): p. 2432-2443.
9. Suloway, C., et al., *Automated molecular microscopy: the new Leginon system.* J Struct Biol, 2005. **151**(1): p. 41-60.
10. Lander, G.C., et al., *Appion: an integrated, database-driven pipeline to facilitate EM image processing.* J Struct Biol, 2009. **166**(1): p. 95-102.
11. Scheres, S.H., *RELION: implementation of a Bayesian approach to cryo-EM structure determination.* J Struct Biol, 2012. **180**(3): p. 519-530.
12. Jackson, R.N., et al., *Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli.* Science, 2014. **345**(6203): p. 1473-1479.
13. Mulepati, S., A. Heroux, and S. Bailey, *Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target.* Science, 2014. **345**(6203): p. 1479-1484.
14. Benda, C., et al., *Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4.* Molecular Cell, 2014. **56**(1): p. 43-54.
15. DiMaio, F., et al., *Refinement of protein structures into low-resolution density maps using rosetta.* J Mol Biol, 2009. **392**(1): p. 181-190.
16. Wiedenheft, B., et al., *Structures of the RNA-guided surveillance complex from a bacterial immune system.* Nature, 2011a. **477**(7365): p. 486-489.
17. Hochstrasser, M.L., et al., *CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference.* Proc Natl Acad Sci U S A, 2014. **111**(18): p. 6618-6623.
18. Chen, Z., H. Yang, and N.P. Pavletich, *Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures.* Nature, 2008. **453**(7194): p. 489-494.
19. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease.* Nature, 2014. **513**(7519): p. 569-573.
20. Nishimasu, H., et al., *Crystal structure of Cas9 in complex with guide RNA and target DNA.* Cell, 2014. **156**(5): p. 935-949.
21. Ryan, O.W., et al., *Selection of chromosomal DNA libraries using a multiplex CRISPR system.* Elife, 2014. **3**.
22. Goldberg, G.W., et al., *Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting.* Nature, 2014. **514**(7524): p. 633-637.

Chapter 5

A flexible seed sequence regulates targeting by the type III CRISPR-Cmr complex

Manuscript in preparation:

Yifan Zhu*, David W. Taylor*, Eline Stroobach, Sanne Klompe, Nirajan Neupane, Anna Sobieraj, Willem M. de Vos, Stan J.J. Brouns, Akeo Shinkai, Jennifer A. Doudna, John van der Oost, Raymond H.J. Staals

*contributed equally

ABSTRACT

CRISPR-Cas systems in prokaryotes provide resistance against mobile genetic elements. Despite the conserved architecture between type I and type III effector complexes, these complexes have very different characteristics in terms of substrate specificity (DNA and/or RNA), crRNA processing and the presence and location of the seed region. Previously, we demonstrated that crRNA processing in type III systems involves a secondary 3' processing step that produces crRNAs of variable lengths. In turn, these differently-sized crRNAs determine the size of the resulting effector complex. In this study, we show that unlike the 5' seed region in type I systems, the seed-like region of the type III-B effector complex from *Thermus thermophilus* (TtCmr) resides at the 3' end. Intriguingly, this seed region shifts towards the 5' end as the length of the crRNA (and therefore, the size of the effector complex) decreases. The positioning of the seed-like sequence is likely important for interaction with the 5' end of a nascent transcript targeted by TtCmr, and its ability to utilize different length guides and flexible seed-region may contribute to avoiding phage escape.

INTRODUCTION

As a widespread prokaryotic adaptive immune system, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated) systems target and cleave mobile genetic elements (MGEs) [1-4]. A CRISPR array is composed of alternating repeat and spacer sequences, with the repeats consisting of the multiple instances of the same sequence and the spacers consisting of highly variable sequence fragments acquired from MGEs, such as invading plasmids and viruses [5-7]. The CRISPR array is accompanied by a set of CRISPR-associated (*cas*) genes encoding the Cas proteins.

Based on the presence of signature Cas proteins, CRISPR-Cas systems have been divided into two major classes (class 1 and 2), six types (I to VI) and multiple sub-types [1, 2, 8, 9]. CRISPR-associated defense proceeds via three mechanistic stages: acquisition, expression and interference [10]. The acquisition of new CRISPR spacers has been demonstrated to resemble recombination by retroviral integrases/transposases and requires the Cas1-Cas2 complex [3, 11, 12]. Following expression of a long precursor-CRISPR transcript (pre-crRNA), initial processing occurs in the repeat sequences by ribonucleases such as Cas6 (type I and III), RNase III (type II) or by a CRISPR effector nuclease (type V and VI, Cas12 and Cas13 respectively). In some CRISPR-Cas systems, a second maturation event occurs where the crRNA is trimmed either at its 5’ (type II) or at 3’ end (type III and some type I systems) [13, 14]. In the interference stage, Cas protein(s) assembled around the matured crRNA guide form a ribonucleoprotein (RNP) complex that specifically binds and degrades complementary target sequences (protospacers).

Unlike other CRISPR-Cas systems that exclusively target either DNA (type I, II and V) or RNA (type VI), many type III systems are endowed with the capacity to target both nucleic acids [15-21]. Structural and biochemical analyses of the multi-subunit type III complexes have revealed the subunits

responsible for these activities [19, 22-28]. These activities are activated in a coordinated fashion, starting with the binding of a target RNA complementary to the effector-bound crRNA. Target RNA binding activates the RNase activity of Cas7, which is present in multiple copies and constitutes the backbone of type III complexes. As such, type III interference complexes have multiple active sites, cleaving the target RNA at 6 nt intervals [16, 19, 22, 29, 30]. 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 [18, 20, 31], while activation of its Palm domain was recently shown to confer oligoadenylate cyclase activity, producing oligo(A) second messenger molecules that allosterically activate a stand-alone, sequence-nonspecific RNase (Csm6 or Csx1 in type III-A or type III-B systems respectively) to provide a robust, transcription-dependent, interference response in type III-A systems [32-35].

Previously, we characterized the structural and enzymatic features of the endogenous type III-B Cmr complex from *T. thermophilus* (TtCmr) [22, 27]. We showed that TtCmr adopts a structure similar to type I (Cascade) complexes: a Cas7 (Cmr4) backbone with multiple copies of the small subunit Cas11 (Cmr5). The complex is capped at one end by the large subunit Cas10 and Cas5 (Cmr3) and is capped at the other end with 2 more Cas7-like subunits (Cmr1 and Cmr6) [27]. The mature crRNA runs along the Cas7 backbone of the complex, with its 5′ repeat-derived end anchored by Cas10/Cas5, and its 3′ end positioned within the Cmr1/Cmr6 end [24, 29, 36]. 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, which produces the non-variable, repeat-derived 5′ end, also referred to as the 5′ handle. 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, as is explained below.

Firstly, the crRNA-content of the endogenous TtCmr complex showed that it indeed co-purifies with mature crRNAs of different sizes (due to their variable 3′ ends), with a distinct 6-nt pattern: 34, 40 & 46 nt [22]. Secondly, our cryo-EM structures revealed that the native population of TtCmr complexes consisted 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). Intriguingly, the distance between these segments shortens the backbone with ~25 Å, which is consistent with a length of ~6 to 7 bp of A-form RNA [27]. Taken together, these data strongly suggest 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 nuclease, such as PNPase, that shortens the unprotected, protruding 3′ end of the bound crRNA [37].

In this study, we set out to understand the biological significance of these differently sized Cmr complexes, revealing an unexpected flexible seed region that might be important for these systems to prevent phage escapees.

RESULTS

RNA cleavage activity of differently sized TtCmr complexes.

Our previous studies showed that native Cmr complexes purified from *T. thermophilus* HB8 were loaded with range of mature crRNA guides of different lengths, and that the crRNA-4.5 (*T. thermophilus* CRISPR array 4, spacer 5) was most abundant [22]. The endogenous Cmr complex specifically cleaves complementary 5′ labeled target RNAs (4.5 target RNA) at 6 nt intervals, resulting in 5′ labeled degradation fragments of mostly 21, 27, 33 and 39 nucleotides (Figure 1B). However, the heterogeneous nature of the crRNA-content of the endogenous Cmr complexes [22, 27], complicates the interpretation of these results. Therefore, to further probe the mechanism of target RNA cleavage, we replaced the endogenous Cmr

complex with a reconstituted Cmr complex bound to a single crRNA (crRNA-4.5) of different lengths. We chose to include crRNA lengths of 34 (TtCmr-34), 40 (TtCmr-40) and 46 nt (TtCmr-46) based on their abundance in our previous RNA sequencing data [22]. Opposed to the mixture of 5′ labeled degradation products observed with the endogenous complex, each of the reconstituted complexes yielded defined degradation products of mostly one size (Figure 1C). This is consistent with the idea that the length of the complex is determined by the length of the crRNA, with bigger complexes (e.g. TtCmr-46) cleaving more closely to the 5′ (labeled) end of the target RNA. Smaller complexes (i.e. TtCmr-40 and TtCmr-34) lack one or two Cas7–Cas11 (Cmr4-Cmr5) backbone segments respectively and therefore cleave the target RNA at 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.

The seed sequence of TtCmr

To investigate the significance of these type III complexes with different stoichiometries, we set up activity assays to probe for differences in targeting and seed requirements. In the structurally similar type I effector complexes (i.e. the Cascade complex) targeting is governed by two factors: the PAM (protospacer adjacent motif) and the seed [38-40]. In type III systems however, self-discrimination is conferred by a so-called rPAM, which checks for complementarity between the 8 nt 5′ handle of the crRNA (referred to as nucleotides -8 to -1) and the corresponding 3′ region flanking the protospacer [18, 20, 41]. In case of complementarity, the DNase activity of Cas10 is abolished. Since TtCmr is devoid of DNase activity [22], due to a N-terminally truncated Cas10 protein lacking the HD domain (data not shown), we tested whether RNase activity is affected by target RNAs with matches to the 5′ handle of the crRNA (Figure 2A). However, the activity assays showed that these substrates had no discernable

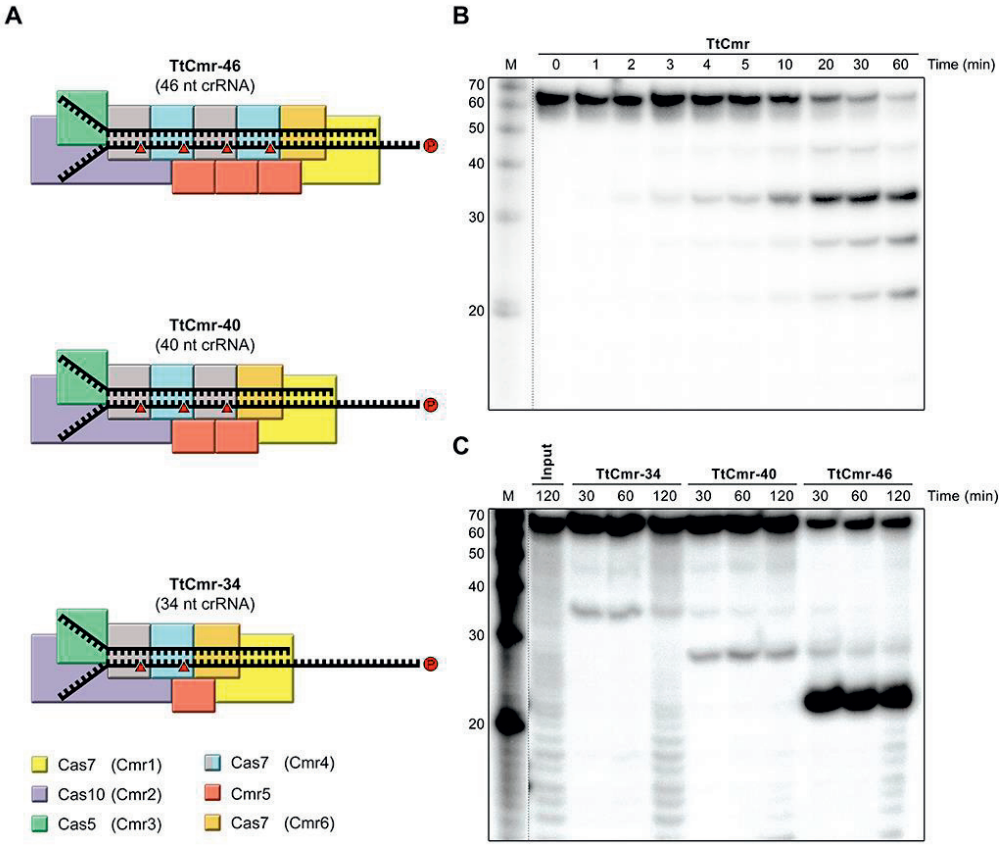


Figure 1. *In vitro* RNase activity assays with the endogenous and reconstituted TtCmr complexes.

(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) mature crRNA (top strand). Red triangles indicate the anticipated cleavage sites in the 4.5 target RNA (bottom strand, table S1) by the endoribonuclease activity of the Cas7 subunits. The target RNA was labeled at the 5' end with ^{32}P ("P" in the red circle) (B) Denaturing PAGE analysis of the activity assay using a 5' ^{32}P labeled 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 A, but using the reconstituted complexes.

effect on RNA cleavage activity, suggesting that RNA targeting by TtCmr does not check for targeting of self RNAs (e.g. antisense transcripts from the CRISPR array), similar to other type III systems [16, 17] .

Next, we tested whether TtCmr utilizes a seed similar to type I systems by setting up activity assays with mutations in the first 8 nt of the guide portion of the crRNA (the non-repeat segment of the crRNA base pairing with the

protospacer). The results showed that RNA targeting was not affected by these mutations, although a mismatch at position 5 blocked abolished the adjacent cleavage site, as demonstrated by the missing 39 nt degradation product (Figure 2B). These results suggesting that full complementarity is unnecessary in this region and that the seed might be located in a different region instead. Indeed, earlier work on the type III-B Cmr complex from *Sulfolobus islandicus* [42] showed strict base-pairing requirements at the 3' end of the crRNA.

To investigate this possibility, we performed activity assays with the endogenous TtCmr complex and RNA targets with different mismatching segments (Figure 3A). In agreement with the previous findings, RNA targets with mismatches in the first segment (nucleotides 1-5) did not interfere with target degradation, despite skipping one cleavage site downstream of the mismatched segment. Similarly, the second and third mismatched segments (nucleotides 7-12 and 13-17) resulted in skipping both the up- and downstream cleavage sites, whereas the other cleavage sites were unaffected. When mismatches were introduced in the fourth and fifth segments (nucleotides 19-23 and 25-29) however, RNA degradation was completely abolished. Mismatches in the last segment (nucleotides 31-35) had no effect on RNA degradation other than skipping the upstream cleavage site. This indicates that base pairing in the region spanning the fourth and fifth segments are crucial for target recognition and degradation.

Since the endogenous complex is a mixture of longer and shorter complexes, we switched to using the reconstituted complexes with 46 (TtCmr-46) or 40 nt crRNA (TtCmr-40) in order to pinpoint this crucial region with more precision. The TtCmr-46 complex almost completely mirrored the results obtained with the endogenous complex, showing the sensitivity to mismatches within the fourth and fifth segments (Figure 3B). However, this essential region appeared to have shifted 1 segment in the TtCmr-40 complex, with strict base pairing requirements in the third and fourth segments (Figure 3C). Taken together, these results strongly

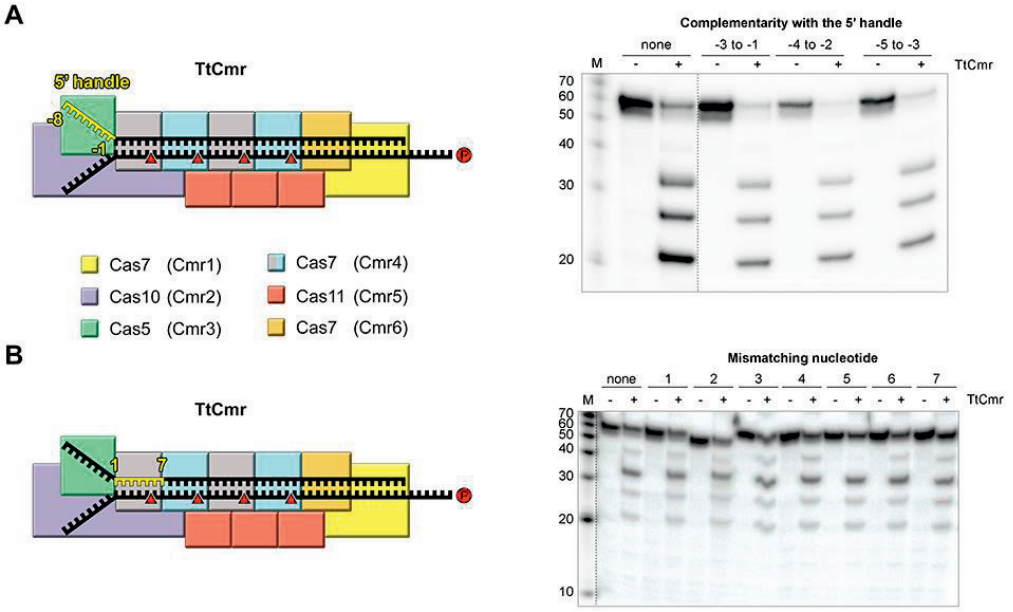


Figure 2. RNAse activity of TtCmr is not affected by targets with complementarity to the 5' handle, nor by targets with mismatches in the first guide nucleotides of crRNA.

(A) Schematic illustration of the TtCmr complex bound to a target RNA (4.5 target RNA, table S1), showing the different subunits, crRNA (top strand) and target RNA (bottom strand). The target RNA was labeled at the 5' end with ^{32}P ("P" in the red circle). Red triangles indicate the cleavage sites within TtCmr. Highlighted in white is the 5' handle of the crRNA (spanning nucleotides -1 to -8). Different target RNAs (table S1) with matches to the 5' handle were used in an activity assay and analyzed on denaturing gel (B) Similar activity assay as in panel A but with RNA targets (table S1) with a single mismatch to one the first 7 nucleotides of the guide portion of the crRNA.

suggest a 3' located seed region within TtCmr that, remarkably, shifts towards that the 5' end of the guide with smaller crRNA (and therefore complex) sizes. We propose that these regions, together or independently, act as a seed sequence in TtCmr.

We hypothesized that this seed sequence is responsible for initiating binding during recognition of a complementary RNA target. For this reason, we tested the binding affinity of the same mismatched RNA targets in an EMSA using the endogenous TtCmr complex (Figure 4A). In accordance with the activity assays, we observed that targets with mismatches in the first three segments (nucleotides 1-5, 7-11, and 13-17) did not interfere with binding to the TtCmr complex and migrated similar to the fully complementary RNA

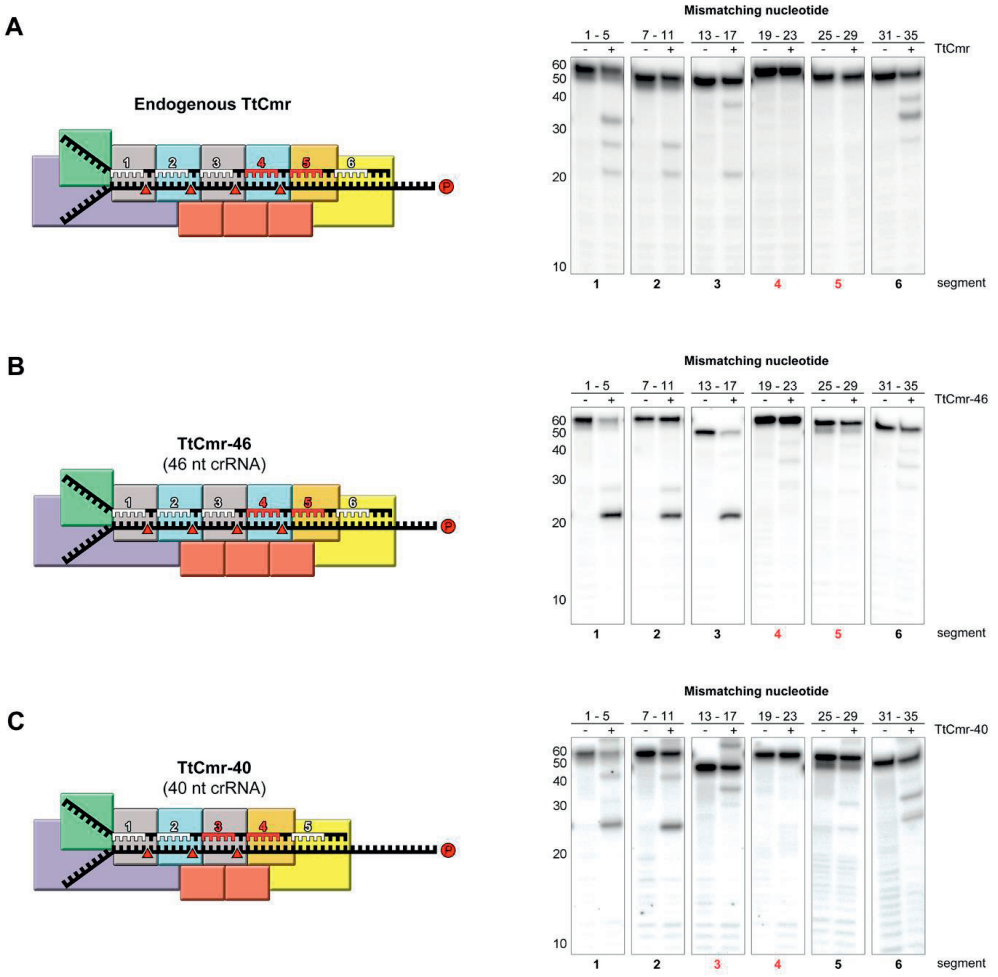


Figure 3. A flexible seed region at the 3′ end of the crRNA.

Activity assay with the (A) endogenous TtCmr ("TtCmr) or the reconstituted TtCmr complex with the (B) 46 nt crRNA ("TtCmr-46") or the 40 nt crRNA ("TtCmr-40") incubated with target RNAs (table S1) with mismatches in the indicated segments. Segments in red indicate the location of the seed sequence.

target control. Mismatches in the fourth and fifth segments (nucleotides 19-23 and 25-29) however substantially affected the migration TtCmr-target RNA tertiary complex on the gel. Although the migration was different from the unbound state, we anticipate that this might represent partial binding to the other downstream (complementary) parts of the target RNA.

To further establish that base pairing is initiated at the 3′ end of the crRNA, we designed short 11 nt target RNAs in an EMSA. Whereas targets complementary to the 5′ parts of the crRNA were unable to base pair with the crRNA, the targets base pairing with the proposed 3′ seed region were able to bind (Figure 4B). These results indicated that the 5′ end of the guide is somehow shielded from base pairing interactions with its cognate target RNA. Rather, crRNA:target duplex formation appears to be initiated at the 3′ end of the guide before propagating towards the 5′ end.

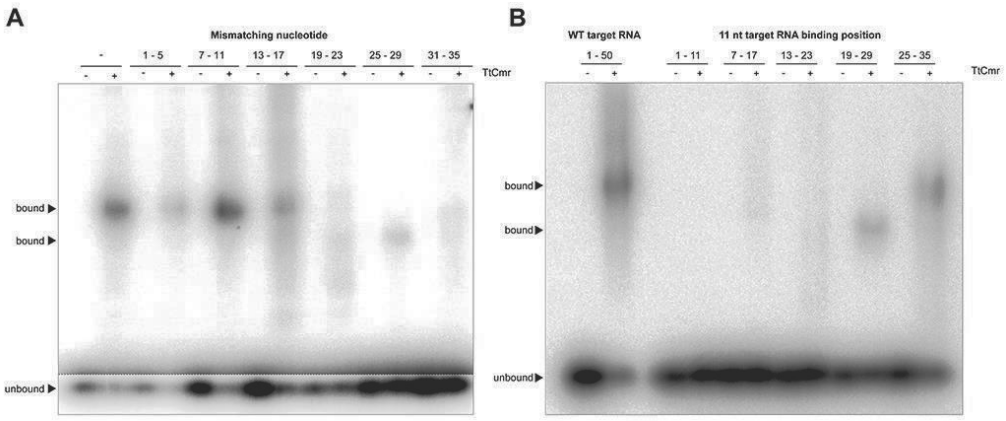


Figure 4. Base pairing of the target RNA with the TtCmr-bound crRNA is initiated at the 3′ end of the crRNA.

(A) 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. (B) EMSA analysis of the endogenous TtCmr complex incubated with short 11 nt target RNAs (table S1) complementary to the indicated nucleotides of the TtCmr-bound crRNAs.

Seed region in TtCmr structure is required for concerted structural rearrangements

We previously solved the cryo-EM structures of differently sized apo- and ssRNA target bound TtCmr complexes [27]. We observed that the complex undergoes a concerted rearrangement of subunits upon target RNA binding. Most notably, the Cas11 (Cmr5) filament rotates away from the center of the complex, thereby exposing the more 5′ located parts of the crRNA in a channel that allows further propagation of the crRNA:target RNA duplex. These observations are in good agreement with the identified seed region

in this study. For example, the proposed seed resides in a region of higher accessibility and has interactions with Cmr1, Cmr6, and a Cmr4 thumb. It's also the first segment of crRNA that is adjacent to a Cmr5 from the top (Figure 5). This suggests that initial target binding could begin at the most accessible 'open' region of the Cmr1 head; however, the crucial seed sequence residues are required to initiate the conformational changes within the complex by intercalating between the 'closed' major (Cmr4) and minor (Cmr5) backbones. Once the seed region binds, the channel of TtCmr can open, the rest of the substrate can base pair, thereby propagating the concerted conformational changes.

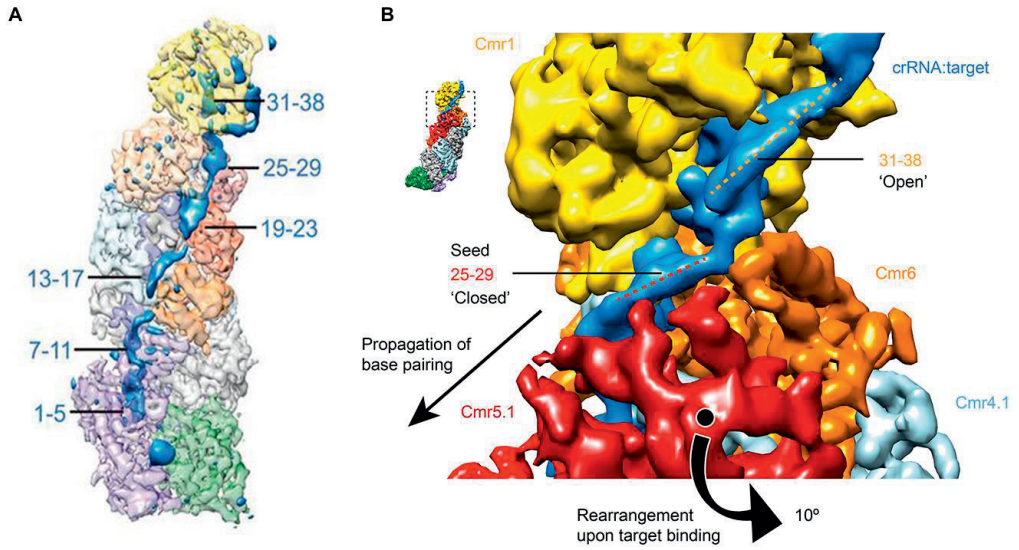


Figure 5. Seed binding leads to structural rearrangements required for target binding.

