IDENTIFICATION AND FUNCTIONAL CHARACTERIZATION OF TWO DNA-TARGETING CRISPR-CAS IMMUNE EFFECTOR COMPLEXES FROM *PYROCOCCUS FURIOSUS*

by

SONALI MAJUMDAR

(Under the Direction of Michael P. Terns and Rebecca M. Terns)

ABSTRACT

There are multiple CRISPR-Cas immune systems that silence invader nucleic acids via diverse effector complexes in prokaryotes. CRISPR-Cas effector complexes are comprised of CRISPR RNAs (crRNAs) that contain short sequences captured from invaders and Cas proteins specific to the immune system type. Invader nucleic acids are recognized by crRNAs and cleaved by the complexes. The work presented here indicates that three distinct CRISPR-Cas effector complexes co-exist in the hyperthermophilic archaeon *Pyrococcus furiosus*. We previously characterized the *P. furiosus* Cmr (Type III-B) effector complex, which cleaves target RNAs. Here, we have isolated two novel complexes in *P. furiosus* by immunoprecipitation. Tandem mass spectrometry and immunoblotting analyses revealed complexes comprised of distinct modules of Cas proteins - Csa (Type I-A) and Cst (Type I-G) as well as Cmr (Type III-B). Deep sequencing and Northern analysis revealed that processed crRNAs derived from all seven of the CRISPR loci found in *P. furiosus* associate with all three complexes. Furthermore, we have reconstituted active Type I-A and Type I-G effector complexes from

recombinant Cas proteins and synthetic crRNA. Types I-A and Type I-G effector complexes each target and cleave non-self, double-stranded DNA (that contains a canonical Protospacer Adjacent Motif or PAM) in a crRNA-guided manner. We determined that six Csa proteins [Csa 4-1 (Cas8a2), Cas3", Cas3', Cas5a, Csa2 (Cas7), Csa5] and three Cst proteins [Cst1 (Cas8a2), Cas5t and Cst2 (Cas7)] are each essential components of active Types I-A complexes and I-G complexes, respectively. Mutational analysis shows Cas3" and Cas3, respectively are the effector nucleases of Type I-A and I-G complexes. We have mapped the sites of cleavages mediated by Type I-A and I-G complexes and developed new models for the molecular mechanisms of DNA silencing. Finally, we delineated some features of the Csa and Cst complex architecture. Native gel shift analysis and UV-induced RNA-protein crosslinking reveal that the Cas5 and Cas7 family proteins of the Csa and Cst complexes constitute the structural core of each complex, making direct contacts with the crRNA. The 8-nt repeat-derived 5' tag of the crRNA with 5'-OH group is critical for the assembly of Csa and Cst complexes.

INDEX WORDS: CRISPR, Cas, Csa, Cst, RNA, crRNA, *Pyrococcus furiosus*

IDENTIFICATION AND FUNCTIONAL CHARACTERIZATION OF TWO DNA-TARGETING CRISPR-CAS IMMUNE EFFECTOR COMPLEXES FROM *PYROCOCCUS FURIOSUS*

by

SONALI MAJUMDAR

BSc., Bangalore University, India, 2002

MSc., Bangalore University, India, 2004

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTORAL OF PHILOSOPHY

ATHENS, GEORGIA

2014

© 2014

Sonali Majumdar

All Rights Reserved

IDENTIFICATION AND FUNCTIONAL CHARACTERIZATION OF TWO DNA-TARGETING CRISPR-CAS IMMUNE EFFECTOR COMPLEXES FROM *PYROCOCCUS FURIOSUS*

by

SONALI MAJUMDAR

Major Professor:	Michael P. Terns Rebecca M. Terns
Committee:	Stephen L. Hajduk Sidney R. Kushner

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
December 2014

DEDICATION

I dedicate this dissertation to my family, especially my parents who have been constant pillars of strength and kept me motivated through this fairly long and adventurous journey in a distant land.

ACKNOWLEDGEMENTS

I would like to express sincere gratitude to my academic mentors Drs. Michael Terns and Rebecca Terns for helping in my professional growth and providing me the opportunity for this enriching experience. I would like to thank my committee members Dr. Stephen Hajduk and Dr. Sidney Kushner for their support and advice. The confidence I have in my scientific abilities currently, I owe it my mentors and committee member's guidance. I would also like to thank my co-workers especially Dr. Caryn Hale, Dr. Jason Carte and Joshua Elmore. We built the foundation of CRISPR research in the lab and learned a lot from each other through interactions, discussions and peer reviewing along the way. I would like to thank the previous and current members of Terns' lab for maintaining a healthy professional atmosphere and resource sharing. I would like to acknowledge the talented undergraduate researchers I had the opportunity to mentor and assist with a special mention to Marianne Ligon and William Colby Skinner for being tour de force in the Cst project.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

CHAPTER

1 INTRODUCTION AND LITERATURE REVIEW 1

2 THREE CRISPR-CAS IMMUNE EFFECTOR COMPLEXES CO-EXIST IN *PYROCOCCUS FURIOSUS* 59

3 CRISPR RNA-GUIDED DNA CLEAVAGE BY TYPE I-A CRISPR-CAS IMMUNE EFFECTOR COMPLEX IN *PYROCOCCUS FURIOSUS* 117

4 CRISPR RNA-GUIDED, CST COMPLEX OF *PYROCOCCUS FURIOSUS* GUIDES CAS3 NUCLEASE OF TYPE I-G SYSTEMS TO CLEAVE DNA TARGETS 181

5 DISCUSSION 233

APPENDIX

A ESSENTIAL FEATURES AND RATIONAL DESIGN OF CRISPR RNAS THAT FUNCTION WITH THE CAS RAMP MODULE COMPLEX TO CLEAVE RNAS 295

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Prokaryotes (bacteria and archaea) are under constant threat of invasion by viruses or phages, plasmids and other mobile genetic elements like transposons. Most bacteria and archaea have acquired one or more of diverse defense mechanisms to combat such invasions (Reviewed in (Chopin et al. 2005; Tock and Dryden 2005; Labrie et al. 2010; Stern and Sorek 2011). The prokaryotic host can restrict phage adsorption and entry of invader genome as the first line of defense. Restriction-Modification system encoded in 90% of sequenced bacterial and in some archaeal genome is another mode of defense. This system encodes methyl-transferases (that methylate the host genome as signatures for host recognition) and restriction endonucleases (that restrict invader genetic elements by cleaving non-methylated DNA in a sequence-specific manner). In case the phage enters the host cell and bypasses other lines of defense, some hosts can trigger a final suicide signal called as Abortive Infection mechanisms by stopping replication, transcription and translational machinery. Recently CRISPR (Clustered Regularly-Interspaced Short Palindromic Repeats)– Cas (CRISPR-Associated) systems have been found in approximately 40% sequenced bacteria and 90% sequenced archaea (Makarova et al. 2006b). CRISPR-Cas systems are RNA-based, adaptive and heritable immune systems that confer defense against genome invaders to prokaryotic hosts (Bolotin et al. 2005; Godde and Bickerton 2006; Grissa et al. 2007).

The length and number of CRISPR loci in a host varies (Jansen et al. 2002b; Bolotin et al. 2005; Godde and Bickerton 2006; Lillestol et al. 2006). A typical CRISPR locus contains an AT-rich leader region, followed by multiple units of short direct repeats, alternated by similar-sized variable and unique sequences (Figure 1.1). Multiple *in silico* studies revealed host genes associated with CRISPR loci (*cas*) encode putative nucleases, helicases and RNA-binding proteins (Jansen et al. 2002b; Haft et al. 2005a; Makarova et al. 2006b). *In silico* analyses showed the unique sequences in host CRISPR loci mapped to genetic elements of invaders like viruses and plasmids (Bolotin et al. 2005; Mojica et al. 2005a; Pourcel et al. 2005). These unique sequences will be referred to as 'invader-derived sequences'. These findings led to the hypothesis that CRISPR-Cas constituted a prokaryotic adaptive immune system. The first experimental evidence confirming this hypothesis came from phage challenge studies done in *Streptococcus thermophilus* (Barrangou et al. 2007). This study showed the host CRISPR loci actively acquired fragments of phage DNA, which in-turn conferred resistance to *S. thermophilus* against the corresponding phage. Multiple Cas proteins were identified as essential components for acquisition and phage defense.

Mechanism of CRISPR-Cas mediated immunity

Subsequent research has supported the proposed mechanistic model of CRISPR-Cas mediated immunity. CRISPR-Cas mediated immune mechanisms are categorized into three stages: adaptation, crRNA biogenesis and invader silencing (Figure 1.1). In the adaptation step fragments of invader DNA called protospacers are processed and integrated in the host CRISPR locus in a leader-proximal polarized manner. During the

biogenesis stage promoter elements in leader region drive transcription of CRISPR locus to form precursor CRISPR RNA (pre-crRNA) transcripts. The pre-crRNAs are processed to ultimately form small, mature CRISPR RNAs (crRNA). Typically mature crRNAs contain invader-derived sequences (hereafter referred as 'guide' in biogenesis and silencing steps) flanked by portions of repeat-derived sequences (tag) on 5' and/ or 3' ends. Finally, in the invader silencing step each mature crRNA assembles into an effector complex (crRNP) and guides homology-dependent recognition and cleavage of invader nucleic acids. Various Cas proteins mediate molecular mechanisms of adaptation, crRNA biogenesis and invader silencing.

Classification of CRISPR-Cas systems

In silico comparisons revealed considerable diversity in repeat sequences (Kunin et al. 2007) and *cas* gene content (Haft et al. 2005a; Makarova et al. 2006b) of CRISPR-Cas modules. As a result, 12 distinct CRISPR-Cas systems were identified. After exhaustive discussions, the field came to a consensus on a classification scheme. CRISPR-Cas systems are categorized into three broad classes (Type I, II and III) comprising 12 systems: Types I A-G, Types II A-C, Types III A and B (Makarova et al. 2011c; Vestergaard et al. 2014). Each CRISPR-Cas system corresponds to a specific Cas protein subtype (e.g Type I-A for Csa, Type I-E for Cse) (Figure 1.2). The Cas protein subtype names are based on their archetype organism (e.g. Csa is Cas subtype Apern (*Aeropyrum pernix*) and Cse is Cas subtype *E. coli*) (Haft et al. 2005a; Makarova et al. 2011c; Vestergaard et al. 2014). Cas1 and Cas2 are conserved across all CRISPR-Cas systems. Each broad class has a conserved 'signature' Cas protein: Cas3 in Type I, Cas9 in Type II

and Cas10 in Type III (teal, Figure 1.2). Distantly related subtype *cas* gene orthologs in different systems were assigned superfamilies: Cas5, Cas6, Cas7, Cas8, Cas9, Cas10 and small subunit (color coded, Figure 1.2). Distinct *cas* genes which did not share similarities with other subtypes retained their sub-type specific annotation (*csy*1-3, *cse*1, *cse*2, etc) (Haft et al. 2005a; Makarova et al. 2011c). For clarity on the identity of each Cas protein, both subtype and superfamily annotations will be specified hereafter [e.g. Csa2 (Cas7)]. Type I and Type III systems share some analogy in their *cas* gene content (Makarova et al. 2011c). While some prokaryotes contain only one CRISPR-Cas system (Type I-A in *A. pernix,* Type I-E in *E. coli*) others harbor multiple CRISPR-Cas systems, each with a distinct repeat class and Cas subtype (e.g. Types I-E, II-A and III-A in *S. thermophilus*). Type I and Type III systems are identified in both bacteria and archaea, but Type II systems have only been identified in bacteria to date (Haft et al. 2005a).

Adaptation in CRISPR-Cas systems

CRISPR-Cas mediated adaptation serves as a molecular memory through acquisition of invader-derived sequences from previous encounters. The molecular mechanisms of this universal step are less understood than crRNA biogenesis and interference. Ongoing research continues to unravel important details on protospacer recognition, the Cas proteins involved and features facilitating adaptation. It is speculated that mechanism of adaptation more conserved across different CRISPR-Cas systems than biogenesis and silencing.

Protospacer recognition

The selection of protospacers for acquisition is not a random process. Studies in multiple CRISPR-Cas systems show there are short, 2-5 nucleotide (nt) sequence motifs called Protospacer Adjacent Motif (PAM), which serve as recognition element for processing and acquisition of adjacent protospacer sequences in invader genome (Figures 1.1, 1.4) (Deveau et al. 2008; Mojica et al. 2009; Deng et al. 2013a). Analyses of adaptation events in *E. coli* (Type I-E) (Datsenko et al. 2012; Swarts et al. 2012; Yosef et al. 2012), *S. solfataricus* (Type I-A, III-B) (Erdmann and Garrett 2012), *S. thermophilus* (Type II-A) (Deveau et al. 2008; Jiang et al. 2013), *H. hispanica* (Type I-B) (Szczelkun et al. 2014) and *P. atrosepticum* (Richter et al. 2014) show invader-derived sequences primarily map to protospacers with adjacent PAM sequences. CRISPR-Cas systems recognize distinct and system-specific PAM sequences which relates to the class of repeat and Cas subtype content (Kunin et al. 2007; Mojica et al. 2009). PAM also serve as recognition motifs for distinguishing self vs invader DNA in CRISPR-Cas mediated silencing across Type I and Type II systems (Deveau et al. 2008; Mojica et al. 2009; Fischer et al. 2012b; Gasiunas et al. 2012b; Jinek et al. 2012; Magadan et al. 2012; van Duijn et al. 2012a; Elmore et al. 2013a).

Cas proteins involved in adaptation

Recent findings show Cas1 and Cas2, the highly conserved Cas proteins across all CRISPR-Cas systems are involved in adaptation in characterized Type I systems (Figure 1.3). Studies in *E. coli* reveal that Cas1 and Cas2 forms a stable complex *in vitro* (Nunez et al. 2014). Cas1 and Cas2 are the minimum protein components required for acquisition

of new invader-derived sequences *in vivo* (Datsenko et al. 2012; Yosef et al. 2012; Nunez et al. 2014). Cas1 homologs exhibit metal-dependent nuclease activity on single stranded RNA and DNA, double stranded linear and circular DNA, as well as branched DNA like holiday junctions *in vitro* (Han et al. 2009; Wiedenheft et al. 2009; Babu et al. 2011; Kim et al. 2013). Cas2 homologs are metal-dependent nucleases shown to cleave either single-stranded RNA (Beloglazova et al. 2008) or double-stranded DNA (Nam et al. 2012a) *in vitro*. Catalytic-inactive mutant of Cas1, but not Cas2 prohibits acquisition of invader-derived sequences in *E. coli* implying Cas1 nuclease plays a role in molecular mechanism of protospacer processing and steps involved in integration of new invader-derived sequences in between repeats of CRISPR locus.

In addition to Cas1 and Cas2, Csn2 mediates acquisition of invader-derived sequences in Type II systems (Figure 1.3). Genetic studies in *S. thermophilus* (Type II-A) demonstrate deletion of *csn2* inhibits acquisition (Barrangou et al. 2007; Garneau et al. 2010). Structural and biochemical analyses show Csn2 proteins are ring-like, tetrameric proteins in solution and bind double-stranded DNA *in vitro* (Nam et al. 2011; Ellinger et al. 2012; Koo et al. 2012; Lee et al. 2012). There is also indirect evidence of Cas4 playing a role in adaptation (Figure 1.3). In Type I-A system of *Thermoproteus tenax*, complex containing fused Cas1/2, Cas4 and Csa1 was reconstituted (Plagens et al. 2012a). In some Type I systems *cas1* and *cas4* genes are fused implying the proteins co-function (Figure 1.2) (Makarova et al. 2011c). Finally, *in vivo* analysis in Type I-B system of *Haloarcula hispanica* indicates native adaptation requires Cas4 in addition to Cas1 and Cas2 (Szczelkun et al. 2014). Elucidated Cas4 structure reveals a RecB-like protein, containing iron-sulfur clusters and bearing similarity with *B. subtilis* DNA repair

protein AddB. Cas4 exhibits metal-dependent DNA endnonuclease, DNA unwinding and 5' to 3' single-stranded RNA and DNA exonuclease activities *in vitro* (Zhang et al. 2012a; Lemak et al. 2013). Further, there is no evidence that Cas4 is involved in either crRNA biogenesis or invader silencing mechanisms in characterized Type I, II or III CRISPR-Cas system.

Important genetic elements of adaptation

Consistent features of adaptation in studied CRISPR-Cas systems involve acquisition of new invader-derived sequences in a leader-proximal position and each acquisition event resulting in duplication of repeat in a host CRISPR locus. Systematic acquisition assays done following sequence deletions and mutations reveal 60 nucleotides of the leader (directly upstream of the repeat) and adjacent repeat sequences in the host CRISPR locus constitute the minimal components essential for acquisition of invader-derived sequences in *E. coli* (Yosef et al. 2012). Additional studies performed by this group indicate leader-adjacent repeat is replicated during each acquisition event. Results suggest sequence and/ or structure elements in the leader-repeat regions may be essential for recognition and processing events required for integrating invader-derived sequences. Supporting evidence in *E. coli* show that Cas1 and Cas2 complex doesn't bind sequences mimicking the CRISPR locus, when 60 nucleotides of the leader are mutated, implying structural elements in leader-repeat as a potential determinant for acquisition steps (Nunez et al. 2014). Another recent study in *E. coli* detected and characterized transient acquisition intermediates *in vivo* (Arslan et al. 2014). In the presence of Cas1 and Cas2, both strands of leader proximal repeat are nicked in a staggered manner. The results

indicate processed invader-derived sequence is ligated between the single-stranded regions of the nicked repeat (Arslan et al. 2014). These studies provide important insight on molecular mechanism of adaptation.

In the ensuing arms race between a CRISPR-bearing host and an invader, a previously encountered invader can escape the CRISPR-Cas immunity of host by mutating its protospacer (Deveau et al. 2008; Datsenko et al. 2012). The partially homologous invader-derived sequence in CRISPR locus can no longer guide invader targeting. But this degenerate sequence can cause positive reinforcement by accelerating new acquisition events of multiple invader-derived sequences in order to silence the escaping invader on future encounters in a step termed "priming". This feature is currently observed in Types I-B, I-E and I-F systems (Datsenko et al. 2012; Swarts et al. 2012; Savitskaya et al. 2013; Fineran et al. 2014; Richter et al. 2014; Szczelkun et al. 2014).

Mechanistic model of adaptation

The more conserved Cas proteins are either characterized (Cas1, Cas2) or expected (Cas4) to play a role in adaptation, thereby indicating the molecular mechanisms of adaptation may be less diverse compared to crRNA biogenesis or invader silencing mediated by the diverse Cas subtype proteins. Based on collective genetic and biochemical evidence a mechanistic model of adaptation is proposed (Figure 1.4). The invader protospacer sequences adjacent to canonical PAMs are recognized and processed to form an invader-derived sequence. Simultaneously, sequence and/ or structure elements in the leader and adjacent repeat region of host CRISPR locus results in opening

up of the region and Cas1 mediated nicking in both strands of the first repeat followed by single-stranded resection by an exonoculease (Cas4 is a predicted candidate). The processed invader-derived sequence is integrated between the two repeat cleavage products. The single-stranded 3' overhangs are filled-in by either Cas proteins or non-Cas DNA repair proteins or polymerases resulting in repeat duplication and expansion of CRISPR locus by a repeat-spacer unit in a leader proximal manner.

crRNA biogenesis and processing

Specificity of CRISPR-Cas mediated invader silencing is provided by guide sequences of crRNAs. One of the key steps to assess an active CRISPR-Cas immune system in a host is done by evaluation of crRNA expression and processing. Similar processing mechanisms generate crRNAs in Types I and III systems and will be discussed together. Characterized Type II systems show distinct mechanism of crRNA biogenesis.

crRNA biogenesis in Type I and Type III systems

Early studies indicate the entire CRISPR locus is transcribed from promoter elements in the leader (Lillestol et al. 2006; Lillestol et al. 2009). Processing of the primary transcript involves a single cleavage in each repeat to generate mature crRNAs containing an invader-derived sequence flanked by partial repeats. In some systems these crRNAs are further trimmed to form shorter forms (Figure 1.5) (Tang et al. 2005; Brouns et al. 2008; Hale et al. 2008).

Cas proteins involved

Processing of the pre-crRNA transcripts in majority of the characterized Type I and Type III systems (except Type I-C) is catalyzed by Cas6 superfamily endonucleases (Figure 1.3, 1.5) (Brouns et al. 2008; Carte et al. 2008b; Haurwitz et al. 2010; Gesner et al. 2011b; Lintner et al. 2011b; Przybilski et al. 2011; Richter et al. 2012; Brendel et al. 2014; Carte et al. 2014b). Cas6 members and Cas6 variants like Cas6e (Cse3, Type I-E), typically contain two conserved ferredoxin-like folds or RRM domains and a C-terminal glycine rich loop (Carte et al. 2008b; Gesner et al. 2011b; Makarova et al. 2011a; Nam et al. 2012b; Reeks et al. 2013c; Shao and Li 2013). Cas6f proteins (Csy4, Type I-F) contain only one conserved ferredoxin-like fold or RRM domain (Haurwitz et al. 2010). Besides these conserved features, the Cas6 superfamily proteins show considerable variation in sequence and structure.

Biochemical characterizations reveal Cas6 superfamily proteins are RNA endonucleases that cleave within the repeat to generate crRNAs containing repeat-derived 8-nt 5'-tag and 3'-tag of variable length that flank (invader-derived) guide sequences. Cas6 homologs and Cas6e, Csa6f variants specifically recognize and bind distinct sequences and/or structure elements of repeat types (Kunin et al. 2007) associated with a particular CRISPR-Cas system. To specifically bind structured and/ or unstructured repeat RNAs, Cas6 of different systems adapt multiple modes of RNA binding, as well as RNA-induced conformation changes (Haurwitz et al. 2010; Gesner et al. 2011b; Sashital et al. 2011; Wang et al. 2011; Nam et al. 2012b; Sternberg et al. 2012; Shao and Li 2013; Niewoehner et al. 2014).

Cas6 are metal-independent endoribonucleases that generate 5' hydroxyl and 3' phosphate or 2'-3' cyclic phosphate ends on crRNAs (Brouns et al. 2008; Carte et al. 2008b; Haurwitz et al. 2010; Gesner et al. 2011b; Lintner et al. 2011b; Carte et al. 2014b). Although, an invariant histidine, tyrosine or serine and sometimes lysine constitute catalytic residues of most characterized Cas6 superfamily enzymes, their positions as well and active-site chemistry varies (Carte et al. 2008b; Gesner et al. 2011b; Sashital et al. 2011; Haurwitz et al. 2012; Wiedenheft et al. 2012; Reeks et al. 2013a; Brendel et al. 2014). Cas6 proteins in *S. solfataricus* lack the conserved histidine, tyrosine or serine residues and use positively charged amino acids for orientation and catalysis (Reeks et al. 2013c; Shao and Li 2013). An interesting feature is while Cas6 superfamily proteins exhibit diversity in structure, RNA substrates, modes of RNA binding and active-site catalysis they all generate the canonical 8-nt, 5'-tag.

Type I-C systems lack Cas6 superfamily endonuclease and use Cas5d, a catalytic variant of Cas5 superfamily for crRNA processing. Cas5d proteins contain one N-terminal RRM domain (ferredoxin folds) and the glycine rich loop. Cas5d homologs are metal-independent endonucleases that contain variable catalytic triads and cleave Type I-C associated repeats to generate 11-nt, 5'-tag (Garside et al. 2012; Nam et al. 2012b; Elmore et al. 2013a).

crRNA maturation

Cas6e (Type I-E) and Cas6f (Type I-F) proteins remain tightly bound to the 3'end of cleaved crRNA. As a result, these proteins are also seen as stable components of the Type I-E and I-F effector crRNPs sequestering the 3'end of crRNAs (Haurwitz et al. 2010;

Gesner et al. 2011b; Wiedenheft et al. 2011a; Haurwitz et al. 2012; Nam et al. 2012b; Sternberg et al. 2012; Niewoehner et al. 2014). Therefore, in Type I-E and I-F systems the primary cleavage within repeats generate mature crRNAs that assemble into invader targeting complexes (Figure 1.5). A recent study presents evidence of Cas6 containing stable crRNP in a Type I-B system of *H. volcani* (Brendel et al. 2014).

Cleavage products of Cas6 homologs in other Type I and all Type III systems are termed 1X intermediate RNAs that undergo further trimming of repeats on 3' end by unknown mechanisms (Figure 1.5). crRNAs associated with Type I-A (Lillestol et al. 2006; Lillestol et al. 2009; Lintner et al. 2011b; Plagens et al. 2014b), Type I-B (Richter et al. 2012; Nickel et al. 2013b), Type I-D (Scholz et al. 2013) systems contain heterogenous and variable amounts of 3' repeat sequences. However, mature crRNAs of Type III systems constitute distinct size forms independent of guide length, lack any 3' repeat-derived sequence and are sometimes truncated at 3' end of guide (Hale et al. 2009; Hatoum-Aslan et al. 2011b; Hale et al. 2012a; Zhang et al. 2012c; Hatoum-Aslan et al. 2013b). Although the mechanism of 3' end processing is not known, evidence in Type III-A system indicate involvement of few subtype-specific Cas proteins (Csm2, Csm3, Csm5) (Hatoum-Aslan et al. 2011b; Hatoum-Aslan et al. 2013b).

crRNA biogenesis in Type II systems

crRNA biogenesis in Type II systems involve a distinct mechanism. Type II-A associated crRNA contains truncated (~2/3rd) invader-derived guide sequence followed by 17-22 nt repeat-derived, 3'-tag. These crRNAs completely lack any repeat-derived, 5'-tag sequences (Deltcheva et al. 2011; Carte et al. 2014b). Two proteins, Cas9 and

endogenous RNase III and a non-coding RNA called trans-activating crRNA (tracrRNA) are essential for crRNA biogenesis in Type II systems (Figure 1.3) (Deltcheva et al. 2011). tracrRNA is transcribed upstream of the associated CRISPR array-Cas module and bears almost perfect complementarity to the associated repeat RNAs. The first processing step involves RNase III cleavage of tracrRNA-repeat RNA duplexes. The processed tracrRNA-crRNA intermediates are further trimmed in the guide sequence by an unknown nuclease that generates processed tracrRNA-mature crRNA hybrids. The precise role of Cas9 in this step is not elucidated but recent observation provides further insight. Cas9 exhibits specificity to its associated CRISPR locus and doesn't function in crRNA biogenesis of other Type II-associated CRISPR loci in the same organism that contain different repeat and tracrRNA sequences. Also, Cas9 deletion does not affect the profile of RNAse III-cleaved tracrRNA-crRNA intermediate (Carte et al. 2014b). This implies Cas9 probably binds sequence-specifically, positions tracrRNA-pre-crRNAs duplexes and directs downstream processing events to generate processed tracrRNA-mature crRNA species.

CRISPR-Cas mediated invader silencing

Mature crRNAs associate with subtype-specific Cas proteins in different systems to form invader (DNA or RNA) silencing effector complexes (crRNPs). While Type I and Type III systems form multi-protein crRNPs, Type II systems form a single-protein (Cas9) crRNP. Type I and Type III crRNPs show some similarities in complex topology and architecture but involve distinct mechanisms for invader silencing.

Invader silencing in Type II systems

Type II systems mediate DNA silencing via crRNP that contains Cas9 and crRNA-tracrRNA duplex (Garneau et al. 2010; Deltcheva et al. 2011; Gasiunas et al. 2012b). Like Type I systems invader DNA recognition is PAM-dependent in Type II systems (Deveau et al. 2008). In Type II systems the PAMs are positioned upstream of protospacer sequences on crRNA-complementary (target) strand of DNA and is located opposite to 3'-repeat-derived tag of crRNA (Figure 1.3). Targeted invader silencing is determined by complementarity of PAM-adjacent protospacer sequences with a minimal 12-nt seed sequence at the 3' end of crRNA guide and located upstream of 3'-tag (Sternberg et al. 2014).

Structure elucidation of Cas9 enodonucleases bound to crRNA-tracrRNA and DNA heteroduplex provide mechanistic insight (Anders et al. 2014; Jinek et al. 2014; Nishimasu et al. 2014). Two arginine residues in C-terminal domain of Cas9 in *S. pyogenes* recognize and interact with PAM nucleotides (Anders et al. 2014). Crosslinking experiments in another study implicated two tryptophan residues (residing in the loop region) at C-terminal domain of Cas9 are likely transiently involved in PAM recognition (Jinek et al. 2014). Crystal structure and EM reconstruction analysis indicate Cas9 contain two lobes: nucleic acid recognition lobe and nuclease lobe (Anders et al. 2014; Jinek et al. 2014; Nishimasu et al. 2014). Association with crRNA-tracrRNA triggers an active conformation of Cas9. crRNA-DNA heteroduplex is loaded in a central channel at the interface of two lobes. crRNA-guided DNA cleavage is catalyzed by two active site domains in the nuclease lobe of Cas9: HNH-like domain and RuvC-like domain. The HNH-type domain catalyzes cleavage in the target strand (crRNA-complementary) and

RuvC-like domain catalyzes the opposite non-target strand cleavage that results in two blunt-end cleavages (Garneau et al. 2010; Jinek et al. 2012; Magadan et al. 2012; Chen et al. 2014; Jinek et al. 2014). Type II crRNP mediated DNA cleavage has been harnessed as a powerful tool for targeted genome engineering and regulation of gene expression in eukaryotes (Recently reviewed in (Hsu et al. 2014; Terns and Terns 2014)).

Type I invader-silencing crRNPs

Assembly and structure of effector complexes

Cas7 superfamily proteins. High-resolution structures by electron-microscopy (EM) (Wiedenheft et al. 2011a) and X- ray crystallography (Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014) of *E. coli* Type I-E crRNP, low-resolution structures of *B. halodurans* Type I-C by preliminary EM and single-particle scattering (Nam et al. 2012b) and *P. aeruginosa* Type I-F crRNP by native-MS, EM and X-ray scattering (SAXS) (Wiedenheft et al. 2011a; van Duijn et al. 2012a) are elucidated. All crRNP structures reveal a similar helical backbone consisting of multiple (six) Cas7 units (Cse4 in Type I-E, Csd2 in Type I-C) and Csy3 (Type I-F) that wrap the crRNA (Figure 1.6). Reconstitution of Type I-A (Lintner et al. 2011b; Plagens et al. 2012a) and Type I-B crRNPs (Brendel et al. 2014) indicate the formation of a similar backbone formed by their Cas7 proteins (Csa2, Csh2 respectively). Structure of *S. solfataricus* Csa2 (Cas7 in Type I-A) reveals a three-domain protein (a modified RRM central domain and two additional domains as insertions into central domain) with a crescent-shaped appearance (Lintner et al. 2011b). Two conserved clusters were seen on the concave surface of this Cas7 (Csa2) protein similar to the-crRNA binding sites of Cas7 (Cse4) in Type I-E

crRNP (Wiedenheft et al. 2011a; Hochstrasser et al. 2014a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014). Structure of Cas7 (Csc2) in Type I-D system of *Thermofilum pandens* reveals a similar RRM-like core domain and three insertion domains (Hrle et al. 2014). crRNA binding was mapped to a positively charged surface patch by crosslinking and mass spectrometry (Hrle et al. 2014) similar to observations in Type I-A and I-E Cas7 proteins. Comparative analyses of Cas7 proteins (Csa2, Csc2, Cse4) of different Type I systems indicate homology in core (RRM) and insertion domains structures despite lack of sequence similarity and similarity in mode of crRNA binding.

Cas5 superfamily proteins. Another shared feature among characterized Type I crRNPs is the formation of a core crRNP that comprises crRNA, Cas5 superfamily proteins (Cas5a in I-A, Cas5e in I-E or its counterpart Csy2 in I-F) and Cas7 members. Resolved structures of crRNPs show Cas5e in Type I-E and Csy2 in Type I-F crRNPs interact with 5'-tag of crRNA (Figure 1.6) (Jore et al. 2011a; Wiedenheft et al. 2011a; van Duijn et al. 2012a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014). The structure of Type I-E crRNP indicates Cas5e has a similar core RRM domain like Cas7 (Cse4). Conserved polar residues in Cas5 interact with nucleotides of 5'-tag (Mulepati et al. 2014a). Among Cas5 superfamily proteins only Cas5d is a known catalytic active variant and acts as the biogenesis enzyme of Type I-C systems. Type I-C crRNP contains two copies of Cas5d, one at the 5'-tag and other sequestering the 3' end of the crRNA (Nam et al. 2012b).

Cas6 superfamily proteins. In Type I-E and I-F systems the biogenesis enzymes (Cas6e, Cas6f) remain bound to the 3'-tag of crRNA after processing and sequesters the crRNA from further 3'-end trimming. This size form is mature crRNA that associates with Type I-E and I-F crRNPs with Cas6e and Cas6f (respectively) as structural components of the complex (located at the 3' end of crRNA). (Figure 1.6) (Jore et al. 2011a; van Duijn et al. 2012a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014). Cas6 is also identified in a stable *H. volcani* Type I-B crRNP (Brendel et al. 2014). Unlike these complexes there is no evidence of Cas6 as a component of recently characterized *T. tenax* Type I-A crRNP (Plagens et al. 2012a; Plagens et al. 2014b).

Large-subunit proteins. Additionally, *cas* gene clusters of Type I systems encode large subunit proteins (Cas8 superfamily proteins of Types I-A, I-B, I-C systems or Cse1 in Type I-E and Csy1 in Type I-F systems) (purple, Figure 1.2). Resolved structures of Cse1 homologs reveal a novel α/β fold domain and four helix bundles on C-terminus (Mulepati et al. 2012; Sashital et al. 2012b). Structural studies reveal Cse1 is a less-stable component of Type I-E crRNP (Figure 1.6), (Jore et al. 2011a). Large subunit proteins Cas8a2 (Csa4 in Type I-A) (Lintner et al. 2011b; Plagens et al. 2012a; Plagens et al. 2014b) and Csy1 (Type I-F) (Wiedenheft et al. 2011a) are also identified in assembled crRNPs.

Small-subunit proteins. Some Type I systems encode small subunit-proteins (proposed to be classified as Cas11 superfamily, colored orange in Figure 1.6). Structures of small subunit proteins Csa5 (Type I-A) (Reeks et al. 2013a)

and Cse2 (I-E) (Agari et al. 2008; Nam et al. 2012c) reveal homology between α-helical domains of Csa5 (N-terminus) and Cse2 (C-terminus). Small-subunit protein (Cse2) dimer is an integral part of Type I-E crRNP making contacts with Cas6e on one end, Cas7 (Cse4) units in the middle and large subunit (Cse1) and Cas7 (Cse4) at the other end (Figure 1.6) (Jore et al. 2011a; Wiedenheft et al. 2011a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014). Small subunit protein (Cse2) doesn't directly interact with crRNA in the complex. Evidence indicates Csa5 is a component of active Type I-A crRNP in *T. tenax* (Plagens et al. 2012a; Plagens et al. 2014b) and Type I-A complex in *S. solfataricus* (Reeks et al. 2013a). Based on *in silico* analyses it is predicted Type I systems that lack small subunit proteins (Type I-B, I-C, I-D) contain a α-helical domain fused at the C-terminus of large subunit (Cas8) proteins (Makarova et al. 2011c).

In summary, remarkable homology is observed with respect to subunit composition, complex architecture, mode of crRNA binding and complex assembly among characterized effector complexes of Type I systems.

Invader silencing by Type I crRNPs

Genetic determinants of targeting. Invader DNA silencing characterized *in vivo* in few Type I (I-A, I-B, I-E and I-F) systems (Brouns et al. 2008; Cady and O'Toole 2011b; Fischer et al. 2012b; Elmore et al. 2013a; Maier et al. 2013b; Peng et al. 2013b) and predicted in others are expected to employ a common mechanism due to similarities in crRNP architecture. Recognition and targeting of invader DNA is determined by type-specific PAM and homology of adjacent protospacer of invader DNA with guide of crRNAs associated with Type I surveillance crRNPs (Wiedenheft et al. 2011a; Elmore et

al. 2013a; Maier et al. 2013b; Peng et al. 2013b; Westra et al. 2013). Homology of protospacer to a minimal seed sequence on 5' ends of crRNA (downstream of 5'-tag) constitutes minimal sequence determinant of DNA silencing by characterized Type I-B (Maier et al. 2013b), Type I-E (Semenova et al. 2011) and Type I-F (Wiedenheft et al. 2011a) systems. PAM in Type I systems occurs downstream of protospacer on crRNA-complementary (target) strand that positions it opposite to 5'-tag of crRNAs (Figure 1.6). Therefore, the protospacer sequences that are homologous to crRNA seed are directly adjacent to PAM.

Role of large-subunit proteins. Extensive biochemical investigations of Type I-E crRNP-mediated DNA silencing have provided important mechanistic insights (Figure 1.6). Large-subunit (Cse1) of Type I-E crRNPs scans and its disordered loop L1 interacts with the PAM (Sashital et al. 2012b; Hochstrasser et al. 2014a). Cse1 interaction with PAM further causes destabilization of the invader DNA duplex resulting in single-stranded regions in the adjacent protospacer (Jore et al. 2011a; Hochstrasser et al. 2014a). Additionally, interactions between Cse1 and two Cas7 (Cse4) subunits (5, 6), placed at the expected crRNA seed region are observed (Hochstrasser et al. 2014a; Hrle et al. 2014; Jackson et al. 2014a; Mulepati et al. 2014a). The interaction between PAM and Cse1 and Cse1 and Cas7 possibly nucleates R-loop formation of DNA through crRNA strand invasion and subsequent base-pairing (Figure 1.6). A recent study of Type I-E effector system provides evidence of directional R-loop formation starting from PAM and expanding towards PAM-distal protospacer end (Szczelkun et al. 2014). Large subunit

proteins (Cas8 superfamily proteins) are expected to be PAM interacting candidates in other Type I systems.

Cas3 superfamily proteins. Cas3 superfamily members that contain fused histidine, aspartate (HD) nuclease and Superfamily II (SF2) helicase domains (Haft et al. 2005a) are the signature proteins of Type I systems (Makarova et al. 2011a). Cas3 homologs exhibit metal-dependent, single-stranded DNA endonuclease and 3' to 5' exonuclease activity catalyzed by conserved adjacent histidine, aspartate (HD) residues of nuclease domains and ATP-dependent helicase activity *in vitro* (Beloglazova et al. 2011; Mulepati and Bailey 2011; Sinkunas et al. 2011). The metal-dependent nuclease combined with ATP-dependent helicase activities of Cas3 enzymes mediate complete degradation of target DNA duplexes and plasmids in *in-vitro* assembled Type I-A (Plagens et al. 2014b) and Type I-E (Westra et al. 2012c; Mulepati and Bailey 2013b) crRNPs. Structures of Type I-E Cas3 proteins from *Thermobifida fusca* (Huo et al. 2014a) and *Thermobaculum terrenum* (Gong et al. 2014) reveal a multi-domain architecture. In addition to N-terminal HD nuclease domain, the SF2 helicase contains two RecA-like domains (RecA1 and RecA2) and two Cas3-specific domains (linker helix domain and C-terminal domain (CTD)). The active site residues (HD) of nuclease domain are coordinated transition metals. Apart from the catalytic HD (His, Asp) residues of nuclease domain and DE (Asp, Glu) residues of DExH helicase domain, mutation of other conserved residues in the RecA1 and RecA2 domains and other SF2 helicase motifs (Ia, Ib, Id, III, IV, IVa, V) and Cas3-specific motif I-C (that coordinates nuclease and helicase activities) significantly reduced invader silencing by Type I-E system *in vivo*.

Cas3 proteins in Type I systems (except Type I-A) are not structural components of DNA surveillance crRNPs. After PAM scanning and interaction large-subunit protein (Cse1) of Type I-E crRNPs recruits Cas3 to PAM-adjacent protospacer sequences for cleavage (Hochstrasser et al. 2014a; Huo et al. 2014a) (Figure 1.6). It has been observed that PAM recognizing loop L1 of Cse1 is required for Cas3 nuclease activity in Type I-E crRNP-mediated invader DNA cleavage (Hochstrasser et al. 2014a). A unique feature of Type I-A systems is that nuclease and helicase domains of Cas3 reside in separate proteins Cas3" and Cas3' respectively (teal and aqua, Figure 1.2) (Haft, Makarova). Cas3" and Cas3' are seen as structural components of recently characterized *T. tenax* Type I-A crRNP (Plagens et al. 2014b). Despite these subtle distinctions we expect effector crRNPs of different Type I systems employ similar molecular mechanisms of invader silencing.

Invader silencing in Type III systems

Structure of effector complexes

Effector complexes of Type III systems (Type III-A or Csm and Type III-B or Cmr) bear some structural similarities with Type I crRNPs (Figure 1.7) but exhibit distinct mechanism of invader silencing. High resolution structures of Type III-A crRNP from *S. solfataricus* (Rouillon et al. 2013) and two Type III-B crRNPs from *T. thermophilus* (Staals et al. 2013a) and *P. furiosus* (Spilman et al. 2013a) are elucidated. The crRNP structures show multiple units of Cas7 superfamily proteins (Csm3 (III-A) and Cmr4 (Type III-B)) bind along the length of crRNA guide. Like Cas5 proteins in Type I systems, Csm4 (Type III-A) and Cmr3 (Type III-B) bind 5'-tag of crRNA. Large subunit

signature protein members of Cas10 (Csm1 and Cmr2) are found associated with Cas5-superfamily proteins (Csm4 and Cmr3 respectively) at 5'tag region. Multiple units of small subunit proteins (Csm2 (Type III-A) and Cmr5 (Type III-B)) interact with the Cas7-like units (Csm3 and Cmr4, except the terminal units). Small-subunit protein (Cmr5) contains a α-helical domain homologous to N-terminal domain of small-subunit protein (Cse2) in Type I-E system (Reeks et al. 2013a). In accordance with the observed interaction in structure of Cmr complex, small-subunit protein (Cmr5) interacts with Cas7 family protein (Cmr4) *in vitro* (Park et al. 2013).

Mechanism of invader silencing

Invader silencing mechanisms in Type III systems are not only different from Type I systems but exhibit distinct properties between Type III-A and III-B systems. One similarity is both Type III-A and Type III-B crRNP mediate PAM-independent invader nucleic acid targeting (Hale et al. 2009; Marraffini and Sontheimer 2010b). Type III-A (Csm) crRNP mediates invader DNA silencing, although the mechanistic details are yet to be elucidated (Marraffini and Sontheimer 2008; Hatoum-Aslan et al. 2014). In Type III-A system homology of protospacer-adjacent sequences with 5'-tag of crRNA determines of self vs. non-self discrimination (Marraffini and Sontheimer 2010b).

Type III-B (Cmr) crRNP mediates homology-dependent target RNA cleavage *in vitro* (Hale et al. 2009; Hale et al. 2012a; Hale et al. 2014) and *in vivo* (Zebec et al. 2014a). Recent evidence indicates Cas7 superfamily protein (Cmr4) is the effector nuclease that cleaves target RNA ((Benda et al. 2014) Ramia et al., Cell Reports, accepted) An interesting twist to this theme is provided by one of the Type III-B crRNPs in *S*.

islandicus. This Cmr crRNP mediates PAM-independent invader DNA silencing (Deng et al. 2013a). The Cmr crRNP mediated DNA silencing activity requires an additional protein that belongs to Csx1 superfamily (Haft et al. 2005a) and a transcript that is complementary to crRNA (Deng et al. 2013a). The effector nuclease of DNA silencing is unknown but based on *in-silico* predictions HD-nuclease containing large subunit protein (Cmr2 or Cas10) is predicted to cleave DNA (Makarova et al. 2011c). These studies provide evidence of Cmr crRNP/s mediated RNA and/ or DNA targeting via distinct mechanisms and effector nucleases.

CRISPR-Cas systems in *Pyrococcus furiosus* and overview of dissertation

Pyrococcus furiosus is a hyperthermophilic euryarchaeon with an optimal growth temperature of 100° C. Notable work by fellow lab members, Dr. C. Hale and Dr. J. Carte laid the foundation of CRISPR-Cas research in *P. furiosus* leading to this work. *P. furiosus* contains seven CRISPR loci encoding ~200 crRNAs (Hale et al. 2008) and ~twenty-nine Cas proteins (encoded from two separate gene clusters) with predicted roles in adaptation, biogenesis and immune-effector systems (Types I-A (Csa), I-G (Cst) and III-B (Cmr)) (Terns and Terns 2013). Cas6 is the primary RNA endonuclease that cleaves 22-nt into each repeat of all CRISPR loci to generate 8-nt 5'-tag on crRNA intermediates (Carte et al. 2008b; Carte et al. 2010). Shorter crRNAs with trimmed 3' ends are processed by unknown mechanisms and constitute stable, mature forms (Hale et al. 2008). Biochemical fractionations of *P. furiosus* cell extract by chromatography, revealed multiple crRNPs containing distinct crRNA composition (Hale et al. 2009). Compositional characterization of one of the fractionated crRNPs revealed Cmr 1-6 as

Cas protein composition and two distinct mature crRNA size forms (45 nt, 39 nt), containing Cas6 generated 5'-tag followed by variable amount of guide sequences. Functional characterization demonstrated Cmr crRNP uses either crRNA size form to guide homology-dependent RNA target cleavage *in vitro* (Hale et al. 2009; Hale et al. 2012a; Hale et al. 2014). This work aims to investigate the composition of other crRNP/s detected in fractionated *P. furiosus* cell extract and understand their role in CRISPR-Cas pathway. Due to stable association of crRNAs we hypothesized the other crRNPs constitute Type I-A and I-G immune-effector systems and likely target invader DNA.

We used immunoprecipitation approach to isolate and purify Csa, Cst and Cmr containing, native complexes (Chapter 2, Majumdar et al., unpublished). Through tandem mass spectrometry analyses we identified three complexes with distinct protein compositions that comprised Csa (I-A), Cst (I-G) and Cmr (III-B) proteins. Northern blot analyses of immunopurified Csa and Cst complexes showed similar crRNA profiles as those associated with the uncharacterized crRNPs fractionated by chromatography (Hale et al. 2009). Through RNA sequencing we determined the molecular identities of crRNAs associated with all three complexes. We observe that crRNAs expressed from all seven CRISPR loci are associated with all three complexes. Immunopurified Cmr crRNP showed the expected crRNA profile and homology-dependent cleavage of RNA targets. We unraveled further details of crRNA features essential for crRNP assembly and provided *in vivo* evidence of Cmr crRNP mediated endogenous target RNA cleavage. (Appendix, (Hale et al. 2012a)). A structure of *P. furiosus* Cmr crRNP is elucidated by electron microscopy (EM) ((Spilman et al. 2013a), Ramia et al., Cell Reports, accepted).

We investigated the role of Csa (I-A) crRNP (Chapter 3, Majumdar et al., unpublished) and Cst (I-G) crRNP (Chapter 4, Majumdar et al., unpublished) by *in vitro* reconstitution of complexes with a synthetic crRNA that was detected in native complexes and recombinant Csa or Cst proteins respectively. Our results show both Types I-A and I-G crRNPs cleave target DNA duplexes in crRNA-guided and PAM-dependent fashion with few distinct properties. Further biochemical characterizations delineate the essential components of active Type I-A and I-G mediated DNA recognition and targeting. We have successfully determined that the Cas3 nucleases (Cas3" for I-A and Cas3 for I-G) associated with each system are the effector nucleases by loss of function caused by mutation of the conserved histidine, aspartate (HD) residues. We observe the Cas3 nuclease (Cas3") and helicase (Cas3') proteins are structural components of active Type I-A crRNP. Finally, by native gel electrophoresis and UV induced RNA-protein crosslinking studies we delineate the structural cores and attempt to understand the organization of Csa (I-A) and Cst (I-G) crRNPs.

In summary, this work shows *P. furiosus* CRISPR-Cas system contains three compositionally distinct immune-effector complexes: Types I-A (Csa), I-G (Cst) and III-B (Cmr). Our data provides evidence that in addition to RNA-targeting Type III-B (Cmr) system *P. furiosus* contains two independent and co-existing DNA targeting systems (Types I-A and I-G). Investigations of *P. furiosus* CRISPR-Cas pathway has provided key insights about co-existing, multiple effector systems that can use common crRNA species encoded from all CRISPR loci yet feature distinct silencing mechanisms.

References

Agari, Y., Yokoyama, S., Kuramitsu, S., and Shinkai, A. (2008). X-ray crystal structure of a CRISPR-associated protein, Cse2, from *Thermus thermophilus* HB8. Proteins *73*, 1063-1067.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature.

Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, U. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. Nucleic acids Research *42*, 7884-7893.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., *et al*. (2011). A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Molecular Microbiology *79*, 484-502.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., *et al*. (2008). A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. The Journal of Biological Chemistry *283*, 20361-20371.

Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., and Yakunin, A.F. (2011). Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. The EMBO Journal *30*, 4616-4627.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551-2561.

Brendel, J., Stoll, B., Lange, S.J., Sharma, K., Lenz, C., Stachler, A.E., Maier, L.K., Richter, H., Nickel, L., Schmitz, R.A., *et al*. (2014). A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in *Haloferax volcanii*. The Journal of Biological Chemistry *289*, 7164-7177.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

Cady, K.C., and O'Toole, G.A. (2011). Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. Journal of Bacteriology *193*, 3433-3445.

Carte, J., Christopher, R.T., Smith, J.T., Olson, S., Barrangou, R., Moineau, S., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M., and Terns, M.P. (2014). The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in *Streptococcus thermophilus*. Molecular Microbiology *93*, 98-112.

Carte, J., Pfister, N.T., Compton, M.M., Terns, R.M., and Terns, M.P. (2010). Binding and cleavage of CRISPR RNA by Cas6. RNA *16*, 2181-2188.

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

Chen, H., Choi, J., and Bailey, S. (2014). Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. The Journal of Biological Chemistry *289*, 13284-13294.

Chopin, M.C., Chopin, A., and Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. Current Opinion in Microbiology *8*, 473-479.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nature Communications *3*, 945.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Deng, L., Garrett, R.A., Shah, S.A., Peng, X., and She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. Molecular Microbiology *87*, 1088-1099.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. Journal of Bacteriology *190*, 1390-1400.

Ebihara, A., Yao, M., Masui, R., Tanaka, I., Yokoyama, S., and Kuramitsu, S. (2006). Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8

reveals a new protein family with an RNA recognition motif-like domain. Protein science : a publication of the Protein Society *15*, 1494-1499.

Ellinger, P., Arslan, Z., Wurm, R., Tschapek, B., MacKenzie, C., Pfeffer, K., Panjikar, S., Wagner, R., Schmitt, L., Gohlke, H., *et al*. (2012). The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*. Journal of Structural Biology *178*, 350-362.

Elmore, J.R., Yokooji, Y., Sato, T., Olson, S., Glover, C.V., 3rd, Graveley, B.R., Atomi, H., Terns, R.M., and Terns, M.P. (2013). Programmable plasmid interference by the CRISPR-Cas system in *Thermococcus kodakarensis*. RNA Biology *10*, 828-840.

Erdmann, S., and Garrett, R.A. (2012). Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Molecular Microbiology *85*, 1044-1056.

Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H., and Brouns, S.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proceedings of the National Academy of Sciences of the United States of America *111*, E1629-1638.

Fischer, S., Maier, L.K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., Dyall-Smith, M., and Marchfelder, A. (2012). An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. The Journal of Biological Chemistry *287*, 33351-33363.

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

Garside, E.L., Schellenberg, M.J., Gesner, E.M., Bonanno, J.B., Sauder, J.M., Burley, S.K., Almo, S.C., Mehta, G., and MacMillan, A.M. (2012). Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. RNA *18*, 2020-2028.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences of the United States of America *109*, E2579-2586.

Gesner, E.M., Schellenberg, M.J., Garside, E.L., George, M.M., and Macmillan, A.M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. Nature Structural & Molecular Biology *18*, 688-692.

Godde, J.S., and Bickerton, A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. Journal of molecular evolution *62*, 718-729.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics *8*, 172.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Computational Biology *1*, e60.

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

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M., *et al*. (2012). Essential features and

rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Molecular Cell *45*, 292-302.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945-956.

Han, D., Lehmann, K., and Krauss, G. (2009). SSO1450--a CAS1 protein from *Sulfolobus solfataricus* P2 with high affinity for RNA and DNA. FEBS Letters *583*, 1928-1932.

Hatoum-Aslan, A., Maniv, I., and Marraffini, L.A. (2011). Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. Proceedings of the National Academy of Sciences of the United States of America *108*, 21218-21222.

Hatoum-Aslan, A., Maniv, I., Samai, P., and Marraffini, L.A. (2014). Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. Journal of Bacteriology *196*, 310-317.

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

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

Haurwitz, R.E., Sternberg, S.H., and Doudna, J.A. (2012). Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. The EMBO Journal *31*, 2824-2832.

Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proceedings of the National Academy of Sciences of the United States of America *111*, 6618-6623.

Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. Journal of Bacteriology *190*, 1401-1412.

Hrle, A., Maier, L.K., Sharma, K., Ebert, J., Basquin, C., Urlaub, H., Marchfelder, A., and Conti, E. (2014). Structural analyses of the CRISPR protein Csc2 reveal the RNA-binding interface of the type I-D Cas7 family. RNA Biology *11*.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell *157*, 1262-1278.

Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M.D., Jr., Zhou, S., Rajashankar, K., Kurinov, I., *et al*. (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nature Structural & Molecular Biology.

Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science.

Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Molecular Microbiology *43*, 1565-1575.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., *et al*. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science *343*, 1247997.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., *et al*. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nature Structural & Molecular biology *18*, 529-536.

Kim, T.Y., Shin, M., Huynh Thi Yen, L., and Kim, J.S. (2013). Crystal structure of Cas1 from Archaeoglobus fulgidus and characterization of its nucleolytic activity. Biochemical and Biophysical Research Communications *441*, 720-725.

Koo, Y., Jung, D.K., and Bae, E. (2012). Crystal structure of *Streptococcus pyogenes* Csn2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure. PloS One *7*, e33401.

Koo, Y., Ka, D., Kim, E.J., Suh, N., and Bae, E. (2013). Conservation and variability in the structure and function of the Cas5d endoribonuclease in the CRISPR-mediated microbial immune system. Journal of Molecular Biology *425*, 3799-3810.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biology *8*, R61.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nature Reviews Microbiology *8*, 317-327.

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

Lemak, S., Beloglazova, N., Nocek, B., Skarina, T., Flick, R., Brown, G., Popovic, A., Joachimiak, A., Savchenko, A., and Yakunin, A.F. (2013). Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S] cluster containing Cas4 nuclease SSO0001 from *Sulfolobus solfataricus*. Journal of the American Chemical Society *135*, 17476-17487.

Levin, B.R., Moineau, S., Bushman, M., and Barrangou, R. (2013). The population and evolutionary dynamics of phage and bacteria with CRISPR-mediated immunity. PLoS Genetics *9*, e1003312.

Li, M., Wang, R., Zhao, D., and Xiang, H. (2014). Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. Nucleic acids Research *42*, 2483-2492.

Lillestol, R.K., Redder, P., Garrett, R.A., and Brugger, K. (2006). A putative viral defence mechanism in archaeal cells. Archaea *2*, 59-72.

Lillestol, R.K., Shah, S.A., Brugger, K., Redder, P., Phan, H., Christiansen, J., and Garrett, R.A. (2009). CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. Molecular Microbiology *72*, 259-272.

Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q., Copie, V., *et al*. (2011). Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-

associated complex for antiviral defense (CASCADE). The Journal of Biological Chemistry *286*, 21643-21656.

Magadan, A.H., Dupuis, M.E., Villion, M., and Moineau, S. (2012). Cleavage of phage DNA by the *Streptococcus thermophilus* CRISPR3-Cas system. PloS One *7*, e40913.

Maier, L.K., Lange, S.J., Stoll, B., Haas, K.A., Fischer, S., Fischer, E., Duchardt-Ferner, E., Wohnert, J., Backofen, R., and Marchfelder, A. (2013). Essential requirements for the detection and degradation of invaders by the *Haloferax volcanii* CRISPR/Cas system I-B. RNA Biology *10*, 865-874.

Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E.V. (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biology Direct *6*, 38.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biology Direct *1*, 7.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., *et al*. (2011b). Evolution and classification of the CRISPR-Cas systems. Nature reviews Microbiology *9*, 467-477.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843-1845.

Marraffini, L.A., and Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568-571.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733-740.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of Molecular Evolution *60*, 174-182.

Mulepati, S., and Bailey, S. (2011). Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). The Journal of Biological Chemistry *286*, 31896-31903.

Mulepati, S., and Bailey, S. (2013). In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. The Journal of Biological Chemistry *288*, 22184-22192.

Mulepati, S., Heroux, A., and Bailey, S. (2014). Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science.

Mulepati, S., Orr, A., and Bailey, S. (2012). Crystal structure of the largest subunit of a bacterial RNA-guided immune complex and its role in DNA target binding. The Journal of Biological Chemistry *287*, 22445-22449.

Nam, K.H., Ding, F., Haitjema, C., Huang, Q., DeLisa, M.P., and Ke, A. (2012a). Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. The Journal of Biological Chemistry *287*, 35943-35952.

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

Nam, K.H., Huang, Q., and Ke, A. (2012c). Nucleic acid binding surface and dimer interface revealed by CRISPR-associated CasB protein structures. FEBS Letters *586*, 3956-3961.

Nam, K.H., Kurinov, I., and Ke, A. (2011). Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA binding activity. The Journal of Biological Chemistry *286*, 30759-30768.

Nickel, L., Weidenbach, K., Jager, D., Backofen, R., Lange, S.J., Heidrich, N., and Schmitz, R.A. (2013). Two CRISPR-Cas systems in *Methanosarcina mazei* strain Go1 display common processing features despite belonging to different types I and III. RNA Biology *10*, 779-791.

Niewoehner, O., Jinek, M., and Doudna, J.A. (2014). Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. Nucleic acids Research *42*, 1341-1353.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935-949.

Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. Nature Structural & Molecular Biology *21*, 528-534.

Park, J.H., Sun, J., Park, S.Y., Hwang, H.J., Park, M.Y., Shin, M., and Kim, J.S. (2013). Crystal structure of Cmr5 from *Pyrococcus furiosus* and its functional implications. FEBS Letters *587*, 562-568.

Peng, W., Li, H., Hallstrom, S., Peng, N., Liang, Y.X., and She, Q. (2013). Genetic determinants of PAM-dependent DNA targeting and pre-crRNA processing in *Sulfolobus islandicus*. RNA Biology *10*, 738-748.

Plagens, A., Tjaden, B., Hagemann, A., Randau, L., and Hensel, R. (2012). Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. Journal of Bacteriology *194*, 2491-2500.

Plagens, A., Tripp, V., Daume, M., Sharma, K., Klingl, A., Hrle, A., Conti, E., Urlaub, H., and Randau, L. (2014). In vitro assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. Nucleic acids Research *42*, 5125-5138.

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

Przybilski, R., Richter, C., Gristwood, T., Clulow, J.S., Vercoe, R.B., and Fineran, P.C. (2011). Csy4 is responsible for CRISPR RNA processing in *Pectobacterium atrosepticum*. RNA Biology *8*, 517-528.

Reeks, J., Graham, S., Anderson, L., Liu, H., White, M.F., and Naismith, J.H. (2013a). Structure of the archaeal Cascade subunit Csa5: relating the small subunits of CRISPR effector complexes. RNA Biology *10*, 762-769.

Reeks, J., Sokolowski, R.D., Graham, S., Liu, H., Naismith, J.H., and White, M.F. (2013b). Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing. The Biochemical Journal *452*, 223-230.

Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H., and Fineran, P.C. (2014). Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic acids Research.

Richter, H., Zoephel, J., Schermuly, J., Maticzka, D., Backofen, R., and Randau, L. (2012). Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. Nucleic acids Research *40*, 9887-9896.

Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C.V., Spagnolo, L., and White, M.F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. Molecular Cell *52*, 124-134.

Sashital, D.G., Jinek, M., and Doudna, J.A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nature Structural & Molecular Biology *18*, 680-687.

Sashital, D.G., Wiedenheft, B., and Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. Molecular Cell *46*, 606-615.

Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., and Severinov, K. (2013). High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. RNA Biology *10*, 716-725.

Scholz, I., Lange, S.J., Hein, S., Hess, W.R., and Backofen, R. (2013). CRISPR-Cas systems in the cyanobacterium Synechocystis sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. PloS One *8*, e56470.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proceedings of the National Academy of Sciences of the United States of America *108*, 10098-10103.

Shah, S.A., Erdmann, S., Mojica, F.J., and Garrett, R.A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. RNA Biology *10*, 891-899.

Shao, Y., and Li, H. (2013). Recognition and cleavage of a nonstructured CRISPR RNA by its processing endoribonuclease Cas6. Structure *21*, 385-393.

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

Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., and Stagg, S. (2013). Structure of an RNA silencing complex of the CRISPR-Cas immune system. Molecular Cell *52*, 146-152.

Staals, R.H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K.*, et al*. (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. Molecular Cell *52*, 135-145.

Stern, A., and Sorek, R. (2011). The phage-host arms race: shaping the evolution of microbes. BioEssays : news and reviews in molecular, cellular and developmental biology *33*, 43-51.

Sternberg, S.H., Haurwitz, R.E., and Doudna, J.A. (2012). Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. RNA *18*, 661-672.

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

Swarts, D.C., Mosterd, C., van Passel, M.W., and Brouns, S.J. (2012). CRISPR interference directs strand specific spacer acquisition. PLoS One *7*, e35888.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proceedings of the National Academy of Sciences of the United States of America *111*, 9798-9803.

Tang, T.H., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., Bachellerie, J.P., and Huttenhofer, A. (2005). Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. Molecular Microbiology *55*, 469-481.

Terns, R.M., and Terns, M.P. (2014). CRISPR-based technologies: prokaryotic defense weapons repurposed. Trends in Genetics : TIG *30*, 111-118.

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

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nature Reviews Microbiology *12*, 479-492.

van Duijn, E., Barbu, I.M., Barendregt, A., Jore, M.M., Wiedenheft, B., Lundgren, M., Westra, E.R., Brouns, S.J., Doudna, J.A., van der Oost, J., *et al*. (2012). 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*. Molecular & Cellular Proteomics : MCP *11*, 1430-1441.

Vestergaard, G., Garrett, R.A., and Shah, S.A. (2014). CRISPR adaptive immune systems of Archaea. RNA Biology *11*, 156-167.

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

Wang, R., Zheng, H., Preamplume, G., Shao, Y., and Li, H. (2012). The impact of CRISPR repeat sequence on structures of a Cas6 protein-RNA complex. Protein science : a publication of the Protein Society *21*, 405-417.

Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K., and Brouns, S.J. (2013). Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. PLoS Genetics *9*, e1003742.

Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., *et al*. (2012). CRISPR immunity relies on

the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Molecular Cell *46*, 595-605.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011a). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature *477*, 486-489.

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

Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J., Dickman, M.J., *et al*. (2011b). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proceedings of the National Academy of Sciences of the United States of America *108*, 10092-10097.

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W., and Doudna, J.A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure *17*, 904-912.

Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic acids Research *40*, 5569-5576.

Zebec, Z., Manica, A., Zhang, J., White, M.F., and Schleper, C. (2014). CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*. Nucleic acids Research *42*, 5280-5288.

Zhang, J., Kasciukovic, T., and White, M.F. (2012a). The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. PloS One *7*, e47232.

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.V., *et al*. (2012b). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Molecular Cell *45*, 303-313.

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. Nature.

Figure 1.1. **Mechanism of CRISPR-Cas immune system**. Upon invasion, fragments of invader genome called protospacer (red), adjacent to a signature recognition element called PAM (white) are processed and integrated in a host's CRISPR locus downstream of a leader (grey) in the adaptation stage. A host's CRISPR locus contains multiple units of direct repeats (black boxes) flanking variant invader-derived sequences (multiple colored boxes), acquired via adaptation. During biogenesis stage precursor CRISPR transcripts are encoded and processed to form mature CRISPR RNAs (crRNAs), each containing guide sequences (multiple colors) flanked by repeat-derived 'tags' (black). In invader silencing stage, each mature crRNA assembles into Cas-effector crRNPs to guide recognition and targeting of homologous invader sequences (red). Figure adapted from (Terns and Terns 2014)

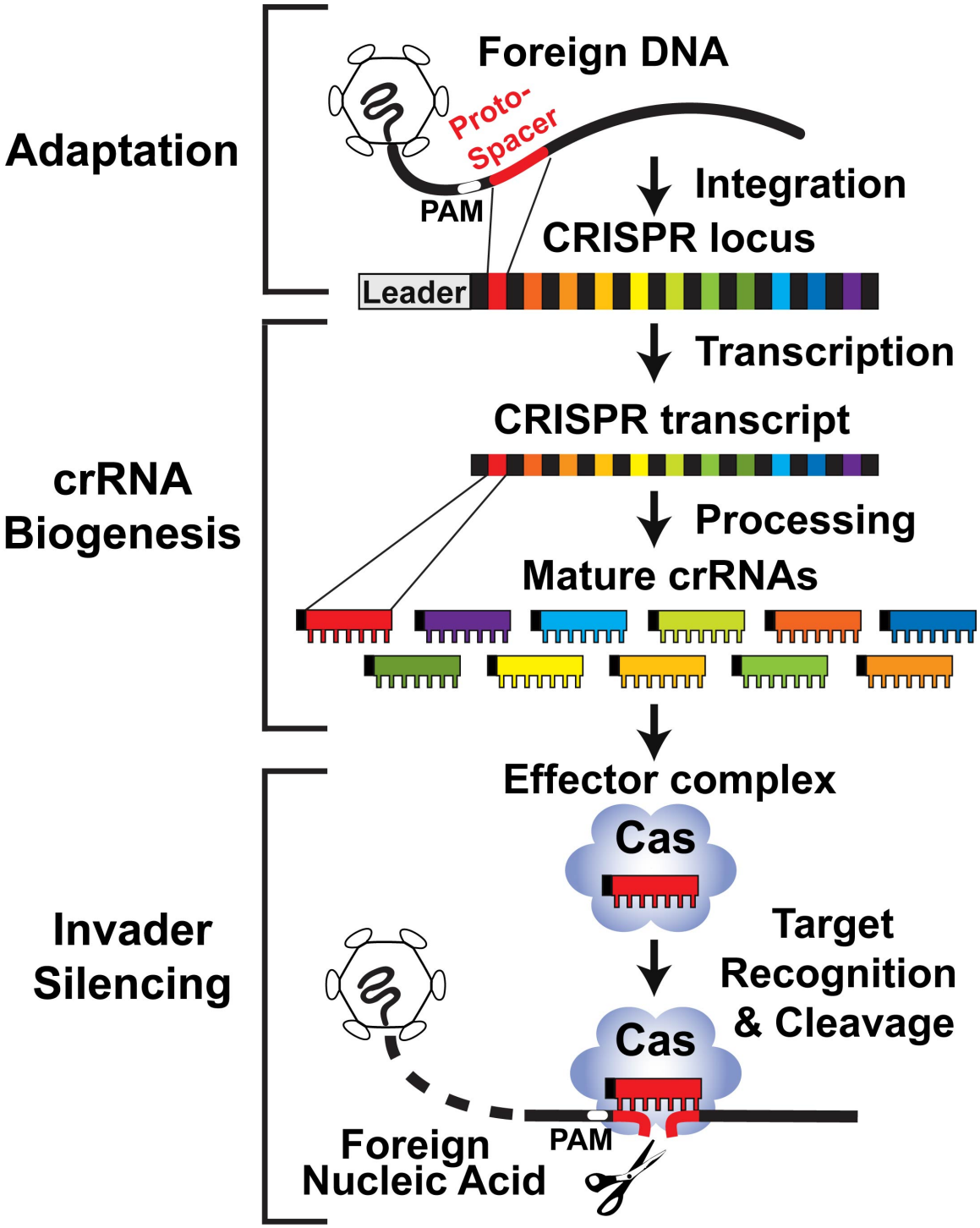

Adaptation
Foreign DNA
Proto-
Spacer
PAM
Integration
CRISPR locus
Leader
Transcription
crRNA
Biogenesis
CRISPR transcript
Processing
Mature crRNAs
Effector complex
Cas
Invader
Silencing
Target
Recognition
& Cleavage
Cas
Foreign
Nucleic Acid
PAM

Figure 1.2. *cas* gene content and distribution across CRISPR-Cas systems. CRISPR-Cas immune systems are categorized into three major types: Types I, II and III. Each system (e.g. I-A, (Makarova et al. 2011c)) and its corresponding Cas subtype (e.g. Csa, (Haft et al. 2005a)) are specified on the left side. Names of *cas* gene is labeled below each box in italics. Genes are color-coded based either on superfamily and/or functional analogy. Genes encoding characterized or predicted adaptation proteins are shown grey, biogenesis proteins of Cas6 superfamily or Cas5d are shown blue and effector nuclease proteins of Type I (Cas3) and Type II (Cas9) systems are shaded teal. Genes encoding Cas superfamilies, their distant relatives or structural analogues in effector complexes are shown: Cas5 (yellow); Cas7, Csy3, Csm3, Cmr4 (green); large-subunit protein superfamilies Cas8 and Cas10 (purple); small-subunit proteins (orange), structural analogues of Type III systems- Csm5 Cmr1, Cmr6 (berry) and CARF-domain containing Csm6 (white). Combinations of colors are used to depict fusion proteins (Cas2-3 in I-F) or functional variant of superfamily (Cas5d). Figure adapted and modified from (Makarova et al. 2011c).

