VILNIUS UNIVERSITY

Tomas Šinkūnas

In vitro reconstitution of DNA interference in a type I CRISPR-Cas system

Doctoral dissertation

Physical science, biochemistry (04 P)

Vilnius, 2015

The work presented in this doctoral dissertation has been carried out at the Institute of Biotechnology, Vilnius University during 2009-2013.

Supervisor

Prof. dr. **Virginijus Šikšnys** (Vilnius University, physical sciences, biochemistry - 04 P)

VILNIAUS UNIVERSITETAS

Tomas Šinkūnas

I tipo CRISPR-Cas sistemos DNR interferencijos atkūrimas *in vitro*

Daktaro disertacija

Fiziniai mokslai, biochemija (04 P)

Vilnius, 2015

Disertacija rengta 2009-2013 m. Vilniaus universiteto Biotechnologijos institute.

Mokslinis vadovas

Prof. dr. **Virginijus Šikšnys** (Vilniaus universitetas, fiziniai mokslai, biochemija - 04 P)

CONTENTS

LIST OF ABBREVIATIONS

Abi	abortive infection
Ap	ampicillin
bp	base pair
BSA	bovine serum albumin
Cas	CRISPR associated
Cascade	CRISPR-associated complex for antiviral defense
Cm	chloramphenicol
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR ribonucleic acid
CTD	C-terminal domain
ds	double-stranded
DTT	1,4-dithiothreitol
EDTA	ethylenedinitrilotetraacetic acid
EM	electron microscopy
EMD	electron microscopy data bank
EMSA	electrophoretic mobility shift assay
ESI-MS	electrospray ionization mass spectrometry
HPLC	high-performance liquid chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Luria-Bertani
MGE	mobile genetic element
nt	nucleotide
NTD	N-terminal domain
PAA	polyacrylamide
PAGE	polyacrylamide gel electrophoresis
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PDB	protein data bank
PNK	T4 polynucleotide kinase
pre-crRNA	precursor crRNA
RAMP	repeat associated mysterious protein
R-loop	DNA-RNA hybrid which is formed by RNA hybridization to a complementary strand of DNA duplex leaving the non-complementary DNA strand unpaired

RM	restriction - modification
RNP	ribonucleoprotein
SDS	sodium dodecyl sulfate
Sie	superinfection exclusion
ss	single-stranded
Str	streptomycin
TA	toxin-antitoxin
TEMED	N,N,N′,N′-tetramethylethylenediamine
tracrRNA	*trans*-activating crRNA
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
WT	wild type

INTRODUCTION

Prokaryotic cell is a subject for virus (phage) infection therefore prokaryotes have developed a multilayered defense barrier interfering with every step of the phage attack (Labrie et al., 2010; Samson et al., 2013). One of the layers is an adaptable defense system of prokaryotes which is comprised of clustered regularly interspaced short palindromic repeat (CRISPR) locus and CRISPR associated (*cas*) genes (Barrangou et al., 2007). CRISPR is an array of repetitive sequences (repeats) interspaced by unique non-repetitive sequences of extracellular origin called spacers (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). A cluster of *cas* genes encode proteins that execute CRISPR-Cas immunity (Brouns et al., 2008; Gasiunas et al., 2012; Hale et al., 2009).

Mechanism of CRISPR-Cas defense systems can be dissected into tree stages: (i) adaptation, (ii) expression and processing and (iii) interference (van der Oost et al., 2014). Cell intruder is recognized and its small DNA fragment is integrated into the CRISPR locus as a new spacer during the adaptation stage (Barrangou et al., 2007). In the expression and processing stage, small CRISPR RNA (crRNA) molecules are generated from a long transcript of CRISPR array that is cut within the repeat sequences (Carte et al., 2008; Deltcheva et al., 2011). In the interference stage, mature crRNA together with Cas proteins assemble into an effector complex that targets and destroys foreign nucleic acid containing a crRNA-complementary sequence termed protospacer (Brouns et al., 2008; Gasiunas et al., 2012; Hale et al., 2009; Jinek et al., 2012; Tamulaitis et al., 2014).