(A) Cross-section of the TtCmr complex with the 46 nt crRNA. The numbers in blue indicate the position of the RNA:target nucleotides, referred to in figures 2,3 and 4. (B) Close-up view of the top of TtCmr. The seed region is in a region that is more accessible than the more 5' located parts of crRNA and this is primed for propagating conformational changes along the complex upon binding. Images were adapted from [27].

DISCUSSION

Recent advancements in our understanding of type III CRISPR-Cas systems have highlighted that there may still be many aspects left uncovered about

these atypical type III CRISPR-Cas systems. Examples of this include the requirement for reverse-transcriptase activity for some type III systems in the adaptation phase [43, 44] and the unexpected involvement of signaling molecules in the interference phase [33-35, 45].

Furthermore, type III systems are unique in the sense that they are the only CRISPR-Cas system characterized to date capable of targeting both DNA and RNA. 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 motif in Cas10 [46]. 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* [15] and *T. thermophilus* presented here. Another typical feature of type III systems is the variable crRNA length with a typical 6 nt periodicity, which, in turn, determines by the number of Cas7 subunits that constitute the backbone of type III-A and type III-B complexes [16, 22, 24, 47-50]. Consequently, the cellular content of type III complexes most likely exists out of a heterogenous mixture of larger and smaller complexes [16, 36], although the biological significance of these observations remained unclear. Lastly, studies addressing the seed in type III systems are conflicting, one report hinting at a complete absence of a seed [51], whereas others are indicative of a seed region in either the 5′ [52, 53] or 3′ end [42, 54]. Here, we set out to characterize the RNA targeting requirements of larger and smaller type III complexes, revealing that RNA targeting is dependent on a flexible seed region at the 3′ end of the crRNA.

We showed that in contrast to the accessible 5′ seed region in type I systems, the type III TtCmr complex shields this region from initiating base pairing with its target RNA (Figure 4B and Figure 5). Instead, base pairing appears to be initiated at the exposed 3′ region of the crRNA, which coincides with the first region of the crRNA leaving the Cas7/Cas11 backbone that extends to Cmr1 and Cmr6 on the top of the complex. In agreement with our observations, previous studies showed that target RNA

binding in type III-B system requires Cmr1 and Cmr6 [29] and that base pairing in a confined region within the 3’ end of the crRNA is crucial for interference [42]. 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. In this model, proper base pairing in 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. 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 to the 5’ end (Figure 3C). Although not tested in this study, we previously observed even shorter complexes (presumably with a 34 nt crRNA), indicating that the seed in these complexes is located one more segment towards the 5’ end [27]. Taken together, these results strongly suggest that TtCmr complexes *in vivo* have a variable seed sequence on the length of the TtCmr backbone.

Although the biological significance of our findings still awaits further testing *in vivo*, we anticipate that the flexibility of the seed region will complicate MGEs that evade CRISPR-Cas targeting by introducing mutations the protospacer. On one hand, the ability to recognize heavily mutated MGEs might be beneficial for the host to provide a robust interference response towards MGEs [51, 55]. In *Marinomonas mediterranea* for example, a horizontally-acquired type III-B systems was recently shown to effectively bolster the immune response of its native type I-F systems to cope with phage escapees [32]. On the other hand, this same flexibility might come at the cost of a higher risk of self-targeting, especially since RNA degradation does not seem to be governed by the rPAM (Figure 2A). 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 HD domain (a sequence non-specific DNase) and Palm domain (resulting in the production of the second messenger that allosterically activates

promiscuous, sequence non-specific RNAse activity) might have more pronounced consequences [19, 20]. TtCmr appears to be devoid of DNase activity (due to a missing HD domain in the N-terminally truncated Cas10 protein). However, its Palm domain is still intact, raising the question whether the flexible seed region governing RNA targeting also applies to the production of the secondary messenger. The presence of (at least) two Csx1 homologues in *T. thermophilus* [30] opens up possibilities to address this and question in future studies.

EXPERIMENTAL PROCEDURES

Purification of the Cmr complex and identification of the Cmr Proteins

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 (affinity) chromatography steps as detailed previously [22]. The purified TtCmr sample was further concentrated using Vivaspin 20 centrifugal concentrator.

In vitro Activity Assays

RNA substrates (listed in Table S1) were 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). *In vitro* 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 labeled RNA substrate and 400nM TtCmr. Unless stated otherwise, the reaction was incubated at 65°C for 1 hour. RNA loading dye (containing 95% formamide) was added to the samples after incubation boiled at 95°C for 5 minutes. The samples were run over a 20% denaturing polyacrylamide gel (containing 7 M urea) for about 3-4 hours at 15mA or overnight at a constant of 4mA. The image was visualized using phosphorimaging.

EMSA

EMSAs were performed by incubating 400nM endogenous TtCmr complex with labeled target RNAs (Table 1) in Cmr binding buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl). All reactions were incubated for one hour at 65°C hour before electrophoresis on a native 5% (w/v) polyacrylamide gel (PAGE) for 2.5 hours at 15mA or overnight at a constant of 4mA. The image was visualized via phosphorimaging

AUTHOR CONTRIBUTIONS

Y.Z., J.v.d.O. and R.H.J.S. contributed to the initial design of this study. A.S. contributed to the production and purification of the Cmr complex. Y.Z., E.S., S.K., N.N., A.S., R.H.J.S., S.J.J.B., W.d.V. contributed to biochemical analyses. D.W.T., and J.A.D. contributed to electron microscopy and 3D structure determination. Y.Z., D.W.T., J.v.d.O. and R.H.J.S. contributed to the writing of manuscript.

ACKNOWLEDGEMENTS

We thank Aimi Osaki for construction of the recombinant *T. thermophilus* strain. This work was financially supported by a Veni grant to R.H.J.S. (016.Veni.171.0147) from the Dutch Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek). Y.Z. and J.v.d.O. received financial support from the Netherlands Organisation for Scientific Research (NWO), via a Gravitation grant to the Soehngen Institute for Anaerobic Microbiology (024.002.002 to W.d.V.) and an ALW-TOP project (854.10.003), respectively. A.S was supported by JSPS KAKENHI Grant Numbers 25440013 and 16K07285. D.W.T. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2218-15), and J.A.D. is Howard Hughes Medical Institute Investigators.

REFERENCES

1. Makarova, K.S., et al., *An updated evolutionary classification of CRISPR-Cas systems.* Nat Rev Microbiol, 2015. **13**(11): p. 722–736.
2. Mohanraju, P., et al., *Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems.* Science, 2016. **353**(6299): p. aad5147.
3. Jackson, S.A., et al., *CRISPR-Cas: Adapting to change.* Science, 2017. **356**(6333): p. eaal5056.
4. Hille, F., et al., *The Biology of CRISPR-Cas: Backward and Forward.* Cell, 2018. **172**(6): p. 1239-1259.
5. Bolotin, A., et al., *Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin.* Microbiology, 2005. **151**: p. 2551-2561.
6. Mojica, F.J., et al., *Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements.* J Mol Evol, 2005. **60**(2): p. 174-182.
7. Pourcel, C., G. Salvignol, and G. Vergnaud, *CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies.* Microbiology, 2005. **151**: p. 653-663.
8. Makarova, K.S., et al., *Evolution and classification of the CRISPR-Cas systems.* Nat Rev Microbiol, 2011. **9**(6): p. 467-477.
9. Shmakov, S., et al., *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems.* Molecular Cell, 2015. **60**(3): p. 385-397.
10. van der Oost, J., et al., *CRISPR-based adaptive and heritable immunity in prokaryotes.* Trends Biochem Sci, 2009. **34**(8): p. 401-407.
11. Nuñez, J.K., et al., *Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity.* Nature, 2015. **519**(7542): p. 193-198.
12. Nuñez, J.K., et al., *Foreign DNA capture during CRISPR-Cas adaptive immunity.* Nature, 2015. **527**: p. 535–538.
13. Hochstrasser, M.L., et al., *CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference.* Proc Natl Acad Sci U S A, 2014. **111**(18): p. 6618-6623.
14. Charpentier, E., et al., *Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity.* FEMS Microbiol Rev, 2015. **39**(3): p. 428-441.
15. Deng, L., et al., *A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus.* Mol Microbiol, 2013. **87**(5): p. 1088-1099.
16. Tamulaitis, G., et al., *Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus.* Molecular Cell, 2014. **56**(4): p. 506-517.
17. Samai, P., et al., *Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity.* Cell, 2015. **161**(5): p. 1164-1174.
18. Elmore, J.R., et al., *Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system.* Genes Dev, 2016. **30**(4): p. 447-459.
19. Estrella, M.A., F.T. Kuo, and S. Bailey, *RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex.* Genes Dev, 2016. **30**(4): p. 460-470.
20. Kazlauskiene, M., et al., *Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition.* Molecular Cell, 2016. **62**(2): p. 295-306.
21. Han, W., et al., *A type III-B CRISPR-Cas effector complex mediating massive target DNA destruction.* Nucleic Acids Research, 2017. **45**(4): p. 1983-1993.
22. Staals, R.H., et al., *Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus.* Molecular Cell, 2013. **52**(1): p. 135-145.
23. Ramia, N.F., et al., *Staphylococcus epidermidis Csm1 is a 3′-5′ exonuclease.* Nucleic Acids Res, 2014a. **42**(2): p. 1129-1138.
24. Benda, C., et al., *Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4.* Molecular Cell, 2014. **56**(1): p. 43-54.
25. Jung, T.Y., et al., *Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity.* Structure, 2015. **23**(4): p. 782-790.
26. Osawa, T., et al., *Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog.* Molecular Cell, 2015. **58**(3): p. 418-430.
27. Taylor, D.W., et al., *Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning.* Science, 2015. **348**(6234): p. 581-585.
28. Zhu, X. and K. Ye, *Cmr4 is the slicer in the RNA-targeting Cmr CRISPR complex.* Nucleic Acids Res, 2015. **43**(2): p. 1257-1267.
29. Hale, C.R., et al., *Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex.* Genes Dev, 2014. **28**(21): p. 2432-2443.
30. Staals, R.H.J., et al., *RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus.* Molecular Cell, 2014. **56**(4): p. 518-530.
31. You, L., et al., *Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference.* Cell, 2018. **176**(1-2): p. 239-253.
32. Silas, S., et al., *Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems.* Elife, 2017. **6**: p. e27601.
33. Athukoralage, J.S., et al., *Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate.* Nature, 2018. **562**(7726): p. 277-280.
34. Kazlauskiene, M., et al., *A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems.* Science, 2017. **357**(6351): p. 605-609.

35. Niewoehner, O., et al., *Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers.* Nature, 2017. **548**(7669): p. 543–548.
36. Spilman, M., et al., *Structure of an RNA silencing complex of the CRISPR-Cas immune system.* Molecular Cell, 2013. **52**(1): p. 146-152.
37. Walker, F.C., et al., *Molecular determinants for CRISPR RNA maturation in the Cas10–Csm complex and roles for non-Cas nucleases.* Nucleic Acids Res, 2016. **45**(4): p. 2112-2123.
38. Mojica, F.J., et al., *Short motif sequences determine the targets of the prokaryotic CRISPR defence system.* Microbiology, 2009. **155**(3): p. 733-740.
39. 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, 2011. **108**(25): p. 10098-10103.
40. Wiedenheft, B., et al., *RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions.* Proc Natl Acad Sci U S A, 2011b. **108**(25): p. 10092-10097.
41. Marraffini, L.A. and E.J. Sontheimer, *Self versus non-self discrimination during CRISPR RNA-directed immunity.* Nature, 2010. **463**(7280): p. 568-571.
42. Peng, W., et al., *An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference.* Nucleic Acids Res, 2015. **43**(1): p. 406-417.
43. Silas, S., et al., *Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein.* Science, 2016. **351**(6276): p. aad4234.
44. Toro, N., F. Martinez-Abarca, and A. Gonzalez-Delgado, *The Reverse Transcriptases Associated with CRISPR-Cas Systems.* Sci Rep, 2017. **7**(1): p. 7089.
45. Rouillon, C., et al., *Control of cyclic oligoadenylate synthesis in a type III CRISPR system.* Elife, 2018. **7**: p. e36734.
46. Koonin, E.V., K.S. Makarova, and F. Zhang, *Diversity, classification and evolution of CRISPR-Cas systems.* Curr Opin Microbiol, 2017. **37**: p. 67-78.
47. Carte, J., et al., *Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes.* Genes Dev, 2008. **22**(24): p. 3489-3496.
48. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex.* Cell, 2009. **139**(5): p. 945-956.
49. Zhang, J., et al., *Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity.* Molecular Cell, 2012. **45**(3): p. 303-313.
50. Hatoum-Aslan, A., et al., *A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs.* J Biol Chem, 2013. **288**(39): p. 27888-27897.
51. Goldberg, G.W., et al., *Incomplete prophage tolerance by type III-A CRISPR-Cas systems reduces the fitness of lysogenic hosts.* Nat Commun, 2018. **9**(1).
52. Cao, L., et al., *Identification and functional study of type III-A CRISPR-Cas systems in clinical isolates of Staphylococcus aureus.* Int J Med Microbiol, 2016. **306**(8): p. 686-696.
53. Wang, L., et al., *Dynamics of Cas10 Govern Discrimination between Self and Non-self in Type III CRISPR-Cas Immunity.* Molecular Cell, 2018. **73**(2): p. 278-290.
54. Manica, A., et al., *Unexpectedly broad target recognition of the CRISPR-mediated virus defence system in the archaeon Sulfolobus solfataricus.* Nucleic Acids Res, 2013. **41**(22): p. 10509-10517.
55. Pyenson, N.C. and L.A. Marraffini, *Type III CRISPR-Cas systems: when DNA cleavage just isn't enough.* Curr Opin Microbiol, 2017. **37**: p. 150-154.

SUPPLEMENTAL INFORMATION

Table S1. Oligonucleotides used in this study

Oligo name	Sequence (5' to 3')	Comments
P1	AUUGCGACCCGUAGAUAAGGCGCCCGGGGACGACCACGUCAAGGCGCAGUUGC	crRNA 4.5, 46nts (spacer region labeled as red)
P2	AUUGCGACCCGUAGAUAAGGCGCCCGGGGACGAC	crRNA 4.5, 34nts (spacer region labeled as red)
P3	AUUGCGACCCGUAGAUAAGGCGCCCGGGGACGACCACGUC	crRNA 4.5, 40nts (spacer region labeled as red)
P4	AUUGCGACCCGUAGAUAAGGCGCCCGGGGACGACCACGUCAAGGCG	crRNA 4.5, 46nts (spacer region labeled as red)
P5	GUUGUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGGCAGCGU	Target RNA 1 (red part complementary to crRNA 4.5)
P6	GUUGUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGGGUCCGU	Matched -3/-1 target RNA (green part)
P7	GUUGUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGGCUCGGU	Matched -4/-2 target RNA (green part)
P8	GUUGUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGGCACGCU	Matched -5/-3 target RNA (green part)
P9	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGCCCAUCG	Mismatched +1 RNA target (blue part)
P10	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACCGCCAUCG	Mismatched +2 RNA target (blue part)
P11	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUAGGGCCAUCG	Mismatched +3 RNA target (blue part)
P12	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUUCGGCCAUCG	Mismatched +4 RNA target (blue part)
P13	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCAACGGCCAUCG	Mismatched +5 RNA target (blue part)
P14	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUGUACGGCCAUCG	Mismatched +6 RNA target (blue part)
P15	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAACUACGGCCAUCG	Mismatched +7 RNA target (blue part)
P16	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCAUGCCCCAUCG	Mismatched +1-5 RNA target (blue part)
P17	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCGAAUACUACGGCCAUCG	Mismatched +7-11 RNA target (blue part)
P18	GAACUGCGCCUUGACGUGGUCGUCCCCCCCGCCCUUAUCUACGGCCAUCG	Mismatched +13-17 RNA target (blue part)
P19	GAACUGCGCCUUGACGUGGUCCAGGGCGGGGCGCCUUAUCUACGGCCAUCG	Mismatched +19-23 RNA target (blue part)
P20	GAACUGCGCCUUGACCACCACGUCCCCGGGCGCCUUAUCUACGGCCAUCG	Mismatched +25-29 RNA target (blue part)
P21	GAACUGCGCGAACUCGUGGUCGUCCCCGGGCGCCUUAUCUACGGCCAUCG	Mismatched +31-35 RNA target (blue part)
P22	CUUAUCUACGG	1-11 base pairing target RNA
P23	GGGCGCCUUAU	7-17 base pairing target RNA
P24	GUCCCCGGGCG	13-23 base pairing target RNA
P25	GUGGUCGUCCC	19-29 base pairing target RNA
P26	CUUGACGUGGU	25-35 base pairing target RNA
P27	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGGCCAUCG	Target RNA 2 (red part complementary to crRNA 4.5)
P28	GUUGUGCGCCUUGACGUGGUCGUCCCCGGGCGCGUUAUCUACGGCAGCGU	Mismatched +11 RNA target (blue part)

P29	GUUGUGCGCCUUGACGUGGUCGUCCCCCGGCGCCUUAUCUACGGCAGCGU	Mismatched +17 RNA target (blue part)
P30	GUUGUGCGCCUUGACGUGGUCCUCCCCGGGCGCCUUAUCUACGGCAGCGU	Mismatched +23 RNA target (blue part)
P31	GUUGUGCGCCUUGACCUGGUCGUCCCCGGGCGCCUUAUCUACGGCAGCGU	Mismatched +29 RNA target (blue part)
P32	GUUGUGCGCGUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGGCAGCGU	Mismatched +35 RNA target (blue part)

Chapter 6

RNA-targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*

This Chapter has been published as:

Staals, R.H.J.*, Zhu, Y.*, Taylor, D.W.*, Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K., Sakamoto, K., Suzuki, T., Dohmae, N., Yokoyama, S., Schaap, P.J., Urlaub, H., Heck, A.J., Nogales, E., Doudna, J.A., Shinkai, A., van der Oost, J. (2014) RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. Mol Cell.56(4):518-530

*contributed equally

ABSTRACT

CRISPR-Cas is a prokaryotic adaptive immune system that provides sequence-specific defense against foreign nucleic acids. Here, we report the structure and function of the effector complex of the type III-A CRISPR-Cas system of *Thermus thermophilus*: the Csm complex (TtCsm). TtCsm is composed of 5 different protein subunits (Csm1-5) with an uneven stoichiometry and a single crRNA of variable size (35-53 nt). The TtCsm crRNA content is similar to the type III-B Cmr complex, indicating that crRNAs are shared among different subtypes. A negative stain EM structure of the TtCsm complex exhibits the characteristic architecture of type I and type III CRISPR-associated ribonucleoprotein complexes. crRNA-protein cross-linking studies show extensive contacts between the Csm3 backbone and the bound crRNA guide. We show that, like TtCmr, TtCsm cleaves complementary target RNAs at multiple sites. Unlike type I complexes, interference by the TtCsm complex does not proceed via initial base pairing by a seed sequence.

INTRODUCTION

The arsenal of prokaryotic defense mechanisms against mobile genetics elements (MGE), such as bacteriophages and (conjugative) plasmids, includes adaptive immunity that serves as a sequence-specific memory of prior infections [1-5]. These systems are made up of arrays of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (*cas*) genes that are present in approximately half of sequenced bacteria and most archaea [6-8]. CRISPR-Cas systems are categorized into three major types (type I, II and III) based on their specific Cas proteins [8, 9].

CRISPR arrays are short repeated sequences (24-50 bp) interspaced by similar-sized sequences with homology to MGE (spacers). The array is preceded by a leader sequence, which contains the promoter for transcription of the array. The spacers are acquired from the MGE and inserted in the chromosomal CRISPR array of the host by a process called 'acquisition' that requires Cas1 and Cas2 proteins [10-12], although the requirement for these proteins in type III systems have not been demonstrated so far. Transcription of the CRISPR array generally results in a long pre-crRNA. In type I and III systems, this pre-crRNA is subsequently processed by a Cas6-type endoribonuclease into separate crRNAs, containing part(s) of the repeat and a single spacer sequence [13, 14]. Type II systems use Cas9, the host factor RNase III and a *trans*-activating crRNA (tracrRNA) with complementarity to the repeat for crRNA maturation [15, 16]. In some CRISPR-Cas systems, unknown nucleases trim the 5' or 3' ends of the crRNA [15, 17-19]. In the interference stage, Cas protein(s) and the mature crRNA associate to form a ribonucleoprotein (RNP) complex that targets nucleic acid sequences complementary to the crRNA (the protospacer) for degradation by a trans-acting nuclease (Cas3) in type I systems [20], or by intrinsic nuclease activity in type II and type III-B crRNP complexes [21-23]. While type I and type II complexes target DNA, the

type III-B complex is the only CRISPR-Cas system characterized to date that targets RNA.

Despite these differences, recent studies have highlighted key similarities in the architecture of type I (Cascade-like) and type III complexes [4, 24-27], suggesting that these complexes have evolved from a common ancestor. These complexes share a 'backbone' consisting of 4-6 copies of Cas7(-like) proteins and contain a smaller Cas5-like protein, which is thought to be involved in binding the 5' end of the crRNA. An important distinction between type I and type III complexes concerns the large subunit positioned at base of the backbone, which is Cas8 in most type I systems and Cas10 in type III systems [4, 5].

Thermus thermophilus HB8 is a convenient model organism to study CRISPR-Cas systems, since it has eleven CRISPR arrays (CRISPR8 is not considered a genuine CRISPR array, since it solely consists out of a single repeat sequence and no spacers) and 4 different CRISPR-Cas systems: type I-E, type III-A, type III-B and an unclassified type I system [28] (Figure 1). We previously characterized the RNA-targeting Cmr complex of the type III-B system of *T. thermophilus* [27]. As opposed to the well-characterized type I-E system, the type III-A system has hardly been analyzed biochemically.

The Csm operon of the type III-A system encodes five Csm proteins (Csm1-5) that form an RNP complex with the mature crRNA and sometimes an additional protein (called Csm6 or Csx1) [25, 29]. The Cas7-like Csm3 forms the backbone of the complex and binds RNA in a sequence-independent fashion [9, 25, 30]. After primary cleavage of the pre-crRNA, guide maturation in the type III-A system involves secondary trimming of the 3' end by a ruler-like mechanism [18], in which each Csm3 subunit binds and extends 6 nt segments of the mature crRNA and exposes unbound 3' ends for cleavage by an unknown nuclease [29].

The type III-A system prevents autoimmunity (e.g. targeting the CRISPR array) by a self versus non-self discrimination mechanism based on

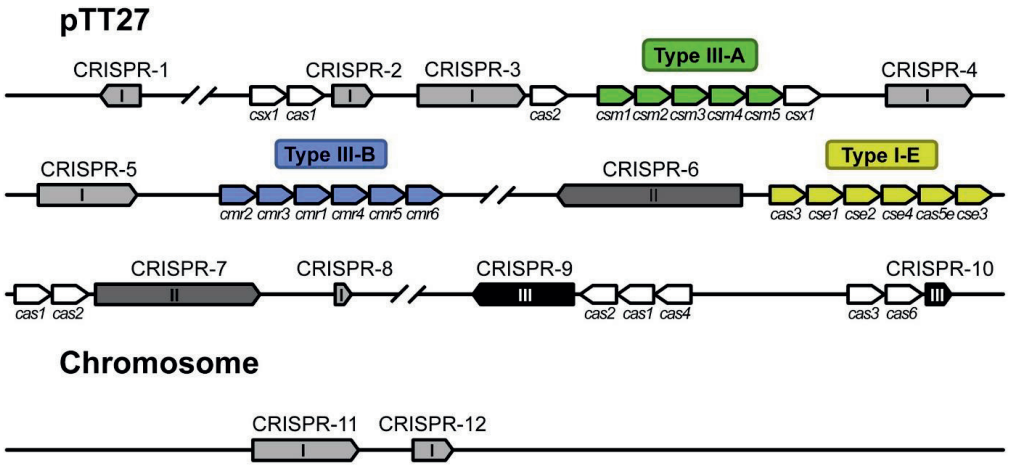
complementarity with the 5′ repeat-derived fragment of the crRNA [14, 31], i.e. by "self inactivation" [5]. In contrast, type I systems use a PAM-dependent targeting mechanism, which does not rely on base pairing [32], i.e. by "non-self activation".

Despite these initial genetic and structural insights, no *in vitro* activity of the type III-A Csm complex has yet been reported. In this study, we investigated the structural and biochemical properties of the native type III-A Csm complex isolated from *T. thermophilus*. Unexpectedly, and in contrast to previous findings, we show that Csm exhibits endoribonuclease activity using a cleavage mechanism similar to the type III-B CRISPR-Cas complex Cmr.

RESULTS

Purification and protein composition of the TtCsm complex

The *csm* genes of *T. thermophilus* HB8 are located in the vicinity of the CRISPR4 region on megaplasmid pTT27, comprising an operon composed of *csm1*, *-2*, *-3*, *-4*, *-5* and *csx1* (Figure 1). We constructed a recombinant *T. thermophilus* strain that produces the Csm5 protein fused with a $(His)_6$ tag at its C-terminus. The protein complex was purified to homogeneity using five subsequent column chromatography steps as described in the Supplemental Experimental Procedures. The purified protein complex was composed of five proteins (Figure 2A). We confirmed by mass spectrometry-based analyses (not shown) that the proteins corresponded to TTHB147 (Csm1/Cas10), TTHB148 (Csm2), TTHB149 (Csm3), TTHB150 (Csm4) and TTHB151 (Csm5). Blue native polyacrylamide gel electrophoresis (BN-PAGE) analysis of the Csm complex revealed two major bands of 430-450 kDa, suggesting minor heterogeneity of the purified Csm complex (Figure 2B). We also constructed a recombinant *T. thermophilus* strain that produces the Csx1 protein fused with a $(His)_6$ tag at its C-terminus. Under the conditions we used, Csx1 did not co-purify with the Csm complex.