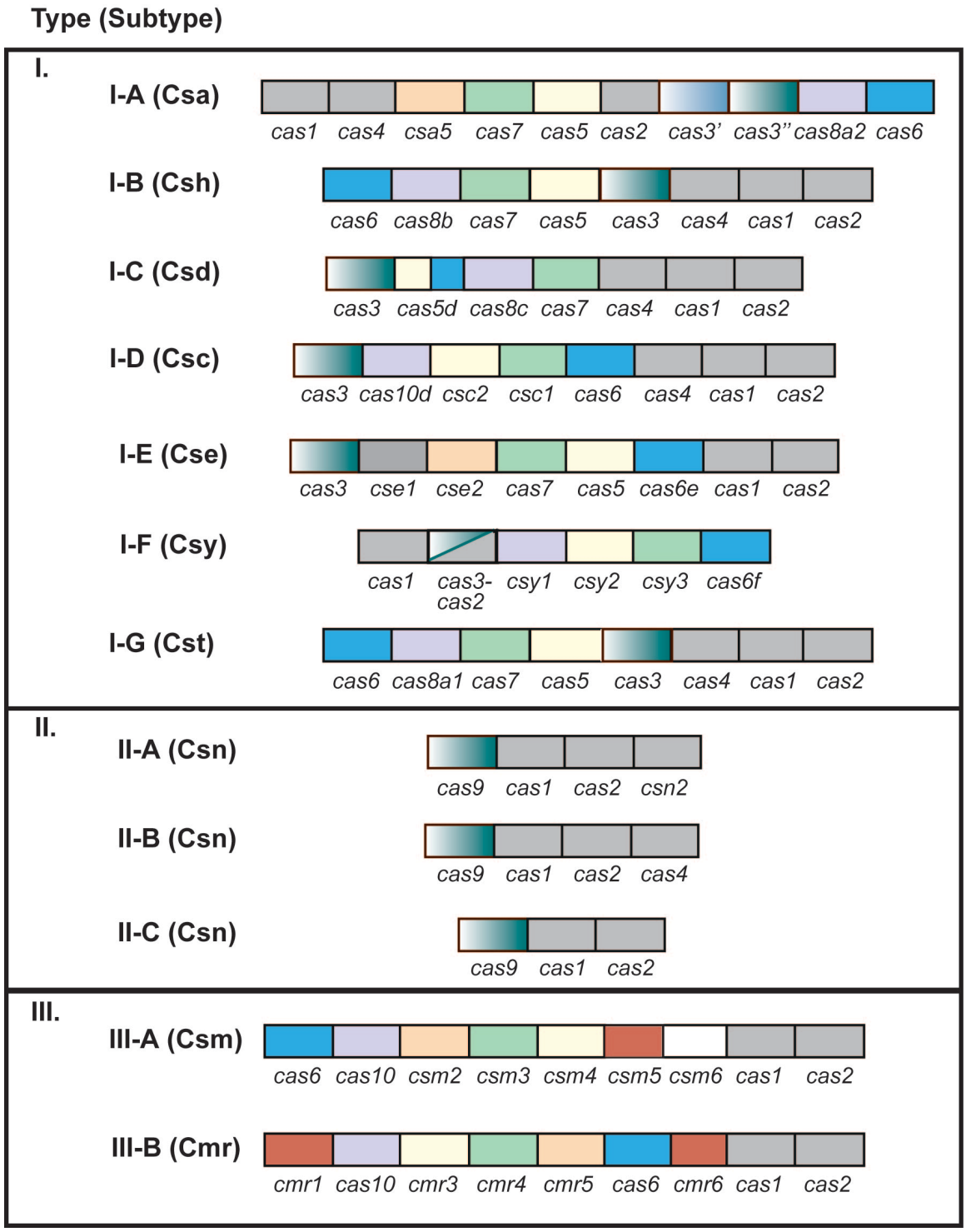

Type (Subtype)
I.
I-A (Csa)
cas1 cas4 csa5 cas7 cas5 cas2 cas3' cas3'' cas8a2 cas6
I-B (Csh)
cas6 cas8b cas7 cas5 cas3 cas4 cas1 cas2
I-C (Csd)
cas3 cas5d cas8c cas7 cas4 cas1 cas2
I-D (Csc)
cas3 cas10d csc2 csc1 cas6 cas4 cas1 cas2
I-E (Cse)
cas3 cse1 cse2 cas7 cas5 cas6e cas1 cas2
I-F (Csy)
cas1 cas3-cas2 csy1 csy2 csy3 cas6f
I-G (Cst)
cas6 cas8a1 cas7 cas5 cas3 cas4 cas1 cas2
II.
II-A (Csn)
cas9 cas1 cas2 csn2
II-B (Csn)
cas9 cas1 cas2 cas4
II-C (Csn)
cas9 cas1 cas2
III.
III-A (Csm)
cas6 cas10 csm2 csm3 csm4 csm5 csm6 cas1 cas2
III-B (Cmr)
cmr1 cas10 cmr3 cmr4 cmr5 cas6 cmr6 cas1 cas2

Figure 1.3 Overview of Cas proteins involved in mechanism of CRISPR-Cas immune systems. The three mechanistic stages are specified on the right. Cas1, Cas2, Cas4 and Csn2 (Type II) are characterized or predicted to be involved in adaptation (grey). Primary processing endonucleases (blue) of crRNA biogenesis across Types I and III systems are shown. Mature crRNAs contain repeat-derived tags (black) flanking invader-derived guide (red). crRNA biogenesis in Type II systems involving Cas9 (teal), RNase III (brown), tracrRNA (black, thin lines) are summarized. Effector crRNPs of Type I systems (green) target invader DNA, catalyzed by associated-Cas3s (teal). Type II crRNP-comprising Cas9 (teal) targets and cleaves invader DNA. Types I and II crRNPs rely on PAM for invader recognition. Cognate protospacer (red) and recognition element, PAM (cyan) of invader DNA is shown. Both Type III-A (Csm, orange) and III-B (Cmr, orange) crRNPs can silence invader DNA via PAM independent mechanisms. Type III-B crRNP (Cmr, blue) cleaves complementary target RNAs.

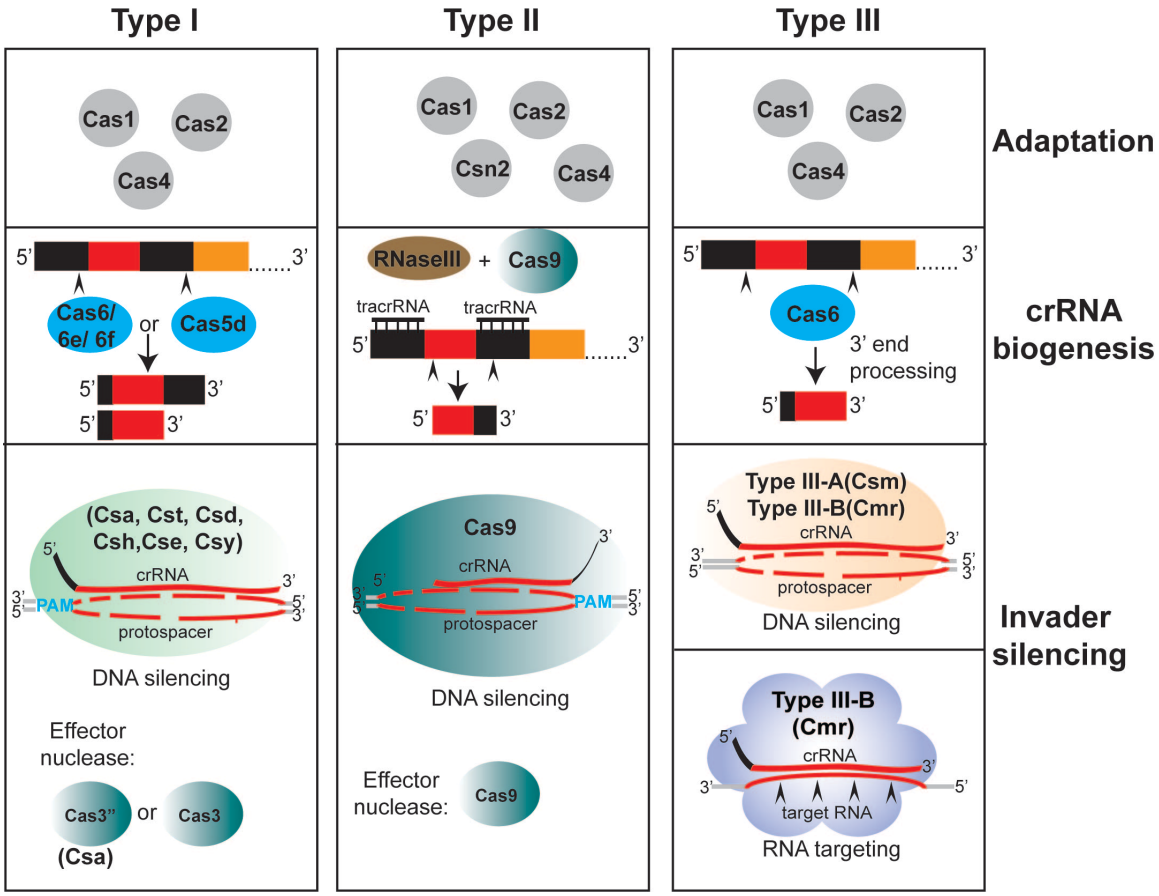

Type I
Type II
Type III
Adaptation
Cas1
Cas2
Cas4
Csn2
crRNA biogenesis
5'
3'
Cas6/ 6e/ 6f
or
Cas5d
RNaseIII
+
Cas9
tracrRNA
Cas6
3' end processing
Invader silencing
(Csa, Cst, Csd, Csh,Cse, Csy)
crRNA
PAM
protospacer
DNA silencing
Effector nuclease:
Cas3"
Cas3
(Csa)
Type III-A(Csm)
Type III-B(Cmr)
Type III-B (Cmr)
target RNA
RNA targeting

Figure 1.4. Mechanistic model of CRISPR-Cas mediated adaptation. Recognition of PAM (white) leads to protospacer (red, labeled) selection in invader genome and subsequent processing into invader-derived sequence. In a host CRISPR locus, leader proximal repeat (black box) is cleaved on both strands (shown by arrowheads), followed by resection to form single-stranded flanks. New invader-derived sequence (red) is integrated between flanks, followed by filling-in leading to duplication of repeat. Cas protein candidates, proposed to mediate combination of steps are shown (grey). Figure adapted from (van der Oost et al. 2014)

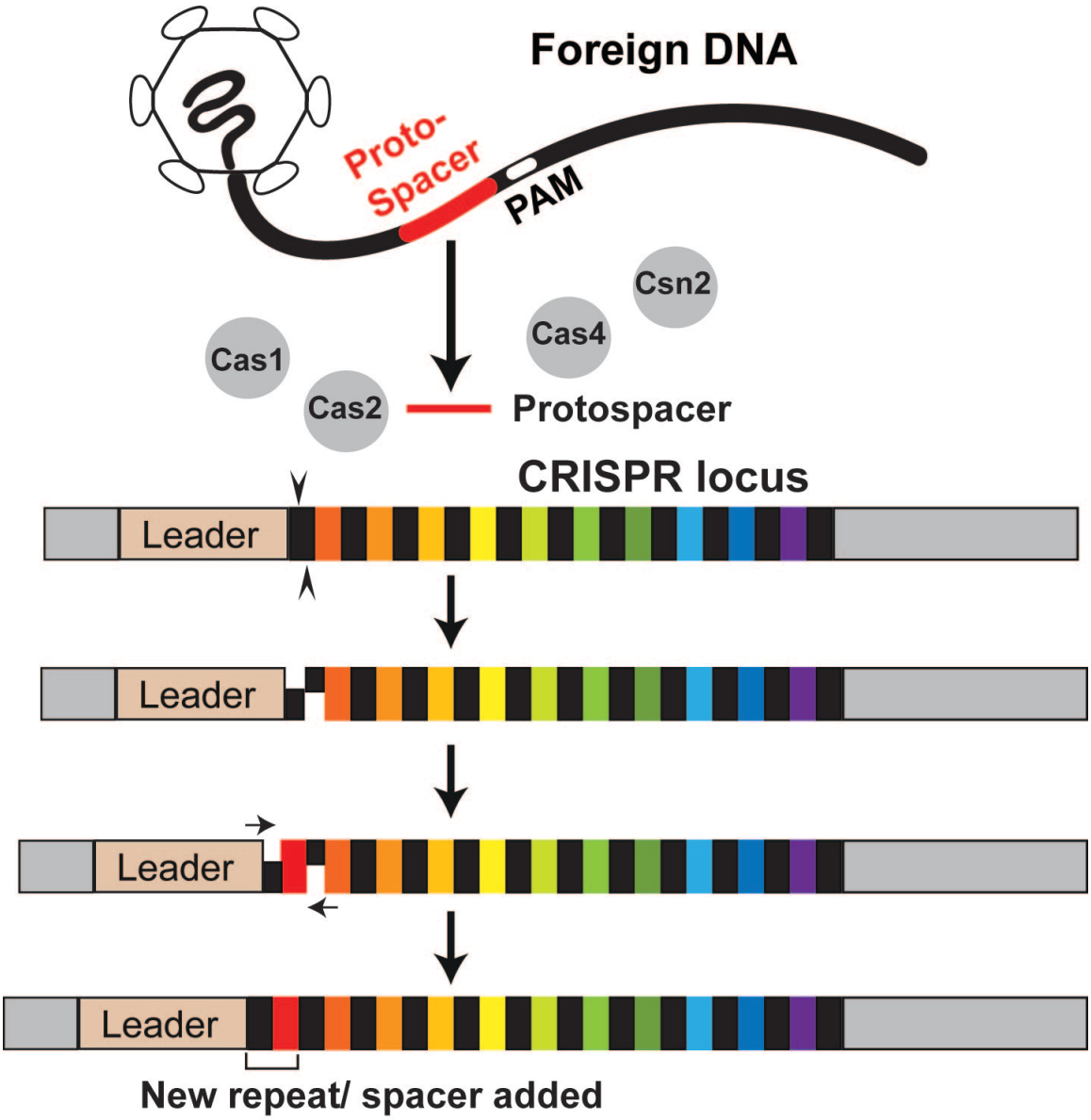

Foreign DNA
Proto-
Spacer
PAM
Csn2
Cas4
Cas1
Cas2
Protospacer
CRISPR locus
Leader
Leader
Leader
Leader
New repeat/ spacer added

Figure 1.5. crRNA biogenesis in Type I and III CRISPR-Cas systems. Long crRNA transcripts are expressed from promoter elements (brown) in leader region of CRISPR locus. Primary processing Cas endonucleases cleave in the repeat RNAs (shown by arrowheads) generating 1X intermediate form containing invader-derived, guide (red) flanked by repeat-derived, 5'-tag and 3'-tags (black). This form acts as mature crRNA in Type I-E, I-F systems and assembles into effector complexes. In other Type I systems and Type III systems further processing of 3' repeat RNA constitute mature crRNAs, either lacking or containing variable amounts of 3' repeats. Each of these crRNA size forms can assemble into effector crRNPs.

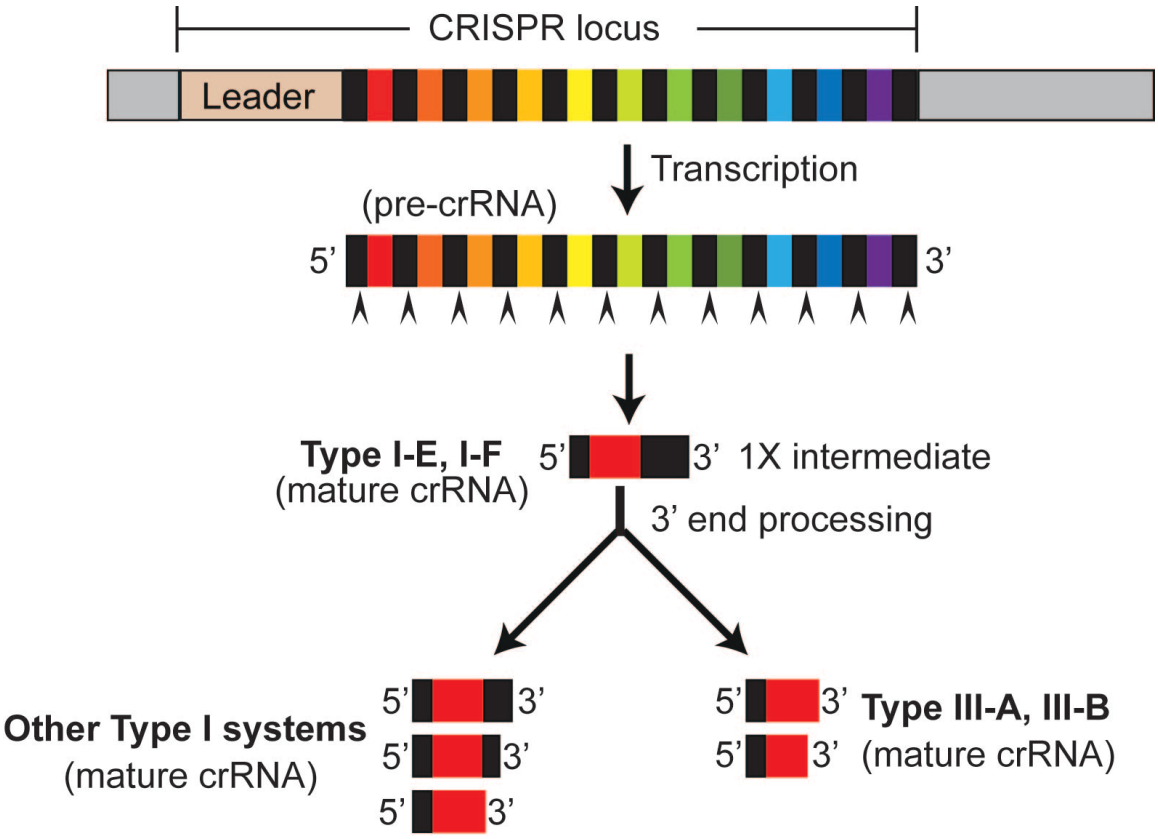

CRISPR locus
Leader
Transcription
(pre-crRNA)
5'
3'
Type I-E, I-F
(mature crRNA)
5'
3'
1X intermediate
3' end processing
Other Type I systems
(mature crRNA)
5'
3'
5'
3'
5'
3'
5'
3'
5'
3'
Type III-A, III-B
(mature crRNA)

Figure 1.6. Invader silencing in Type I CRISR-Cas systems. Invader silencing complexes (crRNPs) in Type I systems comprise subtype-specific proteins and crRNA. Structure of Type I-E (Cse) and I-F (Csy) surveillance crRNPs and their targeting mechanism are represented. The Cas superfamilies are color coded as Fig 1.2. Six units of either Cse4 (Cas7, I-E) or Csy3 (I-F) proteins (green) bind along crRNA guide (red), Cas5e (I-E) or Csy2 (I-F) proteins (yellow) binds 5'-tag (black) and Cas6 superfamily proteins (Cas6e, Cas6f; blue) binds 3'-tag (black hook) of crRNA. A dimer of small-subunit protein Cse2 (orange) interacts with the Cse4 units in Type I-E crRNP. Large subunit proteins (purple), Cse1 (I-E) and Csy1 (I-F) associate weakly with the surveillance complex, recognize and interact with PAM (cyan) in invader DNA. This nucleates crRNA strand invasion, base-pairing with target strand of protospacer (red) and formation of R-loop (shown as bubble). Large sub-unit proteins of crRNPs recruit Cas3 (nuclease-helicase protein, teal) to progressively unwind and cleave invader DNA.

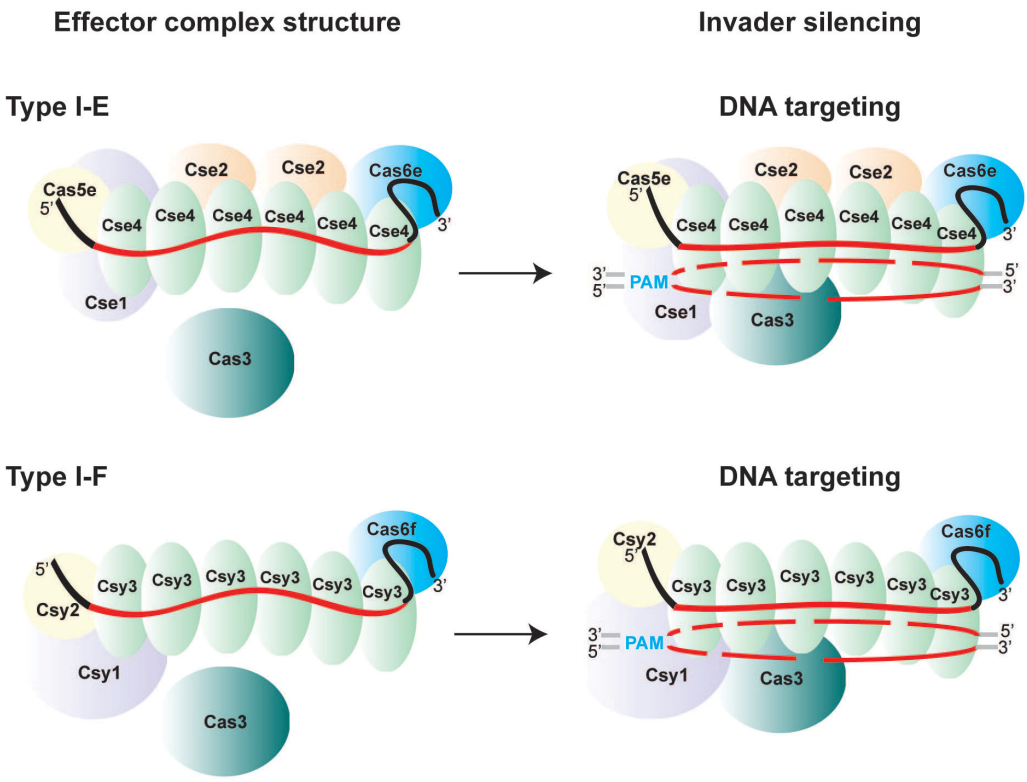
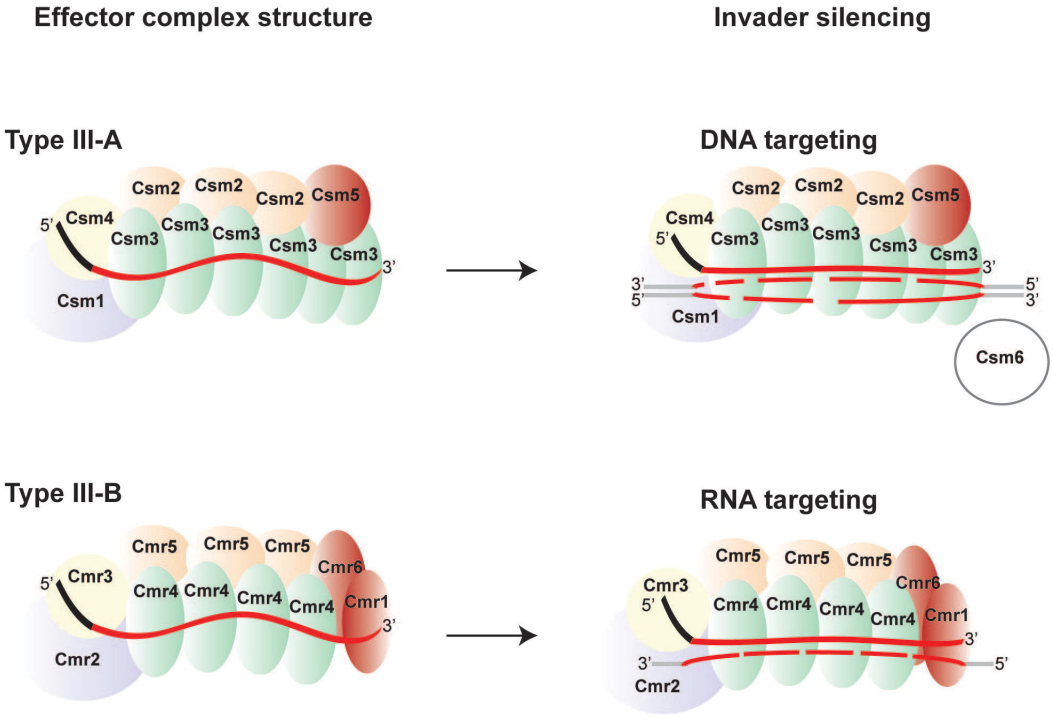

Effector complex structure
Invader silencing
Type I-E
DNA targeting
Cas5e
Cse2
Cse2
Cas6e
5'
Cse4
Cse4
Cse4
Cse4
Cse4
Cse4
3'
Cse1
Cas3
PAM
Type I-F
DNA targeting
Cas6f
Csy2
Csy3
Csy1

Figure 1.7 Invader silencing complexes of Type III CRISR-Cas systems. crRNAs bearing 5'-tag (black) and guide (red) associate with Csm (III-A), Cmr (III-B) sub-type proteins to form Type III effector complexes. The Cas superfamilies or structural analogs are color coded as Fig 1.2. Multiple units of Csm3 or Cmr4 proteins (green) bind along the length of crRNA guide. Csm4 or Cmr3 (yellow) associate with 5'tag of crRNA. Multimers of small subunit proteins (orange), Csm2 and Cmr5 interact with Csm3 and Cmr4 units respectively. Csm5 (berry) interacts with terminal Csm2 units on 3' end of crRNA. Large subunit proteins (purple), Cm2 and Csm1 are found around the 5' end of crRNA, close to tag. Csm crRNP recognizes invader DNA based on non-complementarity to 5'tag. Representation of RNA targeting Cmr crRNP is shown. Cmr4 is the effector nuclease of Type III-B (Cmr) system. The detailed mechanism of invader DNA/ RNA cleavage and corresponding effector nucleases of Type III-A is currently unknown.


Effector complex structure
Invader silencing
Type III-A
DNA targeting
Csm2
Csm2
Csm2
Csm5
5'
Csm4
Csm3
Csm3
Csm3
Csm3
Csm3
3'
Csm1
3'
5'
5'
3'
Csm6
Type III-B
RNA targeting
Cmr5
Cmr5
Cmr5
Cmr6
5'
Cmr3
Cmr4
Cmr4
Cmr4
Cmr4
Cmr1
3'
Cmr2
3'
5'

CHAPTER 2

THREE CRISPR-CAS IMMUNE EFFECTOR COMPLEXES CO-EXIST IN *PYROCOCCUS FURIOSUS*

---

[1]Sonali Majumdar, Peng Zhao, Neil Pfister, Mark Compton, Sara Olson, Claiborne V. C. Glover, Lance Wells, Brenten R. Graveley, Rebecca M. Terns, Michael P. Terns. Submitted to RNA.

Abstract

CRISPR-Cas immune systems function to defend prokaryotes against potentially harmful mobile genetic elements including viruses and plasmids. The multiple CRISPR-Cas systems (Types I, II, III) each target destruction of foreign nucleic acids via structurally and functionally diverse effector complexes (crRNPs). CRISPR-Cas effector complexes are comprised of CRISPR RNAs (crRNAs) that contain sequences homologous to the invading nucleic acids and Cas proteins specific to each immune system type. We have previously characterized a crRNP in *Pyrococcus furiosus* (Pfu) that contains Cmr (Type III-B) Cas proteins associated with one of two size classes of crRNAs and cleaves complementary target RNAs. Here, we have isolated and characterized two additional native Pfu crRNPs containing either Csa (Type I-A) or Cst (Type I-G) Cas proteins and distinct profiles of associated crRNAs. For each complex, the Cas proteins were identified by mass spectrometry and immunoblotting and the crRNAs by RNA sequencing and Northern blot analysis. The crRNAs associated with both the Csa and Cst complexes originate from all seven Pfu CRISPR loci and contain identical 5' ends (8-nt repeat-derived 5' tag sequences) but heterogeneous 3' ends (containing variable amounts of downstream repeat sequences). These crRNA forms are distinct from Cmr-associated crRNAs, indicating different 3' end processing pathways following primary cleavage of common pre-crRNAs. Like other previously characterized Type I CRISPR-Cas effector complexes, we predict that the newly identified Pfu Csa and Cst crRNPs each function to target invading DNA, adding an additional layer of protection beyond that afforded by the previously characterized RNA targeting Cmr complex.

Introduction

CRISPR-Cas systems confer adaptive immunity to prokaryotic hosts against viruses, conjugative plasmids, and other potential genome invaders (Terns and Terns 2011b; Wiedenheft et al. 2012; Reeks et al. 2013b; Gasiunas et al. 2014; van der Oost et al. 2014). A host CRISPR (clustered regularly interspaced short palindromic repeats) locus contains a leader region (typically 100-500 base pairs) followed by multiple copies of direct DNA repeat units (~30-40 base pairs) separated by similar-sized, variable invader-derived sequences (Bolotin et al. 2005; Mojica et al. 2005b; Pourcel et al. 2005; Lillestol et al. 2006; Makarova et al. 2006c; Grissa et al. 2007). CRISPR RNAs (crRNAs) are generally produced by transcription from promoter elements in the leader region (Jansen et al. 2002a), and each crRNA contains a guide region comprising invader-derived sequences for direct targeting of invader nucleic acids by base-pairing (Barrangou et al. 2007; Marraffini and Sontheimer 2008; Hale et al. 2009; Garneau et al. 2010; Jore et al. 2011b; Lintner et al. 2011c; Wiedenheft et al. 2011a; Fischer et al. 2012a; Gasiunas et al. 2012a; Hale et al. 2012b; Jinek et al. 2012; Westra et al. 2012a; Elmore et al. 2013b; Maier et al. 2013a; Mulepati and Bailey 2013a; Peng et al. 2013a; Sinkunas et al. 2013b; Plagens et al. 2014a).

CRISPR-associated (Cas) proteins provide enzymatic machinery and structural components to carry out distinct phases of the CRISPR-Cas pathway including acquisition and integration of invader-derived DNA fragments into CRISPR loci (adaptation), crRNA biogenesis, and crRNA-guided invader silencing. CRISPR-Cas systems are grouped into Type I (A-G), Type II (A-C), and Type III (A-B) systems based primarily on Cas protein superfamily composition (Makarova et al. 2011b; Vestergaard et

al. 2014). Each system corresponds to a Cas protein subtype (e.g., Type I-A for Csa, Type I-E for Cse, Type III-B for Cmr, etc.) and contains a signature Cas superfamily protein (Cas3 in Type I, Cas9 in Type II, Cas10 in Type III) (Haft et al. 2005b; Makarova et al. 2006c; Makarova et al. 2011b). The previously recognized Cst and Csh Cas protein subtypes (Haft et al. 2005b) were recently consolidated into a single Type I-B immune system in one Cas protein nomenclature scheme (Makarova et al. 2011b). However, a recent bioinformatics analysis implicates the Cst system (including Cst of *P. furiosus*) as belonging to a distinct, Type I-G immune system (Vestergaard et al. 2014).

The Cas6 superfamily of endoribonucleases play a key role in crRNA biogenesis in many Type I and Type III systems through cleaving pre-crRNA transcripts once within each repeat region generating crRNAs with short (typically 8 nucleotides) 5' repeat-derived sequences (5' tags), followed by invader-derived (guide) sequences and the remaining (typically 20-30 nucleotide) 3' repeat-derived sequences (3' tags) (Brouns et al. 2008; Carte et al. 2008a; Carte et al. 2010; Haurwitz et al. 2010; Gesner et al. 2011a; Lintner et al. 2011c; Garside et al. 2012; Nam et al. 2012b; Nickel et al. 2013a; Carte et al. 2014a). In several Type I and Type III systems, crRNAs undergo further trimming of 3' repeat sequences (via unclear mechanisms) to form mature crRNAs (Brouns et al. 2008; Hale et al. 2008; Haurwitz et al. 2010; Gesner et al. 2011a; Hatoum-Aslan et al. 2011a; Hale et al. 2012b; Elmore et al. 2013b; Maier et al. 2013a; Nickel et al. 2013a; Peng et al. 2013a).

Mature crRNAs associate with different subtype-specific Cas proteins to form the effector complexes of different immune systems. The crRNAs guide the effector complexes to target invader RNA or DNA (depending on the system) by base pairing.

The Cas protein effector nuclease of the system (which may or may not be a stable component of the effector complex) cleaves the nucleic acid of the virus or other trespasser (Brouns et al. 2008; Hale et al. 2009; Garneau et al. 2010; Jore et al. 2011b; Lintner et al. 2011c; Wiedenheft et al. 2011a; Fischer et al. 2012a; Gasiunas et al. 2012a; Hale et al. 2012b; Jinek et al. 2012; Nam et al. 2012b; Plagens et al. 2012b; Mulepati and Bailey 2013a; Sinkunas et al. 2013b; Staals et al. 2013b; Plagens et al. 2014a; Zebec et al. 2014b).

Type I systems have a similar *cas* content and crRNP composition and organization (Makarova et al. 2011d; Wiedenheft et al. 2011a; Wiedenheft et al. 2011c; Nam et al. 2012b). Characterized Type I crRNPs contain Cas5, Cas7, Cas8 (large subunit) and Cas11 (small subunit) Cas proteins (the latter sometimes fused to the large subunit as a domain) as structural components (Brouns et al. 2008; Lintner et al. 2011c; Wiedenheft et al. 2011a; Wiedenheft et al. 2011c; Nam et al. 2012b; Plagens et al. 2012b; van Duijn et al. 2012b; Brendel et al. 2014). Effector crRNPs of Type I systems target invader DNA with the signature Cas3 (HD nuclease-DExH helicase) protein catalyzing DNA cleavage (Cady and O'Toole 2011a; Westra et al. 2012a; Mulepati and Bailey 2013a; Sinkunas et al. 2013b; Hochstrasser et al. 2014b; Plagens et al. 2014a).

Pyrococcus furiosus (*Pfu*), a hyperthermophilic archaeaon, constitutively expresses approximately 200 mature crRNAs transcribed and processed from seven CRISPR loci (with each locus having essentially the same CRISPR repeat sequence) (Hale et al. 2008; Hale et al. 2009; Hale et al. 2012b; Terns and Terns 2013). *Pfu* Cas proteins are encoded at two gene clusters and comprise predicted adaptation proteins (Cas1, Cas2, Cas4), a crRNA biogenesis and processing protein (Cas6) and three predicted subtype-specific

Cas immune system proteins (Csa or Type I-A, Cst or Type I-G, and Cmr or Type III-B) (Figure 1A). Cas6 is the sequence-specific, metal-independent endonuclease that cleaves within each repeat of a pre-crRNA transcript to generate 1X intermediate RNAs containing an 8 nt, repeat-derived 5'-tag (Carte et al. 2008a; Hale et al. 2008; Carte et al. 2010; Wang et al. 2011). Following variable 3' end processing by unknown mechanisms, mature crRNAs are generated (Hale et al. 2008). Cmr proteins (Cmr1-6) interact with two small crRNA species that are 45 and 39 nucleotides in length differing at their 3' ends, and form an effector complex (Cmr or Type III-B) that cleaves complementary target RNAs (Hale et al. 2009; Hale et al. 2012b; Spilman et al. 2013b; Hale et al. 2014; Ramia et al. 2014).

In this work, we have isolated and identified two distinct native complexes comprised of the Csa (Type I-A) or Cst (Type I-G) Cas proteins and associated crRNAs originating from all seven *Pfu* CRISPR loci. These crRNAs have common 5' ends generated by Cas6 cleavage (5' tag element) but possess different processed 3' ends that often include a portion of the downstream repeat.

We propose that these two newly identified Type I crRNPs, Csa (Type I-A) and Cst (Type I-G), each function by carrying out crRNA-guided DNA targeting of invading mobile genetic elements.

Results

Isolation of native Csa and Cst containing crRNP complexes

Our previous biochemical fractionation analyses of *P. furiosus* (*Pfu*) cell extract revealed that crRNAs are components of chromatographically distinct RNA-protein complexes

(Hale et al. 2008). One of the mature crRNA-containing complexes (crRNPs) was isolated and identified as the Type III-B (Cmr) complex that contains one of two major crRNA size forms (45 and 39 nucleotides) and cleaves complementary RNA target (Hale et al. 2009; Hale et al. 2012b; Spilman et al. 2013b; Hale et al. 2014; Ramia et al. 2014). The other crRNPs identified by chromatography remained uncharacterized. In this work, we investigated the molecular composition of these uncharacterized effector complexes (crRNPs).

To gain insight into complexes containing Csa and Cst proteins, polyclonal antibodies were raised against Csa (Csa2 - a Cas7 superfamily protein, Cas3" - Cas3 subunit with predicted effector nuclease activity) and Cst (Cst2 and Cas5t - Cas7 and Cas5 superfamily proteins, respectively) proteins (Figure 1A). The antibody generated against Cmr2 was used as a positive control to isolate the previously characterized Cmr crRNP by this method. Pre-immune (PI) antibodies (isolated before antigen injection) served as specificity controls. Csa, Cst, and Cmr polyclonal antibodies were first validated by Western blotting with *Pfu* cellular extract (Figure 1B). All five antibodies recognized a protein of the expected size in Western blots confirming expression of each of these Cas proteins in *Pfu*.

Cas protein composition of the isolated complexes

We used these antibodies to immunopurify Csa, Cst, and Cmr-containing complexes from *Pfu* cell extract (Figure 2). The protein components of the immunopurified complexes were identified by tandem mass spectrometry of tryptic peptides obtained from bands excised from silver-stained gels or shotgun analysis of samples eluted from

the immunobeads (Figure 2A and 2B, Supplemental Figure S1 and Table 1). Mock immunoprecipitations, performed with matching pre-immune antibodies, were employed to filter for proteins specifically recognized by immune antibodies. As expected, Cmr2 antibodies selectively co-immunoprecipitated each of the six Cmr proteins (Cmr1-6) that were previously identified in the Cmr complex isolated through successive column chromatography (Hale et al. 2009; Hale et al. 2012b). Immunopurifications performed with antibodies against two Csa (Type I-A) proteins (Csa2 (Cas7) and Cas3" (HD nuclease); see Figure 1A), resulted in selective co-immunoprecipitation of each of the Csa proteins encoded in cas gene cluster 2 (Figure 2 and Supplemental Figure S1 and Table 1) with the exception of the Csa3 protein (whose gene is configured in the opposite orientation to the rest of the genes in the csa operon) (Figure 1A). Likewise, the two antibodies against the Cst proteins (Cst2 (Cas7) and Cas5t) selectively immunopurified only Cst proteins adjacently encoded in gene cluster 1 (Figure 2 and Supplemental Figure S1 and Table 1). In this case, either antibody isolated both Cst2 and Cas5t proteins while two other predicted components of the Cst (Type I-G) immune system (Cst1 (Cas8a1) and Cas3; see Figure 1A), were not identified. A few non-Cas proteins were specifically detected in some of these immunopurifications (Supplemental Table1).

To verify results of our mass spectrometry analyses, Western blotting was carried out for all five immunoprecipitation reactions (Csa2, Cas3", Cst2, Cas5t and Cmr2) with each blot being probed with one of the antibodies (Figure 3). The untreated *Pfu* cellular extract was used as a control in each blot to detect the antigen of expected molecular weight as in Figure 1B. As expected from the mass spectrometry results, Csa2 and Cas3" were both detected in complexes co-immunoprecipitated with Csa (Csa2, Cas3")

antibodies. Cst2 and Cas5t were seen specifically in complexes isolated with Cst (Cst2 and Cas5t) antibodies, and Cmr2 was detected only in the Cmr complex. In summary, the findings (Figures 2 and 3 and Supplemental Figure S1 and Table 1) provide evidence for three compositionally distinct, *Pfu* complexes containing Csa, Cst, or Cmr proteins.

Northern analysis of crRNAs associated with the Csa and Cst complexes.

To determine if the isolated Csa and Cst complexes also contained crRNAs, we performed Northern blot analysis (Figure 4). RNAs extracted from the total RNA samples as well as the immunopurified complexes were probed for the first crRNA in CRISPR locus 7 (7.01 crRNA). The immunopurified Cmr complex shows the expected doublet of mature crRNAs of two major size forms (45 and 39 nucleotides) (Hale et al. 2009; Hale et al. 2012b). In contrast, the immunopurified Csa (Csa2, Cas3" immunoprecipitations) and Cst (Cst2, Cas5t immunoprecipitations) complexes show a profile that consists of more diverse crRNAs size forms ranging from 45 to ~65 nt in length with the 45 nt size form being the most prominent species in each sample. The smaller, Cmr-associated mature crRNA (39 nt) is not detected in association with either the Csa or Cst complexes.

RNA sequencing of the associated crRNAs

To gain a comprehensive understanding of the sequence features of crRNAs associated with Csa and Cst complexes, we sequenced both the total RNAs and complex-associated RNAs. crRNAs associated with immunopurified Csa (Csa2, Cas3"), Cst (Cst2, Cas5t) and Cmr (Cmr2) complexes map to all seven *Pfu* CRISPR loci (Figure 5 and

Supplemental Figures S2-S7) revealing that each expressed crRNA becomes integrated into three distinct CRISPR-Cas complexes. Figure 5 shows a representative profile of crRNAs from CRISPR locus 5 that associate with the Csa, Cst, and Cmr complexes compared to those in the *Pfu* cell extract (total RNA). As previously reported, the vast majority of crRNA reads in total RNA are on the sense strand (transcribed from promoters in leader sequences), and crRNA expression is generally higher at the leader-proximal end of the CRISPR locus (Figure 5 and (Hale et al. 2008; Hale et al. 2012b)). These same trends hold true for crRNAs associated with the Csa, Cst, and Cmr complexes (Figure 5). The profile of individual crRNA reads associated with Csa and Cst complexes look similar to the total crRNA profiles and differ from Cmr-associated crRNAs, which show very sharp 3' end boundaries. Similar crRNA expression trends are observed across all seven CRISPR loci (Supplemental Figures S2-S7 show the profiles for the other CRISPR loci).

To map the 5' and 3' ends of crRNAs associated with immunopurified Csa (Csa2, Cas3" immunoprecipitations), Cst (Cst2, Cas5t immunoprecipitations), and Cmr (Cmr2 immunoprecipitation) complexes compared to crRNA population in total RNA, we plotted 5' and 3' ends of crRNAs from RNA sequencing results (Figure 6). The 5' ends of the vast majority of crRNAs from the total crRNA samples as well as the complex-associated crRNAs contain the 8-nt (-8) repeat-derived 5'-tag (AUUGAAAG), generated by Cas6 cleavage of pre-crRNA transcripts (Carte et al. 2008a). In contrast, the 3' end of the crRNAs associated with different complexes shows considerable heterogeneity. The crRNAs associated with Cmr complex exhibit two distinct peaks at the 3' end, one ending at the guide-repeat junction (-1) and the other peak 6 nt upstream (-7) and

correspond to the previously characterized 45 and 39 nt size forms of Cmr complex-associated crRNAs, respectively (Hale et al. 2008; Hale et al. 2009; Hale et al. 2012b). The crRNAs associated with the Csa (Csa2, Cas3") and Cst (Cst2, Cas5t) complexes exhibit heterogeneity at the 3' end, containing variable amounts of 3' repeat sequences. Both Csa and Cst associated crRNAs show one peak at the guide-repeat junction (-1). There are other peaks at the downstream 3' repeat (+5, +14), corresponding to bigger crRNA forms. Another peak corresponds to smaller crRNA forms, with its 3' extremity at 3 nt upstream (Csa) and 4 nt upstream (Cst) of the guide-repeat junction. Overall, the crRNA sizes obtained from mapping 5' and 3' ends (Figure 6) are in accord with the crRNA species detected in Csa, Cst, and Cmr complexes by Northern blotting (Figure 4). Northern blotting showed a 45 nt 7.01 crRNA in all three complexes, which would correspond to the -1 species at the 3' guide-repeat junction. The 39-nt crRNA (-7 form) specifically associates with the Cmr complex. The Csa, Cst complex blots probed for 7.01 detect longer crRNA species, which would correspond to +5 and +14 peaks seen in the sequencing data.

In summary, Csa, Cst and Cmr complexes associated with a subset of processed crRNAs transcribed from all seven CRISPR loci. All complex-associated crRNAs harbor a 5' tag sequence and appear to share one species of crRNA, ending at the guide-3' repeat junction. In general, the Csa- and Cst-associated crRNA forms share more similarity with each other than the Cmr-associated crRNAs.

Discussion

Three distinct crRNPs co-exist in *Pyrococcus furiosus*

The *Pfu* genome encodes three potential CRISPR-Cas immune effector systems, Type I-A (Csa), Type I-G (Cst), and Type III-B (Cmr) systems (Figure 1A). Here, we provide direct evidence that all three systems are expressed and assemble into compositionally distinct crRNPs. We hypothesize that each of these three co-existing crRNPs (Csa, Cst, and Cmr) can independently act as immune effector complexes that collectively mediate invader interference at both the RNA (Cmr) and DNA (Csa and Cst) levels (Figure 7).

Distinct profiles of crRNA size forms in Csa, Cst, and Cmr complexes

Interestingly, the ~200 crRNAs which are expressed from seven *Pfu* CRISPR loci each sort into the three distinct (Csa, Cst, and Cmr) Cas protein-containing complexes (Figure 5 and Supplemental Figures S2-S7) (these seven CRISPR loci are distributed throughout the genome including at sites distal from the two cas gene clusters) (Terns and Terns 2013). Furthermore, detailed analysis of the crRNA size forms specifically associated with each crRNP revealed that different size forms of each crRNA are integrated into the three distinct complexes (Figures 4 and 6). The crRNAs present in each complex have a common 5' end (CRISPR repeat-derived 8 nt 5' tag generated by Cas6 cleavage of primary CRISPR transcripts) but differ at their 3' termini. While two major size forms of each crRNA (39 and 45 nt in length) are associated with the Cmr complex, a diverse collection of crRNA size forms containing various numbers of 3' CRISPR repeat nucleotides occur in both the Csa and Cst complexes (Figures 4 and 6). A crRNA size form (a 45 nt crRNA species) common in all three complexes in relatively high levels

contains the entire guide sequence and no 3' repeat sequence (the average encoded crRNA guide sequence (invader-derived) is 37 nt in *Pfu*) (Figures 4 and 6). In the case of Cmr-associated crRNAs, both the 45 and 39 nt crRNA forms are functional suggesting that the precise 3' terminus is not critical for activity (Hale et al. 2009; Hale et al. 2012b; Hale et al. 2014). Structural studies reveal that each Cmr crRNP accommodates a single crRNA species (Zhang et al. 2012b; Spilman et al. 2013b; Staals et al. 2013b; Ramia et al. 2014), and this is very likely the case for the Csa and Cst complexes, based on recent structural information of other Type I systems (Jackson et al. 2014b; Mulepati et al. 2014b; Zhao et al. 2014). Further experiments are required to determine if all, or a subset, of the multiple size forms of crRNAs within the Csa and Cst crRNPs are functional or if crRNA 3' trimming leads to functional activation for these specific complexes.

Molecular mechanisms governing crRNA 3' end trimming in *Pfu* are unknown, but work from other systems suggest that non-Cas ribonucleases likely execute the 3' trimming (Hatoum-Aslan et al. 2013a). This finding suggests that the reason why each of the three crRNP complexes harbor crRNAs with distinct profiles of 3' termini relates to differences in protein protection afforded by each assembled Cas complex against further 3' trimming into the body of the crRNA (Hatoum-Aslan et al. 2011a; Hatoum-Aslan et al. 2013a). These results further indicate that 3' end trimming most likely occurs coincident with or after Cas protein assembly.

crRNA biogenesis and sorting into the three distinct Cas complexes

The mechanism(s) governing the sorting of the same crRNA into three distinct complexes in the cell is unknown. All seven CRISPR loci in *Pfu* have highly similar repeat

sequences, and the Cas6 endoribonuclease cleaves primary transcripts expressed from all CRISPR loci into 1X intermediate RNAs that contain the entire crRNA guide region flanked by 5' and 3' repeat sequences (8-nt repeat at 5' end (5' tag) and 22 nt repeat at 3' end that eventually undergo trimming) (Figure 7 and (Carte et al. 2008a; Hale et al. 2008; Hale et al. 2012b)). Our previous Cas6 immunoprecipitation analysis indicated that *in vivo*, Cas6 stably associates with 1X intermediate crRNA but not with 3' trimmed crRNAs (3' end trimming destroys the Cas6 binding site located at the 5' end of the repeat sequence found 3' to the guide RNA of each 1X crRNA intermediate) (Carte et al. 2010; Wang et al. 2011). This conclusion is further strengthened since Cas6 was not observed by mass spectrometry to be a component of isolated Csa, Cst, or Cmr crRNPs (Figure 2 and Supplemental Table 1). Taken together, the data indicate that Cas6 dissociates before or during crRNP assembly and crRNA 3' end trimming, and thus that Cas6 may play a transient role in directly or indirectly facilitating sorting of crRNAs into the three distinct crRNPs (Carte et al. 2008a).

The 5' tag sequence likely plays a critical role in nucleating crRNA assembly into each of the three complexes. This 8 nt 5' tag sequence (generated by Cas6 processing) is the only common sequence element of each of the 200 or so crRNAs in the total RNA pool and is critical for Cmr crRNP assembly and function (Hale et al. 2008; Hale et al. 2012b; Zhang et al. 2012b; Hale et al. 2014). It is likely that different proteins present within the Cmr, Csa, and Cst complexes recognize the 5' tag, and recent work supports a role for the Cas5 superfamily of proteins in mediating direct contact with 5' tag sequences of crRNAs (Spilman et al. 2013b; Hale et al. 2014; Jackson et al. 2014b; Mulepati et al. 2014b; Zhao et al. 2014). For the Cmr complex, the Cmr3 protein is the

Cas5 superfamily member that fulfills the role of direct 5' tag binding (Spilman et al. 2013b; Ramia et al. 2014). Presumably, the Cas5 superfamily proteins of the Csa and Cst complex (Cas5a and Cas5t, respectively) recognize the 5' tag sequences in the context of Csa and Cst crRNPs. Interestingly, our data suggest that these Cas5 proteins may recognize slightly different features of these 5' tag sequences. For example, the first ten crRNAs from CRISPR 8 are underrepresented in Cmr complex as we have reported earlier (Hale et al. 2012b), but not in Csa and Cst complexes (Figure S7). These crRNAs have a single nucleotide variation in the repeat-derived 5' tag sequence that apparently accounts for their ability to associate with the Csa and Cst complex but not the Cmr complex.

Cas protein composition and functional roles of Csa and Cst crRNPs

Our data strongly support the existence of three separate crRNPs each containing non-overlapping Cas protein subunits (Figures 2 and 3, Supplemental Figure S1 and Table 1). The immunopurified Type III-B Cmr complexes contain all Cmr proteins (Cmr1-6) that have been previously identified by conventional column chromatography purification (Hale et al. 2009) and shown to be individually required for Cmr RNA targeting activity *in vitro* using recombinant proteins (Hale et al. 2009; Hale et al. 2012b; Hale et al. 2014). Likewise, immunopurified Type I-A (Csa) crRNPs were found to contain each of the encoded Csa proteins with the exception of Csa3, a protein thought to be a transcriptional regulator of Csa genes rather than a component of the Csa effector complex (Lintner et al. 2011a; Makarova et al. 2014). The first experimental evidence for the existence of native Csa crRNPs came from work with the thermophilic archaeaon *Sulfolobus solfataricus*,

where affinity purifications performed with tagged Csa2 (Cas7) provided evidence for a stable interaction of Csa2 (Cas7) with Cas5a (Cas5) and crRNA with possible additional (less stable) interactions with Csa5 (Cas11 or small subunit protein), Csa4 (Cas8), and Cas6 (Lintner et al. 2011c). Similar to our findings, Type I-A (Csa) crRNP reconstituted with recombinant proteins in *T. tenax* contained Cas3" and Cas3' as stable components of the complex (Plagens et al. 2012b; Plagens et al. 2014a). The recombinant *T. tenax* Csa crRNP targets DNA, and Cas3" (HD-nuclease protein) is the effector nuclease (Plagens et al. 2014a). We expect the stably-associated Cas3" is likely the effector nuclease of the predicted DNA-targeting Type I-A (Csa) effector complex in *Pfu*.

The predicted Cas proteins required for a functional Type I-G (Cst) crRNP include Cst1, Cst2, Cas5t, and Cas3 (Haft et al. 2005b; Makarova et al. 2011b; Vestergaard et al. 2014). Our immunopurified complexes contained Cst2 and Cas5t but lacked Cst1 and Cas3. Cst2 and Cas5t are Cas7 and Cas5 superfamily proteins of Type I-G systems that likely make up the stable crRNA-binding core of the complex. Analysis of Type I-E (Cse) crRNPs provided support that the large subunit protein Cse1 (Cas8 superfamily) is a relatively weakly-associated and peripheral component of the complex (Jore et al. 2011b). Likewise, Cst1 (also a Cas8 superfamily protein) may be less stably associated with Type I-G (Cst) crRNP. While we failed to identify the Cst1 subunit by our immunopurification approach (Figure 2), *Pfu* Cst1 was found in complex with both Cst2 and Cas5t following sequential non-denaturing column chromatography fractionations (Menon et al. 2009a). Collectively, the data suggest that all three Cst proteins (Cst1, Cst2, Cas5t) are integral subunits of Type I-G crRNPs. We predict that Cas3 (PF1120), which occurs immediately adjacent to cst genes in the gene cluster (Figure 1A) will co-

function with Cst complex as the effector nuclease/helicase. Based on expectations from other Type I systems, the Cst crRNP may function as a surveillance complex that recruits Cas3 nuclease to cleave invader DNA (Brouns et al. 2008; Cady and O'Toole 2011a; Jore et al. 2011b; Wiedenheft et al. 2011a; Wiedenheft et al. 2011c; Nam et al. 2012b; Sashital et al. 2012a; van Duijn et al. 2012b; Westra et al. 2012d; Mulepati and Bailey 2013a; Sinkunas et al. 2013b; Hochstrasser et al. 2014b; Huo et al. 2014b). We likely did not detect Cas3 in our isolated Cst crRNP complexes (Figure 2) because the Cas3 protein may not associate with Cst complexes in the absence of (invader) DNA targets (Hochstrasser et al. 2014b; Huo et al. 2014b).

A small number of non-Cas proteins were identified in the Csa, Cst, and/or Cmr complex, but further work is required to understand if any of these proteins are functionally significant for crRNP regulation or function (Supplemental Table 1). For example, the thermosome subunit (PF1974), identified in Csa immunoprecipitations, is a heat shock protein (Hsp60 homolog) and could act as a chaperone in proper folding of some Cas proteins (Klumpp and Baumeister 1998; Shockley et al. 2003). By analogy, previous work suggests that an *E. coli* chaperonin protein (high temperature protein G) is essential for maintaining functional levels of Cas3 and thereby CRISPR-mediated silencing (Yosef et al. 2011). Moreover, PF1563, a subunit of DNA-dependent RNA polymerase was also identified in Csa immunoprecipitations. Interestingly, *T. kodakarensis* Csa2 (TK0462) was also shown to co-purify with RNA Polymerase (Santangelo et al. 2007) indicating a potential conserved interaction between Csa components and transcriptional machinery.

Advantages of multiple CRISPR-Cas systems

While some organisms have a single CRISPR-Cas immune system, many harbor more than one system. The three co-existing complexes of *Pfu* very likely target invaders at both RNA (Cmr) and DNA (Csa, Cst) levels for silencing (Figure 7). Having multiple effector complexes offers potential advantages. We now know that crRNAs from all seven loci associate with each of the three (Csa, Cst, Cmr) co-existing complexes in *Pfu* (Figures 4, 5 and Supplemental Figures 2-7). Therefore, a single crRNA can potentially guide targeting by each of the three complexes. Having both RNA and DNA-targeting systems could provide silencing of different types of viruses and other mobile genetic elements that for *Pfu* remain largely uncharacterized. Furthermore, we speculate that having multiple, and potentially redundant, CRISPR-Cas immune systems would be highly beneficial in the event that a virally encoded "anti-CRISPR" gene functions to inactivate a specific CRISPR-Cas type as has been observed for phages that infect *Pseudomonas aeruginosa* (Bondy-Denomy et al. 2013; Pawluk et al. 2014). The identification of the Csa and Cst crRNPs described in this study lays the foundation for a more detailed structural and functional analysis of these complexes akin to those that have been performed to understand the previously identified Cmr complex (Hale et al. 2009; Hale et al. 2012b; Spilman et al. 2013b; Hale et al. 2014; Ramia et al. 2014). Further investigation of these three potentially independent and co-functioning complexes will provide insights into multiple layers of CRISPR-Cas mediated invader defense in *P. furiosus*.

Materials and Methods

Preparation of polyclonal antibodies against Cas proteins

For antigen production, *csa2*, *cas3"* (pET24D, Novagen), *cas5t*, *cst2*, and *cmr2* (pET200D, Invitrogen) genes were cloned and transformed into *E. coli* (BL21 RIPL cells, Agilent Technologies). N-terminal, 6X His-tagged recombinant proteins were expressed in *E. coli* by inducing with 1 mM IPTG at mid-log growth phase (OD_{600} at 0.4-0.7) and incubated overnight at room temperature with shaking and purified by nickel chromatography as described (Hale et al. 2009). Polyclonal antibodies against each antigen were raised in chickens and purified as previously described (Carte et al. 2010).

Preparation of *Pfu* cell extract

Approximately 5 grams of *Pfu* cells were resuspended in 20 ml of 50 mM Tris (pH 8.0) in the presence of 100 U of RQ1 DNase (Promega) and 0.1 mM phenylmethanesulfonylfluoride (PMSF) and lysed by stirring at room temperature. The cell extract was then centrifuged at 100,000 g for 90 minutes at 4°C using a SW41 Ti rotor (Beckman). The resulting S100 extract, with a typical total protein concentration of 20 mg/ml, was stored at -80°C until further use.

Immunoprecipitation of Csa, Cst, and Cmr-containing complexes

Immunoprecipitations were performed using anti-Csa2, anti-Cas3", anti-Cst2, anti-Cas5t, and anti-Cmr2 IgY antibodies conjugated either directly to CarboLink-coupled beads (Pierce) or indirectly coupled to anti-IgY-conjugated agarose beads (Gallus Immunotech). The coupling was performed as per manufacturer's recommended

protocol and efficiency verified by recording A280 readings (before and after coupling reaction) and subsequent analysis via SDS-PAGE. Immunoprecipitation reactions with the direct conjugated method were used for mass spectrometry analysis of protein samples and those done with the indirect (IgY) coupling method were used for Northern analysis and RNA sequencing.

An individual immunoprecipitation reaction typically contained 270-550 µg of the antibody-conjugated beads, 200 µl of *Pfu* S100 extract (~4 mg of total protein), 1X Complete Mini protease inhibitor (Roche), and 50 U of SUPERase-IN RNase inhibitor (Applied Biosystems), brought up to a final volume of 1 ml with IPP 300 (10 mM Tris (pH 8.0), 300 mM NaCl, 0.05 % Igepal). The reaction was then incubated at room temperature for 2 hours with end-over-end rotation followed by centrifugation at 3,000 g for 2 minutes. The supernatant was stored at 4°C for any further analysis (e.g., Western or Northern blotting). The pellets were washed four times in 1 ml of IPP 300.

Mass spectrometry sample preparation

For in-solution mass spectrometry analysis, the immunoprecipitated samples (pre-immune and immune) were washed 2-4 times in 1 ml of 1X PBS to remove excess detergent and eluted in 100 µl of 100 mM glycine at pH 2.8 for 2-10 minutes. The eluted material was quickly neutralized in 10 µl of 1 M sodium phosphate buffer (pH 8.8). For in-gel mass spectrometry analysis, the immunoprecipitated samples (pre-immune and immune) were resuspended in SDS loading buffer (without reducing agent) and heated at 65 °C for 5 minutes followed by electrophoretic separation on 12% Mini-PROTEAN

TGX gels (Biorad). The gels were either silver stained (Blum et al. 1987; Shevchenko et al. 1996) and bands were excised or unstained gel portions were excised.

Protein Assignment by Tandem Mass Spectrometry

The data generated comes from 2-5 separate immunoprecipitation replicates. In-gel and in-solution tryptic digests were performed essentially as described previously (Wells et al. 2002; Lim et al. 2008; Hale et al. 2009). The desalted tryptic peptides were analyzed by nanoESI-LC-MS/MS on a linear ion trap mass spectrometer (LTQ XL, Thermo Fisher Scientific) as described previously (Lim et al. 2008). The acquired data was searched against a *Pyrococcus furiosus* specific database (Uniprot, updated on December 21, 2009) using SEQUEST (Thermo Fisher Scientific). The search output was filtered to obtain a 1% protein false discovery rate (99% confidence) using the ProValT algorithm as deployed in PROTEOIQ (www.nusep.com). The proteins identified in Table 1 were not present in the corresponding pre-immune control samples and contained two or more peptides.

Western Blotting

Pfu S100 extract (~40 μg total protein) and one-third of an immunoprecipitation reaction were typically used. Western blotting was performed by standard procedures. The blots were incubated with 1:2,500-1:100,000 dilution of polyclonal IgY immune antibodies and 1:5,000 dilution of HRP-conjugated Anti-IgY secondary antibody (Gallus Immunotech). The protein bands on the blots were detected using an enhanced

chemiluminescent substrate for HRP (horse radish peroxidase) activity (ECL, Thermo Scientific Pierce).

Northern analysis

Total RNAs were extracted from *Pfu* S100 extract and immunoprecipitated samples (pre-immune and immune) using TRIzol LS Reagent (Invitrogen) as per the manufacturer's recommendation. Total RNA (2 µg) from S100 extract and one-fourth (Csa2), one-half (Cas5t, Cst2, Cmr2), and 2 reactions (Cas3") of immunoprecipitation samples were used. Northern blots were performed as described previously (Hale et al. 2008). Blots were probed with 5' ^{32}P-labeled crRNA 7.01 antisense oligo (Hale et al. 2008) and exposed on a phosphorimager screen overnight.

Illumina sequencing of crRNAs

Total RNA (50 µg) from *Pfu* S100 extract and Csa, Cst and Cmr immunoprecipitation samples (immune) were extracted using TRIzol LS reagent (Life Technologies). RNAs were electrophoresed on a 10% PAGE with 7M urea alongside oligonucleotide size standards, stained with SYBR-gold (Life Technologies), and gel slices in the size range of 19-70 nt were excised. The libraries were prepared as described (Hale et al. 2012b). Briefly, the RNAs were phosphatase-treated to remove 3' phosphates and ligated to an adaptor. These products were treated with T4 polynucleotide kinase before ligation of the 5' end to another adaptor. The products were reverse transcribed, followed by PCR. The samples were diluted to a concentration of 24 pM and subjected to 76 cycles of sequencing on an Illumina Genome Analyzer II. Reads were trimmed of the linker

sequences and aligned to the genome using Bowtie (Langmead et al.) and the end analysis performed as previously described (20).

Acknowledgements

We thank Joshua Elmore for his contribution in preparing total *P. furiosus* RNA for RNA sequencing. We thank members of Terns' lab for scientific discussions and technical input including team-effort preparation of antibodies. This work was supported by a National Institutes of Health Grant R01 GM54682 to M.P.T. and R.M.T. and in part by the Georgia Research Alliance and the National Center for Biomedical Glycomics P41GM103490 to L.W., Hatch Funds to M.C., and grants R01 GM095296 and R01 GM067842 to B.R.G.

References

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.

Blum H, Beier H, Gross HJ. 1987. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *ELECTROPHORESIS* **8**: 93-99.

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551-2561.

Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. 2013. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**: 429-432.

Brendel J, Stoll B, Lange SJ, Sharma K, Lenz C, Stachler AE, Maier LK, Richter H, Nickel L, Schmitz RA et al. 2014. A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (CRISPR)-derived RNAs (crRNAs) in *Haloferax volcanii*. *J Biol Chem* **289**: 7164-7177.

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.

Cady KC, O'Toole GA. 2011. Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J Bacteriol* **193**: 3433-3445.

Carte J, Christopher RT, Smith JT, Olson S, Barrangou R, Moineau S, Glover CV, 3rd, Graveley BR, Terns RM, Terns MP. 2014. The three major types of CRISPR-Cas

systems function independently in CRISPR RNA biogenesis in *Streptococcus thermophilus*. *Mol Microbiol*.

Carte J, Pfister NT, Compton MM, Terns RM, Terns MP. 2010. Binding and cleavage of CRISPR RNA by Cas6. *RNA* **16**: 2181-2188.

Carte J, Wang R, Li H, Terns RM, Terns MP. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* **22**: 3489-3496.

Elmore JR, Yokooji Y, Sato T, Olson S, Glover CV, 3rd, Graveley BR, Atomi H, Terns RM, Terns MP. 2013. Programmable plasmid interference by the CRISPR-Cas system in *Thermococcus kodakarensis*. *RNA Biol* **10**: 828-840.

Fischer S, Maier LK, Stoll B, Brendel J, Fischer E, Pfeiffer F, Dyall-Smith M, Marchfelder A. 2012. An Archaeal Immune System Can Detect Multiple Protospacer Adjacent Motifs (PAMs) to Target Invader DNA. *J Biol Chem* **287**: 33351-33363.

Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67-71.

Garside EL, Schellenberg MJ, Gesner EM, Bonanno JB, Sauder JM, Burley SK, Almo SC, Mehta G, Macmillan AM. 2012. Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *RNA* **18**: 2020-2028.

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

Gasiunas G, Sinkunas T, Siksnys V. 2014. Molecular mechanisms of CRISPR-mediated microbial immunity. *Cellular and Molecular Life Sciences : CMLS* **71**: 449-465.

Gesner EM, Schellenberg MJ, Garside EL, George MM, Macmillan AM. 2011. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat Struct Mol Biol* **18**: 688-692.

Grissa I, Vergnaud G, Pourcel C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**: 172.

Haft DH, Selengut J, Mongodin EF, Nelson KE. 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1**: e60.

Hale C, Kleppe K, Terns RM, Terns MP. 2008. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* **14**: 2572-2579.

Hale CR, Cocozaki A, Li H, Terns RM, Terns MP. 2014. Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex. *Genes Dev* **28**: 2432-2443.

Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover CV, 3rd, Graveley BR, Terns RM et al. 2012. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol Cell* **45**: 292-302.

Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP. 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945-956.

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

Hatoum-Aslan A, Samai P, Maniv I, Jiang W, Marraffini LA. 2013. A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *J Biol Chem* **288**: 27888-27897.

Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. 2010. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**: 1355-1358.

Hochstrasser ML, Taylor DW, Bhat P, Guegler CK, Sternberg SH, Nogales E, Doudna JA. 2014. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc Natl Acad Sci U S A* **111**: 6618-6623.

Huo Y, Nam KH, Ding F, Lee H, Wu L, Xiao Y, Farchione MD, Jr., Zhou S, Rajashankar K, Kurinov I et al. 2014. Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat Struct Mol Biol* **21**: 771-777.

Jackson RN, Golden SM, van Erp PB, Carter J, Westra ER, Brouns SJ, van der Oost J, Terwilliger TC, Read RJ, Wiedenheft B. 2014. Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* **345**: 1473-1479.

Jansen R, Embden JD, Gaastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**: 1565-1575.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821.

Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R et al. 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **18**: 529-536.

Klumpp M, Baumeister W. 1998. The thermosome: archetype of group II chaperonins. *FEBS Lett* **430**: 73-77.

Lillestol RK, Redder P, Garrett RA, Brugger K. 2006. A putative viral defence mechanism in archaeal cells. *Archaea* **2**: 59-72.

Lim JM, Sherling D, Teo CF, Hausman DB, Lin D, Wells L. 2008. Defining the regulated secreted proteome of rodent adipocytes upon the induction of insulin resistance. *J Proteome Res* **7**: 1251-1263.