Composition of *cas* genes as well as action mechanisms diverge between different CRISPR-Cas systems therefore these systems are classified into three major types, which are further subdivided into subtypes (Makarova et al., 2011b). Type I and type II systems interfere with double-stranded (ds) DNA,

while single-stranded (ss) RNA is targeted by the type III system. Multiple Cas protein subunits assemble on the crRNA molecule, forming a ribonucleoprotein (RNP) complex termed Cascade, which is responsible for DNA target recognition in type I systems (Jore et al., 2011). However, an accessory Cas3 protein is required for DNA interference in the type I (Brouns et al., 2008). In contrast to type I, the effector complex of type II systems is formed by a single Cas9 protein and two RNA molecules: crRNA and tracrRNA (Jinek et al., 2012; Karvelis et al., 2013a). Similarly to type I, multisubunit RNP effector complexes, termed Csm and Cmr, are formed in type III-A and III-B systems, respectively (Rouillon et al., 2013; Staals et al., 2013). In contrast to type I, effector complexes of type II and type III have intrinsic nucleases therefore they do not use accessory proteins to destroy target nucleic acids (Brouns et al., 2008; Sapranauskas et al., 2011; Tamulaitis et al., 2014).

Besides complementary protospacer, DNA targets have a protospacer adjacent motif (PAM), which is the main determinant discriminating foreign from self DNA (Deveau et al., 2008). PAM sequences are diverse and in most cases specific for CRISPR-Cas system (Mojica et al., 2009). DNA target recognition leads to an R-loop formation where target DNA strand of the protospacer is base-paired with the crRNA while non-target strand is displaced as an ssDNA (Jore et al., 2011). The R-loop formation triggers target DNA destruction.

Cascade complex together with the type I signature protein, Cas3, confers resistance against the target DNA (Brouns et al., 2008); however, molecular mechanism of DNA interference stage remained to be elucidated. Therefore, the main objective of this study was to establish the mechanistic details of DNA interference stage in type I CRISPR-Cas systems. Type I-E CRISPR4-Cas from *Streptococcus thermophilus* DGCC7710 strain was used as a model system (Horvath & Barrangou, 2010). It is comprised of eight *cas* genes and a CRISPR region bearing twelve spacers. The *cse1-cse2-cas7-cas5-cas6e* gene cassette encodes proteins that form *S. thermophilus* Cascade (St-Cascade)

complex while *cas3* gene encodes a large multidomain protein (St-Cas3), which contains both helicase/ATPase and nuclease motifs.

The specific aims of this study were:

(i). To identify the biochemical activities of the stand-alone St-Cas3 protein;

(ii). To determine the composition of the St-Cascade complex;

(iii). To analyse the PAM preference of the St-Cascade;

(iv). To analyse the mechanism of R-loops formation by single molecules of the Cascade;

(v). To reconstitute DNA interference mechanism of the *S. thermophilus* type I-E CRISPR4-Cas system *in vitro*.

Scientific novelty.

In this work, we for the first time provide the biochemical characterization of the Cas3 protein that is a hallmark protein for type I systems. We show here that Cas3 combines both an ssDNA nuclease and ATPase/helicase catalytic activities. We have isolated and established the molecular composition of the *S. thermophilus* Cascade complex, and demonstrated that it binds DNA targets containing promiscuous PAM sequence. Furthermore, we provide the first experimental evidences for the R-loop formation by a single Cascade molecule that revealed the locking step of the stable R-loops. Last but not least, we show Cascade-triggered directional degradation of the DNA target by the Cas3 protein. Summarising our results, we propose detailed mechanism of DNA interference stage for the type I CRISPR-Cas systems.

Practical value.

The Cas9-based CRISPR-Cas system recently came to prominence as a versatile molecular tool for genome editing. Due to the complex multisubunit structure, Cascade complexes of type I CRISPR-Cas systems are less suitable for genome editing applications. Nonetheless, they have been already employed for elimination of antibiotic resistant bacterial strains (Gomaa et al.,

2014), engineering of a bacteriophage genome (Kiro et al., 2014) and regulation of gene repression (Luo et al., 2014). Cascade fusions with other proteins such as gene expression regulators or fluorescent proteins may provide programmable tools for analysis of gene expression or DNA labelling, respectively. Detailed understanding of the molecular mechanisms of the type I systems may open new application possibilities.