Locus	Genomic location	Consensus repeat sequence (5' to 3')	Repeat type	# Spacers
CRISPR-1	18303-18044	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	3
CRISPR-2	133766-133954	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	2
CRISPR-3	135156-136099	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	12
CRISPR-4	144129-144842	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	9
CRISPR-5	146042-146983	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	12
CRISPR-6	190947-189507	GTAGTCCCCACGCACGTGGGGATGGACCG	II	23
CRISPR-7	200811-202078	GTAGTCCCCACGCGTGTGGGGATGGACCG	II	20
CRISPR-8	210807-210842	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGATAC	I	0
CRISPR-9	228237-227324	GTTGCAAACCCCGTCAGCCTCGTAGAGGATTGAAAC	III	12
CRISPR-10	239108-239214	GTTGCAAACCTCGTTAGCCTCGTAGAGGATTGAAAC	III	1
CRISPR-11	872101-873199	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	14
CRISPR-12	874397-874734	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	4

Figure 1. Schematic Representation of CRISPR Arrays and *cas* Genes on the Chromosome and Plasmid pTT27 of *T. thermophilus* HB8

CRISPR arrays (1–12) are indicated in different grayscales, depending on the repeat type (I, II, or III). *Cas*(-related) genes belonging to a particular CRISPR-Cas subtype are colored in green (subtype III-A), blue (subtype III-B), or yellow (subtype I-E). Additional *cas* genes are indicated in white. For each of these CRISPR arrays, the bottom panel summarized the genomic location, the consensus repeat sequence, repeat type, and the number of spacers. The 5' tag sequences, as found by our RNA-seq analysis, are underlined. The figure and legend are adapted from [27].

crRNA content of the TtCsm complex

Denaturing gel analysis of the co-purifying nucleic acids revealed that TtCsm binds crRNAs of variable lengths (Figure 3A). RNA-seq was performed to determine the size-distribution and nature of these

sequences. RNA-seq confirmed that the most abundant crRNAs varied in length from 35 to 53 nt

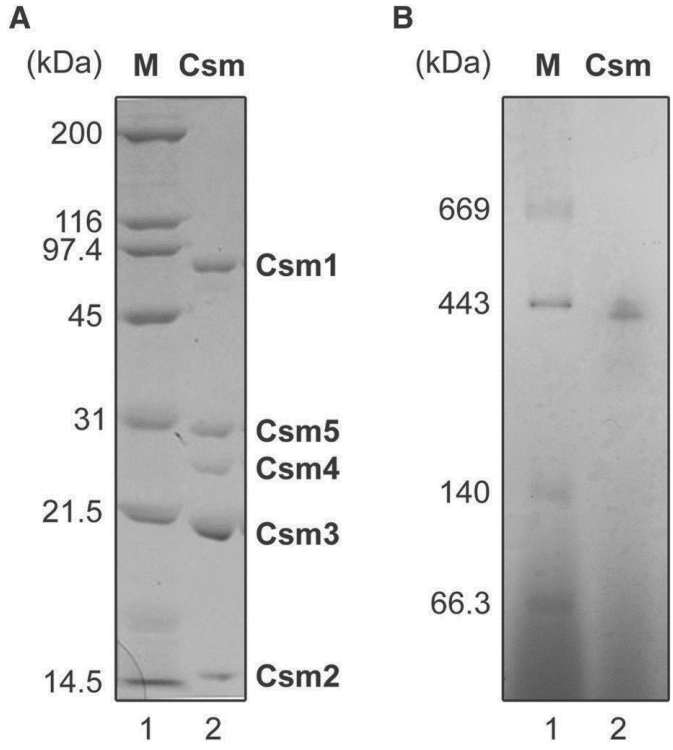


Figure 2. Purification of the Native *T. thermophilus* Type III-A Csm Complex

(A) SDS-PAGE analysis of the *T. thermophilus* Csm complex. A representative sample of the purified protein (2 μg) was analyzed on a 15% polyacrylamide gel (lane 2), followed by staining with Coomassie Brilliant Blue R-250. Each subunit is indicated. The Csm5 has a (His)6-tag at its C terminus. Lane 1, molecularmass markers. (B) BN-PAGE analysis of the *T. thermophilus* Csm complex. Two μg of the representative sample was analyzed on a 4%–16% linear polyacrylamide gradient gel in the presence of 0.02% Coomassie Brilliant Blue G-250 (lane 2). Lane 1, molecular-mass markers. Protein concentration used was determined by the Bradford method [35], using bovine serum albumin as a standard.

and were enriched for the 45 and 53 nt species. Although not as clear as in other studies of type III systems [29, 33, 34], there is a trend of 5 or 6 nt steps in the size distribution of crRNAs isolated from TtCsm (35-40-45 nt and 35-41-47-53 nt) (Figure 3B). Similar to our previous observation with crRNAs from the *T. thermophilus* Cmr complex, the majority of the Csm-bound crRNAs (71%) contained an 8 nt repeat-derived 5′ handle: 5′-AUUGCGAC [34]. Csm-bound crRNAs retained either the complete or a

truncated spacer region (39-42 nt), with the larger species containing a few nucleotides (3-6 nt) of the downstream repeat sequence.

The vast majority of the reads of the RNA-seq dataset could be mapped to the genome of *T. thermophilus* HB8 (94.18%) and revealed that most crRNAs were derived from CRISPRs 1, 4 and 11 (84.73%), while CRISPRs 6, 7, 9 and 10 were highly underrepresented (0.13%) (Figure 3C, Figure S1A). This bias strongly correlates with the different classes of repeat sequences (Figure 1). In addition, major variation of Csm-bound crRNAs occurs among sequences derived from the same CRISPR array: e.g. crRNA 4.5 (spacer 5 from CRISPR array 4, 15.5%) was one of the most abundant guides in the complex, while levels of crRNA 4.4 (0.1%) were extremely low (Figure 3D). Strikingly, the observed bias in the Csm-bound crRNA population closely resembles that of the recently established Cmr-bound crRNAs [34] (Figure S1B), indicating that similar crRNAs can be shared among the effector complexes of different CRISPR-Cas systems within one host. Despite these similarities, TtCsm-bound crRNAs are somewhat longer than those found in the TtCmr complex [34, 36]. This finding most likely suggests that crRNAs initially assemble with pre-mature protein complexes that differ in number of backbone subunits (Csm3 in TtCsm, Cmr4 in TtCmr), and that trimming of their 3’ overhang results in the aforementioned 5-6 nt size differences. These data showed that the Csm complex binds crRNA species of multiple lengths with a conserved 8 nt 5’ handle from a subset of CRISPR arrays and spacers.

crRNA-protein interactions

To study the protein-crRNA interactions within the Csm complex, we used UV-induced protein-RNA crosslinking [37, 38]. UV crosslinking was followed by enzymatic digestion of the protein and RNA moiety, enrichment of cross-linked peptide-RNA oligonucleotides and LC-MS/MS analysis. Peptide-RNA oligonucleotides were identified with their cross-linked amino acid and nucleotide by dedicated database searches [38]. Using this approach twelve

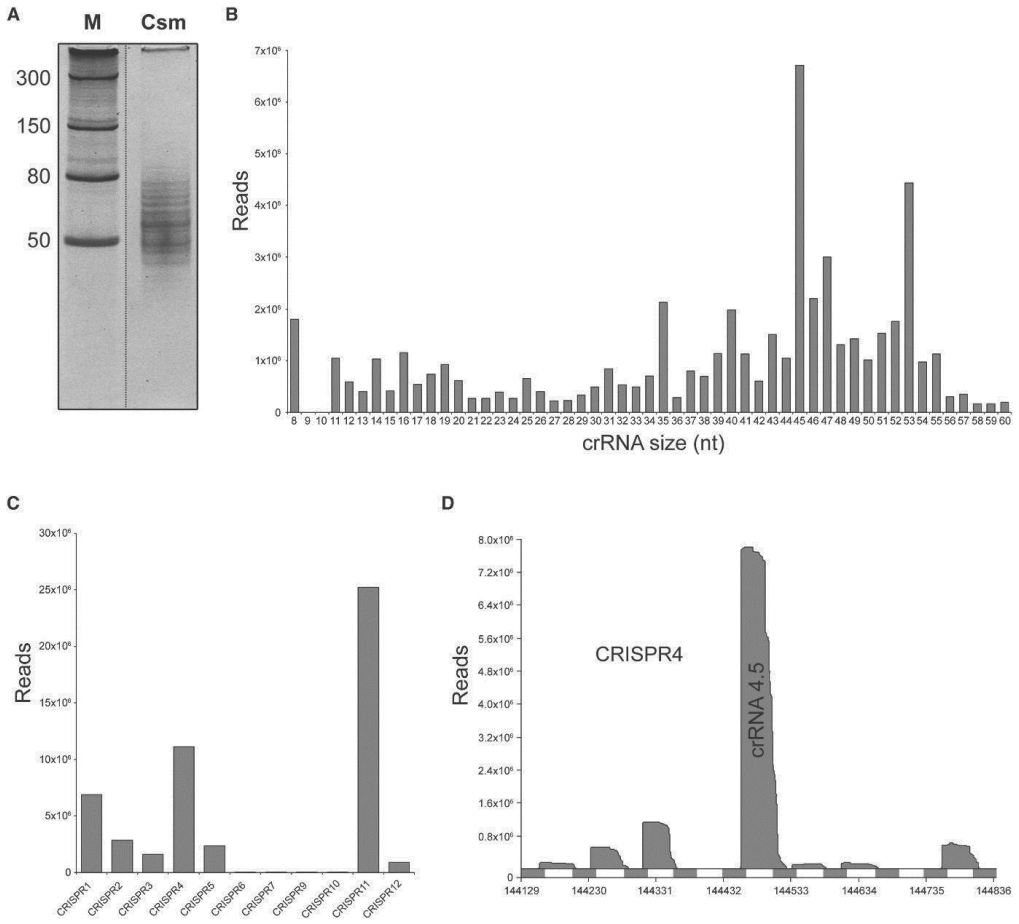


Figure 3. RNA-seq Analysis of TtCsm-Bound crRNAs

(A) crRNAs were isolated by phenol-chloroform-isoamyl alcohol extraction and analyzed by denaturing PAGE (20% AA, 7 M urea). Discontinuous gel lanes are indicated by dashed lines. (B) Histogram of the size-distributions of the Csm-bound crRNAs. (C) Histogram of the distribution of the Csm-bound crRNAs over the 11 CRISPR arrays. (D) Mapping of the Csm-bound crRNAs on CRISPR4. Overview of all mapped crRNAs and comparison of the crRNA content of the TtCsm and TtCmr complexes are provided in Figure S1.

peptides were found cross-linked to different mono- di- and tri-nucleotides from the crRNAs in the TtCsm complex (Table 1). For each of the five protein subunits (Csm1 to Csm5), at least one cross-linked peptide was identified. Remarkably, six different cross-linked peptides were identified in the Csm3 subunit. By inspection of the MS/MS fragment ion spectra we identified different amino acids as cross-linking sites (Table 1, Figure S2). The cross-linking site on the RNA was always uracil, because this is the most UV-

reactive nucleotide for this technique [39]. In most cases, the same peptide sequence was found to be cross-linked to di- and tri-nucleotide RNAs of various compositions. However, because the sample contained a mixture of natural guides isolated from *T. thermophilus*, unambiguous identification of the exact cross-linking site on the crRNAs was not possible.

Table 1. List of RNA-protein crosslinks identified in the TtCsm complex.

Protein	**Peptide Sequence**	**RNA moieties observed cross-linked to the peptide**
Csm1	^{371} RLHEALAR ^{378}	**UUA**
Csm2	^{35} LKSSQFR ^{41}	U, **U-H_2O**
Csm3	^{21} IGMSRDQMAIGDLDNPVVR ^{39}	U, **UU**, UG
Csm3	^{40} NPLTDEPYIPGSSLK ^{54}	U, U-H_2O, **UG**, UA
Csm3	^{91} IFGLAPENDER ^{101}	**U**, UU, UC, UG
Csm3	^{136} GGLYTEIKQEVFIPR ^{150}	U, UU, **UC**, UG, UCG, UUC, UUG
Csm3	^{151} LGGNANPR ^{158}	**UA**, UC, UG UGG, UCA, UUA
Csm3	^{159} TTERVPAGAR ^{168}	**U**, UG, UGG, UUG
Csm4	^{69} LPPVQVEETTLRK ^{81}	U, **UG**, UUA
Csm4	^{126} TRVGVDR ^{132}	UC, **UU**
Csm5	^{132} SPLGAYLPGSSVK ^{144}	**U**, UA, UG, UUA
Csm5	^{255} MVLLAETFR ^{263}	**U**, U-H_2O, UG

Overview of peptide-RNA oligonucleotide cross-links identified in TtCsm complex. The positions and sequence of the cross-linked peptides as identified by MS is shown and the cross-linked amino acids are underlined. For cross-linked peptides of Csm1 (positions 371 – 378) and of Csm3 (positions 21-39) the cross-linked amino acid could not be unambiguously identified due to the lack of corresponding fragment ions in the MSMS carrying a nucleotide moiety (for details, see Figure S2). For peptides cross-linked to mono-, di- or tri-nucleotides shown in bold, the corresponding MS/MS spectra are given in Figure S2.

Stoichiometry of the TtCsm complex

To investigate the architecture of the TtCsm complex, we determined the composition of the Csm protein complex using native mass spectrometry as performed previously [34, 40, 41]. Denaturing and tandem MS analyses provided accurate mass measurements for each protein subunit of TtCsm. The measured masses of the individual subunits were consistent with the theoretical values based on their amino acid sequence (Table S1). Analysis of the intact assembly by native MS revealed the presence of two major species. From their well-resolved charge state distributions, we accurately determined their masses as 426,998 ± 217 Da and 381,896 ± 261 Da (Figure 4A and Table S1), in agreement with the estimate from native gel electrophoresis (430-450 kDa, Figure 2B). Although we could measure the masses quite accurately, the stoichiometry of the Csm subunits could not be resolved unambiguously.

The two most abundant TtCsm complexes observed (427 and 382 kDa) most likely represent the intact Csm and a Csm subcomplex lacking Csm5, respectively. The measured mass difference between the two assemblies is 46.1 kDa (the mass of a Csm5 monomer is 44,286 Da). Previously, we used collision-induced dissociation on mass-selected ions of intact Cascade complexes [41]. However, selection and activation did not result in substantial fragmentation of TtCsm due to the exceptional intrinsic stability of the complex. As the TtCsm assembly could not be disrupted by tandem MS, we sought to further explore the TtCsm structure using a variety of in solution dissociation experiments on intact TtCsm. As previously described for the *Sulfolobus solfataricus* Csm complex [25], we lowered the pH of the solvent used for electrospray and also added organic modifiers to the spray solution.

From the results of all mass spectrometric experiments, we obtained a plethora of masses for subcomplexes of TtCsm formed by elimination of individual subunits (Table S2). Using these data, the stoichiometry of the

complex was narrowed down to two possible solutions, $Csm1_1Csm2_3Csm3_6Csm4_2Csm5_1crRNA_1$ (model 1, Figure 4B) or $Csm1_1Csm2_3Csm3_2Csm4_4Csm5_2crRNA_1$ (model 2). Model 1 contains multiple copies of Csm3 and has an expected mass of 427,040 Da, whereas model 2 shows a more diverse stoichiometry and has an expected mass of 427,462 Da (Table S2). Both of these models have masses that are in reasonable agreement with the measured mass of 426,998.1 Da. Although, it is not possible to distinguish between these two models based on the MS data alone, we favor model 1 for two reasons. Firstly, model 1 suggests that Csm3 (rather than Csm4) is present in multiple copies, in better agreement with the abundances observed by SDS/PAGE (Figure 2A). Secondly, the structural similarities with other type I and type III CRISPR-Cas complexes [4, 5] suggests that the Cas7-like Csm3 protein forms the backbone of the complex, as previously proposed [25, 30].

Enzymatic activity of the TtCsm complex

The Csm complex of *Staphylococcus epidermidis* has previously been shown *in vivo* to provide resistance against conjugative plasmids [42, 43], and it has been proposed to rely on a DNA-targeting mechanism. However, its cleavage activity has not yet been demonstrated *in vitro*. In the present study, we performed *in vitro* Csm activity assays with radiolabeled ssDNA and dsDNA oligos, as well as plasmid targets complementary to abundant crRNAs in the complex. Despite numerous attempts in the presence of potential co-factors, such as Mg^{2+}, Mn^{2+}, Fe^{2+}, Cu^{2+}, Co^{2+}, Ni^{2+} and Zn^{2+} (data not shown) and different topologies of the DNA target (oligos and plasmids), no specific activity could be detected in these assays (Figure S3A-B).

This prompted us to investigate the ribonuclease capabilities of the complex. A 50 nt, 5’ radiolabeled ssRNA substrate complementary to crRNA 4.5 (the most abundant Csm-bound crRNA, Figure 3D, Figure S1A) was incubated with the Csm complex in a buffer containing different co-factors

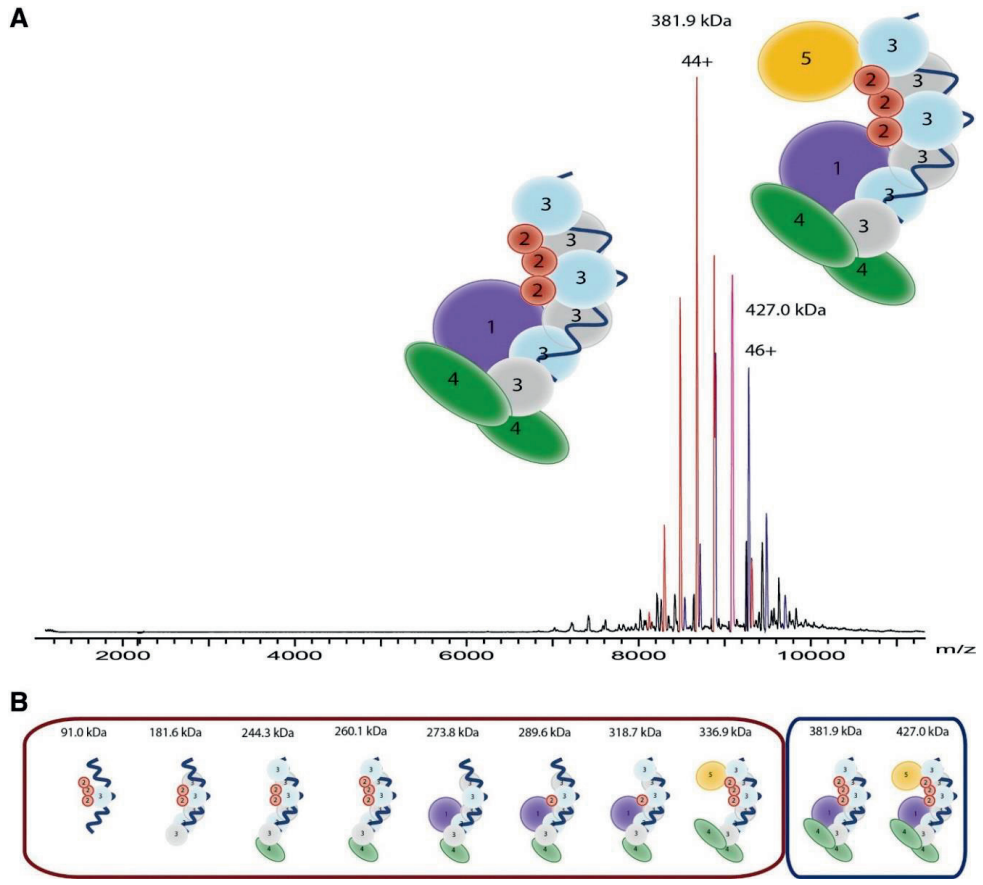


Figure 4. Subunit Composition of TtCsm

(A) Native nanoelectrospray ionization mass spectrum of the native TtCsm complex. Two main well-resolved charge state distributions are present at high m/z values, corresponding to complexes of 427 kDa (blue) and 382 kDa (red). (B) TtCsm (sub)complexes analyzed by native mass spectrometry. The subcomplexes were formed after in-solution dissociation with 30% DMSO v/v or 175 mM ammoniumacetate acidified with acetic acid (pH 3.6–4). More in-depth calculations of the different (sub)complexes can be found in Table S1 and Table S2.

(Mg^{2+}, Mn^{2+}, Zn^{2+} and Cu^{2+}) followed by denaturing gel analysis. Surprisingly, we observed Mg^{2+} and Mn^{2+}-stimulated endoribonuclease activity, while Zn^{2+} and Cu^{2+} did not stimulate specific activity (Figure S3C). For this reason, we performed all subsequent *in vitro* activity assays in the presence of Mg^{2+}. Multiple degradation products of different sizes accumulated, with 21, 15 and 9 nt being the most predominant sizes observed.

Experiments with different concentrations of TtCsm (Figure 5A) or a low Mg^{2+} concentration (Figure S3D) showed that the larger degradation products were short-lived, which suggests that the target RNA is cleaved in a step-wise fashion starting at its 3′ end. To establish that this activity was Csm-specific, we tested three RNA substrates (Figure S3D). The results showed that only RNA targets complementary to crRNAs loaded in the Csm complex (crRNA 4.5 and 11.3) were degraded, while an unrelated target RNA with no complementary to Csm-bound crRNAs remained unaffected. These results demonstrated that the endoribonuclease activity in our assays was specific for Csm, and not due to any co-purifying contaminant nuclease. The degradation products of the 4.5 and 11.3 target RNAs had similar sizes, indicating a sequence-independent cleavage mechanism.

The pattern of cleavage products and their 6 nt periodicity had a striking resemblance to those observed with the type III-B TtCmr complex, i.e. 33, 27, 21 and 15 nt (Figure S3D and Figure 5A) [34]. These results suggested that TtCsm, like TtCmr, cleaves complementary RNAs with a 5′ ruler-like mechanism, cleaving its target RNA at 6 nt intervals measured from the 5′ end, and progressing from the 3′ end. To confirm this, we performed a similar activity assay with either TtCsm or TtCmr and followed the cleavage activity in time (Figure 5B). Although TtCsm appeared to favor the formation of the smaller degradation products more quickly as compared to TtCmr, both complexes indeed had similar cleavage patterns. In further support, activity assays with a 3′ labeled 50 nt ssRNA target (complementary to crRNA 4.5) with either TtCsm or TtCmr resulted in the accumulation of a degradation product of predominantly 12 nt (Figure S3E). This indicates that the cleavage of the target RNA by both complexes is initiated at the 3′ end, followed by 6 nt interspaced, periodic cleavage events progressing towards its 5′ end (the 3′ labeling reaction adds one additional nucleotide at the 3′ end). These results indicate that TtCsm cleaves complementary target RNAs with a 5′ ruler-like mechanism analogous to TtCmr.

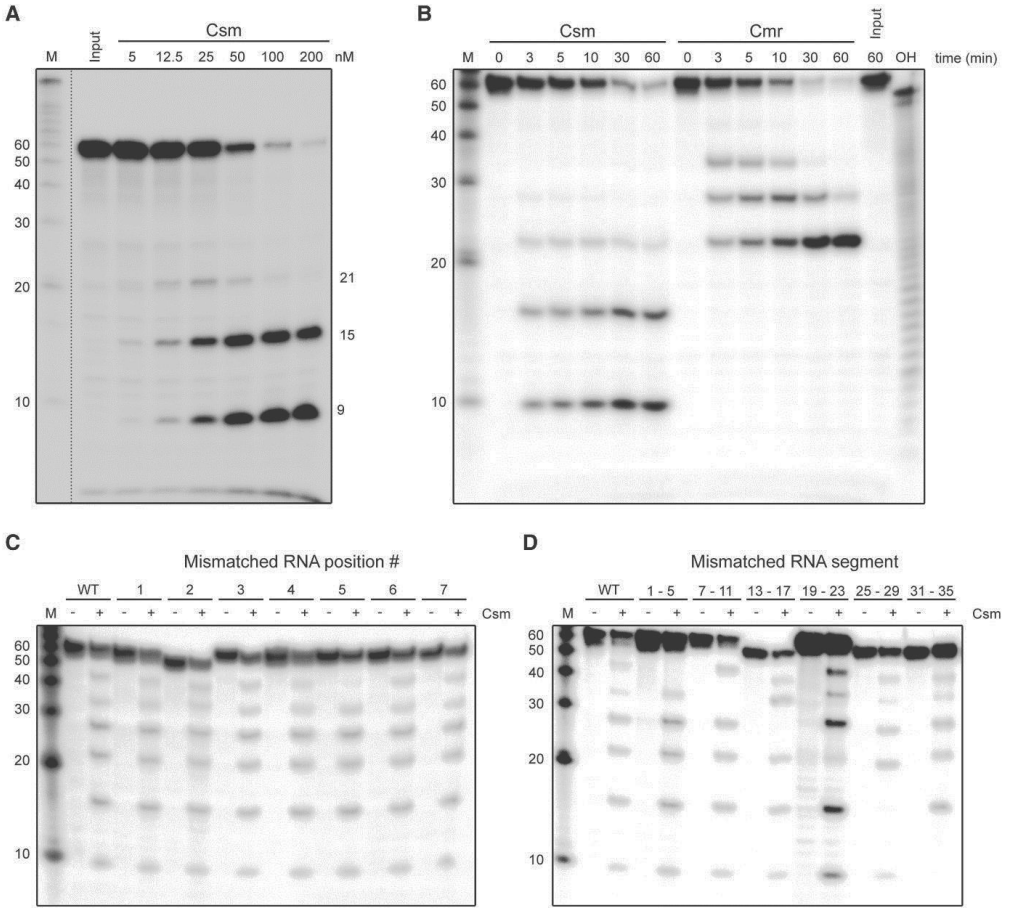


Figure 5. In Vitro Activity Assays with the TtCsm Complex

(A) A 5′-labeled ssRNA target complementary to crRNA 4.5 was incubated with different concentration of the purified, endogenous TtCsm complex in a buffer containing 2 mM Mg^{2+}. Samples were analyzed by denaturing PAGE (20% AA, 7 M urea), followed by phosphorimaging. (B) The ssRNA target was incubated with 100 nM of the endogenous Csm or Cmr complex for the indicated amount of time. OH, alkaline hydrolysis ladder of the 5′ nt RNA target. (C) RNA targets (complementary to crRNA 4.5) with base-pair-disrupting mutations at the indicated positions (also see Figure 6A) were incubated in the absence (''–'') or presence (''+'') of TtCsm. In order to visualize more (transient) degradation products, the assay was performed with a lower (10 μM) Mg^{2+} concentration. ''WT'' refers to the unmodified ''wild-type'' target RNA. (D) RNA targets (complementary to crRNA 4.5) with mutated, base-pair-disrupting regions at the indicated positions (also see Figure 6A) were incubated in the absence (''–'') or presence (''+'') of TtCsm in a buffer containing 10 μM Mg^{2+}. Additional Csm activity assays with different RNA or DNA substrates and different cofactors are provided in Figure S3. Discontinuous gel lanes are indicated by the dashed line.