Lintner NG, Frankel KA, Tsutakawa SE, Alsbury DL, Copie V, Young MJ, Tainer JA, Lawrence CM. 2011a. The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. *J Mol Biol* **405**: 939-955.

Lintner NG, Kerou M, Brumfield SK, Graham S, Liu H, Naismith JH, Sdano M, Peng N, She Q, Copie V et al. 2011b. Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J Biol Chem* **286**: 21643-21656.

Maier LK, Lange SJ, Stoll B, Haas KA, Fischer S, Fischer E, Duchardt-Ferner E, Wohnert J, Backofen R, Marchfelder A. 2013. Essential requirements for the detection and degradation of invaders by the *Haloferax volcanii* CRISPR/Cas system I-B. *RNA Biol* **10**: 865-874.

Makarova KS, Anantharaman V, Grishin NV, Koonin EV, Aravind L. 2014. CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Frontiers in Genetics* **5**: 102.

Makarova KS, Aravind L, Wolf YI, Koonin EV. 2011a. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct* **6**: 38.

Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**: 7.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF et al. 2011b. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9**: 467-477.

Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**: 1843-1845.

Menon AL, Poole FL, 2nd, Cvetkovic A, Trauger SA, Kalisiak E, Scott JW, Shanmukh S, Praissman J, Jenney FE, Jr., Wikoff WR et al. 2009. Novel multiprotein complexes identified in the hyperthermophilic archaeon *Pyrococcus furiosus* by non-denaturing fractionation of the native proteome. *Mol Cell Proteomics* **8**: 735-751.

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**: 174-182.

Mulepati S, Bailey S. 2013. *In Vitro* Reconstitution of an *Escherichia coli* RNA-guided Immune System Reveals Unidirectional, ATP-dependent Degradation of DNA Target. *J Biol Chem* **288**: 22184-22192.

Mulepati S, Heroux A, Bailey S. 2014. Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* **345**: 1479-1484.

Nam KH, Haitjema C, Liu X, Ding F, Wang H, Delisa MP, Ke A. 2012. Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System. *Structure* **20**: 1574-1584.

Nickel L, Weidenbach K, Jager D, Backofen R, Lange SJ, Heidrich N, Schmitz RA. 2013. Two CRISPR-Cas systems in *Methanosarcina mazei* strain Go1 display common processing features despite belonging to different types I and III. *RNA Biol* **10**: 779-791.

Pawluk A, Bondy-Denomy J, Cheung VH, Maxwell KL, Davidson AR. 2014. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. *MBio* **5**: e00896.

Peng W, Li H, Hallstrom S, Peng N, Liang YX, She Q. 2013. Genetic determinants of PAM-dependent DNA targeting and pre-crRNA processing in *Sulfolobus islandicus*. *RNA Biol* **10**: 738-748.

Plagens A, Tjaden B, Hagemann A, Randau L, Hensel R. 2012. Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *J Bacteriol* **194**: 2491-2500.

Plagens A, Tripp V, Daume M, Sharma K, Klingl A, Hrle A, Conti E, Urlaub H, Randau L. 2014. *In vitro* assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. *Nucleic Acids Res*.

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

Ramia N, Spilman M, Tang L, Shao Y, Elmore J, Hale C, Cocozaki A, Bhattacharya N, Terns R, Terns M et al. 2014. Essential Structural and Functional Roles of the Cmr4 Subunit in RNA Cleavage by the Cmr CRISPR-Cas Complex *Cell Reports* **In Press**.

Reeks J, Naismith JH, White MF. 2013. CRISPR interference: a structural perspective. *Biochem J* **453**: 155-166.

Santangelo TJ, Cubonova L, James CL, Reeve JN. 2007. TFB1 or TFB2 is sufficient for *Thermococcus kodakaraensis* viability and for basal transcription in vitro. *J Mol Biol* **367**: 344-357.

Sashital DG, Wiedenheft B, Doudna JA. 2012. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol Cell* **46**: 606-615.

Shevchenko A, Wilm M, Vorm O, Mann M. 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical chemistry* **68**: 850-858.

Shockley KR, Ward DE, Chhabra SR, Conners SB, Montero CI, Kelly RM. 2003. Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. *Appl Environ Microbiol* **69**: 2365-2371.

Sinkunas T, Gasiunas G, Waghmare SP, Dickman MJ, Barrangou R, Horvath P, Siksnys V. 2013. *In vitro* reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *Embo J* **32**: 385-394.

Spilman M, Cocozaki A, Hale C, Shao Y, Ramia N, Terns R, Terns M, Li H, Stagg S. 2013. Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Mol Cell* **52**: 146-152.

Staals RH, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K et al. 2013. Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell* **52**: 135-145.

Terns MP, Terns RM. 2011. CRISPR-based adaptive immune systems. *Curr Opin Microbiol* **14**: 321-327.

Terns RM, Terns MP. 2013. The RNA- and DNA-targeting CRISPR-Cas immune systems of *Pyrococcus furiosus*. *Biochem Soc Trans* **41**: 1416-1421.

van der Oost J, Westra ER, Jackson RN, Wiedenheft B. 2014. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol* **12**: 479-492.

van Duijn E, Barbu IM, Barendregt A, Jore MM, Wiedenheft B, Lundgren M, Westra ER, Brouns SJ, Doudna JA, van der Oost J et al. 2012. Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in CRISPR-associated protein complexes from *Escherichia coli* and *Pseudomonas aeruginosa*. *Mol Cell Proteomics*.

Vestergaard G, Garrett RA, Shah SA. 2014. CRISPR adaptive immune systems of Archaea. *RNA Biol* **11**: 156-167.

Wang R, Preamplume G, Terns MP, Terns RM, Li H. 2011. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure* **19**: 257-264.

Wells L, Vosseller K, Cole RN, Cronshaw JM, Matunis MJ, Hart GW. 2002. Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* **1**: 791-804.

Westra ER, Nilges B, van Erp PB, van der Oost J, Dame RT, Brouns SJ. 2012a. Cascade-mediated binding and bending of negatively supercoiled DNA. *RNA Biol* **9**.

Westra ER, van Erp PB, Kunne T, Wong SP, Staals RH, Seegers CL, Bollen S, Jore MM, Semenova E, Severinov K et al. 2012b. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* **46**: 595-605.

Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, Doudna JA, Nogales E. 2011a. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**: 486-489.

Wiedenheft B, Sternberg SH, Doudna JA. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**: 331-338.

Wiedenheft B, van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A, Westphal W, Heck AJ, Boekema EJ, Dickman MJ et al. 2011b. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* **108**: 10092-10097.

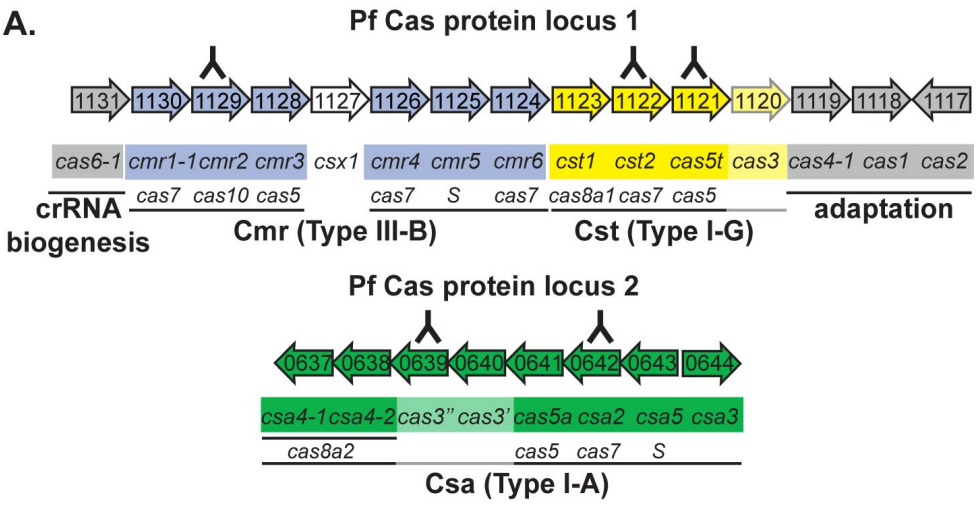
Yosef I, Goren MG, Kiro R, Edgar R, Qimron U. 2011. High-temperature protein G is essential for activity of the *Escherichia coli* clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. *Proc Natl Acad Sci U S A* **108**: 20136-20141.

Zebec Z, Manica A, Zhang J, White MF, Schleper C. 2014. CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res*.

Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, Graham S, Reimann J, Cannone G, Liu H, Albers SV et al. 2012. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* **45**: 303-313.

Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, Gong W, Wei Z, Wang Y. 2014. Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature*.

Figure 2.1. A. Organization of cas genes in *Pyrococcus furiosus*. Cas genes are mainly distributed in two gene clusters in Pfu. Genes encoding predicted adaptation Cas proteins (Cas1, Cas2, Cas4, grey), crRNA biogenesis protein (Cas6, grey) and immune-effector modules: Type I-A (Csa, green), Type I-G (Cst, yellow), Type III-B (Cmr, blue) are labeled. Genes encoding Cas3 (light yellow) and Cas3', Cas3" (light green) correspond to their subtype association. Inverted Y indicates polyclonal antibodies were raised against the specified Csa, Cst and Cmr proteins. **B. Western blotting of Pfu S100 extract with Csa, Cst and Cmr antibodies.** The indicated antibodies detect a major band (asterisk) of the expected molecular weight for each antigen. The predicted molecular weights for each protein are: Csa2 (37.4 kDa), Cas3" (22.9 kDa), Cst2 (39.7 kDa), Cas5t (23.8 kDa) and Cmr2 (100.8 kDa).


A.
Pf Cas protein locus 1
1131 1130 1129 1128 1127 1126 1125 1124 1123 1122 1121 1120 1119 1118 1117
cas6-1 cmr1-1 cmr2 cmr3 csx1 cmr4 cmr5 cmr6 cst1 cst2 cas5t cas3 cas4-1 cas1 cas2
cas7 cas10 cas5
cas7 S cas7
cas8a1 cas7 cas5
crRNA biogenesis
Cmr (Type III-B)
Cst (Type I-G)
adaptation
Pf Cas protein locus 2
0637 0638 0639 0640 0641 0642 0643 0644
csa4-1 csa4-2 cas3'' cas3' cas5a csa2 csa5 csa3
cas8a2
cas5 cas7 S
Csa (Type I-A)

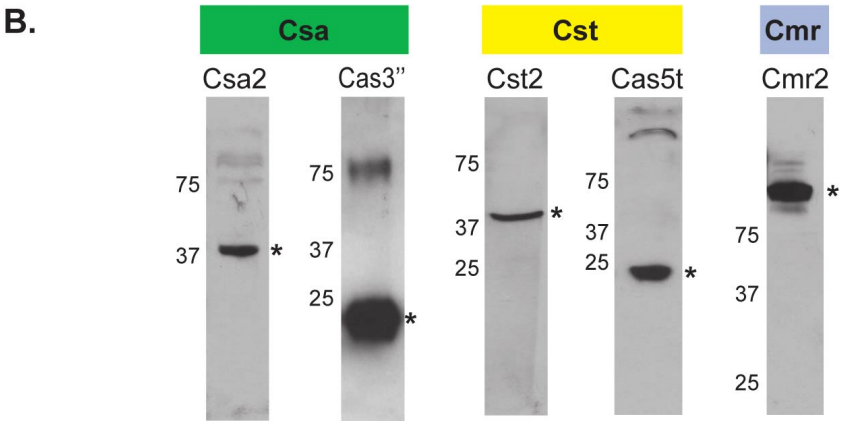

B.
Csa
Cst
Cmr
Csa2
Cas3''
Cst2
Cas5t
Cmr2
75
37
25
*

Figure 2.2 Cas protein composition of immunopurified Csa, Cst, Cmr complexes. A. Cas proteins identified by tandem mass spectrometric analysis of Csa (green), Cst (yellow), Cmr (blue) immounoprecipitated complexes are indicated with a check. Some Cas proteins predicted to be part of a complex but not identified by mass spectrometry are also listed. **B**. Silver stained SDS-polyacrylamide gels of Csa (Csa2), Cst (Cst2) and Cmr (Cmr2) immunoprecipitated pellets with pre-immune (PI) or immune (I) antibodies as labeled. Cas proteins identified by tandem mass spectrometry are indicated based on expected molecular weight. The antigens are labeled in bold. Cas proteins identified by mass spectrometry that do not correlate to a visible band are marked with an asterisk. M is the protein standard in kilodaltons (kDa).

A.

Cas Super-family	Cas Proteins	Csa IPs Csa2	Csa IPs Cas 3"	Cst IPs Cst2	Cst IPs Cas5t	Cmr IP Cmr2
Csa8a2	Csa4-1	√	√			
Csa8a2	Csa4-2	√				
Cas3	Cas3"	√	√			
Cas3	Cas3'	√	√			
Cas5	Cas5a	√	√			
Cas7	Csa2	√	√			
S	Csa5	√	√			
	Csa3					
Cas8a1	Cst1					
Cas7	Cst2			√	√	
Cas5	Cas5t			√	√	
Cas3	Cas3					
Cas7	Cmr1-1					√
Cas7	Cmr1-2					√
Cas10	Cmr2					√
Cas5	Cmr3					√
Cas7	Cmr4					√
S	Cmr5					√
Cas7	Cmr6					√

B.

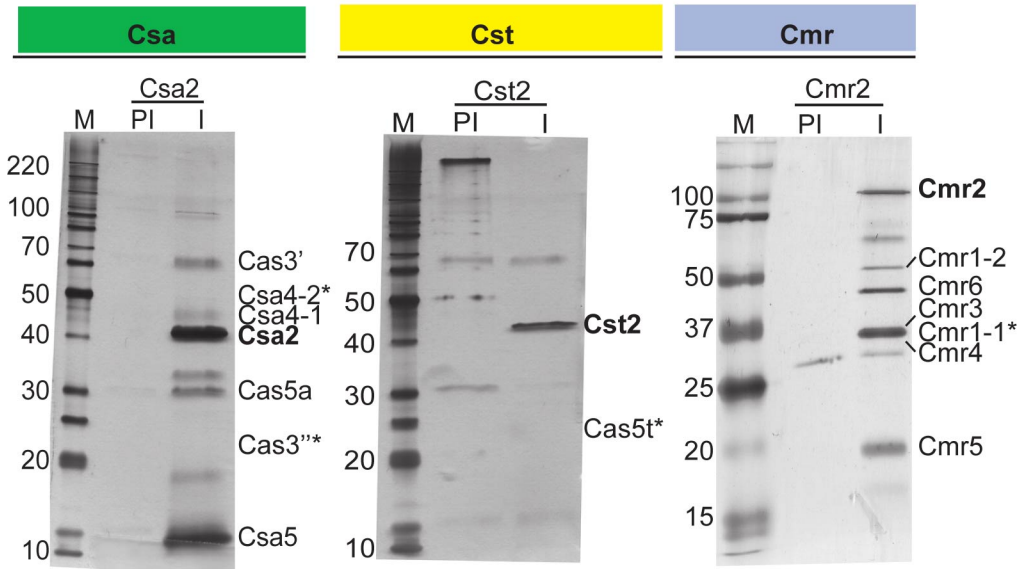


Figure 2.3. **Western blot analysis of Cas proteins associated with Csa, Cst and Cmr complexes.** All five immunopurified samples were blotted and probed with each of the five immune antibodies (Csa2, Cas3", Cst2, Cas5t, Cmr2) individually. The first lane of each blot is untreated Pfu S100 extract used as a reference. An asterisk indicates the expected position of antigens based on their molecular weights. Molecular weights are indicated in kilodaltons (kDa).

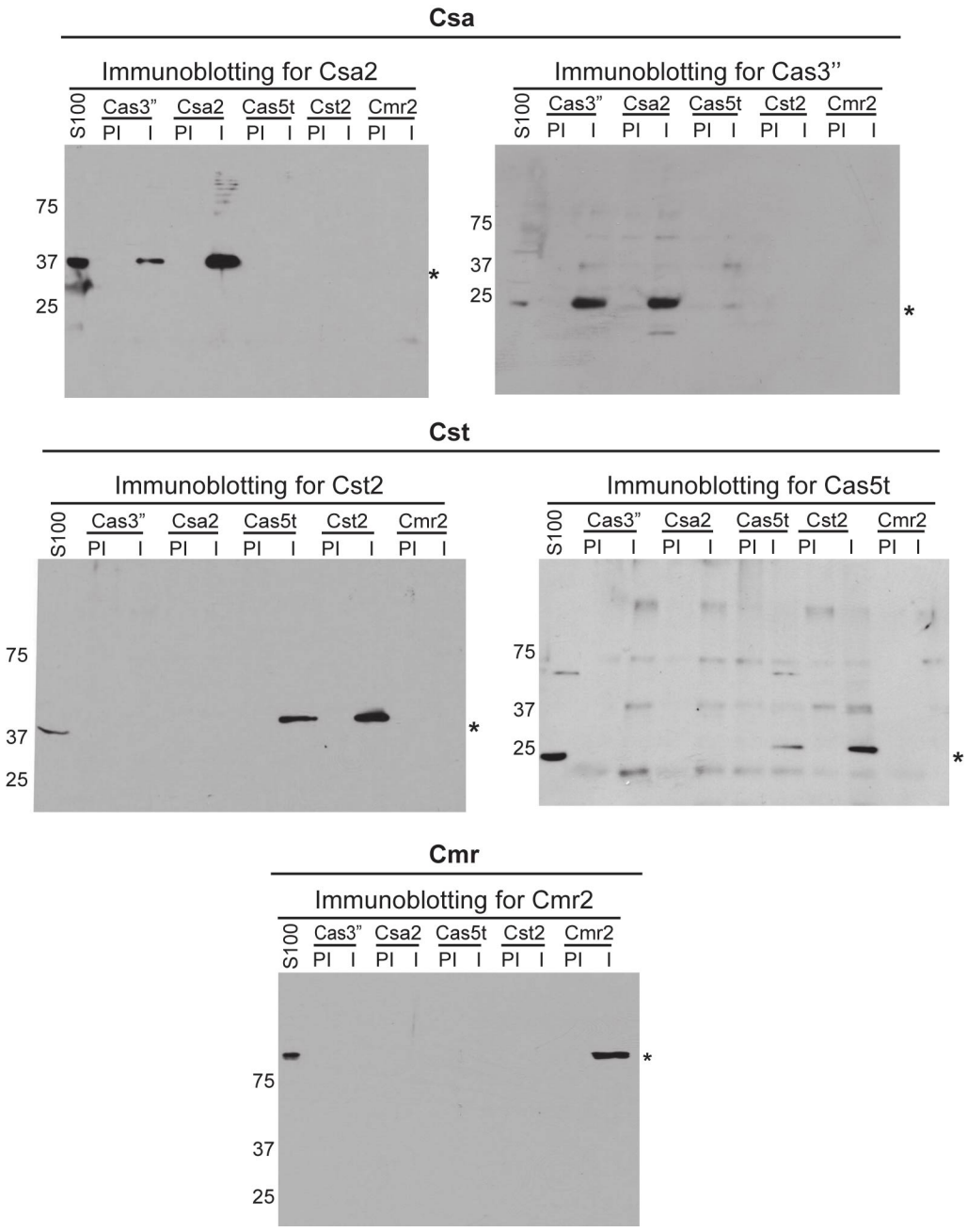

Csa
Immunoblotting for Csa2
S100
Cas3"
Csa2
Cas5t
Cst2
Cmr2
PI
I
75
37
25
*
Immunoblotting for Cas3"
Cst
Immunoblotting for Cst2
Immunoblotting for Cas5t
Cmr
Immunoblotting for Cmr2

Figure 2.4. Northern blot analysis of crRNAs associated with Csa, Cst and Cmr complexes. Northern blotting of RNAs extracted from Pfu S100 extract (T), pre-immune pellets (PI) and immune pellets (I) of Csa (Csa2, Cas3"), Cst (Cst2, Cas5t), Cmr (Cmr2). The blots were probed for the first crRNA of locus 7 (7.01). Asterisk indicates the prominent 45-nt crRNA band detected in all three complexes. Radiolabeled decade marker in nucleotides (M).

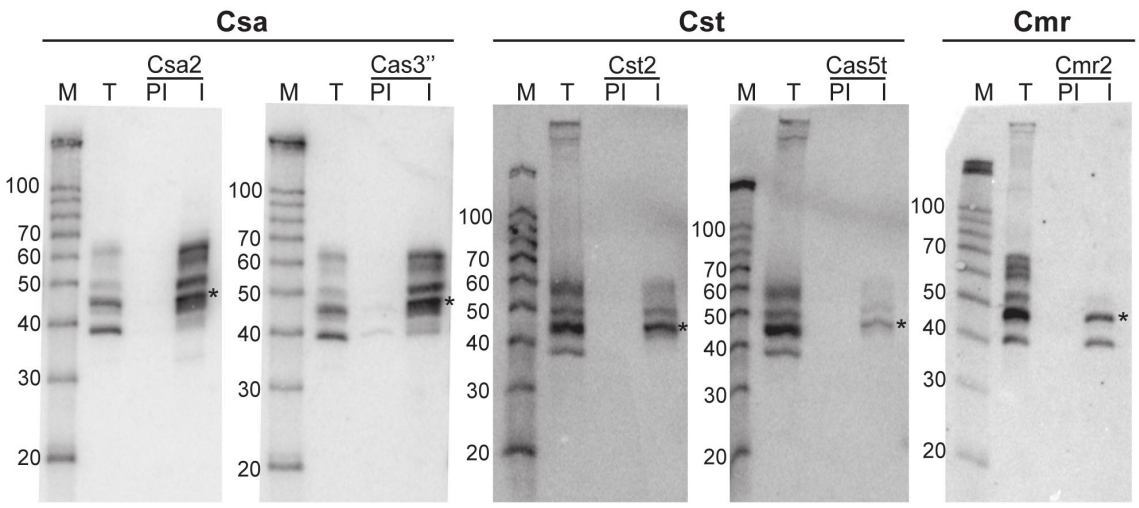

Csa
Cst
Cmr
Csa2
Cas3”
Cst2
Cas5t
Cmr2
M T PI I
100
70
60
50
40
30
20
*

Figure 2.5. RNA-seq profiles of crRNAs associated with the Csa, Cst and Cmr complexes and in total RNA. The number of reads (in thousands) of crRNAs-associated with each complex (extracted from immune pellets of Cas3", Csa2, Cas5t, Cst2, Cmr2) compared crRNAs in total RNA (extracted from untreated S100 extract) that map to CRISPR locus 5 are shown. A higher resolution image of three crRNAs is highlighted in the box on right. Reads mapping to both the sense (red, transcribed from upstream leader) and antisense (blue) strand are indicated. The black boxes at the top correspond to CRISPR repeats interspersed by guide sequences shown as black lines.

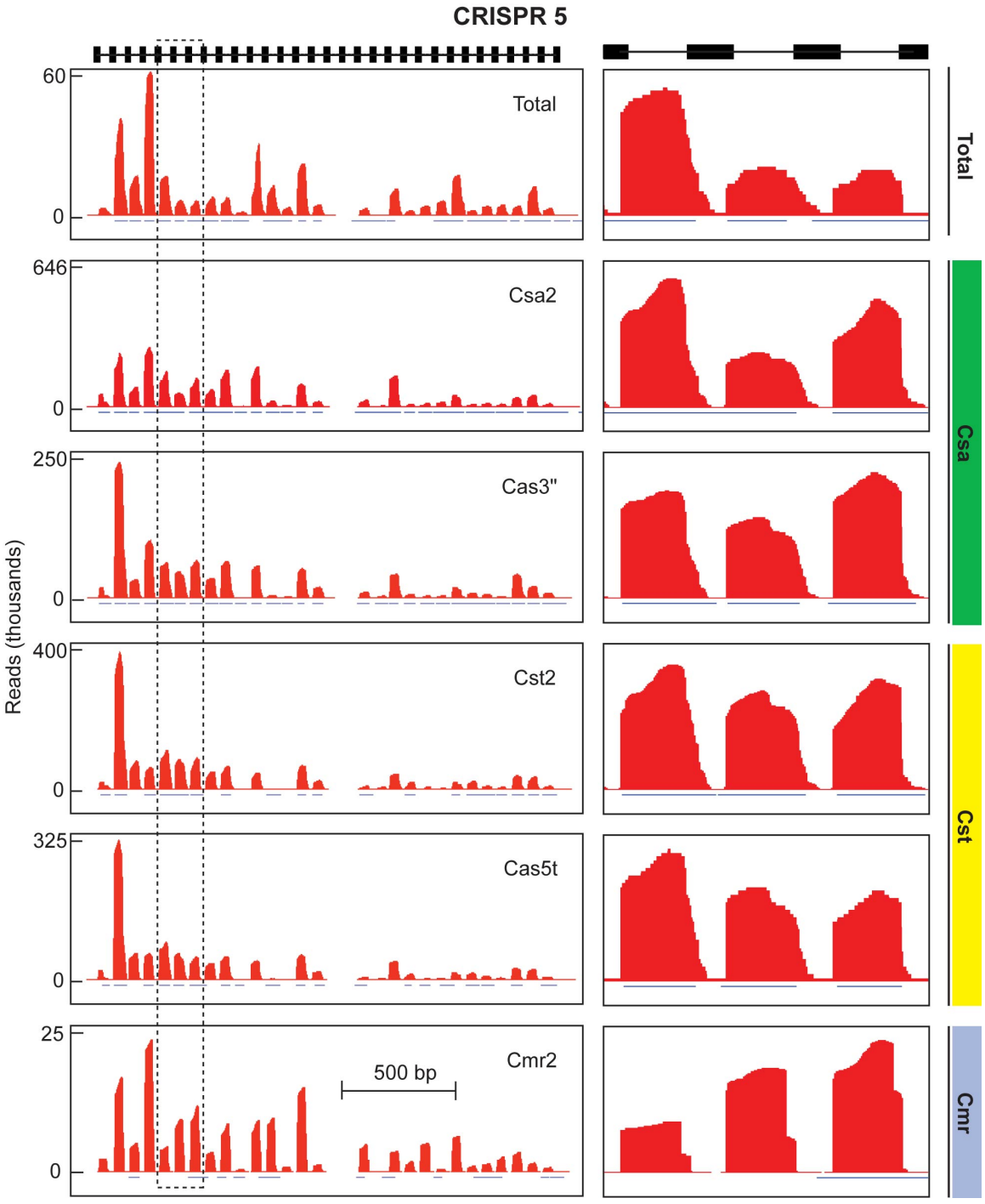

CRISPR 5
Reads (thousands)
60
0
Total
646
0
Csa2
250
0
Cas3"
400
0
Cst2
325
0
Cas5t
25
0
Cmr2
500 bp
Total
Csa
Cst
Cmr

Figure 2.6. Mapping ends of crRNAs in total RNA and crRNAs associated with Csa, Cst, Cmr complexes. 5' and 3' ends analysis of all sequenced crRNAs in untreated S100 extract (total RNA) and crRNAs associated with Csa (Csa2, Cas3"), Cst (Cst2, Cas5t) and Cmr (Cmr2) complexes. The percentages of crRNAs starting and ending at a specific point are marked on the y-axis. The nucleotides are counted relative to their position from guide-repeat junction. The nucleotides upstream of the guide-repeat junction are counted from -1 and those downstream are counted from +1. The peaks in 5' repeat region are shown in blue. The peaks downstream of 5' repeat –guide junction are shown in orange. The schematic on top indicates where the peaks map (guide- green and repeats -blue)

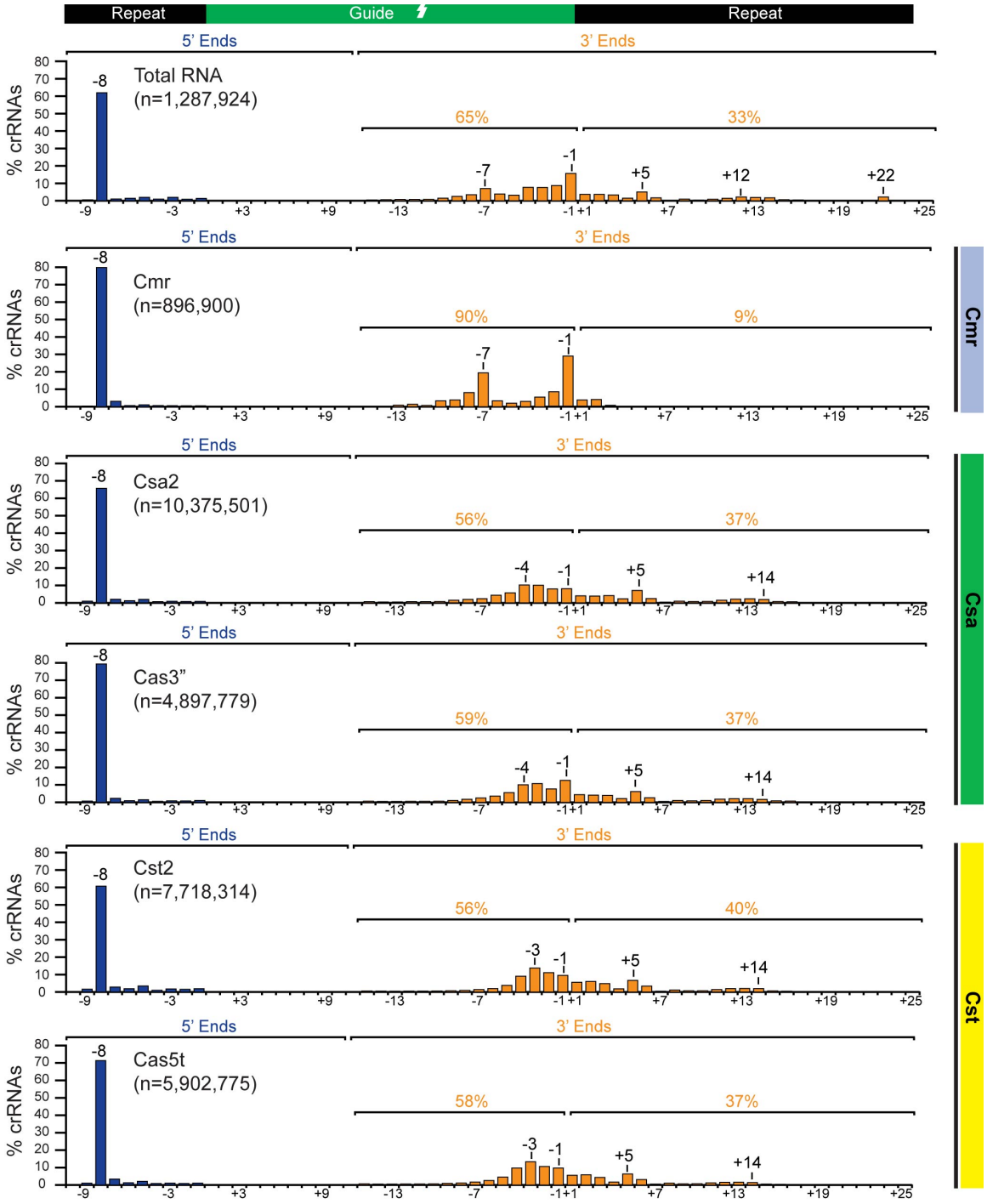

Repeat
Guide
Repeat
5' Ends
3' Ends
% crRNAs
Total RNA
(n=1,287,924)
65%
33%
Cmr
(n=896,900)
90%
9%
Csa2
(n=10,375,501)
56%
37%
Cas3"
(n=4,897,779)
59%
37%
Cst2
(n=7,718,314)
56%
40%
Cas5t
(n=5,902,775)
58%
37%
Cmr
Csa
Cst

Figure 2.7. Model for integration of crRNAs into three distinct crRNP effector complexes. The CRISPR locus (repeat: black rectangles, invader-derived guide sequences: multiple colors) is transcribed from promoters in the leader (brown) to form the pre-crRNA and following processing by Cas6, the 1X intermediate is generated starting at the 5' tag. The 1X intermediate is sorted and loaded onto separate effector complexes (crRNPs) following 3' end processing in the repeat region. Csa (green) and Cst (yellow) complexes harbor crRNAs that contain variable amounts of 3' repeats and one crRNA species without any 3' repeat. Cmr complex associates with two forms of crRNAs lacking any 3' repeat. Each crRNA size-form can assemble into the corresponding effector complexes and guide targeting. The Cmr complex cleaves complementary RNA target. The Csa and Cst effector complexes are predicted to target DNA homologous to the guide sequences in the crRNAs.

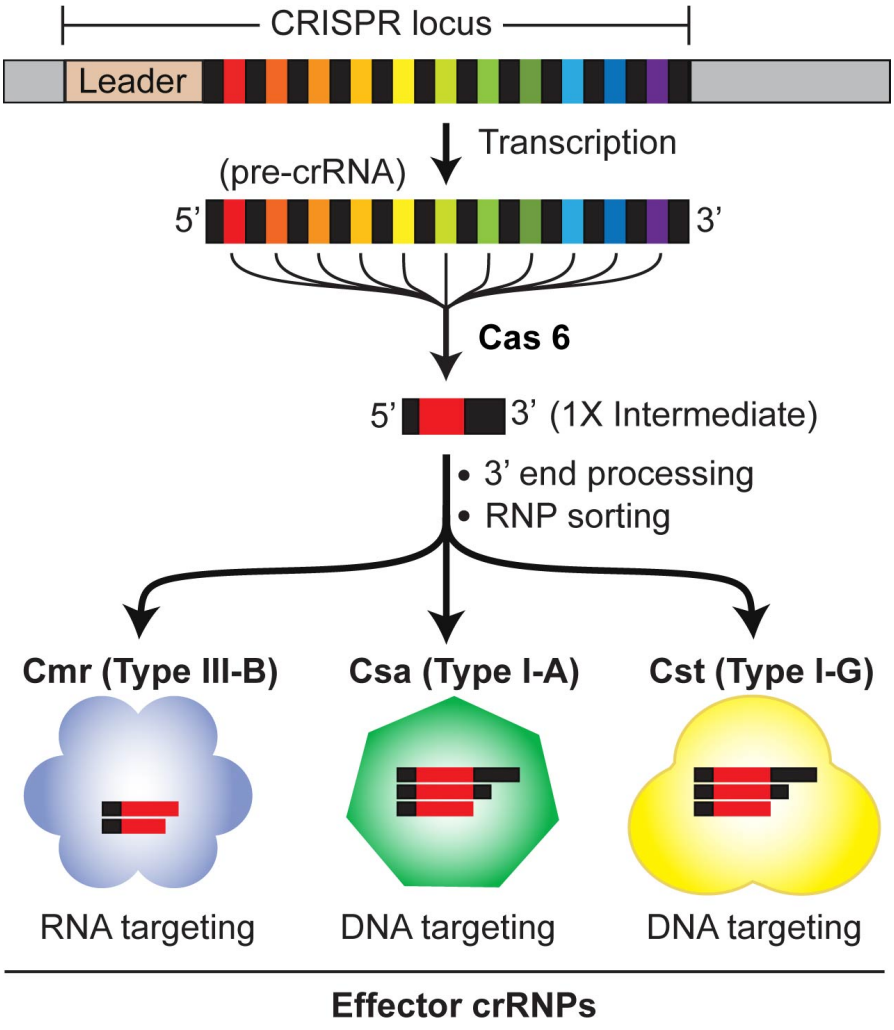
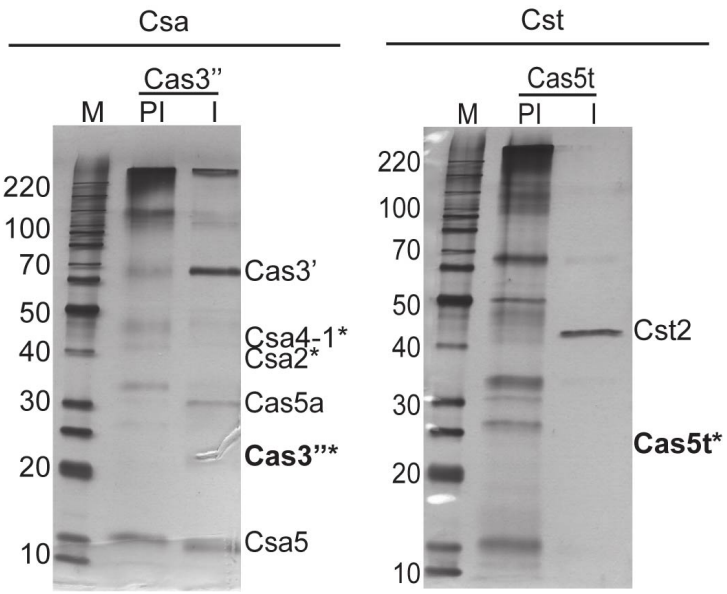

CRISPR locus
Leader
Transcription
(pre-crRNA)
5'
3'
Cas 6
5'
3' (1X Intermediate)
• 3' end processing
• RNP sorting
Cmr (Type III-B)
Csa (Type I-A)
Cst (Type I-G)
RNA targeting
DNA targeting
DNA targeting
Effector crRNPs

Table 2.1. Protein assignment in Csa, Cst, Cmr immunoprecipitated complexes by tandem mass spectrometry. Percent sequence coverage and unique peptides of Cas and non-Cas proteins identified in Csa (Csa2, Cas3"), Cst (Cst2, Cas5t) and Cmr (Cmr2) immunoprecipitations are specified. The data was generated from either in-solution or in-gel trypsin digested samples from 2-5 replicates. Proteins listed were subjected to pre-immune background and are specifically identified in immune pellets. Proteins identified in one replicate are indicated by asterisk.

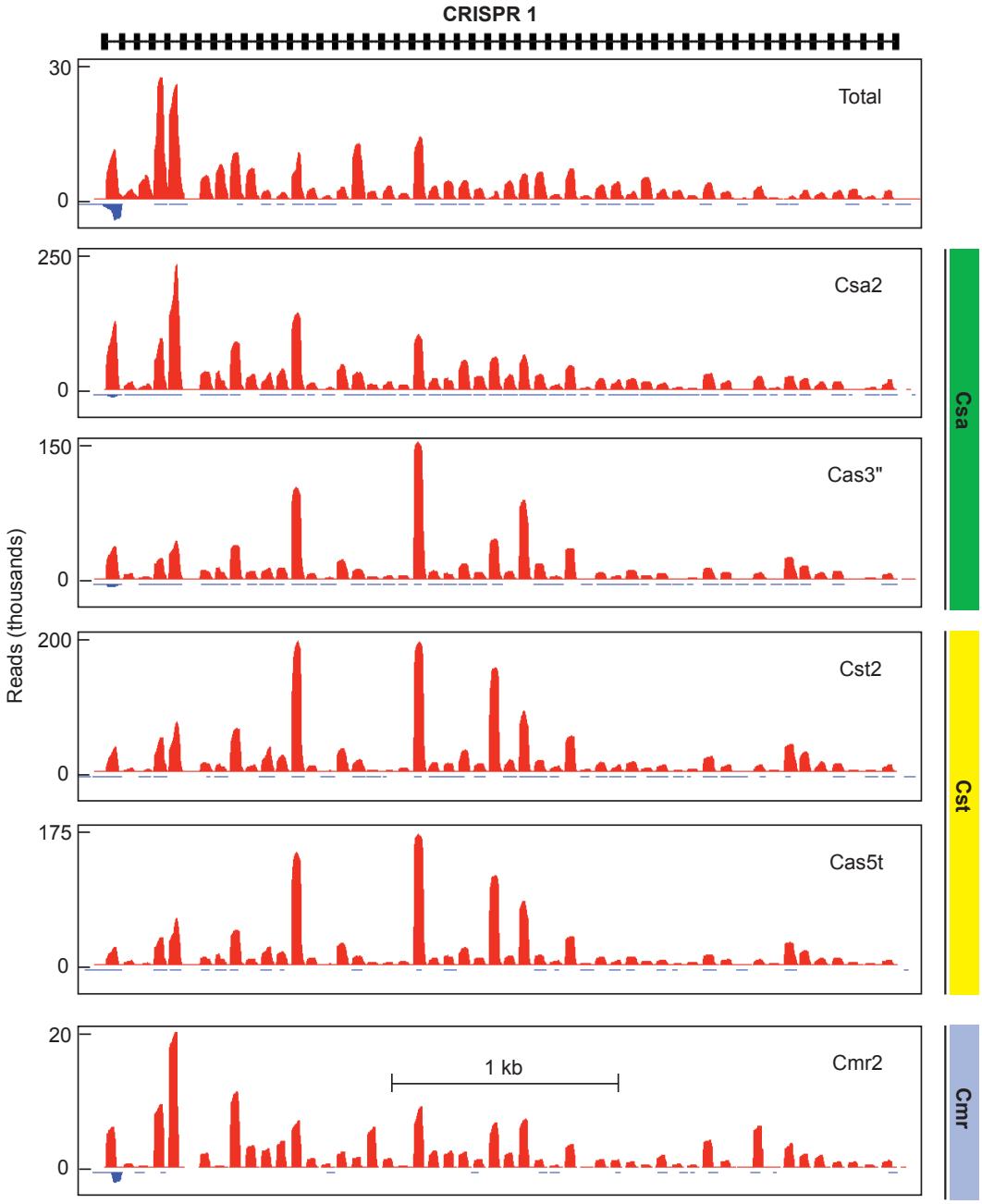
			Csa				Cst				Cmr		
	Protein	Mol. Wt.	Csa2 IP		Cas3" IP		Cst2 IP		Cas5t IP		Cmr2 IP		Annotation
			Unique Peptides	%SeqCov	Unique Peptides	%SeqCov	Unique Peptides	%SeqCov	Unique Peptides	%SeqCov	Unique Peptides	%SeqCov	
Csa	PF0637	39.6	4*	12	5*	13							Csa4-1 (Cas8a2)
	PF0638	44.3	7	25									Csa4-2 (Cas8a2)
	PF0639	22.9	7	32	13	59							Cas3" (nuclease)
	PF0640	75.2	19	35	15	24							Cas3' (helicase)
	PF0641	29.8	21	84	6	21							Cas5a (Cas5)
	PF0642	37.4	22	95	12	48							Csa2 (Cas7)
	PF0643	13.2	3	48	2*	20							Csa5 (Cas11)
Cst	PF1121	23.8					3*	9	7	53			Cas5t (Cas5)
	PF1122	29.7					8	17	15	42			Cst2 (Cas7)
Cmr	PF1124	38.7									11	26	Cmr6 (Cas7)
	PF1125	19.7									7	30	Cmr5 (Cas11)
	PF1126	32.6									15	67	Cmr4 (Cas7)
	PF1128	36.3									16	57	Cmr3 (Cas5)
	PF1129	100.8									38	39	Cmr2 (Cas10)
	PF1130	34.9									6	17	Cmr1-1 (Cas7)
	PF0352	21.7									4	25	Cmr1-2 (Cas7)
Non-Cas proteins	PF0440	200.8	5*	5			8	6	6*	5			Ribonucleotide diphosphate reductase subunit A (nr)
	PF0566	127.6	7	4							4*	3	Putative uncharacterized protein PF0566
	PF0924	16.7			2*	26							Putative uncharacterized protein PF0924
	PF0994	131.0	15	19							10	12	Type II secretion system proteins
	PF1142	31.4	12	53	10	42	2*	6					Putative uncharacterized protein PF1142
	PF1351	73.2					4*	13					Threonyl-tRNA synthetase
	PF1462	71.2	2*	5									Glutamyl-tRNA(Gln) amidotransferase subunit E (gatE)
	PF1563	103.0	4*	6	13	19					5*	5	DNA-directed RNA polymerase subunit A
	PF1882	89.1									6	6	Cell division protein 48
	PF1974	59.9	3*	8	25	64							Thermosome, single subunit

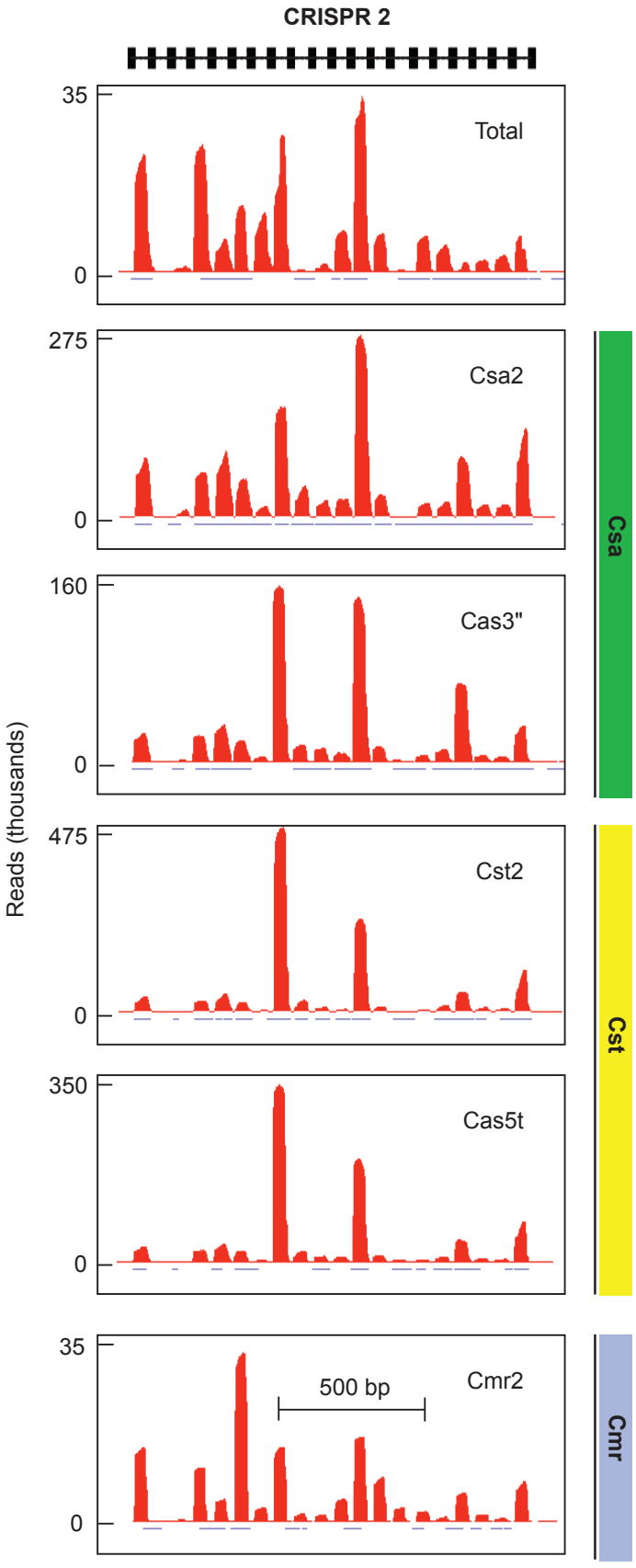
Makarova et al. Supplemental Table 1

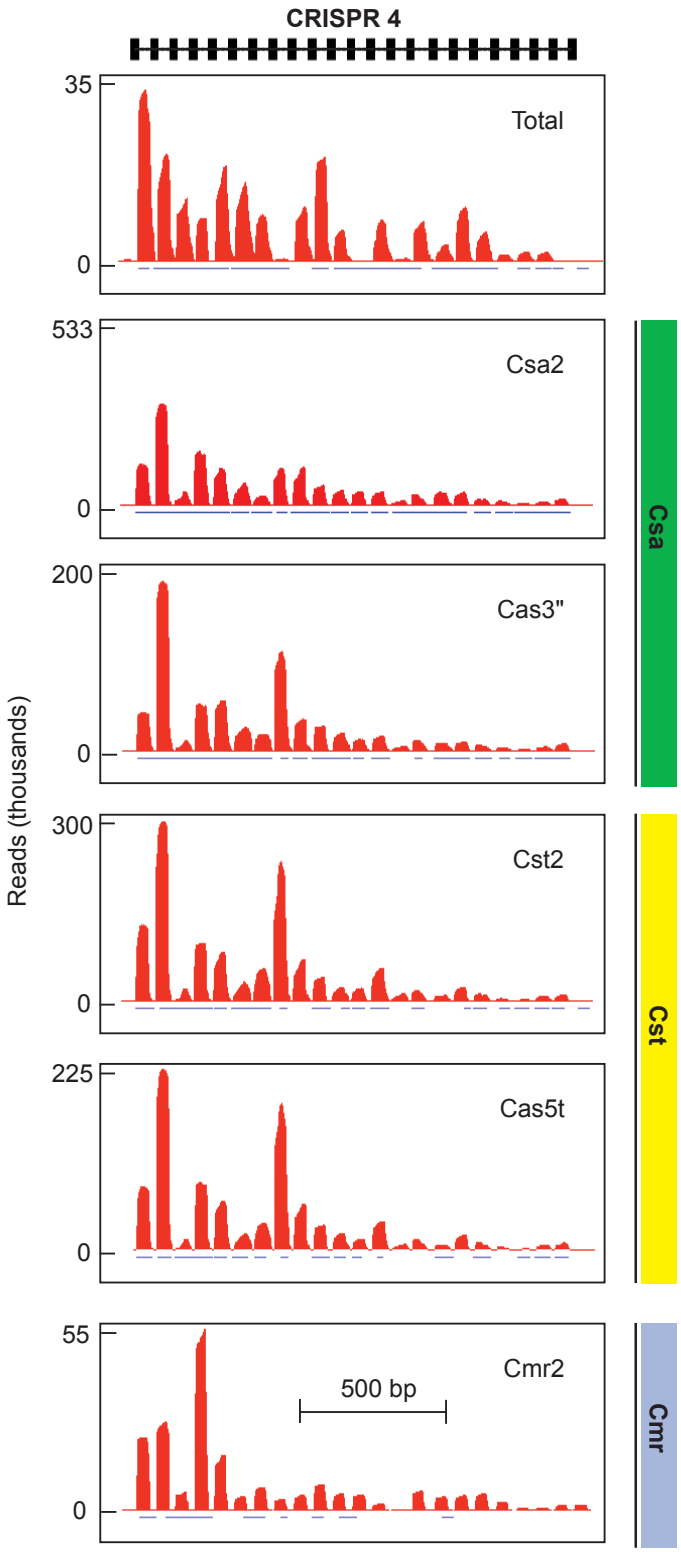
Figure 2.S1. Cas protein composition of immunopurified complexes. Silver stained SDS-polyacrylamide gels of Csa (Cas3") and Cst (Cas5t) immunopurified pellets with pre-immune (PI) or immune (I) antibodies as labeled. Cas proteins identified by tandem mass spectrometry are indicated based on expected molecular weight. The antigens are labeled in bold. Cas proteins identified by mass spectrometry that do not correlate to a visible band are marked with an asterisk. M is the protein standard in kilodaltons (kDa).

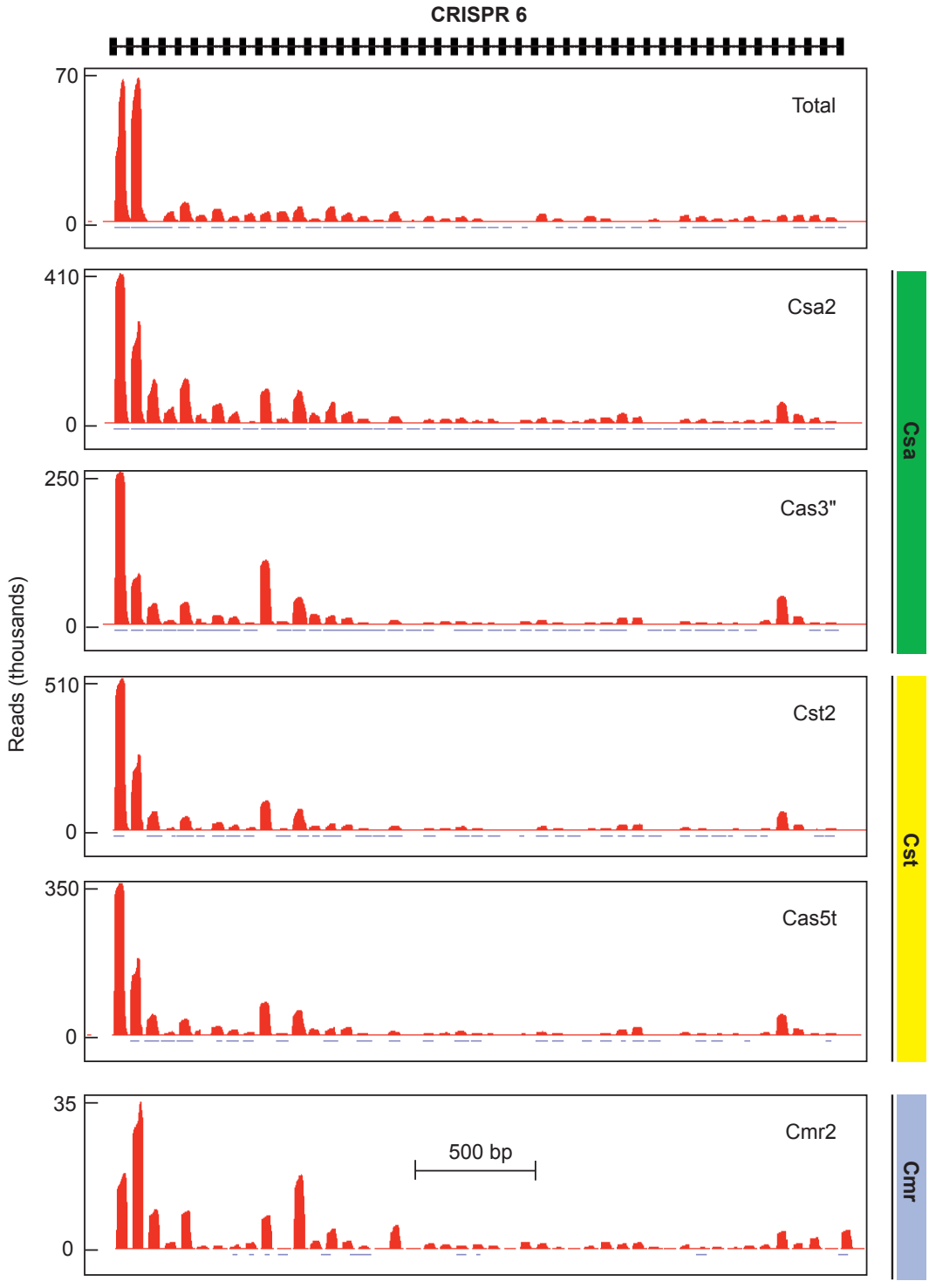

Csa
Cst
Cas3”
Cas5t
M
PI
I
M
PI
I
220
100
70
50
40
30
20
10
220
100
70
50
40
30
20
10
Cas3’
Csa4-1*
Csa2*
Cas5a
Cas3”*
Csa5
Cst2
Cas5t*

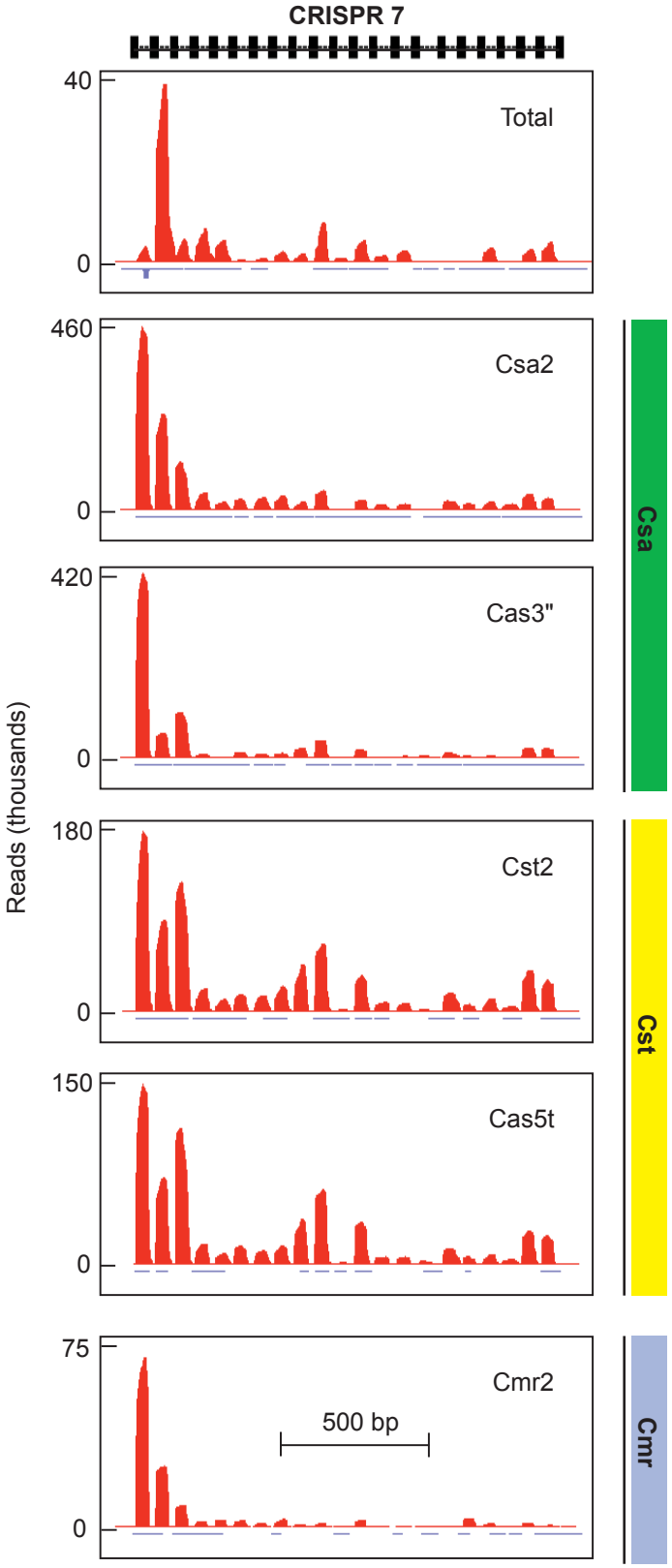
Figures 2.S2-2.S7. RNA-seq profiles of crRNAs associated with the Csa, Cst and Cmr complexes and in total RNA. The number of reads (in thousands) of crRNAs-associated with each complex (extracted from immune pellets of Cas3", Csa2, Cas5t, Cst2, Cmr2) compared crRNAs in total RNA (extracted from untreated S100 extract) that map to CRISPR locus 1 **(S2)**, CRISPR locus 2 **(S3)**, CRISPR locus 4 **(S4)**, CRISPR locus 6 **(S5)**, CRISPR 7 **(S6)** and CRISPR 8 **(S7)** are shown. Reads mapping to both the sense (red, transcribed from upstream leader) and antisense (blue) strand are indicated. The black boxes at the top correspond to CRISPR repeats interspersed by guide sequences shown as black lines.

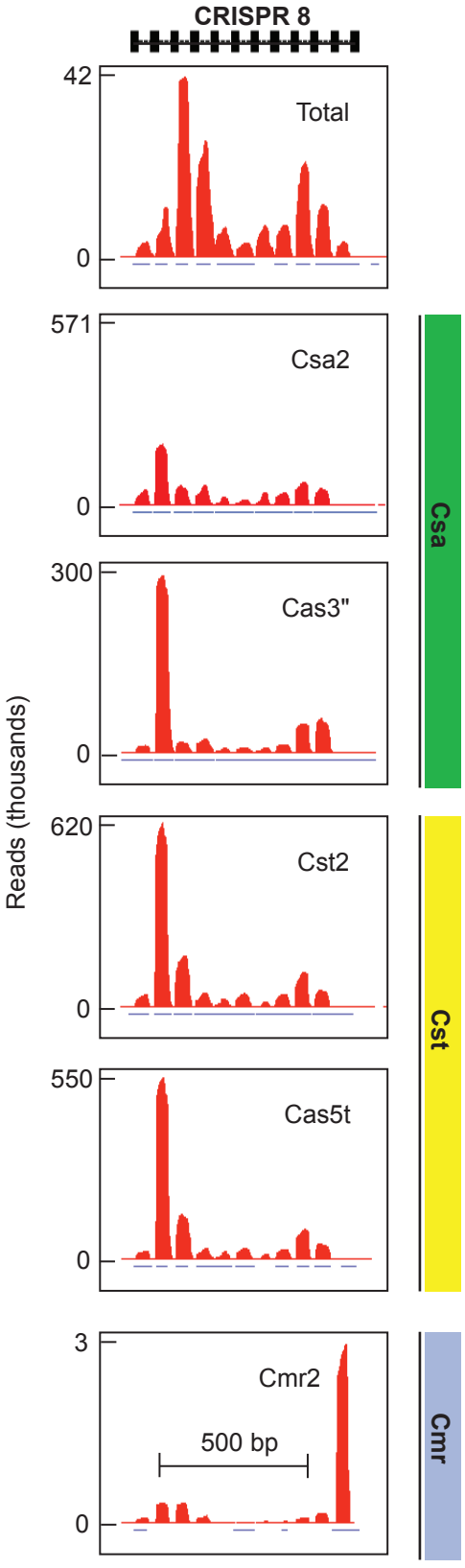

CRISPR 1
Reads (thousands)
30
0
Total
250
0
Csa2
150
0
Cas3"
Csa
200
0
Cst2
175
0
Cas5t
Cst
20
0
1 kb
Cmr2
Cmr


CRISPR 2
35
Total
0
275
Csa2
0
160
Cas3"
0
Reads (thousands)
475
Cst2
0
350
Cas5t
0
35
500 bp
Cmr2
0
Csa
Cst
Cmr


CRISPR 4
35
Total
0
533
Csa2
0
200
Cas3"
0
Reads (thousands)
Csa
300
Cst2
0
225
Cas5t
0
Cst
55
Cmr2
500 bp
0
Cmr


CRISPR 6
Reads (thousands)
70
0
Total
410
0
Csa2
250
0
Cas3"
Csa
510
0
Cst2
350
0
Cas5t
Cst
35
0
Cmr2
500 bp
Cmr


CRISPR 7
40
Total
0
460
Csa2
0
420
Cas3"
0
Csa
Reads (thousands)
180
Cst2
0
150
Cas5t
0
Cst
75
Cmr2
500 bp
0
Cmr


CRISPR 8
42
Total
0
571
Csa2
0
300
Cas3"
0
Csa
Reads (thousands)
620
Cst2
0
550
Cas5t
0
Cst
3
Cmr2
500 bp
0
Cmr

CHAPTER 3

CRISPR RNA-GUIDED DNA CLEAVAGE BY TYPE I-A CRISPR-CAS IMMUNE EFFECTOR COMPLEX IN *PYROCOCCUS FURIOSUS*

---

[1]Sonali Majumdar, Rebecca M. Terns, Michael P. Terns. To be submitted to *RNA*.

Abstract

CRISPR-Cas are RNA-based immune systems (Type I, II, III) that mediate silencing of invader nucleic acids through structurally and functionally diverse effector complexes. CRISPR RNAs (crRNAs) containing invader-derived sequences, associate with different Cas protein subtypes to form effector complexes and guide targeting. In *Pyrococcus furiosus* we have previously identified three compositionally distinct, native effector complexes by immunoprecipitation: Type I-A (Csa), Type I-G (Cst) and Type III-B (Cmr). We have shown that the Type III-B effector complex mediates crRNA-guided cleavage of target RNAs. In the current study, we have reconstituted an active Type I-A (Csa) complex from recombinant Cas proteins and synthetic crRNA that targets and cleaves double-stranded DNA in a crRNA-guided, PAM dependent manner. Our results show six Csa proteins (Csa 4-1 (Cas8a2), Cas3", Cas3', Cas5a, Csa2 (Cas7), Csa5) are each essential components of the active complex. Mutational analysis shows Cas3" (HD) is the effector nuclease of the complex. Cleavage profiles show multiple cleavages on target and non-target strands of DNA. Mapping of cleavage sites suggest that cleavages occur in the protospacer of non-target strand due to crRNA-induced, strand-displacement followed by ATP-dependent unwinding and cleavage on both strands in 3' to 5' direction. Finally, we delineate some features of the crRNA-Csa complex architecture. Native gel shift analysis and UV-induced RNA-protein crosslinking reveal Cas5a (Cas5) and Csa2 (Cas7) constitute the structural core making direct contacts with crRNA. Larger complex (holoenzyme) with six Csa proteins is seen in gel shift analysis. Cas6 generated 5'-OH and 8-nt repeat-derived 5' tag of crRNA are critical for assembly of Csa complex.

Introduction

CRISPR (Clustered Regularly-Interspaced Short Palindromic Repeats)-Cas (CRIPSR-associated) systems are RNA-based adaptive immune systems that defend bacteria and archaea against viruses, plasmids and other genome invaders. A CRISPR locus of the host genome typically contains arrays of direct repeat sequences (30-40 base pairs), interspersed by similar-sized, variable, "invader-derived" sequences (called protospacers). Precursor CRISPR RNAs are transcribed from CRISPR locus and processed by Cas endonucleases to generate mature CRISPR RNAs (crRNAs). Mature crRNAs are comprised of invader-derived (guide) sequences flanked by 5' and/ or 3' repeat-derived sequences (known as 5' tag and 3' tag). Mature crRNAs associate with Cas proteins to form effector complexes (crRNPs) and guide silencing of invader DNA or RNA in a sequence-specific manner (Reviewed in (Makarova et al. 2011c; Terns and Terns 2011a; Westra et al. 2012b; Wiedenheft et al. 2012; Gasiunas et al. 2014; van der Oost et al. 2014)). CRISPR-Cas systems have been grouped into Type I (A-G), Type II (A-C) and Type III (A-B or A-D) (Haft et al. 2005a; Makarova et al. 2011c; Vestergaard et al. 2014) primarily based on their Cas protein composition. Each group specifically relates to a Cas-subtype (for example, Type I-A (Csa), Type I-E (Cse), Type III-B (Cmr)). All characterized Type I, Type II and Type III crRNPs target invader DNA, whereas Types III-B (Cmr) crRNPs typically target RNAs (Barrangou et al. 2007; Brouns et al. 2008; Marraffini and Sontheimer 2008; Hale et al. 2009; Garneau et al. 2010; Garrett et al. 2011a; Manica et al. 2011; Fischer et al. 2012b; Gasiunas et al. 2012b; Jinek et al. 2012; Magadan et al. 2012; van Duijn et al. 2012a; Zhang et al. 2012c; Deng et al.

2013a; Elmore et al. 2013a; Maier et al. 2013b; Mulepati and Bailey 2013b; Sinkunas et al. 2013a; Staals et al. 2013a; Plagens et al. 2014b; Zebec et al. 2014a).

Characterized Type I systems possess similarities in Cas protein composition, crRNP structure as well as functional parallels (Makarova et al. 2011c). Cas5, Cas7, large subunit (Cas8) and small subunit superfamily proteins assemble with crRNA to form the structure of crRNP in studied Types I-A, I-C, I-E, I-F (Jore et al. 2011a; Lintner et al. 2011d; Wiedenheft et al. 2011a; Nam et al. 2012b; Plagens et al. 2012a; van Duijn et al. 2012a; Brendel et al. 2014; Jackson et al. 2014a; Mulepati et al. 2014a; Plagens et al. 2014b; Vestergaard et al. 2014; Zhao et al. 2014) (Majumdar et al. unpublished). Cas3 superfamily signature proteins of Type I systems contain fused DExH helicase and HD nuclease domains and are the effector nucleases that cleave invader DNA for silencing (Brouns et al. 2008; Cady and O'Toole 2011b; van Duijn et al. 2012a; Mulepati and Bailey 2013b; Sinkunas et al. 2013a; Huo et al. 2014a; Plagens et al. 2014b). Type I-A systems are unique and contain two Cas3 proteins, Cas 3" (HD nuclease), Cas 3' (DExH helicase) (Haft et al. 2005a; Makarova et al. 2011c). Cas3" catalyzes DNA cleavage in *T. tenax* Type I-A system (Plagens et al. 2014b).

The hyperthermophilic euryarchaeon, *Pyrococcus furiosus* has seven CRISPR loci encoding ~200 crRNAs and twenty-nine *cas* genes primarily arranged in two gene clusters. These genes encode Cas proteins for protospacer acquisition (Cas1, Cas2, Cas4), crRNA biogenesis (Cas6) and three immune-effector systems (Type I-A (Csa), Type I-G (Cst) and Type III-B (Cmr)) (Figure 3.1A). We have previously identified three distinct native *P. furiosus* crRNPs (Csa (Type I-A), Cst (Type I-G) and Cmr (Type III-B)) (Hale et al. 2012a) (Majumdar et al. unpublished). All three complexes associate with mature

crRNAs from all seven CRISPR loci. We have elucidated structural organization of Cmr crRNP (Spilman et al. 2013a) and investigated molecular mechanism of crRNA-guided and Type III-B (Cmr) mediated cleavage of target RNA (Hale et al. 2009; Hale et al. 2012a). Based on parallels with other Type I systems and genetic evidence of DNA targeting by Type I-A and I-G systems in closely related *T. kodakarensis (Elmore et al. 2013a)*, we hypothesized Csa and Cst crRNPs are independent and co-functioning effector complexes targeting invader DNA in *P. furiosus*, providing multiple lines (RNA and DNA) of CRISPR-Cas mediated invader defense.