The major findings presented for defense in this thesis:

- St-Cas3 posesses both nuclease and ATPase/helicase activities localised on separate domains.
- St-Cascade is composed of multiple Cas protein subunits and the mature crRNA molecule.
- St-Cascade binds to DNA targets with a promiscuous PAM sequence generating the R-loop.
- Complementarity at the PAM-distal end of the protospacer determines stability of the R-loops.
- Cascade-mediated R-loop formation triggers directional target DNA degradation by the Cas3.

1. LITERATURE OVERVIEW

1.1. Defense barriers in prokaryotes

Prokaryotes are the most spread cellular life form on Earth. They populate diverse ecological niches and have to adapt to various abiotic stresses. Furthermore, they are constantly attacked by viruses (bacteriophages or archaeal phages) which outnumber prokaryotic cells by 10-fold (Bergh et al., 1989). In order to propagate, phages adsorb to the surface of the cell and inject nucleic acids into the cytoplasm where cellular resources are hijacked for multiplication of phage particles (Sturino & Klaenhammer, 2006). The prokaryotes have developed multiple defense barriers to avoid the phage predation (Dy et al., 2014b; Labrie et al., 2010), however, the phages develop new strategies to overcome resistance of the host (Dy et al., 2014b; Samson et al., 2013) (Figure 1). Therefore, prokaryotes and phages are in continuous arms race which drives co-evolution of these species.

To penetrate a cell, the phage has to attach to a specific receptor (protein, polysaccharide or lipopolysaccharide), which is localized at the surface of the host cell (Labrie et al., 2010). Prokaryotes use several strategies to inhibit the phage adsorption. First, phage adhesion receptors can be masked by a host protein. For example, the outer-membrane protein TraT masks an OmpA (outer-membrane protein A) protein that serves as the receptor for T-even-like *Escherichia coli* phages (Achtman et al., 1977; Riede & Eschbach, 1986). Moreover, mutations that alter OmpA surface can restrict the phage binding (Koebnik, 1999). Second, bacteria synthesize extracellular polymers, such as alginate and hyaluronan, which protect bacteria from abiotic stress and act as physical barriers against phage adsorption (Hanlon et al., 2001; Hynes et al., 1995). Third, prokaryotes produce molecules that bind to the receptor, thus blocking the phage attachment site. Iron transporter FuhA is the receptor of *E. coli* phage T5. A small peptide, MccJ25 (Microcin J25), binds to FuhA and

competitively inhibits adsorption of T5 phage (Destoumieux-Garzon et al., 2005). Last, cell surface protein (including phage receptors) content can be altered during a stochastic process termed phase variation (Hoskisson & Smith, 2007). *Bordetella* species have a two-component regulatory system termed BvgAS, which turn-on or turn-off the expression of various genes. One of these genes encode the cell surface protein pertactin (Prn) that is expressed during the Bvg^+ phase and serves as the receptor for the phage BPP-1 adhesion. However, Prn is absent in the Bvg^- phase, thus susceptibility of phage infection is reduced by 10^6 folds (Liu et al., 2002; Stibitz et al., 1989).