Since complementarity to the Csm-bound crRNAs is required for activity (Figure S3D), we analyzed the determinants of target recognition in more detail, by testing activity on target RNAs with single base pair-disrupting

mismatches at position 1 to 7 (Figure 5C, also see Figure 6A). Only the mismatch at position 5 hampered cleavage at the site directly downstream of it, as is reflected by the less abundant 39 nt band. Apart from that, it was found that single nucleotide mismatches did not affect degradation of these target RNAs, showing that perfect complementarity is not required at these positions. These results are in contrast with the stringent requirement for base pairing interactions

Structural analysis of the TtCsm complex

We used single-particle electron microscopy (EM) and three-dimensional reconstruction (3D) of negatively stained TtCsm complexes to gain structural information about this crRNP. Raw EM micrographs showed mono-dispersed, elongated particles with a length of ~220 Å in the largest direction (Figure 7A). Using the automated data collection program Leginon [44] and the Appion image-processing pipeline [45], we recorded ~420 micrographs and extracted a stack of ~60,000 individual particle images. Reference-free two-dimensional (2D) alignment and classification produced class averages with striking features, clearly showing a worm-like architecture of two intertwined filaments with a base resembling a foot (Figure 7B). Using the structure of *E. coli* Cascade (EMDB-5314) [46] low-pass filtered to 60 Å as an initial model, we performed iterative projection-matching refinement to generate a final 3D electron density map at 17 Å (using the 0.5 Fourier Shell Correlation criterion) (Figure 7C, Figure S4). The 3D structure of TtCsm resembles a "sea worm" similar to TtCmr [34], composed of two intertwined filaments that terminate in a foot-like base. We segmented the TtCsm structure based on visual inspection of the map, our mass spectrometry analyses, and comparison to our previous segmentation of TtCmr. Of the two filaments, one is clearly larger and appears to be composed of identical, repeating subunits. This larger filament is likely composed of Csm3 based on homology to the Cmr4 backbone subunit of TtCmr and our native mass-spectrometry analyses (Figure 4). The smaller filament is likely composed of Csm2 subunits based

on a similar analysis. The head of the complex is most likely capped by Csm5 and the foot-like base contains Csm1 at the toe based on comparison to our previous Cmr structure and homology between these respective subunits (Figure S5A-B).

At the current resolution, it is difficult to unambiguously segment and dock atomic structures into the EM density; nevertheless, we constructed a model that is consistent with our biochemical results above. Using Cas7 from the *E. coli* Cascade crystal structure [47], the Csm3 crystal structure from *M. kandleri* (35% identity) [30] and the PHYRE automatic fold recognition server [48], structures for TtCsm3 were generated that fit reasonably well into the larger of the two filaments, especially for the three Csm3 subunits near the head (Figure S5C). The three Csm3 subunits near the foot appear to have substantial heterogeneity and/or adopt a different helical geometry. We hypothesize that an additional Csm3 backbone subunit in a sub-population of the purified TtCsm sample (isolated from *Thermus*) may contribute to the observed heterogeneity. Additionally, the crRNA-protein cross-linking experiments described above suggest a likely path of the crRNA along the docked TtCsm3 homology structures. In this model the crRNA would bind along the ~25 Å wide channel located between the Csm3 and Csm2 filaments and engage the conserved thumb domain of Csm3 (Figure S5D-H).

Interestingly, the two Csm3 structure predictions show hand-like characteristics similar to those attributed to Cas7 in the recent crystal structures of Cascade [47, 49, 50] (Figure S5D-H). This thumb domain is particularly evident in the PHYRE model (Figure S5D) based on Cas7 [47], whose crRNA binding regions are known from the structure of the intact Cascade. We generated a multiple sequence alignment (MSA) for Csm3 using ClustalO and showed that most of the proposed cross-linked residues are highly conserved (Figure S6). Further structural studies will be necessary to verify the accuracy of our model.

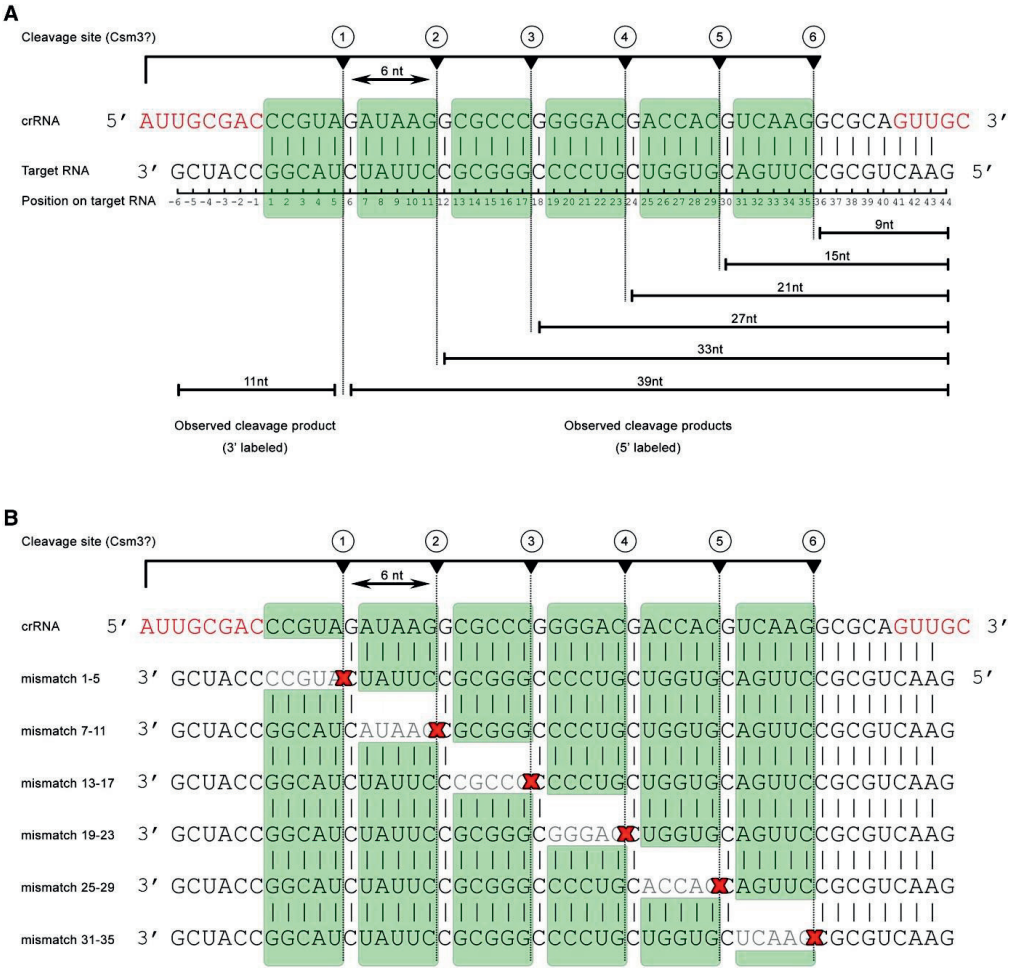


Figure 6. Model of Cleavage Activity of the TtCsm Complex

(A) Schematic representation of the cleavage activity of the TtCsm complex. (B) Schematic representation of the impact of base-pair-disrupting mutations in regions of the target RNA on activity (also see Figure 5D). Cleavages observed in this study are indicated by dotted lines. Skipped cleavage sites are indicated with red crosses.

Nucleotides in red indicate repeat-derived sequences of the crRNA.

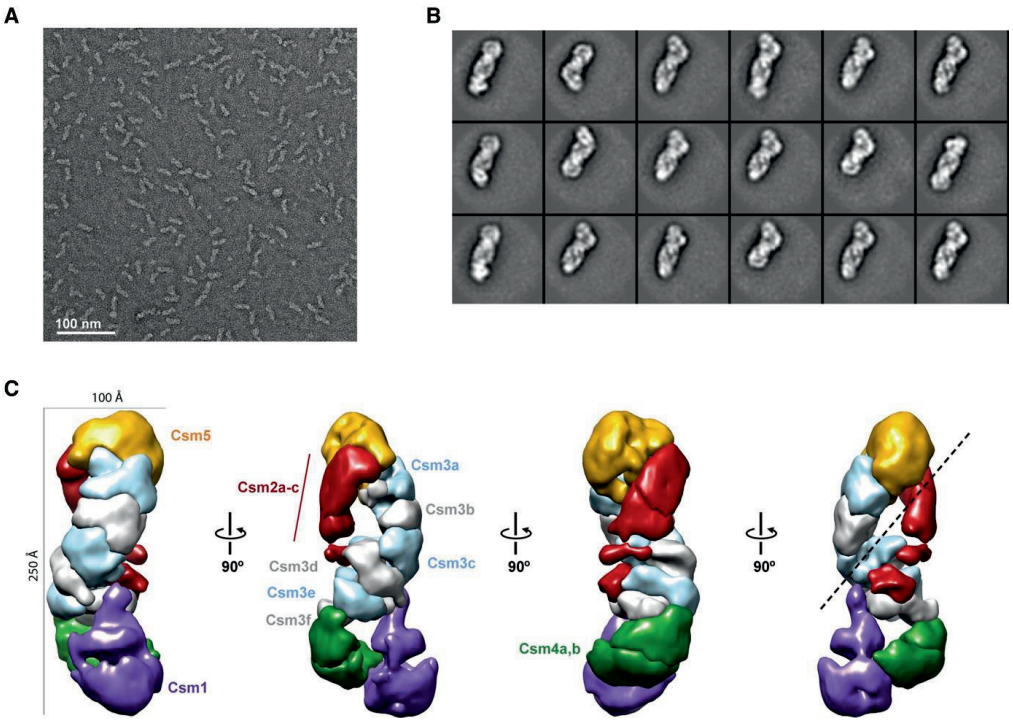


Figure 7. Molecular Architecture of the *T. thermophilus* Csm Complex

(A) Raw micrograph of negatively stained TtCsm complexes. Scale bar, 100 nm. (B) Representative reference-free 2D class averages of TtCsm complexes. The width of the boxes is ~400 Å. (C) Working segmentation of the TtCsm complex reconstruction at 17 Å resolution highlighting the "sea worm" architecture. Segmented regions are colored and labeled as Csm1 (purple), Csm2 (red), Csm3 (alternating light blue and gray), Csm4 (green), and Csm5 (orange).

DISCUSSION

By studying the native type III-A Csm complex, purified from *T. thermophilus*, we have revealed several important features of this CRISPR-Cas system, including its composition, structure and activity *in vitro*.

The TtCsm complex consists of five Csm proteins (Csm1-5) and one crRNA of variable sizes. Our RNA-seq analysis revealed that the Csm-bound crRNAs contain an invariable 5' handle: 5'-AUUGCGAC, which is consistent with the primary, Cas6-mediated cleavage of the pre-crRNA at this position [14]. The presence of the tag in the mature crRNA may play a role in initiating RNP complex formation, with the tag being specifically recognized and bound by one of the protein subunits of the complex. By analogy with

other CRISPR-Cas complexes, it is most likely that the Cas5-like Csm4 subunit might perform this function [5, 46]. The fact that mature Csm-bound crRNAs have a variable 3' end, hints at a processing step for trimming these crRNAs that appears to be a typical feature for type III CRISPR-Cas complexes [18, 21, 24, 25, 34]. Interestingly, we found more extensive 3' end processing of Csm-bound crRNAs with a non-canonical 5' handle (e.g. crRNAs from CRISPR6 and CRISPR7). This indicates that proper crRNP complex assembly (including recognition of the 5' tag sequence) is directly coupled to mature 3' end crRNA formation, and suggests that 3' processing occurs when bound within the surveillance complex. Hence, failure to properly assemble the complex (in the case of a non-canonical tag sequence) would result in crRNAs with more exposed 3' ends that are susceptible to 3' trimming [51]. These observations are consistent with a model where the sizes of mature crRNAs are determined by the dimensions of the complex. Indeed, TtCsm crRNAs (45/53 nt) were somewhat longer than the TtCmr-bound ones (40/46 nt), due to a complex with a slightly more extended backbone and therefore more protection of their 3' ends (Figure S7). Our model where Csm3 forms the backbone of the complex (model 1) is therefore in good agreement with the previously reported 6 nt binding periodicity of this protein [29].

The crRNA content of the TtCsm complex showed a remarkable bias for particular spacers, both in terms of variety and abundance. These results were strikingly similar to our previous observations with the TtCmr complex [34] (Figure S1B), suggesting that crRNAs can be shared among complexes from different CRISPR-Cas subtypes. In stark contrast, *Sulfolobus solfataricus* Csm and Cmr complexes interact with crRNAs derived from different CRISPR arrays [24, 25]. This sorting phenomenon might be accounted for by the differences in repeat specificity of these complexes or the different repertoire of Cas6 paralogues in these species and the hand-over mechanisms of these paralogues [52, 53]. It is possible that bias in spacer selection occurs at the level of primary processing by Cas6. Indeed,

in the Cascade-associated Cas6e/6f, single turnover processing of pre-crRNAs results in delivery to the appropriate crRNP complexes [4, 5, 52].

Type III-A CRISPR-Cas systems have been implicated in providing protection against plasmid conjugation and transformation *in vivo* [31, 42, 43]. The results of these studies were interpreted as evidence for a DNA targeting mechanism, although no confirmation of this activity has yet been obtained *in vitro*. Interestingly, the *in vitro* analyses presented here reveal that (under the tested conditions) the TtCsm complex harbors *in vitro* RNase rather than DNase activity. TtCsm catalyzes the cleavage of complementary target RNAs with a 5' ruler-like endoribonuclease mechanism similar to that of type III-B systems [21, 27]. This ruler mechanism defines 6 cleavage sites at the target RNA, each separated by 6 nt distances. Analogous to the type I Cascade complex, the target RNA is most likely base pairing with the crRNA guide along the backbone of the TtCsm, a model that is supported by the specificity of the TtCsm complex for complementary RNA targets and the extensive crRNA-Csm3 crosslinks observed in this study. We therefore hypothesize that the Cas7-like subunits, constituting the backbone of these type III complexes (Csm3 and Cmr4 in type III-A and III-B, respectively), harbor the active sites. Our *in vitro* activity assay with partially mismatching target RNAs showed that adjacent active sites were impaired when base pairing of the guide's upstream nucleotide(s) was disrupted. Nevertheless, these partially mismatching target RNAs were still degraded at the more distantly located active sites, which indicates that the TtCsm crRNA guides lack a defined seed sequence as is present in at least a subset of the type I systems [54-56]. Although a more detailed analysis on the boundaries of target recognition is still required, these results indicate that RNA targeting by type III-A systems is quite flexible. This flexibility could also explain why target interrogation in type III systems does not rely on a PAM (as in type I systems), since target versus non-target discrimination should be dispensable for RNA. Whether or not RNA targeting relies on other motifs

outside of the protospacer region is an interesting task for further investigations.

The intriguing discrepancy between the apparent DNA targeting activities of the Csm complex *in vivo* [42] and its RNA targeting activity *in vitro* (this study) opens up the question of whether DNA, RNA or both are the natural targets of the type III-A system (and possibly of the III-B system). Several recent studies may provide pieces of the puzzle. (1) A type III system of *Sulfolobus islandicus* has been reported to result in degradation of plasmid DNA. Interestingly, DNA interference appeared to be dependent on both transcription of the target sequence, and on the presence of Csx1 (Csm6) [57]. (2) Csx1/Csm6 are members of a highly variable protein family sharing a CARF domain (CRISPR-Cas Associated Rossmann Fold) and are strongly associated with type III CRISPR-Cas systems [58]. Furthermore, Csx1 of *Pyrococcus furiosus* has been demonstrated to associate with both dsDNA and dsRNA [59]. (3) The large subunit of the Csm complex (Cas10/Csm1) of *Staphylococcus epidermidis* degrades single-stranded DNA and RNA *in vitro* [60]. (4) In *S. epidermidis,* deletion of the *csm*6 gene (encoding a Csx1 homolog) and mutations of conserved residues in the Palm polymerase domain of Cas10/Csm1 prevented CRISPR immunity *in vivo*, without affecting either complex formation or crRNA production, strongly suggesting their involvement in target degradation [43]. (5) Given the clustering of the gene encoding Csx1/Csm6 with the 5 genes encoding the Csm complex (Figure 1) [58], it is possible the type III-A system uses a similar (Csm6/Csx1-dependent and transcription-dependent) DNA targeting mechanism as well. Indeed, during the revision of this manuscript, a new study by the Marraffini group showed that interference by the type III-A Csm complex in *S. epidermidis* proceeds in a transcription-dependent fashion, which was shown to confer resistance against lytic viruses [61]. These observations, together with the *in vitro* RNase activities from this study strongly suggest a role of the type III-A system in degrading transcriptionally active MGEs. Future *in vivo* and *in vitro* analyses are

required to fully understand how this intriguing CRISPR-Cas variant functions to protect its host from MGE invasions.

EXPERIMENTAL PROCEDURES

Purification of the Csm complex and identification of the Csm proteins

The TtCsm complex was from a $(His)_6$-tagged Csm complex expressing *T. thermophilus* HB8 strain, as described in detail in the Supplemental experimental procedures.

RNA-seq analysis

crRNAs were purified and sequenced essentially as described previously [27], the details of which can be read in the Supplemental experimental procedures.

In vitro activity assays

In vitro activity assays were performed essentially as described previously [27], details are described in the Supplemental experimental procedures.

UV-crosslinking and identification of crRNA-protein interactions by LC-MS/MS

Protein-RNA crosslinking was performed using UV irradiation at 254 nm and the cross-linked peptides were enriched as described previously [37, 39]. The sample was analyzed by LC-MS/MS essential according to [38]. A more detailed description is provided in the Supplemental experimental procedures.

Native mass spectrometry

Native MS was performed as described in detail previously [41]. Details about measurements of TtCsm subcomplexes and individual TtCsm proteins under denaturing conditions can be found in the Supplemental experimental procedures.

Single particle electron microscopy and analysis

TtCsm complexes diluted to ~25-50 nM were applied immediately to a glow-discharged continuous-carbon grid and then negatively stained with four consecutive droplets of 2% uranyl acetate. The sample was examined using a Technai-20 electron microscope equipped with a field emission gun and operated at 120-kV acceleration voltage. Image processing and 3D reconstruction were performed as described in the Supplementary experimental procedures.

ACCESSION NUMBERS

The EM-derived density map of the TtCsm complex has been deposited in the Electron Microscopy Data Bank under accession number EMD-6122. The RNA-seq data set of the Csm-bound crRNAs has been deposited in the European Nucleotide Archive database under accession number PRJEB7461.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2014.10.005.

AUTHOR CONTRIBUTIONS

A.S., K. Sakamoto, T.S., N.D., S.Y., and J.v.d.O. contributed to the initial design of this study. K. Sakamoto, S.Y., and A.S. contributed to the production and purification of the Csm complex. T.S. and N.D. contributed to the MS identification of the Csm proteins. R.H.J.S., J.J.K., and P.J.S. contributed to RNA-seq analysis. R.H.J.S., Y.Z., K.V., N.N., and M.V. contributed to biochemical analyses. D.W.T., J.E.K., E.N., and J.A.D. contributed to electron microscopy and 3D structure determination. K. Sharma and H.U. contributed to MS-based protein-RNA crosslinking. A.B. and A.J.R.H. contributed to MS-based stoichiometry analysis. R.H.J.S., Y.Z., D.W.T., K. Sharma, A.B., A.S., and J.v.d.O., with input from all, contributed to the writing of manuscript.

ACKNOWLEDGEMENTS

The authors thank Virgis Siksnys (Vilnius) for sharing data before publication. We thank Aimi Osaki for construction of the recombinant *T. thermophilus* strain. This work was supported by an ALW grant (820.02.003 to J.v.d.O.) from the Netherlands Organization for Scientific Research (N.W.O.), and by JSPS KAKENHI Grant Number 25440013 (to A.S.). A.B. and A.J.R.H. were supported by the Netherlands Proteomics Centre. E.N. and J.A.D. are Howard Hughes Medical Institute Investigators.

REFERENCES

1. Barrangou, R. and L.A. Marraffini, *CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity.* Molecular Cell, 2014. **54**(2): p. 234-244.
2. Gasiunas, G., T. Sinkunas, and V. Siksnys, *Molecular mechanisms of CRISPR-mediated microbial immunity.* Cell Mol Life Sci, 2014. **71**(3): p. 449-465.
3. Terns, R.M. and M.P. Terns, *CRISPR-based technologies: prokaryotic defense weapons repurposed.* Trends Genet, 2014. **30**(3): p. 111-118.
4. Reeks, J., J.H. Naismith, and M.F. White, *CRISPR interference: a structural perspective.* Biochem J, 2013. **453**(2): p. 155-166.
5. van der Oost, J., et al., *Unravelling the structural and mechanistic basis of CRISPR-Cas systems.* Nature Reviews Microbiology, 2014. **12**(7): p. 479-492.
6. Grissa, I., G. Vergnaud, and C. Pourcel, *The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats.* BMC Bioinformatics, 2007. **8**: p. 172.
7. Haft, D.H., et al., *A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes.* PLoS Comput Biol, 2005. **1**(6): p. e60.
8. Makarova, K.S., et al., *Evolution and classification of the CRISPR-Cas systems.* Nat Rev Microbiol, 2011. **9**(6): p. 467-477.
9. Koonin, E.V. and K.S. Makarova, *CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes.* RNA Biol, 2013. **10**(5): p. 679-686.
10. Arslan, Z., et al., *Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system.* Nucleic Acids Res, 2014. **42**(12): p. 7884-7893.
11. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-1712.
12. Yosef, I., M.G. Goren, and U. Qimron, *Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli.* Nucleic Acids Res, 2012. **40**(12): p. 5569-5576.
13. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-964.
14. Carte, J., et al., *Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes.* Genes Dev, 2008. **22**(24): p. 3489-3496.
15. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-607.
16. Carte, J., et al., *The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in Streptococcus thermophilus.* Mol Microbiol, 2014. **93**(1): p. 98-112.
17. Scholz, I., et al., *CRISPR-Cas systems in the cyanobacterium Synechocystis sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein.* PLoS One, 2013. **8**(2): p. e56470.
18. Hatoum-Aslan, A., I. Maniv, and L.A. Marraffini, *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, 2011. **108**(52): p. 21218-21222.
19. Richter, H., et al., *Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis.* Nucleic Acids Res, 2012. **40**(19): p. 9887-9896.
20. Westra, E.R., et al., *CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3.* Molecular Cell, 2012. **46**(5): p. 595-605.
21. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex.* Cell, 2009. **139**(5): p. 945-956.
22. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.* Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-E2586.
23. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-821.
24. Zhang, J., et al., *Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity.* Molecular Cell, 2012. **45**(3): p. 303-313.
25. Rouillon, C., et al., *Structure of the CRISPR interference complex CSM reveals key similarities with cascade.* Molecular Cell, 2013. **52**(1): p. 124-134.
26. Spilman, M., et al., *Structure of an RNA silencing complex of the CRISPR-Cas immune system.* Molecular Cell, 2013. **52**(1): p. 146-152.
27. Staals, R.H., et al., *Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus.* Molecular Cell, 2013. **52**(1): p. 135-145.
28. Agari, Y., et al., *Transcription profile of Thermus thermophilus CRISPR systems after phage infection.* J Mol Biol, 2010. **395**(2): p. 270-281.
29. Hatoum-Aslan, A., et al., *A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs.* J Biol Chem, 2013. **288**(39): p. 27888-27897.
30. Hrle, A., et al., *Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3.* RNA Biol, 2013. **10**(11): p. 1670-1678.
31. Marraffini, L.A. and E.J. Sontheimer, *Self versus non-self discrimination during CRISPR RNA-directed immunity.* Nature, 2010. **463**(7280): p. 568-571.
32. Westra, E.R., et al., *Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition.* PLoS Genet, 2013. **9**(9): p. e1003742.
33. Hale, C.R., et al., *Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs.* Molecular Cell, 2012. **45**(3): p. 292-302.

34. Westra, E.R., et al., *CRISPR-Cas systems preferentially target the leading regions of MOBF conjugative plasmids.* RNA Biol, 2013. **10**(5): p. 749-761.
35. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal Biochem, 1976. **72**(1-2): p. 248-254.
36. Juranek, S., et al., *A genome-wide view of the expression and processing patterns of Thermus thermophilus HB8 CRISPR RNAs.* RNA, 2012. **18**(4): p. 783-794.
37. Schmidt, C., K. Kramer, and H. Urlaub, *Investigation of protein-RNA interactions by mass spectrometry--Techniques and applications.* J Proteomics, 2012. **75**(12): p. 3478-3494.
38. Kramer, K., et al., *Mass-spectrometric analysis of proteins cross-linked to 4-thio-uracil- and 5-bromo-uracil-substituted RNA.* International Journal of Mass Spectrometry, 2011. **304**(2–3): p. 184-194.
39. Kramer, K., et al., *Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins.* Nat Methods, 2014. **11**(10): p. 1064-1070.
40. Jore, M.M., et al., *Structural basis for CRISPR RNA-guided DNA recognition by Cascade.* Nat Struct Mol Biol, 2011. **18**(5): p. 529-536.
41. van Duijn, E., et al., *Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced shot-palindromic-repeats (CRISPR)-associated protein complexes from Escherichia coli and Pseudomonas aeruginosa.* Mol Cell Proteomics, 2012. **11**(11): p. 1430-1441.
42. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA.* Science, 2008. **322**(5909): p. 1843-1845.
43. Hatoum-Aslan, A., et al., *Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system.* J Bacteriol, 2014. **196**(2): p. 310-317.
44. Suloway, C., et al., *Fully automated, sequential tilt-series acquisition with Leginon.* J Struct Biol, 2009. **167**(1): p. 11-18.
45. Lander, G.C., et al., *Appion: an integrated, database-driven pipeline to facilitate EM image processing.* J Struct Biol, 2009. **166**(1): p. 95-102.
46. Wiedenheft, B., et al., *Structures of the RNA-guided surveillance complex from a bacterial immune system.* Nature, 2011a. **477**(7365): p. 486-489.
47. Jackson, R.N., et al., *Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli.* Science, 2014. **345**(6203): p. 1473-1479.
48. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server.* Nat Protoc, 2009. **4**(3): p. 363-371.
49. Zhao, H., et al., *Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli.* Nature, 2014. **515**(7525): p. 147-150.
50. Mulepati, S., A. Heroux, and S. Bailey, *Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target.* Science, 2014. **345**(6203): p. 1479-1484.
51. Brendel, J., et al., *A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in Haloferax volcanii.* J Biol Chem, 2014. **289**(10): p. 7164-7177.
52. Niewoehner, O., M. Jinek, and J.A. Doudna, *Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases.* Nucleic Acids Res, 2014. **42**(2): p. 1341-1353.
53. Sokolowski, R.D., S. Graham, and M.F. White, *Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system.* Nucleic Acids Res, 2014. **42**(10): p. 6532-6541.
54. Wiedenheft, B., et al., *RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions.* Proc Natl Acad Sci U S A, 2011b. **108**(25): p. 10092-10097.
55. Maier, L.K., et al., *Essential requirements for the detection and degradation of invaders by the Haloferax volcanii CRISPR/Cas system I-B.* RNA Biol, 2013. **10**(5): p. 865-874.
56. 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, 2011. **108**(25): p. 10098-10103.
57. Deng, L., et al., *A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus.* Mol Microbiol, 2013. **87**(5): p. 1088-1099.
58. Makarova, K.S., et al., *CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems.* Front Genet, 2014. **5**: p. 102.
59. Kim, Y.K., Y.G. Kim, and B.H. Oh, *Crystal structure and nucleic acid-binding activity of the CRISPR-associated protein Csx1 of Pyrococcus furiosus.* Proteins, 2013. **81**(2): p. 261-270.
60. Ramia, N.F., et al., *Staphylococcus epidermidis Csm1 is a 3'-5' exonuclease.* Nucleic Acids Res, 2014a. **42**(2): p. 1129-1138.
61. Goldberg, G.W., et al., *Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting.* Nature, 2014. **514**(7524): p. 633-637.