In this study, we have reconstituted active Type I-A complex with recombinant Csa proteins and synthetic mature crRNA. The complex binds and cleaves non-self double-stranded DNA in crRNA-guided manner. We observe that six Csa proteins (Cas8a2, Cas3", Cas3', Cas5a, Cas7 and small subunit protein Csa5) are essential and minimal protein components of Csa complex for target DNA binding and cleavage. Consistent with findings in other Type I systems, our results indicate Cas3" (HD nuclease) is the effector nuclease that cleaves target DNA. Based on analysis of cleavage profiles, we propose a mechanistic model of Csa-mediated DNA cleavage. Finally, we have delineated some key molecular features of Csa crRNP assembly.

Results

Reconstituted Type I-A (Csa) complex cleaves double-stranded DNA targets *in P. furiosus*.

Type I effector complexes (crRNPs) target invaders via base pairing of crRNA guide with cognate protospacer sequences of invader DNA. Moreover, in order to disciriminate

self vs non-self DNA, Type I systems recognize a signature motif called Protospacer Adjacent Motif or PAM (present upstream or downstream of protospacer) on invader DNA (Deveau et al. 2008; Mojica et al. 2009; Fischer et al. 2012b; Gasiunas et al. 2012b; Jinek et al. 2012; Sashital et al. 2012b; Mulepati and Bailey 2013b; Sinkunas et al. 2013a; Westra et al. 2013; Hochstrasser et al. 2014a; Nishimasu et al. 2014). To investigate the function of Type-I-A system in *P. furiosus*, we reconstituted the complex with recombinant Csa proteins and a crRNA identified in native Csa complex (Majumdar et al., unpublished). The synthetic crRNA is 45-nt crRNA and is comprised of 8-nt, 5' tag (black, Figure 3.1B) and entire 7.01 guide sequence (37-nt, first invader-derived sequence of locus 7) (red, Figure 3.1B). *P. furiosus csa* gene cluster encodes eight proteins namely Csa4-1 and Csa4-2 (both large subunit proteins of Cas8a2 superfamily), Cas3" (HD-nuclease), Cas3' (DExH helicase), Cas5a (Cas5), Csa2 (Cas7), Csa5 (small subunit) and Csa3 (putative transcriptional regulator (Lintner et al. 2011b; Makarova et al. 2014)) (Figures 3.1A, 3.S1A). To test activity of the reconstituted Csa complex, we used (117-nt) double-stranded DNA (Figure 3.1B) with sequences bearing homology to 7.01 protospacer (red) wild-type PAM (blue) and random flanking sequence (grey). Based on bioinformatics prediction of PAM-association with repeat classes and Cas subtypes (Kunin et al. 2007; Mojica et al. 2009) and *in vivo* studies of PAM-dependent, CRISPR-Cas mediated invader DNA silencing in related archaeon, *Thermococcus kodakarensis* (Elmore et al. 2013a), we predicted 5'-NGG-3' (target strand) as the functional consensus PAM for Type I-A system in *P. furiosus*. The reconstituted Csa crRNP cleaved the double-stranded DNA substrate with cleavages seen in both target strand (complementary to crRNA guide) and non-target strand (Figure 3.1B).

Minimal components of Type I-A effector complex essential for DNA targeting.

To determine the minimal protein components of Type I-A (Csa) effector complex essential for DNA-cleavage, we excluded one Csa protein at a time from the active, reconstituted Csa crRNP. Csa4-2 and Csa3 were dispensable for DNA cleavage activity (Figure 3.S1C). Excluding any one of the six Csa proteins (Csa 4-1, Cas3", Cas3' Cas5a, Csa2, Csa5) abolished DNA cleavage in both non-target and target strand (Figure 3.2A). In the absence of crRNA, all six Csa proteins did not cleave DNA. Therefore, mature crRNA and six Csa proteins (Csa 4-1, Cas3", Cas3' Cas5a, Csa2, Csa5) are the essential components of the catalytically active, DNA-cleaving Type I-A effector complex in *P. furiosus*. All assays in the paper (except Figure 3.S1C) are done with the active complex reconstituted with six Csa proteins and 45-nt, 7.01 crRNA.

To further investigate the minimal components of Csa crRNP essential for DNA recognition and binding, we performed native gel shift assays with all six Csa proteins and observed the effect of excluding a single protein on DNA binding (Figure 3.2B). DNA binding (with either strand labeled) was observed with six Csa proteins that form the catalytically active complex (Csa4-1, Cas3", Cas3', Cas5a, Csa2, Csa5) and was either greatly reduced (without Cas3"or Cas3') or completely lost (without Csa4-1, Cas5a, Csa2 and Csa5) when one of six Csa proteins, was missing. Csa complex did not bind DNA in the absence of crRNA. Therefore, the minimal Csa crRNP essential for cleavage is also critical for DNA recognition and binding.

Cas3" is the effector nuclease of active Type I-A system in *P. furiosus*.

Unlike other characterized Type I systems where Cas3 has distinct nuclease and helicase domains in the same polypeptide, Type I-A has two Cas3 proteins. Cas3" contains the predicted nuclease (HD) domain and Cas3' contains the (DExH) helicase domain. Based on sequence homology with other HD domain-containing phopshohydrolases and biochemical properties of *M. janaschii* and *P. furiosus* Cas3" (Beloglazova et al. 2011), we predicted histidine and aspartate (HD) amino acids of Cas3" (H60, D61) are the active site residues catalyzing Csa crRNP-mediated DNA cleavage. We performed site-directed mutagenesis (HD-AA) of Cas3" and reconstituted Csa complex with either wild type or mutant Cas3" (Figure 3.3A). Csa crRNP only cleaved DNA in the presence of wild type Cas3", but not mutant Cas3" (Figure 3.3B). Csa crRNP reconstituted with either wild-type or mutated Cas3" bound DNA with similar efficiency (Figure 3.3C). So the lack of DNA cleavage can't be attributed to loss of DNA binding, but to catalytic inactivity. These results indicate that Cas3" is the effector nuclease of the active Type I-A system in *P. furiosus*.

Active Type I-A system recognizes and targets DNA in a crRNA-guided and PAM-dependent fashion.

To determine the elements of target DNA essential for Type I-A mediated DNA recognition and targeting (by binding and cleavage), we mutated components (protospacer or predicted PAM) of the DNA substrates (Figure 3.4A). We mutated wild-type PAM (5'-GGG-3' on target strand; wt. PAM, blue) to a mutated variant (5'-CCC-3' on target strand; mut. PAM, green) and made compensatory mutations on non-target

strand to maintain complementarity (Figure 3.4A). The reconstituted Csa crRNP cleaved DNA with wild-type PAM (wt. PAM) but not mutated PAM (mut. PAM) (Figure 3.4B). Similar to DNA cleavage, there is a specific PAM requirement for DNA binding. The Csa crRNP did not bind DNA when the PAM was mutated (wt. PAM vs mut. PAM, Figure 3.4C). Moreover, Type III systems recognize non-self DNA by lack of complementarity with 5'-tag of crRNA with DNA targets (Marraffini and Sontheimer 2010b). We tested this paradigm by mutating the PAM and adjacent nucleotides spanning the PAM, such that they bear complementarity to the 8-nt repeat-derived 5' tag of crRNA (tc. PAM, Figure 3.4B). Type I-A mediated DNA cleavage (tc. PAM, Figure 3.4B) and DNA binding (tc. PAM, Figure 3.4C) were both lost when the 5' tag of crRNA was complementary to corresponding DNA sequences on target strand. We also mutated the protospacer sequence (orange, mut. protospacer, Figure 3.4A) thereby disrupting the ability of 7.01 crRNA guide sequence to base-pair with target strand of the DNA. Type I-A mediated DNA cleavage (Figure 3.4B) and DNA binding (Figure 3.4C) was abolished when the protospacer of DNA was non-cognate (mut. protospacer). Therefore, Type I-A (Csa) complex-mediated DNA recognition and targeting is crRNA-guided and exhibits PAM-specificity.

We further tested whether RNA (single-stranded and double-stranded) could serve as targets for Csa complex (Figure 3.S2A). Csa crRNP in *P. furiosus* appears to be specific to DNA and does not cleave single-stranded or double-stranded RNA targets.

***P. furiosus* Type I-A system recognizes double-stranded PAM for DNA targeting.**

In order to probe whether the PAM (G-C, wild type, Figure 3.5A) recognized by Type I-A complex for DNA targeting was a single-stranded or double-stranded entity, we made DNAs with hybrid PAMs (G-G PAM, C-C PAM, Figure 3.5A). We mutated one strand of the PAM (non-target strand in G-G, target strand in C-C, colored green) and did not make compensatory mutations in the other strand leaving a region of non-complementarity in the PAM region. Csa complex bound DNA with single-stranded PAM on target strand (G-G) with reduced efficiency compared to double-stranded, wild-type PAM (GC, wt. PAM) (Figure 3.5B). While the active Csa complex cleaved DNA with double-stranded PAM (GC, wt. PAM), reduced binding of DNA with single-stranded, target strand PAM (G-G) did not result in detectable cleavage (Figure 3.4C). DNA with single-stranded PAM on the non-target strand (C-C) was neither bound nor cleaved by Csa complex. (Figure 3.4B, 3.4C). This data suggests that Type I-A (Csa) complex may recognize PAM sequences on target strand and bind with reduced efficiency, but requires double-strandedness at PAM region for efficient DNA binding and cleavage.

Csa crRNP cleaves within protospacer of non-target followed by ATP-dependent multiple cleavages in 3' to 5' direction on both strands.

During silencing, crRNA (guide sequences) of effector complexes base-pair with target strand of duplex DNA (in the protospacer region) leading to formation of R-loop and displacement of non-target strand. For cleavage of long DNA targets and plasmids, Type I CRISPR-Cas effector complexes are predicted to progressively unwind DNA

catalyzed by Cas3 helicase using ATP as energy source (Sinkunas et al. 2011; Mulepati and Bailey 2013b; Sinkunas et al. 2013a). We tested the requirement of ATP as energy source by observing the effect on DNA cleavage in the absence of ATP (Figure 3.6A). Without ATP, no cleavages were observed in the target strand of DNA, whereas the non-target strand cleavage was restricted to protospacer, not flanking sequences (cleavages and mapped sites marked with asterisk in Figures 3.6A, 3.6C). Csa complex bound DNA with similar efficiency with or without ATP (Figure 3.S3). So partial lack of cleavage in the absence of ATP was not a consequence of reduced DNA binding.

We mapped the cleavage products with two different-sized double-stranded DNA substrates (117 and 67 nucleotides) (Figure 3.6B, 3.6C). For both substrates, cleavages on the non-target strand map approximately to the middle of protospacer (red, Figure 3.6C). On the target strand (complementary to crRNA guide) the 3'-end proximal cleavages map 0-5 sequences upstream of PAM (GGG, blue, Figure 3.6C). The nucleotides downstream of PAM on target strand are not cleaved. On both target and non-target strands profile of cleavage product and its mapping suggest cleavage occurs in 3' to 5' direction.

Protein components and crRNA features essential for Csa complex formation.

Our findings indicate six Csa proteins are required for Csa crRNP mediated DNA cleavage. We wanted to investigate whether each of these proteins were stable components of active Csa complex. We did native gel shift assays to study the contribution of protein components in Csa crRNP/s formation (Figure 3.7A). Native gel shifts with Csa proteins and 3' radiolabeled 45-nt mature 7.01 crRNA (used in reconstitution assays) show crRNPs (indicated by larger shifts) are formed with the six

protein components of catalytically active Csa complex. Smaller-sized shifts were observed when either Csa 4-1, Cas3", Cas3', Csa5 were excluded from the six Csa proteins, indicating formation of subcomplex/es. crRNP shifts were not detected when either Cas5a (Cas5) or Csa2 (Cas7) were absent. We observed that while none of the proteins specifically bound crRNA individually, a minimal crRNP core is formed with Cas5a (Cas5) and Csa2 (Cas7) (Figure 3.S4A).

Cas6 endonuclease cleaves within the repeats of CRISPR RNA transcripts from all seven loci to generate IX intermediate crRNAs, beginning with 8-nt, 5' tag (AUUGAAAG). Cas6 is a metal-independent endonuclease, which cleaves on the 3' side of phosphodiester bond to leave 5'-OH and 2'-3' phosphate on the cleaved crRNA (Carte et al. 2008b; Carte et al. 2010). Similar to some Type I and Type III systems (Hale et al. 2008; Lillestol et al. 2009; Hatoum-Aslan et al. 2011b; Bernick et al. 2012; Hale et al. 2012a; Plagens et al. 2012a; Richter et al. 2012; Zhang et al. 2012c; Deng et al. 2013b), further processing of 3' repeats by unknown mechanisms generate mature crRNAs in *P. furiosus* (Hale et al. 2008) (Majumdar et al., unpublished). We examined the essential features of crRNA required for Csa crRNP formation. As crRNAs associated with native *P. furiosus* Csa complex bear Cas6- generated, repeat-derived 5' tag (Majumdar et. al, unpublished), we tested if this 8-nt, 5' tag is important for crRNP assembly. The core crRNP (with Cas5a and Csa2/Cas7) and the holoenzyme (crRNP with six Csa proteins) are formed only in the presence of wild-type 5' tag of crRNA but not when the 5' tag is mutated (Figure 3.7B). Moreover, we wanted to decipher what chemical ends of crRNA were amenable for Csa complex assembly. We either 5' end-labeled to generate 5'-P and 3'-OH bearing crRNAs, or 3' end-labeled to generate 3'-P and 5'-OH bearing crRNAs,

or internally-labeled to generate 5'-OH and 3'-OH bearing crRNAs. Only crRNAs with 5'OH were bound to form the core crRNP and holoenzyme (Figure 3.7C). Therefore, the 5' tag and 5'-OH of crRNA (generated by Cas6 cleavage) are critical features for Csa complex formation.

Cas5a interacts with 5'-tag and Csa2 (Cas7) binds guide region of crRNA.

We did UV-induced crosslinking of internally radiolabeled, 4-thiouridine containing crRNA and Csa proteins to decipher which proteins in Csa complex (core or holoenzyme) make direct contacts with crRNA (5' tag and guide) (Figure 3.8). To visualize interactions at specific crRNA regions (5' tag or guide), we selectively radiolabeled the desired region. Crosslinking data showed that Cas5a contacts the 5' tag region whereas Csa2 (Cas7) contacted the guide region of crRNA. Crosslinking patters at the guide region indicate association with large subunit protein Cas8a2 (Csa4-1). We did not observe any other Csa proteins interacting with the crRNA in our experimental conditions. As expected core crRNP and bigger crRNPs were observed by native gel shift with each of the substrates tested for crosslinking (Figure 3.S4B).

Discussion

DNA-targeting Type I-A system in *P. furiosus*.

In this paper, we have reconstituted an active Type I-A system in *P. furiosus* with recombinant Csa proteins and synthetic crRNA (containing 5'-tag and 7.01 guide). Guided by invader-derived sequence in crRNA, Type I-A (Csa) complex cleaves double-

stranded DNA targets with wild-type PAM (GGG on target strand). Our results indicate Cas3" is the effector nuclease of the complex.

Proposed mechanism of DNA cleavage by *P. furiosus* Type I-A complex.

Based on the findings of this paper, we propose a model of crRNA-guided Type I-A mediated DNA cleavage in *P. furisosus* (Figure 3.9). Six Csa proteins (large subunit (Cas8a2), Cas3', Cas3", Cas5a, Cas7 and small subunit (Csa5)) are essential components of the active Csa complex that recognizes and binds non-self DNA with a cognate protospacer and wild-type PAM. We expect that base pairing between crRNA guide and target strand of protospacer results in R-loop formation of DNA duplex and displacement of non-target strand. Based on cleavage mapping and ATP-dependence of certain cleavages (Figures 3.6A, 3.6B), we speculate Cas3" (nuclease) initiates ATP-independent cleavages in the protospacer region of displaced non-target strand mapping approximately up to the middle of the protospacer. Further strand unwinding, probably mediated by Cas3' (helicase), requires ATP and results in degradation of the non-target strand by Cas3" in a 3' to 5' direction. Non-target strand degradation leaves single-stranded regions in the target strand, which may cause disruption in crRNA-protospacer base pairing, fostered by ATP-dependent, Cas3' helicase-mediated unwinding. This triggers cleavage of displaced target strand in the vicinity of PAM sequence and progressing cleavages in 3' to 5' direction. *P. furiosus* Cas3" shows both endonuclease and exonuclease (3'-5') activity *in vitro* (Beloglazova et al. 2011). We cannot conclude if the Cas3" catalyzed cleavages in context of Csa complex are multiple endonucleolytic cuts or 3'-5' exonuclease degradation or a combination of both. The uncleaved DNA

sequences on both target and non-target strands are most likely protected by the bound Csa complex, shielding it from the active site.

Cas3" and Cas3' are structural and functional components of type I-A complex.

A unique feature of Type I-A systems is HD nuclease and DExH helicase domains are encoded in two Cas3 proteins, Cas3" and Cas3' respectively. Also, an interesting thing to note is Cas3" and Cas3' are seen as components of the native Csa complexes in Type I-A systems of *S. islandicus*, *T. tenax*, *P. furiosus* (Plagens et al. 2012a; Peng et al. 2013b; Plagens et al. 2014b) (Majumdar et al., unpublished). This is in contrast to the other characterized Type I systems where Cas3 is recruited for cleavage and not a structural component of DNA-binding surveillance complex (Jore et al. 2011a; Wiedenheft et al. 2011a; Sashital et al. 2012b; Westra et al. 2012c; Mulepati and Bailey 2013b; Sinkunas et al. 2013a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014). In Type I-E systems interaction of larget subunit protein Cse1 with PAM leads to recruitment of Cas3 at the site of cleavage downstream (Hochstrasser et al. 2014a; Huo et al. 2014a). It will be interesting to investigate if the structural uniqueness of Type I-A crRNPs that contain Cas3 proteins as stable components translates to a distinct molecular mechanism of target DNA interaction and silencing. Like Cas3s of other Type I systems (Mulepati and Bailey 2011; Sinkunas et al. 2011; Westra et al. 2012c; Huo et al. 2014a), Cas3" (*M. janschii, P. furiosus*) exhibits metal-dependent cleavage of single-stranded RNAs and DNAs *in vitro* (Beloglazova et al. 2011). In our biochemical assays, Cas3" cleaves single-stranded DNA in the presence of nickel, but does not cleave double-stranded DNA (Figure 3.S2B).

Non-self DNA recognition and targeting by Type I-A complex.

Based on bioinformatics predictions and functional evidence in *T. kodakarensis*, we predicted the trinucleotide motif, 5'-NGG-3' (target strand) as PAM sequence for invader DNA targeting by Type I-A system in *P. furiosus*. Our DNA binding and cleavage data reinforces that predicted PAM (NGG) is important for non-self DNA recognition and targeting by Type I-A system in *P. furiosus* (Figure 3.4). A matrix of mutations would be required to further delineate which PAM nucleotide position/s are critical and whether any partial mutations are tolerated for binding and/ or cleavage.

Based on functional parallels with Type I-E system, (Sashital et al. 2012b; Hochstrasser et al. 2014a; Huo et al. 2014a) large subunit protein Cas8a2 (Csa4-1) is a strong candidate for scanning and interacting with PAM in Type I-A mediated DNA recognition. In *E. coli* Type I-E complex, following PAM scanning and interaction large subunit protein Cse1 interacts with and recruits Cas3 at the site of cleavage (Hochstrasser et al. 2014a; Huo et al. 2014a). We have preliminary evidence of interaction between Cas8a2 (Csa4-1) and the two Cas3 proteins (Cas3", Cas3') (Figure S5). Biochemical evidence suggests Csa5 (small subunit) in *T. tenax* binds ssDNA in a sequence non-specific manner (Daume et al. 2014). We predict small subunit protein (Csa5) likely associates with the displaced non-target strand of bound duplex DNA and stabilizes the R-loop formed. Structure of Csa complex bound to DNA targets will elucidate details of molecular mechanism of DNA recognition and targeting.

In our attempt to decipher whether PAM is recognized as a single or double-strand entity, we introduced bubbles in the DNA duplex at PAM positions. Our data indicates Type I-A complex bound DNA with single-stranded PAM on target strand with reduced

efficiency (G-G, Figure 3.5B) but cleavage was only detected with substrates with double-stranded wild-type PAMs (Figure 3.5). Recent findings in *E. coli* Cascade (Type I-E) show similar evidence that while target strand PAM can determine binding to double-stranded DNA (Westra et al. 2012c), a structural duplex at PAM region is required for stable binding and cleavage (Hochstrasser et al. 2014a; Huo et al. 2014a). Also, recent evidence indicates PAM scanning and interaction of PAM with Cse1 of Type I-E complex triggers R-loop formation of DNA at the PAM proximal end (Szczelkun et al. 2014). We can't rule out the possibility that the non-complementarity at PAM region introduced in our hybrid PAM substrates (G-G, C-C Figure 3.5) somehow resulted in a distortion of duplex hindering formation of R-loops and thereby resulting in reduced binding and no detectable cleavage. Further investigations would be required to delineate whether the PAM sequence or complementarity at PAM region or both are recognized by Type I-A complex for DNA silencing.

Type I-A mediated DNA targeting (binding and cleavage) requires homology of protospacer with (invader-derived) guide sequence of crRNA (Figure 3.4). In characterized Type I-B, Type I-E and Type I-F there is a minimal crRNA guide sequence (downstream of 5' tag) called the seed sequence and base-pairing of the seed sequence with PAM proximal protospacer sequences on target strand of DNA is a determinant of CRISPR-Cas mediated DNA silencing (Jore et al. 2011a; Semenova et al. 2011; Wiedenheft et al. 2011b; Cady et al. 2012; Maier et al. 2013b). Whether there is a minimal crRNA seed sequence, which would be critical for DNA recognition and targeting by Type I-A system in *P. furiosus* needs to be explored.

Molecular features of Type I-A (Csa) complex in *P. furiosus*.

Analogous to other characterized Type I and Type III systems, our data indicates the structural core of the complex is formed by molecular interactions between crRNA and Cas5a (Cas5), Cas7 (Csa2). The crystal structures and cryo-EM structures of Type I-E (Cse), Type III-A (Csm) and Type III-B (Cmr) systems show that multiple Cas7 proteins bind along crRNA guide and 5' end is capped by Cas5 superfamily proteins binding to repeat-derived 5' tag (Wiedenheft et al. 2011a; Rouillon et al. 2013; Spilman et al. 2013a; Staals et al. 2013a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014). Native gel shift results in this paper, suggest a similar structural core and crosslinking data further elucidates that Cas5a binds crRNA tag and Csa2 (Cas7) binds crRNA-guide (Figure 3.7 and 3.8).

Gel shift analysis further suggests formation of larger subcomplex/es with core crRNP (Cas5a, Csa2) and Csa5 (Figure 3.7A). The structure of *S. solfataricus* Csa5 reveals α-helical protein that bears structural similarities with C-terminal α-helical domain of Type I-E small subunit protein, Cse2 in *E. coli*. Csa5 crystals show its propensity to form long helical multiimers (Reeks et al. 2013a). Based on other resolved structures of *E. coli* Type I-E and *P. furiosus* and *T. thermophilus* Type III-B complexes (Wiedenheft et al. 2011a; Spilman et al. 2013a; Staals et al. 2013a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014), we predict small subunit Csa5 multimers are part of the crRNP structure, very likely through interaction with Cas7 (Figure 3.9). Crosslinking results (Figure 3.8) suggest association of large subunit protein Csa4-1 (Cas8a2) with crRNA guide. Similar to Type I-E crRNP, we speculate large subunit protein localizes at the 5' tag-proximal guide sequences and associates with the Cas7

subunit at that region (Hochstrasser et al. 2014a; Mulepati et al. 2014a). In *P. furiosus*, there are two copies of Cas8a2 (Csa4-1, Csa4-2) (Figure 13.A). Both Csa4-1 and Csa4-2 were found in native Csa crRNP (Majumdar et al., unpublished). In our reconstitution assays Csa4-2 was dispensable for DNA binding and cleavage (Figure 3.S1C). Also, Csa4-2 could not substitute for Csa4-1 (data not shown). We can't conclude if the recombinant protein is non-functional or there is gene redundancy. *S. solfataricus* Csa3 structure reveals a N-terminal di-nucleotide binding domain (CARF) and a C-terminus domain winged helix (Lintner et al. 2011b; Makarova et al. 2014). Csa3 family of proteins predicted to be transcriptional regulators (Haft et al. 2005a; Makarova et al. 2014). Csa3 was not identified in native *P. furiosus* Csa complex isolated by immunoprecipitation in our previous work (Majumdar et al, unpublished). Our current finding validates the prediction that Csa3 is not a component of Type I-A effector complex in *P. furiosus*.

In summary, Csa complex observed in native gel shifts with six proteins (Cas8a2 (Csa4-1), Cas3', Cas3", Cas5a, Csa2 and Csa5) constitutes the minimal DNA-targeting holoenzyme (Figure 3.2). We speculate protein-protein interactions around the crRNP core builds-up the holoenzyme (Figure 3.9). Further structural studies of Csa complex will delineate the details of molecular organization.

RNA sequencing of total RNAs and complex-associated crRNAs (Csa, Cst and Cmr) in *P. furiosus*, showed that crRNAs contain repeat-derived, 5' tag generated by Cas6 cleavage (Majumdar et al, unpublished). Csa complex-associated crRNAs show heterogeneity at 3' end suggesting further processing of 3' end to produce shorter crRNAs. One of the most abundant crRNA size forms that associate with all three

complexes in *P. furiosus* are comprised of 5' tag and full-length invader-derived guide (~37 nt for most guide) (Majumdar et al, unpublished). We reconstituted the active complex with this crRNA size-form with 7.01 invader-derived guide sequence. Native gel shift result reveals a strict requirement for the Cas6 generated (5'-OH) chemical end on the 5' end of crRNA as well as the 5' tag (Figure 3.7B, 3.7C). Both Csa core and holoenzyme display remarkable specificity for mature crRNA, especially on the 5'-end. Similar to Cmr complex (Hale et al. 2012a), our current data implies 5' tag and 5'-OH chemical end constitute signatures that are likely recognized by Cas5 superfamily proteins for assembly of Csa complex.

Concluding Remarks.

P. furiosus has been a model organism to investigate three distinct and co-existing CRISPR-Cas systems. In our previous work, we have identified three native immune systems (Type I-A, Type I-G and Type III-B) and delineated features of Type III-B mediated target RNA silencing. In this work, we have explored the molecular basis of invader DNA silencing by Type I-A system *in vitro*. We predict that both Type I-A and Type I-G would act as independent DNA-silencing immune-effector systems in *P. furiosus*. Further *in vivo* studies are necessary for genetic evidence of invader DNA silencing by co-functioning Type I-A and Type I-G systems in *P. furiosus*. The availability of active, *in vitro* reconstituted Type I-A complex would enable structure-function studies to dissect detailed molecular mechanisms of DNA silencing.

Materials and methods

Expression and Purification of recombinant Csa proteins.

Clones containing genes encoding N-terminal 6X his-tag of Cas3", Cas5a, Csa2 (pET24) (gift of Michael Adams, University of Georgia), Csa4-1, Cas3', Csa3 (pET200), Csa5 (pET24) and C-terminal 6X his-tagged Csa4-2 (pET101) were transformed into *E. coli* BL21 RIPL. 1-4 liter cultures were grown up to OD_{600} of 0.4-0.6 at 37°C and induced with 1mM IPTG at room temperature overnight. Cells were lysed by sonication (Misonix Sonicator 3000) in buffer A (25 mM TRIS, 250 mM NaCl pH 7.6) with 10 mM imidazole and 0.1 mM phenylmethylsulfonyl fluoride (PMSF). The lysed cells were incubated at 70°C for 30 minutes with frequent mixing, centrifuged at 14,000 rpm at 4°C and soluble fraction was filtered (0.8 μm filter pore size Millex filter unit, Millipore). His-tagged proteins were affinity purified over nickel columns followed by washing with buffer A with 25 mM imidazole and eluting with buffer A and increasing imidazole (50 mM, 100 mM, 200 mM, 500 mM). Peak elution fractions were dialyzed in buffer A using Slide-a-lyzer mini dialysis cassettes (ThermoScientific), quantified with Qubit 2.0 fluorometer (Life Technologies) and stored at 4°C, prior to use for functional assays.

Preparation of substrates for functional assays.

Synthetic RNA and DNA oligo sequences are listed in table 1. 7.01 crRNAs (wild type tag and mutant tag) were described (Hale et al. 2009; Hale et al. 2012a). DNA oligos were purchased from MWG, Eurofins. DNA and RNA (Figures 3.1-3.6, 3.S1-3.S3) were 5' labeled with T4 polynucleotide kinase (PNK) and [γ-^{32}P] ATP as described (Carte et

al. 2008b). DNA oligos were annealed by mixing unlabeled DNA and 5' end-labeled DNA in a molar ratio of 2:1 in 10 mM TRIS (pH 7.5 – 8), 50 mM NaCl, 1 mM EDTA and incubating at 95°C for 5 minutes followed by slow cooling until 23°C (at the rate of 1°C per minute). Annealing was confirmed and substrates were extracted from 8% non-denaturing PAGE. The annealed DNA duplex was extracted with phenol/ chloroform/ isoamyl alcohol (pH-8) and ethanol precipitation, resuspended in buffer A and stored at -20°C for future use in functional assays. Details of oligonucleotide combinations to prepare each RNA and DNA duplex are listed in table 2.

crRNAs were 3' end labeled (Figure 3.7, 3.S4) with T4 RNA ligase (New England Biolabs) and [α-^{32}P] pCp (3000 Ci/ mmol, Perkin Elmer) as described (Carte et al. 2010). Internally labeled 7.01 crRNA was prepared (Fig 3.5B) by *in vitro transcription* of repeat-spacer unit with [α-^{32}P]-GTP (800 Ci/mmol, 10 mCi/mL, Perkin Elmer) followed by Cas6 digestion as described (Spilman et al. 2013a).

cRNP reconstitution and functional assays.

For reconstitution of complex, 1 μM each of recombinant Csa proteins were incubated with 0.5 pmoles of unlabeled crRNA (Figure 3.1-3.6, 3.S1-3.S3) or 0.05 pmoles of labeled crRNA (Figure 3.7, 3.S4) in buffer A (25 mM TRIS, pH- 7.6, 250 mM NaCl), 2mM ATP, 1.5 mM Mg, 1 unit of SUPERase-In ribonuclease inhibitor (Applied Biosystems) and 1 μg of *E. coli* tRNA (competitor) at 70°C for 30 minutes.

Functional assays involved addition of 0.02 pmoles - 0.05 pmoles of DNA substrate and 100 μM Ni (nickel chloride) to the reconstituted reactions (Figure 3.1-3.6, 3.S1-3.S3) followed by incubation at 70°C for 3 hours. For native gel shifts (Figures 3.2B, 3.3C,

3.4C, 3.5B, 3.S3) half of the reactions were run on 6% native polyacrylamide gels (containing 0.5X TBE and 8% glycerol). Electrophoresis was performed in 0.5X TBE for 2 h at 110 V. For assaying cleavages, remaining half of reactions were treated with 1 µg of Proteinase K (New England Biolabs) at 37°C for 15 mins followed by PCI extraction (pH 8) and ethanol precipitation and resolved on 15% denaturing (7 M urea) polyacrylamide sequencing or mini gels. Gels were dried and the radiolabeled substrates and cleavages visualized by phosphor imaging. 5' end-labeled low molecular weight marker (10-100, Affymetrix) was used to determine the sizes of the observed products.

Site-directed mutagenesis of Cas3".

To generate the HD-AA mutant of Cas3", QuikChange site-directed mutagenesis was performed as per manufacture's guidelines (Agilent Technologies). Briefly, primers were designed and ordered from MWG Eurofins (listed in table 3.1). The mutants were generated by PCR amplification of wild type cas3" (pET24) clone followed by DpnI digestion and transformation in to *E. coli* BL21 RIPL. Mutation was confirmed by sequencing.

UV induced RNA-protein crosslinking

crRNA-Csa proteins crosslinking assays, using photoactivatable 4-thio uridine and internally (^{32}P) radiolabeled crRNAs were done as described (Spilman et al. 2013a) with following modifications. crRNP reconstitution reactions were scaled up 2 times of standard reconstitutions done for crRNP formation (except no RNasin added). A quarter of the reactions was assessed for crRNA binding by native gel electrophoresis (Figure

S4B) Following crosslinking at 312 nm UV light and complexes were disrupted by addition of 1% SDS treatment and boiling for 10 mins. crRNAs (body) were treated with RNase T1 (1000 U, Thermo), crRNAs (tag) were treated with RNase T1 (1000 U, Thermo) and RNase A (10 µg, Thermo) and crRNAs (guide) were treated with RNase A (10 µg, Thermo) at 37 degrees for 1 hour. Laemmli Buffer was added and the samples were boiled and separated on 12.5% SDS-PAGE gels. The gels were visualized by G-250 Coomassie staining, dried, and subjected to phosphor imaging and autoradiography to allow for alignment of protein bands with autoradiographs.

Protein- protein interaction studies

1 liter culture of N- terminal, 6X- His tagged Cas3' (pET 200) and and 500 ml cultures each of untagged Csa4-1, Cas3" (modified pET 100) were grown and induced with IPTG as mentioned before. Before sonication three proteins (2:1:1 of Cas3'+Cas3"+Csa4-1) and two proteins (2:1 of Cas3'+Cas3") cultures were mixed and purified as mentioned before. The samples were affinity purified over nickel columns. The interaction between N-tagged protein (bait) and untagged proteins (prey) was visualized by running the elutions on 12.5% SDS-PAGE followed by staining with Coomassie (R-250) staining. The peak elutions are shown (Figure S5A). The Cas3'+Cas3" sub-complex was further purified by gel filtration (Superdex 200, GE Healthcare Hong Li's lab) (FigureS5B).

Acknowledgments

We would like to thank Dr. Hong Li (Florida State University) for contributing towards Cas3", Cas3' interaction studies (Figure 3.S5B) and Dr. Michael Adams (University of Georgia) for some of the *csa* clones. We thank members of Terns' lab for their technical inputs and peer review of the paper. This work was supported by National Institute of Health (NIH) to M.P.T. and R.M.T.

References

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., and Yakunin, A.F. (2011). Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. The EMBO Journal *30*, 4616-4627.

Bernick, D.L., Cox, C.L., Dennis, P.P., and Lowe, T.M. (2012). Comparative genomic and transcriptional analyses of CRISPR systems across the genus *Pyrobaculum*. Frontiers in Microbiology *3*, 251.

Brendel, J., Stoll, B., Lange, S.J., Sharma, K., Lenz, C., Stachler, A.E., Maier, L.K., Richter, H., Nickel, L., Schmitz, R.A.*, et al.* (2014). A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in *Haloferax volcanii*. The Journal of Biological Chemistry *289*, 7164-7177.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R., and O'Toole, G.A. (2012). The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. Journal of Bacteriology *194*, 5728-5738.

Cady, K.C., and O'Toole, G.A. (2011). Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. Journal of Bacteriology *193*, 3433-3445.

Carte, J., Pfister, N.T., Compton, M.M., Terns, R.M., and Terns, M.P. (2010). Binding and cleavage of CRISPR RNA by Cas6. RNA *16*, 2181-2188.

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

Daume, M., Plagens, A., and Randau, L. (2014). DNA binding properties of the small cascade subunit csa5. PloS One *9*, e105716.

Deng, L., Garrett, R.A., Shah, S.A., Peng, X., and She, Q. (2013a). A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. Molecular Microbiology *87*, 1088-1099.

Deng, L., Niu, S., Liu, X., Che, Y., and Li, E. (2013b). Coniochaetones E-I, new 4H-chromen-4-one derivatives from the Cordyceps-colonizing fungus *Fimetariella sp.* Fitoterapia *89*, 8-14.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. Journal of Bacteriology *190*, 1390-1400.

Elmore, J.R., Yokooji, Y., Sato, T., Olson, S., Glover, C.V., 3rd, Graveley, B.R., Atomi, H., Terns, R.M., and Terns, M.P. (2013). Programmable plasmid interference by the CRISPR-Cas system in *Thermococcus kodakarensis*. RNA Biology *10*, 828-840.

Fischer, S., Maier, L.K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., Dyall-Smith, M., and Marchfelder, A. (2012). An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. The Journal of Biological Chemistry *287*, 33351-33363.

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

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences of the United States of America *109*, E2579-2586.

Gasiunas, G., Sinkunas, T., and Siksnys, V. (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. Cellular and Molecular Life sciences : CMLS *71*, 449-465.

Gudbergsdottir, S., Deng, L., Chen, Z., Jensen, J.V., Jensen, L.R., She, Q., and Garrett, R.A. (2011). Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. Molecular Microbiology *79*, 35-49.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Computational Biology *1*, e60.

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

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M.*, et al.* (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Molecular Cell *45*, 292-302.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945-956.

Hatoum-Aslan, A., Maniv, I., and Marraffini, L.A. (2011). Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. Proceedings of the National Academy of Sciences of the United States of America *108*, 21218-21222.

Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proceedings of the National Academy of Sciences of the United States of America *111*, 6618-6623.

Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M.D., Jr., Zhou, S., Rajashankar, K., Kurinov, I.*, et al.* (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nature Structural & Molecular Biology.

Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. Science.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R.*, et al.* (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nature Structural & Molecular Biology *18*, 529-536.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biology *8*, R61.

Lillestol, R.K., Shah, S.A., Brugger, K., Redder, P., Phan, H., Christiansen, J., and Garrett, R.A. (2009). CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. Molecular Microbiology *72*, 259-272.

Lintner, N.G., Frankel, K.A., Tsutakawa, S.E., Alsbury, D.L., Copie, V., Young, M.J., Tainer, J.A., and Lawrence, C.M. (2011a). The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. Journal of Molecular Biology *405*, 939-955.

Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q., Copie, V.*, et al.* (2011b). Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). The Journal of Biological Chemistry *286*, 21643-21656.

Magadan, A.H., Dupuis, M.E., Villion, M., and Moineau, S. (2012). Cleavage of phage DNA by the Streptococcus thermophilus CRISPR3-Cas system. PloS One *7*, e40913.

Maier, L.K., Lange, S.J., Stoll, B., Haas, K.A., Fischer, S., Fischer, E., Duchardt-Ferner, E., Wohnert, J., Backofen, R., and Marchfelder, A. (2013). Essential requirements for the detection and degradation of invaders by the Haloferax volcanii CRISPR/Cas system I-B. RNA Biology *10*, 865-874.

Makarova, K.S., Anantharaman, V., Grishin, N.V., Koonin, E.V., and Aravind, L. (2014). CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. Frontiers in Genetics *5*, 102.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F*., et al.* (2011). Evolution and classification of the CRISPR-Cas systems. Nature Reviews Microbiology *9*, 467-477.

Manica, A., Zebec, Z., Teichmann, D., and Schleper, C. (2011). In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. Molecular Microbiology *80*, 481-491.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843-1845.

Marraffini, L.A., and Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568-571.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733-740.

Mulepati, S., and Bailey, S. (2011). Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). The Journal of Biological Chemistry *286*, 31896-31903.

Mulepati, S., and Bailey, S. (2013). In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. The Journal of Biological Chemistry *288*, 22184-22192.

Mulepati, S., Heroux, A., and Bailey, S. (2014). Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science.

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

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of cas9 in complex with guide RNA and target DNA. Cell *156*, 935-949.

Peng, W., Li, H., Hallstrom, S., Peng, N., Liang, Y.X., and She, Q. (2013). Genetic determinants of PAM-dependent DNA targeting and pre-crRNA processing in *Sulfolobus islandicus*. RNA Biology *10*, 738-748.

Plagens, A., Tjaden, B., Hagemann, A., Randau, L., and Hensel, R. (2012). Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. Journal of Bacteriology *194*, 2491-2500.

Plagens, A., Tripp, V., Daume, M., Sharma, K., Klingl, A., Hrle, A., Conti, E., Urlaub, H., and Randau, L. (2014). In vitro assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. Nucleic acids Research.

Reeks, J., Graham, S., Anderson, L., Liu, H., White, M.F., and Naismith, J.H. (2013). Structure of the archaeal Cascade subunit Csa5: relating the small subunits of CRISPR effector complexes. RNA Biology *10*, 762-769.

Richter, H., Zoephel, J., Schermuly, J., Maticzka, D., Backofen, R., and Randau, L. (2012). Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. Nucleic acids Research *40*, 9887-9896.

Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C.V., Spagnolo, L., and White, M.F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. Molecular Cell *52*, 124-134.

Sashital, D.G., Wiedenheft, B., and Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. Molecular Cell *46*, 606-615.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proceedings of the National Academy of Sciences of the United States of America *108*, 10098-10103.

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

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). *In vitro* reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. The EMBO Journal *32*, 385-394.

Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., and Stagg, S. (2013). Structure of an RNA silencing complex of the CRISPR-Cas immune system. Molecular Cell *52*, 146-152.

Staals, R.H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K.*, et al.* (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. Molecular Cell *52*, 135-145.

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

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proceedings of the National Academy of Sciences of the United States of America.

Terns, M.P., and Terns, R.M. (2011). CRISPR-based adaptive immune systems. Current Opinion in Microbiology *14*, 321-327.

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nature Reviews Microbiology *12*, 479-492.

van Duijn, E., Barbu, I.M., Barendregt, A., Jore, M.M., Wiedenheft, B., Lundgren, M., Westra, E.R., Brouns, S.J., Doudna, J.A., van der Oost, J.*, et al.* (2012). 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*. Molecular & Cellular Proteomics : MCP *11*, 1430-1441.

Vestergaard, G., Garrett, R.A., and Shah, S.A. (2014). CRISPR adaptive immune systems of Archaea. RNA Biology *11*, 156-167.

Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K., and Brouns, S.J. (2013). Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. PLoS Genetics *9*, e1003742.

Westra, E.R., Swarts, D.C., Staals, R.H., Jore, M.M., Brouns, S.J., and van der Oost, J. (2012a). The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. Annual Review of Genetics *46*, 311-339.

Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K.*, et al.* (2012b). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Molecular Cell *46*, 595-605.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011a). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature *477*, 486-489.

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

Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J., Dickman, M.J.*, et al.* (2011b). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proceedings of the National Academy of Sciences of the United States of America *108*, 10092-10097.

Zebec, Z., Manica, A., Zhang, J., White, M.F., and Schleper, C. (2014). CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*. Nucleic acids Research.

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.V.*, et al.* (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Molecular Cell *45*, 303-313.

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. Nature.

Figure 3.1. Reconstitution of active, DNA targeting, Type I-A system in *P. furiosus*

A. *cas* gene organization in *Pyrococcus furiosus*. Genes encoding predicted adaptation proteins (*cas 1, cas2, cas4*) and biogenesis (*cas6*) are colored grey. The three effector modules- Type I-A (*csa*) indicated in a box, is colored green, Type I-G (*cst*) is colored yellow, and. Type III-B (*cmr*) is colored blue. **B.** Csa crRNP reconstituted with six Csa proteins (Csa4-1, Cas3", Cas3', Cas5a, Csa2, Csa5, labeled +) and 45-nt crRNA consisting 8-nt 5'-tag (black in schematic) and 37-nt 7.01 guide (red in schematic) cleaves double-stranded DNA. The DNA substrate contains wild-type PAM (blue in schematic) and 7.01 cognate protospacer (red in schematic). Either target strand (complementary to crRNA guide) or non-target strand of DNA are 5' (^{32}P) radiolabeled as labeled on top. Reactions without any Csa proteins are labeled (-). Major cleavages are indicated with asterisks. 5' radiolabeled DNA size-standard (Affymetrix, low molecular weight marker, 10-100 nts) is labeled (M).

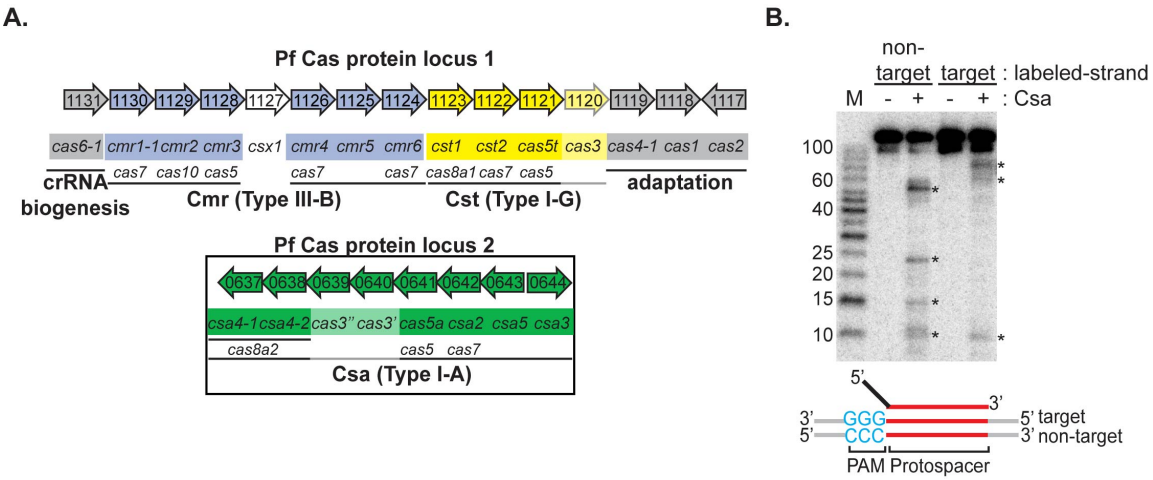
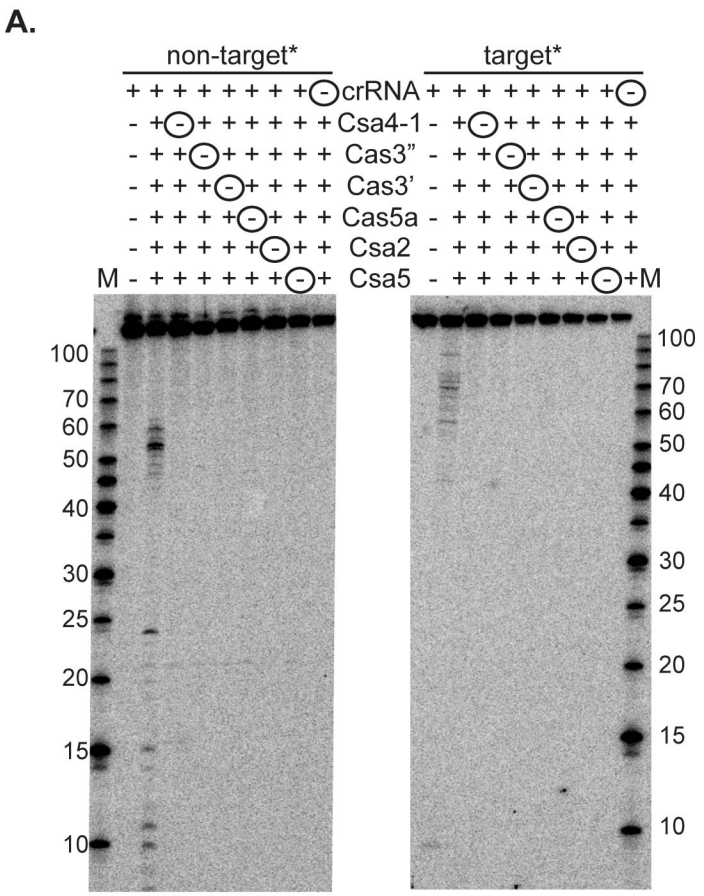

A.
Pf Cas protein locus 1
1131 1130 1129 1128 1127 1126 1125 1124 1123 1122 1121 1120 1119 1118 1117
cas6-1 cmr1-1 cmr2 cmr3 csx1 cmr4 cmr5 cmr6 cst1 cst2 cas5t cas3 cas4-1 cas1 cas2
cas7 cas10 cas5 cas7 cas7 cas8a1 cas7 cas5
crRNA biogenesis
Cmr (Type III-B)
Cst (Type I-G)
adaptation
Pf Cas protein locus 2
0637 0638 0639 0640 0641 0642 0643 0644
csa4-1 csa4-2 cas3'' cas3' cas5a csa2 csa5 csa3
cas8a2 cas5 cas7
Csa (Type I-A)
B.
non-target target target : labeled-strand
M - + - + : Csa
100
60
40
25
20
15
10
5'
3'
3' GGG 5' target
5' CCC 3' non-target
PAM Protospacer

Figure 3.2. Minimal DNA targeting Type I-A complex in *P. furiosus*.

Effect on Type –IA mediated DNA cleavage **(A)** DNA binding **(B)** when one of the components of active complex (crRNA, Csa4-1, Cas3", Cas3', Cas5a, Csa2, Csa5) were excluded at one time. The excluded (-, with a circle) and included (+) components for each reaction are labeled. Lane 1 of each gel shows reactions without any Csa proteins. Either non-target or target strand of DNA is 5'- radiolabeled (^{32}P) and indicated on top with asterisk. 5' radiolabeled DNA size-standard is labeled (M) and sizes (10-100 nt) labeled on left.


A.
non-target*
target*
crRNA
Csa4-1
Cas3"
Cas3'
Cas5a
Csa2
Csa5
M
100
70
60
50
40
30
25
20
15
10

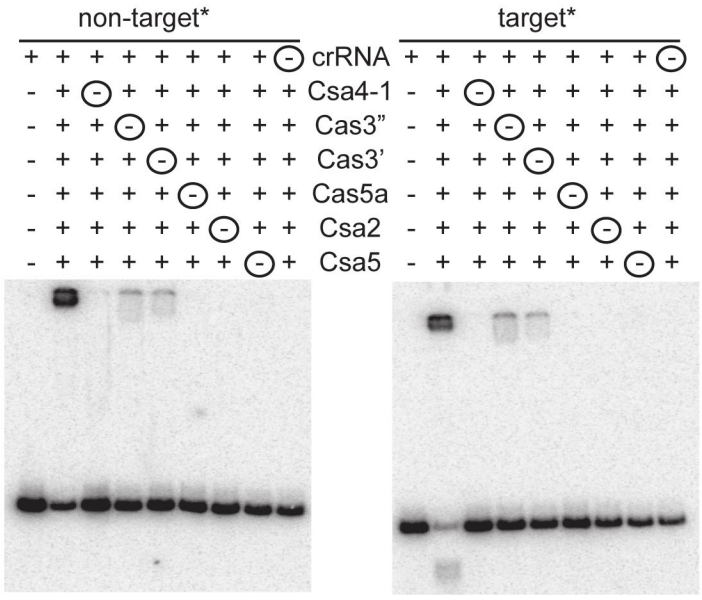

B.
non-target*
target*
crRNA
Csa4-1
Cas3"
Cas3'
Cas5a
Csa2
Csa5

Figure 3.3. Cas3" is the effector nuclease of DNA targeting, Type I-A (Csa crRNP) in *P. furiosus*.

A. SDS PAGE showing recombinant *P. furiosus* Cas3" wild-type (wt.) and Cas3" (HD-AA) mutated (mut.) protein. Molecular weight marker (Precision plus dual color standard, Biorad) is used for size estimation and labeled M. Labeled sizes on the left are in Kilo-Daltons (KD). DNA cleavage **(B)** and DNA binding **(C)** without (-) and with Csa crRNP (+, containing minimal components for active complex) assembled with either wild-type (wt.) Cas3" or mutated (mut.) Cas3". Either non-target or target strand of DNA is 5'- radiolabeled (^{32}P) and indicated on top with asterisk. 5' radiolabeled DNA size-standard is labeled (M) and sizes (10-100 nt) labeled on left.

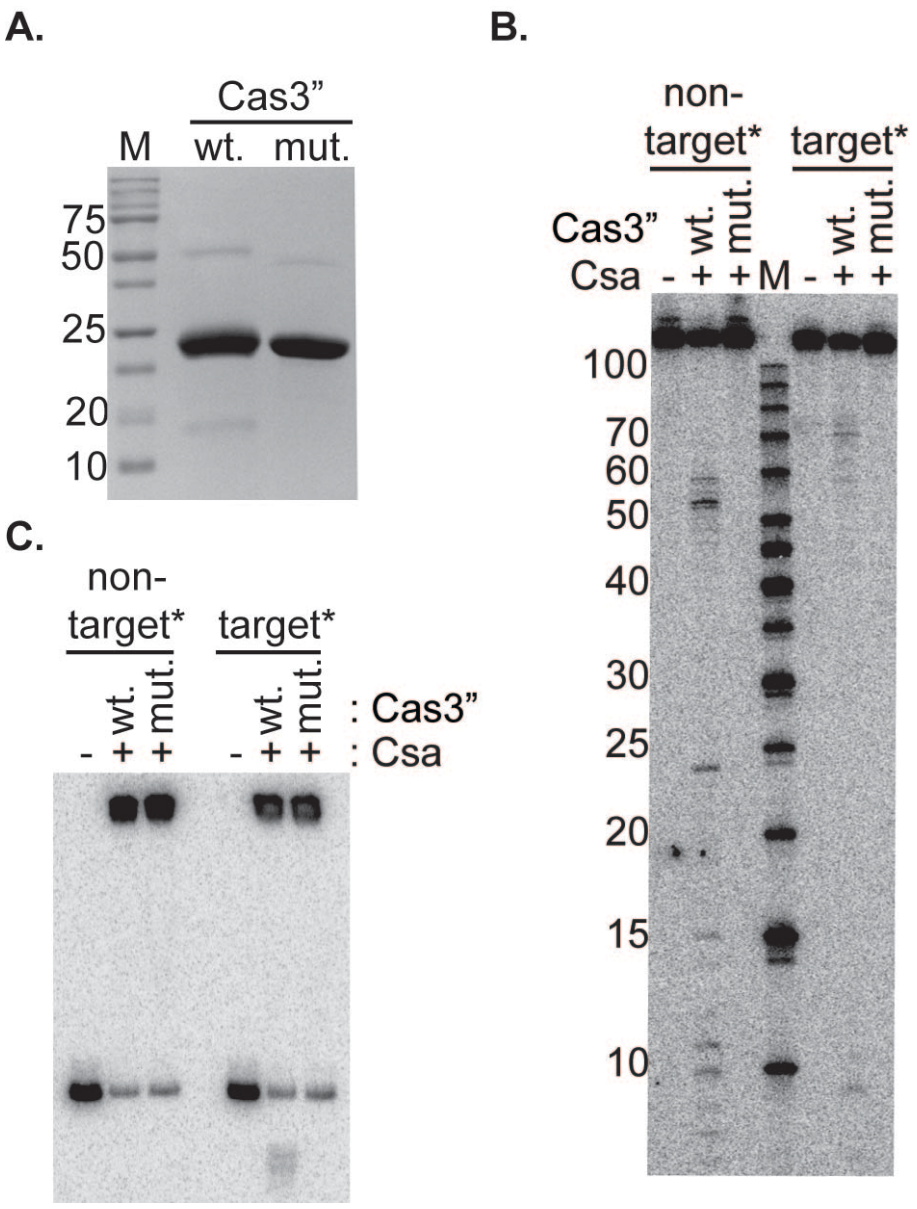

A.
Cas3”
M
wt.
mut.
75
50
25
20
10
B.
non-
target*
target*
Cas3”
wt.
mut.
Csa
-
+
+
M
-
+
+
100
70
60
50
40
30
25
20
15
10
C.
non-
target*
target*
-
+wt.
+mut.
: Cas3”
: Csa

Figure 3.4. Type I-A mediated DNA binding and cleavage is crRNA-guided and PAM-dependent in *P. furiosus*.

DNA substrates used in the assay are depicted (**A)**. crRNA is shown on top with 5'-tag (black) and guide (red). Target and non-target strands of double-stranded DNA with wild-type cognate protospacer (red) and wild-type PAM sequence (blue), mutated (mut.) PAM (green), mutated (mut.) non-cognate protospacer (orange) are marked.

DNA cleavage **(B)** and DNA binding **(C)** by no protein (-) or Csa crRNP (+) and either of the specified substrates, as labeled. Non-target or target strand of DNA is 5'-radiolabeled (^{32}P) as indicated on top (asterisk). 5' radiolabeled DNA size-standard is labeled (M) and sizes (10- 100 nt) labeled on left.

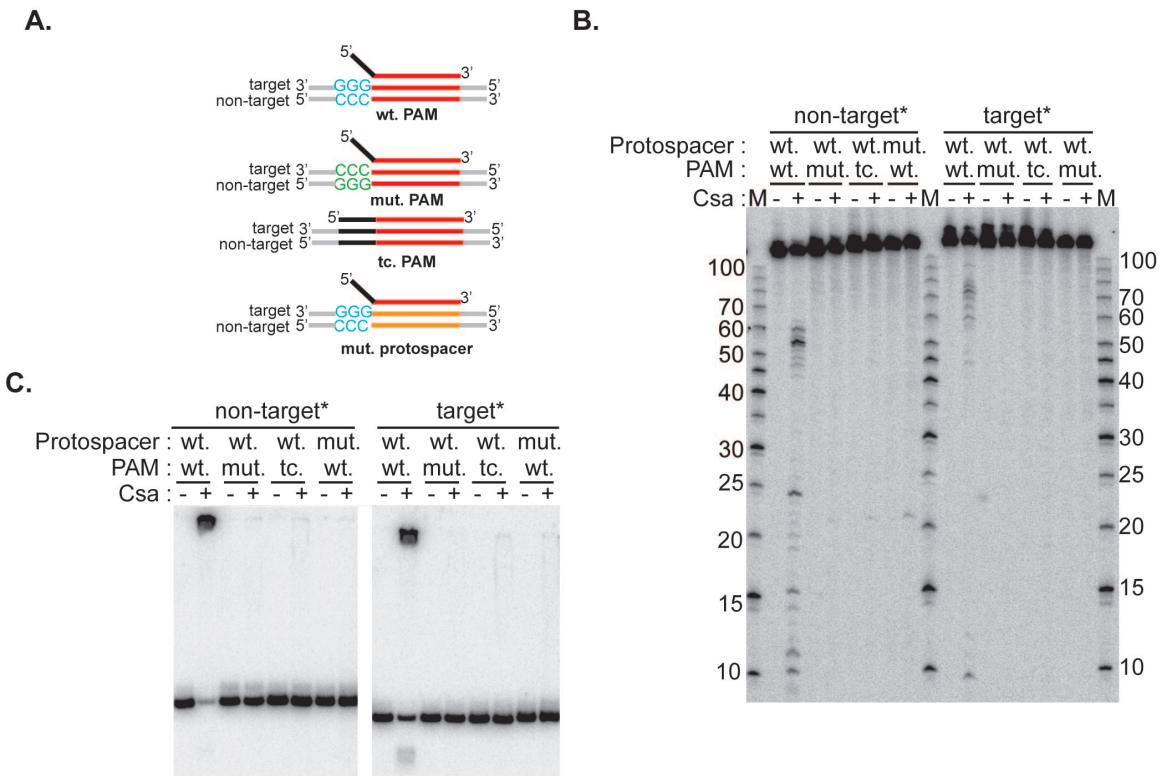

A.
target 3'
non-target 5'
GGG
CCC
wt. PAM
CCC
GGG
mut. PAM
tc. PAM
mut. protospacer
B.
non-target*
target*
Protospacer :
PAM :
Csa :
wt. wt. wt. mut. wt. wt. wt. wt.
wt. mut. tc. wt. wt. mut. tc. mut.
M - + - + - + - + M - + - + - + - + M
100
70
60
50
40
30
25
20
15
10
C.
non-target*
target*
Protospacer :
PAM :
Csa :
wt. wt. wt. mut. wt. wt. wt. mut.
wt. mut. tc. wt. wt. mut. tc. wt.
- + - + - + - + - + - + - + - +

Figure 3.5. Type I-A recognizes double-stranded PAM for efficient DNA targeting in *P. furiosus*.

DNA substrates used in the assay are depicted **(A)**. For each substrate crRNA is shown on top with 5'-tag (black) and guide (red). Target and non-target strands of double-stranded DNA with wild-type cognate protospacer (red) and wild-type PAM sequence (blue) are marked. DNAs with hybrid PAMs (G-G, C-C) have wild-type PAM (blue) on one strand (target strand in G-G, non-target strand in C-C) and mutated PAM (green) on the other strand, leaving PAM sequence non-complementary.

DNA binding **(B)** and DNA cleavage **(C)** of the specified substrates (wt. PAM. G-G PAM, C-C PAM) with no protein (-) or Csa crRNP (+). Non-target strand of DNA is 5'-radiolabeled (^{32}P) as indicated on top (asterisk). 5' radiolabeled DNA size-standard is labeled (M) and sizes (10- 100 nt) labeled on left.

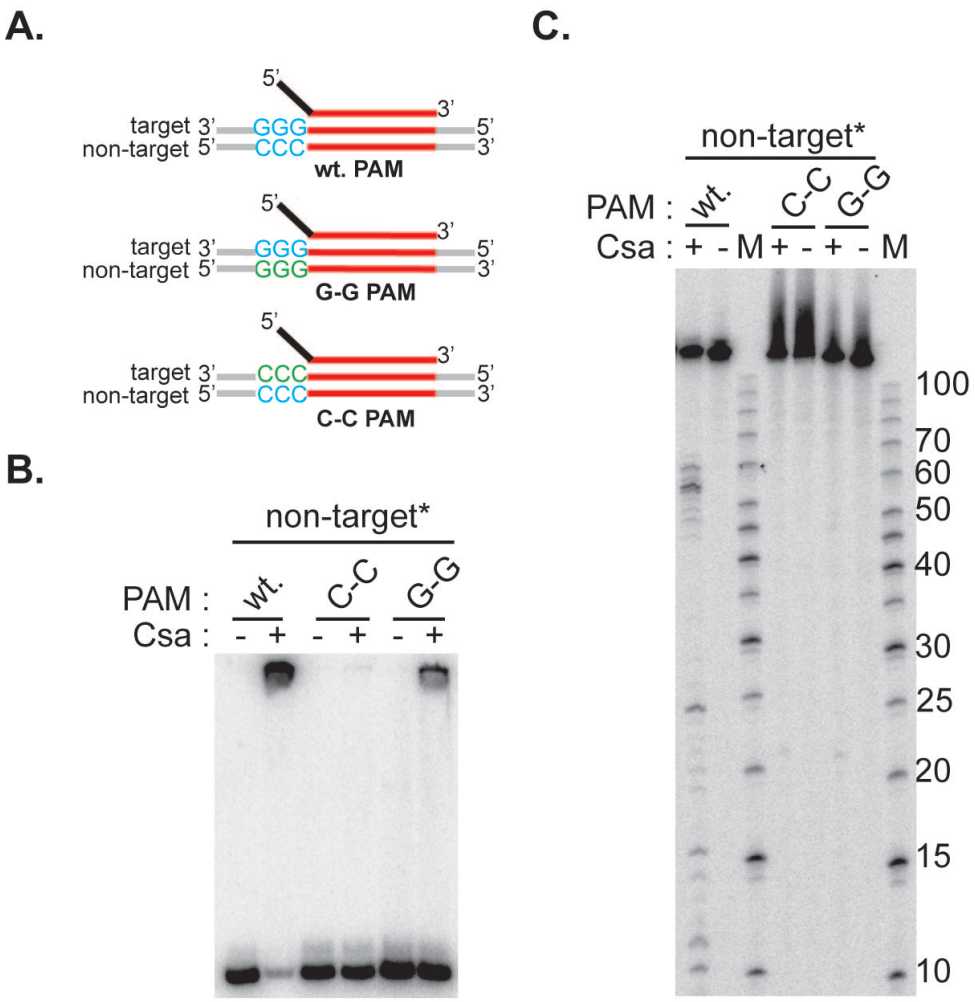

A.
5'
3'
target 3'
non-target 5'
GGG
CCC
5'
3'
wt. PAM
5'
3'
target 3'
non-target 5'
GGG
GGG
5'
3'
G-G PAM
5'
3'
target 3'
non-target 5'
CCC
CCC
5'
3'
C-C PAM
B.
non-target*
PAM :
wt.
C-C
G-G
Csa :
- + - + - +
C.
non-target*
PAM :
wt.
C-C
G-G
Csa :
+ - M + - + - M
100
70
60
50
40
30
25
20
15
10

Figure 3.6. Profile of Type I-A (Csa crRNP) mediated DNA cleavage.

A. Effect of ATP (-/+) on DNA cleavage (117-nt, labeled II in figure 6B) without (-) or with (+) Csa crRNP. Either non-target or target strand of DNA is 5' radiolabeled (P-32) as indicated on top. The ATP-independent cleavage products on non-target strand is indicated with asterisks. 5' radiolabeled DNA size-standard is labeled (M) and sizes (10-100 nt) labeled on left. **B.** Profile of cleavage products without (-) or with (+) Csa-crRNP and two DNA size forms: shorter DNA (67-nt, I) and longer DNA (117-nt, labeled II). Either non-target or target strand of DNA is 5' radiolabeled (P-32) as indicated on top. 5' radiolabeled DNA size-standard is labeled (M) and sizes (10- 100 nt) labeled on left. **C.** Mapping Csa-mediated cleavage sites in shorter DNA (I) and longer DNA (II, tested in A and B) sequences. For each DNA, target and non-target strands strands, protospacer (red, labeled on top), PAM (blue, labeled on top) and 5', 3' flanking regions (grey) are marked. The cleavages sites in the sequences are indicated with arrows and numbered (counting from 5' end of each strand). More abundant products (major products) are indicated with black circles. ATP-independent cleavages seen in non-target strand of longer DNA (II) in 6A are indicated with asterisks. The other cleavages mapped (in both strands of II) are ATP-dependent.

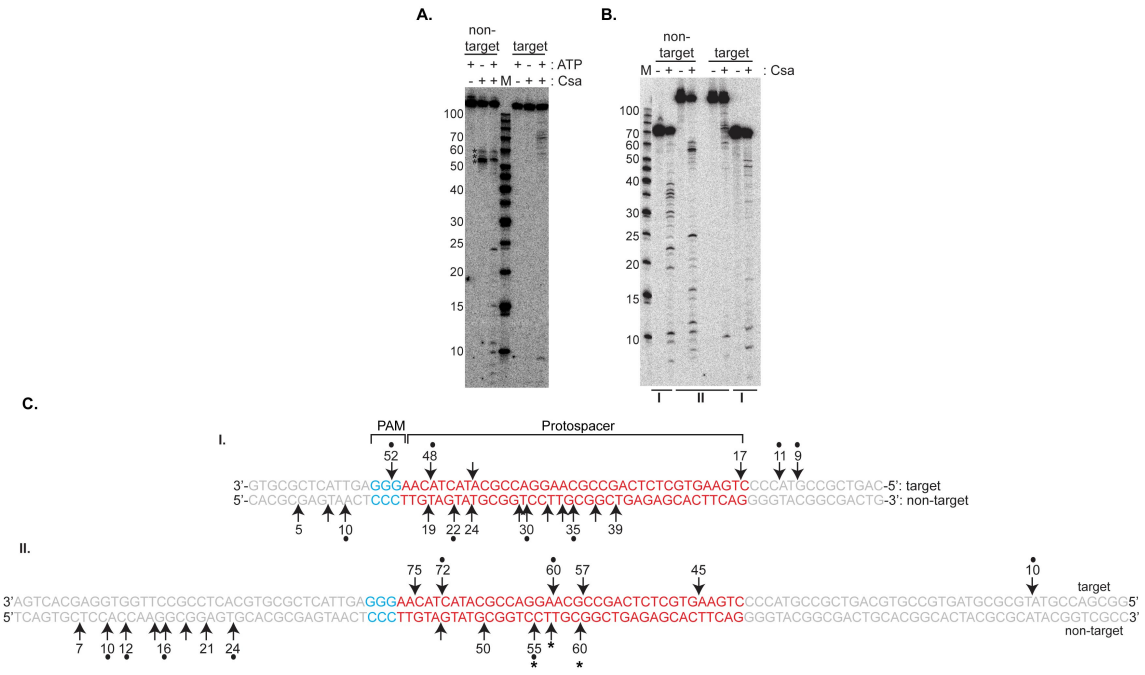

A.
non-target
target
+ - + + - + : ATP
- + + M - + + : Csa
100
70
60
50
40
30
25
20
15
10
B.
non-target
target
M - + - + - + - + : Csa
100
70
60
50
40
30
25
20
15
10
I
II
I
C.
I.
PAM
Protospacer
52
48
17
11
9
3'-GTGCGCTCATTGAGGGAACATCATACGCCAGGAACGCCGACTCTCGTGAAGTCCCCATGCCGCTGAC-5': target
5'-CACGCGAGTAACTCCCTTGTAGTATGCGGTCCTTGCGGCTGAGAGCACTTCAGGGGTACGGCGACTG-3': non-target
5
10
19
22
24
30
35
39
II.
75
72
60
57
45
10
target
3'AGTCACGAGGTGGTTCCGCCTCACGTGCGCTCATTGAGGGAACATCATACGCCAGGAACGCCGACTCTCGTGAAGTCCCATGCCGCTGACGTGCCGTGATGCGCGTATGCCAGCGG5'
5'TCAGTGCTCCACCAAGGCGGAGTGCACGCGAGTAACTCCCTTGTAGTATGCGGTCCTTGCGGCTGAGAGCACTTCAGGGTACGGCGACTGCACGGCACTACGCGCATACGGTCGCC3'
non-target
7
10
12
16
21
24
50
55
60

Figure 3.7. Molecular features of Csa crRNP formation.