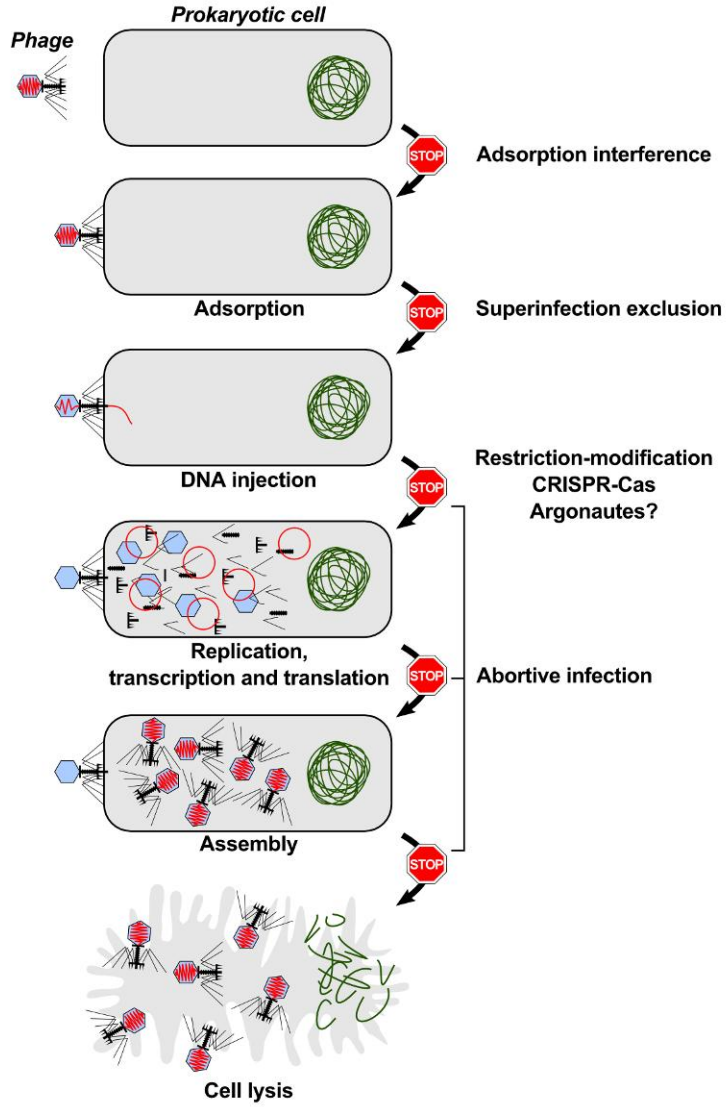

Figure 1. Anti-phage barriers of prokaryotes. Prokaryotes possess diverse defense systems that interfere with every step of phage lytic life cycle. Modified from (Labrie et al., 2010; Sturino & Klaenhammer, 2006).

Despite diverse adsorption-inhibiting defense barriers, phages can overcome them by modifying their receptor binding proteins (RBPs). Mutations in the RBP can tune adsorption to the modified receptor; furthermore, some mutations in the RBP lead to targeting of a new receptor. Stochastic expression of RBPs helps counteract the phase variation. Moreover, phage-encoded hydrolases can "dig in" the extracellular polymers to reach the receptors (Samson et al., 2013).

After attachment, the phage should inject its nucleic acid into the host. The entry of the phage DNA could be blocked by superinfection exclusion (Sie) systems (Dy et al., 2014b). Sie components are often encoded in prophages, thus Sie systems are important for phage-phage interactions rather than phage-host interactions (Labrie et al., 2010). The best characterised Sie system is comprised of Sp and Imm proteins encoded in the *E. coli* phage T4. The Sp prevents degradation of peptidoglycan by inhibiting phage-encoded lysozyme, whereas Imm changes conformation of the injection site. These proteins block DNA injection of concomitant T4 or other T-even phage infections (Lu et al., 1993).

If the phage DNA penetrates the cell, it encounters host defense systems that recognize and destroy foreign DNA. These barriers are restriction-modification (RM), CRISPR-Cas (clustered regularly interspaced short palindromic repeats – CRISPR associated) systems, and probably prokaryotic Argonautes (Dy et al., 2014b).

RM systems are comprised of restriction and modification modules. The restriction module cuts unmodified DNA in a sequence-specific manner, whereas the modification module methylates both DNA strands at the same recognition sequence thus protecting the DNA from nuclease activity of the restriction module. RM systems are classified into four types (type I-IV) by subunit composition, DNA sequence recognition, cleavage position, cofactor requirements, and substrate specificity (Loenen et al., 2014b; Pingoud et al., 2014). When phages inject unmethylated DNA, it encounters type I-III RM

systems. The DNA is either destroyed by nuclease or protected by methyltransferase (MTase) which methylates recognition sequence. The fate of the phage DNA depends on the rate of these reactions. In general, nuclease reaction is faster than methylation of unmodified DNA therefore foreign DNA is prone to be destroyed. Moreover, number of recognition sequences correlates with antiviral efficiency of RM system (Loenen et al., 2014a).

Phages have evolved several strategies to prevail RM systems (Samson et al., 2013). One of them is elimination of recognition sites by point mutations (Meisel et al., 1992). Some phages encode MTase that modifies their own DNA (Drozdz et al., 2012), while other phages have proteins that increase reaction rate of the host MTase and promote its own DNA metylation, thus protecting phage DNA from cleavage by nuclease of RM system (Loenen & Murray, 1986). In addition, phages can incorporate modified nucleotides, such as hydroxymethylcytosine, in their genomes to prevent recognition by the RM systems. However, prokaryotes have evolved type IV RM systems that cleave modified instead of unmodified DNA (Loenen & Raleigh, 2014). On the other hand, phages may encode proteins that inhibit RM systems, for example by mimicking a double-stranded B-DNA structure (Walkinshaw et al., 2002). Finally, some phages inject DNA together with proteins that temporarily protect RM target sites from nuclease cleavage (Streiff et al., 1987).