Chapter 7

Biochemical and structural analysis of DNA targeting by the *Thermus thermophilus* CRISPR-Cas type III-A system

Manuscript in preparation:

Yifan Zhu, David W. Taylor, Eline Stroobach, Sanne Klompe, Jack E. Kornfeld, Anna Sobieraj, Willem M. de Vos, Eva Nogales, Akeo Shinkai, Stan J. J. Brouns, Jennifer A. Doudna, Raymond H.J. Staals, John van der Oost

ABSTRACT

The *Thermus thermophilus* CRISPR-Csm complex (type III-A) has previously been demonstrated to perform crRNA-guided interference of RNA targets. The crRNA guide is positioned in a groove of the backbone subunits of the Cascade-like Csm complex. These backbone subunits each contain a ribonuclease site, resulting in the concerted fragmentation of a complementary target RNA. In addition, Csm complexes have been demonstrated to target DNA as well. Here we show that complementarity between the crRNA guide and a target RNA induces non-specific cleavage of DNA substrate by the TtCsm complex. Analysis of DNA cleavage products reveals a bias of TtCsm for cleaving downstream thymidine nucleotides. Moreover, in the presence of an rPAM motif the 3' in the flanking sequence of a target RNA, allowing base pairing with the 5' handle of the crRNA guide, the activation of the DNase activity of TtCsm is substantially hampered. Comparison of cryo-EM structures of the type I-E Cascade and the type III-A Csm complex, strongly suggests that the DNA nucleases involved (Cas3 and Cas10/Csm1) share a similar position in the effector complexes of both systems.

INTRODUCTION

CRISPR-Cas is an RNA-guided adaptive immune system found in the majority of archaea and in many bacteria. It can be divided into two classes (class 1 and 2) which includes six types (I-VI) based on the variation in *cas* genes, Cas proteins, and the organization of CRISPR loci [1-3]. A major difference concerns the architecture of the crRNA binding module, in class 1 (types I, III, and IV) being a multi-subunit Cascade-like protein complex, while in Class 2 (types II, V, and VI) it rather is a single, Cas9-like multi-domain protein. Despite major structural differences, some of the systems (types I, II, and V) share some mechanistic features, including CRISPR RNA (crRNA)-guided targeting of DNA substrates, and distinguishing self and non-self DNA via a PAM (protospacer adjacent motif).

Type III system are further classified into four subtypes (III A-D) [4]. The first report of an RNA targeting CRISPR-Cas system concerned the type III-B Cmr complex of *Pyrococcus furiosus* (PfCmr) [5]. In subsequent analyses of this RNA targeting by Cmr-like complexes it has been revealed that this ribonucleoprotein (RNP) complex uses a crRNA guide to cleave a complementary RNA target multiple times at 6 nt intervals via Cmr4 subunits that compose the backbone of the Cmr complex [6-8]. Although this RNA-guided RNA targeting mechanism is very different from the RNA-guided DNA targeting of type I Cascade/Cas3 complex, there is convincing evolutionary and structural evidence that indicates a certain level of conservation between these two types [2, 8-12].

In contrast to the aforementioned *in vitro* analyses of the type III-B system, *in vivo* studies of the type III-A Csm complex of *Staphylococcus epidermidis* (SeCsm) indicated that invasion of plasmid DNA through conjugation and transformation was neutralized by cleavage of the plasmid DNA [13]. Subsequently, crRNA-dependent RNA targeting has been demonstrated *in vitro* [14, 15] revealing that type III system (for example, SeCsm, PfCmr, *Thermotoga maritima* Cmr, *Streptococcus thermophilus* Csm, *Sulfolobus*

solfataricus Cmr) can cleave both RNA and DNA in a transcription dependent manner [16-22].

Distinct subunits in type III system are responsible for the RNase and DNase ability. The Csm3 protein, that like Cmr4 constitutes the complex backbone, contains a well conserved aspartic acid residue that is involved in Mg^{2+}-dependent cleavage of target RNA [14]. The HD nuclease domain and the Palm polymerase domain of the Cas10 subunits (Csm1) play key roles in DNA target cleavage and generation of a regulatory cyclic oligonucleotide, respectively [16, 23-26]. With respect to self and non-self discrimination, it has been demonstrated that DNA cleavage by type III systems is inhibited by extended base-pairing between the 5′ repeat region (5′ handle) of the crRNA and the 3′ part of a complementary RNA target [18, 27].

In this study we have focused on the mechanism of DNA interference by the *T. thermophilus* type III-A (TtCsm) complex. We show that cleavage of a ssDNA substrate is non-specific and that it mainly occurs downstream of thymidine nucleotides. Cleavage of dsDNA has been observed, but mainly under conditions in which the two strands are at least partly separated. In addition, the DNase ability of TtCsm is hampered by the base pairing between the 5′ handle of crRNA and an rPAM motif in the 3′ flank of target RNA. Finally, we obtained structures of crRNA bound TtCsm complexes by single particle cryo-electron microscopy (cryo-EM) at ~4.5 Å resolution. Comparison of these structures with a cryo-EM structure of a previously described type I-E Cascade/Cas3 super-complex, reveals that the DNA nucleases involved (Cas10/Csm1 and Cas3) share similar positions in the effector complexes of both systems.

RESULTS

DNA binding and cleavage by TtCsm complex

Previously, we determined the crRNA content of the endogenous TtCsm complex as isolated from *Thermus thermophilus*, and demonstrated that it recognizes and degrades RNA substrates complementarity to the native set

of TtCsm-bound crRNAs [15]. To test whether TtCsm also recognizes ssDNA substrate in a sequence-specific manner, and if so, to compare the possible binding affinity between complementary ssDNA and target RNA, we performed electrophoretic mobility shift assays (EMSAs). We designed potential target RNA and ssDNA sequences based on complementarity with the most abundant native crRNA, spacer 5 from CRISPR array 4 (4.5 crRNA) [15]. When exposed to the TtCsm/crRNA complex independently, it was observed that both ssDNA and RNA cause a mobility shift in the EMSA analysis (Fig. 1A). When both RNA and ssDNA are mixed with the TtCsm complex only binding of RNA is observed (Fig. 1A), indicating that target RNA has a relatively high binding affinity to the complementary crRNA in a TtCsm complex compared with a ssDNA substrate.

To address the potential DNase activity of TtCsm, we incubated the complex with complementary RNA and/or DNA in a buffer containing Mg^{2+}, conditions that were previously found to support RNase activity [15]. As observed before, the target RNA was converted into six nucleotide interspaced degradation products, indicating robust RNase activity (Fig. 1B). However, no DNase activity could be detected under these conditions (Fig. 1B). This prompted us to determine the requirements for stimulating the DNase activity of TtCsm.

Activity assays with TtCsm were performed using a labelled ssDNA strands, either complementary ("target") or non-complementary (reverse-complement of target ("non-target" or “NC-target”)) to the 4.5 crRNA sequence, and unlabelled complementary RNA substrates in a buffer containing different combinations of metal ions (Mg^{2+}, Mn^{2+}) (Fig. 1C, Fig. S1). It was found that both the target strand and the non-target ssDNA strand could be degraded, and that this DNase activity was Mn^{2+}-dependent (Fig. 1C); Mg^{2+} appeared not to be required and, when mixed with Mn^{2+}, it has been observed that Mg^{2+} even inhibited DNase activity at concentrations higher than 0.5 mM (Fig. S2). We determined that buffer containing low Mn^{2+} concentrations in combination with KCl (instead of NaCl)

enhanced the activity even further and was used for subsequent activity assays presented in this study (Fig. S1, S2, S3 and S4). In addition, the presence of target RNA (complementarity to the 4.5 crRNA) was found to be a requisite for DNase activity, as a non-target RNA failed to stimulate DNase activity (Fig. 1D). Note that several DNA degradation products were observed in our assays (Fig. 1C & 1D), the sizes of which are consistent with cleavages of the ssDNA substrate outside of the protospacer region (details will be described below). Taken together, these results demonstrated that TtCsm, as has previously been demonstrated for other type III-A (SeCsm, StCsm) and type III-B (PfCmr) systems, is endowed with DNA endonuclease activity, activated by binding a complementary target RNA.

To determine whether complementarity between the crRNA and the target DNA is important for cleavage of the latter, activity assays were performed with or without complementary target RNA in combination with either one of the following ssDNA substrates: target, NC-target or random (no sequence homology to any of the complex-bound crRNAs) (Fig. 2A). Although it was noted that minor DNA degradation occurred in the absence of complementary target RNA, all three DNA substrates were efficiently degraded in the presence of complementary target RNA (Fig. 2A). This suggest that the DNase activity activated by binding a complementary target RNA is sequence independent. Although all three DNA substrates were degraded, the resulting degradation products differed markedly.

To determine whether TtCsm can cleave dsDNA substrates as well, the target or NC-target ssDNA were PNK labelled at their 5' end, gel-purified, annealed with an unlabelled complementary sequence, and then used in an activity assay with or without complementary target RNA (Fig. 2B). In the presence of complementary RNA, TtCsm/4.5 crRNA complex has low dsDNA cleavage activity. Minor degradation activity could be observed, which could indicate that either a minor fraction of the annealed DNA still existed in a

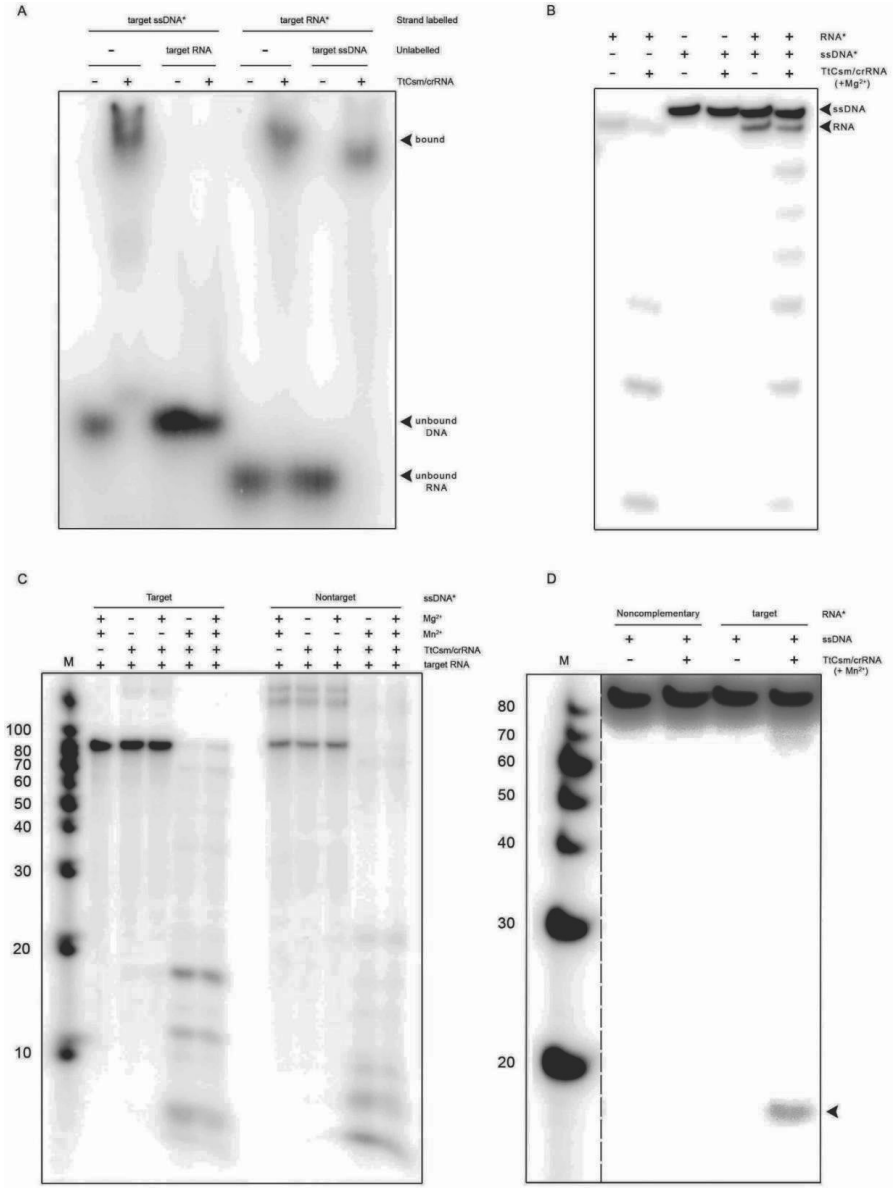


Figure 1. *In vitro* activity assays for optimal DNA cleavage system

(A) The complementary single strand target RNA or DNA sequence (template strand) was incubate with TtCsm complex in absence of metals. (B) A 5′ γ-^{32}P-ATP labelled target ssDNA or 5′ labelled target RNA complementary to crRNA 4.5 was incubated with endogenous TtCsm (with or without target RNA) in a buffer containing Mg^{2+}. (C) Cleavage of 5′ γ-^{32}P-ATP labelled ssDNA strand by the endogenous TtCsm complex with Mg^{2+} or/and Mn^{2+}. Radiolabelled Decade Marker RNAs (M) were used for size determination. (D) Cleavage of 5′ γ-^{32}P-ATP labelled complementary strand by the endogenous TtCsm complex with random RNA sequence or 4.5 target RNA. Radiolabelled linearized ssDNA primers (M) were utilized for size determination. The cleavage product is indicated (black triangle). Target: complementary, Nontarget: noncomplementary, *: 5′ γ-^{32}P-ATP labelled strand.

single-stranded state, or that the incubation temperature of 65°C melted (part of) the substrate, or TtCsm can cleave dsDNA in the presence of target RNA. In order to test whether TtCsm could cleave single-stranded regions within a dsDNA substrate, a 'bubble DNA' substrate was generated, which consist of the coding strand annealed to the template strand containing a 38 nt region mismatching with the protospacer region on the coding strand. The resulting annealed DNA substrate contains a central, 38 nt region of single-stranded DNA, flanked by 22 nt and 20 nt long duplex DNA at the 5' and 3' end. Efficient degradation of both strands of this substrate was observed when incubated with TtCsm and complementary RNA (Fig. 2B), although it may well be that the short complementary sequences melted under the used assay conditions. This further confirms that TtCsm harbours RNA-activated, sequence-unspecific endonuclease activity, as no free 3' or 5' ssDNA ends were available in this substrate. Multiple degradation bands (80-20 nt) are observed in case of the labelled NC-target strand as part of the bubble dsDNA in the absence of target RNA (Fig.2B), might be caused by binding of the target DNA strand of the bubble dsDNA to the crRNA, triggering (partial) DNase activity. Based on data presented previously with another Cmr complex [19], and on the estimated size of the obtained DNA cleavage products, we assume the thymidine (T) has the high cleavage possibility. Analysis of 5' labelled (single strand) target DNA variants (Fig. 2C), in which a stepwise substitution of thymidines has been performed, appears to agree with this assumption (Fig. 2D). For example, two degradation products between 20 nt and 10 nt were disappeared in the slot where five Ts were replaced by guanine, and instead an additional product of approximately 70 nt fragment was detected. The scenario that thymidine in some positions can be recognized and cleaved by TtCsm was confirmed by introducing two thymidines at position 9 and 10 (G9T, C10T) in the five thymidines mutation (Fig. 2D).

Self and non-self discrimination

Preventing the potential autoimmunity problem in type I and type III

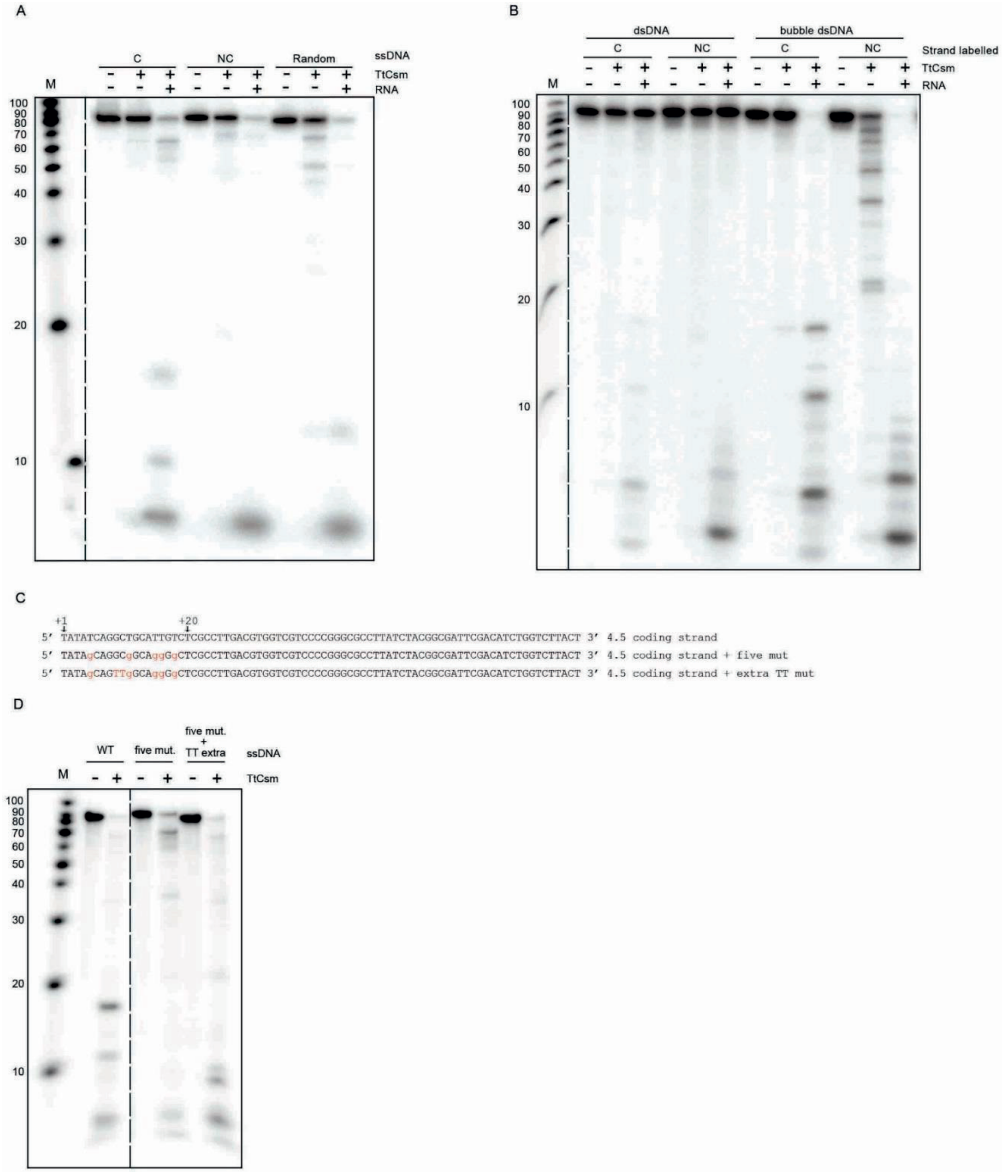


Figure 2. Endogenous TtCsm reaction with single strand, double strand DNA, and bubble dsDNA sequence

(A and B) DNA cleavage by various substrates. The ssDNA substrates were tested in (A), the dsDNA or bubble dsDNA sequences with the same size and order were tested in (B). (C) DNA sequences utilized in Fig. 2C. Red lowercase guanine is the mutant in the sequence to test the possible cleavage site. (D) *In vitro* cleavage of DNA substrates with less T on the sequence by the TtCsm complex (all the slots add target RNA as default). Figure 2A-C use radiolabelled linearized ssDNA primers (M) for size determination. Target: complementary (C), Nontarget: noncomplementary (NC) *: 5′ γ-^{32}P-ATP labelled_strand.

systems is fundamentally different. DNA targeting CRISPR-Cas systems (type I, II, V) use a protospacer adjacent motif (PAM) sequence on the foreign DNA to distinguish between target and non-targets [28-31]. PAM recognition by these systems relies on the interaction of the PAM nucleotides by specific amino acid residues of the different effector proteins, resulting in conformational changes that triggers the dsDNA target's unwinding and/or the repositioning of the nuclease in a cleavage-compatible mode [32-35]. Unlike this "non-self cleavage activating" mechanism, type III systems employ a "self cleavage inactivation" strategy to differentiate between self and non-self. Initially it was found that base-pairing interactions between the 5’ handle of the crRNA and the 3’ tag of the target (then assumed to be DNA) resulted in inactivation of target nuclease activity [27]. Later, it has been revealed that base pairing occurs between the 3' flanking sequence of the target RNA (the so-called rPAM motif) and the 5’ handle of the crRNA guide [18, 36-39]. Since our previous data showed that TtCsm can specifically target RNA (not DNA), we checked whether base pairing of the 5’ handle region of the crRNA guide with the RNA target would also affect RNA and/or DNA cleavage activities. Two RNA target substrates were generated, either with a 3 nucleotides sequence (matching positions -3/-1; 3'-CUG) complementary to the guide's 5' handle (5'-GAC), or a non-matching sequence ('wild type' positions -3/-1; 3'-ACC) (Fig. 3A). *In vitro* activity assays with these substrates reveals that the presence of matching nucleotides does not affect the typical 6 nt-interval cleavage of the RNA substrate (Fig. 3B).

To also assess the impact of crRNA/target-RNA base pairing on DNase activity, activity assays were performed with fully complementary dsDNA as well as with bubble dsDNA (Fig. 3C). As observed earlier, degradation of the substrate DNA is observed in the presence of the ‘WT’ activating RNA (“WT”, no matches with the 5’ handle of the crRNA). However, when the activating target RNA contains nucleotides (“-3/-1”) that can base pair with the guides 5' handle, the DNase activity of TtCsm is substantially affected.

Again, elevated DNA cleavage is observed in case of the bubble dsDNA construct, most likely because of melting of the strands under the used assay conditions.

These results show that in type III-A the autoimmune-like cleavage of self DNA (i.e. the CRISPR on the host genome) is prevented by inactivation of the DNase activity is controlled by a base pairing interaction between the 5′ handle of the crRNA and the target RNA (which would be an anti-sense CRISPR transcript).

Three-Dimensional Structure of the Cmr Complex

We performed cryo-electron microscopy (cryo-EM) on the endogenous TtCsm complex to further study the molecular basis of Csm complex assembly/architecture. Cryo-EM micrographs of the Csm complex showed mono-disperse, sea cucumber-shaped particles with a length of ~220 Å (Fig. 4A). Using the automated data collection program LEGINON [40], we acquired ~3,500 micrographs and automatically picked ~211,000 particles using Appion [41]. The particles were then subjected to 2D and 3D classification in RELION [42]. 2D class averages showed particles with an elongated architecture consisting of two intertwined filaments with a base (Fig. 4B). 3D classification of ~171,000 particles with three classes revealed one predominate class and two classes that consisted of junk and bad particles. Using a final set of ~71,000 particles, 3D single-particle reconstruction in RELION was performed yielding a structure of intact Csm at ~4.5 Å resolution (Fig. 4C) (using the 0.143 gold standard Fourier Shell Correlation - calculated from two independent half sets-criterion) [42-44].

The 3D structure of intact TtCsm (major species with the largest number of particles) resembles TtCmr [6, 8, 12], composed of a right handed, double helical backbone of five TtCsm3 (TtCsm3.1-3.5) subunits and four Csm2 (Csm2.1-2.4) subunits. The backbone is capped at one end by TtCsm5, and at the other end by a Csm1-Csm4 heterodimer (Fig. 4C). The Csm1–Csm4

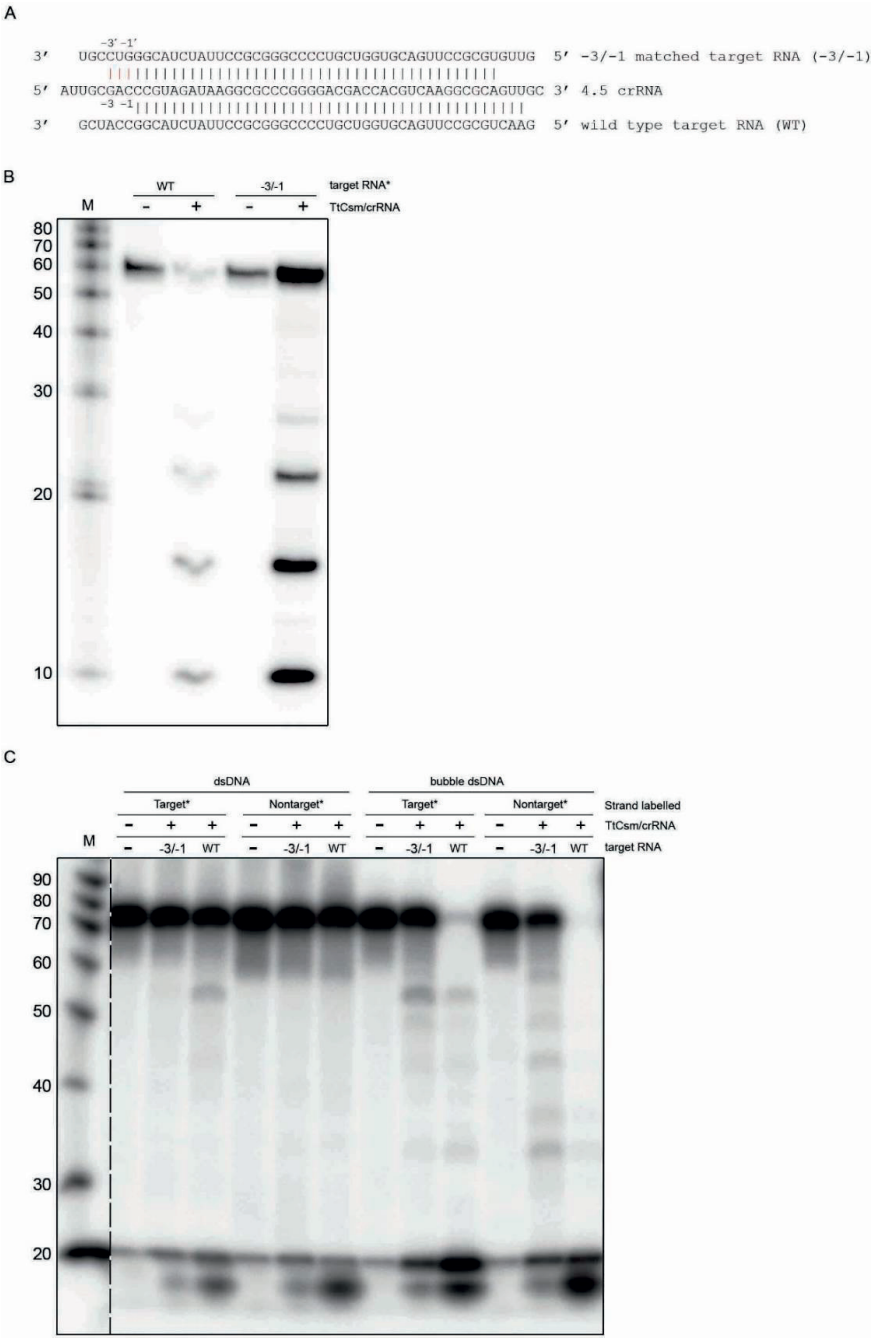


Figure 3. Activator RNA requirements for the dsDNA cleavage by the TtCsm

(A) The RNA sequences utilized in Fig. 3B and 3C. (B) Target RNA cleavage result with additional 3 bp base pairing in the 3′-flanking region of the target RNA. (C) DNA substrate cleavage by the TtCsm system with same target RNA which described in Fig. 3A. Target: complementary, Nontarget: noncomplementary *: 5′ γ-^{32}P-ATP labelled strand.

heterodimer interacts with and fixes the invariant, 5′ handle of the crRNA. Because of a slight difference in stoichiometry of this TtCsm complex structure with a previous study [15], and because of structural heterogeneity in the top TtCsm2 (TtCsm2.1) and the TtCsm5 subunit, it was decide to run a second round of 3D classification. This resulted in three species of TtCsm were obtained as resolution ~11.0- (smaller), ~6.8- (normal), ~7.6- Å (larger) (Fig. 4D). The differently sized TtCsm species likely reflect their assembly with differently sized crRNAs: smaller-TtCsm with 41 nt guide, intermediate-TtCsm with 47 nt guide and larger-TtCsm with 53 nt guide. The presence of different sized crRNAs bound to TtCsm complexes is in agreement with RNA-seq data in a previous study [15].