A. Gel mobility shift assay to visualize crRNPs formed with 3'- radiolabeled (^{32}P) crRNA and no proteins (-), six Csa proteins (Csa4-1, Cas3", Cas3', Cas3', Cas5a, Csa2, Csa5), and combinations of five Csa proteins (one protein excluded at a time, indicated with a circled minus). The free crRNA and crRNP/s formed are labeled on right. **B.** Gel mobility shift assay to visualize Csa crRNPs formed without proteins (-), six Csa proteins (Csa4-1, Cas3", Cas3', Cas3', Cas5a, Csa2, Csa5), two Csa proteins (Cas5a, Csa2) and 3'-radiolabeled (^{32}P) crRNAs, with either wild type 5' tag (marked black in schematic, lanes 1-3) or mutated 5' tag (marked red in schematic, lanes 4-6). Both crRNAs contain the same guide sequence (7.01, marked orange in schematic). The free crRNA and specific crRNP/s formed are labeled on left. **C.** Gel mobility shift assay to visualize Csa crRNPs formed without proteins (-), six Csa proteins (Csa4-1, Cas3", Cas3', Cas3', Cas5a, Csa2, Csa5), two Csa proteins (Cas5a, Csa2) and either crRNAs with 5'-hydroxyl, 3' phosphate (3' radiolabeled, lanes 1-3), crRNAs with 5'- hydroxyl, 3' hydroxyl (internally radiolabeled, lanes 4-6), crRNAs with 5'- phosphate, 3' hydroxyl (5' radiolabeled, lanes 7-9). Free crRNAs and specific crRNP shifts are indicated on left.

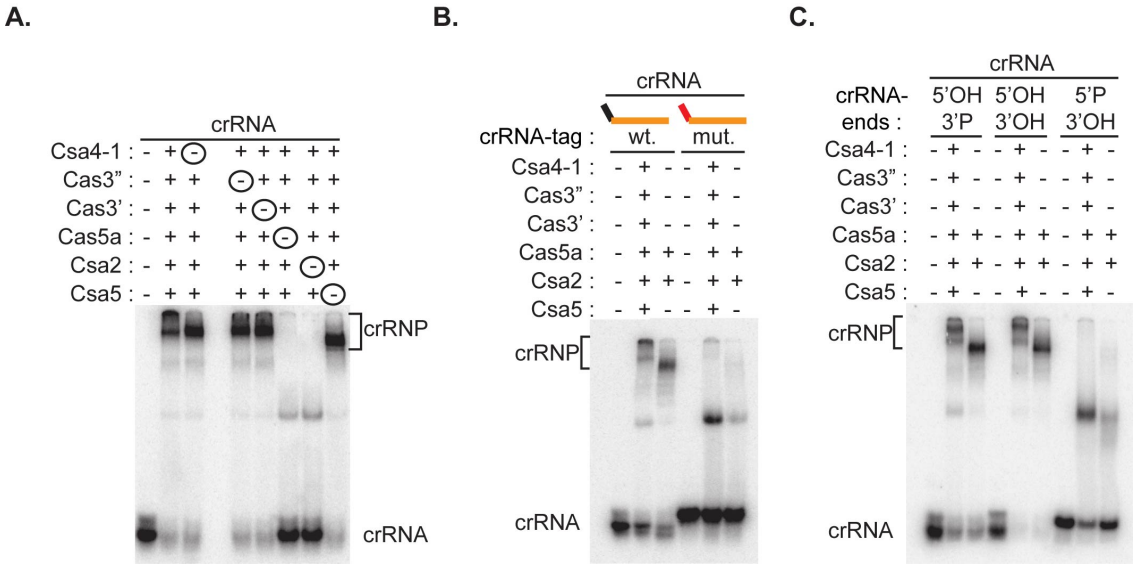

A.
crRNA
Csa4-1 :
Cas3" :
Cas3' :
Cas5a :
Csa2 :
Csa5 :
crRNP
crRNA
B.
crRNA
crRNA-tag :
wt.
mut.
Csa4-1 :
Cas3" :
Cas3' :
Cas5a :
Csa2 :
Csa5 :
crRNP
crRNA
C.
crRNA
crRNA-
ends :
5'OH
3'P
5'OH
3'OH
5'P
3'OH
Csa4-1 :
Cas3" :
Cas3' :
Cas5a :
Csa2 :
Csa5 :
crRNP
crRNA

Figure 3.8. Mapping interactions of Csa proteins at crRNA regions and by UV crosslinking. SDS gel (G-250) showing total proteins and aligned phosphor images (rad) showing the crosslinked protein profiles of six Csa proteins (Csa4-1, Cas3", Cas3, Cas5a, Csa2, Csa5) and two Csa (Cas5a, Csa2) with 4-SU crRNAs, radio-labeled (^{32}P, marked with asterisks in schematic) either all across (crRNA- body) or in 5'-tag (crRNA-tag), or in the guide region (crRNA-guide). 5' tag (black) and guide (orange) regions of crRNAs tested for each set are indicated on top. RNAse T1 and/or RNase A used in the reactions are indicated in SDS (G250) gels. The aligned protein bands on G-250 and rad images are marked with asterisks.

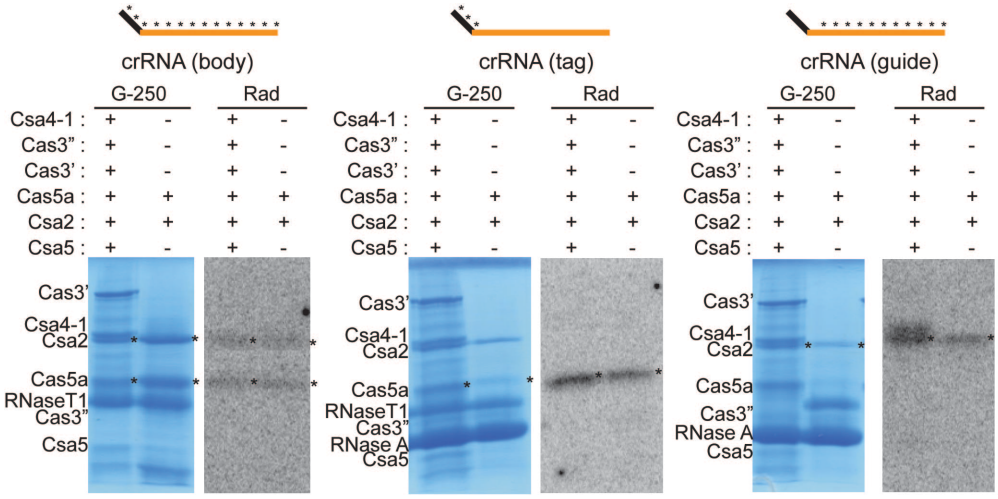
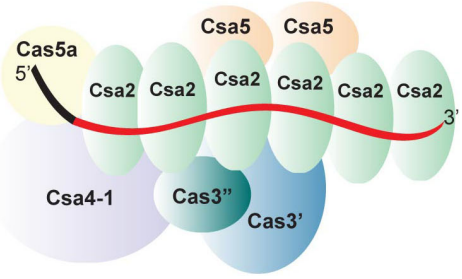

crRNA (body)
crRNA (tag)
crRNA (guide)
G-250
Rad
Csa4-1 :
Cas3" :
Cas3' :
Cas5a :
Csa2 :
Csa5 :
Cas3'
Csa4-1
Csa2
Cas5a
RNaseT1
Cas3"
RNase A
Csa5

Figure 3.9. Mechanistic model of Type I-A mediated DNA targeting.

A. Six proteins: Csa4-1 (purple), Cas3" (teal), Cas3' (blue), Cas5a (yellow), Csa2 (green), Csa5 (orange) form the Csa crRNP architecture. Cas5a contacts 5'-tag (black) and Csa2 (Cas7) binds along the guide (red) of crRNA and form the core. **B.** Csa4-1 recognizes and interacts with PAM (cyan), leading to Csa crRNP-mediated DNA binding via crRNA-guide basepairing with target strand of protospacer. Strand invasion results in displacement of non-target strand (R-loop formation) and Cas3" nuclease makes ATP-independent cleavages in the protospacer region of displaced non-target strand. ATP-dependent, Cas3'–mediated unwinding results in Cas3" catalyzed cleavages outside protospacer of non-target strand in 3' to 5' direction. ATP-dependent, Cas3'-mediated unwinding of target strand and crRNA conformation changes contributes to Cas3" catalyzed cleavages of target strand in 3'-5' direction.

A. Csa crRNP complex architecture



B. crRNA guided-PAM dependent DNA targeting and cleavage by Csa crRNP

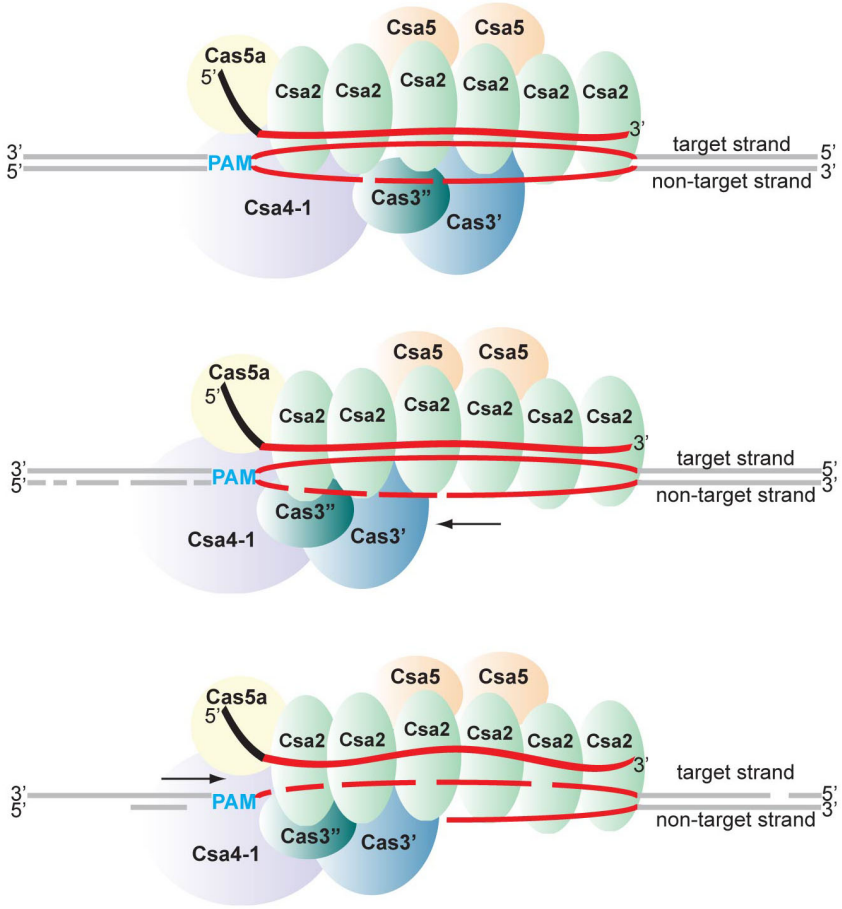
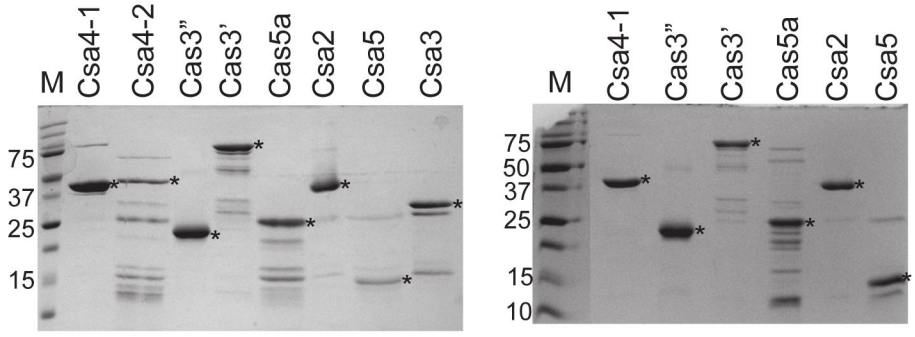


Figure 3.S1. A. SDS PAGE showing all eight Csa proteins: Csa4-1, Csa4-2, Cas3", Cas3', Cas5a, Csa2, Csa5, Csa3. The second gel shows six Csa proteins (Csa4-1, Cas3", Cas3', Cas5a, Csa2, Csa5) that were used in this paper. Precision plus dual color standard (Biorad) is used for size estimation and labeled M. Labeled sizes are in Kilo-Daltons (KD). Protein bands correspond to expected size. **B.** Effect on reconstituted Csa crRNP mediated DNA cleavage without any protein (-), with all eight Csa proteins, and combinations of seven, six Csa protein (either Csa4-2, Csa3 or both excluded, marked by minus with a circle). 5' radiolabeled DNA size-standard (10-100 nt) is labeled (M).

A.


M
Csa4-1
Csa4-2
Cas3”
Cas3’
Cas5a
Csa2
Csa5
Csa3
75
37
25
15
M
Csa4-1
Cas3”
Cas3’
Cas5a
Csa2
Csa5
75
50
37
25
15
10

B.

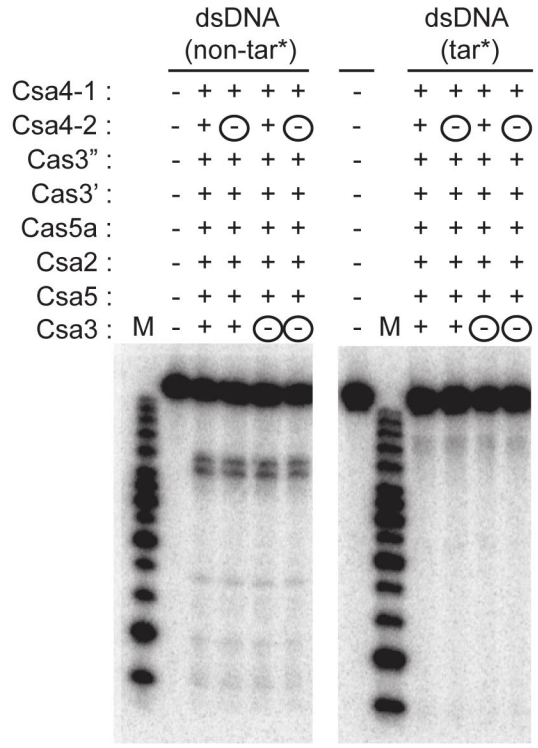
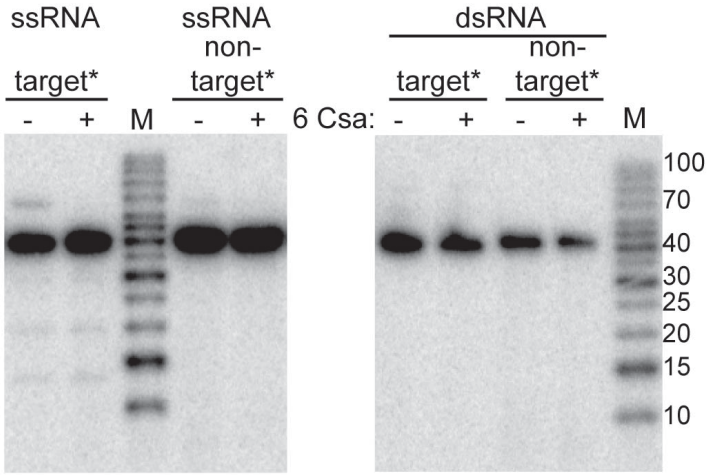

dsDNA (non-tar*)
dsDNA (tar*)
Csa4-1 :
Csa4-2 :
Cas3” :
Cas3’ :
Cas5a :
Csa2 :
Csa5 :
Csa3 :
M
M

Figure 3.S2. Type I-A complex does not cleave RNA.

A. Testing RNA (single-stranded or ss, double stranded or ds) as substrate for Csa crRNP mediated cleavage. **B.** Assaying cleavage of DNA (single-stranded or ss, double stranded or ds) without protein (-), with Cas3" alone (-/+ unlabeled crRNA) and with Csa complex with 6 proteins (-/+crRNA). Either target or non-target strands are 5' radiolabeled (^{32}P) as labeled and indicated with asterisk.

A.


ssRNA
target*
ssRNA
non-
target*
dsRNA
target*
dsRNA
non-
target*
- + M - + 6 Csa: - + - + M
100
70
40
30
25
20
15
10

B.

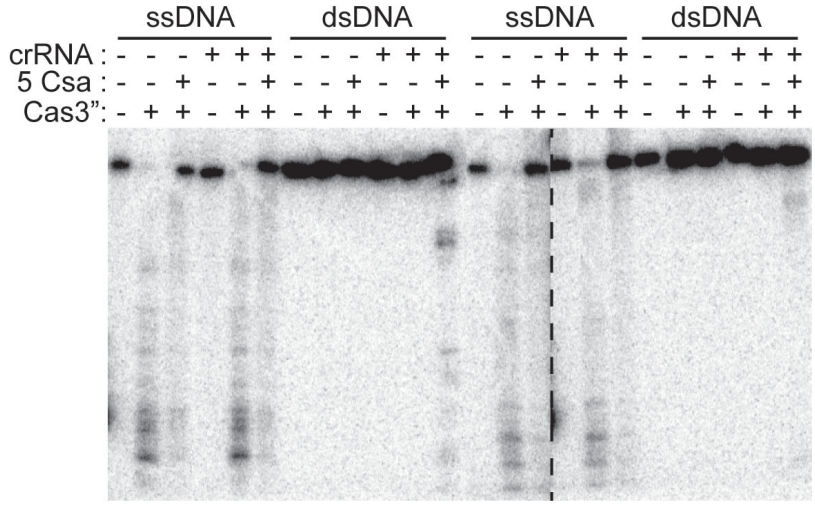

ssDNA
dsDNA
ssDNA
dsDNA
crRNA : - - - + + + - - - + + + - - - + + + - - - + + +
5 Csa : - - + - - + - - + - - + - - + - - + - - + - - +
Cas3": - + + - + + - + + - + + - + + - + + - + + - + +

Figure 3.S3. Effect of ATP (-/+) on DNA binding, without (-) or with (+) Csa crRNP. Either non-target or target strand of DNA is 5' radiolabeled (^{32}P) as labeled on top (indicated with asterisk).

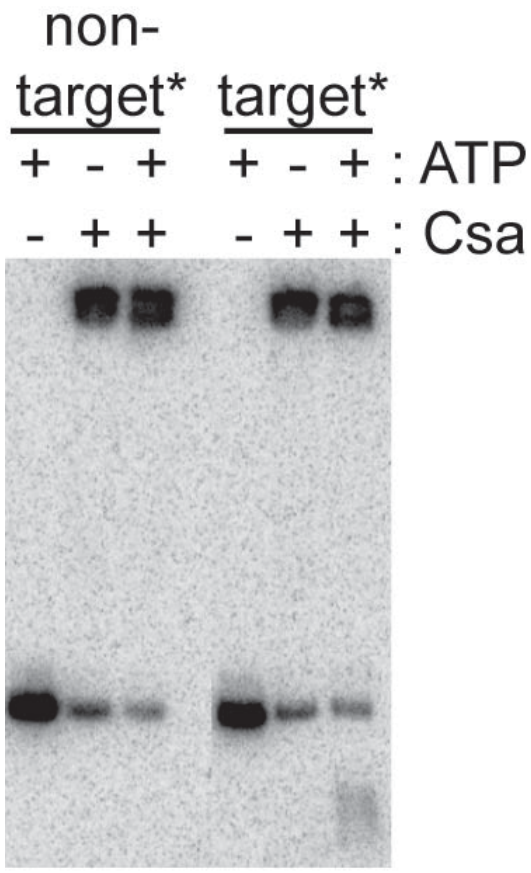

non-
target*
target*
+ - + + - + : ATP
- + + - + + : Csa

Figure 3.S4. **A.** Gel mobility shift assay to visualize crRNPs formed with 3'-radiolabeled (^{32}P) crRNA and no proteins (-), six individual Csa proteins (Csa4-1, Cas3", Cas3', Cas5a, Csa2, Csa5), all six Csa proteins and two Csa proteins (Csa5a, Csa2). Free crRNA and specific crRNP shifts are indicated. **B.** Gel mobility shifts showing the crRNPs formed during UV crosslinking (Figure 8) with no proteins (-),six Csa proteins Csa4-1, Cas3", Cas3', Cas5a, Csa2, Csa5) and two Csa proteins (Csa5a, Csa2) and different 4-SU crRNAs; radio-labeled (^{32}P, marked with asterisks) either all across (crRNA- body) , 5' tag (crRNA-tag), guide region (crRNA-guide). 5' tag (black) and guide (orange) regions of crRNA are marked.

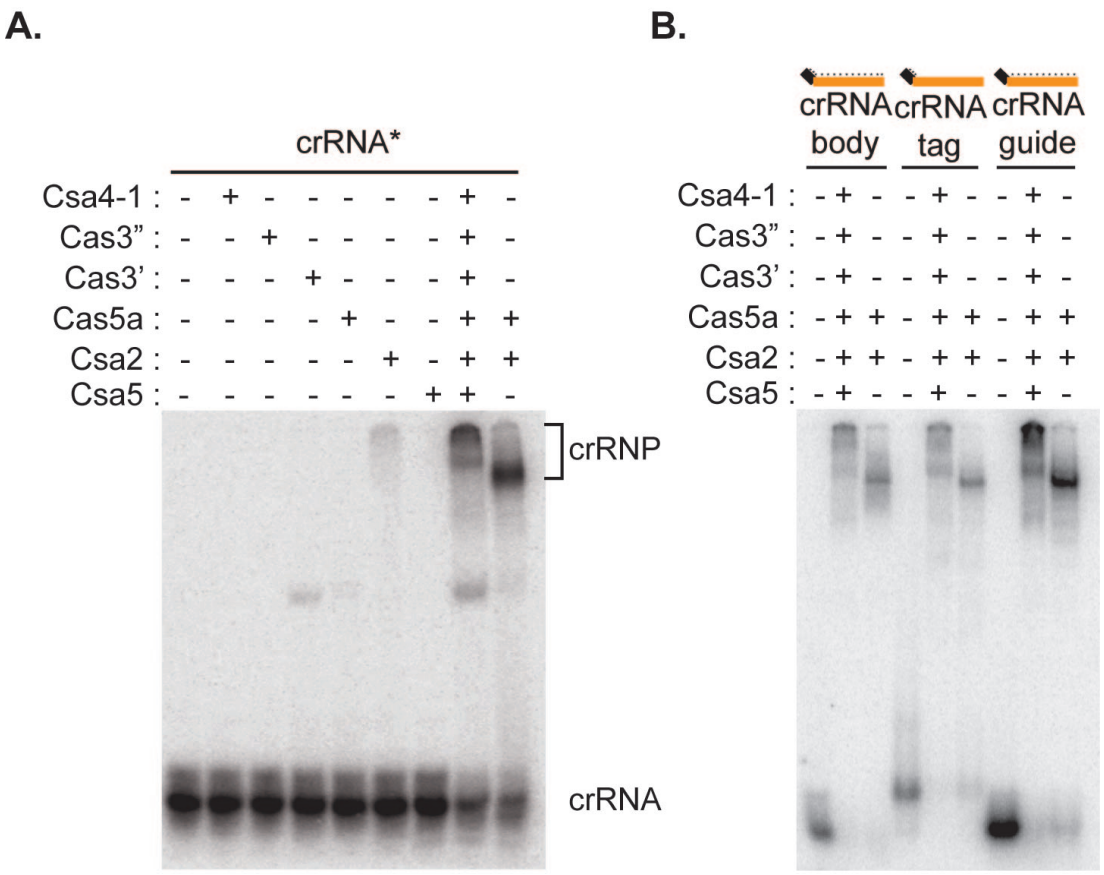
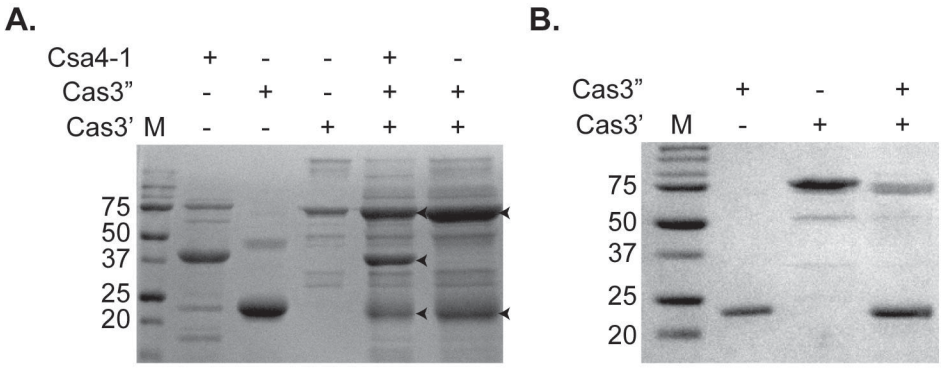

A.
crRNA*
Csa4-1 : - + - - - - - + -
Cas3" : - - + - - - - + -
Cas3' : - - - + - - - + -
Cas5a : - - - - + - - + +
Csa2 : - - - - - + - + +
Csa5 : - - - - - - + + -
crRNP
crRNA
B.
crRNA body
crRNA tag
crRNA guide
Csa4-1 : - + - - + - - + -
Cas3" : - + - - + - - + -
Cas3' : - + - - + - - + -
Cas5a : - + + - + + - + +
Csa2 : - + + - + + - + +
Csa5 : - + - - + - - + -

Figure 3.S5. Csa4-1 (Cas8a2), Cas3", Cas3' interaction studies.

SDS polyacrylamide gel showing interaction between Csa4-1 (untag), Cas3" (untag), Cas3' (His-tagged) and Cas3" (untag), Cas3' (His-tagged) affinity purified over Ni-column **(A)**; interaction between Cas3" (untag)+Cas3' (His-tagged) affinity purified followed by gel filtration **(B)**. Arrows indicate protein bands of expected molecular weight, suggesting true interaction. Single proteins are shown as size controls. Precision plus dual color standard (Biorad) is used for size estimation and labeled M. Labeled sizes are in Kilo-Daltons (KD). Protein bands correspond to expected size.


A.
Csa4-1 + - - + -
Cas3" - + - + +
Cas3' M - - + + +
75
50
37
25
20
B.
Cas3" + - +
Cas3' M - + +
75
50
37
25
20

CHAPTER 4

CRISPR RNA-GUIDED, CST COMPLEX OF *PYROCOCCUS FURIOSUS* GUIDES CAS3 NUCLEASE OF TYPE I-G SYSTEMS TO CLEAVE DNA TARGETS

---

[1]Sonali Majumdar, [1]Marianne Ligon, William C. Skinner, Rebecca M. Terns, Michael P. Terns. To be submitted to *Nucleic Acids Research*.

Abstract

CRISPR-Cas systems are RNA-based immune systems protecting prokaryotes against genome invaders like viruses and plasmids, etc. CRISPR RNAs (crRNAs) contain invader-derived sequences and associate with Cas protein subtypes to form multiple immune-effector complexes (Type I, II, III). We have previously identified three native, effector complexes (crRNPs) in *Pyrococcus furiosus*: Type I-A (Csa), Type I-G (Cst) and Type III-B (Cmr). Our previous work investigated Type III-B mediated homology-dependent cleavage of target RNA in Pfu. In the current study, we have reconstituted Type I-G (Cst) crRNP and elucidated key elements for crRNP formation. Crosslinking and gel shift assays show Cas5t and Cst2 (Cas7) interact with crRNA to form the minimal Cst crRNP. Cst crRNP (consisting Cas5t, Cst2, Cst1 (Cas8a1)) binds double-stranded DNA targets and recruits associated Cas3 to cleave DNA in the protospacer region of non-target strand. DNA binding and cleavage is crRNA-guided and PAM dependent. Cst crRNP recognizes a single stranded PAM (5'-NGG-3' on target strand) for DNA binding (and cleavage?). Site-directed mutagenesis (HD residues) of Cas3 results in loss of cleavage. Finally, we show Cst crRNP acts as a surveillance complex and can recruit Type I-G-associated Cas3 from *Thermococcus kodakerensis* for DNA cleavage.

Introduction

CRISPR (Clustered Regularly-Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated) systems are RNA-based adaptive immune systems found in bacteria and archaea, defending them against viruses, plasmids and other genome invaders. A CRISPR

locus in the host genome consists arrays of multiple repeats of 30-40 nucleotides (nt) interspersed by similar-sized, "invader-derived" variable sequences (Grissa et al. 2007). These variable sequences are acquired from genomic regions of invaders called protospacers during adaptation step. CRISPR loci are transcribed to form precursor CRISPR RNAs (crRNAs), which are processed to smaller mature crRNAs during biogenesis. Mature crRNAs contain invader-derived (guide) sequences flanked on 5' and/ or 3' ends by repeat-derived (tag) sequences. The crRNA guide sequence mediates homology-dependent silencing of invader DNA or RNA. Cas proteins provide enzymatic machinery for protospacer acquisition into host CRISPR locus (adaptation), crRNA biogenesis and processing and invader silencing. (Recently reviewed in (Terns and Terns 2011a; Barrangou and Horvath 2012; Sorek et al. 2013; Gasiunas et al. 2014; Heler et al. 2014; van der Oost et al. 2014). CRISPR-Cas immune systems are grouped into Type I (A-G), Type II (A-C) and Type III (A-B) systems primarily based on Cas protein composition (Haft et al. 2005a; Makarova et al. 2006b; Makarova et al. 2011c; Vestergaard et al. 2014). Each system corresponds to a Cas protein subtype (for e.g. Type I-A for Csa, Type I-E for Cse, Type III-B for Cmr). Mature crRNAs associate with Cas subtype proteins to form effector complexes (crRNPs) that mediate silencing of invader RNA (Type III-B) or DNA targets (Types I, II, III-A and some III-B) (Brouns et al. 2008; Marraffini and Sontheimer 2008; Hale et al. 2009; Garneau et al. 2010; Cady and O'Toole 2011b; Gasiunas et al. 2012b; Jinek et al. 2012; Magadan et al. 2012; Nam et al. 2012b; Westra et al. 2012c; Zhang et al. 2012c; Deng et al. 2013a; Maier et al. 2013b; Mulepati and Bailey 2013b; Sinkunas et al. 2013a; Staals et al. 2013a; Plagens et al. 2014b; Zebec et al. 2014a).

Type I systems have a similar *cas* content and crRNP composition and organization. Characterized Type I crRNPs contain members of Cas5, Cas7, Cas8 (large subunit) and small subunit (sometimes fused to the large subunit as a domain) superfamily Cas proteins as structural components (Brouns et al. 2008; Lintner et al. 2011d; Makarova et al. 2011c; Wiedenheft et al. 2011a; Nam et al. 2012b; Plagens et al. 2012a; van Duijn et al. 2012a; Brendel et al. 2014; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014) (Majumdar et al., unpublished). Cas3 superfamily proteins that contain nuclease (HD) and helicase (DExH) domains are typically the effector nucleases of DNA silencing Type I effector complexes (Cady and O'Toole 2011b; Westra et al. 2012c; Mulepati and Bailey 2013b; Sinkunas et al. 2013a; Huo et al. 2014a; Plagens et al. 2014b). Initially Cst and Csh subtypes (Haft et al. 2005a) were both grouped as Type I-B systems, based on their Cas superfamily protein similarities (Makarova et al. 2011c). However, recent *in silico* analysis indicates considerable diversity between large-subunit proteins of archaeal Cst systems (Cas8a1, including *Pyrococcus furiosus*) and Csh systems (Cas8b). Therefore, it is proposed that Cst subtype be classified into a separate CRISPR-Cas immune system (Type I-G). In support of the evidence for the proposed nomenclature, we refer to Pfu Cst as Type I-G system in this work. Type I-G (Cst) systems are three Cas subtype-component systems that contain genes encoding Cas5t, Cas7 (Cst2) and Cas8a1 (Cst1). Type I-G systems have a gene encoding Cas3 adjacent to *cst* genes, which is predicted to co-function like other characterized Type I systems.

Pyrococcus furiosus (Pfu) contains seven CRISPR loci that encode approximately 200 crRNAs. The *cas* genes are organized in two separate operons and encode for adaptation proteins (Cas1, Cas2, Cas4), crRNA biogenesis protein (Cas6) and three

effector modules (Csa or Type I-A, Cst or Type I-G and Cmr or Type III-B) (Figure 4.1A). We have previously identified three distinct native crRNP complexes in Pfu (Csa (Type I-A), Cst (Type I-G) and Cmr (Type III-B)) (Hale et al. 2012a) (Majumdar et al. unpublished). All three effector complexes associate with mature crRNAs from all seven CRISPR loci. Pfu Type III-B (Cmr) mediates crRNA-guided cleavage of target RNAs *in vitro* (Hale et al. 2009; Hale et al. 2012a). Based on parallels with other Type I systems and genetic evidence of DNA targeting by Type I-A and I-G systems in closely related *Thermococcus kodakarensis* (Tko) *(Elmore et al. 2013a)*, we predict Csa and Cst crRNPs are co-existing and co-functioning effector complexes targeting invader DNA in Pfu. In this study, we have reconstituted an active Type I-G or Cst crRNP with synthetic crRNA and recombinant Cst proteins. We delineated crRNA features and protein components for Cst crRNP assembly. We investigated elements of target DNA important for Cst crRNP-mediated DNA targeting. Like other characterized Type I systems, Cst crRNP serves as a surveillance complex and binds non-self DNAs in a crRNA homology-dependent manner. Cas3 encoded adjacent to *cst* genes in Type I-G system of Pfu (Figure 4.1A) mediates cleavage of DNA bound by Cst crRNP, thereby indicating Cas3 co-functions with Cst crRNP. We further presented evidence that Cst crRNP specifically binds DNA targets and can direct Cas3 associated with Type I-G system of Tko for DNA cleavage.

Results

Proteins required for Cst crRNP formation

To investigate composition of Cst crRNP, we reconstituted complex/es with recombinant Cst proteins (Cas5t, Cst2 or Cas7, Cst1 or Cas8a1) (Figure 4.1 and 4.S1) and

a synthetic 45-nt crRNA, containing 5' 8-nt repeated-derived tag and 7.01 guide sequence (from first invader-derived sequence of CRISPR locus 7). This crRNA species was detected in native Pfu Cst complex through Northern blotting and RNA sequencing (Majumdar et. al, unpublished). To determine which proteins were required for crRNP formation, we did gel mobility shift assay with 3' radiolabeled crRNA and different combinations of Cst proteins (Figure 4.1B and 4.1C). Neither of the individual Cst proteins (Cas5t (Cas5), Cst2 (Cas7) and Cst1 (Cas8a1) or Cas3 (encoding gene adjacent to *cst* (Figure 1A)) bound the crRNA (Figure 4.1B). crRNP shifts were seen with Cas5t (Cas5) and Cst2 (Cas7) but none of the other combinations of 2 Cst proteins (Figure 4.1B and 4.1C). All three Cst proteins (Cas5t, Cst2 or Cas7 and Cst1 or Cas8a1) form stronger / tighter crRNP shifts. Adding Cas3 did not affect the shifts formed with all 3 Cst proteins (Figure 4.1B and 4.1C).

crRNA features for assembly of Cst crRNP

We further sought to understand the features of crRNA, essential for Cst crRNP formation. In Pfu, Cas6 endonuclease cleaves all seven CRISPR RNA precursor transcripts within the repeats to generate IX crRNA intermediates that contain 8-nt 5' tag (AUUGAAAG) (Carte et al. 2008b; Carte et al. 2010). Similar to some Type I and Type III systems (Hale et al. 2008; Lillestol et al. 2009; Deltcheva et al. 2011; Hatoum-Aslan et al. 2011b; Bernick et al. 2012; Hale et al. 2012a; Plagens et al. 2012a; Richter et al. 2012; Zhang et al. 2012c; Hatoum-Aslan et al. 2014) (Majumdar et al., unpublished), further processing of 3' repeats by unknown mechanisms generate mature crRNAs in Pfu (Hale et al. 2008). Our previous analyses revealed majority of native Cst complex-

associated crRNAs bear Cas6-generated, repeat-derived 5' tag (Majumdar et. al, unpublished). Therefore, we tested if this 8-nt 5'-tag is important for crRNP formation. We mutated the sequence of the 5' tag and compared binding of crRNAs that contain wild-type tag with crRNAs that contained mutated tag by 2, 3, 4 Cst proteins (all combinations which showed crRNP formation in Figure 4.1). We observed that none of the protein combinations bound crRNA that contained mutated-tag (Figure 4.2A). Therefore Cas6-genertated 5'tag is a critical component for recognition and assembly of Cst crRNPs.

Cas6 cleavage generates 1X intermediate crRNAs with 5' OH and 3' P (presumably 2'-3' phosphate). These are further processed on 3'- end to mature crRNAs that associate with the effector complexes (Csa, Cst and Cmr). We wanted to uncover what chemical ends of crRNA were amenable for Cst crRNP assembly. We either 5' end-labeled (^{32}P) to generate 5'P and 3'OH bearing crRNAs, or 3' end-labeled (^{32}P) 3'P and 5'OH bearing crRNAs, or internally-labeled (^{32}P) to generate 5'OH and 3'OH bearing crRNAs. We tested each crRNA with different Cst protein combinations (2 Cst, 3 Cst, 4 Cst) (Figure 4.2B). All different Cst protein combinations bound crRNA with 3' labeled or internally labeled crRNAs (containing 5' OH), but not 5' labeled crRNAs to form Cst crRNPs were formed with. This indicates Cas6-generated 5' OH on crRNAs is critical for Cst complex assembly but both hydroxyl and phosphate groups on 3' end are amenable.

Cas5t and Cst2/Cas7 make direct interactions with crRNA

We wanted to further investigate which Cst protein/s directly contact the crRNA-tag and crRNA-guide regions. We performed UV crosslinking of internally radiolabeled

(^{32}P), 4-thiouridine containing crRNA or 8-nt 5' tag assembled with 2 Cst (Cas5t+Cst2) or all 3 Cst proteins (Cas5t+Cst2+Cst1) (Figure 4.3). To visualize interactions at specific crRNA regions (tag or guide), we selectively radiolabeled the desired region. Cas5t (Cas5) and Cst2 directly contact crRNA labeled in both tag and guide region (crRNA body), as well the labeled 8-nt 5' tag. The profiles of crRNP shifts formed in the crosslinking reactions are shown and look similar to the shifts seen with end-labeled crRNA (Figure 4.3A). Additionally, we saw Cas5t+Cst2 can bind the stand-alone 8-nt 5' tag. More Cst2 than Cas5t crosslinked to the entire crRNA (crRNA body), but we saw a higher signal of Cas5t than Cst2 crosslinked with the 8-nt 5' tag (Figure 4.3B). Only Cst2 crosslinked with the crRNA labeled at the guide region (crRNA guide). There was no evidence of Cst1 crosslinking to any region of the crRNA or 5' tag, as all samples containing two Cst proteins had identical crosslinking patterns to samples containing all three Cst proteins. These results indicate that Cst2 (Cas7) contacts the crRNA guide region, while Cas5t (Cas5) primarily contacts the 5' tag with some interaction coming from Cst2 (Cas7).

All three Cst proteins and crRNA are required for DNA binding

Similar to effector complexes of other Type I systems, we predicted Cst crRNP targets invader DNA. DNA-silencing effector complexes of Type I systems recognize a signature motif called Protospacer Adjacent Motif (PAM) that are found directly upstream or downstream of protospacers of invader DNA (Deveau et al. 2008; Mojica et al. 2009; Fischer et al. 2012b; Elmore et al. 2013a; Shah et al. 2013; Hochstrasser et al. 2014a; Nishimasu et al. 2014; Sternberg et al. 2014). Therefore, PAMs serve as a

recognition signature for differentiating self vs non-self DNA. Based on bioinformatics prediction of PAM-association with repeat classes and Cas subtypes (Kunin et al. 2007; Mojica et al. 2009) and functional PAM in CRISRP-Cas mediated invader DNA targeting in Tko (Elmore et al. 2013a), we predict 5'-NGG-3' (crRNA complementary DNA target strand) as the functional PAM for Type I-G system in Pfu. We tested double-stranded DNA that mimics invader DNA containing the 7.01 protospacer sequence and including the predicted PAM sequence (5'-NGG-3' on the target strand and 5'-CCN-3' on the non-target strand; Figure 4.4A). To determine the minimal protein components important to recognize and bind this DNA, we performed DNA binding assay with crRNPs formed with different combinations (2,3,4) of Cst proteins. The minimal crRNP formed with Cas5t (Cas5) and Cst2 (Cas7) did not bind DNA, but crRNP formed with all three Cst proteins (Cas5, Cas7, Ca8a1) bound both strands of DNA (Figure 4.4B). The addition of Cas3 to the DNA-binding Cst crRNP shows additional bigger shift. Furthermore, removing crRNA from the reactions abolished DNA binding. We conclude that although Cas5t and Cst2 (Cas7) form a minimal crRNP, crRNP formed with all three Cst proteins (Cas5t, Cst2, and Cst1) are required to bind DNA targets.

Cst crRNP directs Cas3 to cleave DNA

We further investigated the minimal components required to cleave the bound DNA in either target or non-target stranded (Figure 4.4C). Our data shows that while crRNPs containing 3 Cst proteins (Cas5t, Cst and Cst1) can bind DNA, it is not sufficient to cleave the DNA. Similar to other Type I systems, the Cas3 associated with Type I-G system in Pfu, is the predicted effector nuclease and likely an important component for

DNA cleavage. As predicted, addition of Cas3 to Cst crRNP shows cleavage of bound double-stranded DNA in the non-target strand (Figure 4.4C, lane 4), but no cleavage products were observed in the target strand. Cas3 by itself did not cleave double-stranded DNA (Figure 4.4C, lane 6). No cleavage is seen in the absence of crRNA, due to loss of DNA binding (Figure 4.4C, lane 5). Therefore, Cst crRNP targets DNA by binding and recruits Cas3 nuclease to cleave. The cleavages were only seen in non-target strand of double stranded DNA, but not in target strand (complementary to crRNA guide).

Cst-Cas3 mediated DNA binding and cleavage is crRNA-guided and PAM dependent

In order to investigate which components of DNA determine specificity of Cst crRNP-Cas3 mediated DNA binding and cleavage, we made mutations either in the protospacer or PAM regions (Figure 4.5A). We mutated 5'-GGG-3' PAM on target strand to 5'-CCC-3' and made compensatory mutations on non-target strand to maintain complementarity (mut. PAM, Figure 4.5A). Cst crRNP mediated DNA binding and Cst crRNP-Cas3 mediated cleavage of non-target strand of DNA was lost when PAM was mutated (wt. PAM vs. mut. PAM, Figure 4.5B, 4.5C). Next, we kept the PAM wild type and made random mutations along the protospacer region of DNA, such that crRNA guide bears no homology to the protospacer (mut. protospacer, Figure 4.5A). Both Cst crRNP-Cas3 mediated DNA binding (target and non-target strand, Figure 4.5B) and cleavage of DNA (non-target strand, Figure 4.5C) was abolished. Therefore Cst -Cas3 mediated DNA binding and cleavage is guided by crRNA homology and dependent on PAM sequence.

Cst crRNP recognizes a single-stranded PAM sequence for DNA targeting

We sought to delineate whether Cst crRNP recognizes single-strand or double-stranded PAM for DNA binding and Cas3-mediated cleavage. To analyze this, we made mutations on one strand of the PAM and did not make compensatory mutations in the other strand leaving a region of non-complementarity in the PAM region (G-G PAM, C-C PAM; Figure 4.5A). When we mutated the PAM on target strand (5'-GGG-3' to 5'CCC-3') but not on non-target strand (5'-CCC-3') of DNA, Cst crRNP-Cas3 did not bind or cleave the DNA substrate (C-C PAM, Figure 4.5B, 4.5C). Cst crRNP could bind DNA with wild-type PAM on target strand (5'GGG-3') and mutations in the non-target strand PAM (5'-CCC-3 to 5'-GGG-3') (G-G PAM, Figure 4.5B). The efficiency of binding was comparable to DNA with wild-type PAM. When we tested Cst crRNP-Cas3 mediated cleavage of G-G PAM substrate, we observed a different cleavage pattern (not sure if I should mention this although it looks apparently different) and reduction in efficiency (G-G PAM, Figure 4.5C). There was more accumulation of intermediate products that did not process down to the major product (~50 nts) seen with wild-type PAM substrate. Our data suggests Cst crRNP recognizes PAM sequence on target strand for DNA binding. Cst crRNP-Cas3 can cleave DNA with single-stranded PAM on target strand although the cleavage profile is different.

Cst crRNP can recruit related Cas3 effector nucleases for DNA cleavage

Our data indicates Cst cRNP binds DNA targets and recruits Cas3 for DNA cleavage (Figure 4.4). We wanted to investigate whether Cst crRNP can recruit Cas3 associated with Type I-G system of a different organism to cleave DNA. We made recombinant

Cas3 associated with Type I-G (Cst) system of Tko (Figure 4.S2A). We tested Cst crRNP mediated DNA cleavage by using either Pfu or Tko Cas3 (Figure 4.6A). Cst crRNP binds DNA with similar efficiency with either Cas3 (Figure 4.S2B) and can recruit either Cas3 to cleave DNA target (Figure 4.6A; lanes 2,4). Tko Cas3 appears to exhibit more efficient cleavage.

Conserved residues of nuclease domain of Pfu and Tko Cas3 are critical for Cst crRNP directed DNA cleavage

Cas3 are signature proteins of Type I CRISPR-Cas systems containing nuclease-helicase domains. Histidine and aspartate (HD) amino acids of nuclease domain of characterized Cas3s are catalytic residues for DNA cleavage (Mulepati and Bailey 2013b; Sinkunas et al. 2013a; Plagens et al. 2014b) add belzegova. We performed site-directed mutagenesis (H53, D54-AA) of both Pfu and Tko Cas3 and expressed and purified the mutant proteins (Figure S2A). Cst crRNP mediated DNA cleavage was lost with either Pfu or Tko mutant Cas3 (Figure 4.6A; lanes 3,5). Cst crRNP bound DNA with similar efficiency with either Pfu or Tko mutant Cas3s and the loss of cleavage was not due to any reduction in binding (Figure 4.S2B). Our data suggests that Cst crRNP can recruit various Type I-G associated Cas3 effector nucleases to cleave DNA. HD amino acids of both Cas3s form the catalytic basis of cleavage.

Mapping DNA cleavage by Cst crRNP -Pfu Cas3 or Cst crRNP -Tko Cas3

We pursued to further understand the molecular mechanism of DNA cleavage meditated by Cst crRNP and catalyzed by Cas3. We mapped and compared the sites of

cleavages by both Pfu Cas3 and Tko Cas3 following recruitment by Pfu Cst crRNP (Figure 4.6B). Although the relative abundance of cleavage products by both Cas3s is different, the cleavages map to similar positions. All cleavages occur in the protospacer region of non-target strand. The largest cleavage products by both Cas3s are ~66 nts (~10 nts from the 3' end of the protospacer). The smallest cleavages seen with Pfu Cas3 are the most abundant (~47 nt, ~48 nt and ~50 nt). The same cleavages are seen with Tko Cas3 along with smaller cleavages mapping close to PAM sequence on non-target strand (~2-5 nts from PAM sequence). With Tko Cas3 there are detectable levels of intermediate sites of cleavage between 50 and 66 nts that are low abundant in cleavages by Pfu Cas3. In summary both Pfu and Tko Cas3 cleave the DNA bound by Cst crRNP in similar sites but show some distinct properties.

Discussion

Molecular features of Type I-G (Cst) crRNP

Our crRNP reconstitution results (Figure 4.1 and 4.2) indicate the structural core of Cst crRNP in Pfu is made of crRNA and Cas5t, Cas7 (Cst2). Crosslinking results further delineate that Cas7 (Cst2) interacts with the guide region of crRNA whereas Cas5t primarily contacts the 8-nt 5' tag with some Cas7 (Cst2) interaction (Figure 4.3B). We have some evidence of protein-protein interaction between Cas5t and Cst2 (Cas7) (Figure 4.S3). Similar to characterized structures of Type I and Type III crRNPs (Wiedenheft et al. 2011a; van Duijn et al. 2012a; Rouillon et al. 2013; Spilman et al. 2013a; Staals et al. 2013a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014), we expect Cst2 (Cas7) multimers binds along the crRNA and an interaction between Cas5t, Cst2 (Cas7)

and 5’ tag caps the 5’ end. Crosslinking results do not show any evidence of direct interaction between Cas8a1 (Cst1) and crRNA (Figure 4.3). Yet the crRNP shifts formed with addition of Cas8a1 (Cst1) do show some distinct properties (Figure 4.1, 4.2). Cst1 may contribute to crRNP structure via protein-protein interactions. Cas5t and Cst2, but not Cst1 were identified as stable protein components of immunopurified native Cst crRNP in Pfu (Majumdar et al., unpublished) implying these two proteins constitute the more stable components. An earlier study identified a complex consisting Cas5t, Cst2 and Cst1 isolated by series of chromatographic fractionation of Pfu cell extract (Menon et al. 2009b). Structural studies would uncover the details of the Cst crRNP organization that currently elude us.

crRNAs associated with native Type I-A, Type I-G and Type III-B effector complexes in Pfu contain a conserved Cas6 generated, 8-nt repeat-derived, 5’ tag (Majumdar et al., unpublished). Similar to Type III-B (Cmr) complex in Pfu (Hale et al. 2012a), results in this work indicate that Cas6 generated 5’ tag as well as 5’ OH chemical ends of crRNA are important for formation of both the structural core and DNA binding surveillance Cst complex. Systematic single-nucleotide mutations of the tag sequence would further our understanding about which sequences in the 5’-tag are absolutely critical for crRNP assembly and whether any mutations are tolerated.

DNA recognition and targeting by Type I-G (Cst) complex in Pfu

We have reconstituted an active Pfu Type I-G complex *in vitro* and investigated detailed mechanism of DNA targeting. Guided by crRNA, Cst complex binds non-self DNA targets that contain functional PAM (5’ NGG-3’ on target strand) (Figure 4.4A) and

recruits associated Cas3 (Figure 4.4B) to cleave the bound DNA. Our data provides evidence that Cst1 (Cas8a1, large subunit protein) is an essential component of Cst crRNP mediated DNA binding. Our results implicate functional parallels with *E. coli* Type I-E (Cse) crRNP in which the large subunit protein Cse1 scans and recognizes invader DNA by interacting with PAM (Sashital et al. 2012b). Cse1 in turn positions Cas3 adjacent to PAM (in the protospacer) and activates Cas3 to ensure DNA cleavage (Westra et al. 2012c; Hochstrasser et al. 2014a; Huo et al. 2014a). We expect Cst1 (Cas8a1) to scan and interact with PAM and recruit Cas3 to the protospacer site for cleavage (Figure 4.7).

Here, we investigated the requirement of a functional PAM for DNA binding by Cst crRNP and cleavage by Cas3. Mutation of single nucleotide at PAM will give us further insights on which positions are critical for DNA recognition and which PAM mutations are tolerated. We further sought to understand whether Cst crRNP recognizes single-stranded or double-stranded PAM and designed hybrid PAM DNAs (Figure 4.5). Our results indicate Cst crRNP recognizes single-stranded PAM sequence on target strand (complementary to crRNA-guide) for DNA binding (Figure 4.5B). This is similar to single-stranded PAM (on target strand) recognition for DNA binding exhibited by *E. coli* Type I-E (Cse) crRNP (Westra et al. 2012c). A more recent study on *E. coli* Type I-E crRNP shows that while a single-stranded PAM (on target strand) is sufficient for binding, base pairing at PAM region is required for efficient cleavage (Hochstrasser et al. 2014a).Our results indicate Cst crRNP-Cas3 mediated cleavage of DNA (in non-target strand) with single-stranded PAM on target strand (Figure 4.5C, G-G) with reduced efficiency compared to double-stranded PAM (Figure 4.5C, wt. PAM). We cannot rule

out the possibility that non-complementarity (G-G) at PAM may cause distortion of DNA duplex-structure, thereby resulting in loss of cleavage efficiency. We need to investigate the effect of PAM sequence mutations without compromising duplex formation to elucidate whether sequence or structure is the determinant feature for DNA cleavage. Also, structural studies would provide further understanding in this regard.

DNA binding and cleavage depended on homology with guide of crRNA (Figure 4.5, mut. protospacer) and DNA substrates with non-cognate protospacer (no homology to crRNA guide) was not bound or cleaved. In characterized Type I-B, Type I-E and Type I-F base-pairing between crRNA "seed" sequence (downstream of 5'-tag in the guide region) and the PAM-proximal protospacer sequences on target strand of DNA is a minimal determinant of CRISPR-Cas mediated DNA silencing (Jore et al. 2011a; Semenova et al. 2011; Wiedenheft et al. 2011a; Cady et al. 2012; Maier et al. 2013b). Whether there is a minimal PAM proximal protospacer region (homologous to crRNA seed) critical for DNA recognition in Type I-G system of Pfu needs to be explored.

Cst crRNP directs Type I-G associated Cas3s from Pfu and Tko to cleave DNA targets

Our results provide evidence that Cas3 is the effector nuclease of Type I-G system in Pfu (Figure 4.6) and conserved his and asp residues of Cas3 nuclease domain are the catalytic components of DNA cleavage. Our earlier work showed genetic evidence of active, CRISPR-Cas dependent, invader plasmid silencing in Tko (Elmore et al. 2013a). We were curious to see if Cst crRNP could direct Cas3 associated with Type I-G system of *T. kodakerensis* (Tko) to cleave DNA targets and how the cleavages compare with the

Pfu Cas3. Type I-G associated Pfu Cas3 (PF 1120) share 58% identity in amino-acid sequence with Tko Cas3 (TK 0460). Our results indicate DNA interference by Pfu Cst crRNP can be complemented by Type I-G associated Cas3 from Tko (Figure 4.6) and possibly other organisms containing Type I-G system. It will be interesting to investigate if Cas3 of other Type I systems (like Type I-E) can be catalytic substitutes.

Mechanism of Cst crRNP-Cas3 mediated DNA cleavage

Based on our findings we propose a mechanistic model of Type I-G (Cst) mediated DNA targeting in Pfu (Figure 7). Cas5t and Cst2 (Cas7) form a minimal core crRNP. Guided by crRNA, Cas5t, Cst2 (Cas7) and Cst1 (Cas8a1, large subunit) mediate DNA binding in a PAM-dependent fashion. We predict Cst1 (Cas8a1) scans and interacts with PAM sequences and recruits Cas3 to the site of cleavage by protein- protein interaction. crRNA strand invasion and base-pairing of crRNA-guide with protospacer region of target strand of DNA results in displacement of the non-target strand (in the protospacer region). Cas3 (HD nuclease) catalyzes nickel-dependent, ATP-independent cleavages in the protospacer region of the displaced non-target strand, possibly in 3' to 5' direction. The cleavages could be either multiple endonucleolytic cleavages or exnoucleolytic degradation or a combination of both. The cleavages map to the middle of the protospacer sequence. The downstream sequence on non-target strand is probably bound by the Cst complex and therefore protected.

Progressive cleavage in target strand and non-target strands outside protospacer would potentially require ATP-dependent, Cas3 helicase-mediated unwinding of DNA, as seen in Type I-E system (Mulepati and Bailey 2013b; Sinkunas et al. 2013a). We did

not see any cleavages in the target strand of DNA or outside the protospacer region in non-target strand even on addition of ATP with Pfu Cas3 or Tko Cas3 (data not shown). We cannot exclude the possibility that our reaction conditions may be missing a cofactor required for Cas3 helicase activity. Alternatively, Type I-G (Cst)-associated Cas3 helicase may be non-functional (as seen in both Pfu and Tko). In such a scenario, nicking of DNA, followed by Cst-crRNP (and/or Cas3) could remain bound to the invader DNA and confer potent DNA silencing.

Concluding remarks: *Pyrococcus furiosus* is a model organism to investigate three distinct and co-existing CRISPR-Cas systems (Type I-A, Type I-G, Type III-B). In our previous work, we have delineated features of Type III-B (Cmr) mediated target RNA silencing. This work provides the first evidence of DNA targeting by Type I-G CRISPR-Cas immune system. We have investigated the molecular features essential for Cst core and effector complex formation and mechanism of DNA binding and cleavage.

Methods

Expression and purification of recombinant Cst proteins

Clones (pET 24) containing the genes encoding N-terminal, 6x His-tagged Cst proteins (Cas5t, Cst2, Cst1) and Cas3 (Pfu Cas3 and Tko Cas3) were transformed into BL21 RIPL *E. coli* cells. 1 to 2 liters LB cultures were grown at 37 °C under antibiotic selection, until they reached an OD_{600} between 0.4 and 0.7. Protein expression was induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) overnight at room temperature. Cells pellets were resuspended in buffer A (25 mM TRIS, 250 mM NaCl pH

7.6) with 10 mM imidazole and 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and lysed by sonication (Misonix Sonicator 3000). The lysate was incubated at 70 °C for 30 minutes with frequent mixing to precipitate heat-labile proteins, then centrifuged at 14,000 rpm, 4°C and soluble fraction was filtered (0.8 μm filter pore size Millex filter unit, Millipore).

His-tagged proteins were purified via Ni^{2+}-affinity chromatography, washed with buffer A + 20 mM imidazole, and eluted in buffer A with increasing imidazole concentration (50 mM, 100 mM, 200 mM, and 500 mM). Peak elution fractions were dialyzed in buffer A using Slide-a-lyzer mini dialysis cassettes (ThermoScientific). Cst2 (Cas7) and Pfu Cas3 were concentrated with Vivaspin 500 centrifugal filters (GE Healthcare). Proteins were quantified with Qubit 2.0 flurometer (Life Technologies) and SDS-PAGE and stored at 4°C, prior to use for functional assays.

Preparation of RNA and DNA substrates

Synthetic RNA and DNA oligo sequences are listed in table 4.1. 7.01 crRNAs (wild type tag and mutant tag) sequences were previously described (Hale et al. 2009; Hale et al. 2012a). crRNAs were 3'-end labeled (figures 4.1, 4.S1, 4.2A and 4.2B, 5'-OH, 3'-P) with T4 RNA ligase (New England Biolabs) and [α-^{32}P] pCp (3000 Ci/ mmol, Perkin Elmer) as described (Carte et al. 2010). Internally labeled 7.01 crRNA was prepared (Fig 4.2B, 5'-OH/ 3'-OH) by *in vitro* transcription of repeat-spacer unit with [α-^{32}P]-GTP (800 Ci/mmol, 10 mCi/mL, Perkin Elmer) followed by Cas6 digestion as described (Spilman et al. 2013a). crRNA (Figure 4.2B, 5'-P/ 3'-OH) and DNA oligos (MWG

Eurofins) were 5'-labeled with T4 polynucleotide kinase (PNK) and [γ-^{32}P] ATP as described (Carte et al. 2008b).

To make double-stranded DNA substrates (figures 4.4, 4.5 and 4.6), DNA oligos were annealed by mixing cold DNA and radio-labeled DNA in a molar ratio of 2:1 in 10 mM TRIS (pH 7.5 – 8), 50 mM NaCl, 1 mM EDTA and incubating at 95°C for 5 minutes followed by slow cooling for 23°C (at the rate of 1°C per minute). Annealing was confirmed and substrates were extracted from 8% non-denaturing PAGE. The annealed DNA duplex was extracted with phenol/ chloroform/ isoamyl alcohol (pH-8) and ethanol precipitation, resuspended in buffer A and stored at -20°C for future use in functional assays. Details of oligo combinations to prepare each DNA duplex are listed in table 4.2.

Gel shift assays to study crRNP formation

Cst-crRNP formation (figures 4.1, 4.2 and 4.S1) was assayed as follows. 1 µM each of recombinant Cst proteins (as indicated) were incubated with 0.01- 0.05 pmoles of radiolabeled crRNA (end labeled or internally labeled as specified) in buffer A (25 mM TRIS, pH- 7.6, 250 mM NaCl), 1.5 mM Mg, 1 unit of SUPERase-In ribonuclease inhibitor (Applied Biosystems) and 5 µg of *E. coli* tRNA (competitor) at 70°C for 30 minutes. Reactions were mixed with 10x DNA loading dye (40% glycerol, 0.5% bromophenol blue, 0.5% xylene cyanol in TBE) to a final concentration of 1X, and loaded onto a 6% native PAGE (containing 0.5X TBE and 8% glycerol). The gel was run at 110V for 70 min, dried, and exposed on a phosphor-imaging screen.

Cst crRNP-mediated DNA binding and cleavage assays

crRNPs were formed as specified above except with 0.5 pmoles unlabeled crRNA (instead of 0.01-0.05 pmoles radiolabeled crRNA). After 30 minutes incubation at 70 °C for crRNP formation, 0.01~0.05 pmoles of radiolabeled double-stranded DNA (as specified in figures 4.4-4.6) and 100 μM of Ni were added to the reactions. Samples were further incubated at 70 °C for 3 hours. To assess DNA binding, half of each reaction was mixed with 10X DNA loading dye (final concentration- 1X), and run on 6% native PAGE at 110V for 110 min, dried, and exposed on a phosphor-imaging screen.
For assaying DNA cleavage, remaining half of the reactions were treating with 1 μg of Proteinase K (New England Biolabs) and incubated at 37 °C for 15 min, followed by PCI extraction (pH 8) and ethanol precipitation and resolved on 15% denaturing (7 M urea) polyacrylamide sequencing gels. Gels were dried and the radiolabeled RNAs visualized by phosphor imaging. 5' end-labeled low molecular weight marker (10-100, Affymetrix) was used to determine the sizes of the observed products.

RNA-Protein Crosslinking

crRNA-Cst protein UV-induced crosslinking, using photoactivatable 4-thio uridine and internally (^{32}P) radiolabeled crRNAs, was performed as described (Spilman et al. 2013a) with the following modifications. crRNP reconstitution reactions were scaled up 2 times relative to standard Cst-crRNP reconstitution experiments (except no RNasin added) . A quarter of each reaction was assessed for crRNA binding by native gel electrophoresis (Figure 4.3A). Following crosslinking with 312 nm UV light, complexes were disrupted with 1% SDS treatment and boiled for 10 minutes, followed by RNAse

treatment at 37 °C for 1 hour. Reactions containing crRNA-body, 8-nt 5'-tag were treated with RNase T1 (1,000 U, Thermo) and crRNA-guide were treated with RNase A (10 µg, Thermo). Laemmli Buffer was added and the samples were boiled and separated on 12.5% SDS-PAGE gels. The gels were visualized by G-250 Coomassie staining, dried, and subjected to phosphor imaging and autoradiography to allow for alignment of protein bands with autoradiographs (Figure 4.3B).

Site-Directed Mutagenesis

To make HD-AA mutants of Pfu (His 53, Asp 54) and Tko Cas3 (His 53, Asp 54), QuikChange site-directed mutagenesis was performed as per manufacture's guidelines (Agilent Technologies). Briefly, primers were designed and ordered from MWG Eurofins (listed in table 4.1). The mutants were generated by PCR amplification of wild type Pfu *cas3* and Tko *cas3* (pET24) clones followed by DpnI digestion and transformation in to *E. coli* BL21 RIPL. Mutations were confirmed by sequencing. The proteins were expressed, purified as alongside wild type Pfu and Tko Cas3 as shown (Figure 4.S2).

Protein- protein interaction

N- terminal, 6X- His tagged Cas5t (pET 24) and untagged Cst2 (modified pET 21) were purified as mentioned before. The supernatant obtained after cell-lysis, thermal purification and centrifugation was used. 30 µg of Cas5t, 100 µg of Cst2 and 30 µg of BSA (used as internal control) were mixed and incubated at 70 °C for 30 minutes with frequent mixing. One half of the samples were precipitated with TCA/ DOC and acetone, resuspended in SDS loading buffer and used as 'input'. The other half was affinity

purified with Ni-NTA (Qiagen) by incubating with 15 µl of Nickel beads (binding capacity upto 50 mg/ ml) at room temperature for 20 minutes with end on end rotation, followed by washing four times in buffer A with 25 mM Imidazole and eluting by boiling in SDS loading dye. Reactions with individual proteins served as reference. The proteins were separated by running on 12.5% SDS-PAGE gels at 200V and visualized Coomassie R-250 staining.

Acknowledgement

We thank Lesha Spencer (and Julie Tischer) for their contribution in protein-protein interaction studies. We extend our appreciation to Dr. Caryn Hale for intellectual and technical inputs towards initial progress of the work. This work was supported by National Institutes of Health (NIH) to M.P.T. and R.M

References

Barrangou, R., and Horvath, P. (2012). CRISPR: new horizons in phage resistance and strain identification. Annual Review of Food Science and Technology *3*, 143-162.

Bernick, D.L., Cox, C.L., Dennis, P.P., and Lowe, T.M. (2012). Comparative genomic and transcriptional analyses of CRISPR systems across the genus *Pyrobaculum*. Frontiers in Microbiology *3*, 251.

Brendel, J., Stoll, B., Lange, S.J., Sharma, K., Lenz, C., Stachler, A.E., Maier, L.K., Richter, H., Nickel, L., Schmitz, R.A.*, et al.* (2014). A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in *Haloferax volcanii*. The Journal of Biological Chemistry *289*, 7164-7177.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R., and O'Toole, G.A. (2012). The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. Journal of Bacteriology *194*, 5728-5738.

Cady, K.C., and O'Toole, G.A. (2011). Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. Journal of Bacteriology *193*, 3433-3445.

Carte, J., Pfister, N.T., Compton, M.M., Terns, R.M., and Terns, M.P. (2010). Binding and cleavage of CRISPR RNA by Cas6. RNA *16*, 2181-2188.

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

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Deng, L., Garrett, R.A., Shah, S.A., Peng, X., and She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. Molecular Microbiology *87*, 1088-1099.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. Journal of Bacteriology *190*, 1390-1400.

Elmore, J.R., Yokooji, Y., Sato, T., Olson, S., Glover, C.V., 3rd, Graveley, B.R., Atomi, H., Terns, R.M., and Terns, M.P. (2013). Programmable plasmid interference by the CRISPR-Cas system in *Thermococcus kodakarensis*. RNA Biology *10*, 828-840.

Fischer, S., Maier, L.K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., Dyall-Smith, M., and Marchfelder, A. (2012). An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. The Journal of Biological Chemistry *287*, 33351-33363.

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

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences of the United States of America *109*, E2579-2586.

Gasiunas, G., Sinkunas, T., and Siksnys, V. (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. Cellular and Molecular Life Sciences : CMLS *71*, 449-465.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics *8*, 172.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Computational Biology *1*, e60.

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

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M.*, et al.* (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Molecular Cell *45*, 292-302.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945-956.

Hatoum-Aslan, A., Maniv, I., and Marraffini, L.A. (2011). Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. Proceedings of the National Academy of Sciences of the United States of America *108*, 21218-21222.

Hatoum-Aslan, A., Maniv, I., Samai, P., and Marraffini, L.A. (2014). Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. Journal of Bacteriology *196*, 310-317.

Heler, R., Marraffini, L.A., and Bikard, D. (2014). Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. Molecular Microbiology.

Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proceedings of the National Academy of Sciences of the United States of America *111*, 6618-6623.

Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M.D., Jr., Zhou, S., Rajashankar, K., Kurinov, I.*, et al.* (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nature Structural & Molecular Biology.

Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. Science.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R.*, et al.* (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nature Structural & Molecular Biology *18*, 529-536.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biology *8*, R61.