Recently, bacterial Argonautes were shown to interfere with invading foreign DNA. These Argonautes are homologous to eukaryotic counterpart that in association with small RNA molecules target destruction of complementary RNA in a process known as RNA interference (Swarts et al., 2014b). Prokaryotic Argonautes assemble with small RNA or DNA and cut complementary RNA or DNA, respectively (Sheng et al., 2014; Swarts et al., 2014a). However, direct evidences for anti-phage resistance of Argonoutes in prokaryotes are still missing.

When all other defense systems fail, infected cell can execute “molecular harakiri”. This altruistic act of one cell suicide may rescue population of the

same kin. Abortive infection (Abi) and toxin-antitoxin (TA) systems are responsible for suicide of the infected cell (Dy et al., 2014b; Labrie et al., 2010). Although mechanisms of Abi systems differ, the common feature of these systems is a dormant protein which is activated by phage infection and then shuts down the essential metabolic pathways of the cell. Diversity of Abi mechanisms can be exemplified by four *E. coli* Abi systems. First, the RexAB system aborts T4 phage infection by depolarization of the cell membrane. RexA is an intracellular sensor, which is activated by intermediate of phage recombination and replication. Binding of the activated RexA to the ion channel RexB leads to ion flux across the membrane and consequent inhibition of ATP synthesis, causing the cell death (Parma et al., 1992). Second, T7 phage proteins activate a transmembrane protein, PifA, that causes leakage of ATP molecules (Schmitt et al., 1991). Third, a capsid protein of T4 phage activates Lit protease, which cleaves the elongation factor-Tu of ribosom by stopping protein synthesis (Georgiou et al., 1998). Last, a T4 protein alters the interaction between EcoprrI and PrrC, releasing PrrC protein that cleaves $tRNA^{Lys}$ anti-codon loop, thus inhibiting protein synthesis (Levitz et al., 1990). On the other hand, phages can escape the abortive infection by mutating proteins that activate Abi systems (Samson et al., 2013).

TA systems are composed of a toxic molecule, which is neutralized by an antitoxic counterpart. Five types of TA systems are classified by the mode of toxin repression (Mruk & Kobayashi, 2014). The balance of TA systems is tightly regulated. However, particular stimulus may reduce the amount of antitoxin, thus liberating the toxin. Released toxin interferes with the vital cellular processes that lead to the cell death. In some cases such stimulus can be a phage infection (Dy et al., 2014a). The ToxIN system from *Pectobacterium atrosepticum* is the best characterized case for the TA-mediated abortive infection. The cytotoxic endorybonuclease ToxN is neutralized by the repetitive noncoding ToxI RNA (Fineran et al., 2009). Various phages can activate ToxN, which degrades cellular and phage RNA, causing abortive infection. However, some phages manage to outwit ToxIN

defense system by encoding a pseudo-antitoxin or even hijacking the *toxI* gene (Blower et al., 2012).

1.2. CRISPR-Cas systems

The adaptable immunity system of prokaryotes is composed of clustered regularly interspaced short palindromic repeat (CRISPR) locus and CRISPR associated (*cas*) genes. CRISPR is an array of repetitive sequences interspaced by unique non-repetitive sequences of similar length called spacers (Figure 2).

The repeat sequences within the CRISPR locus are conserved but sequence and length of the repeats in other CRISPR loci may vary greatly. Repeats are usually between 23 and 47 bp in length. CRISPRs are clustered into 12 families based on sequence similarity of the repeats (Kunin et al., 2007). Some repeats have palindromic sequence elements and therefore generate RNAs with stable hairpin structures, whereas others are non-palindromic (Kunin et al., 2007).

Bioinformatic analyzes of spacer sequences have traced them to mobile genetic elements (MGE), such as bacteriophages, plasmids, or transposons (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). Sequence stretch in the MGE, identical to the spacer sequence, is termed a protospacer.