As the subtypes belong to same class (class 1), CRISPR-Cas type I Cascade, type III complexes (Csm and Cmr) all form multisubunit complexes around a crRNA scaffold. Remarkably, these complexes share a core backbone consisting of 3-6 copies of the Cas7 (-like) proteins (Cas7, Csm3, and Cmr4, respectively). Both in the type I and type III complexes, the crRNA guides interact with the thumb clasping the crRNA and the palm of the adjacent subunit (Fig. S5B). The backbone of the TtCsm complex also shares a common inter-subunit (Csm2-Csm3) with TtCmr (Cmr4-Cmr5) and Cascade (Cas7-Cse1). Moreover, another important structural feature shared by Cascade, Csm and Cmr is the large subunit positioned at the base of each of these complexes (Cas8e, Csm1, and Cmr2, respectively). Intriguingly, there is one important distinction between these two subunits: TtCsm1 contains an HD nuclease domain, while TtCmr2 lacks this module (Fig. 4E). This difference may explain why TtCsm1 has DNA cleavage activity, whereas TtCmr2 has no DNase activity (Zhu, Van der Oost & Staals, unpublished results in chapter 5). Another interesting observation is that the position of the TtCsm1 HD nuclease domain is similar to the site of the trans-acting Cas3 nuclease's docking with Cas8e within Cascade (Fig. 4E).

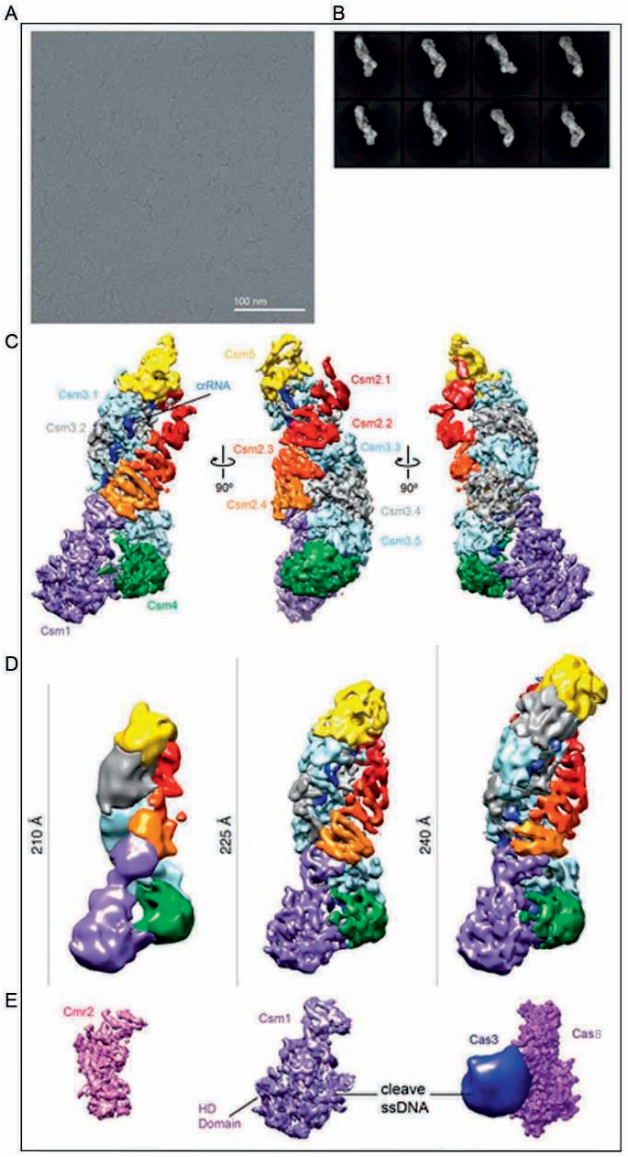


Figure 4. Cryo-EM of CRISPR-Cas ribonucleoprotein particles

(A) Drift corrected cryo-EM micrographs of TtCsm. Scale bar indicates 100 nm. (B) Reference-free 2D class averages of TtCsm show the elongated, worm-like particles, with α-helices clearly visible. The width of the boxes is ~352 Å. (C) Cryo-EM reconstruction of intact TtCsm at 4.5 Å resolution with subunits colored and labelled. (D) Comparison of the three different sized TtCsm complexes identified by 3D classification in RELION and their proposed stoichiometries. From left to right: the smaller-Csm at ~11.0 Å resolution ($Csm1_1Csm2_3Csm3_4Csm4_1Csm5_1crRNA_1$); the normal-Csm at ~6.8 Å resolution ($Csm1_1Csm2_4Csm3_5Csm4_1Csm5_1crRNA_1$); the larger-Csm at ~7.6 Å resolution ($Csm1_1Csm2_5Csm3_6Csm4_1Csm5_1crRNA_1$). (E) TtCmr2 (left) and TtCsm1 (middle) share core structural features, but Cmr2 has no HD nuclease domain. Csm1 structurally resembles Cas3:Cas8e (CasA) (right). TtCsm1 and Cas3 have an HD nuclease domain that cleaves ssDNA.

DISCUSSION

Several key characteristics of CRISPR-Cas adaptive immune systems have been uncovered after the study of CRISPR-Cas type III system which is common in bacteria and archaea [5, 13, 27, 45]. In addition, several unique features of type III systems have been elucidated (for review, see [39]). *In vivo* and *in vitro* studies have revealed transcriptional-dependent DNA targeting activity as a general mechanism of type III systems. However, some details of correct target recognizing require additional analysis. In a previous study, we addressed the *in vitro* mechanism of target RNA cleavage and ribonucleotides mismatch tolerance ability of TtCsm. The fact that a homologous complex (SeCsm, StCsm, PfCmr, TmCmr, or *Thermococcus onnurineus* Csm) has been demonstrated to catalyse shown *in vitro* DNA cleavage [16, 18, 19, 21, 46], prompted us to revisit the nuclease activities of the TtCsm complex using a combined biochemical and structural approach.

Using TtCsm isolated from its native host, we showed that the TtCsm complex cleaves ssDNA. We also discovered that complementarity between the 5' handle of the crRNA guide and the 3’ end of the target RNA stimulates the DNase activity of TtCsm. We unveiled the cryo-EM structure of the TtCsm and revealed how the different length of crRNA works as a scaffold to form three different sized RNP complexes.

The RNase activity of this defense complex is metal ion dependent and the enzyme is active in the presence of both Mg^{2+} or Mn^{2+} [15]. However, in an *in vitro* cleavage assay it was observed that the TtCsm complex does not cleave DNA fragments when Mg^{2+} is utilized as divalent cation. Intriguingly, this RNP complex started to cleave DNA when Mn^{2+} is used as metal ion. This activity is hampered by the high concentration of Mg^{2+}, which suggests that these two metal ions compete for the same metal ion binding site most likely located in the HD domain of the Csm1 subunit [47]. Analysis of the *in vitro* assay result of endogenous TtCsm revealed that the complementary

target RNA bound to the crRNA guide is required for inducing the DNase activity of the TtCsm complex, which is similar to previous reports [18, 19, 48]. In the presence of target RNA, ssDNA substrates are easily cleaved while dsDNA fragments of the same length and sequence are not a good substrate for TtCsm in the condition we tested, corroborating recent publications that dsDNA is not a favourable substrate for the Csm complex [37, 38]. The cleavage products of ssDNA and dsDNA substrates not only suggest a nonspecific ssDNA cleavage preference of TtCsm, but also show that the used 80 nt ssDNA strands are most frequently cleaved near their 5’ or 3’ end, which is not consistent with previous studies [16, 20, 21]. Further systematic analyses of the DNA substrate cleavage mechanism is required to rule out that the observed differences relate to the different sequences used, or rather to the different complexes used.

We also describe the feature of TtCsm’s preferred cleavage site in ssDNA sequence. The *T. maritima* Cmr complex is reported to have thymidine cleavage specificity caused by Cmr2 subunit [19]. Assuming that the homologous protein TtCsm1 (Cas10) most probably has a similar function, there would be a bias for cleavage at thymidine sites. This has been addressed either by substituting 5 thymidine bases (T ->G) resulting in reduction of cleavage products, and by introducing 2 thymidines (GC -> TT) resulting in increased cleavage.

As an acquired immune system, also DNA-degrading CRISPR-Cas systems must be able to discriminate self from non-self to avoid degradation of the host genome. In type I, II and V crRNA-guided DNA cleavage CRISPR-Cas systems, a protospacer adjacent motif (PAM) on the non-self target DNA sequence is required to trigger cleavage activity [28-31, 49-51]. In the type III systems, DNA is targeted non-specifically. Marraffini and Sontheimer originally proposed that complementarity between the 5’ handle of the crRNA and the 3’ flank of a target DNA prevent CRISPR-Cas type III-A Csm complex from cleaving its own CRISPR array on the host genome [27]. Later it turned out that crRNA guided RNA interference [5, 6, 15, 18], after which

non-specific DNase activity is induced. This transcription-dependent (target RNA-dependent) Csm interference system [16, 52] will encounter exposed, partially single stranded DNA in the transcription bubble, that will be cleaved in a non-specific manner (Figure 5). We indeed find that in case of the formation of 3 base pairs between the 5′ handle of the crRNA guide and the 3′ flanking sequence of the target RNA, not the target DNA, substantially lowers the DNase activity of TtCsm. In other words, TtCsm mediated DNA cleavage requires the dissociation between the 5′ handle of the crRNA and the 3′ tag of the target RNA. This will protect the host genome if antisense CRISPR transcripts are generated [53]. Most likely, the interactions of the 3′ tag of the target RNA with several amino acids in Csm1 trigger the DNase activity of the Csm1 (Cas10) subunit. In case of extended interactions between two RNA sequences will result in de-activation of the DNA cleavage. One recent report indeed demonstrates that the 3′ tag of target RNA interacts with part of the Csm1 subunit that is critical for activating DNase [38]. At the same time, the target RNA with additional complementation is still cleaved by TtCsm. This indistinctive cleavage of target RNA may also shorten the time of the (target RNA-dependent) DNase activity, reducing potential damage of the host genome.

Endogenous TtCsm complex contains three species of Csm which can be clustered as smaller, normal, and larger particles with different sized crRNAs. This finding is consistent with former RNA-seq data we reported, and in line with one more pair of Csm2-Csm3 subunits in longer TtCsm or one less pair in smaller complex relative to normal TtCsm [15]. These different sized complexes likely reflect the generation of different size crRNAs produced by Cas6 mediated pre-crRNA maturation. However, in the *T. onnurineus* type III-A complex (ToCsm), the reconstituted Csm complex only has one Csm2 subunit and harbours a 38 nt crRNA [46]. In future studies, it will be interesting to test how the size of crRNA affects the protein assembly and substrate cleavage [38]. The type III-A Csm complex is composed of two backbone filaments, with features analogous to the structures of the

Cascade (type I-E) and Cmr (type III-B) complexes. They all revealed interesting similarities in the large subunit (Cas8e and Cas10) at one end of the complex. The position of the TtCsm1 HD domain is strikingly similar to the site of the Cas3 nuclease's docking with Cas8e within Cascade. Importantly, the HD domain present in Cas3 is required for DNA cleavage. Along with recent findings that suggest transcription dependent ssDNA cleavage ability of type III-A complexes, our cryo-EM data support a possible model for DNA targeting by the Csm complex [16, 18, 19, 21]. Formation of a duplex between crRNA and a nascent transcript might position the Csm1 HD domain adjacent to DNA substrate for cleavage (Figure 5). Negative supercoiling induced in the wake of the passing of the transcription bubble would contribute to local melting of the dsDNA, which is likely required for Csm1 ssDNA cleavage [16].

In summary, our results elaborate on the mechanism of DNA cleavage in type III-A TtCsm complex and establish a link between target RNA and DNA. Future work will explore the relationship between type III system and co-expressed Cas protein, for example two Csx1 proteins, in *T*. *thermophilus*.

EXPERIMENTAL PROCEDURES

Endogenous TtCsm proteins

Construction and cultivation of the *T*. *thermophilus* HB8 strain producing the $(His)_6$-tagged Csm protein was described in previous article [15]. Protein purification and identification part also can be found in this article.

Target binding assay

For target binding assay, the endogenous TtCsm complexes (400 nM) were incubated with the 5' terminal γ-^{32}P-ATP labelled nucleic acid (DNA or RNA) for 1 h at 65°C in 40 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM DTT, and 1 mM ATP. All the reactions were mixed with 6X gel loading dye (NEB) and loaded into 5% nondenaturing polyacrylamide gel.

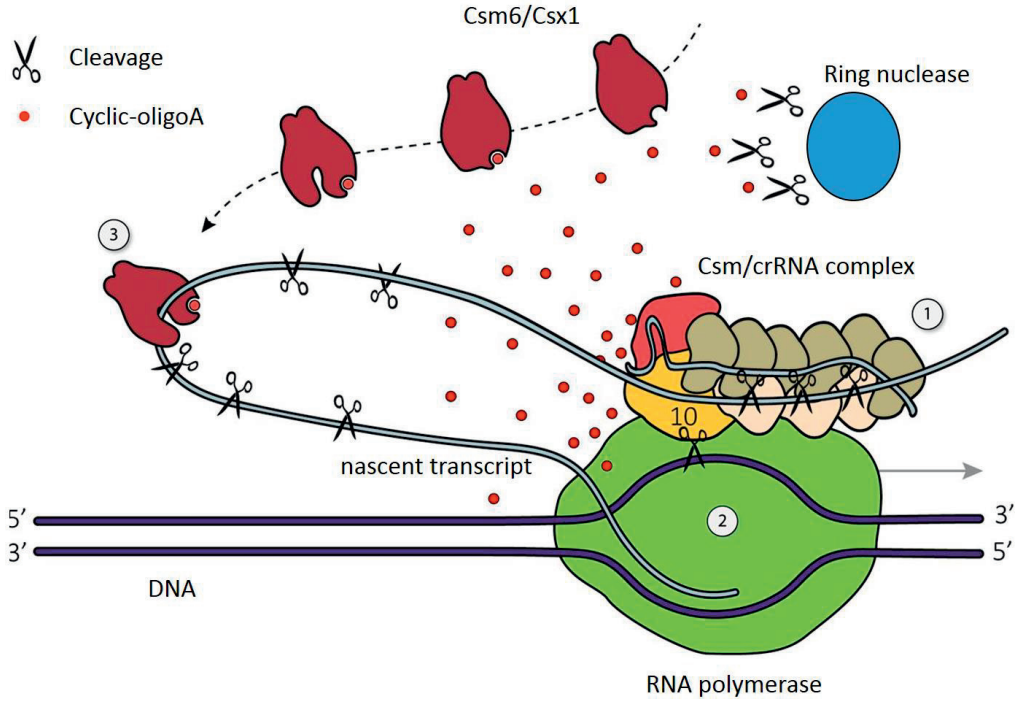


Figure 5. Model for type III and Cs6 RNA and DNA targeting

(1) Base pairing between the crRNA and target RNA induces RNA cleavages by the multiple copies of Cas7 homologues (Csm3 or Cmr4) present in the RNP backbone. (2) Non-complementarity between the 5′ handle of the crRNA and the 3′ tag of the protospacer region of the target RNA result in the activation of the HD and palm domains of Cas10. The HD domain of Cas10 non-specifically degrades ssDNA that is likely to be available in the transcription bubble caused by RNA polymerase (green) during transcription. The Cas10 palm domain converts ATP to c(OA)s or short poly(A) tail that function as diffusible second messengers. (3) Second messenger bind to the CRISPR-A-associated Rossmann Fold (CARF) domains of the Csm6 homodimer. This interaction allosterically activates the sequence non-specific ssRNase activity of its higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domains. The second messenger will be degraded by a Cas protein via its CARF domain.

In vitro activity assays

All DNA and RNA substrates were purchased from Intefrated DNA Technologies (IDT) or Eurogentic. A full list of all the oligonucleotides is provided in Table S1. The 5′ terminal of DNA or RNA substrates were labelled by T4 polynucleotide kinase (PNK) and γ-^{32}P-ATP, and then purified from denaturing PAGE. *In vitro* activity assays were achieved by incubating the substrate with TtCsm complex at 65°C for 0.5 or 1 h in a buffer containing: 40 mM Tris-HCl, 40 mM NaCl or 150 mM NaCl, 5 mM DTT, and 0.125 mM $MnCl_2$ (unless indicated otherwise). After incubation, an equal

volume of formamide RNA loading buffer was loaded and incubated for 5 min at 95°C. Samples and 5′ labelled ssDNA markers or ssRNA Decade Markers (Ambion) were analysed by denaturing PAGE and visualized by autoradiography.

Single Particle Electron Microscopy and Analysis

Three microliter droplets of the TtCsm (~0.1 mg ml-1) were placed onto glow-discharged continuous-carbon grid and then plunged into liquid ethane. Data were acquired via a FEI Titan Krios transmission electron microscope operating at 300 keV. Data processing and analysis was performed using a method similar to those described previously for TtCmr [12].

ACCESSION NUMBERS

The EM-derived density map of the TtCsm complex has been deposited in the Electron Microscopy Data Bank under accession number EMD-6122.

AUTHOR CONTRIBUTIONS

Y.Z., W.d.V., R.H.J.S., and J.v.d.O. contributed to the initial design of this study. A.S. contributed to the production and purification of the TtCsm complex. Y.Z., E.S., S.K., S.J.J.B, and A.S. contributed to biochemical analyses. D.W.T., J.E.K., E.N., and J.A.D. contributed to electron microscopy and 3D structure determination. Y.Z., R.H.J.S., and J.v.d.O., with input from all, contributed to the writing of manuscript.

ACKNOWLEDGMENTS

The authors thank Daria Lavysh and Konstantin Severinov for helpful discussion. We thank Aimi Osaki for construction of the recombinant *T. thermophilus* strain. This work was supported by an Gravitation grant (6162500400 to W.d.V.) from the Dutch governments and by JSPS KAKENHI Grant Numbers 25440013 and 16K07285 (to A.S.). R.H.J.S. is supported by an NWO/Veni grant (016.Veni.171.047), and J.v.d.O. by an NWO/TOP grant (714.015.001). S.J.J.B. is supported by the Netherlands

Organization for Scientific Research VIDI Grant 864.11.005. D.W.T. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2218-15)., E.N. and J.A.D. are Howard Hughes Medical Institute Investigators.

REFERENCES

1. Makarova, K.S., et al., *Evolution and classification of the CRISPR-Cas systems.* Nat Rev Microbiol, 2011. **9**(6): p. 467-477.
2. Makarova, K.S., et al., *An updated evolutionary classification of CRISPR-Cas systems.* Nat Rev Microbiol, 2015. **13**(11): p. 722–736.
3. Shmakov, S., et al., *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems.* Molecular Cell, 2015. **60**(3): p. 385-397.
4. Koonin, E.V. and K.S. Makarova, *Mobile genetic elements and evolution of CRISPR-Cas systems: all the way there and back.* Genome Biology and Evolution, 2017. **9**(10): p. 2812-2825.
5. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex.* Cell, 2009. **139**(5): p. 945-956.
6. Staals, R.H., et al., *Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus.* Molecular Cell, 2013. **52**(1): p. 135-145.
7. Ramia, N.F., et al., *Essential structural and functional roles of the Cmr4 subunit in RNA cleavage by the Cmr CRISPR-Cas complex.* Cell Rep, 2014b. **9**(5): p. 1610-1617.
8. Osawa, T., et al., *Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog.* Molecular Cell, 2015. **58**(3): p. 418-430.
9. Jackson, R.N., et al., *Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli.* Science, 2014. **345**(6203): p. 1473-1479.
10. Mulepati, S., A. Heroux, and S. Bailey, *Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target.* Science, 2014. **345**(6203): p. 1479-1484.
11. Zhao, H., et al., *Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli.* Nature, 2014. **515**(7525): p. 147-150.
12. Taylor, D.W., et al., *Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning.* Science, 2015. **348**(6234): p. 581-585.
13. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA.* Science, 2008. **322**(5909): p. 1843-1845.
14. Tamulaitis, G., et al., *Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus.* Molecular Cell, 2014. **56**(4): p. 506-517.
15. Staals, R.H.J., et al., *RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus.* Molecular Cell, 2014. **56**(4): p. 518-530.
16. Samai, P., et al., *Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity.* Cell, 2015. **161**(5): p. 1164-1174.
17. Zhang, J., et al., *Multiple nucleic acid cleavage modes in divergent type III CRISPR systems.* Nucleic Acids Res, 2016. **44**(4): p. 1789-1799.
18. Elmore, J.R., et al., *Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system.* Genes Dev, 2016. **30**(4): p. 447-459.
19. Estrella, M.A., F.T. Kuo, and S. Bailey, *RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex.* Genes Dev, 2016. **30**(4): p. 460-470.
20. Jiang, W., P. Samai, and L.A. Marraffini, *Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity.* Cell, 2016. **164**(4): p. 710-721.
21. Kazlauskiene, M., et al., *Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition.* Molecular Cell, 2016. **62**(2): p. 295-306.
22. Liu, T.Y., A.T. Iavarone, and J.A. Doudna, *RNA and DNA Targeting by a Reconstituted Thermus thermophiles Type III-A CRISPR-Cas System.* PLoS One, 2017. **12**(1): p. e0170552.
23. Hatoum-Aslan, A., et al., *Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system.* J Bacteriol, 2014. **196**(2): p. 310-317.
24. Ramia, N.F., et al., *Staphylococcus epidermidis Csm1 is a 3'-5' exonuclease.* Nucleic Acids Res, 2014a. **42**(2): p. 1129-1138.
25. Kazlauskiene, M., et al., *A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems.* Science, 2017. **357**(6351): p. 605-609.
26. Niewoehner, O., et al., *Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers.* Nature, 2017. **548**(7669): p. 543–548.
27. Marraffini, L.A. and E.J. Sontheimer, *Self versus non-self discrimination during CRISPR RNA-directed immunity.* Nature, 2010. **463**(7280): p. 568-571.
28. Deveau, H., et al., *Phage response to CRISPR-encoded resistance in Streptococcus thermophilus.* J Bacteriol, 2008. **190**(4): p. 1390-1400.
29. Mojica, F.J., et al., *Short motif sequences determine the targets of the prokaryotic CRISPR defence system.* Microbiology, 2009. **155**(3): p. 733-740.
30. Westra, E.R., et al., *Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition.* PLoS Genet, 2013. **9**(9): p. e1003742.
31. Zetsche, B., et al., *Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System.* Cell, 2015. **163**(3): p. 759-771.
32. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease.* Nature, 2014. **513**(7519): p. 569-573.
33. Jinek, M., et al., *Structures of Cas9 endonucleases reveal RNA-mediated conformational activation.* Science, 2014. **343**(6176): p. 1247997.
34. Blosser, T.R., et al., *Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex.* Molecular Cell, 2015. **58**(1): p. 60-70.

35. Swarts, D.C., J. van der Oost, and M. Jinek, *Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a.* Molecular Cell, 2017. **66**(2): p. 221-233.e4.
36. Jia, N., et al., *Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity.* Molecular Cell, 2018. **73**(2): p. 264-277.
37. Wang, L., et al., *Dynamics of Cas10 Govern Discrimination between Self and Non-self in Type III CRISPR-Cas Immunity.* Molecular Cell, 2018. **73**(2): p. 278-290.
38. You, L., et al., *Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference.* Cell, 2018. **176**(1-2): p. 239-253.
39. Zhu, Y., et al., *Shooting the messenger: RNA-targeting CRISPR-Cas systems.* Biosci Rep, 2018. **38**(3): p. 1-11.
40. Suloway, C., et al., *Fully automated, sequential tilt-series acquisition with Leginon.* J Struct Biol, 2009. **167**(1): p. 11-18.
41. Lander, G.C., et al., *Appion: an integrated, database-driven pipeline to facilitate EM image processing.* J Struct Biol, 2009. **166**(1): p. 95-102.
42. Scheres, S.H., *RELION: implementation of a Bayesian approach to cryo-EM structure determination.* J Struct Biol, 2012. **180**(3): p. 519-530.
43. Bai, X.C., et al., *Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles.* Elife, 2013. **2**: p. e00461.
44. Scheres, S.H., *Beam-induced motion correction for sub-megadalton cryo-EM particles.* Elife, 2014. **3**: p. e03665.
45. Crawley, A.B., J.R. Henriksen, and R. Barrangou, *CRISPRdisco: An Automated Pipeline for the Discovery and Analysis of CRISPR-Cas Systems.* The CRISPR Journal, 2018. **1**(2): p. 171-181.
46. Park, K.H., et al., *RNA activation-independent DNA targeting of the Type III CRISPR-Cas system by a Csm complex.* EMBO Rep., 2017. **18**(5): p. 826-840.
47. Jung, T.Y., et al., *Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity.* Structure, 2015. **23**(4): p. 782-790.
48. Han, W., et al., *A type III-B CRISPR-Cas effector complex mediating massive target DNA destruction.* Nucleic Acids Research, 2017. **45**(4): p. 1983-1993.
49. 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, 2011. **108**(25): p. 10098-10103.
50. Wiedenheft, B., et al., *RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions.* Proc Natl Acad Sci U S A, 2011b. **108**(25): p. 10092-10097.
51. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-821.
52. Deng, L., et al., *A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus.* Mol Microbiol, 2013. **87**(5): p. 1088-1099.
53. Hale, C.R., et al., *Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs.* Molecular Cell, 2012. **45**(3): p. 292-302.