Lillestol, R.K., Shah, S.A., Brugger, K., Redder, P., Phan, H., Christiansen, J., and Garrett, R.A. (2009). CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. Molecular Microbiology *72*, 259-272.

Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q., Copie, V.*, et al.* (2011). Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). The Journal of Biological Chemistry *286*, 21643-21656.

Magadan, A.H., Dupuis, M.E., Villion, M., and Moineau, S. (2012). Cleavage of phage DNA by the *Streptococcus thermophilus* CRISPR3-Cas system. PloS One *7*, e40913.

Maier, L.K., Lange, S.J., Stoll, B., Haas, K.A., Fischer, S., Fischer, E., Duchardt-Ferner, E., Wohnert, J., Backofen, R., and Marchfelder, A. (2013). Essential requirements for the detection and degradation of invaders by the *Haloferax volcanii* CRISPR/Cas system I-B. RNA Biology *10*, 865-874.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis

of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biology Direct *1*, 7.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F.*, et al.* (2011). Evolution and classification of the CRISPR-Cas systems. Nature Reviews Microbiology *9*, 467-477.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843-1845.

Menon, A.L., Poole, F.L., 2nd, Cvetkovic, A., Trauger, S.A., Kalisiak, E., Scott, J.W., Shanmukh, S., Praissman, J., Jenney, F.E., Jr., Wikoff, W.R.*, et al.* (2009). Novel multiprotein complexes identified in the hyperthermophilic archaeon *Pyrococcus furiosus* by non-denaturing fractionation of the native proteome. Molecular & Cellular Proteomics : MCP *8*, 735-751.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733-740.

Mulepati, S., and Bailey, S. (2013). *In vitro* reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. The Journal of Biological Chemistry *288*, 22184-22192.

Mulepati, S., Heroux, A., and Bailey, S. (2014). Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science.

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

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935-949.

Plagens, A., Tjaden, B., Hagemann, A., Randau, L., and Hensel, R. (2012). Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. Journal of Bacteriology *194*, 2491-2500.

Plagens, A., Tripp, V., Daume, M., Sharma, K., Klingl, A., Hrle, A., Conti, E., Urlaub, H., and Randau, L. (2014). *In vitro* assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. Nucleic Acids Research *42*, 5125-5138.

Richter, H., Zoephel, J., Schermuly, J., Maticzka, D., Backofen, R., and Randau, L. (2012). Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. Nucleic Acids Research *40*, 9887-9896.

Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C.V., Spagnolo, L., and White, M.F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. Molecular Cell *52*, 124-134.

Sashital, D.G., Wiedenheft, B., and Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. Molecular Cell *46*, 606-615.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proceedings of the National Academy of Sciences of the United States of America *108*, 10098-10103.

Shah, S.A., Erdmann, S., Mojica, F.J., and Garrett, R.A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. RNA Biology *10*, 891-899.

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. The EMBO Journal *32*, 385-394.

Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annual Review of Biochemistry *82*, 237-266.

Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., and Stagg, S. (2013). Structure of an RNA silencing complex of the CRISPR-Cas immune system. Molecular Cell *52*, 146-152.

Staals, R.H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K.*, et al.* (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. Molecular Cell *52*, 135-145.

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

Terns, M.P., and Terns, R.M. (2011). CRISPR-based adaptive immune systems. Current Opinion in Microbiology *14*, 321-327.

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nature Reviews Microbiology *12*, 479-492.

van Duijn, E., Barbu, I.M., Barendregt, A., Jore, M.M., Wiedenheft, B., Lundgren, M., Westra, E.R., Brouns, S.J., Doudna, J.A., van der Oost, J.*, et al.* (2012). 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*. Molecular & Cellular Proteomics : MCP *11*, 1430-1441.

Vestergaard, G., Garrett, R.A., and Shah, S.A. (2014). CRISPR adaptive immune systems of Archaea. RNA Biology *11*, 156-167.

Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K.*, et al.* (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Molecular Cell *46*, 595-605.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature *477*, 486-489.

Zebec, Z., Manica, A., Zhang, J., White, M.F., and Schleper, C. (2014). CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*. Nucleic Acids Research *42*, 5280-5288.

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.V.*, et al.* (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Molecular Cell *45*, 303-313.

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. Nature.

Figure 4.1. Protein components of Cst crRNP/s.

cas gene organization in *Pyrococcus furiosus* **(A)**. Genes encoding predicted adaptation proteins (*cas 1, cas2, cas4*) and crRNA biogenesis protein (*cas6*) are colored grey . The three effector modules- Type III-B (*cmr*) is colored blue, Type I-G (*cst*) is colored yellow, indicated in a box and Type I-A (*csa*) is colored green.

Gel mobility shift assay to visualize Cst crRNPs formed with 3'- radiolabeled (P-32) crRNA and no protein (-), individual Cst proteins (Cas5t, Cst2, Cst1, Cas3); 2 proteins (Cas5t+ Cst2), 3 proteins (Cas5t+ Cst2+ Cst1) and 4 proteins (Cas5t+Cst2+Cst1+Cas3) as labeled **(B)** or crRNPs formed with 3'- radiolabeled (P-32) crRNA and different combinations of 2 protein (Cas5t+Cst2, Cas5t+Cst1, Cst2+Cst1), 3 proteins (Cas5t+ Cst2+ Cst1) and 4 proteins (Cas5t+Cst2+Cst1+Cas3) **(C)** as labeled. Unbound crRNA and crRNP shift are indicated.

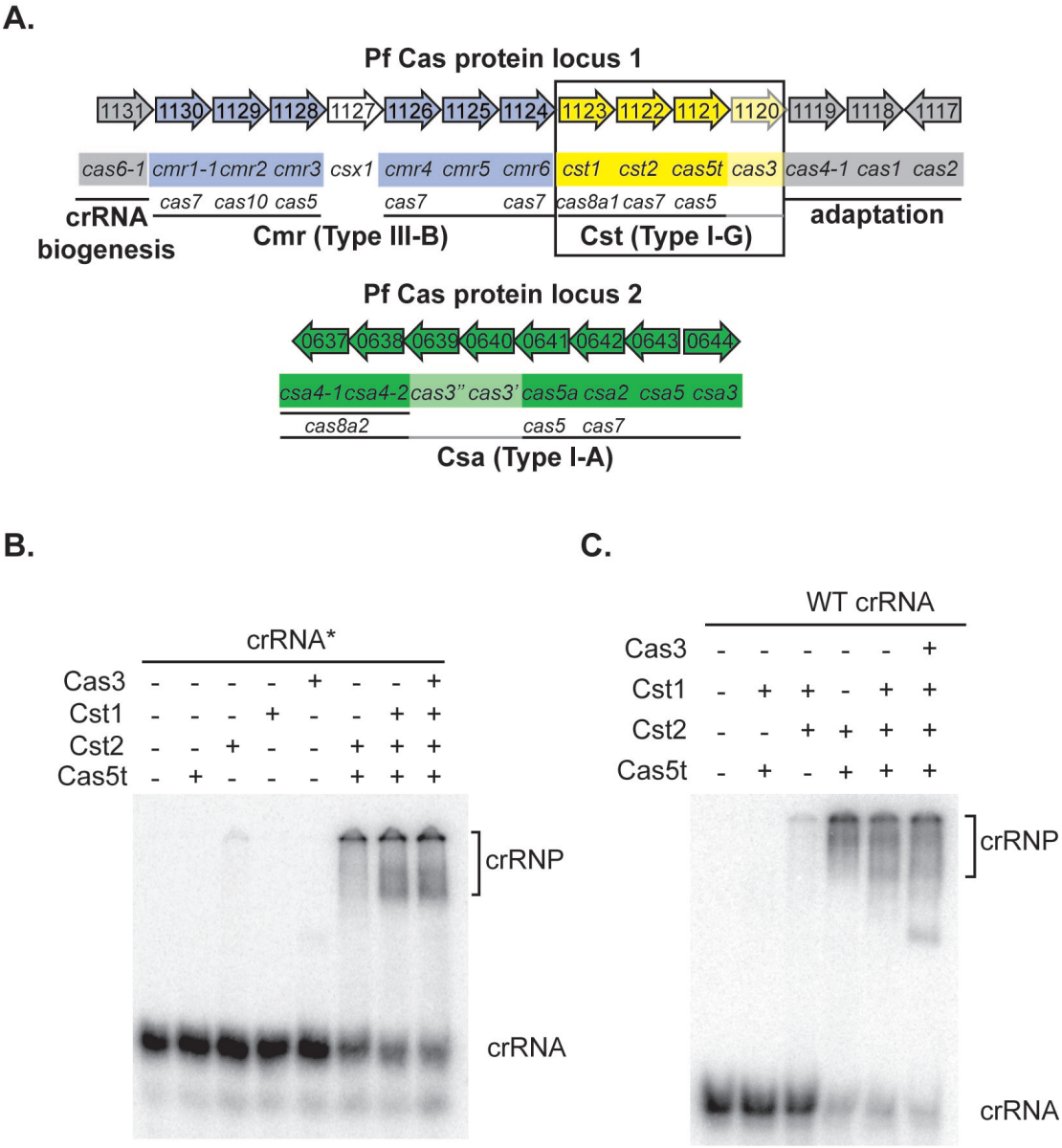

A.
Pf Cas protein locus 1
1131 1130 1129 1128 1127 1126 1125 1124 1123 1122 1121 1120 1119 1118 1117
cas6-1 cmr1-1 cmr2 cmr3 csx1 cmr4 cmr5 cmr6 cst1 cst2 cas5t cas3 cas4-1 cas1 cas2
cas7 cas10 cas5
cas7
cas7
cas8a1 cas7 cas5
crRNA biogenesis
Cmr (Type III-B)
Cst (Type I-G)
adaptation
Pf Cas protein locus 2
0637 0638 0639 0640 0641 0642 0643 0644
csa4-1 csa4-2 cas3" cas3' cas5a csa2 csa5 csa3
cas8a2
cas5 cas7
Csa (Type I-A)
B.
crRNA*
Cas3 - - - - + - - +
Cst1 - - - + - - + +
Cst2 - - + - - + + +
Cas5t - + - - - + + +
crRNP
crRNA
C.
WT crRNA
Cas3 - - - - - +
Cst1 - + + - + +
Cst2 - - + + + +
Cas5t - + - + + +
crRNP
crRNA

Figure 4.2. crRNA features essential for Cst crRNP formation.

A. Gel mobility shift assay to visualize Cst crRNP formation with no protein (-), 2 proteins (Cas5t+ Cst2), 3 proteins (Cas5t+ Cst2+ Cst1), 4 proteins (Cas5t+Cst2+Cst1+Cas3) and 3'-radiolabeled (^{32}P) crRNA with either wild type 5'-tag (labeled black) (lanes 1-4) or mutated 5'-tag, labeled red (lanes 6-9). The crRNA-guide is wild-type (labeled orange). **B.** Cst crRNP formation with no protein (-), 2 proteins (Cas5t+ Cst2), 3 proteins (Cas5t+ Cst2+ Cst1), 4 proteins (Cas5t+Cst2+Cst1+Cas3) and either crRNAs with 5'-hydroxyl, 3' phosphate (3' radiolabeled, lanes 1-4), crRNAs with 5'- phosphate, 3' hydroxyl (5' radiolabeled, lanes 5-8), crRNAs with 5'- hydroxyl, 3' hydroxyl (internally radiolabeled, lanes 9-12). Unbound crRNAs and crRNP shifts are indicated.

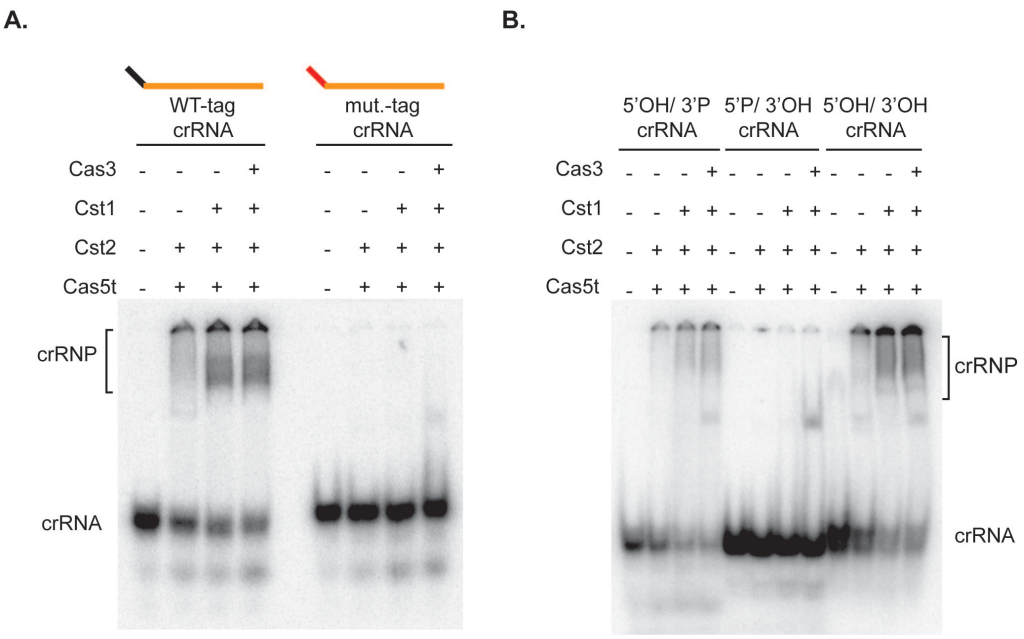
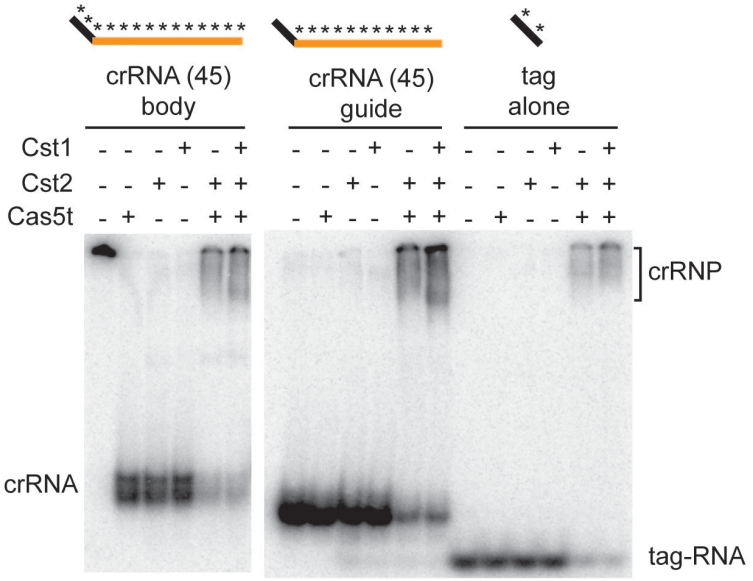

A.
WT-tag crRNA
mut.-tag crRNA
Cas3 - - - + - - - +
Cst1 - - + + - - + +
Cst2 - + + + - + + +
Cas5t - + + + - + + +
crRNP
crRNA
B.
5'OH/ 3'P crRNA
5'P/ 3'OH crRNA
5'OH/ 3'OH crRNA
Cas3 - - - + - - - + - - - +
Cst1 - - + + - - + + - - + +
Cst2 - + + + - + + + - + + +
Cas5t - + + + - + + + - + + +
crRNP
crRNA

Figure 4.3. Mapping interactions of Cst proteins with crRNA regions by UV crosslinking. A. Gel mobility shifts showing the crRNPs formed with no proteins (-), individual Cst proteins, 2 proteins (Cas5t+ Cst2), 3 proteins (Cas5t+Cst2+Cst1) and different crRNAs; radio-labeled (P-32, marked with asterisks) either all across (crRNA-body) or guide region (crRNA-guide) or radiolabeled (P-32, marked with asterisks) 8-nt 5' tag. 5'-tag (black) and guide (orange) regions of crRNA or 8-nt tag are marked. **B.** SDS gel (G-250) showing total proteins and aligned phosphor images (rad) showing the crosslinked protein profiles with Cas5t+Cst2, Cas5t+Cst2+Cst1 lanes and crRNA (body or guide) or 5'-tag. RNAse T1 and/or RNase A used in the reactions are indicated in SDS gels. The substrates (crRNA, 5'-tag) for each set is indicated on top.

A.


crRNA (45)
body
crRNA (45)
guide
tag
alone
Cst1
Cst2
Cas5t
crRNP
crRNA
tag-RNA

B.

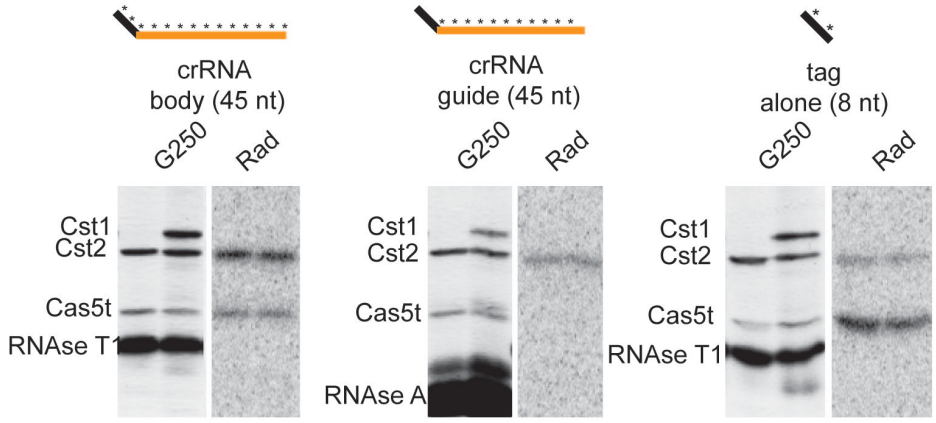

crRNA
body (45 nt)
crRNA
guide (45 nt)
tag
alone (8 nt)
G250
Rad
Cst1
Cst2
Cas5t
RNAse T1
RNAse A

Figure 4.4. Proteins essential for Cst crRNP-mediated DNA binding and cleavage.

A. Double-stranded DNA oligo used for DNA binding and cleavage assays. Wild-type (wt.) protospacer (red) and PAM (blue) are marked and labeled. 5'-tag (black) and guide (red) of crRNA assembled in Cst crRNP is shown on top. Target strand (complementary to crRNA-guide) and non-target strand of DNA are indicated. DNA binding **(B)** and DNA cleavage **(C)** by no protein (-) or Cst crRNP formed with either Cas5t+Cst2, Cas5t+Cst2+Cst1, Cas5t+Cst2+Cst1+Cas3 as labeled. Effect of all 4 proteins without crRNA (lanes 5, 11) and Cas3 + crRNA (but no Cst proteins, lanes 6, 12) are shown, labeled. Either target strand or non-target strand of DNA are 5'- radiolabeled (^{32}P) as indicated on top. 5' radiolabeled DNA size-standard (Affymetrix, low molecular weight marker, 10-100) is labeled (M).

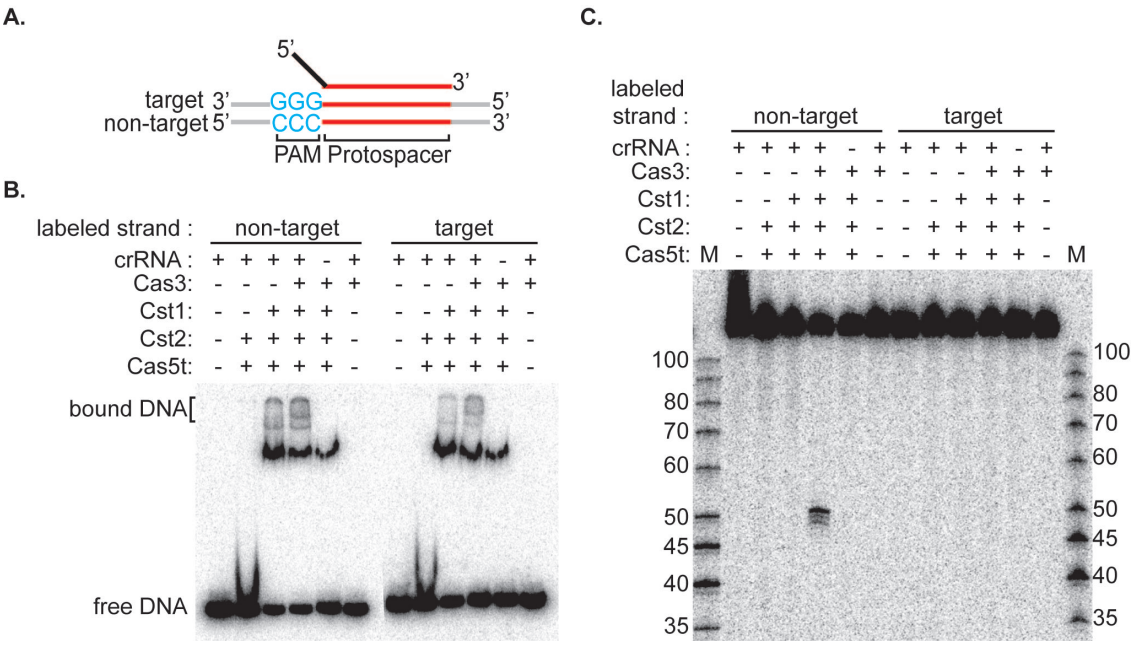

A.
5'
3'
target 3'
GGG
5'
non-target 5'
CCC
3'
PAM Protospacer
B.
labeled strand :
non-target
target
crRNA :
Cas3:
Cst1:
Cst2:
Cas5t:
bound DNA[
free DNA
C.
labeled strand :
non-target
target
crRNA :
Cas3:
Cst1:
Cst2:
Cas5t:
M
100
80
70
60
50
45
40
35

Figure 4.5. crRNA-guided, PAM-specific DNA binding and cleavage by Cst-Cas3. **A.** DNA substrates used in the assay are depicted. crRNA is shown on top with 5'-tag (black) and guide (red). Target and non-target strands of double-stranded DNA with wild-type protospacer (red) and wild-type PAM sequence (blue), mutated (mut.) PAM (green), mutated (mut.) protospacer (orange) are marked. DNAs with hybrid PAMs (G-G, C-C) have wild-type PAM (blue) on one strand (target strand in G-G, non-target strand in C-C) and mutated PAM (green) on the other strand, leaving PAM sequence non-complementary. DNA binding **(B)** and DNA cleavage **(C)** by no protein (-) or Cst crRNP formed with Cas5t+Cst2+Cst1+Cas3 and either of the specified substrates, as labeled. Non-target strand of DNA is 5'-radiolabeled (^{32}P). 5' radiolabeled DNA size-standard is labeled (M).

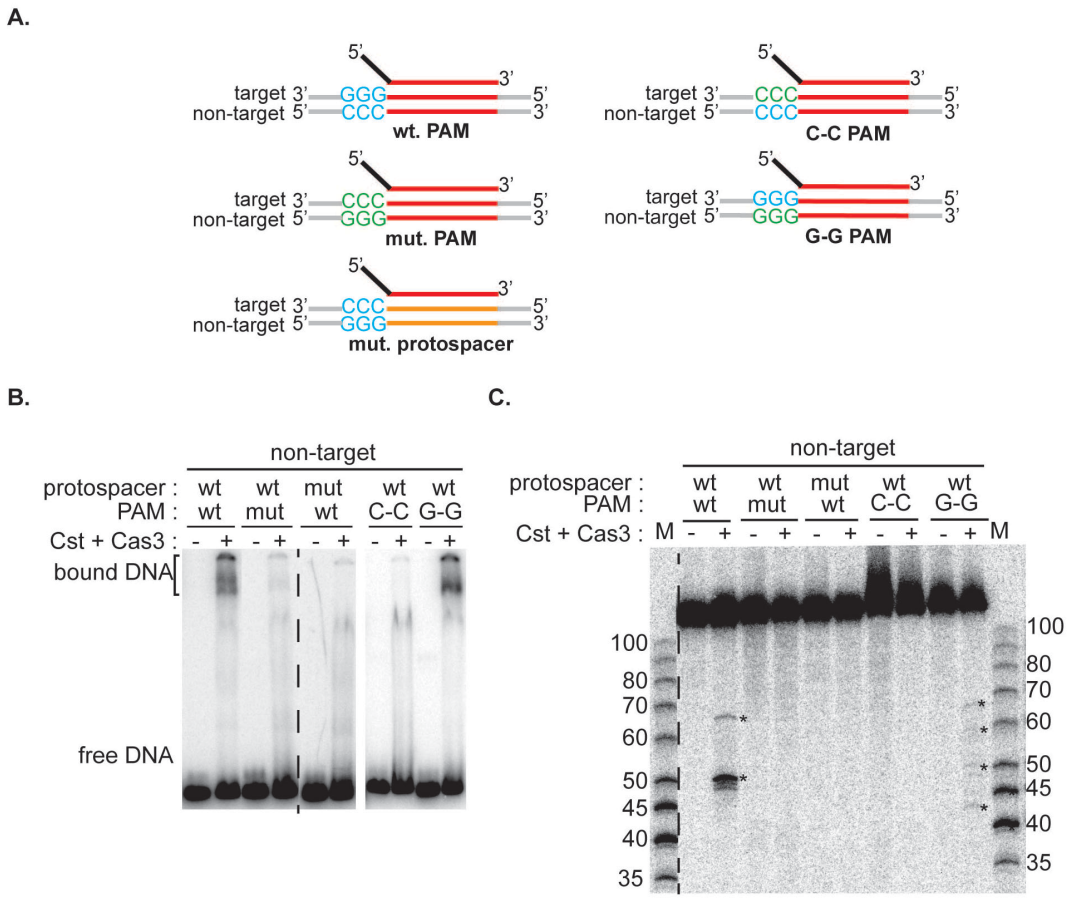
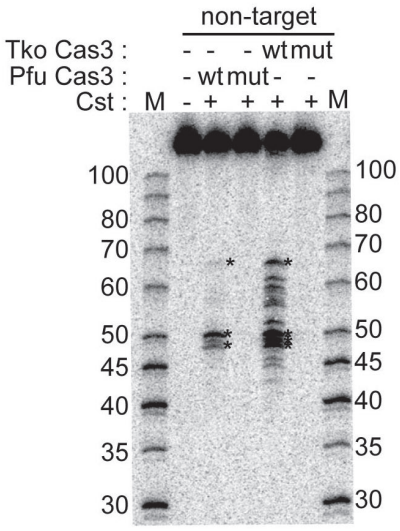

A.
5'
3'
target 3'
non-target 5'
GGG
CCC
5'
3'
wt. PAM
C-C PAM
G-G PAM
mut. PAM
mut. protospacer
B.
C.
non-target
protospacer :
PAM :
Cst + Cas3 :
wt
mut
C-C
G-G
M
bound DNA
free DNA
100
80
70
60
50
45
40
35

Figure 4.6. Cst crRNP directs Pfu and Tko Type I-B Cas3s to cleave DNA in protospacer of non-target strand.

A. DNA cleavage by no protein (-) or Cst crRNP (formed with Cas5t+Cst2+Cst1) and either wild-type Cas3 (wt.) or HD-AA mutated Cas3 (mut.), from Pfu (lanes 2,3) or Tko (lanes 4,5). Non-target strand of DNA is 5'-radiolabeled (^{32}P). 5' radioabeled DNA size-standard is labeled (M). **B.** Mapping of cleavage site by Pfu or Tko Cas3 in Fig 6A**.** Protospacer (red) and PAM (blue) sequences are marked. Cleavages are indicated by arrows; short arrows mark faint bands and stronger bands are marked by longer arrows. Numbers indicate size of major products. Same-sized products by Pfu and Tko Cas3 are marked by asterisk on the gel (A) and in the sequence (B). All cleavages are detected in protospacer region of non-target strand.

A.



B.

Cleavage by Cst+Pfu Cas3 :

3'-(N_{37})GGGAACATCATACGCCAGGAACGCCGACTCTCGTGAAGT(N_{40})-5' target strand

5'-(N_{37})CCCTTGTAGTATGCGGTCCTTGCGGCTGAGAGCACTTCA(N_{40})-3' non-target strand

47 * 50 66

Cleavage by Cst+Tko Cas3 :

3'-(N_{37})GGGAACATCATACGCCAGGAACGCCGACTCTCGTGAAGT(N_{40})-5' target strand

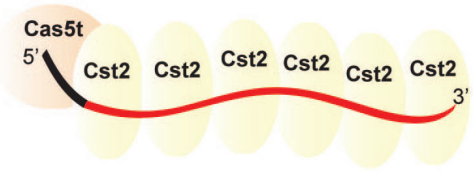
5'-(N_{37})CCCTTGTAGTATGCGGTCCTTGCGGCTGAGAGCACTTCA(N_{40})-3' non-target strand

42 45 47 * 50 55 59 66

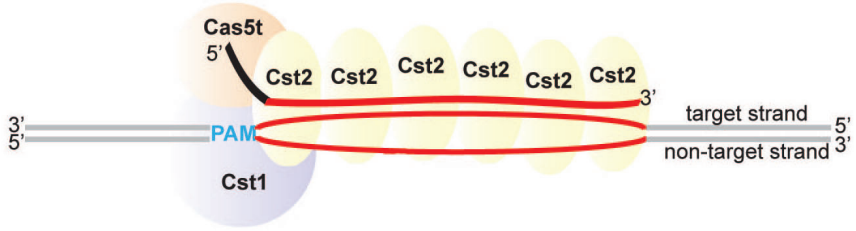
Figure 4.7. A mechanistic model of Cst-crRNP mediated DNA silencing.

Cas5t (orange) and Cst2 (yellow) interact with 5' tag (black) and Cst2 interacts with guide (red) of crRNA to form minimal crRNP. Guided by guide crRNA, Cas5t+Cst2+Cst1 forms a surveillance complex and binds DNA. Cst1 (purple) scans and interacts with PAM (blue) and recruits Cas3 (teal). crRNA-guide basepairs with target strand of cognate protospacer (red) displacing non-target strand, forming R-loop. Cas3 nuclease cuts single-stranded region of protospacer of non-target strand in the R-loop.

A. Cst crRNP assembly



B. DNA targeting by Cst crRNP



C. DNA cleavage by Cst crRNP-recruited Cas3

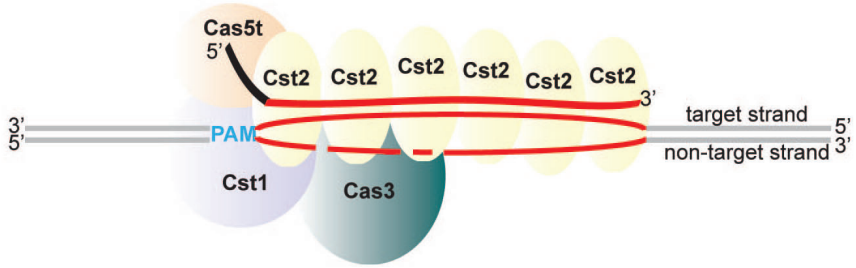


Figure 4.S1. SDS-PAGE showing recombinant Cas5t, Cst2, Cst1 and Cas3 (as labeled). Precision plus dual color standard (Biorad) is used for size estimation and labeled M. Labeled sizes are in Kilo-Daltons (kDa). Protein bands correspond to expected size.

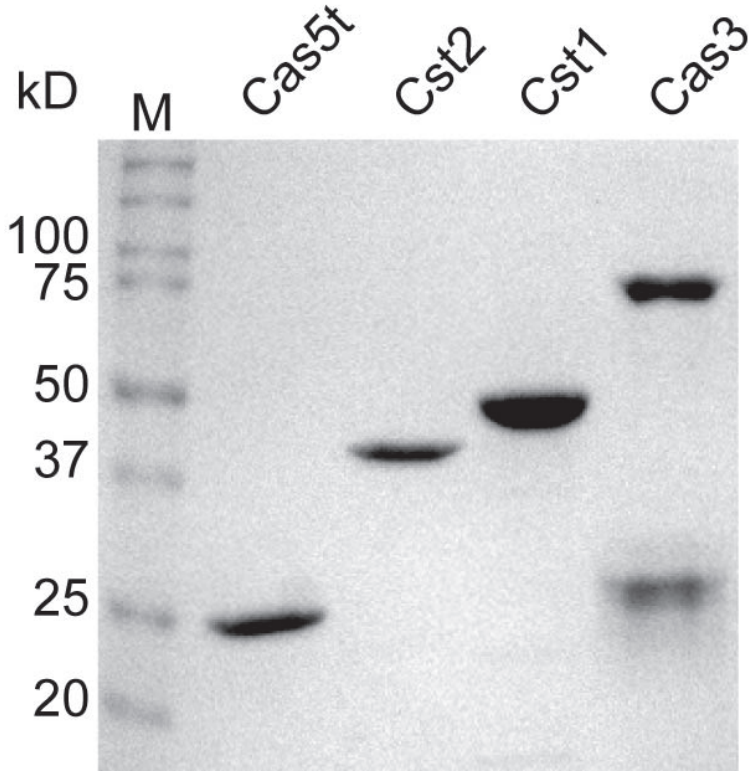

kD
M
Cas5t
Cst2
Cst1
Cas3
100
75
50
37
25
20

Figure 4.S2. A. SDS-PAGE showing recombinant wild-type (wt.) Cas3 and mutated (mut.) Cas3 proteins from Pfu or Tko (as labeled)**.** Precision plus dual color standard (Biorad) is used for size estimation and labeled M. Labeled sizes are in Kilo-Daltons (kDA). Protein bands marked by asterisk correspond to expected size. **B.** DNA binding by no protein (-), Cst crRNP (formed with Cas5t+Cst2+Cst1) and either wild-type (wt.) Cas3 or mutated (mut.) Cas3 from Pfu or Tko. Unbound and Cst crRNP-bound DNA are indicated. Relates to assay in Figure 6.

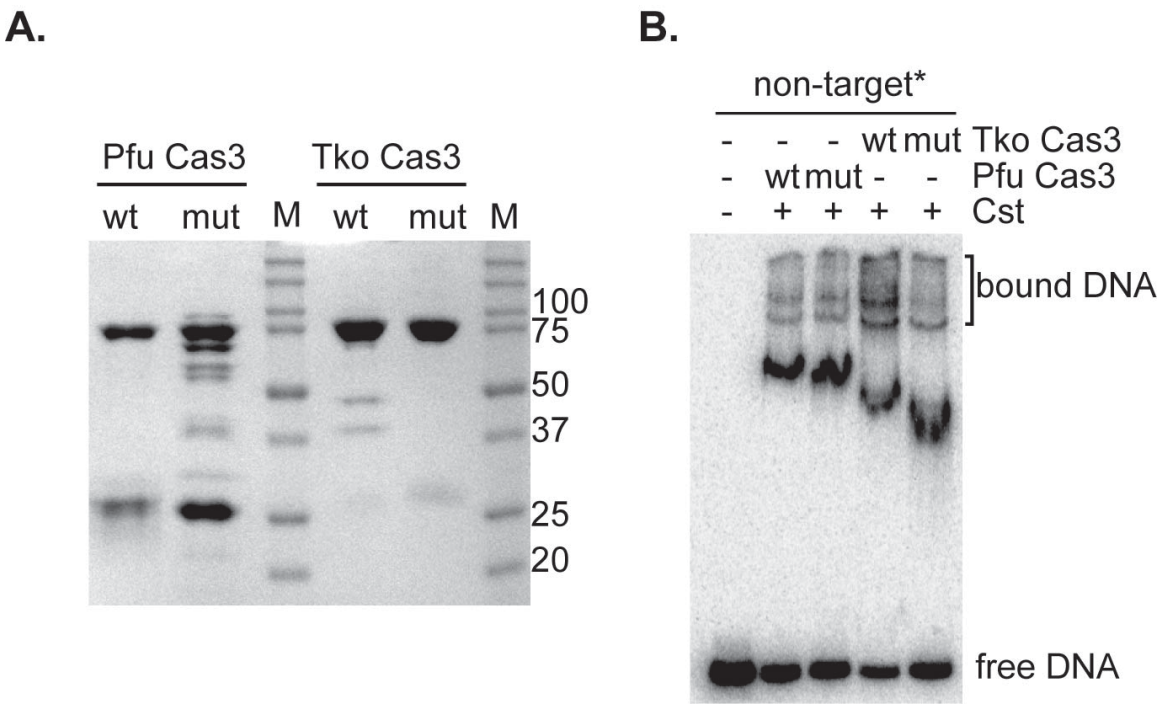
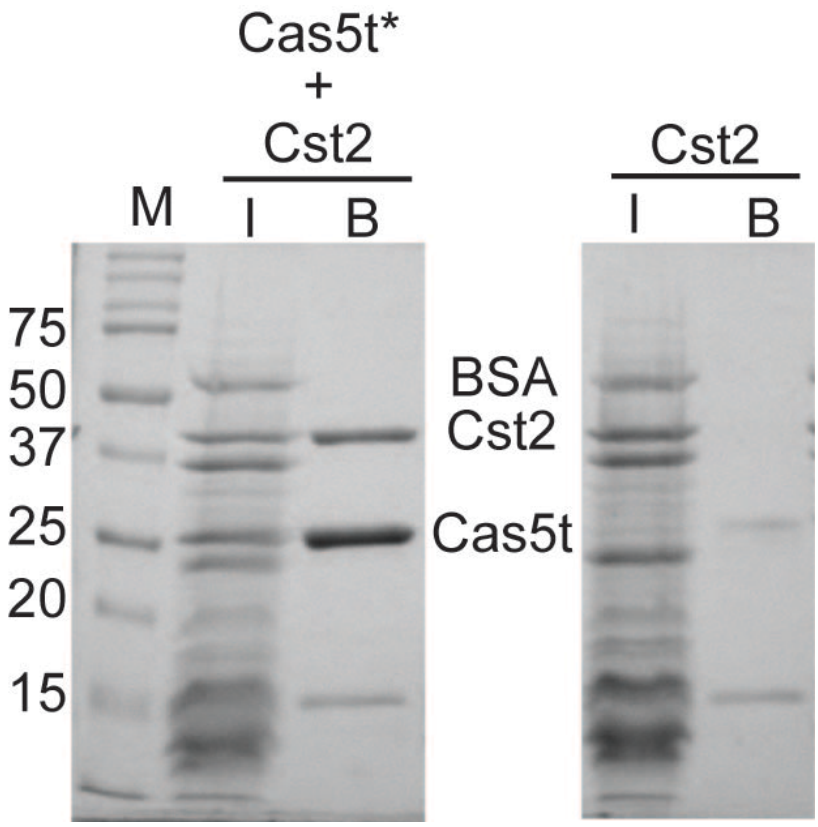

A.
Pfu Cas3
Tko Cas3
wt
mut
M
wt
mut
M
100
75
50
37
25
20
B.
non-target*
- - - wt mut Tko Cas3
- wt mut - - Pfu Cas3
- + + + + Cst
bound DNA
free DNA

Figure 4.S3. **Protein-protein interaction between Cas5t and Cst2 by affinity pull-down.** SDS-PAGE showing protein profiles of input (I) and bound (B) samples in Cas5t (His-tagged, marked by asterisk) + Cst2 (untagged) and Cst2 (untagged background, negative control). BSA (seen in input lanes) is used as an internal control. Presence of Cst2 in bound lane of Cas5t+Cst2 reaction, but not in bound lane of Cst2 alone is indicative of a true interaction. Precision plus dual color standard (Biorad) is used for size estimation and labeled M. Labeled sizes are in Kilo-Daltons (kDA). Protein bands correspond to expected size.


Cas5t*
+
Cst2
Cst2
M
I
B
I
B
75
50
37
25
20
15
BSA
Cst2
Cas5t

CHAPTER 5

DISCUSSION

Since the first experimental evidence of CRISPR-Cas mediated adaptive immunity in prokaryotes (Barrangou et al. 2007), collective research endeavors have provided mechanistic insights at a remarkable pace. Multiple CRISPR-Cas immune systems (Types I-III) (Haft et al. 2005a; Makarova et al. 2011c) have been identified that silence invader nucleic acids via diverse effector complexes in bacteria and archaea (Brouns et al. 2008; Marraffini and Sontheimer 2008; Hale et al. 2009; Garneau et al. 2010; Cady and O'Toole 2011b; Fischer et al. 2012b; Gasiunas et al. 2012b; Jinek et al. 2012; Zhang et al. 2012c; Deng et al. 2013a; Elmore et al. 2013a; Maier et al. 2013b; Mulepati and Bailey 2013b; Peng et al. 2013b; Sinkunas et al. 2013a; Staals et al. 2013a; Plagens et al. 2014b).

Effector complexes in *Pyrococcus furiosus*

Effector complexes (crRNPs) are composed of processed, mature CRISPR RNAs (crRNAs) containing fragments of invader-derived sequences and immune system-specific subtype Cas proteins. Cas gene clusters in *Pyrococcus furiosus* (Pfu) encode three putative immune systems, Types I-A, I-G and III-B. Initial biochemical fractionation of Pfu cell extract indicated the presence of multiple crRNPs (Hale et al. 2008). One of the crRNPs was previously characterized that contained Cmr proteins

(Cmr1-6) of Type III-B system and mature crRNA forms and mediated homology dependent cleavage of target RNAs *in vitro* ((Hale et al. 2009) and Chapter 6, outlined in Figure 5.1). Here, we have additionally identified two compositionally-distinct, native effector complexes corresponding to associated immune systems: Csa crRNP of Type I-A system and Cst crRNP of Type I-G system (Chapter 2) (Figure 5.1). Prior to this work, there was no insight on the molecular composition and functional role of Csa and Cst crRNPs in *P. furiosus*. Our biochemical investigations indicate that Csa and Cst complexes constitute independent effector complexes in Pfu and cleave target DNA in homology and PAM-dependent manner, similar to other Type I systems (Chapters 3 and 4 respectively, Figure 5.1). Through the course of this work we have unraveled some important molecular features of Csa and Cst effector complex assembly and mechanistic basis of invader DNA silencing.

Essential features of crRNAs associated with Csa, Cst and Cmr complexes in Pfu

5' tag of crRNAs

Our findings indicate that 8-nt repeat-derived 5'tag of crRNAs is important for assembly of all three (Csa, Cst and Cmr) crRNPs in Pfu (outlined in Figure 5.2). RNA sequencing analyses indicated processed crRNAs originated from all seven CRISPR loci could associate with all three complexes (Chapter 2). This is a unique feature typically not found in organisms containing multiple CRISPR-Cas systems where each immune system associates with crRNAs derived from the associated CRISPR locus (e.g. *S. thermophilus* (Carte et al. 2014b)). This is due to the fact that each Cas subtype associates with its own specific repeat types that are processed by distinct crRNA biogenesis

mechanisms and generate different 5' tag sequences. All seven CRISPR loci of Pfu have highly similar repeat sequences (Hale et al. 2008) that belong to same class of repeats (Kunin et al. 2007). Cas6 cleaves every repeat RNA in Pfu (Carte et al. 2008b; Carte et al. 2010) to generate the same 5' tag found in all mature crRNAs (Hale et al. 2008). RNA sequencing analyses confirms a vast majority of processed total crRNAs and complex-associated crRNAs in Pfu retain the same 8-nt, 5'-tag (Chapter 2 and (Hale et al. 2012a) Figure 5.2).

5' tag recognition and effector complex assembly

Reconstitution of crRNPs by gel shift analyses corroborates 5'-tag is a critical recognition sequence for assembly of Csa, Cst crRNPs (the structural core and the functional complex (holoenzyme)) (Chapters 3 and 4) and Cmr crRNP (Hale et al. 2012a; Hale et al. 2014). Additionally, we determined the 5'-OH chemical end of crRNAs (generated by Cas6 cleavage) is important for Cmr, Csa and Cst crRNP reconstitution (Chapter 3 and 4, (Hale et al. 2012a; Hale et al. 2014).

While 5'-tag sequence is essential component for assembly of all three complexes in Pfu (Chapters 2, 3 and 4 and (Hale et al. 2012a)), evidence implicates distinct modes of recognition. RNA sequencing results show Cmr complex does not associate with ten crRNAs of locus 8, while Csa and Cst complexes do ((Hale et al. 2012a) and Chapter 2). The 5' tag of these crRNAs shows sequence variation and appears to be a 7-nt tag due to deletion of the guide-adjacent nucleotide. This sequence variation is not tolerated by Cmr complex specifically implicating distinct mechanism of tag recognition and interaction for complex assembly.

3’ ends of crRNAs

Contrary to stringent requirements of 5’ ends of crRNA, Northern blot and RNA sequencing analyses suggest crRNAs with heterogenous 3’ ends can associate with Csa, Cst and Cmr complexes (outlined in Figure 5.2). While Csa and Cst associated crRNAs either lack or contain variable amount of 3’ repeats, Cmr-associated crRNA size forms do not contain 3’ repeat and subsets of crRNA species contain guide sequences truncated at the 3’ end (Chapter 2 and (Hale et al. 2012a). The 39-nt crRNA size forms exclusively associate with Cmr complex (Chapter 2) while the 45-nt crRNA size form, one of the most abundant crRNA species comprised of 8-nt 5’-tag and 37-nt of guide associate with all three complexes (Chapters 2, 3 and 4 and (Hale et al. 2009; Hale et al. 2012a)).

Previous biochemical characterization of Cmr complex and functional assays with reconstituted Csa and Cst complexes here show, the 45-nt crRNA (that contains 5’ tag and 7.01 guide sequence) can assemble into all three complexes and guide cleavage of target RNA (Cmr) and target DNA (Csa and Cst) via distinct mechanisms (Chapters 3 and 4 and (Hale et al. 2009; Hale et al. 2012a). We have no current evidence whether the Csa and Cst complex-associated crRNA size forms that contain 3’ repeats are functional in guiding target cleavage.

Sorting of crRNAs in Pfu effector complexes

We analyzed the distinct features of Pfu Csa, Cst and Cmr complex-associated crRNAs (outlined in Figure 5.2) and propose a model of crRNA sorting into different complexes. Cas6e and Cas6f (Types I-E and I-F systems, respectively) and Cas5d of

Type I-C systems define the 5' and 3' boundary of complex-associated mature crRNAs and also remain bound to 3' end of crRNAs in the effector complexes (Jore et al. 2011a; Wiedenheft et al. 2011a; Wiedenheft et al. 2011b; Nam et al. 2012b; van Duijn et al. 2012a). Compositional analyses of native Csa and Cst and Cmr crRNPs in Pfu indicate Cas6 is not a component of Csa, Cst or Cmr effector complexes that contain shorter, processed crRNAs (Hale et al. 2009; Hale et al. 2012a).

Cmr-associated crRNAs are specific in length (45 nt, 39 nt) irrespective of the length of the (invader-derived) guide sequence (Hale et al. 2012a). Similar findings were reported in Type III-A system, where two specific mature crRNA size forms (43 and 37 nt) are formed independent of the length of the guide indicative of a ruler mechanism of processing that involves counting from the 5' end of crRNA (Hatoum-Aslan et al. 2011b). This correlates to the number of Csm3 (Cas7 of Type III-A) multimers binding along the guide to direct 3' end processing by unknown nuclease/s and thereby determining the length of mature crRNA (Hatoum-Aslan et al. 2013b). Likewise, we expect Cmr4 (Cas7 of Type III-B) to determine the length of Cmr-associated mature crRNAs. We speculate binding of Csa, Cst, Cmr proteins to Cas6-generated 1X intermediate RNAs triggers dissociation of Cas6 from the 3' repeat sequences. The Cas7 proteins of the different complexes (Csa2, Cst2 and Cmr4 respectively) that bind along the guide likely directs the site of 3' end processing in the three complexes by unknown mechanisms. It is possible Type I crRNPs (Csa, Cst) accommodate more Cas7 multimers to assemble with longer crRNAs that contain 3' repeat sequences. In summary, collective evidence implicates Type I systems either contain Cas6 or possess a distinct mode of defining the 3' boundary of crRNAs amenable to complex assembly.

Cas protein composition of co-existing effector complexes in Pfu

Immunopurified native Pfu crRNPs that were identified through tandem mass spectrometry and immunoblotting revealed distinct Cas protein compositions. Each complex comprised subtype-specific Csa, Cst and Cmr proteins encoded by the three immune-systems (Types I-A, I-G and III-B respectively) (Chapters 2 and (Hale et al. 2012a)).

Cmr complex

Cmr complex islolated by immunopurification had the same Cas protein composition (Cmr1-6) as Cmr complex fractionated by chromatography (Chapter 2 and (Hale et al. 2009; Hale et al. 2012a)). The architecture of Pfu Cmr complex is elucidated through UV induced RNA-protein crosslinking and electron microscope analyses (Spilman et al. 2013a; Hale et al. 2014) Ramai et al, Cell reports, accepted).

Csa complex

The Cas protein composition results obtained from the native Csa complex (Chapter 2) were corroborated in active Csa complex reconstitution *in vitro* (Chapter 3). Native Csa crRNP contained seven Csa proteins encoded by the Csa gene cluster except Csa3 (a predicted transcriptional regulator containing CARF domain) (Lintner et al. 2011b; Makarova et al. 2014). In addition to Csa3, Csa complex immunopurified with Cas3" (nuclease) antibody did not contain Csa4-2 (one copy of Cas8a2, large subunit). Csa3 and Csa4-2 were not required for DNA targeting in our reconstitution assays. The six Csa

proteins (Csa4-1 (Cas8a2, large subunit), Cas3", Cas3', Cas5a, Csa2 (Cas7) and Csa5 (small subunit)) identified in native Csa complex (Chapter 2) were essential components of DNA targeting Csa crRNP (holoenzyme) (Chapter 3).

Cst complex

Cas5t and Cst2 (Cas7) were the only Cas proteins identified in native Cst crRNP by immunopurification (Chapter 2). Similar to large subunit protein (Cse1) of Cse crRNP (Type I-E) (Jore et al. 2011a; Wiedenheft et al. 2011a), Cas8 superfamily large-subunit protein (Cst1) of Type I-G system is likely less stably associated with Cst crRNP and wasn't identified likely due to stringent immunopurification conditions. Our reconstitution studies indicate Cas5t (Cas5) Cst2 (Cas7) and Cst1 (Cas8a1) are essential essential components of Cst surveillance complex that binds specific DNA targets (Chapter 4).

Predicted roles of Cas proteins in effector complex structure

Cas5 and Cas7 proteins. Native gel shift and UV-induced RNA-protein crosslinking studies provide insights of molecular organization of Csa and Cst crRNPs and draws parallels with other Type I and III complexes with resolved structures (Figure 5.3). Analogous to crRNPs of Types I-A (Lintner et al. 2011b; Plagens et al. 2012a), I-B (Brendel et al. 2014), I-C (Csd) (Nam et al. 2012b), I-E (Cse) (Jore et al. 2011a; Wiedenheft et al. 2011a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014) and I-F (Csy) (van Duijn et al. 2012a) systems, the structural core of both Csa and Cst

crRNPs involve interactions between crRNA, Cas5 (Cas5a in I-A and Cas5t in I-G) and Cas7 (Csa2 in I-A and Cst2 in I-G) superfamily proteins (Chapters 3, 4).

Cas7 superfamily proteins (Csa2 in Type I-A and Cst2 in Type I-G) associate with guide region of crRNA and most likely bind along the length to form a helical backbone. Native gel shift assays indicate Cas7 proteins (Csa2 in *S. solfataricus* (Lintner et al. 2011d) and *T. tenax* (Plagens et al. 2014b), Csc2 in *T. pandens* (Hrle et al. 2014) and Csa2, Cst2 in Pfu (data not shown) bind RNA in sequence-independent manner.

Our crosslinking data suggests Cas5 superfamily proteins (Cas5a in Type I-A, Cas5t in Type I-G) interact with 5' tag of crRNA. Based on resolved structures of analogous Type I systems (I-E, I-F) (Wiedenheft et al. 2011a; van Duijn et al. 2012a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014) and Type III systems ((Rouillon et al. 2013; Spilman et al. 2013a; Staals et al. 2013a) Ramia et al., Cell Reports, accepted), we speculate Cas5 proteins cap the 5' end and restrict further oligomerization of Cas7 units. We expect Cas5 superfamily proteins specifically recognize sequence or structure elements of 5'-tag and 5'-OH for complex assembly (Figure 5.3)

Similar to Type I crRNPs, resolved structures of Type III crRNPs (Rouillon et al. 2013; Staals et al. 2013a), and RNA- protein crosslinking and EM reconstruction of Pfu Cmr crRNP (Spilman et al. 2013a; Hale et al. 2014) show a similar architecture that involves Cas5 superfamily members (Csm4 and Cmr3) binding 5' tag and multimers of Cas7 superfamily members (Csm3 and Cmr4) binding along the length. These findings indicate Pfu Csa, Cst and Cmr crRNPs contain a similar structural core.

Small-subunit proteins. Native gel shifts of Csa crRNPs suggest Cas5, Cas7 and small subunit protein (Csa5) constitute sub-complex/es larger than the core (Chapter 3). Structure of *S. solfataricus* small subunit protein Csa5 (Type I-A) contains a β-strand rich domain insertion into conserved α helical domain. The α helical domain bears homology to C-terminal domain of small subunit protein of Type I-E system (Cse2) (Reeks et al. 2013a). In keeping with the structural analogy between Type I and Type III crRNP components, domain modeling analysis indicate N-terminal α helical domain of small subunit protein Cse2 (Type I-E) bears homology with that of small subunit protein Cmr5 (Type III-B) (Reeks et al. 2013a).

Multimers of small subunit proteins (Cse2 in Type I-E, Csm2 in Type III-A and Cmr5 in Type III-A) interact with Cas7 proteins (Cse4, Csm3 and Cmr4 respectively) to form a second helical bridge around the Cas7-crRNA backbone (Rouillon et al. 2013; Spilman et al. 2013a; Staals et al. 2013a). Moreover, Type I-E (Cse) crRNP structure reveals small subunit protein (Cse2) forms a head to tail connecting dimer, interacting with multimers of Cas7 (Cse4) at the body, large subunit (Cse1) closer to the 5'end of crRNA and Cas6e at 3' end of crRNA (Jore et al. 2011a; Wiedenheft et al. 2011a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014). The above-mentioned Type I and III crRNPs structures further reveal small subunit proteins (Cse2, Csm2, Cmr5) do not directly interact with crRNA. Small subunit protein Csa5 (Type I-A) has a propensity of forming helical oligomers in solution (Reeks et al. 2013a). Therefore, we expect multimers of small subunit protein Csa5 interact with multiple Csa2 (Cas7) units in Pfu Csa (Type I-A) crRNP.

Large-subunit proteins. Structural studies show the large-subunit proteins of Type I crRNPs (Csd1 (Cas8c) in Type I-C (Nam et al. 2012b), Cse1 in Type I-E (Wiedenheft et al. 2011a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014), Csy2 in Type I-F (van Duijn et al. 2012a)) and Type III crRNPs (Csm1 in Type III-A (Rouillon et al. 2013) and Cmr2 in Type III-B (Spilman et al. 2013a; Staals et al. 2013a), both Cas10 superfamily proteins) are localized close to 5'-tag of crRNAs and Cas5 superfamily proteins. This positions the large subunit proteins opposite to known PAM sequences in crRNPs bound to target DNA of characterized Type I systems (Hochstrasser et al. 2014a; Mulepati et al. 2014a) (the functional context is discussed in the next section). We expect large subunit proteins (Csa4-1 (Cas8a2) and Cst1 (Cas8a1) respectively) of Type I-A (Csa) and Type I-G (Cst) crRNPs in *P. furiosus* will have a similar location close to 5' tag of crRNAs and Cas5 proteins of the Csa and Cst complex.

Cas3 proteins. In Pfu Cas3 proteins (Cas3" and Cas3') of Type I-A and Cas3 of Type I-G system make different structural contributions in effector complexes (Csa and Cst). Unlike other Type I systems, nuclease and helicase domains of Cas3 in archaeal-specific Type I-A systems are encoded in two separate proteins (Cas3" and Cas3' respectively) (Haft et al. 2005a; Makarova et al. 2011c). Similar to Type I-A system in *T. tenax* (Plagens et al. 2012a; Plagens et al. 2014b) our findings indicate Cas3" and Cas3' are structural components of the reconstituted Csa complex in Pfu that recognizes and binds invader DNA before cleavage (Chapter 3). In contrast, the Type I-G -associated Cas3 is not a structural component of the Pfu Cst surveillance complex and is recruited downstream for DNA cleavage (Chapter 4), similar to role of Cas3 in other characterized

Type I system (I-B, I-C, I-E and I-F) (Brouns et al. 2008; Cady and O'Toole 2011b; Wiedenheft et al. 2011a; Wiedenheft et al. 2011b; Nam et al. 2012b; van Duijn et al. 2012a; Mulepati and Bailey 2013b; Sinkunas et al. 2013a; Brendel et al. 2014; Hochstrasser et al. 2014a; Huo et al. 2014a) (Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014).

To summarize, we predict Pfu Csa complex (Type I-A) likely bears homology to Cse (Type I-E), Csm (Type III-A) and Cmr (Type III-B) complex architecture (Figure 5.3). The three Cas component (Cas5-Cas7-large subunit Cas8) Cst surveillance complex, likely bears closer resemblance to Csd (Type I-C) and Csy (Type I-F) complex topology (Figure 5.3). Structural studies of these complexes will provide details that are currently elusive.

DNA silencing by Type I-A and Type I-G systems in Pfu

Multiple Type I and Type II immune systems mediate DNA targeting through similar step-wise mechanisms (Wiedenheft et al. 2012; Sorek et al. 2013; van der Oost et al. 2014). Typically, the effector complexes recognize non-self (invader) DNA by scanning and recognition of PAM sequences. PAM recognition triggers strand invasion, crRNA-homology scanning and base-pairing between guide of crRNA and target strand of protospacer. This leads to R-loop formation and displacement of non-target strand of DNA. Finally, the effector nucleases of the immune systems (trans-activated or intrinsic) catalyze DNA cleavage. In contrast, Type III immune systems employ distinct invader silencing mechanisms.

In vitro reconstituted Csa (Type I-A) and Cst (Type I-G) crRNPs can independently mediate cleavage of duplex DNA targets in a crRNA-guided and PAM-dependent manner (Chapters 3 and 4). Type I-A holoenzyme requires six Csa proteins (Cas5 (Cas5a), Cas7 (Csa2), Cas3 (Cas3", Cas3'), Cas8a2 (Csa4-1, large subunit) and small subunit (Csa5)) for DNA binding and cleavage. Intrinsic Cas3" (HD-nuclease protein) is the effector nuclease. In comparison, Cas5 (Cas5t), Cas7 (Cst2) and Cas8a1 (Cst1, large subunit) proteins are essential components of DNA-binding Cst surveillance complex . Unlike Type I-A system, the Cst complex bound to target DNA recruits associated Cas3 protein to catalyze cleavage in Type I-G system of Pfu. Based on findings in this work and mechanistic parallels with DNA silencing by other Type I systems, we discuss and propose models to delineate each mechanistic step of DNA targeting by Type I-A and I-G systems in Pfu.

Invader DNA recognition and scanning

Types I and II systems scan Protospacer Adjacent Motif (PAM) on invader DNA to recognize and differentiate non-self from self DNA (CRISPR loci) (Mojica et al. 2009). We predicted 5'-NGG-3' (on crRNA-complementary, target strand of protospacer) as PAM in Pfu based on *in-silico* analyses of PAMs associated with the class of repeat (Kunin et al. 2007) and corresponding Cas subtypes (Csa, Cst) (Mojica et al. 2009) and genetic evidence of PAM-dependent Types I-A and I-G mediated DNA silencing in highly related *Thermococcus kodakarensis* (Elmore et al. 2013a). Our findings reveal this predicted PAM sequence determines target DNA recognition for binding and cleavage by both Type I-A and I-G systems (Figures 3.5 and 4.5). Like other characterized Type I

systems, Pfu Type I-A and I-G recognize PAMs on target DNAs that are positioned opposite to the 5'-tag of crRNAs assembled in Csa and Cst crRNPs.

Cas proteins involved. We predict Cas protein candidate of Pfu Types I-A and I-G systems that scans PAMs, based on parallels with characterized Type I-E (Cse) system. Structure of Cse crRNP (Type I-E) bound to target DNA elucidates molecular interactions at PAM region (Hochstrasser et al. 2014a; Mulepati et al. 2014a). Large subunit protein (Cse1, that is weakly associated with the Cse crRNP) (Jore et al. 2011a) interacts with PAM nucleotides through its L1 loop (Mulepati et al. 2012; Sashital et al. 2012b; Hochstrasser et al. 2014a). Further insight is gathered from crystal structure of E. coli Type I-E crRNP bound to ssDNA (Mulepati et al. 2014a). It corroborates earlier observations that N-terminal loop of Cse1 (large subunit) protein interacts with PAM. Therefore we expect large subunit proteins Cas8a2 (Csa4-1) of Csa crRNP and Cas8a1 (Cst1) of Cst crRNP as the functional analogues involved in PAM scanning and interaction in Pfu. Similar to Type I-E system, compositional analysis of immunopurified native Cst complex suggested Cas8a1 (Cst1) associates with Cst crRNP though weak interactions. But Cas8a1 (Cst1) is required for Cst crRNP mediated target DNA binding, making it a strong candidate for PAM interacting protein.

Features of PAM recognized by Pfu DNA targeting complexes. We were further interested to investigate details of the molecular nature of PAM recognized by these effector complexes. Specifically, we asked if PAM was recognized as a single-stranded or double-stranded entity. Our data indicates that double-strandedness at PAM region is

important for efficient DNA cleavage by Csa and Cst crRNPs. We observed single-stranded PAM on target strand greatly reduces DNA binding efficiency and no detectable cleavage by Csa crRNP (Figure 3.5). On the other hand, single-stranded PAM on target strand does not affect DNA binding efficiency by Cst crRNP, but the cleavage efficiency is greatly reduced (Figure 4.5). Similar to our observation, structural and biochemical investigation of DNA targeting by Cse (Type I-E) crRNP indicates Cse1 of Csa crRNP can recognize and bind single-stranded PAM sequences on target strand, but base-pairing at PAM is required for DNA cleavage (Westra et al. 2012c; Hochstrasser et al. 2014a).

DNA strand-invasion and R-loop formation

We have no experimental evidence for mechanisms and dynamics of R-loop formation. Based on some mechanistic insights in other Type I systems, structure-function analysis in Type I-E system and drawing parallels with RecA-mediated strand-invasion during recombination, we propose mechanistic features of this step.

Rearrangement of effector complex structure. Electron microscopic (EM) reconstruction of *E. coli* Cse crRNP bound to DNA shows Cse1 interaction with PAM leads to structural rearrangement of the Cse complex (Hochstrasser et al. 2014a). Large subunit protein (Cse1) rotates by 30 degrees and small-subunit proteins (Cse2) slide closer to the 5'-tag-and PAM proximal end of crRNA-DNA hetroduplex. The crystal structure of Type I-E crRNP bound to ssDNA further reveals C-terminal four-helix bundle of Cse1 (large subunit) interacts with the proximal Cse2 (small subunit) unit of the Cse2 dimer (Mulepati et al. 2014a). The PAM-adjacent protospacer sequences of

target strand (crRNA complementary) are mapped to a groove in the crRNP formed by Cas7, Cse1 (large subunit) and Cse2 (small subunit) (Mulepati et al. 2014a). Furthermore, a recent study investigates the dynamics of R-loop formation in Cse crRNP bound DNA (Szczelkun et al. 2014). It provides evidence of a directional R-loop formation moving from the PAM-proximal end to PAM-distal end. Cse1 interaction at PAM destabilizes DNA duplex in adjacent protospacer for crRNA-homology scanning. Structural evidence indicate large subunit protein (Cse1) and small subunit protein (Cse2) could interact and stabilize the displaced non-target strand (R loop) (Mulepati et al. 2014a) and in turn facilitating homology scanning of target strand of protospacer. Likewise, we predict PAM scanning and interaction would lead to structural rearrangement of Csa and Cst crRNP and large subunit proteins Cas8a2 (Csa4-1) and Csa5 (small subunit) in Type I-A and Cas8a1 (Cst1) in Type I-G systems of Pfu will stabilize the R-loop formed after displacement of non-target strand of DNA.

crRNA seed sequence. There is evidence of 8-12 nt crRNA 'seed sequence' in the guide of crRNA, adjacent to 5'-tag in some Type I systems. This seed sequences acts as minimal determinant of homology-dependent targeting in characterized Types I-B, I-E and I-F systems (Semenova et al. 2011; Wiedenheft et al. 2011b; Maier et al. 2013b; Fineran et al. 2014). Structural studies of DNA-bound Cse crRNP reveals the PAM-adjacent protospacer sequences enter the Cse complex between 5^{th} and 6^{th} unit of Cas7 (Cse2) protein positioned next to 5' tag (Hochstrasser et al. 2014a). This co-relates to an initial seed-dependent homology scanning. It is speculated that initial base-pairing between crRNA-seed and target strand of protospacer would lead to R-loop expansion

towards PAM distal end of protospacer. Whether there is a minimal seed sequence determinant for DNA silencing by Type I-A and I-B in *P. furiosus*, needs to be determined.

Predicted roles of Cas proteins. Cas7 proteins that bind along the length of crRNA guide likely mediate strand invasion of protospacer, similar to strand invasion of duplex DNA by RecA-bound filament (Story and Steitz 1992; Chen et al. 2008). Structure of *S. solfataricus* Type I-A Cas7 (Csa2) protein shows two positively charged clusters (one Asn and other Gly rich) in the two domains which insert into central modified RRM domain (Lintner et al. 2011d). Cas7 superfamily proteins Csc2 of Type I-D system and Csm3 of Type III-A system have similar structure that contains a central modified RRM domain and additional domains inserting into the RRM domain (Hrle et al. 2013; Hrle et al. 2014). Basic patches with surface exposed positive charged residues are involved in RNA binding in Csc2 and Csm3, similar to Sso Csa2 (all Cas7 superfamily members). Mutation of conserved His in Asn-rich patch of Csa2 (Lintner et al. 2011d) and Arg in Arg-rich cluster of Csm3 (Hrle et al. 2013) results in loss of RNA binding. It can be speculated that the other glycine-rich clusters in Cas7 proteins can act as binding pockets for dsDNA.

Features of complex-bound heteroduplex. Structure of target-bound Cse (Type I-E) crRNP provides evidence of discontinuous helix (Jore et al. 2011a; Wiedenheft et al. 2011a; Mulepati et al. 2014a) and biochemical and genetic evidences show base pairing at the Cas7-unbound, kinked guide nucleotides of crRNA (every 6th nucleotide from 5'

end of guide-repeat junction) are not important for DNA silencing in Types I-B and I-E systems (Semenova et al. 2011; Maier et al. 2013b; Stoll et al. 2013; Fineran et al. 2014). Structure of Type I-E crRNP bound to ssDNA further reveals crRNA-target DNA hybrid resembles an under-wound ribbon rather than a twisted helix. The flexibility is provided by this kinked 6th nucleotide of crRNA and its target DNA complement. At those positions the crRNA and target DNA nucleotides are rotated 90° at opposite sides (Mulepati et al. 2014a). Based on current understanding, Type I crRNP mediated strand invasion of dsDNA is predicted to involve similar mechanisms employed by RecA-ssDNA filament where the dsDNA is globally under wound and stretched but locally adapts a B-form conformation that results in base-pairing-dependent heteroduplex formation

To summarize, some structural and mechanistic parallels involving nucleation of strand invasion and R-loop formation between Type I systems (including Pfu Types I-A and I-G) are likely with some distinct features. It has to be considered that in Types I-E and I-F crRNPs Cas6e and Cas6f anchor the 3'-end of crRNA (Wiedenheft et al. 2011a; Wiedenheft et al. 2011b; van Duijn et al. 2012a; Jackson et al. 2014a; Mulepati et al. 2014a; Zhao et al. 2014) but our findings in Pfu Type I-A and I-G crRNPs indicate otherwise. Although the Type I-E crRNP crystal structure suggests there is some flexibility at the Cas6e-3' tag end due to a fairly open conformation (Mulepati et al. 2014a), we expect Types I-A, I-G crRNPs would have greater flexibility afforded at that end due to lack of Cas6 and presumably any anchor. Therefore, the dynamics and stability of R-loop formed during Csa and Cst crRNP mediated strand invasion may have some distinctions.