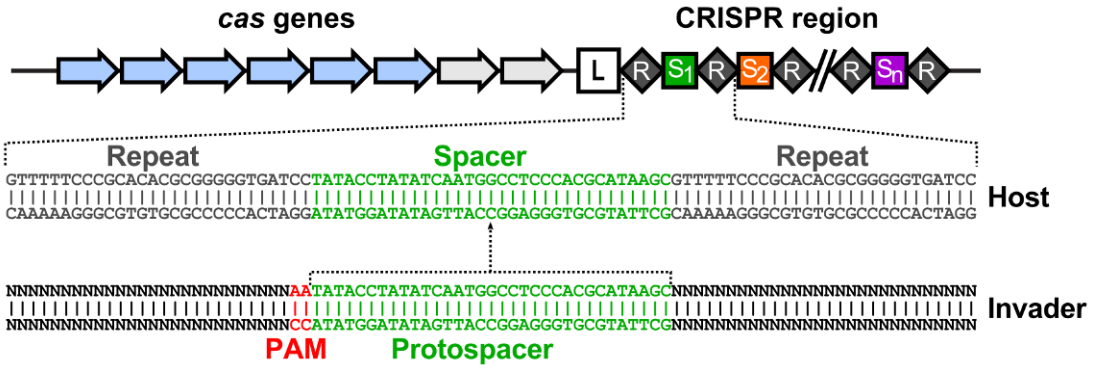

Figure 2. Schematic representation of CRISPR-Cas system. Cluster of *cas* genes (arrows) is usually localised in the vicinity of CRISPR region which is comprised of repeat sequences (R) interspaced with non-repetitive spacer sequences (S_1-S_n). CRISPR locus is flanked by A/T-rich leader sequence (L). Spacer sequence is originated from invading DNA where it is termed a protospacer. Protospacer adjacent motive (PAM) is a short DNA sequence that is present in the vicinity of protospacer.

Usually, short DNA sequence of 2-5 bp is located near the protospacer and denoted as a protospacer adjacent motive (PAM); it plays important role in discrimination of foreign DNA (Deveau et al., 2008; Mojica et al., 2009; Shah et al., 2013) (Figure 2). It was shown that the phage challenge survivors had inserted new spacers into the CRISPR region. Furthermore, spacer sequences identical for the protospacer sequences conferred resistance to the phage (Barrangou et al., 2007). Hence, the CRISPR loci functions like a recordable memory storage which keeps track about invaders of previous infections.

Spacer insertions are polarized (Barrangou et al., 2007) and occur at the CRISPR end, which is proximal to a 200–500 bp long adenine/thymine (AT)-rich sequence named leader (Jansen et al., 2002) (Figure 2). It often has promoter elements (Hale et al., 2012; Lillestol et al., 2009; Pougach et al., 2010; Pul et al., 2010) and binding sites for regulatory proteins (Pougach et al., 2010; Pul et al., 2010).

CRISPRs are widespread among prokaryotes. According to the CRISPRdb database (Grissa et al., 2007) 84% of archeal and 45% of bacterial genomes possess CRISPRs. Furthermore, often a single prokaryotic chromosome carries multiple CRISPR loci (e.g., 18 CRISPRs in *Methanocaldococcus sp*. FS406-22 chromosome); the number of repeat-spacer units in the CRISPR locus may reach a few hundreds (e.g., *Sulfolobus tokodaii* str. 7 contains 458 units in a single CRISPR region).

Cluster of *cas* genes is most frequently found adjacent to the CRISPR locus (Jansen et al., 2002) (Figure 2). These genes encode diverse proteins that contain helicase, nuclease, polymerase, or RNA-binding domains (Makarova et al., 2002). Cas proteins are the executors of microbial adaptive immunity and participate in each step of CRISPR-Cas action.