SUPPLEMENTARY FIGURES

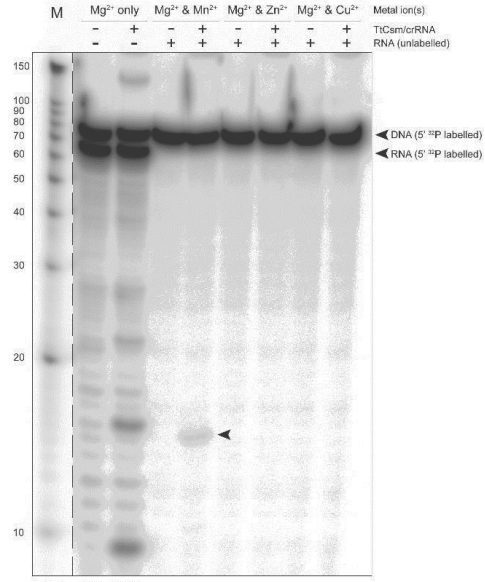


Figure S1. TtCsm *in vitro* DNase activity assay with metal ions

Four different metal ions were utilized to test the catalytic function in DNA *in vitro* cleavage assay.

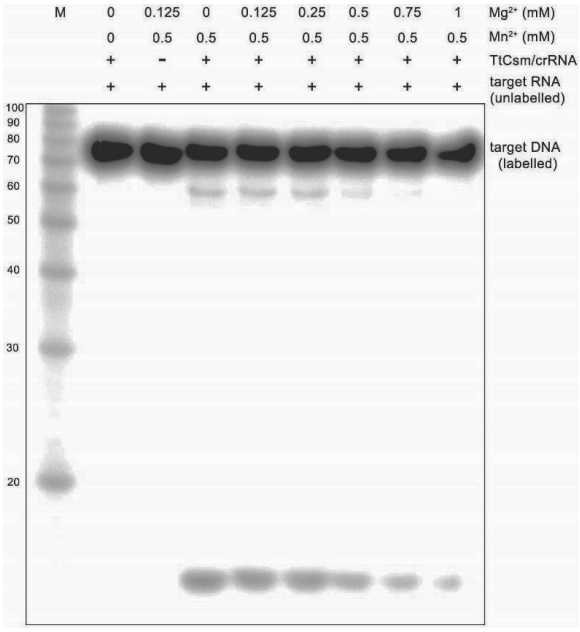


Figure S2. TtCsm *in vitro* DNA cleavage assay with Mg^{2+} and Mn^{2+}

Different molarity Mg^{2+} (from 0.125 mM to 1 mM as final concentration) and 0.5 mM Mn^{2+} were mixed to test the ssDNA cleavage ability with target RNA.

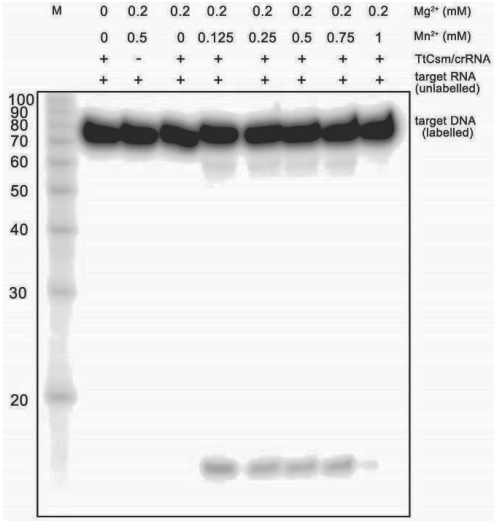


Figure S3. TtCsm in vitro DNA cleavage assay with Mg^{2+} and Mn^{2+} (based on the result from Figure S2)

Different molarity Mn^{2+} (from 0.125 mM to 1 mM as final concentration) and 0.2 mM Mg^{2+} were mixed to test the ssDNA cleavage ability with target RNA.

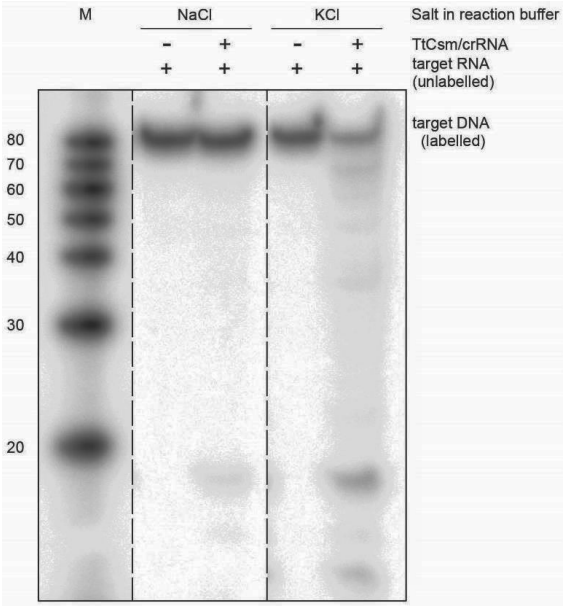


Figure S4. The effect of salt in in vitro reaction buffer for ssDNA cleavage

NaCl and KCl was tested to determine the optimal *in vitro* reaction buffer for TtCsm DNA cleavage assay.

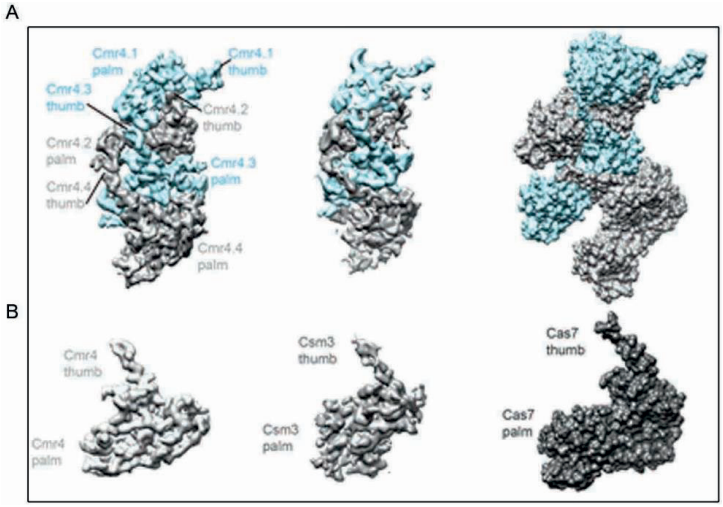


Figure S5. Structural homology across the backbone of type I and type III CRISPR-Cas systems

(A) Comparison of the thumb to palm backbone interactions of TtCmr4 (left), TtCsm3 (middle), and type I-E Cas7 (right) based on cryo-EM struture. (B) Individual backbone subunits from Cmr, Csm and Cascade with the thumb and palm indicated.

Chapter 8

General discussion

Over twenty years have passed since the first report of the repeat-spacer nucleotides sequence in *Escherichia coli* [1]. Initially, this research was first restricted to European and American groups that were interested in characterising this prokaryotic defence system in bacteria and archaea. However, when the genome editing properties of *Streptococcus pyogenes* Cas9 (SpCas9) were elucidated [2-4], this field became more popular and quickly expanded, especially regarding potential applications of the CRISPR-Cas technology. Consequently, it is important to further elucidate the molecular mechanism of CRISPR-Cas systems and their implications for further applications. Thus, in this thesis, I elucidated the molecular mechanisms of the type III CRISPR-Cas systems in *Thermus thermophilus* HB8.

MECHANISM

(r)PAM or PFS in invader sequence

The interference stage mediated by the CRISPR-Cas ribonucleoprotein (RNP) complexes have been studied in considerable detail in different bacteria and archaea [5-7]. The CRISPR-Cas RNP complex will recognize and degrade foreign nucleic acid but to avoid degradation of the host genome, this system employs several mechanisms to distinguish between self and non-self nucleic acid, the two most common mechanisms will be explored below. First, the most common mechanism, which exists in every dsDNA targeting CRISPR-Cas system, is the protospacer adjacent motif (PAM) [8-10]. PAM, a very short stretch of conserved nucleotides in the immediate vicinity (5’ or 3’) of the protospacer on the invader DNA is required to discriminate target versus non-target DNA [8, 10-12]. After forming a PAM-protein interaction, the dsDNA target will unwind and start base pairing to the crRNA [13]. An artificial PAM, PAM-presenting oligonucleotides (denoted “PAMmers”), mediating *in vitro* ssRNA cleavage is also observed in type II-A SpCas9 [14].

The second mechanism used by CRISPR-Cas to distinguish between host and foreign nucleic acid is the 3’ protospacer flanking site (PFS) found in type VI-A *Leptotrichia shahii* Cas13a (LshCas13a) and *L. buccalis* Cas13a (LbuCas13a) [15, 16]. For example, cleavage of RNA targets by LshCas13a requires a non-guanine residue directly flanking the 3’ end of the protospacer [15]. Hydrogen bonding between a 3’ guanine and a specific amino acid in Cas13a will trap this nucleotide in a cleft formed between the NTD and the Helical-1 domain of Cas13a. Consequently, base pairing between the RNA target and crRNA cannot be achieved, thereby negatively affecting RNA target cleavage [16].

Unlike the preceding CRISPR-Cas systems, type III-A systems recognize a PAM within the target RNA (denoted rPAM) to initiate DNA cleavage. Marraffini and Sontheimer were the first to report on the importance of mismatches with the 5’ handle of the crRNA to license mobile genetics elements (MGE) cleavage in the type III-A system of *S. epidermidis* [17]. In other words, base-pairing at this region between the 5’ handle crRNA and the DNA target will prevent autoimmunity [17, 18]. However, it was later discovered that this base pairing occurs between the crRNA and RNA target to block the ssDNA target cleavage [19, 20]. The latest structural and biochemical studies of StCsm confirms that a 4 nt stretch non-complementarity (position -2 to -5) between nucleotides in the 3’ tag of target RNA and the 5’ handle of the crRNA is critical for activating ssDNase activity. This could be because the mismatches will affect the conformation of the HD domain in the Cas10 (Csm1) subunit [21]. This fits with the results mentioned above and further suggests the universality of rPAM in dual-cleavage (both RNase and DNase activities) systems (chapter 7 and [22]).

This dual-cleavage activity might be especially relevant in the context of a prophage. In this scenario, a type III system will allow integration of a temperature phage as a prophage, but will prevent conversion to the lytic cycle by surveilling the production of viral transcripts [23].

Seed sequences in crRNA-guided immune complexes

A short fragment (denoted "seed sequence") located in the crRNA requires full base pairing with the corresponding sequence in the protospacer and contributes to target DNA recognition [24-27]. Such seed sequences have been analysed in detail by cryo-EM for type I, II, and V systems [2, 24, 25, 28, 29]. The RNP complexes in all these systems have variable protein structures that induce different conformations of the complex-bound crRNAs which may affect their target recognition and seed requirements. In type I Cascade systems for example, the 5′ and 3′ repeat handles of crRNA are tightly encased by the different subunit in these multisubunit complexes [30-38]. The spacer portion of the crRNA however, is distorted by the thumb domain of the Cas7 subunit, which intercalates every 6^{th} base of the crRNA sequence to form the 5 nt segments that can bind to the target DNA. Only the two most 5′ segments on contain a seed sequence (since they are easy accessible), while the rest of the segments are partly blocked by the Cse2 dimer [35]. Indeed, only positions 1-5 and 7-8 mediated the initial base pairing of the crRNA with the protospacer of the invader DNA [24].

In the type II system, only the 3′ repeat of crRNA is base-paired with the anti-repeat part of the tracrRNA and strongly bound to Cas9, while the 5′ part of the crRNA does not tightly interact with the protein. This causes the guide fragment adjacent to the 3′ repeat to form an 8-12 nt seed sequence [2, 36]. As a class 2 system, the CRISPR-Cas12a (formerly Cpf1) also employs seed sequences to recognize dsDNA targets, although variations in different Cas12a homologues can be discerned. For example, *Francisella novicida* Cas12a uses a pre-ordered seed for DNA binding [28], and *Lachnospiraceae bacterium* and *Acidaminococcus sp*. Cas12a both contain disordered seeds and undergo conformational rearrangement upon DNA target binding [28, 31, 37]. The type V CRISPR-Cas14a enzyme contains an internal seed sequence near the middle of the crRNA [39]. As an extremely compact RNA-mediated endonuclease (400 – 700 amino acids),

the structural (crystal or cryo-EM) study of Cas14 enzymes will reveal new insights into how the seed is positioned in these CRISPR-Cas systems.

In the RNA-targeting type VI-A LshCas13a RNP complex, a defined 2nt mismatch-sensitive region in the crRNA was identified [15]. However, further studies using Cas13a orthologues from *L. buccalis* suggests that the seed sequence spans a longer stretch of nucleotides (positions 9 to 15) in the central region of crRNA [16]. The A-form helical structure of the crRNA:target RNA duplex lies on the surface of Helical-2 domain of LbuCas13a and is exposed towards the solvent to facilitate the crRNA-target RNA base-pairing [16].

Despite our advances in our understanding how the seed region(s) are organized in type I, II, V and VI CRISPR-Cas system, the characterization of the seed in type III has remained enigmatic and, at times, conflicting for quite some years [19, 40-43]. In chapter 5 we have identified a potential answer for this, by showing that the type III-B TtCmr complex utilizes a flexible seed region, depending on the stoichiometry of the complex. Using our high-resolution cryo-EM structure (chapter 4), it was observed that the 3' seed-like region in the crRNA in TtCmr is easily accessible, which facilitates initiation of base pairing with the target RNA to trigger a concerted rearrangment of the remaining TtCmr subunits [44]. Future studies on other type III-B systems will have to reveal whether this is a common characteristic of these defense systems.

In conclusion, the seed(-like) sequence is a major determinant for DNA or in RNA target cleavage in both Class 1 and Class 2 CRISPR-Cas systems. In general, the exact position of the seed region seems to be determined by the exposed portions of the crRNA within these complexes.

Target cleavage ability

The type III system was initially considered to cleave the DNA target based on the *in vivo* assay of type III-A Csm complexes [45]. However, researchers rapidly found a subtype of this system (type III-B Cmr complex)

that degrades the RNA target *in vitro* [46]. Further studies concluded that both RNA and ssDNA targets can be cleaved by different subunits of the type III effector (chapter 3, 6 and 7) [18, 20, 22, 47-50]. The RNA target is cleaved at 6 nt intervals via the Csm3/Cmr4 subunit or between the UA nucleotides when the RNA target is aplenty [51], whereas the ssDNA target, is cleaved by the HD domain in the Cas10 protein (Csm1 or Cmr2) in type III systems [18, 20, 22, 52, 53].

This target DNA cleavage pattern appears to be different for type III subtypes. Currently, three ssDNA target cleavage modes have been reported: 1) RNA-activated sequence non-specific endonuclease activity, 2) RNA-activated T-site specific endonuclease activity (chapter 7 and [18, 20-22, 50, 54]), 3) RNA-independent DNase activity [55]. The first DNA cleavage mode is based on the *in vitro* result of SeCsm suggesting that the strand in target DNA with complementary to the crRNA can only be cleaved twice at its 3' region [18]. And further demonstrated via artificial bubbles (up to 38 nt mismatches between two DNA strands) to mimic the transcription bubble, while the real prokaryotic transcription bubble is around 8-10 nt [56]. As a result, this might cause the activity of Cas10 to become abnormal. More importantly, this design does not involve RNAP which makes the system simpler to operate but also omits the potential protection of unwounded ssDNA by RNAP. The potential protection may hamper the ability of the HD domain to cleave the DNA sequence.

The various types of DNA substrates (ssDNA plasmid, dsDNA plasmid, linearized DNA fragment) utilized in these modes contain ambiguous results, which makes generating a uniform interference model difficult. For example, five clear degradation bands are indicative for endonuclease activity associated with the TmCmr complex, whereas the diffused degradation products are reminiscent of exonuclease-like activity the StCsm system [49, 50]. Given the above mentioned biochemical activities of type III RNPs, it is clear that DNase activity relies on the target RNA binding in most cases. Further studies, including structural, biochemical, and

biophysical analysis, are required to better understand the interactions within the Csm/Cmr-crRNA-target RNA ternary complex and the allosteric activation in type III complexes [21, 40].

CRISPR-Cas AND PROGRAMMED CELL DEATH

Another unresolved issue is the promiscuous DNase or RNase activity in CRISPR-Cas systems. The type VI Cas13 RNP proteins, which exclusively cleave RNA, all contain two higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domains [15, 57-60]. The current speculation is that the indiscriminate RNA sequence cleavage activity of Cas13 RNPs elicits programmed cell death (PCD) [61].

Similarly, the indiscriminate DNA and RNA sequence cleavage observed in CRISPR-Cas type III systems further suggest a potential connection between antivirus immunity and PCD or dormancy induction [20, 22, 62-65]. In these systems, the Cas10 proteins (Csm1 or Cmr2) modulate the activity of Csm6 or Csx1 proteins via the production of cyclic oligoadenylate (cOA) second messenger molecules [53, 66]. The cOA synthesized by Cas10 can bind the CRISPR-Associated Rossmann Fold (CARF) domain of Csm6 to trigger its promiscuous RNase conferred by its HEPN domain. Interestingly, it has been reported that a dedicated nuclease (containing the CARF-domain in *S. solfataricus*) can deactivate Csm6 by degrading cOA to control the activation of unselective RNA cleavage [67, 68]. Consider the evidence that HEPN superfamily as dormancy- or PCD-inducing toxins [69, 70], and type III-associated proteins with a CARF domain [71], have resulted in a bold hypothesis: CRISPR-Cas linked CARF domain proteins elicit dormancy or PCD when immunity fails, most likely via various signal transduction pathways.

POSSIBLE INHIBITORS OF CRISPR-Cas TYPE III SYSTEMS

Almost all cellular organisms have evolved diverse, multilayer defense mechanisms to deal with the presence of ubiquitous viruses and other mobile genetic elements (MGEs), one of which, the adaptive immunity

effector CRISPR-Cas has been the focus of this research. However, even this highly potent, precise, and elaborate defense system is not without competitors, and they occur in the form of anti-CRISPRs. The first case of anti-CRISPR protein (Acr) against CRISPR-Cas system was found in the genomes of bacteriophages infecting *Pseudomonas aeruginosa* [72]. Because there are two CRISPR-Cas classes, six types, and around 29 subtypes discovered to date [73], Acrs also evolved specifically for certain CRISPR-Cas types or subtypes.

So far, all discovered Acrs are against type I, II, V CRISPR-Cas systems and function through a distinct mechanism [72, 74-80]. The discovered Acr proteins lack sequence similarity to other proteins with a known structure and function. Furthermore, they do not share conserved sequences with each other making identification difficult [81]. However, *acr* genes can form clusters in virus genomes and on prophage regions [81, 82], analogous to the way that host organisms often form genomic islands that encode antivirus related proteins [83, 84]. And a remarkable feature of *acr* genes is that they are frequently followed by highly conserved anti-CRISPR associated (aca) genes encode proteins containing helix-turn-helix (HTH) DNA-binding domain [85]. Thus, we can use this feature to predict new Acrs and further demonstrated the prediction via biochemical or molecular biology approaches. Using this knowledge, a bioinformatic and biochemical combinatorial approach should be used to screen for potential type III anti-CRISPR proteins, which are currently lacking.

An alternative strategy is to screen bacterial genomes, that contain self-targeting spacers, for genes encoding for proteins with anti-CRISPR activity [76, 80]. The hypothesis begin this is that host cells are still alive when the bacterial genome (most likely a prophage region) possess a target DNA sequence which perfectly matches with a spacer in a CRISPR array in the same genome (called genomic self-targeting) if the CRISPR-Cas effector is inactivated by inhibitor. The metagenome data can be screened based on

these two approaches to predict potential novel Acrs against the CRISPR-Cas type III system.

The third possible approach is to enrich the virus-like particles (VLPs) from environmental samples via centrifugation and polyethersulfone (PES) membrane filtration. The inoculate can then be filtrated with different microbial strains containing CRISPR-Cas type III systems. After a few successive rounds of enrichment, cell-free supernatants will be subjected to metagenomic sequencing. Analysis of the assembled sequencing reads will demonstrate which supernatant is dominated by bacteriophage. This working pipeline was utilized to find the crAssphage virion [86].

EVOLUTION

The systematic analyses of (meta)genomic sequencing data and protein structures has shed light on the origins and evolution of CRISPR-Cas systems. In general, it appears that several classes of MGEs have contributed to the evolution of the CRISPR-Cas systems. For example, the newly identified superfamily of self-synthesizing transposons – the Casposon, likely share the same ancestry with the adaptation module of CRISPR-Cas system as it may integrate next to genes involved in innate immunity [87-90]. Perhaps, the Cas2 protein, prototype CRISPR repeats, and the leader sequence have originated from these Casposon, thus forming the current CRISPR-Cas adaptation module [90-92].

The class 1 CRISPR-Cas effectors likely evolved from an RRM (RNA Recognition Motif) domain-containing nuclease (polymerase or cyclase), as is suggested by the ancestry of Cas10 [93, 94]. The RAMP (Repeat Associated Mysterious Protein) superfamily proteins (Cas5, Cas6, and Cas7) are proposed to originate from an ancestral Cas10, giving rise to the ancestral innate immunity module [94-97]. Several rearrangements of this innate immunity module and a random insertion of a Casposon nearby might have led to the emergence of the class 1 CRISPR-Cas effectors [97, 98]. The recently discovered cOA-Cas10 signalling pathway in CRISPR-Cas

type III systems not only suggests that CRISPR-Cas systems may have evolved from a signal transduction mechanism associated to cell dormancy or PCD [99, 100], but also gives rise to the hypothesis that the core unit of ancestral CRISPR-Cas consists of an ancestral protein containing the CARF-HEPN domain which is homologous to Csm6 or Csx1 and Cas10.

In the class 2 CRISPR-Cas system, it is clear that the RuvC-like domains in type II and type V are homologous to different types of TnpB-like transposases. The type II proteins evolved from ISC-like (Insertion Sequences Cas9-related) transposons [101], whereas the type V effector could be traced to IS605-like transposons [102]. Different TnpB-encoding transposons randomly insert close to CRISPR arrays resulting in the matured type II and type V effectors. This occurs through acquisition of additional domains that facilitate the interaction between crRNA and the target DNA in these two systems [7, 103]. In contrast, the ancestry of type VI system is hard to determine because of the profound sequence divergence of the HEPN superfamily [98]. It is likely that the type VI effector Cas13 and the type III Csm6 and Csx1 proteins share a common ancestor as they share promiscuous RNA degradation ability and both possess the HEPN domain. It is hypothesized that the type VI systems acquired a HEPN-containing Cas protein from a class 1/type III system and then evolved to a matured class 2 effector. However, an alternative scenario that type VI system may originate from HEPN-containing toxin-antitoxin (TA) modules, cannot be ruled out [98]. The cell growth inhibition ability of one HEPN-containing toxin protein in a TA module has been demonstrated [104], and similar results were obtained when the type VI Cas13a protein was co-expressed with its crRNA and target RNA [15]. Hence, further studies are required to reveal the origin of type VI systems, and to help us better understand the connection between immunity and dormancy and/or PCD in prokaryotes.

Taken together, these observation support the idea that different classes of MGE contributed to the evolution of the CRISPR-Cas immune system. The

widespread existence of class 1 systems in archaea and bacteria [105], along with the rearrangement and evolution of ancient RRM domain containing Cas10 protein in class 1 systems, suggest that the CRISPR-Cas systems are likely evolved from an ancestral class 1 type. The class 2 systems most likely appeared after that from replacement of some class 1 effectors with MGE-derived functional genes [7].

APPLIED USE OF CRISPR-Cas TYPE III SYSTEMS

Actually, in my opinion, the first experimental data to support the idea that CRISPR-Cas type III systems can be utilized for genome editing was reported by Terns and their work on the type III-B Cmr effector [106]. Combine the data from other two groups, type III complexes were demonstrated for programmed knock down mRNA of a selected gene in *S. solfataricus* and *S. islandicus* [107, 108]. These two successful cases demonstrated that the CRISPR-Cas type III system is a potential tool that interferes with the function of certain genes in archaea akin to the utility of RNAi in eukaryote. However, considering the structural complexity of type III systems, expressing the active complexes for genome editing in eukaryotes requires a substantial amount of laboratory work. Meanwhile, purified reconstituted type III complexes with designed crRNA are reported numerous times in the literature and thus, an alternative approach is to deliver reconstituted type III protein with engineered or synthesized crRNA against specific mRNA via microinjection, amphipathic peptide, or using viral vectors as has been described for Cas9 and Cas12a (Cpf1) [109-111]. Another technical issue concerns type III complexes is the dual nuclease activity. Inactivating RNase (Cas7 family) or DNase (Cas10 family) activity of this multiple function RNP via site-directed mutagenesis to eliminate unexpected side effect may be necessary for dedicated applications. For RNA editing, one good candidate could be the TtCmr complex since the HD domain is deleted in the native Cmr2 subunit.

Single subunit of the type III complex or even single Cas protein could be selectively harnessed for application purposes. For example, the Cas10 subunit or the Csm6/Csx1 protein. However, *in vitro* activity analyses suggest that these proteins will specifically cleave DNA or RNA [21, 53, 65, 66, 112]. The Csm6 protein has been developed to increase the signal sensitivity and for a detection platform called SHERLOCK (specific high-sensitivity enzymatic reporter unlocking) [113, 114]. Also, given the binding affinity of short linear oligoadenylate molecules to the N-terminal CARF domain in Csx1 protein [65], generating a fusion protein containing the Csx1 and then controlling its activity via poly(A) for specific application is possible.

The CRISPR-Cas type III system, combined with RNA targeting type II and type VI systems, created novel molecular tools for scientific study, although they only recently have been exploited as RNA programmable tools. They expand the boundary of the CRISPR-Cas based genome editing tool box, which previously focused solely on DNA targeting, and create new opportunities to discover the function of gene or non-coding RNA. Furthermore, editing genomes at the RNA level has fewer off-target consequences than on the DNA level. Both currently-established and future CRISPR-based RNA editing systems hold exciting potential to serve as novel diagnostic or therapeutic agents for the treatment of a wide range of human infectious and genetic diseases.

CONCLUDING REMARKS

The mechanism of CRISPR-Cas adaptive immunity in bacteria and archaea has been unravelled at a spectacular pace during the last decade. Many key questions concerning this unique defense system have been answered. Based on these insights, tremendous applications have been developed. Sometimes I have a feeling that we are reaching the edge of this area. However, new systems, new mechanisms and new applications of these systems are still emerging, making this field even more exciting and

surprising than ever. We currently know that these systems have multiple roles beyond immunity, hence genome editing or RNA editing will not be the only application direction.