Mechanism of DNA cleavage

DNA targeting and cleavage in other characterized Type I systems (e.g. I-A, I-E) and Types I-A and I-G systems in Pfu involve similar molecular mechanisms (Figure 5.4). **Cas3 nucleases.** Type I signature proteins of Cas3 superfamily including Pfu Cas3" exhibit metal-dependent endonuclease as well as 3' to 5' exonuclease activity *in vitro* (Beloglazova et al. 2011; Sinkunas et al. 2011). Conserved histidine and aspartate (HD) amino acids, adjacently located in the nuclease domain of Cas3 (Cas3" in Type I-A) are the active site catalytic residues of DNA cleavage. Typically Cas3 enzymes are dependent on divalent transition metals (Fe, Ni, Co). (Beloglazova et al. 2011; Sinkunas et al. 2011; Mulepati and Bailey 2013b; Sinkunas et al. 2013a; Plagens et al. 2014b). Reconstituted Pfu Types I-A and Type I-G systems exhibit Ni-dependent cleavage of homologous, double stranded DNA targets *in vitro*. Cas3" of Csa crRNP and Cst-associated Cas3 are Ni-dependent effector nucleases and conserved His, Asp (HD) are the catalytic residues.

Cas3 proteins from two Type I-G systems (but not Type I-A) can substitute function. We observe Type I-G-associated Cas3 from *T. kodakarensis* (Tko) that bears 58% amino acid sequence similarity to *P. furiosus* Type I-G Cas3, can functionally substitute and cleave DNA bound by Pfu Cst crRNP (Figure 4.6). Both Pfu and Tko Type I-G associated Cas3 nucleases have conserved catalytic residues (HD) and exhibit similar DNA cleavage products (Figure 4.6). Moreover, genetic evidence indicates that Tko Type I-G Cas3 can complement invader silencing for Pfu Type I-G system (that lacks Cas3).

However, Cas3 proteins of Types I-A and I-G systems can not complement function and are exclusive for their associated system (data not shown). The findings make sense due to their different structural roles in the respective effector complexes (Csa and Cst). Our findings implicate archaeal-specific Type I-A system is likely one of the more distinct systems among Type I systems.

Cas3 localization and recruitment. Extensive structure-function studies of Cse crRNP bound to DNA followed by recruitment of Cas3 in *E. coli* provide mechanistic insights of Cas3 localization and activity (Hochstrasser et al. 2014a; Huo et al. 2014a). Cas3 interacts with large subunit protein Cse1 in a region that undergoes rearrangement post PAM binding. Moreover, mutating residues in the loop region of Cse1 results in loss of DNA cleavage, although the Cse1 mutant can still interact with PAM and the reconstituted Cse crRNP can bind the target DNA and recruit Cas3 (Hochstrasser et al. 2014a). Similarly, in Type I-E system of *T. fusca* Cas3 was recruited to DNA-bound Cse crRNP only when Cse1 (large subunit) was stably associated (Huo et al. 2014a). These observations indicate interaction between PAM and large subunit protein triggers rearrangement and physical association of large subunit protein with Cas3 results in Cas3 recruitment and downstream DNA cleavage. In support of this hypothesis, we have preliminary evidence of large subunit protein Cas8a2 (Csa4-1) interaction with Cas3" and Cas3' proteins of Type I-A system in Pfu.

Target DNA cleavage by Type I-A system in Pfu

Reconstituted Csa crRNP (Type I-A) of Pfu cleaves both target and non-target strands of DNA (Figures 3.1, 3.6 and 5.4). The cleavages in the protospacer of non-target strand are ATP-independent (Figures 3.6 and 5.4). The cleavages outside protospacer on non-target strand and all cleavages on target strand are dependent on ATP (Figures 3.6 and 5.4). Similar cleavage properties are observed with reconstituted Type I-A system of *T. tenax* (Plagens et al. 2014b) and Type I-E systems of *E. coli* (Mulepati and Bailey 2013b) and *S. thermophilus* (Sinkunas et al. 2013a) (Figure 5.4, elaborated later).

Mechanistic model of DNA cleavage. Based on collective structural (Hochstrasser et al. 2014a; Huo et al. 2014a) and biochemical evidence in Type I-E systems (Mulepati and Bailey 2013b; Sinkunas et al. 2013a) and our data on ATP-independent cleavages by Type I-A crRNP (Figures 3.6 and 5.4), we deduce that mechanisms of PAM interaction with large subunit protein followed by homology scanning, crRNA strand invasion and initial R-loop formation due to displacement of non-target strand of protospacer are ATP-independent. We expect large subunit proteins (Cse1, Cas8 superfamily proteins and Csy1) interaction with Cas3 proteins position them to single-stranded displaced non-target strand of protospacer in the R-loop for ATP-independent initial cleavages.

Cas3 helicases further unwinds DNA by using ATP as energy source and thereby facilitate progressive degradation of non-target strand beyond protospacer by Cas3 nucleases in 3’ to 5’ direction. The degradation of non-target strand and unwinding by Cas3 helicase may cause conformational rearrangement and lead to disruption of base pairing between crRNA guide and protospacer of target strand, thereby exposing single-

stranded regions of target strand for cleavage by Cas3 nucleases. In support of this theory, mutation of conserved aspartate in DExH motif of helicase domain of Cas3 associated with *S. thermophilus* Type I-E leads to only one cleavage in protospacer of non-target strand even in the presence of ATP (Sinkunas et al. 2013a). The cleavage site maps to the ATP-independent cleavage on non-target strand by Cas3 nuclease (Sinkunas et al. 2013a) (Figure 5.4). We couldn't locate a canonical DExH motif in Pfu Cas3' (helicase). Mutation of other conserved residues in Type I-E Cas3 (in the helicase domains) result in loss or significant reduction of invader silencing (Huo et al. 2014a). It is to be determined if Pfu Cas3' (helicase of Type I-A) share conserved residues and motifs with Type I-E Cas3 and whether they contribute to catalytic activity of the proteins.

Recent structure of Type I-E Cas3 bound to ssDNA provides detailed mechanistic insights and supports our model (Huo et al. 2014a). A snapshot of ssDNA trapped in Cas3 structure reveals anchor points in RecA1 and RecA2 of helicase domains of Cas3 that likely facilitates a classical 'inchworm' movement. Moreover, the path of ssDNA in Cas3 structure suggests Cas3 helicase threads the DNA substrate towards the nuclease domain resulting in 3' to 5' degradation. Furthermore, the electron density of DNA suggests the substrate rotates by ~ 90° at the nuclease domain, which likely prevents steric hindrance (Huo et al. 2014a). As the biochemical properties of Csa crRNP-mediated cleavage are highly similar to Type I-E (Cse crRNP-Cas3) system, we expect a similar catalytic mechanism.

Type I-G mediated target DNA cleavage in Pfu

In contrast to Type I-A system, we observed no ATP-dependent cleavages of target and non-target strand by Cst crRNP with either Pfu or Tko Cas3s (data not shown). Either the Cas3 helicase in these Type I-G systems are not functional or our biochemical conditions were sub-optimal. Alignment of helicase motifs of functional Cas3 of Type I-E system (Huo et al. 2014a) with Pfu and Tko Cas3 proteins show similar conserved motifs and catalytic residues (DExH). In case the Cas3 helicases are non-functional, *in vivo* evidence of Type I-G mediated DNA silencing in Pfu and implicative evidence of Type I-G mediated silencing in Tko (Elmore et al. 2013a) suggest that cleavages in the protospacer of non-target strand are sufficient for silencing. In that case, it is likely that either ATP-independent non-target strand cleavages are sufficient or Cst crRNP remains bound to DNA heteroduplex after the non-target strand cleavage to ensure silencing. In support of the second theory, helicase mutant of Type I-E Cas3 as well use of non-hydrolysable analog AMP-PNP leads to more stable interaction of Cse crRNP-Cas3 with the bound DNA in comparison to Cas3 wild-type and inclusion of ATP (Huo et al. 2014a).

Profiles of DNA cleavage mediated by Type I effector complexes

We further analyzed biochemical parallels by comparing the sites of cleavages by Types I-A, I-G in Pfu, I-A in *T. tenax*, I-E in *E. coli* and *S. thermophilus* (Figure 5.4 compiled from Figures 3.6, 4.6 and (Mulepati and Bailey 2013b; Sinkunas et al. 2013a; Plagens et al. 2014b)). All cleavages follow a similar trend. The ATP-independent cleavages in protospacer of non-target strand determine the 3’ extremity of cleavage on

that strand. To put it in perspective, the bound effector complexes possibly protect the downstream sequences of non-target strand towards the PAM distal 3' end. The site of this ATP-independent, non-target strand cleavage mapped from the PAM-protospacer junction, varies (Figure 5.4). It maps to 26 nt in case of Type I-G of *P. furiosus* (Figure 4.6), 20 nt in Type I-A of *P. furiosus* (Figures 3.6), 15-19 nt in Type I-A of *T. tenax* (Plagens et al. 2014b), 11 nts in Type I-E of *E. coli* (Mulepati and Bailey 2013b), 15 nt in Type I-E of *S. thermophilus* (Sinkunas et al. 2013a) (compiled in Figure 5.4). The position of this cleavage may co-relate to location of Cas3 proteins at the displaced non-target strand of R-loop formed after initial crRNA strand invasion and base-pairing of guide with protospacer. Multiple factors can influence the variation in positioning of Cas3, like dynamics of R-loop initiation and stability with respect to the flexibility at the PAM distal end. Like mentioned before, in case of Cse (Type I-E) complex Cas6 anchors 3' end of crRNA that provides less flexibility at the PAM-distal end of heteroduplex and in combination with the seed sequence interaction with protospacer, the initial cleavage in non-target is merely 11 nts from the PAM in *E. coli*. On the other hand Types I-A and I-G systems may provide more flexibility to R-loop towards PAM distal in the absence of Cas6 anchor and the initial R-loop formed due to strand invasion could be larger, positioning Cas3 farther into the protospacer. An added layer of complexity arises from Cas3" and Cas3' being structural components of Csa crRNP in Type I-A system.

Apart from the initial cleavage event, all cleavages on non-target strand by characterized Types I-A, I-G and I-E systems seem to progress in 3' to 5' direction, supporting the mechanistic model. Similarly, the overall trend of cleavages on target

strand by characterized Types I-A and I-E systems follow a similar pattern, in accordance with the mechanistic model.

Future directions

The findings described in this dissertation successfully identified three co-existing, compositionally-distinct, CRISPR-Cas immune-effector complexes (Types I-A, I-G and III-B) in Pfu and uncovered the functional role of Types I-A and I-G immune-effector complexes in CRISPR-Cas mediated DNA silencing (Figure 5.1). We were able to delineate some important features of complex assembly, DNA recognition, target DNA binding and cleavage. Further investigations are required to unravel other uncharacterized mechanistic details, specifically pertaining to non-self DNA recognition and proteins involved. Structures of Csa and Cst crRNP need to be elucidated to understand stoichiometric molecular composition of the complexes. Structure of DNA-bound Csa or Cst crRNP reconstituted with catalytic mutants of Cas3" and Cas3' (Type I-A) and Cas3 (Type I-G) will provide details of molecular interaction and mechanism of DNA targeting before cleavage. Residues contributing to Cas3' (Type I-A) helicase activity needs to be determined. Assaying loss of function with active-site mutants of Cas3' helicase will provide answers on the precise role of helicase component in DNA silencing. Comparing with helicase motifs of Type I-E Cas3 may provide some insights on conserved residues. Although we have performed initial analyses of PAM requirements, more systematic analyses of PAM sequence and structure required for DNA silencing by Types I-A and I-G systems are important to investigate similarity and distinction of invader DNA recognition with other Type I systems. Systematic sequence mutations of crRNA guide or

complementary target-strand protospacer are required to analyze whether there is a minimal crRNA seed sequence determinant for DNA silencing by Type I-A and I-G systems. We investigated DNA targeting by Types I-A and I-G system by *in vitro* reconstitution approach. Native Csa and Cst complexes need to be tested for similar activity as well.

Recently a genetic system in *P. furiosus* was developed (Lipscomb et al. 2011; Farkas et al. 2012) and it provides us an avenue to test *in vivo* functions of Type I-A and I-G systems and investigate features of DNA silencing. Similar to some elements tested for functional characterization described here and mentioned for future investigation need to be assayed *in vivo*. A fellow graduate student, Joshua Elmore has undertaken these investigations for his thesis. *In vivo* evidence should support observation obtained from biochemical characterization of Type I-A and I-G systems and circumvent the limitations of biochemical investigations.

Initial studies implicated a generic concept of PAM for recognition of invader sequence for both acquisition and interference but recent findings suggest certain distinctions. From analyses of primed adaptation in Types I-E systems (Fineran et al. 2014), we deduce that while there are strict PAM requirements for invader defense, some mutations in PAM can be tolerated to trigger acquisition events. Genetic analyses to distinguish features of PAM in adaptation vs. invader defense need to be investigated. Moreover, there is evidence in Types I-B and I-E system that interference is coupled to adaptation (Swarts et al. 2012; Szczelkun et al. 2014) and Cas3 generated cleavage products could be substrates for acquisition. We need to determine whether interference

by Type I-A or I-G immune effector systems is coupled to adaptation in Pfu and further whether it involves the associated Cas3 nucleases.

Concluding remarks

The findings described here provide evidence that Types I-A and I-G DNA targeting CRISPR-Cas systems co-function in *Pyrococcus furiosus*. Each immune-effector complex mediates DNA targeting by independent mechanisms. We have recent *in vivo* evidence in Pfu that supports the *in vitro* findings. While multiple immune-effector systems sharing crRNAs appears redundant from the outset, there is a selective advantage. In case either system is compromised the other can still function especially because the crRNAs are shared. Our findings indicate mechanistic parallels between Types I-A and I-G systems overall with some distinctions. Moreover, *in vivo* analysis indicates Cas3 nucleases of one system can not complement the other system. Cas3 proteins of Type I-A are stable structural components, whereas Cas3 of Type I-G is transacting. We expect some differences in complex assembly and structure with likelihood of subtle mechanistic distinctions. Furthermore, if Cas3 helicases of Type I-G systems are inactive, it would involve deviations in mechanism of DNA silencing. Further investigations are needed to determine structural and mechanistic details before we compare conclusively.

In conclusion, CRISPR-Cas systems in *Pyrococcus furiosus* possess interesting combination of resource (crRNA) sharing, seemingly redundant DNA-targeting systems (Type I-A and Type I-G) and additional line of defense by targeting RNAs (Type III-B). This only reinforces Pfu as a model organism to study the dynamic properties of multiple

CRISPR-Cas immune systems. Further structural and mechanistic understanding will add to the foundation laid in this work. Uncovering fascinating aspects of CRISPR-Cas biology will add to the advances made in repurposing this prokaryotic defense system as a potent tool and continue to benefit research, industry and medicine.

References

Agari Y, Yokoyama S, Kuramitsu S, Shinkai A. 2008. X-ray crystal structure of a CRISPR-associated protein, Cse2, from *Thermus thermophilus* HB8. *Proteins* **73**: 1063-1067.

Anders C, Niewoehner O, Duerst A, Jinek M. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*.

Arslan Z, Hermanns V, Wurm R, Wagner R, Pul U. 2014. Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic acids research* **42**: 7884-7893.

Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, Gagarinova A, Pogoutse O, Brown G, Binkowski A et al. 2011. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Molecular microbiology* **79**: 484-502.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.

Barrangou R, Horvath P. 2012. CRISPR: new horizons in phage resistance and strain identification. *Annual review of food science and technology* **3**: 143-162.

Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M, Kochinyan S, Wang S, Chruszcz M, Minor W et al. 2008. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *The Journal of biological chemistry* **283**: 20361-20371.

Beloglazova N, Petit P, Flick R, Brown G, Savchenko A, Yakunin AF. 2011. Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *The EMBO journal* **30**: 4616-4627.

Benda C, Ebert J, Scheltema RA, Schiller HB, Baumgartner M, Bonneau F, Mann M, Conti E. 2014. Structural Model of a CRISPR RNA-Silencing Complex Reveals the RNA-Target Cleavage Activity in Cmr4. *Molecular cell* **56**: 43-54.

Bernick DL, Cox CL, Dennis PP, Lowe TM. 2012. Comparative genomic and transcriptional analyses of CRISPR systems across the genus *Pyrobaculum*. *Frontiers in microbiology* **3**: 251.

Blum H, Beier H, Gross HJ. 1987. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *ELECTROPHORESIS* **8**: 93-99.

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551-2561.

Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. 2013. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**: 429-432.

Brendel J, Stoll B, Lange SJ, Sharma K, Lenz C, Stachler AE, Maier LK, Richter H, Nickel L, Schmitz RA et al. 2014. A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in *Haloferax volcanii*. *J Biol Chem* **289**: 7164-7177.

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.

Cady KC, Bondy-Denomy J, Heussler GE, Davidson AR, O'Toole GA. 2012. The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *Journal of bacteriology* **194**: 5728-5738.

Cady KC, O'Toole GA. 2011a. Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J Bacteriol* **193**: 3433-3445.

-. 2011b. Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *Journal of bacteriology* **193**: 3433-3445.

Carte J, Christopher RT, Smith JT, Olson S, Barrangou R, Moineau S, Glover CV, 3rd, Graveley BR, Terns RM, Terns MP. 2014a. The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in *Streptococcus thermophilus*. *Mol Microbiol*.

-. 2014b. The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in *Streptococcus thermophilus*. *Molecular microbiology* **93**: 98-112.

Carte J, Pfister NT, Compton MM, Terns RM, Terns MP. 2010. Binding and cleavage of CRISPR RNA by Cas6. *Rna* **16**: 2181-2188.

Carte J, Wang R, Li H, Terns RM, Terns MP. 2008a. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* **22**: 3489-3496.

-. 2008b. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & development* **22**: 3489-3496.

Chen H, Choi J, Bailey S. 2014. Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. *The Journal of biological chemistry* **289**: 13284-13294.

Chen Z, Yang H, Pavletich NP. 2008. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* **453**: 489-484.

Chopin MC, Chopin A, Bidnenko E. 2005. Phage abortive infection in lactococci: variations on a theme. *Current opinion in microbiology* **8**: 473-479.

Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E. 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature communications* **3**: 945.

Daume M, Plagens A, Randau L. 2014. DNA binding properties of the small cascade subunit csa5. *PLoS One* **9**: e105716.

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602-607.

Deng L, Garrett RA, Shah SA, Peng X, She Q. 2013a. A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Molecular microbiology* **87**: 1088-1099.

Deng L, Niu S, Liu X, Che Y, Li E. 2013b. Coniochaetones E-I, new 4H-chromen-4-one derivatives from the Cordyceps-colonizing fungus Fimetariella sp. *Fitoterapia* **89**: 8-14.

Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of bacteriology* **190**: 1390-1400.

Deveau H, Garneau JE, Moineau S. 2010. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol* **64**: 475-493.

Ellinger P, Arslan Z, Wurm R, Tschapek B, MacKenzie C, Pfeffer K, Panjikar S, Wagner R, Schmitt L, Gohlke H et al. 2012. The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*. *Journal of structural biology* **178**: 350-362.

Elmore JR, Yokooji Y, Sato T, Olson S, Glover CV, 3rd, Graveley BR, Atomi H, Terns RM, Terns MP. 2013a. Programmable plasmid interference by the CRISPR-Cas system in *Thermococcus kodakarensis*. *RNA Biology* **10**: 828-840.

-. 2013b. Programmable plasmid interference by the CRISPR-Cas system in Thermococcus kodakarensis. *RNA Biol* **10**: 828-840.

Erdmann S, Garrett RA. 2012. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Molecular microbiology* **85**: 1044-1056.

Farkas J, Stirrett K, Lipscomb GL, Nixon W, Scott RA, Adams MW, Westpheling J. 2012. Recombinogenic properties of *Pyrococcus furiosus* strain COM1 enable rapid selection of targeted mutants. *Applied and environmental microbiology* **78**: 4669-4676.

Fineran PC, Gerritzen MJ, Suarez-Diez M, Kunne T, Boekhorst J, van Hijum SA, Staals RH, Brouns SJ. 2014. Degenerate target sites mediate rapid primed CRISPR adaptation. *Proceedings of the National Academy of Sciences of the United States of America* **111**: E1629-1638.

Fischer S, Maier LK, Stoll B, Brendel J, Fischer E, Pfeiffer F, Dyall-Smith M, Marchfelder A. 2012a. An Archaeal Immune System Can Detect Multiple Protospacer Adjacent Motifs (PAMs) to Target Invader DNA. *J Biol Chem* **287**: 33351-33363.

-. 2012b. An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. *The Journal of biological chemistry* **287**: 33351-33363.

Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67-71.

Garrett RA, Shah SA, Vestergaard G, Deng L, Gudbergsdottir S, Kenchappa CS, Erdmann S, She Q. 2011a. CRISPR-based immune systems of the *Sulfolobales*: complexity and diversity. *Biochem Soc Trans* **39**: 51-57.

Garrett RA, Vestergaard G, Shah SA. 2011b. Archaeal CRISPR-based immune systems: exchangeable functional modules. *Trends Microbiol* **19**: 549-556.

Garside EL, Schellenberg MJ, Gesner EM, Bonanno JB, Sauder JM, Burley SK, Almo SC, Mehta G, Macmillan AM. 2012. Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *Rna* **18**: 2020-2028.

Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012a. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**: E2579-2586.

-. 2012b. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E2579-2586.

Gasiunas G, Sinkunas T, Siksnys V. 2014. Molecular mechanisms of CRISPR-mediated microbial immunity. *Cellular and molecular life sciences : CMLS* **71**: 449-465.

Gesner EM, Schellenberg MJ, Garside EL, George MM, Macmillan AM. 2011a. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat Struct Mol Biol* **18**: 688-692.

-. 2011b. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nature structural & molecular biology* **18**: 688-692.

Godde JS, Bickerton A. 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *Journal of molecular evolution* **62**: 718-729.

Gong B, Shin M, Sun J, Jung CH, Bolt EL, van der Oost J, Kim JS. 2014. Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proceedings of the National Academy of Sciences of the United States of America*.

Grissa I, Vergnaud G, Pourcel C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**: 172.

Gudbergsdottir S, Deng L, Chen Z, Jensen JV, Jensen LR, She Q, Garrett RA. 2011. Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Molecular microbiology* **79**: 35-49.

Haft DH, Selengut J, Mongodin EF, Nelson KE. 2005a. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology* **1**: e60.

-. 2005b. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1**: e60.

Hale C, Kleppe K, Terns RM, Terns MP. 2008. Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *Rna* **14**: 2572-2579.

Hale CR, Cocozaki A, Li H, Terns RM, Terns MP. 2014. Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex. *Genes Dev* **28**: 2432-2443.

Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover CV, 3rd, Graveley BR, Terns RM et al. 2012a. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Molecular cell* **45**: 292-302.

-. 2012b. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol Cell* **45**: 292-302.

Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP. 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945-956.

Han D, Lehmann K, Krauss G. 2009. SSO1450--a CAS1 protein from *Sulfolobus solfataricus* P2 with high affinity for RNA and DNA. *FEBS letters* **583**: 1928-1932.

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

-. 2011b. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 21218-21222.

Hatoum-Aslan A, Maniv I, Samai P, Marraffini LA. 2014. Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *Journal of bacteriology* **196**: 310-317.

Hatoum-Aslan A, Samai P, Maniv I, Jiang W, Marraffini LA. 2013a. A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *J Biol Chem* **288**: 27888-27897.

-. 2013b. A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *The Journal of biological chemistry* **288**: 27888-27897.

Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. 2010. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**: 1355-1358.

Haurwitz RE, Sternberg SH, Doudna JA. 2012. Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *The EMBO journal* **31**: 2824-2832.

Heler R, Marraffini LA, Bikard D. 2014. Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Molecular microbiology*.

Hochstrasser ML, Taylor DW, Bhat P, Guegler CK, Sternberg SH, Nogales E, Doudna JA. 2014a. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 6618-6623.

-. 2014b. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc Natl Acad Sci U S A* **111**: 6618-6623.

Horvath P, Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**: 167-170.

Hrle A, Maier LK, Sharma K, Ebert J, Basquin C, Urlaub H, Marchfelder A, Conti E. 2014. Structural analyses of the CRISPR protein Csc2 reveal the RNA-binding interface of the type I-D Cas7 family. *RNA biology* **11**.

Hrle A, Su AA, Ebert J, Benda C, Randau L, Conti E. 2013. Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3. *RNA Biology* **10**: 1670-1678.

Hsu PD, Lander ES, Zhang F. 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**: 1262-1278.

Huo Y, Nam KH, Ding F, Lee H, Wu L, Xiao Y, Farchione MD, Jr., Zhou S, Rajashankar K, Kurinov I et al. 2014a. Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nature Structural & Molecular Biology*.

-. 2014b. Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat Struct Mol Biol* **21**: 771-777.

Jackson RN, Golden SM, van Erp PB, Carter J, Westra ER, Brouns SJ, van der Oost J, Terwilliger TC, Read RJ, Wiedenheft B. 2014a. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science*.

-. 2014b. Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* **345**: 1473-1479.

Jansen R, Embden JD, Gaastra W, Schouls LM. 2002a. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**: 1565-1575.

-. 2002b. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* **43**: 1565-1575.

Jiang W, Maniv I, Arain F, Wang Y, Levin BR, Marraffini LA. 2013. Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS Genetics* **9**: e1003844.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821.

Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S et al. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**: 1247997.

Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R et al. 2011a. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology* **18**: 529-536.

-. 2011b. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **18**: 529-536.

Karginov FV, Hannon GJ. 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell* **37**: 7-19.

Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ et al. 2003. The UCSC Genome Browser Database. *Nucleic Acids Res* **31**: 51-54.

Kim TY, Shin M, Huynh Thi Yen L, Kim JS. 2013. Crystal structure of Cas1 from Archaeoglobus fulgidus and characterization of its nucleolytic activity. *Biochemical and biophysical research communications* **441**: 720-725.

Klumpp M, Baumeister W. 1998. The thermosome: archetype of group II chaperonins. *FEBS Lett* **430**: 73-77.

Koo Y, Jung DK, Bae E. 2012. Crystal structure of Streptococcus pyogenes Csn2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure. *PloS One* **7**: e33401.

Kunin V, Sorek R, Hugenholtz P. 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biology* **8**: R61.

Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* **8**: 317-327.

Lee KH, Lee SG, Eun Lee K, Jeon H, Robinson H, Oh BH. 2012. Identification, structural, and biochemical characterization of a group of large Csn2 proteins involved in CRISPR-mediated bacterial immunity. *Proteins* **80**: 2573-2582.

Lemak S, Beloglazova N, Nocek B, Skarina T, Flick R, Brown G, Popovic A, Joachimiak A, Savchenko A, Yakunin AF. 2013. Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S] cluster containing Cas4 nuclease SSO0001 from *Sulfolobus solfataricus*. *Journal of the American Chemical Society* **135**: 17476-17487.

Lillestol RK, Redder P, Garrett RA, Brugger K. 2006. A putative viral defence mechanism in archaeal cells. *Archaea* **2**: 59-72.

Lillestol RK, Shah SA, Brugger K, Redder P, Phan H, Christiansen J, Garrett RA. 2009. CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Molecular Microbiology* **72**: 259-272.

Lim JM, Sherling D, Teo CF, Hausman DB, Lin D, Wells L. 2008. Defining the regulated secreted proteome of rodent adipocytes upon the induction of insulin resistance. *J Proteome Res* **7**: 1251-1263.

Lintner NG, Frankel KA, Tsutakawa SE, Alsbury DL, Copie V, Young MJ, Tainer JA, Lawrence CM. 2011a. The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. *J Mol Biol* **405**: 939-955.

-. 2011b. The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. *Journal of Molecular Biology* **405**: 939-955.

Lintner NG, Kerou M, Brumfield SK, Graham S, Liu H, Naismith JH, Sdano M, Peng N, She Q, Copie V et al. 2011c. Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J Biol Chem* **286**: 21643-21656.

-. 2011d. Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *The Journal of Biological Chemistry* **286**: 21643-21656.

Lipscomb GL, Stirrett K, Schut GJ, Yang F, Jenney FE, Jr., Scott RA, Adams MW, Westpheling J. 2011. Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. *Applied and Environmental Microbiology* **77**: 2232-2238.

Magadan AH, Dupuis ME, Villion M, Moineau S. 2012. Cleavage of phage DNA by the Streptococcus thermophilus CRISPR3-Cas system. *PloS One* **7**: e40913.

Maier LK, Lange SJ, Stoll B, Haas KA, Fischer S, Fischer E, Duchardt-Ferner E, Wohnert J, Backofen R, Marchfelder A. 2013a. Essential requirements for the detection and degradation of invaders by the *Haloferax volcanii* CRISPR/Cas system I-B. *RNA Biol* **10**: 865-874.

-. 2013b. Essential requirements for the detection and degradation of invaders by the *Haloferax volcanii* CRISPR/Cas system I-B. *RNA Biology* **10**: 865-874.

Makarova KS, Anantharaman V, Grishin NV, Koonin EV, Aravind L. 2014. CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Frontiers in Genetics* **5**: 102.

Makarova KS, Aravind L, Wolf YI, Koonin EV. 2011a. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology direct* **6**: 38.

-. 2011b. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct* **6**: 38.

Makarova KS, Grishin NV, Koonin EV. 2006a. The HicAB cassette, a putative novel, RNA-targeting toxin-antitoxin system in archaea and bacteria. *Bioinformatics* **22**: 2581-2584.

Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. 2006b. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct* **1**: 7.

-. 2006c. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**: 7.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF et al. 2011c. Evolution and classification of the CRISPR-Cas systems. *Nature reviews Microbiology* **9**: 467-477.

-. 2011d. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9**: 467-477.

Manica A, Zebec Z, Teichmann D, Schleper C. 2011. In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Mol Microbiol* **80**: 481-491.

Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**: 1843-1845.

-. 2010a. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* **11**: 181-190.

-. 2010b. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**: 568-571.

Menon AL, Poole FL, 2nd, Cvetkovic A, Trauger SA, Kalisiak E, Scott JW, Shanmukh S, Praissman J, Jenney FE, Jr., Wikoff WR et al. 2009a. Novel multiprotein complexes identified in the hyperthermophilic archaeon *Pyrococcus furiosus* by non-denaturing fractionation of the native proteome. *Mol Cell Proteomics* **8**: 735-751.

-. 2009b. Novel multiprotein complexes identified in the hyperthermophilic archaeon *Pyrococcus furiosus* by non-denaturing fractionation of the native proteome. *Molecular & Cellular Proteomics : MCP* **8**: 735-751.

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733-740.

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. 2005a. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution* **60**: 174-182.

-. 2005b. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**: 174-182.

Mulepati S, Bailey S. 2011. Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). *The Journal of biological chemistry* **286**: 31896-31903.

-. 2013a. In Vitro Reconstitution of an *Escherichia coli* RNA-guided Immune System Reveals Unidirectional, ATP-dependent Degradation of DNA Target. *J Biol Chem* **288**: 22184-22192.

-. 2013b. In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *The Journal of Biological Chemistry* **288**: 22184-22192.

Mulepati S, Heroux A, Bailey S. 2014a. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science*.

-. 2014b. Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* **345**: 1479-1484.

Mulepati S, Orr A, Bailey S. 2012. Crystal structure of the largest subunit of a bacterial RNA-guided immune complex and its role in DNA target binding. *J Biol Chem* **287**: 22445-22449.

Nam KH, Ding F, Haitjema C, Huang Q, DeLisa MP, Ke A. 2012a. Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *The Journal of Biological Chemistry* **287**: 35943-35952.

Nam KH, Haitjema C, Liu X, Ding F, Wang H, Delisa MP, Ke A. 2012b. Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System. *Structure* **20**: 1574-1584.

Nam KH, Huang Q, Ke A. 2012c. Nucleic acid binding surface and dimer interface revealed by CRISPR-associated CasB protein structures. *FEBS letters* **586**: 3956-3961.

Nam KH, Kurinov I, Ke A. 2011. Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA binding activity. *The Journal of biological chemistry* **286**: 30759-30768.

Nickel L, Weidenbach K, Jager D, Backofen R, Lange SJ, Heidrich N, Schmitz RA. 2013a. Two CRISPR-Cas systems in Methanosarcina mazei strain Go1 display common processing features despite belonging to different types I and III. *RNA Biol* **10**: 779-791.

-. 2013b. Two CRISPR-Cas systems in Methanosarcina mazei strain Go1 display common processing features despite belonging to different types I and III. *RNA biology* **10**: 779-791.

Niewoehner O, Jinek M, Doudna JA. 2014. Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic acids research* **42**: 1341-1353.

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**: 935-949.

Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. 2014. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature structural & molecular biology* **21**: 528-534.

Park JH, Sun J, Park SY, Hwang HJ, Park MY, Shin M, Kim JS. 2013. Crystal structure of Cmr5 from *Pyrococcus furiosus* and its functional implications. *FEBS letters* **587**: 562-568.

Pawluk A, Bondy-Denomy J, Cheung VH, Maxwell KL, Davidson AR. 2014. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginos*a. *MBio* **5**: e00896.

Peng W, Li H, Hallstrom S, Peng N, Liang YX, She Q. 2013a. Genetic determinants of PAM-dependent DNA targeting and pre-crRNA processing in *Sulfolobus islandicus*. *RNA Biol* **10**: 738-748.

-. 2013b. Genetic determinants of PAM-dependent DNA targeting and pre-crRNA processing in *Sulfolobus islandicus*. *RNA Biology* **10**: 738-748.

Plagens A, Tjaden B, Hagemann A, Randau L, Hensel R. 2012a. Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *Journal of bacteriology* **194**: 2491-2500.

-. 2012b. Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *J Bacteriol* **194**: 2491-2500.

Plagens A, Tripp V, Daume M, Sharma K, Klingl A, Hrle A, Conti E, Urlaub H, Randau L. 2014a. In vitro assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. *Nucleic Acids Res*.

-. 2014b. In vitro assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. *Nucleic acids research* **42**: 5125-5138.

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

Przybilski R, Richter C, Gristwood T, Clulow JS, Vercoe RB, Fineran PC. 2011. Csy4 is responsible for CRISPR RNA processing in Pectobacterium atrosepticum. *RNA Biology* **8**: 517-528.

Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R. 2010. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol Microbiol* **75**: 1495-1512.

Ramia N, Spilman M, Tang L, Shao Y, Elmore J, Hale C, Cocozaki A, Bhattacharya N, Terns R, Terns M et al. 2014. Essential Structural and Functional Roles of the Cmr4 Subunit in RNA Cleavage by the Cmr CRISPR-Cas Complex *Cell Reports* **In Press**.

Reeks J, Graham S, Anderson L, Liu H, White MF, Naismith JH. 2013a. Structure of the archaeal Cascade subunit Csa5: relating the small subunits of CRISPR effector complexes. *RNA Biology* **10**: 762-769.

Reeks J, Naismith JH, White MF. 2013b. CRISPR interference: a structural perspective. *Biochem J* **453**: 155-166.

Reeks J, Sokolowski RD, Graham S, Liu H, Naismith JH, White MF. 2013c. Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing. *The Biochemical journal* **452**: 223-230.

Richter C, Dy RL, McKenzie RE, Watson BN, Taylor C, Chang JT, McNeil MB, Staals RH, Fineran PC. 2014. Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Research*.

Richter H, Zoephel J, Schermuly J, Maticzka D, Backofen R, Randau L. 2012. Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. *Nucleic acids research* **40**: 9887-9896.

Rouillon C, Zhou M, Zhang J, Politis A, Beilsten-Edmands V, Cannone G, Graham S, Robinson CV, Spagnolo L, White MF. 2013. Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Molecular Cell* **52**: 124-134.

Santangelo TJ, Cubonova L, James CL, Reeve JN. 2007. TFB1 or TFB2 is sufficient for *Thermococcus kodakaraensis* viability and for basal transcription in vitro. *J Mol Biol* **367**: 344-357.

Sashital DG, Jinek M, Doudna JA. 2011. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nature Structural & Molecular Biology* **18**: 680-687.

Sashital DG, Wiedenheft B, Doudna JA. 2012a. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol Cell* **46**: 606-615.

-. 2012b. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Molecular Cell* **46**: 606-615.

Savitskaya E, Semenova E, Dedkov V, Metlitskaya A, Severinov K. 2013. High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. *RNA Biology* **10**: 716-725.

Scholz I, Lange SJ, Hein S, Hess WR, Backofen R. 2013. CRISPR-Cas systems in the cyanobacterium *Synechocystis* sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. *PloS one* **8**: e56470.

Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, Severinov K. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 10098-10103.

Shah SA, Erdmann S, Mojica FJ, Garrett RA. 2013. Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biology* **10**: 891-899.

Shao Y, Li H. 2013. Recognition and cleavage of a nonstructured CRISPR RNA by its processing endoribonuclease Cas6. *Structure* **21**: 385-393.

Shevchenko A, Wilm M, Vorm O, Mann M. 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical chemistry* **68**: 850-858.

Shockley KR, Ward DE, Chhabra SR, Conners SB, Montero CI, Kelly RM. 2003. Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. *Appl Environ Microbiol* **69**: 2365-2371.

Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *The EMBO Journal* **30**: 1335-1342.

Sinkunas T, Gasiunas G, Waghmare SP, Dickman MJ, Barrangou R, Horvath P, Siksnys V. 2013a. In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *The EMBO Journal* **32**: 385-394.

-. 2013b. In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *Embo J* **32**: 385-394.

Sorek R, Lawrence CM, Wiedenheft B. 2013. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual Review of Biochemistry* **82**: 237-266.

Spilman M, Cocozaki A, Hale C, Shao Y, Ramia N, Terns R, Terns M, Li H, Stagg S. 2013a. Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Molecular Cell* **52**: 146-152.

-. 2013b. Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Mol Cell* **52**: 146-152.

Staals RH, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K et al. 2013a. Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Molecular Cell* **52**: 135-145.

-. 2013b. Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell* **52**: 135-145.

Stern A, Sorek R. 2011. The phage-host arms race: shaping the evolution of microbes. *BioEssays : news and reviews in molecular, cellular and developmental biology* **33**: 43-51.

Sternberg SH, Haurwitz RE, Doudna JA. 2012. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* **18**: 661-672.

Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**: 62-67.

Stoll B, Maier LK, Lange SJ, Brendel J, Fischer S, Backofen R, Marchfelder A. 2013. Requirements for a successful defence reaction by the CRISPR-Cas subtype I-B system. *Biochemical Society transactions* **41**: 1444-1448.

Story RM, Steitz TA. 1992. Structure of the recA protein-ADP complex. *Nature* **355**: 374-376.

Swarts DC, Mosterd C, van Passel MW, Brouns SJ. 2012. CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7**: e35888.

Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, Pschera P, Siksnys V, Seidel R. 2014. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 9798-9803.

Tang TH, Polacek N, Zywicki M, Huber H, Brugger K, Garrett R, Bachellerie JP, Huttenhofer A. 2005. Identification of novel non-coding RNAs as potential antisense regulators in the archaeon Sulfolobus solfataricus. *Molecular Microbiology* **55**: 469-481.

Terns MP, Terns RM. 2011a. CRISPR-based adaptive immune systems. *Current Opinion in Microbiology* **14**: 321-327.

-. 2011b. CRISPR-based adaptive immune systems. *Curr Opin Microbiol* **14**: 321-327.

Terns RM, Terns MP. 2013. The RNA- and DNA-targeting CRISPR-Cas immune systems of *Pyrococcus furiosus*. *Biochem Soc Trans* **41**: 1416-1421.

-. 2014. CRISPR-based technologies: prokaryotic defense weapons repurposed. *Trends in Genetics : TIG* **30**: 111-118.

Tock MR, Dryden DT. 2005. The biology of restriction and anti-restriction. *Current Opinion in Microbiology* **8**: 466-472.

van der Oost J, Brouns SJ. 2009. RNAi: prokaryotes get in on the act. *Cell* **139**: 863-865.

van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ. 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* **34**: 401-407.

van der Oost J, Westra ER, Jackson RN, Wiedenheft B. 2014. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol* **12**: 479-492.

van Duijn E, Barbu IM, Barendregt A, Jore MM, Wiedenheft B, Lundgren M, Westra ER, Brouns SJ, Doudna JA, van der Oost J et al. 2012a. 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*. *Molecular & Cellular Proteomics : MCP* **11**: 1430-1441.

-. 2012b. Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in CRISPR-associated protein complexes from *Escherichia coli* and *Pseudomonas aeruginosa*. *Mol Cell Proteomics*.

Vestergaard G, Garrett RA, Shah SA. 2014. CRISPR adaptive immune systems of Archaea. *RNA Biology* **11**: 156-167.

Wang R, Preamplume G, Terns MP, Terns RM, Li H. 2011. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure* **19**: 257-264.

Wells L, Vosseller K, Cole RN, Cronshaw JM, Matunis MJ, Hart GW. 2002. Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* **1**: 791-804.

Westra ER, Nilges B, van Erp PB, van der Oost J, Dame RT, Brouns SJ. 2012a. Cascade-mediated binding and bending of negatively supercoiled DNA. *RNA Biol* **9**.

Westra ER, Semenova E, Datsenko KA, Jackson RN, Wiedenheft B, Severinov K, Brouns SJ. 2013. Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genetics* **9**: e1003742.

Westra ER, Swarts DC, Staals RH, Jore MM, Brouns SJ, van der Oost J. 2012b. The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annual review of genetics* **46**: 311-339.

Westra ER, van Erp PB, Kunne T, Wong SP, Staals RH, Seegers CL, Bollen S, Jore MM, Semenova E, Severinov K et al. 2012c. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular Cell* **46**: 595-605.

-. 2012d. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* **46**: 595-605.

Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, Doudna JA, Nogales E. 2011a. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**: 486-489.

Wiedenheft B, Sternberg SH, Doudna JA. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**: 331-338.

Wiedenheft B, van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A, Westphal W, Heck AJ, Boekema EJ, Dickman MJ et al. 2011b. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 10092-10097.

-. 2011c. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* **108**: 10092-10097.

Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA. 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **17**: 904-912.

Yosef I, Goren MG, Kiro R, Edgar R, Qimron U. 2011. High-temperature protein G is essential for activity of the *Escherichia coli* clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. *Proc Natl Acad Sci U S A* **108**: 20136-20141.

Yosef I, Goren MG, Qimron U. 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic acids research* **40**: 5569-5576.

Zebec Z, Manica A, Zhang J, White MF, Schleper C. 2014a. CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Research* **42**: 5280-5288.

-. 2014b. CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res*.

Zhang J, Kasciukovic T, White MF. 2012a. The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. *PLoS One* **7**: e47232.

Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, Graham S, Reimann J, Cannone G, Liu H, Albers SV et al. 2012b. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* **45**: 303-313.

-. 2012c. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Molecular cell* **45**: 303-313.

Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, Gong W, Wei Z, Wang Y. 2014. Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature*.

Figure 5.1. Overview of CRISPR-Cas immune systems in *Pyrococcus furiosus*. Each of seven CRISPR loci in *P. furiosus* is comprised of multiple direct repeats (black boxes) interspersed by unique invader-derived sequences (colored boxes). Precursor CRISPR RNAs (pre-crRNAs) are transcribed from promoters in leader region (brown box). Cas6 endonucleases cleave into every repeat to generate 1X intermediate RNAs followed by 3' end processing to generate multiple mature crRNA size forms. Mature crRNAs contain 5' tag (black), guide sequences (red) and sometimes 3' repeat (black). Diverse size forms of all encoded crRNAs are shared between three immune-effector complexes, Type I-A (Csa), Type I-B (Cst) and Type III-B (Cmr). The Cas superfamily proteins of effector complexes are color-coded as Figure 1.2 (Cas7 superfamily proteins (green), Cas5 superfamily proteins (yellow), large subunit proteins (purple), small subunit proteins (orange) and Cas3 (teal and blue)). Types I-A (Csa) and I-B (Cst) complexes mediate crRNA-guided, PAM-dependent cleavage of duplex DNA and Type III-B complex cleaves RNA targets in a homology dependent manner.

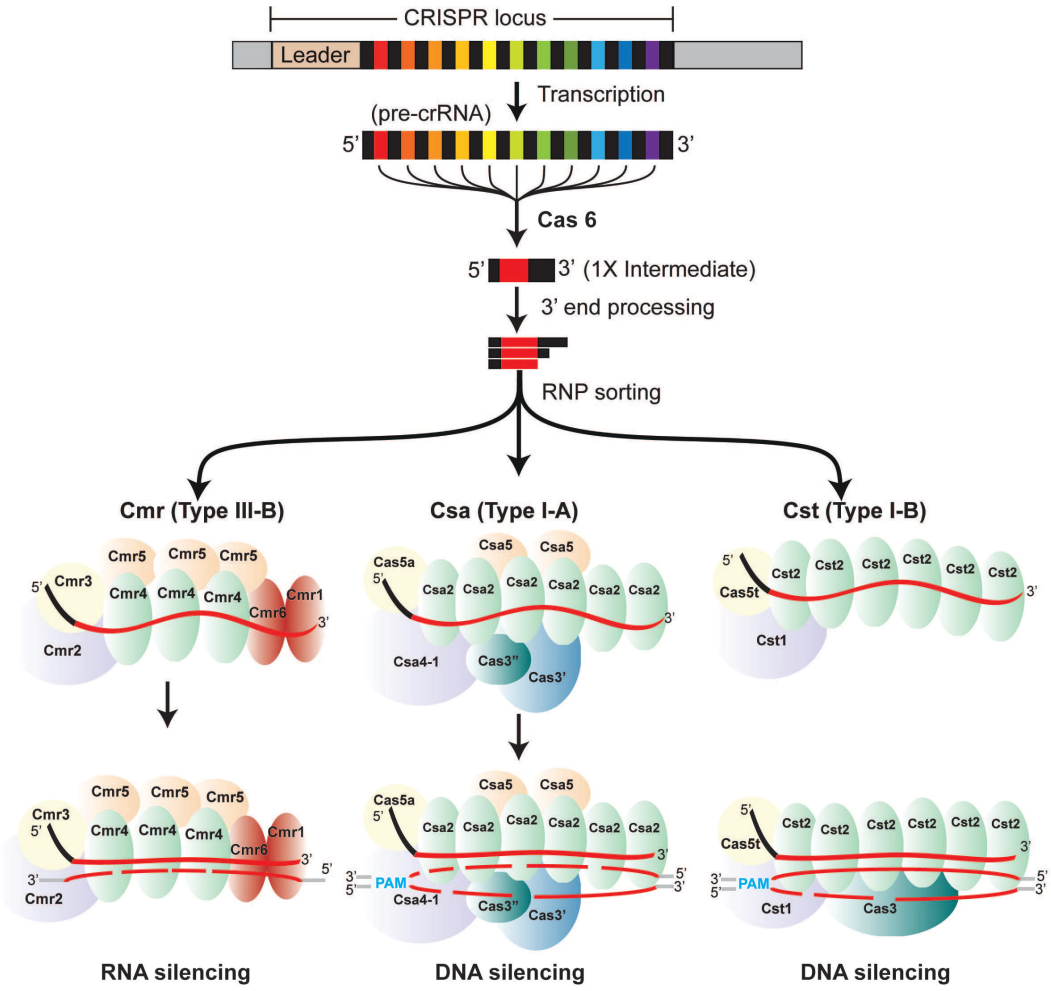

CRISPR locus
Leader
Transcription
(pre-crRNA)
5'
3'
Cas 6
5'
3' (1X Intermediate)
3' end processing
RNP sorting
Cmr (Type III-B)
Csa (Type I-A)
Cst (Type I-B)
Cmr5
Cmr3
Cmr4
Cmr1
Cmr6
Cmr2
Cas5a
Csa5
Csa2
Csa4-1
Cas3"
Cas3'
Cst2
Cas5t
Cst1
PAM
Cas3
RNA silencing
DNA silencing
DNA silencing

Figure 5.2. **Sorting of mature crRNAs into effector complexes in *P. furiosus*.** Pre-crRNAs are transcribed from promoters in leader (brown) and Cas6 (cyan) cleaves into each repeat (black boxes) to generate 1X intermediate RNAs. Processing of 3' repeat sequences generates mature crRNA size forms. Each crRNA size form contains 8-nt 5'-tag (black) followed by invader-derived guide (red) and either no 3' repeat (45 nt crRNA) or variable repeat sequences (black). crRNAs with 3' repeats assemble into Type I-A (Csa) and I-B (Cst) complexes but not Type III-B (Cmr) complex. 45-nt crRNA size forms that contain 5' tag and guide sequences are shared between all three complexes. 39-nt crRNA size form that harbors 5' tag and a truncated guide exclusively associates with Type III-B (Cmr) complex. All three (Csa, Cst, Cmr) complexes share mature crRNAs (45-nt) encoded from all seven CRISPR loci.

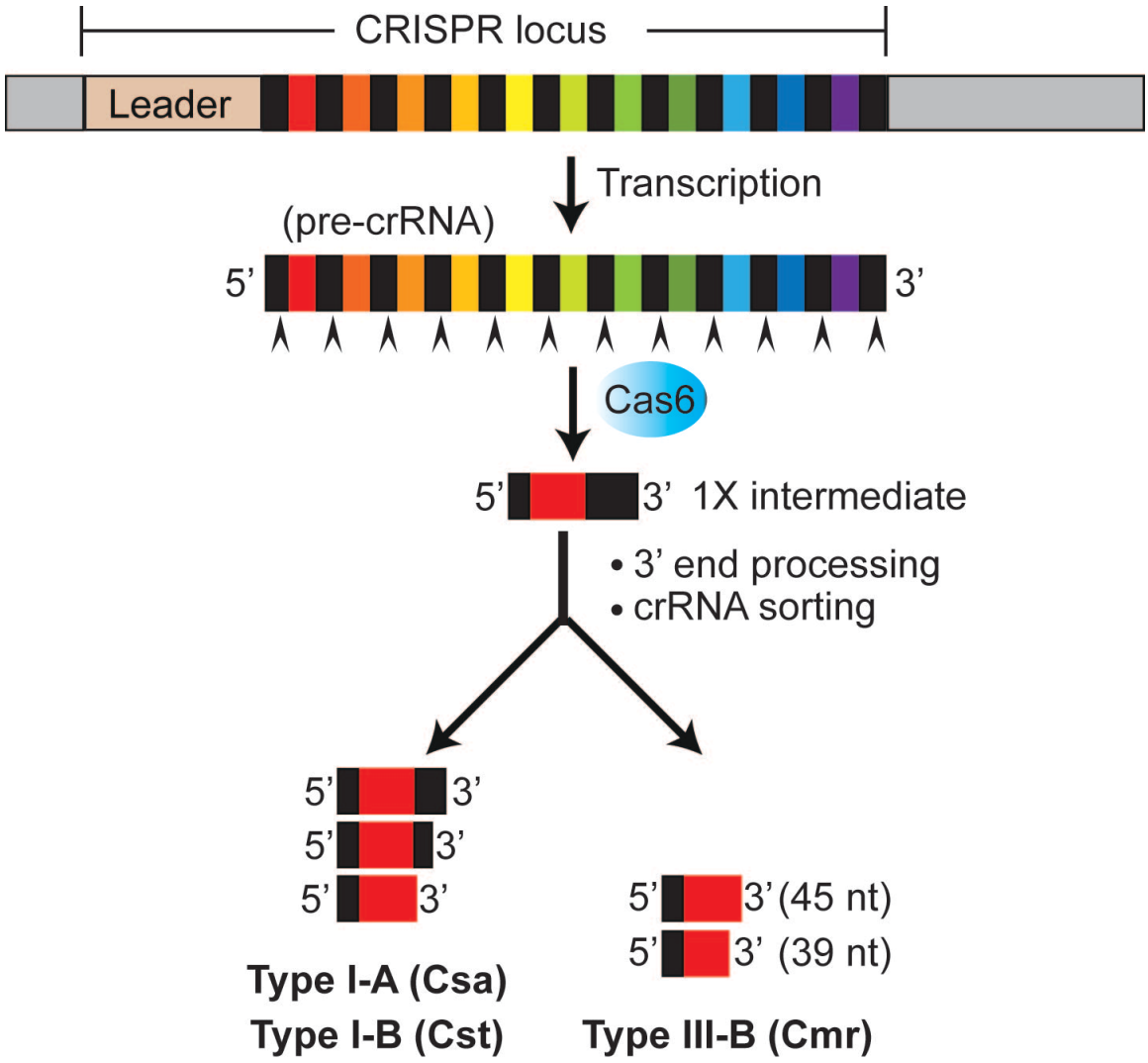

CRISPR locus
Leader
Transcription
(pre-crRNA)
5'
3'
Cas6
5'
3' 1X intermediate
• 3' end processing
• crRNA sorting
5'
3'
5'
3'
5'
3'
5'
3' (45 nt)
5'
3' (39 nt)
Type I-A (Csa)
Type I-B (Cst)
Type III-B (Cmr)

Figure 5.3. Compositional analogy between effector complexes of P. furiosus with other Type I complexes. We predict *P. furiosus* Type I-A (Csa), characterized Type I-E (Cse) and Type III-B (Cmr) complexes contain similar Cas superfamily protein localization. *P. furiosus* Type I-B (Cst) is predicted to share similar complex topology as Type I-F (Csy) complex. Schematics showing association of Cas5 superfamily proteins (yellow), Cas6 superfamily proteins (cyan), Cas7 superfamily proteins (green), large subunit proteins (purple) and small subunit proteins (orange) with crRNA (5'tag –black, guide- red) that is characterized in Types I-E (Cse), I-F (Csy) and III-B (Cmr) complexes and proposed for Types I-A (Csa) and I-B (Cst). Cas3"(effector nuclease, teal) and Cas3' (helicase, blue) are structural components of Type I-A (Csa) complex.

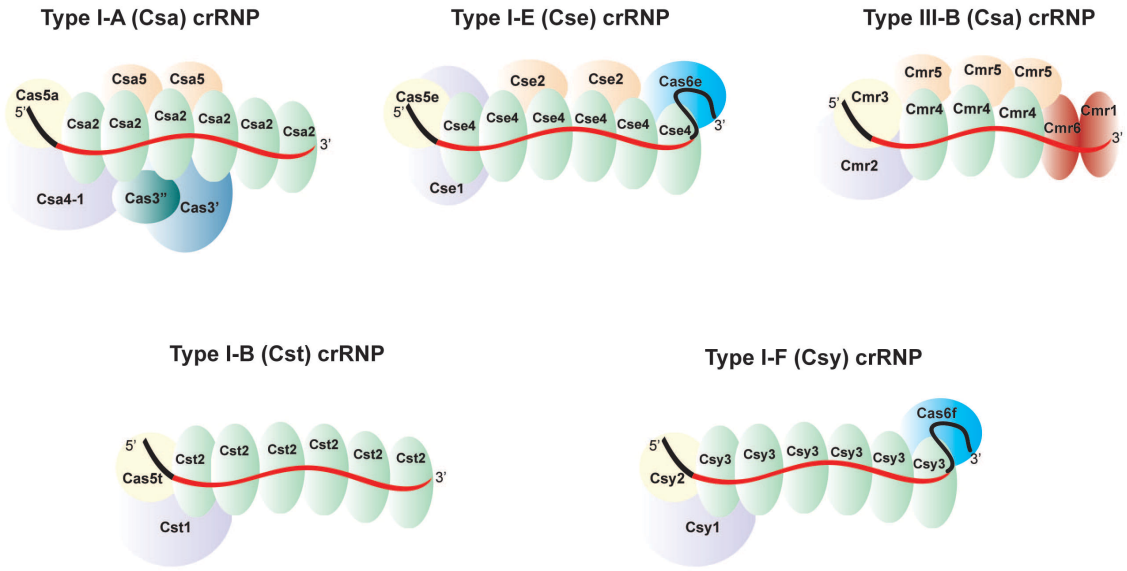
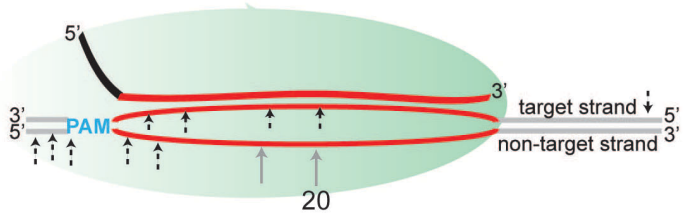
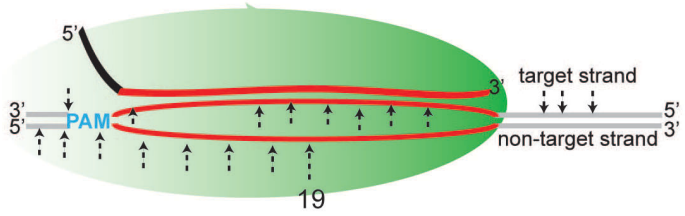

Type I-A (Csa) crRNP
Cas5a
Csa5
Csa5
5'
Csa2
Csa2
Csa2
Csa2
Csa2
Csa2
3'
Csa4-1
Cas3"
Cas3'
Type I-E (Cse) crRNP
Cas5e
Cse2
Cse2
Cas6e
5'
Cse4
Cse4
Cse4
Cse4
Cse4
Cse4
3'
Cse1
Type III-B (Csa) crRNP
Cmr5
Cmr5
Cmr5
5'
Cmr3
Cmr4
Cmr4
Cmr4
Cmr1
Cmr6
3'
Cmr2
Type I-B (Cst) crRNP
5'
Cst2
Cst2
Cst2
Cst2
Cst2
Cst2
Cas5t
3'
Cst1
Type I-F (Csy) crRNP
Cas6f
5'
Csy3
Csy3
Csy3
Csy3
Csy3
Csy3
3'
Csy2
Csy1

Figure 5.4. **Profile of DNA cleavages mediated by reconstituted Type I effector complexes**. Cleavages sites by Type I-A (Csa) complexes of *P. furiosus* (light green) and *T. tenax* (dark green), Type I-E (Cse) complexes of *E. coli* (bright orange) and *S. thermomphilus* (light orange) and Type I-B complex of *P. furiosus* (yellow) are mapped. 5' tag (black), guide (red) sequences of crRNA and PAM (cyan), protospacer (red), flanking sequences (grey) of DNA are shown. Target strand (complementary to crRNA-guide) and non-target strand of DNA are labeled on the right. ATP-independent cleavages (grey arrows with solid lines) on non-target strand and ATP-dependent cleavages (black arrows with dashed lines) on both strand of DNA are mapped. Cleavage sites marking the 3' extremity on non-target strand by each effector complex is labeled with a number (counted from PAM-protospacer junction).

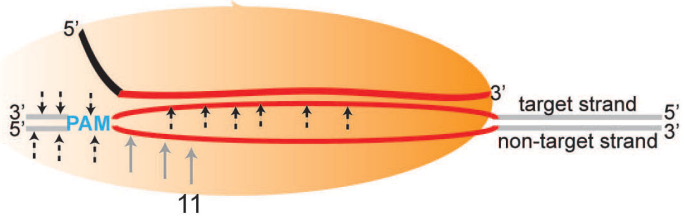
***P. furiosus* Type I-A (Csa)**



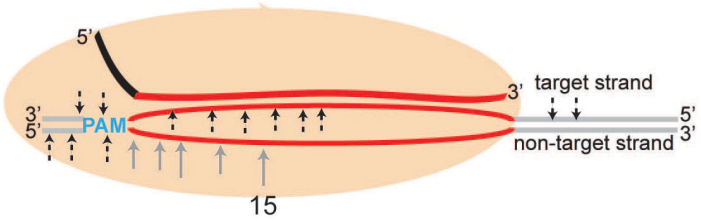
***T. tenax* Type I-A (Csa)**



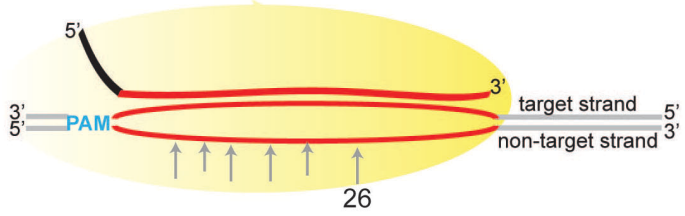
***E. coli* Type I-E (Cse)**



***S. thermophilus* Type I-E (Cse)**



***P. furiosus* Type I-B (Cst)**



APPENDIX

ESSENTIAL FEATURES AND RATIONAL DESIGN OF CRISPR RNAS THAT FUNCTION WITH THE CAS RAMP MODULE COMPLEX TO CLEAVE RNAS

---

[1]Caryn R. Hale, Sonali Majumdar, Joshua Elmore, Neil Pfister, Mark Compton, Sara Olson, Alissa M. Resch, Claiborne V. C. Glover III, Brenton R. Graveley, Rebecca M. Terns and Michael P. Terns. 2012. *Molecular Cell. 45*: 292-302.

Reprinted here with permission of the publisher.

Abstract

Small RNAs target invaders for silencing in the CRISPR-Cas pathways that protect bacteria and archaea from viruses and plasmids. The CRISPR RNAs (crRNAs) contain sequence elements acquired from invaders that guide CRISPR-associated (Cas) proteins back to the complementary invading DNA or RNA. Here, we have analyzed essential features of the crRNAs associated with the Cas RAMP module (Cmr) effector complex, which cleaves targeted RNAs. We show that Cmr crRNAs contain an 8-nucleotide 5' sequence tag (also found on crRNAs associated with other CRISPR-Cas pathways) that is critical for crRNA function and can be used to engineer crRNAs that direct cleavage of novel targets. We also present data that indicates that the Cmr complex cleaves an endogenous complementary RNA in *Pyrococcus furiosus*, providing direct *in vivo* evidence of RNA targeting by the CRISPR-Cas system. Our findings indicate that the CRISPR RNA-Cmr protein pathway may be exploited to cleave RNAs of interest.

Introduction

The CRISPR-Cas systems are RNA-based immune systems that protect prokaryotes from viruses, plasmids, and other invaders (see recent general reviews: (van der Oost and Brouns 2009; Deveau et al. 2010; Horvath and Barrangou 2010; Karginov and Hannon 2010; Marraffini and Sontheimer 2010a; Jore et al. 2011a; Makarova et al. 2011c; Terns and Terns 2011b)). The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) loci found in prokaryotic genomes acquire short fragments of invader sequence (~30-40 base pairs in length) that are inserted between short CRISPR repeat sequences. The CRISPR loci give rise to CRISPR RNAs (crRNAs) that each contain an invader-

derived (guide) sequence and interact with CRISPR-associated or Cas proteins to form effector complexes that recognize and silence the corresponding invader (Brouns et al. 2008; Hale et al. 2008; Hale et al. 2009; Jore et al. 2011a; Lintner et al. 2011d; Wiedenheft et al. 2011a). There are multiple modules of Cas proteins (termed for example Cse and Csy proteins) that can function independently with crRNAs to effect CRISPR-Cas defense (Haft et al. 2005a; Makarova et al. 2006a; Makarova et al. 2011c; Terns and Terns 2011b); thus distinct CRISPR-Cas pathways function in various prokaryotes.

CRISPR-Cas-mediated silencing appears to occur through sequence-specific targeting of RNA or DNA. In *Streptococcus thermophilus* (Barrangou et al. 2007; Garneau et al. 2010), *Escherichia coli* (Brouns et al. 2008; Semenova et al. 2011), *Staphylococcus epidermidis* (Marraffini and Sontheimer 2008), and *Sulfolobus solfataricus* (Gudbergsdottir et al. 2011; Manica et al. 2011), there is evidence that CRISPR-Cas systems target foreign DNA. Indeed, sequence-specific cleavage of both plasmid and bacteriophage DNA was recently documented in *S. thermophilus* (Garneau et al. 2010). At the same time, the Cmr-type crRNA-Cas protein complex from *Pyrococcus furiosus* specifically cleaves complementary RNAs (Hale et al. 2009). Thus, the various CRISPR-Cas systems (formed by distinct modules of Cas proteins) may function by different mechanisms, including DNA and RNA cleavage. Interestingly, many organisms possess more than one module of Cas proteins and may resist invaders by multiple mechanisms to provide robust protection from diverse invaders.

CRISPR RNAs (previously also referred to as prokaryotic silencing (psi)RNAs) are utilized by all versions of the CRISPR-Cas system characterized to date (Brouns et al.

2008; Marraffini and Sontheimer 2008; Hale et al. 2009; Deltcheva et al. 2011; Gudbergsdottir et al. 2011; Lintner et al. 2011d; Manica et al. 2011; Wiedenheft et al. 2011a). The CRISPR loci are transcribed from a leader sequence region of 100-500 basepairs found at one end of each locus (Jansen et al. 2002b), where promoter sequences have been identified in a few organisms (Lillestol et al. 2009; Pul et al. 2010). Insertion of new invader-derived sequences into CRISPR loci most frequently occurs immediately downstream of the leader (Pourcel et al. 2005; Barrangou et al. 2007; Lillestol et al. 2009). The CRISPR locus transcripts are cleaved within the repeat sequence to release the individual embedded crRNAs. A series of structurally related proteins from various CRISPR-Cas systems process CRISPR RNA transcripts: Cas6 (Carte et al. 2008b; Carte et al. 2010; Wang et al. 2011), Cse3 (Brouns et al. 2008; Gesner et al. 2011b; Sashital et al. 2011), and Csy4 (Haurwitz et al. 2010; Przybilski et al. 2011). Cleavage within the repeats by these Cas endonucleases produces 1X intermediate RNAs generally comprised of 8 nucleotides of repeat sequence (called the 5' tag), a guide sequence, and the remaining ~20-25 nucleotides (nt) of repeat sequence at the 3' end (the 3' tag) (Brouns et al. 2008; Hale et al. 2008; Haurwitz et al. 2010; Gesner et al. 2011b). These 1X intermediates undergo various amounts of 3' end trimming (Hale et al. 2008; Marraffini and Sontheimer 2008; Hale et al. 2009; Jore et al. 2011a; Lintner et al. 2011d; Wiedenheft et al. 2011a). However, an equivalent 8-nt 5' tag is present on mature crRNAs from a variety of bacterial and archaeal species including *E. coli* (Brouns et al. 2008), *S. epidermidis* (Marraffini and Sontheimer 2008), *Pseudomonas aeruginosa* (Haurwitz et al. 2010), *P. furiosus* (Hale et al. 2008; Hale et al. 2009), *S. solfataricus* ((Lintner et al. 2011d), our unpublished data), *Sulfolobus tokadai* (our unpublished data),

and *Thermococcus kodakarensis* (our unpublished data). In *S. epidermis*, the 5' tag of the crRNAs provides for discrimination between self (i.e., the complementary DNA strand in the CRISPR locus, which can basepair with the tag) and non-self (invader DNA, which does not basepair with the tag) (Marraffini and Sontheimer 2010b), but other functions of this crRNA signature sequence are not known.

There are 3 modules of Cas proteins (Cmr, Csa, and Cst) and 7 CRISPR loci encoded in the genome of the hyperthermophilic archaeon *P. furiosus.* Previously, we identified and characterized a complex comprised of the Cas module RAMP or Cmr proteins (Cmr1-6) and 2 crRNA species in *P. furiosus* (Hale et al. 2009). The Cmr-crRNA protein complex is unique among CRISPR-Cas systems characterized to date in that it targets RNA; the complex specifically cleaves RNAs that are complementary to the crRNAs, 14 nt from the 3' end of the crRNAs (Hale et al. 2009). Activity was observed *in vitro* using both native and reconstituted complexes, but has not been described *in vivo.* Here we have investigated the functionality of the crRNAs that direct cleavage by the Cmr complex. Deep sequencing of the RNAs associated with the complex as well as total crRNAs unexpectedly revealed the cleavage of a fortuitous endogenous target RNA by the Cmr complex *in vivo*. While multiple species of crRNAs are present in *P. furiosus*, the Cmr complex selectively includes 45- and 39-nt crRNAs regardless of the size of the invader-derived guide sequence encoded in the genome. We show that the 8-nt 5' tag sequence is not only necessary for Cmr-directed cleavage, but can be used to target the complex to cleave novel* RNAs, providing the potential for manipulation of the CRISPR-Cas system for experimental or pharmaceutical gene silencing.

Results

Key features of the Cmr complex RNAs

To delineate functional features of the crRNAs that guide cleavage by the Cmr complex, we isolated active Cmr complexes using antibodies against the Cmr2 protein. Polyclonal antibodies raised against recombinant *P. furiosus* Cmr2 specifically immunoprecipitated Cmr2 (assessed by Western blotting, Figure A.1A), as well as the five other Cmr complex proteins (Cmr1 and Cmr3-6; assessed by mass spectrometry, unpublished data), and 39- and 45-nt species of a crRNA (assessed by Northern blotting, Figure A.1B). Moreover, the immunoprecipitated RNA-protein complex replicated the cleavage activity observed with both the chromatographically-purified and reconstituted Cmr complexes (Hale et al. 2009), cleaving a target RNA (complementary to one of the endogenous crRNAs) at the two sites located 14 nt upstream of the 3' ends of the complementary 39- and 45-nt crRNA species (Figure A.1C).