REFERENCES

1. Ishino, Y., et al., *Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.* J Bacteriol, 1987. **169**(12): p. 5429-5433.
2. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-821.
3. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-823.
4. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-607.
5. Hille, F., et al., *The Biology of CRISPR-Cas: Backward and Forward.* Cell, 2018. **172**(6): p. 1239-1259.
6. Mohanraju, P., et al., *Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems.* Science, 2016. **353**(6299): p. aad5147.
7. Shmakov, S., et al., *Diversity and evolution of class 2 CRISPR-Cas systems.* Nat Rev Microbiol, 2017. **15**(3): p. 169-182.
8. Mojica, F.J., et al., *Short motif sequences determine the targets of the prokaryotic CRISPR defence system.* Microbiology, 2009. **155**(3): p. 733-740.
9. Sapranauskas, R., et al., *The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli.* Nucleic Acids Res, 2011. **39**(21): p. 9275-9282.
10. Shmakov, S., et al., *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems.* Molecular Cell, 2015. **60**(3): p. 385-397.
11. Swarts, D.C., et al., *CRISPR interference directs strand specific spacer acquisition.* PLoS One, 2012. **7**(4): p. e35888.
12. Deveau, H., et al., *Phage response to CRISPR-encoded resistance in Streptococcus thermophilus.* J Bacteriol, 2008. **190**(4): p. 1390-1400.
13. Sashital, D.G., B. Wiedenheft, and J.A. Doudna, *Mechanism of foreign DNA selection in a bacterial adaptive immune system.* Molecular Cell, 2012. **46**(5): p. 606-615.
14. O'Connell, M.R., et al., *Programmable RNA recognition and cleavage by CRISPR/Cas9.* Nature, 2014. **516**(7530): p. 263-266.
15. Abudayyeh, O.O., et al., *C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector.* Science, 2016. **353**(6299): p. aaf5573.
16. Liu, L., et al., *The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a.* Cell, 2017. **170**(4): p. 714-726.
17. Marraffini, L.A. and E.J. Sontheimer, *Self versus non-self discrimination during CRISPR RNA-directed immunity.* Nature, 2010. **463**(7280): p. 568-571.
18. Samai, P., et al., *Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity.* Cell, 2015. **161**(5): p. 1164-1174.
19. Peng, W., et al., *An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference.* Nucleic Acids Res, 2015. **43**(1): p. 406-417.
20. Elmore, J.R., et al., *Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system.* Genes Dev, 2016. **30**(4): p. 447-459.
21. You, L., et al., *Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference.* Cell, 2018. **176**(1-2): p. 239-253.
22. Kazlauskiene, M., et al., *Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition.* Molecular Cell, 2016. **62**(2): p. 295-306.
23. Goldberg, G.W., et al., *Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting.* Nature, 2014. **514**(7524): p. 633-637.
24. 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, 2011. **108**(25): p. 10098-10103.
25. Wiedenheft, B., et al., *RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions.* Proc Natl Acad Sci U S A, 2011b. **108**(25): p. 10092-10097.
26. Fonfara, I., et al., *The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA.* Nature, 2016. **532**(7600): p. 517-521.
27. Zetsche, B., et al., *Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System.* Cell, 2015. **163**(3): p. 759-771.
28. Swarts, D.C., J. van der Oost, and M. Jinek, *Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a.* Molecular Cell, 2017. **66**(2): p. 221-233.e4.
29. Jain, I., et al., *Defining the Seed Sequence of the Cas12b CRISPR-Cas Effector Complex.* RNA Biol, 2018: p. 1-10.
30. Jackson, R.N., et al., *Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli.* Science, 2014. **345**(6203): p. 1473-1479.
31. Dong, D., et al., *The crystal structure of Cpf1 in complex with CRISPR RNA.* Nature, 2016. **532**: p. 522–526.
32. Jinek, M., et al., *Structures of Cas9 endonucleases reveal RNA-mediated conformational activation.* Science, 2014. **343**(6176): p. 1247997.
33. Nishimasu, H., et al., *Crystal structure of Cas9 in complex with guide RNA and target DNA.* Cell, 2014. **156**(5): p. 935-949.
34. Yamano, T., et al., *Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA.* Cell, 2016. **165**(4): p. 949-962.

35. Zhao, H., et al., *Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli.* Nature, 2014. **515**(7525): p. 147-150.
36. Jiang, W., et al., *RNA-guided editing of bacterial genomes using CRISPR-Cas systems.* Nat Biotechnol, 2013. **31**(3): p. 233-239.
37. Gao, P., et al., *Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition.* Cell Res, 2016. **26**(8): p. 901-913.
38. Liu, L., et al., *C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism.* Molecular Cell, 2017. **65**(2): p. 310-322.
39. Harrington, L.B., et al., *Programmed DNA destruction by miniature CRISPR-Cas14 enzymes.* Science, 2018. **362**(6416): p. 839-842.
40. Wang, L., et al., *Dynamics of Cas10 Govern Discrimination between Self and Non-self in Type III CRISPR-Cas Immunity.* Molecular Cell, 2018. **73**(2): p. 278-290.
41. Goldberg, G.W., et al., *Incomplete prophage tolerance by type III-A CRISPR-Cas systems reduces the fitness of lysogenic hosts.* Nat Commun, 2018. **9**(1).
42. Cao, L., et al., *Identification and functional study of type III-A CRISPR-Cas systems in clinical isolates of Staphylococcus aureus.* Int J Med Microbiol, 2016. **306**(8): p. 686-696.
43. Manica, A., et al., *Unexpectedly broad target recognition of the CRISPR-mediated virus defence system in the archaeon Sulfolobus solfataricus.* Nucleic Acids Res, 2013. **41**(22): p. 10509-10517.
44. Taylor, D.W., et al., *Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning.* Science, 2015. **348**(6234): p. 581-585.
45. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA.* Science, 2008. **322**(5909): p. 1843-1845.
46. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex.* Cell, 2009. **139**(5): p. 945-956.
47. Staals, R.H., et al., *Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus.* Molecular Cell, 2013. **52**(1): p. 135-145.
48. Staals, R.H.J., et al., *RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus.* Molecular Cell, 2014. **56**(4): p. 518-530.
49. Tamulaitis, G., et al., *Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus.* Molecular Cell, 2014. **56**(4): p. 506-517.
50. Estrella, M.A., F.T. Kuo, and S. Bailey, *RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex.* Genes Dev, 2016. **30**(4): p. 460-470.
51. Zhang, J., et al., *Multiple nucleic acid cleavage modes in divergent type III CRISPR systems.* Nucleic Acids Res, 2016. **44**(4): p. 1789-1799.
52. Jung, T.Y., et al., *Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity.* Structure, 2015. **23**(4): p. 782-790.
53. Niewoehner, O., et al., *Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers.* Nature, 2017. **548**(7669): p. 543–548.
54. Liu, T.Y., A.T. Iavarone, and J.A. Doudna, *RNA and DNA Targeting by a Reconstituted Thermus thermophiles Type III-A CRISPR-Cas System.* PLoS One, 2017. **12**(1): p. e0170552.
55. Park, K.H., et al., *RNA activation-independent DNA targeting of the Type III CRISPR-Cas system by a Csm complex.* EMBO Rep., 2017. **18**(5): p. 826-840.
56. Vassylyev, D.G., et al., *Structural basis for transcription elongation by bacterial RNA polymerase.* Nature, 2007. **448**(7150): p. 157-162.
57. East-Seletsky, A., et al., *Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection.* Nature, 2016. **538**(7624): p. 270–273.
58. East-Seletsky, A., et al., *RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes.* Molecular Cell, 2017. **66**(3): p. 373-383.e3.
59. O'Connell, M.R., *Molecular Mechanisms of RNA-Targeting by Cas13-containing Type VI CRISPR-Cas Systems.* J Mol Biol, 2018. **431**(1): p. 66-87.
60. Yan, W.X., et al., *Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein.* Molecular Cell, 2018. **70**(2): p. 327-339.e5.
61. Koonin, E.V. and F. Zhang, *Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices.* Bioessays, 2017. **39**(1): p. 1-9.
62. Jiang, W., P. Samai, and L.A. Marraffini, *Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity.* Cell, 2016. **164**(4): p. 710-721.
63. Niewoehner, O. and M. Jinek, *Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6.* RNA, 2016. **22**(3): p. 318-329.
64. Sheppard, N.F., et al., *The CRISPR-associated Csx1 protein of Pyrococcus furiosus is an adenosine-specific endoribonuclease.* RNA, 2016. **22**(2): p. 216-224.
65. Han, W., et al., *Allosteric regulation of Csx1, a type IIIB-associated CARF domain ribonuclease by RNAs carrying a tetraadenylate tail.* Nucleic Acids Research, 2017. **18**(5): p. 826-840.
66. Kazlauskiene, M., et al., *A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems.* Science, 2017. **357**(6351): p. 605-609.
67. Athukoralage, J.S., et al., *Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate.* Nature, 2018. **562**(7726): p. 277-280.
68. Rouillon, C., et al., *Control of cyclic oligoadenylate synthesis in a type III CRISPR system.* Elife, 2018. **7**: p. e36734.
69. Anantharaman, V., et al., *Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing.* Biol Direct, 2013. **8**(15).
70. Jia, X., et al., *Structure-function analyses reveal the molecular architecture and neutralization mechanism of a bacterial HEPN-MNT toxin-antitoxin system.* J Biol Chem, 2018. **293**(18): p. 6812-6823.

71. 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, 2018. **19**: p. 1-13.
72. Bondy-Denomy, J., et al., *Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system.* Nature, 2013. **493**(7432): p. 429-432.
73. Makarova, K.S., Y.I. Wolf, and E.V. Koonin, *Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?* The CRISPR Journal, 2018. **1**(5): p. 325-336.
74. Bondy-Denomy, J., et al., *Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins.* Nature, 2015. **526**(7571): p. 136-139.
75. Pawluk, A., et al., *A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa.* MBio, 2014. **5**(2): p. e00896.
76. Rauch, B.J., et al., *Inhibition of CRISPR-Cas9 with Bacteriophage Proteins.* Cell, 2017. **168**(1-2): p. 150-158.e10.
77. Pawluk, A., et al., *Naturally Occurring Off-Switches for CRISPR-Cas9.* Cell, 2016. **167**(7): p. 1829-1838.e9.
78. Harrington, L.B., et al., *A Broad-Spectrum Inhibitor of CRISPR-Cas9.* Cell, 2017. **170**(6): p. 1224-1233.e15.
79. Marino, N.D., et al., *Discovery of widespread Type I and Type V CRISPR-Cas inhibitors.* Science, 2018. **362**(6411): p. 240-242.
80. Watters, K.E., et al., *Systematic discovery of natural CRISPR-Cas12a inhibitors.* Science, 2018. **362**(6411): p. 236-239.
81. Borges, A.L., A.R. Davidson, and J. Bondy-Denomy, *The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs.* Annu Rev Virol, 2017. **4**(1): p. 37-59.
82. Pawluk, A., A.R. Davidson, and K.L. Maxwell, *Anti-CRISPR: discovery, mechanism and function.* Nat Rev Microbiol, 2018. **16**(1): p. 12-17.
83. Makarova, K.S., et al., *Defense islands in bacterial and archaeal genomes and prediction of novel defense systems.* J Bacteriol, 2011. **193**(21): p. 6039-6056.
84. Makarova, K.S., et al., *Dark matter in archaeal genomes: a rich source of novel mobile elements, defense systems and secretory complexes.* Extremophiles, 2014. **18**(5): p. 877-893.
85. Pawluk, A., et al., *Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species.* Nat Microbiolo, 2016. **1**(8): p. 1-6.
86. Shkoporov, A.N., et al., *PhiCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis.* Nat Commun, 2018. **9**(1): p. 4781.
87. Krupovic, M., et al., *Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity.* BMC Biol, 2014. **12**(36).
88. Koonin, E.V. and M. Krupovic, *Evolution of adaptive immunity from transposable elements combined with innate immune systems.* Nat Rev Genet, 2015. **16**(3): p. 184-192.
89. Béguin P, C.N., Koonin EV, Forterre P, Krupovic M, *Casposon integration shows strong target site preference and recapitulates protospacer integration by CRISPR-Cas systems.* Nucleic Acids Res, 2016. **44**(21): p. 10367-10376.
90. Krupovic, M., P. Béguin, and E.V. Koonin, *Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery.* Current Opinion in Microbiology, 2017. **38**: p. 36-43.
91. Makarova, K.S., et al., *A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action.* Biol Direct, 2006. **1**(7).
92. Kwon, A.R., et al., *Structural and biochemical characterization of HP0315 from Helicobacter pylori as a VapD protein with an endoribonuclease activity.* Nucleic Acids Res, 2012. **40**(9): p. 4216-4228.
93. Makarova, K.S., et al., *Evolution and classification of the CRISPR-Cas systems.* Nat Rev Microbiol, 2011. **9**(6): p. 467-477.
94. Makarova, K.S., Y.I. Wolf, and E.V. Koonin, *The basic building blocks and evolution of CRISPR-CAS systems.* Biochem Soc Trans, 2013. **41**(6): p. 1392-1400.
95. Koonin, E.V. and K.S. Makarova, *CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes.* RNA Biol, 2013. **10**(5): p. 679-686.
96. Makarova, K.S., et al., *Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems.* Biol Direct, 2011. **6**(38).
97. Faure, G., K.S. Makarova, and E.V. Koonin, *CRISPR-Cas: Complex functional networks and multiple roles beyond adaptive immunity.* J Mol Biol, 2018. **431**(1): p. 3-20.
98. Koonin, E.V. and K.S. Makarova, *Mobile genetic elements and evolution of CRISPR-Cas systems: all the way there and back.* Genome Biology and Evolution, 2017. **9**(10): p. 2812-2825.
99. Burroughs, A.M., et al., *Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling.* Nucleic Acids Res, 2015. **43**(22): p. 10633-10654.
100. Koonin, E.V. and K.S. Makarova, *Discovery of oligonucleotide signaling mediated by CRISPR-associated polymerases solves two puzzles but leaves an enigma.* ACS Chem Biol, 2017. **13**(2): p. 309-312.
101. Kapitonov, V.V., K.S. Makarova, and E.V. Koonin, *ISC, a novel group of bacterial and archaeal DNA transposons that encode Cas9 homologs.* J Bacteriol, 2015. **198**(5): p. 797-807.
102. Chylinski, K., et al., *Classification and evolution of type II CRISPR-Cas systems.* Nucleic Acids Res, 2014. **42**(10): p. 6091-6105.
103. Koonin, E.V., K.S. Makarova, and F. Zhang, *Diversity, classification and evolution of CRISPR-Cas systems.* Curr Opin Microbiol, 2017. **37**: p. 67-78.
104. Yao, J., et al., *Identification and characterization of a HEPN-MNT family type II toxin-antitoxin in Shewanella oneidensis.* Microb Biotechnol, 2015. **8**(6): p. 961-973.

105. Crawley, A.B., J.R. Henriksen, and R. Barrangou, *CRISPRdisco: An Automated Pipeline for the Discovery and Analysis of CRISPR-Cas Systems.* The CRISPR Journal, 2018. **1**(2): p. 171-181.
106. Hale, C.R., et al., *Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs.* Molecular Cell, 2012. **45**(3): p. 292-302.
107. Zebec, Z., et al., *CRISPR-mediated targeted mRNA degradation in the archaeon Sulfolobus solfataricus.* Nucleic Acids Res, 2014. **42**(8): p. 5280-5288.
108. Li, Y., et al., *Harnessing Type I and Type III CRISPR-Cas systems for genome editing.* Nucleic Acids Res, 2015. **44**(4): p. e34.
109. Vilarino, M., et al., *CRISPR/Cas9 microinjection in oocytes disables pancreas development in sheep.* Sci Rep, 2017. **7**(1): p. 17472.
110. Shen, Y., et al., *CRISPR-delivery particles targeting nuclear receptor-interacting protein 1 (Nrip1) in adipose cells to enhance energy expenditure.* J Biol Chem, 2018. **293**(44): p. 17291-17305.
111. Jo, D.H., et al., *Long-Term Effects of In Vivo Genome Editing in the Mouse Retina Using Campylobacter jejuni Cas9 Expressed via Adeno-Associated Virus.* Mol Ther, 2018. **27**(1): p. 130-136.
112. Ramia, N.F., et al., *Staphylococcus epidermidis Csm1 is a 3′-5′ exonuclease.* Nucleic Acids Res, 2014a. **42**(2): p. 1129-1138.
113. Gootenberg, J.S., et al., *Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6.* Science, 2018. **360**(6387): p. 439-444.
114. Myhrvold, C., et al., *Field-deployable viral diagnostics using CRISPR-Cas13.* Science, 2018. **360**(6387): p. 444-448.

Appendices

List of publications

Zhu, **Y**.[#], Taylor, D.W. [#], Stroobach, E., Klompe, S.E., Neupane, N., Sobieraj, A., de Vos, W.M., Brouns, S.J.J., Shinkai, Doudna, A.J., van der Oost, J., Staals, H.J.R., A flexible seed sequence regulates targeting by the type III CRISPR/Cmr complex. *Manuscript in preparation*

Zhu, **Y**., Taylor, D.W., Stroobach, E., Klompe, S.E., Kornfeld, J.E., Sobieraj, A., de Vos, W.M., Nogales, E., Shinkai, A., Brouns, S.J.J., Doudna, A.J., Staals, H.J.R., van der Oost, J., Biochemical and structural analysis of DNA targeting by the Thermus thermophilus CRISPR-Cas type III-A system. *Manuscript in preparation*

Künne, T., **Zhu**, **Y**., da Silva, F., Konstantinides, N., McKenzie, R.E., Jackson, R.N., Brouns, S.J.J., 2018. Role of nucleotide identity in effective CRISPR target escape mutations. Nucleic Acids Research 46 (19):10395-10404

Zhu, **Y.**, Klompe, S.E., Vlot, M., van der Oost, J., Staals, H.J.R., 2018. Shooting the messenger: RNA-targeting CRISPR-Cas systems. Bioscience Reports 38(3): 1-11

Swarts, D.C.[#], Szczepaniak, M.[#], Sheng, G.[#], Chandradoss, S.D., **Zhu**, **Y.**, Timmers, E.M., Zhang, Y., Zhao, H.T., Lou, J.Z., Wang, Y.L., Joo, C., van der Oost, J., 2017. Autonomous generation and loading of DNA guides by bacterial Argonaute. Molecular Cell 65(6): 985-998

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

Staals, H.J.R.[#], **Zhu**, **Y.**[#], Taylor, D.W.[#], Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K., Sakamoto, K., Suzuki, T., Dohmae, N., Yokoyama, S., Schaap, P.J., Urlaub, H., Heck, A.J.R., Nogales, E., Doudna, A.J., Shinkai, A., van der Oost, J., 2014. RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. Molecular Cell 56(4): 518-30

Swart, D.C.[#], Jore, M.M.[#], Westra, E.R., **Zhu**, **Y.**, Janssen, J.H., Wang, Y.L., Patel, D.J., Berenguer, J., Brouns, S.J.J., van der Oost, J., 2013. DNA-guided DNA interference by a prokaryotic Argonaute protein. Nature 507(7491): 258-261

Staals, H.J.R.[#], Agari, Y.[#], Maki-Yonekura S.[#], **Zhu**, **Y.**, Taylor, D.W., Duijn, E. van, Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K., Masuda, A., Dohmae, N., Schaap, P.J., Doudna, A.J., Heck, A.J.R., Yonekura, K., van der Oost, J., Shinkai, A., 2013. Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of *Thermus thermophilus*. Molecular Cell 52(1): 135-145

[#]Equal contributions

Overview of completed training activities

Discipline-specific activities

Meeting & conferences

- NWO-ALW molecular genetics metting. Lunteren (NL), 2014*
- CRISPR meeting 2014. Berlin (DE), 2014*
- 2nd Wageningen PhD symposiums. Wageningen (NL), 2015**
- CRISPR meeting 2015. New York (USA), 2015*
- NWO-ALW molecular genetics metting. Lunteren (NL), 2015*
- Viruses of microbes. Liverpool (UK), 2016*
- The Scientific Spring Meeting KNVM & NVMM. Arnhem (NL), 2017

Courses

- The Intestinal Microbiome and Diet in Human and Animal Health. Wageningen (NL), 2014*
- Host-Microbe Iteractomics. Wageningen (NL), 2014
- Biorefinery for Biomolecules. Wageningen (NL), 2015**
- Food and Biorefinery Enzymology. Wageningen (NL), 2015**
- SIAM metagenomics course. Nijmegen (NL), 2016

General courses

- VLAG PhD week. Baarlo (NL), 2014
- Academic Writing II. Wageningen (NL), 2014
- Project & Time Management. Wageningen (NL), 2016
- Scientific Writing. Wageningen (NL), 2016/2017
- Writing Grant Proposal. Wageningen (NL), 2018
- Illustrator for scientists. Wageningen (NL), 2018

Optionals

- PhD study tour MIB-SSB. California (USA), 2015*
- Bacterial Genetics group meetings
- PhD meetings Laboratory of Microbiology
- Microbiology Seminars

*poster presentation, **oral presentation

Acknowledgements

After staying in MIB for almost 7 years (from 4th of May 2012 to 3rd of May 2019), my time here has come to an end. The number 7 is the seeker, the thinker, the searcher of Truth. The number 7 doesn't take anything at face value – it is always trying to understand the underlying, hidden truths. I would not say that the number 7 has secrets but the description of number 7 in terms of numerology - mentioned above - perfectly defines the spiritual temperament of our lab. There have been many good moments that I will never forget and there are so many people I want to say thank you to.

First of all, I would like to express my deep gratitude to my promotors **Willem M**. **de Vos** and **John van der Oost**. **Willem**, I am always eager and excited to try out the new ideas that you suggest during our monthly discussions and I am grateful for the constructive feedback and encouragement you have given me throughout this PhD project. Most importantly, you have provided me with the opportunity to learn how to be an independent scientist. **John**, I have learned many things from you in the past five years; from how to systematically design *in vitro* assays, to how to critically consider results from diverse points of view. I admire your intelligence and your ability to remember intricate details from the literature that I tend to forget.

Many special thanks to my co-promotor **Raymond Staals**. Thank you for supervising me during the last six years. It was a great pleasure to work with you. You have always provided me with intelligent suggestions and this has truly helped me along the way, especially at times when I encountered difficulties during my research. I would like to say an especially big thank you for your encouragement during my search for a postdoctoral position. Besides work, you are also just a really nice guy and I have greatly enjoyed the times that we have tasted whisky together. I am grateful to have you as my supervisor.

Stan J.J. Brouns, thank you for your excellent suggestion which helped me to find the reaction system for the DNA cleavage of type III Csm protein. I also thank you for organizing the CRISPR meetings which have been very useful.

Thank you to all of my colleagues who I have shared an office with. **Marnix**, we were team members during a master course in 2012 and then became colleagues. I am amazed by how many students you can supervise at one time and I admire you for your skilful project managing capabilities. **Sjoerd**, you are the living Wikipedia in our office. I always greatly enjoyed talking to you, not only about science. Thank you for your help in the design of the plasmid in Csm project. **Prarthana**, I have great respect for you. You never give up and you always appear composed, no matter how difficult your project is. **Jorrit**, I wish you success not only in the prokaryotic Argonaute project but also in online gambling. **Wen**, I am constantly curious about how many smart ideas can pop out of your mind every day. **Melvin**, we have had very nice talks about Chinese cities, Chinese food and your amazing hybrid membrane project. I am very happy to work with you in the CRISPR-Cas genome editing project of *Thermoanaerobacter ethanolicus* JW200. **Jurre** and **Stijn**, I am so happy that I could work with you two in the isotope lab for CRISPR-Cas experiments.

During my PhD I collaborated with many experts outside of the WUR and I am grateful for all the exciting moments outside of the lab. Thank you: David W. Taylor, Jack E. Kornfeld, Eva Nogales, Jennifer A. Doudna, Albert J.R. Heck, Henning Urlaub, Kundan Sharma, Akeo Shinkai, and Yanli Wang.

A big thank you to all students whom I have supervised: Nirajan, Anna, Eline, Antoine, Sanne, and Sharon. It is hard to forget the impressive experiments that you have carried out. Some of you find your names in publications as a result of you hard work. As far as I know, at least four of you are now pursuing your own PhD and Antoine you are working in the

Valto BV. I wish all of you the very best! It was a pleasure to work with you all.

Daan, Ioannis, Nico, Tom, Edze, Serve, Richard, Elleke, Jasper, Sudarshan, Joyshree, Mark, Jeroen, Lione, Joep, Catarina, Thomas, Melvin, Constantinos, Despoina, Kees, Alex, James, Thijs, Ismael, Tessa, Caroline, Gerben, and Peter, it was fun to work together and to learn something from you but also to have lunch or dinner with you. Also Anja, Sjon, Wim, Philippe, Steven, Gosse, many thanks for all the support that you have provided for our lab.

To Rick, Aafke, Charlotte, Reina, Guy, Mattew, Gijs, Janella, Rosemarijn, Randi, Pia, Olga, and Jessica, I will always remember the good times we have shared in our corridor.

Xijie Guo, Jun Wang, Jie Tu, Jianping Zhang, Yurong Li, Li Ren, Juzhen Xin, Chunhui Zhang, I thank you all very much for the supervision at school when I was in Zhenjiang. I feel incredibly lucky that you are in my life.

I cannot forget my friends who helped me, or who have encouraged me in the last seven years: Yingdi Zhang, Junyi Zhang, Yuxi Deng, Jie Lian, Yuxin Zhang, Loo Wee Chia, Indraningrat Anak, Ziqin Tang, Taojun Wang, Xinyi Wang, Caifang Wen, Weiwei Shi, Chen Zhang, Ran An.

Pan Xiao, Yimin Deng, Yifei Jia, Yanling Hao, Jun Qiu, Kejia Li, Qin Zhou, Hao Yan, Yanlin Liao, Yan Zhou, Bingzhen Du, Yuehan Dou, Le Fan, Huchen Li, Xiao Dong, Alex Wang, Yuan Feng, Xu Cheng, Peng Peng, Ziying Huang, Ni Zhang, Kun Liu, Qi Liu, Yan Liu, Zhe Zeng, Zhuang Liu, Yue Han, Chunyue Zhang, Yangwenshan Ou, Zhan Huang, Jianxin Wang, Ruiqi Xu, and Zhiliang Liu, I will always remember the great memories we have together.

Last, but certainly not least, I would like to thank my parents for their continuous support! Without your understanding and support, I could not have written this thesis book.

Once again thank you all for your support!

The research described in this thesis was financially supported by the Dutch governments Gravitation grant (SIAM grant 024.002.002 to W.d.V.).

Cover design: Yifan Zhu, Wenjing Wang, Yuran Hu

Lay-out: Yifan Zhu

Printed: Proefschriftmaken||www.proefschriftmaken.nl