We probed the nature of the end groups of the crRNAs immunopurified with Cmr complexes by end radiolabeling. 5' end labeling of the 39- and 45-nt RNAs was not affected by prior enzymatic treatment with phosphatase, whereas phosphatase pre-treatment significantly increased 3' end labeling of the RNAs (Figure A.1D), indicating that the majority of both species of crRNAs in the Cmr complex contain 5' hydroxyl and 3' phosphate end groups.

In order to further characterize the Cmr-associated crRNAs, we deep-sequenced the RNAs immunopurified with the complex, as well as total *P. furiosus* small RNAs. We isolated RNAs smaller than ~70 nt by gel purification, and based on the results of end group analysis (Figure A.1D), treated the RNAs with phosphatase to optimize cloning.

The cDNA libraries were subject to Illumina sequencing. Figure 2A shows the numbers of small RNA sequence reads that map to each of the seven *P. furiosus* CRISPR loci from the total RNA and Cmr complex samples. A total of 2,632,773 uniquely mapped small RNA reads were obtained from the total RNA sample, of which 54.5% mapped to the CRISPR loci. In contrast, 91.2% of the 990,060 uniquely mapped reads from the Cmr complex sample mapped to the CRISPR loci.

The total RNA profiles indicate that RNAs are expressed from all seven *P. furiosus* CRISPR loci (Figure A.2A, total RNA) as limited sequencing data had previously suggested (Hale et al. 2009). We analyzed the leader sequences of the seven CRISPR loci and found canonical BRE/TATA promoter sequences at a conserved position within the leaders (Figure A.2B). A population of RNAs was detected that begins ~21 nt downstream of these promoter elements (and includes ~29 nt of the leader sequence at the 5' end, indicated in red in Figure A.2B). (These RNAs do not represent the major crRNA species, which are described below.) These findings suggest that the CRISPR loci are transcribed by archaeal RNA polymerase via recognition of the BRE/TATA promoters within the leader sequences. The deep sequencing data also indicate that, in general, crRNAs from the leader-proximal regions of the CRISPR loci are more abundant than crRNAs encoded downstream, as had been hypothesized based on previous data (Hale et al. 2008). The molecular basis for the gradient in steady state levels of the crRNAs from each locus is currently unknown, but could reflect differences in transcription, processing, or stability. The expression gradient is noteworthy because the crRNAs that arise from the leader-proximal end of a CRISPR locus generally target the

most recent invaders confronted by the system (Pourcel et al. 2005; Barrangou et al. 2007; Lillestol et al. 2009).

Deep sequencing and Northern analysis indicate that a series of species of crRNAs of different lengths are present in *P. furiosus*, and that the Cmr complex specifically loads two of these crRNA species. The 5' ends of the vast majority of crRNAs in *P. furiosus,* in both the total and Cmr complex samples, are comprised of an identical 5' end tag sequence (AUUGAAAG) derived from the upstream repeat element (Figure A.3A). This end is generated by Cas6 cleavage of CRISPR transcripts at this site (Carte et al. 2008b; Carte et al. 2010; Wang et al. 2011). In total RNA, a mixture of 3' ends corresponding to crRNAs 30 to 70 nt in length is observed (Figure A.3A), consistent with a series of RNA species detected in total RNA by Northern analysis of one of the crRNAs (Figure 1B, T lane). Strikingly, the 39- and 45-nt crRNA species selectively co-purify with the Cmr complex (Figures A.1B and A.3A). The 45-nt crRNAs constitute 46% of the crRNAs associated with the Cmr complex, and the 39-nucleotide species make up 38% (Figure A.3A).

Interestingly, the deep sequence analysis indicates that the 3' ends of the Cmr complex-associated crRNAs are defined by distance from the 5' tag or absolute length of the crRNA (rather than, for example, by the length of the guide sequence encoded in the CRISPR locus or by position relative to the 3' repeat element). Guide sequences in *P. furiosus* CRISPRs are most commonly 37 nt long, in which case the 45-nt crRNA species is comprised of an 8-nt repeat tag sequence and a full 37-nt guide sequence. (The 39-nt species simply lacks 6 nt of the guide sequence at the 3' end, see Figure A.3A.) However some guide sequences are as short as 34 nt and as long as 59 nt in *P. furiosus*, and Cmr

complex crRNAs are nonetheless consistently 39 and 45 nt in length. For example, while three of the first 5 guide sequences in CRISPR locus 4 are 37 nt long, the second guide sequence in the locus (termed 4.02) is 35 nt; yet the 4.02 crRNAs found in the Cmr complex are 45 and 39 nt in length (Figure A.3B). The 45-nt 4.02 crRNA includes 2 nt of the downstream repeat sequence. Conversely, the fifth CRISPR 4 guide sequence is one nt longer than standard (38 nt), and the 45-nt 4.05 crRNA lacks the last encoded nt of the guide sequence (Figure A.3B). Similarly, the first guide sequence in CRISPR 1 is 48 nt long (11 nt longer than is typical), and the 45-nt 1.01 crRNA found in the Cmr complex includes only 37 nt of the 48-nt guide sequence (see Figure A.4B). Thus, regardless of the size of the guide sequence encoded in the genome, the crRNAs of the Cmr complex are comprised of an 8-nt 5' repeat tag and a 37- or 31-nt guide sequence.

Cleavage of a complementary RNA *in vivo*

Very few RNAs were detected that were transcribed from the opposite (non-leader) ends of the CRISPR loci in *P. furiosus*; however, we did detect significant antisense transcription from an apparently fortuitous promoter within CRISPR locus 1. A BRE/TATA promoter sequence is present in the antisense orientation within the second guide sequence of CRISPR locus 1, and a significant number of antisense RNAs that begin distinctly 21 nt downstream of the promoter are present in the sequence from the total RNA sample (Figure A.S1, blue reads). The number of antisense RNA reads is approximately 1/3 the number of reads observed for the sense strand crRNA reads (crRNA 1.01; Figure A.S1, red reads). Northern analysis reveals a series of antisense

RNAs up to ~140 nt in length in total *P. furiosus* RNA (Figure A.4A, 1.01 antisense, total (T) lane).

The antisense CRISPR transcripts found in *P. furiosus* would be predicted to be recognized and targeted by Cmr complexes containing the crRNA encoded on the opposite strand of CRISPR locus 1 (crRNA 1.01). Previous *in vitro* studies established that the Cmr complex cleaves complementary RNAs 14 nt upstream of the 3' ends of the 39- and 45-nt crRNAs (Hale et al. 2008). Accordingly, in this case, the 1.01 crRNAs would be predicted to generate 5' products of 45 and 39 nt (as well as 3' products of various lengths) upon cleavage of the antisense transcripts. And indeed, distinct antisense RNAs of 45 and 39 nt were observed in Northern analysis of the total RNA (with a probe to the 5' region of the antisense transcript; Figure A.4A, 1.01 antisense, total (T) lane), suggesting that the Cmr complex targets and cleaves the endogenous CRISPR1 antisense transcript in *P. furiosus*.

We also found in both RNA sequencing and Northern analysis that the 45- and 39-nt putative 5' cleavage products are specifically co-immunopurified with the Cmr complex (Figure A.4A, 1.01 antisense, immune (I) vs. preimmune (PI) lanes; and Figure A.4B, blue reads vs. Figure A.S1, blue reads), suggesting that the products remain associated with the complex to some extent following cleavage. The 45- and 39-nt antisense RNAs do not possess the 5' repeat tag sequence, which we hypothesize is important for recognition and function of crRNAs with the Cmr complex. However, because of the coincidence in the size of these RNAs with Cmr complex crRNA species, we tested the possibility that the antisense RNAs are functional guide RNAs. We found that the immunopurified Cmr complexes do not cleave a target RNA complementary to the

antisense RNAs, indicating that the antisense RNAs do not guide cleavage by the Cmr complex (Figure A.4C, antisense RNA target). At the same time, the immunopurified complexes do cleave an RNA corresponding to the ~140-nt antisense RNA transcript, generating 5' products of the same sizes observed *in vivo* (45 and 39 nt; Figure A.4C, crRNA target). Together, the findings indicate that the Cmr complex recognizes and cleaves endogenous complementary RNAs *in vivo*.

The CRISPR repeat tag sequence is required for functional Cmr complex crRNAs

Its conservation suggests that the 5' CRISPR repeat tag is an important element of functional Cmr complex crRNAs. We investigated the importance of the tag by testing several tag sequence mutants for the ability to guide cleavage by reconstituted Cmr complexes. Removal of the repeat tag sequence from crRNA 7.01 eliminates cleavage of a complementary target RNA (Figure A.5A). Substitution of the entire tag sequence to its complement (AUUGAAAG → UAACUUUC) or of the first two nucleotides of the tag sequence (AUUGAAAG → GAUGAAAG) also prevents function of the crRNA, as does the addition of a G at the 5' end of the tag (Figure A.5B).

Analysis of CRISPR locus 8 provided further insight into the basis for the importance of the 5' repeat tag *in vivo*. crRNAs from locus 8 were significantly underrepresented in the Cmr complex (Figures A.2A and A.S2). The reduced presence of CRISPR locus 8 RNAs in the Cmr complex correlates with a difference in the 5' repeat tag found on these RNAs. Most of the crRNAs from locus 8 possess a 7-nt AUUGAAG (rather than 8-nt AUUGAAAG) repeat tag due to a variation in the repeat sequences within this CRISPR locus (Figure A.5C). The exception is crRNA 8.11 (the last crRNA encoded by locus 8),

which retains the canonical 8-nt repeat tag sequence, and this crRNA is selectively enriched in the Cmr complex relative to the other crRNAs from CRISPR locus 8 (Figures A.2A and A.5C). Together, these results indicate that the CRISPR repeat tag is critical for the formation of functional Cmr complexes by crRNAs.

For CRISPR-Cas systems that target DNA, the 5' repeat tag on the crRNAs provides a key function in distinguishing self from non-self. In organisms with these systems, it is important for crRNAs to avoid targeting the complementary DNA sequence found in the host CRISPR locus itself, and evidence indicates that, in these systems, potential targets with complementarity to the tag sequence (found in the CRISPR locus but not in invader targets) are not silenced (Marraffini and Sontheimer 2010b). We tested the effect of target complementarity to the tag sequence on RNA cleavage by the Cmr complex, and observed no significant reduction in activity (Figure A.5D). The antisense transcript that is cleaved *in vivo* by the Cmr complex also is complementary to the tag of crRNA 1.01 (Figure A.4). Thus, RNA targeting by the Cmr complex does not require a lack of complementarity between the target and the crRNA repeat tag.

crRNAs can be engineered to direct the Cmr complex to cleave novel target RNAs

Our findings indicate that a 5' repeat tag is necessary for a crRNA to function to guide cleavage by the Cmr complex. To determine whether the Cmr complex can be directed to cleave novel target RNAs by engineered crRNAs, we designed crRNAs with the 5' AUUGAAAG tag sequence and guide sequences complementary to a random target sequence (X') or to a sequence near the 5' end of the β-lactamase (*bla*) mRNA. Cmr complexes reconstituted with the crRNA designed to cleave the random target

sequence specifically cleaved the X' target RNA but not a *bla* target sequence (Figure A.6A). Similarly, the crRNA designed against the 5' region of the *bla* mRNA directed cleavage of the *bla* target sequence but not the X' target RNA (Figure A.6A). The cleavage product sizes are consistent with target cleavage 14 nt from the 5' ends of the crRNAs. The 5' *bla* crRNA and a crRNA designed against an internal region of the *bla* mRNA were also tested for the ability to direct cleavage in the context of the full-length *bla* mRNA. The 5' radiolabeled *bla* mRNA (~860 nt) was cleaved by Cmr complexes reconstituted with both *bla* crRNAs, yielding products of the expected sizes (~17 nt for the 5' *bla* crRNA and ~236 nt for the internal *bla* crRNA) (Figure A.6B). These results demonstrate that the Cmr complex can be directed to cleave novel target RNAs at a desired site by crRNAs designed with an appropriate repeat tag sequence and a guide region complementary to the target cleavage site.

Discussion

The RNA targeting branch of the CRISPR-Cas immune system

There is tremendous diversity in the CRISPR-Cas immune pathways that protect prokaryotes from invaders (recently reviewed in: (van der Oost et al. 2009; Deveau et al. 2010; Garneau et al. 2010; Karginov and Hannon 2010; Marraffini and Sontheimer 2010a; Garrett et al. 2011b; Jore et al. 2011a; Makarova et al. 2011c; Terns and Terns 2011b)). The different CRISPR-Cas systems are populated by distinct modules of proteins, and the mechanisms of key steps in the pathways, including crRNA biogenesis and invader silencing, can be fundamentally different. In this work, we investigated the Cmr effector complex, which is unique among systems characterized to date in that it

targets RNAs rather than DNAs (Hale et al. 2009). This RNA-targeting branch of the CRISPR-Cas immune system appears to play an important role in the war against viruses and other mobile genetic elements given that approximately 70% of the archaea and 30% of the bacteria with CRISPR-Cas systems include the Cmr module (Garrett et al. 2011a). Our results provide direct evidence that the CRISPR-Cmr system cleaves complementary RNAs *in vivo* (Figure A.4). Moreover, we have identified the key structural features of the crRNAs associated with active Cmr complexes (Figures A.1-A.3, A.5) and engineered crRNAs capable of directing Cmr complexes to efficiently cleave selected RNA targets at predictable sites (Figure A.6).

Anatomy of the Cmr complex guide RNAs

The crRNAs associated with the Cmr complex are of two sizes, 39 and 45 nt in length, both with an 8-nt CRISPR repeat sequence at the 5' end (5' tag) followed by either 31 or 39 nt of invader guide sequence (Figures A.1 and A.3, and (Hale et al. 2009)). Unlike other crRNAs found in *P. furiosus*, the Cmr crRNAs contain little or no CRISPR repeat sequence at the 3' end (Figure A.3). Our studies revealed that the crRNAs present in Cmr complexes are essentially restricted to an absolute length of 39 or 45 nt, which may be defined by a mechanism that measures from the 5' end of the crRNA or from the 5' tag sequence, but does not depend on the position of the guide sequence-3' repeat sequence boundary (Figure 3). Both the 39- and 45-nt crRNAs have 5' hydroxyl and 3' phosphate end groups (Figure A.1D). Overall, the 39- and 45-nt forms are found in roughly equal proportion in the purified Cmr complexes (Figure 3A); however, the proportion of the two forms for an individual crRNA can vary significantly (see crRNAs

4.01 and 4.05, Figure 3B), and the basis for this variation is unclear. It is not yet known whether individual Cmr complexes contain one or more crRNAs, however, functional Cmr complexes can be reconstituted with either the 39- or 45-nt crRNA species, indicating that both forms are not simultaneously required for cleavage activity and hinting that each Cmr complex may harbor a single crRNA.

While the 5' ends of the Cmr complex crRNAs are presumably generated by cleavage of CRISPR transcripts by the site-specific riboendonuclease Cas6 (Carte et al. 2008b; Carte et al. 2010; Wang et al. 2011), we do not yet know how the 3' ends of mature crRNAs are generated or how the 39- and 45-nt crRNAs are selectively incorporated into Cmr complexes. This exquisite selectivity may simply be because the Cmr complex proteins preferentially bind crRNAs of these lengths. Alternatively, the proteins may associate with longer crRNA species (e.g., the 1X intermediate generated by Cas6 cleavage that includes an additional 22 nt of the repeat at the 3' end) and the 3' ends may be generated following formation of the complexes. The nuclease(s) responsible for crRNA 3' end trimming could be an intrinsic subunit of the Cmr complex or an extrinsic enzyme that may or may not be specific for crRNA processing. The precisely measured 3' ends of the Cmr complex crRNAs could result, for example, from a combination of non-specific nuclease activity and specific RNA protection by Cmr proteins. The 45-nt crRNA and the other crRNA size forms found in *P. furiosus* (which contain longer 3' ends; see Figures A.1 and A.3A) are components of other RNP complexes currently under investigation ((Hale et al. 2008), and our unpublished data). The *P. furiosus* genome encodes 21 Cas proteins in addition to the 6 Cmr proteins, including Csa and Cst subtype proteins that are also expected to associate with crRNAs to form additional

effector complexes. Future work will elucidate the mechanisms by which distinct crRNA species are sorted among CRISPR-Cas complexes within a cell.

The 5' tag is critical for Cmr complex crRNA function

The results presented here indicate that the 5' tag sequence is necessary for crRNA function in Cmr complexes (Figure A.5). We found that the sequence, and perhaps size, of the 5' tag is important for Cmr complex crRNA function; even subtle mutations are detrimental to Cmr activity. We hypothesize that the 5' tag is important as a binding site for one or more Cmr proteins. Consistent with this idea, we found that crRNAs from CRISPR locus 8 with variant 5' repeat tags are preferentially excluded from Cmr complexes *in vivo* (Figure A.5C). As discussed above, the basis for exclusion of longer crRNA species, which possess the 5' tag, from Cmr complexes in *P. furiosus* is not yet clear. The presence of a 5' tag sequence is a conserved feature of crRNAs in many CRISPR-Cas systems (Brouns et al. 2008; Hale et al. 2008; Marraffini and Sontheimer 2008; Hale et al. 2009; Haurwitz et al. 2010; Lintner et al. 2011d). A correlation exists between specific CRISPR-Cas modules and the CRISPR repeat sequences found together in various organisms, suggesting that the different systems may recognize characteristic 5' tags. However, in *P. furiosus* the crRNAs encoded by the 7 CRISPR loci generally share the same 5' tag sequence, which evidently functions with the Csa and Cst as well as Cmr modules. We hypothesize that the 5' tag is also important in effector complex function in the other systems where it is found. Our results reveal the functional importance of the 5' tag in Cmr effector complex function.

Directing Cmr complexes to cleave RNAs of choice

In this work, we have demonstrated that crRNAs can be designed to target the Cmr complex to cleave novel RNA substrates (Figure A.6). crRNAs comprised of a 37-nt complementary guide sequence and 5'-tag were used to direct specific cleavage of RNAs including the b-lactamase mRNA *in vitro*. Cleavage product sizes indicate that the target was predictably cut 14 nt downstream of the 3' end of the engineered crRNA. These findings will enable efforts to exploit the site-specific cleavage function of Cmr-crRNA complexes for *in vivo* gene knockdown experiments. Our studies have defined the minimal components for Cmr complex function (Cmr1-6 and a crRNA; Figure A.6 and (Hale et al. 2009)), making it conceivable to express functional Cmr complexes in eukaryotes as well as additional prokaryotes to provide a novel method of RNA silencing. Our results indicate that Cmr-crRNA complexes have the potential to target destruction of mRNAs encoding any gene of interest and to become important research tools with industrial and biomedical applications.

RNA targeting vs. DNA targeting CRISPR-based gene silencing systems

Our findings demonstrate that there are significant differences in the modes of target nucleic acid recognition for CRISPR-based gene silencing systems that target RNA and DNA. For example, while DNA targeting by at least some CRISPR-Cas systems requires a short PAM (protospacer adjacent motif) in the invading DNA target (Deveau et al. 2008; Mojica et al. 2009; Garneau et al. 2010; Gudbergsdottir et al. 2011; Semenova et al. 2011), our work suggests that Cmr complexes function independently of PAMs. Cmr-crRNA complexes can cleave target RNAs that completely lack flanking sequences or

contain a variety of flanking sequences (that do not match predicted PAMs; Figures A.1, A.4, and A.6 and (Hale et al. 2009)). For DNA-targeting systems, the requirement for a PAM sequence in the invader avoids damage to the host genome at the site of the corresponding guide sequence in the CRISPR locus (where PAMs are not found in the flanking repeat sequences). The DNA-targeting system of *Staphylococcus epidermidis* further demonstrates an explicit requirement for mismatches between the 5' tag of the crRNA and the target DNA strand, which also prevents targeting the host CRISPR locus (Marraffini and Sontheimer 2010b). In contrast, for the RNA-targeting Cmr complex (where host genome damage is not a risk) target RNAs containing sequences with perfect homology to the 5' tag of the crRNA are efficiently cleaved (Figures A.4 and A.5D). Evidence indicates that there are also differences in the interactions with the guide region of the crRNA. A specific 7-8 nt region of the crRNA located immediately downstream of the 5' tag (dubbed the "seed sequence") was found to be important for initial high affinity recognition of targets for certain DNA-targeting CRISPR-Cas systems (Semenova et al. 2011; Wiedenheft et al. 2011b). Base-pairing between the 3' end of the crRNA and target DNA was dramatically less important. The opposite trend is observed for RNA targeting, where interaction with the 3' end of the crRNA is critical for function and target RNAs that lack the "seed sequence" region are efficiently cleaved (Hale et al. 2009). The molecular bases of these differences in CRISPR-Cas RNA and DNA targeting pathways remain to be explored. The lack of restrictions in the flanking sequence content in RNA targeting provides for more flexibility in substrate interaction.

Experimental procedures

Antibody preparation and co-immunoprecipitation reactions

IgY antibodies were prepared in *Gallus gallus* against recombinant Cmr2 from *Pyrococcus furiosus* as described (Carte et al. 2010). Pre-immune and immune Cmr2 antibodies were either coupled to anti-IgY conjugated agarose beads (Gallus Immunotech, for experiments resulting in Northern analysis, end analysis or small RNA sequencing), or directly linked to CarboLink beads (Pierce, for Western analysis). The resulting beads were used for co-immunoprecipitation of Cmr2-containing complexes from P. furiosus cell extract. See Supplemental Experimental Procedures for details.

Co-IP/Western analysis

For Western analysis, 1/3 of a co-IP reaction was resuspended in non-reducing loading dye. The co-IP samples were eluted at 60°C for 5 minutes, centrifuged briefly, and loaded onto 12.5% SDS-PAGE gels. 1 μl of S100 extract (~15 μg of total protein) was loaded for comparison (T). S100 sample was boiled in reducing loading dye. The resulting gels were subject to Western blotting using standard procedures. See Supplemental Experimental Procedures for details.

Co-IP/Northern analysis

For Northern analysis of co-immunopurified RNAs, Trizol LS reagent (Invitrogen) was used to isolate RNAs directly from the co-IP sample. RNA isolated from S100 was used for the total RNA (T) lanes for all Northerns. Northern analysis was performed as described previously (Hale et al. 2008) with modifications (see Supplemental

Experimental Procedures). Probe sequences can be found in Supplemental Table A.1.

Co-IP/Activity assays

For Cmr activity assays, co-IP beads were washed into 40 mM Hepes pH 7.0, 500 mM KCl, and resuspended to a 50% slurry. 10 μl of the slurry (1/6 of a co-IP reaction) was added to a 20 μl cleavage reaction containing radiolabeled target RNA (see Supplemental Experimental Procedures for details). Reactions were incubated for 3 hours at 70°C, with mixing to resuspend beads every 1 hour. Proteins were removed by proteinase K, and the resulting RNAs were isolated by phenol extraction and ethanol precipitation. The resulting RNAs were separated on 15% 7M urea acrylamide gels and visualized on phosphor imaging screens.

Co-IP/End analysis

For analysis of the end groups of the co-immunopurified small RNAs, Trizol LS was used to isolate small RNAs from one co-IP reaction. One half of the small RNAs were phosphatase treated using Thermostable Shrimp Alkaline Phosphatase (TSAP, Promega). Half of each the treated and non-treated RNAs was subject to either: 1) 5' radiolabeling with gamma-[^{32}P]-ATP (MP Biomedicals) via polynucleotide kinase (PNK, Ambion), or 2) 3' radiolabeling via T4 ssRNA Ligase 1 (NEB) in a reaction containing 1 μl cytidine-3',5'-bisphosphate [5'-^{32}P] (3,000 ci/mmol, MP Biomedicals). Both labeling reactions were carried out for 1 hour at 37°C. The resulting RNAs were isolated, separated by denaturing gel electrophoresis and visualized on phosphor-imaging screens (Amersham).

Small RNA library preparation

For preparation of total and Cmr-associated small RNA libraries, Trizol was used to extract RNAs directly from *P. furiosus* cells or from one co-IP reaction, respectively. 50 μg of total RNA was used for total small RNA library preparation. See Supplemental Experimental Procedures for details about library preparation. Briefly, the RNAs were phosphatase treated to remove 3' phosphates, followed by 3' ligation to an adaptor. The resulting products were gel purified and treated with T4 polynucleotide kinase to enable 5' ligation to an additional adaptor. The 5' and 3' ligated products were then subject to reverse transcription, RNase H digestion and PCR.

Illumina sequencing and analysis

cDNA libraries were visualized by electrophoresis, quantitated by Agilent Bioanalyzer and Nano-drop, diluted to 2 pM, and subject to 76 cycles of sequencing on the Illumina Genome Analyzer IIx. Sequence reads were trimmed of the 3' linker and reads 18-76 nt in length were aligned to the *P. furiosus* genome using Bowtie {Langmead, 2009 #171}. The 5' and 3' ends of reads mapping to the CRISPR loci were calculated using a custom PERL script available upon request.

Recombinant Cmr protein production and activity assays

Expression and purification of recombinant Cmr proteins was performed as described (Hale et al. 2009) with modifications (See Supplemental Experimental Procedures). Recombinant Cmr assays using 500 nM of each Cmr protein were performed as described (Hale et al. 2009). Briefly, recombinant proteins were pre-incubated with

unlabeled crRNA (0.05 pmol) prior to addition of 5' radiolabeled target RNAs. Reactions were incubated for 1-3 hours at 70°C. Following incubation, reactions were subject to proteinase K treatment, PCI and ethanol precipitation, and separated by denaturing gel electrophoresis. The gels were dried and visualized on phosphor imaging screens (Amersham).

RNA preparation

Sequences of RNA/DNA oligos used for RNA preparation are found in Supplemetal Table 1, and detailed information about RNA preparation is found in the Supplemental Experimental Procedures.

Acknowledgments

We thank the UCHC Translational Genomics Core Facility for use of the Illumina Genome Analyzer IIx. This work was supported by National Institutes of Health grants RO1GM54682 (including American Recovery and Reinvestment Act funds) to M.P.T. and R.M.T. and R01GM062516 to B.R.G.

References

Andersson, A.F., and Banfield, J.F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. Science *320*, 1047-1050.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

Carte, J., Pfister, N.T., Compton, M.M., Terns, R.M., and Terns, M.P. (2010). Binding and cleavage of CRISPR RNA by Cas6. RNA *16*, 2181-2188.

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

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol *190*, 1390-1400.

Deveau, H., Garneau, J.E., and Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol *64*, 475-493.

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

Garrett, R.A., Shah, S.A., Vestergaard, G., Deng, L., Gudbergsdottir, S., Kenchappa, C.S., Erdmann, S., and She, Q. (2011). CRISPR-based immune systems of the Sulfolobales: complexity and diversity. Biochem Soc Trans *39*, 51-57.

Gesner, E.M., Schellenberg, M.J., Garside, E.L., George, M.M., and Macmillan, A.M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. Nat Struct Mol Biol.

Gudbergsdottir, S., Deng, L., Chen, Z., Jensen, J.V., Jensen, L.R., She, Q., and Garrett, R.A. (2011). Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. Mol Microbiol *79*, 35-49.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol *1*, e60.

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

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945-956.

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

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science *327*, 167-170.

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

Jore, M.M., Brouns, S.J., and van der Oost, J. (2011a). RNA in Defense: CRISPRs Protect Prokaryotes against Mobile Genetic Elements. Cold Spring Harb Perspect Biol.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R.*, et al.* (2011b). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol.

Karginov, F.V., and Hannon, G.J. (2010). The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol Cell *37*, 7-19.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J.*, et al.* (2003). The UCSC Genome Browser Database. Nucleic Acids Res *31*, 51-54.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

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

Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q., Copie, V.*, et al.* (2011). Structural and functional characterization of an archaeal CASCADE complex for CRISPR-mediated viral defense. J Biol Chem.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct *1*, 7.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F.*, et al.* (2011). Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol *9*, 467-477.

Manica, A., Zebec, Z., Teichmann, D., and Schleper, C. (2011). *In vivo* activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. Mol Microbiol.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843-1845.

Marraffini, L.A., and Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568-571.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733-740.

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

Przybilski, R., Richter, C., Gristwood, T., Clulow, J.S., Vercoe, R.B., and Fineran, P.C. (2011). Csy4 is responsible for CRISPR RNA processing in *Pectobacterium atrosepticum*. RNA Biol *8*.

Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., and Wagner, R. (2010). Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol *75*, 1495-1512.

Sashital, D.G., Jinek, M., and Doudna, J.A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nat Struct Mol Biol.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci U S A *108*, 10098-10103.

Shah, S.A., and Garrett, R.A. (2011). CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. Res Microbiol *162*, 27-38.

Shah, S.A., Hansen, N.R., and Garrett, R.A. (2009). Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. Biochem Soc Trans *37*, 23-28.

Terns, M.P., and Terns, R.M. (2011). CRISPR-based adaptive immune systems. Curr Opin Microbiol.

van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M., and Brouns, S.J. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem Sci *34*, 401-407.

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

Wiedenheft, B., van Duijn, E., Bultema, J., Waghmare, S., Zhou, K., Barendregt, A., Westphal, W., Heck, A., Boekema, E., Dickman, M*., et al.* (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc Natl Acad Sci U S A.

Figure A.1. **Cmr2 antibodies immunopurify active Cmr-crRNA complexes.** A) *Cmr2 antibodies recognize and specifically immunoprecipitate ~100 kD protein. P. furiosus* S100 extract (T) and PfCmr2 antibody immune (I) and pre-immune (PI) immunoprecipitation (IP) samples were analyzed by Western blotting with Cmr2 antibodies. Positions of size markers are indicated. ~100 kD protein specifically recognized by immunoblotting and immunoprecipitated by Cmr2 antibodies is indicated by arrow. B) *45- and 39-nt crRNAs are immunopurified with Cmr2 antibodies.* RNAs extracted from *P. furiosus* S100 extract (T) and PfCmr2 antibody immune (I) and pre-immune (PI) immunoprecipitation (IP) samples were analyzed by Northern blotting with a probe against *P. furiosus* crRNA 7.01. The 45- and 39-nt species of 7.01 that are co-IPed with the Cmr2 antibodies are indicated by arrows. Sizes of RNA markers (M, Decade markers, Ambion) are shown. C) *Immunoprecipitated Cmr complexes cleave a target RNA.* 5' end-labeled 7.01 target RNA (complementary to crRNA 7.01, indicated with arrow) was incubated in the absence (–) of proteins or in the presence of immune (I) and pre-immune (PI) immunoprecipitation samples. Cleavage products are marked with asterisks. D) *Co-immunopurified crRNAs possess 5' OH and 3' phosphate ends.* Immunoprecipitated crRNAs were isolated and 5' or 3' labeled with (+) and without (–) prior phosphatase treatment. The 45- and 39-nt crRNA species are indicated by arrows and RNA size markers (M) are shown.

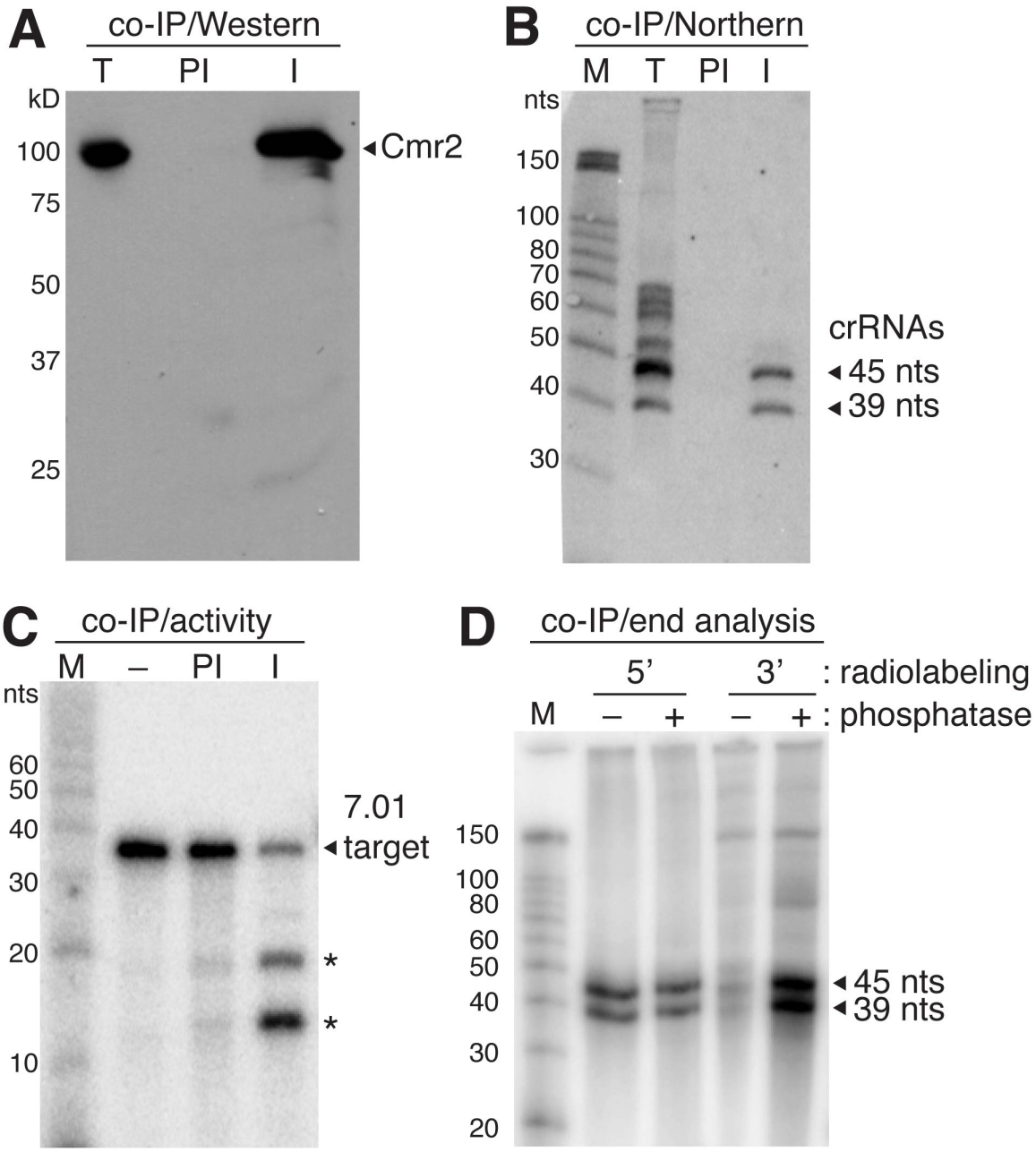
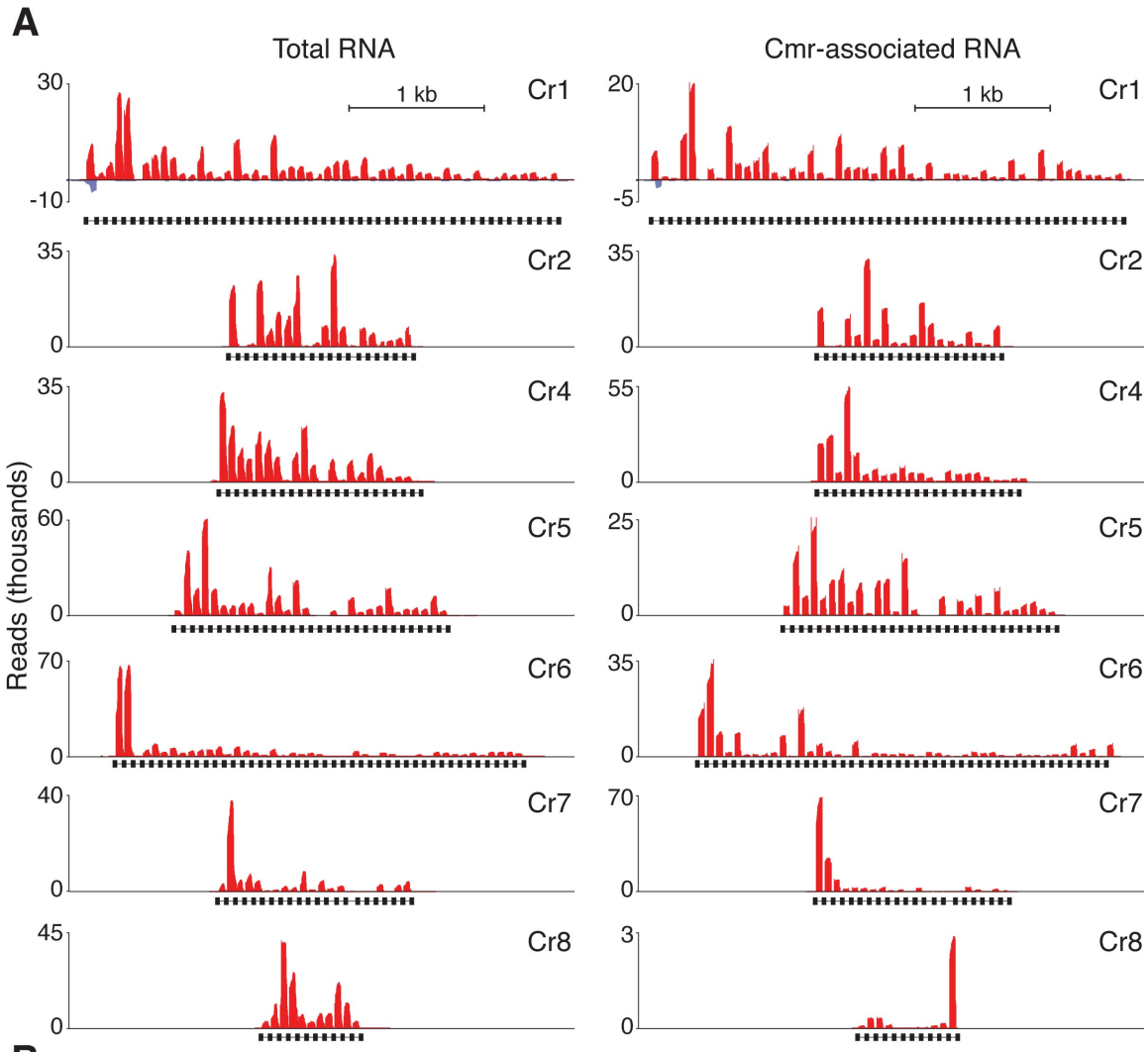

A
co-IP/Western
T
PI
I
kD
100
75
50
37
25
Cmr2
B
co-IP/Northern
M
T
PI
I
nts
150
100
80
70
60
50
40
30
crRNAs
45 nts
39 nts
C
co-IP/activity
M
–
PI
I
nts
60
50
40
30
20
10
7.01
target
*
*
D
co-IP/end analysis
5’
3’
: radiolabeling
M
–
+
–
+
: phosphatase
150
100
80
60
50
40
30
20
45 nts
39 nts

Figure A.2. **Deep sequencing profiles of crRNAs found in total *P. furiosus* RNA and immunopurified Cmr complexes.** A) *Total and Cmr-associated small RNA sequencing reads that map to the seven P. furiosus CRISPR loci.* Y-axis scale bars indicate the number of reads (thousands) that contain nucleotides that map to the indicated position within the CRISPR locus. Red indicates sequence reads that map to the sense strand relative to the CRISPR leader region, and blue corresponds to reads that map to the antisense strand. The positions of CRISPR repeats are shown as black bars below the graph. 1 kb x-axis scale is indicated. Images were generated using the UCSC archaeal genome browser (Karolchik et al. 2003). B) *Putative promoter sequences found in the P. furiosus CRISPR leader regions.* Alignment of the regions upstream of the first repeat of each *P. furiosus* CRISPR locus are shown. The repeats are shown in blue, the initial transcribed regions (detected by sequencing) are shown in red, the potential TATA elements are shown in pink, the potential BRE elements are underlined, and other leader sequences are shown in black. Consensus BRE/TATA elements are shown above the alignment.



B

```
                    BRE    TATA                          RNA reads              Repeat 1
BRE/TATA:           cRnaAntTTTAWAtr<----21/22 bases---->R
Cr1  GTTGGATGCCCCGTAAAGGTTATAAATTCGAGTGATATAGTTACTCCGTAGGAGTATTGGGGCGAAAAAGCCCCCTGTTC
Cr2  GGTTGATGTCCCGTAAGGCTTATAAATTCGAGTGATATAGTTACTCTGTAGGAGTATTGGGGCAAAAAAGCCCCCTGTTC
Cr4  GGTGGATGTCCCGTAAGGCTTATAAATTCGAGTGATGTAATTACCCCAATGGGGAATTGGGGCAAAAAAGCCCCCTGTTC
Cr5  GTTGGATGCCCCGTAAGGGTTATAAATTCAAGTGATATAATTACTCCATAGGAGTATTGGGGCGAAAAAGCCCCCTGTTA
Cr6  GTTGGATGCCCCGTAAGGGTTATAAATTCAAGTGATATAATTACTCCATAGGAGTATTGGGGCGAAAAAGCCCCCTGTTA
Cr7  GGTTGATGCCCCGTAAGGCTTATAAATTCGAGTGATGTAGTTACTCCGTAGGAGGATTGGGGCAAAAAAGCCCCCTGTTC
Cr8  GGTGGATGTCCCGTAAGGCTTATAAATTCGAGTGATATAATTACCCCAATGGGGAATTAGGGCAAAAAAGCCCACTGTTC
```

Figure A.3. **Cmr complex crRNAs contain a conserved 8-nucleotide 5' sequence tag and are of defined 39- and 45-nucleotide lengths.** A) *The 5' and 3' ends of Cmr-associated crRNAs are defined relative to the upstream repeat element.* Graphs show the % of crRNAs sequenced in total *P. furiosus* RNA (upper panel) and immunopurified Cmr complexes (lower) with ends that map at indicated positions relative to the 5' repeat-guide junction (arrowhead). The Cmr-associated crRNA species are illustrated below with 5' repeat tag (black) and guide sequence (green) indicated. RNAs from CRISPR8 were excluded from this analysis due to a repeat sequence polymorphism (see text). Read numbers analyzed (n) are indicated. B) *The length of the Cmr complex crRNAs is independent of the length of the genome-encoded guide sequence.* Deep sequence profiles of crRNAs 4.01 – 4.05 in total *P. furiosus* RNA (upper panel) and immunopurified Cmr complexes (lower) are shown relative to repeat (black) and guide (green) sequence elements (see also Figure 2A). Dashed lines mark the 3' ends of the 45-nucleotide crRNA species. (The 5' end and 3' end of the 39-nucleotide species are also indicated for crRNA 4.01.) The sequence at the 3' guide-repeat junction is shown below for crRNAs 4.01, 4.02 and 4.05. Guide sequences are shown in green, and repeat sequences are shown in black. Arrowheads indicate the locations of the 3' ends of the 45-nucleotide crRNA species.

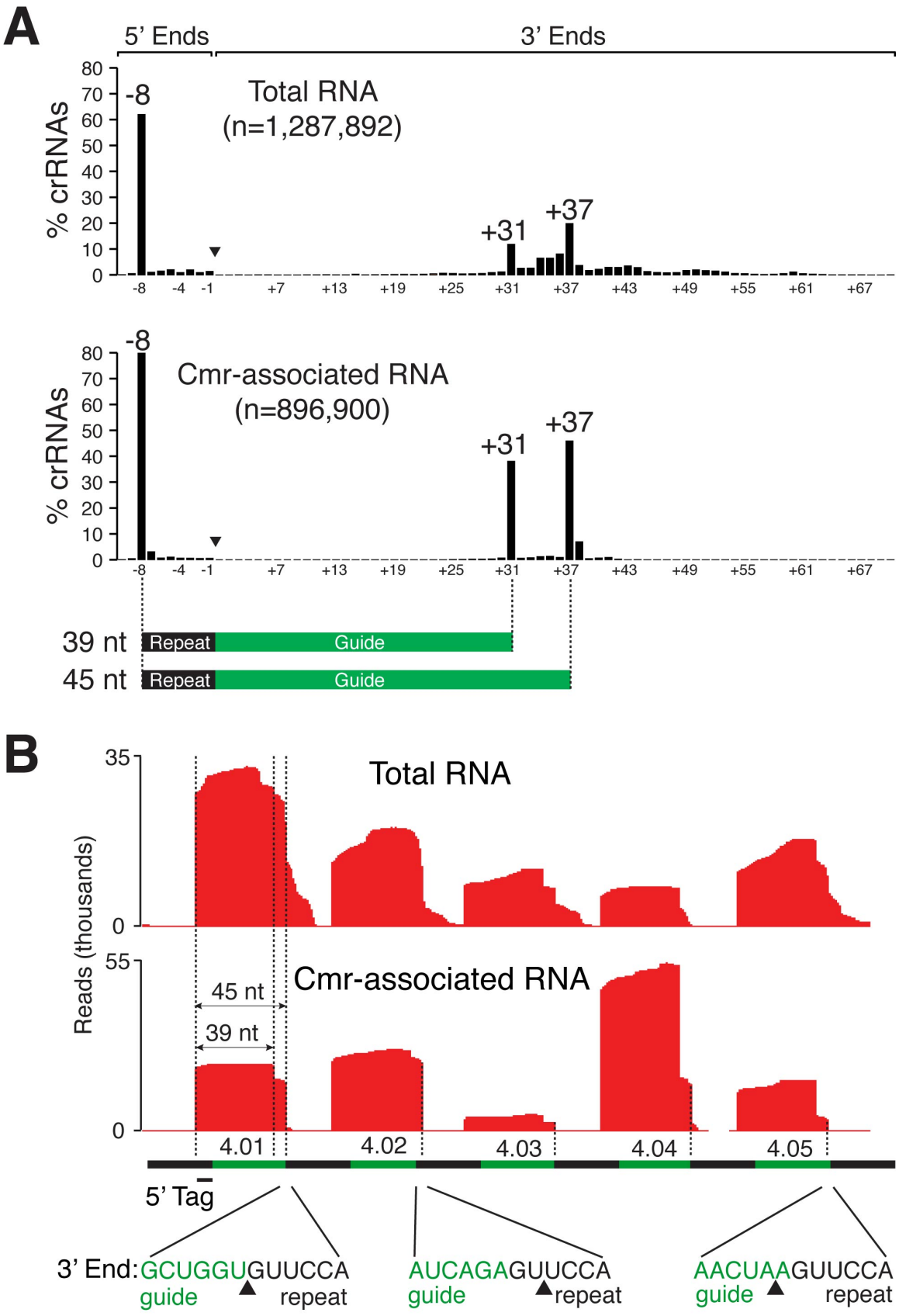

A
5' Ends
3' Ends
Total RNA
(n=1,287,892)
% crRNAs
-8
+31
+37
Cmr-associated RNA
(n=896,900)
39 nt
45 nt
Repeat
Guide
B
Total RNA
Reads (thousands)
Cmr-associated RNA
45 nt
39 nt
4.01
4.02
4.03
4.04
4.05
5' Tag
3' End: GCUGGUGUUCCA
guide
repeat
AUCAGAGUUCCA
AACUAAGUUCCA

Figure A.4. **Antisense RNA found in *P. furiosus* is cleaved by the Cmr-crRNA complex.** A) *Antisense RNA from the region of crRNA 1.01 is detected by Northern analysis, and two species are associated with the Cmr complex.* Northern analysis of total *P. furiosus* RNA (T) and RNAs isolated from immune (I) or pre-immune (PI) Cmr2 immunopurifications was probed for antisense (left panel) or sense 1.01 crRNAs (right panel). Colored dots indicate the primary species of antisense RNAs (blue dots) and crRNAs (red dots) specifically associated with the Cmr complex (and correspond with dots in panel B). Sizes of radiolabeled RNA marker (M) are noted. B) *The two antisense RNAs associated with the Cmr complex map downstream of a potential fortuitous promoter and correspond to the expected products of crRNA 1.01-guided cleavage.* Deep sequencing profiles of Cmr-associated crRNAs in the region encoding crRNA 1.01 are shown. Red indicates RNAs transcribed from the leader region, and blue corresponds to reads from the opposite strand. The positions of CRISPR repeats (black), guide sequences (green) and BRE/TATA promoters (red and blue) are indicated. A blue line represents the apparent full-length antisense RNA observed by deep sequencing total RNA (see Figure S1). Cleavage of the antisense RNA by the Cmr complex 14 nucleotides from the 3' ends of the 45- and 39-nt species of crRNA 1.01 (red dots) would produce antisense RNA products of the sizes observed by deep sequencing and Northern analysis (blue dots). C) *Immunopurified Cmr complexes cleave the antisense RNA*. Radiolabeled RNA targets for the antisense RNAs (left panel) and crRNAs (right panel) were incubated in the absence (–) of proteins or in the presence of immune (I) or pre-immune (PI) immunourified preparations, and the products were analyzed following denaturing gel electrophoresis. (The target for the crRNAs is 140 nucleotides in length

and corresponds to the full-length antisense RNA (see panel C).) The cleavage products obtained by incubation of the crRNA target with the immunopurified Cmr complex (asterisks) are approximately the same sizes as the endogenous antisense RNA species that are immunopurified with the Cmr complexes (A, left panel). Non-contiguous lanes from the same gel are indicated by dashed lines.

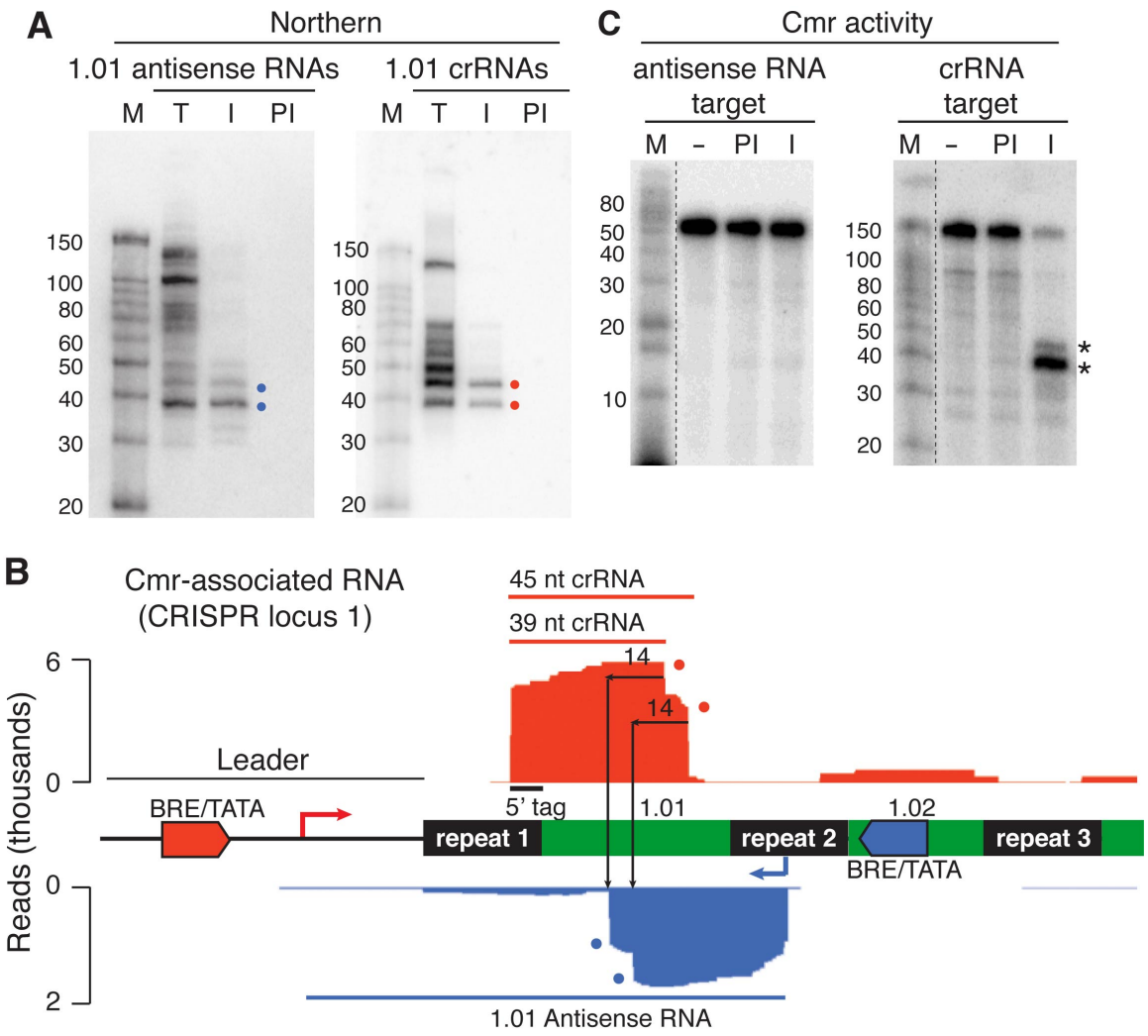

A
Northern
1.01 antisense RNAs
1.01 crRNAs
M T I PI
M T I PI
150
100
80
60
50
40
30
20
C
Cmr activity
antisense RNA target
crRNA target
M – PI I
M – PI I
80
50
40
30
20
10
150
100
80
60
50
40
30
20
*
*
B
Cmr-associated RNA (CRISPR locus 1)
45 nt crRNA
39 nt crRNA
14
14
6
0
Leader
BRE/TATA
5' tag
1.01
1.02
repeat 1
repeat 2
repeat 3
BRE/TATA
Reads (thousands)
0
2
1.01 Antisense RNA

Figure A.5. **The 5' repeat tag is required for formation of functional Cmr complexes.** A) *The 5' repeat tag is required for Cmr complex function.* Reconstitutions were performed with recombinant Cmr proteins (1-1, 2, 3, 4, 5, and 6) and the 45-nucleotide wildtype 7.01 crRNA (AUUGAAAG tag) or crRNA 7.01 lacking the tag (–), and activity was tested against the radiolabeled 7.01 target RNA. RNA marker sizes are indicated. The cleavage product observed with the wildtype crRNA is indicated with an asterisk. B) *Mutations to the tag sequence prevent function.* Reconstitutions were performed with the 45-nucleotide wildtype 7.01 crRNA (AUUGAAAG tag) or crRNA 7.01 with the indicated mutations in the tag sequence (mutated or inserted nucleotides are highlighted in red) and tested in cleavage assays as described in A. The cleavage product observed with the wildtype crRNA is indicated with an asterisk. C) *crRNAs with variant tag sequences are underrepresented in the Cmr complex.* Deep sequencing profiles of Cmr-associated crRNAs in the region encoding crRNAs 8.08 – 8.11 are shown (see also Figure 2A). The positions of CRISPR repeats (black) and guide sequences (green) are indicated. The 5' tag sequence is shown below for crRNAs 8.08 - 8.11. An additional nucleotide present in the wildtype tag of crRNA 8.11 is highlighted in red D) *Complementarity to the tag does not prevent function.* Cmr complexes reconstituted with wildtype crRNA 7.01 were tested for the ability to cleave the 7.01 target RNA or a target with complementarity to the tag sequence (as illustrated). Cleavage products are indicated with asterisks. Non-contiguous lanes from the same gel are indicated with a dashed line..

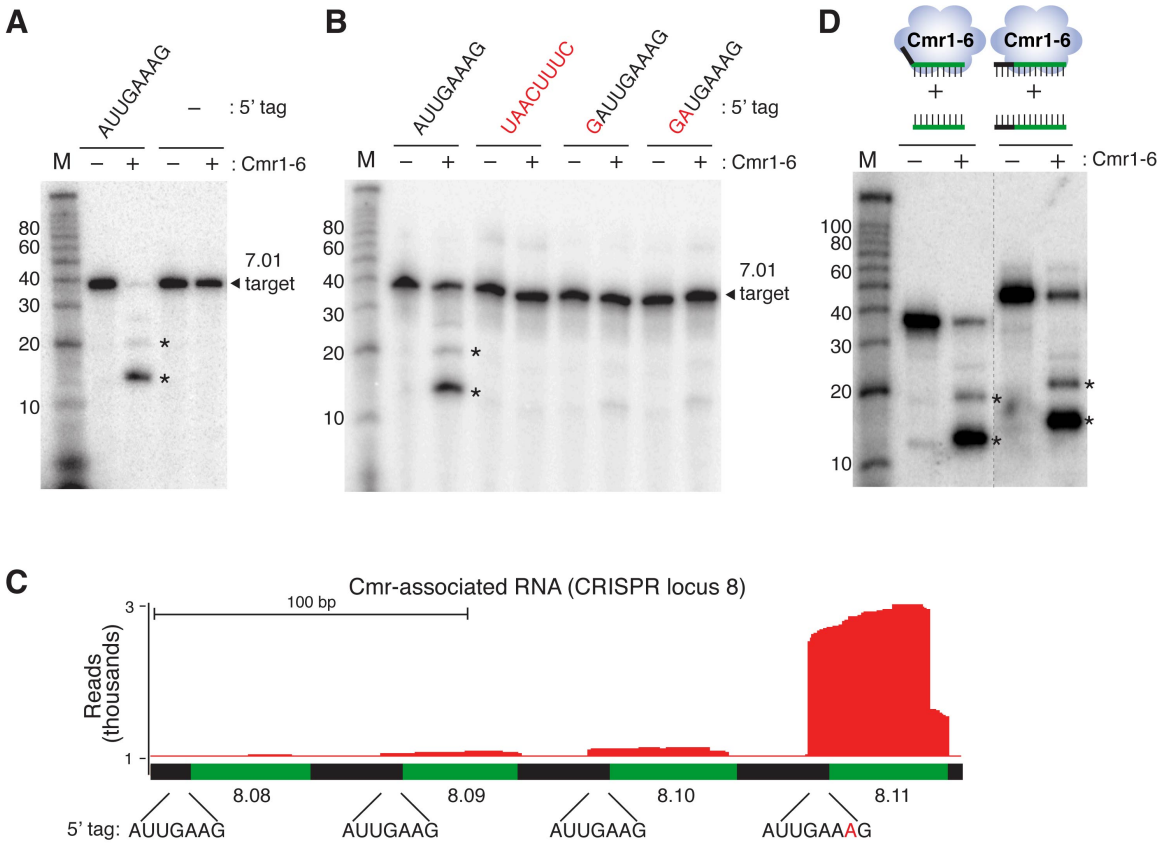

A
AUUGAAAG
– : 5' tag
M – + – + : Cmr1-6
80
60
40
30
20
10
7.01
target
*
*
B
AUUGAAAG
UAACUUUC
GAUUGAAAG
GAUGAAAG
: 5' tag
M – + – + – + – + : Cmr1-6
80
60
40
30
20
10
7.01
target
*
*
D
Cmr1-6
Cmr1-6
+
+
M – + – + : Cmr1-6
100
80
60
40
30
20
10
*
*
*
*
C
Cmr-associated RNA (CRISPR locus 8)
100 bp
3
1
Reads (thousands)
8.08
8.09
8.10
8.11
5' tag: AUUGAAG
AUUGAAG
AUUGAAG
AUUGAAAG

Figure A.6. **Engineered crRNAs direct the Cmr complex to cleave intended target RNAs.** A) *Specific cleavage of novel target RNAs.* Cmr complexes were reconstituted with crRNAs designed to target a random sequence (X crRNA, X' target, orange) or a fragment from the 5' end of the β-lactamase (bla) mRNA (5' *bla* crRNA, *bla* target, red), and tested for the ability to cleave both target RNAs. Cleavage products are indicated with asterisks. Diagrams illustrate engineered crRNAs with 5' repeat tag (black) and guide sequence elements (orange or red) interacting with and cleaving the intended 5' radiolabeled target. B) *Directed cleavage of predicted sites in b-lactamase mRNA.* Engineered crRNAs target two locations on the full-length *bla* mRNA. The full length *bla* mRNA was subject to cleavage by Cmr complexes reconstituted with crRNAs targeting a region near the 5' end of the mRNA (5' *bla* crRNA, red) or a sequence approximately 200 nucleotides from the 5' end of the mRNA (internal *bla* crRNA, blue). Cleavage products are indicated by asterisks. Diagrams illustrate engineered crRNAs interacting with and cleaving the *bla* mRNA.

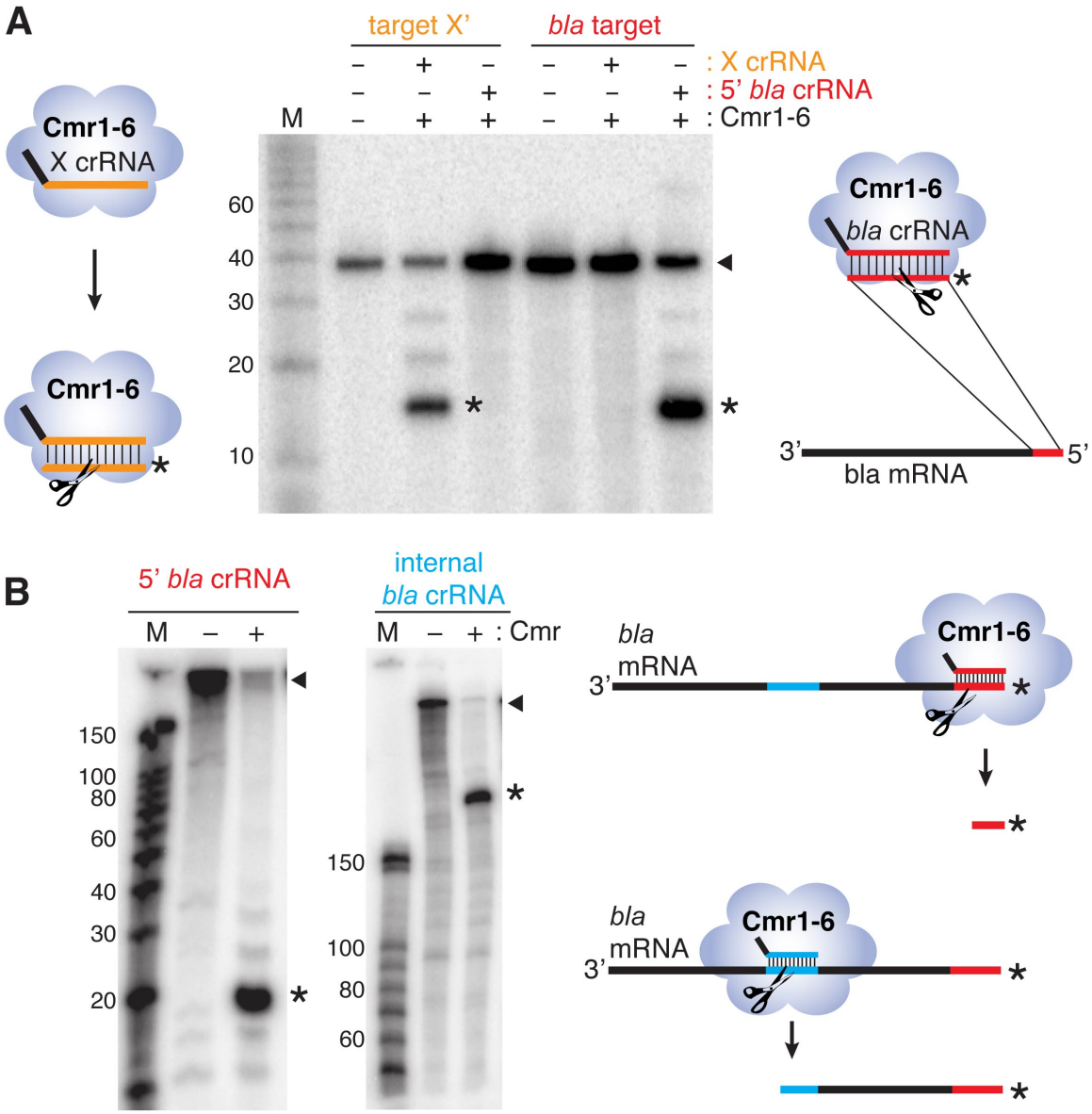

A
target X'
bla target
– + – – + – : X crRNA
– – + – – + : 5' bla crRNA
M – + + – + + : Cmr1-6
60
40
30
20
10
Cmr1-6
X crRNA
Cmr1-6
Cmr1-6
bla crRNA
3'
5'
bla mRNA
B
5' bla crRNA
internal
bla crRNA
M – + M – + : Cmr
150
100
80
60
40
30
20
150
100
80
60
bla
mRNA
3'
Cmr1-6
bla
mRNA
3'
Cmr1-6

Figure A.S1, related to Figure 4. **Antisense RNA complementary to crRNA 1.01 in *P. furiosus*.** Deep sequencing profiles from total *P. furiosus* RNA in the region encoding crRNA 1.01 are shown. Red indicates RNAs transcribed from the leader region, and blue corresponds to reads from the opposite strand. The positions of CRISPR repeats (black), guide sequences (green) and BRE/TATA promoters (red and blue) are indicated. A blue line represents the apparent full-length 1.01 antisense RNA.

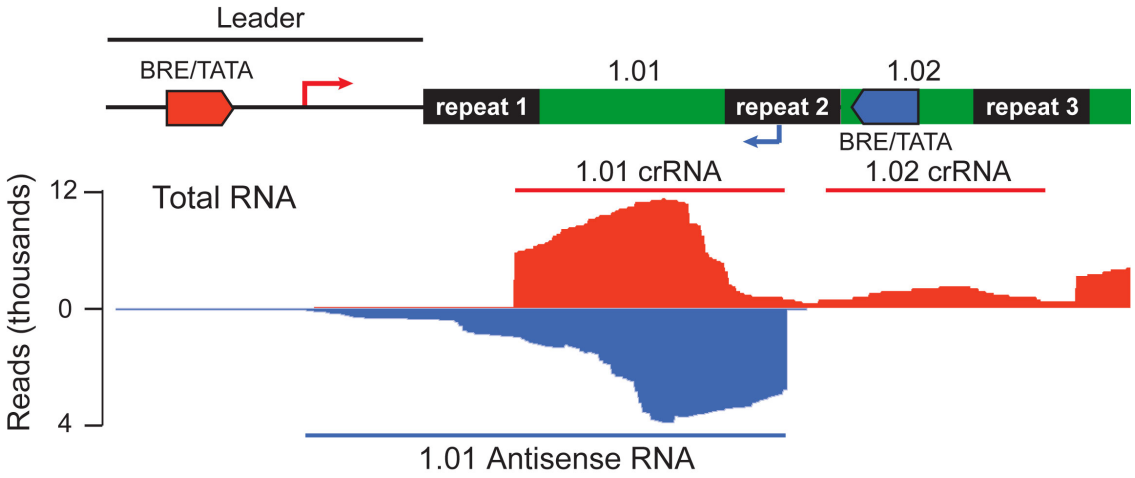
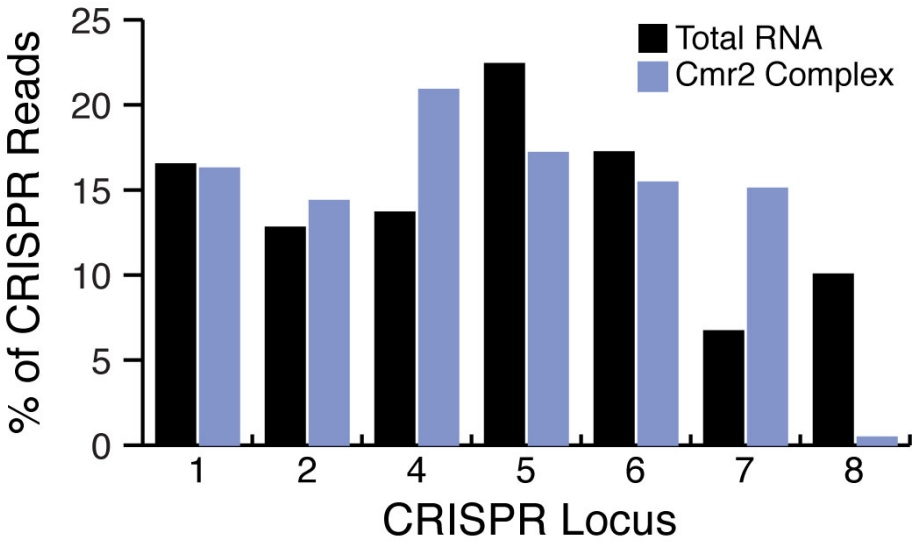

Leader
BRE/TATA
1.01
1.02
repeat 1
repeat 2
repeat 3
BRE/TATA
1.01 crRNA
1.02 crRNA
Total RNA
12
0
4
Reads (thousands)
1.01 Antisense RNA

Figure A.S2, related to Figure 5. **crRNAs from CRISPR 8 are underrepresented in the Cmr complex.** Graph indicates the percentage of CRISPR-derived deep sequencing reads that map to each CRISPR locus in total RNA (black) and Cmr-associated RNAs (blue)


% of CRISPR Reads
25
20
15
10
5
0
Total RNA
Cmr2 Complex
1
2
4
5
6
7
8
CRISPR Locus