1.3. Nomenclature of Cas proteins and CRISPR-Cas systems

More than a decade after the discovery of CRISPR arrays (Ishino et al., 1987) four genes (*cas1-cas4*) associated with CRISPRs have been identified (Jansen et al., 2002). Accumulating data of sequenced genomes and implementation of more sophisticated sequence search engines led to identification of about 45 diverse protein families associated with CRISPR arrays (Haft et al., 2005). Six of them were widely distributed (Cas1-Cas6) among the CRISPR systems and thus classified as core Cas proteins. However, only Cas1 and Cas2 proteins were universally conserved (Jansen et al., 2002). Other genes of Cas proteins were distributed only in fraction of CRISPR-Cas systems. They were clustered into eight subtypes based on the phylogeny of Cas1 protein and the operonic organisation of *cas* genes. These subtype specific *cas* genes were named according to the reference organism of the cluster and numbered by the position in the operon of *cas* genes [e.g., *Escherichia coli* subtype was named cse1-5; other subtypes included *Aeropyrum* (csa), *Desulfovibrio* (csd), *Haloarcula* (csh), *Mycobacterium* (csm), *Neisseria* (csn), *Thermotoga* (cst), and *Yersinia* (csy)].

Later some of the Cas protein families were linked within the same "clusters of orthologous groups" (COG) (Makarova et al., 2006). Therefore, orthologous Cas protein families were combined, renamed, and classified into three types (which could be further subdivided into several subtypes) integrating phylogenies of the most common *cas* genes, the sequence and organization of CRISPR repeats, and the architecture of CRISPR-*cas* loci (Makarova et al., 2011b) (Figure 3). Current nomenclature extends list of the core *cas* genes to ten (*cas1 – cas10*) that are found in more than one subtype. In cases for which significant sequence similarity between Cas proteins is observed, but orthologous relationships cannot be definitively assigned, a letter derived from the subtype label is added (i.e., Cas6e, Cas6f, Cas8a, Cas8b, Cas8c and Cas10d). Previously assigned names are retained for less common

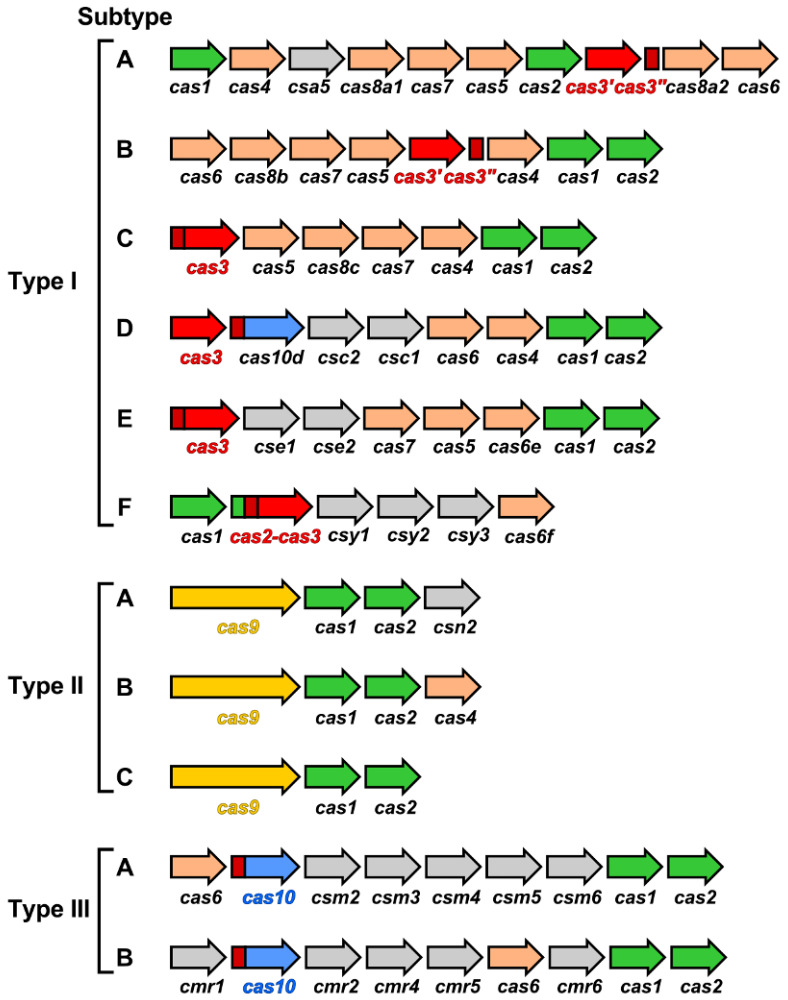

Figure 3. Diversity of CRISPR-Cas systems. Three types of CRISPR-Cas systems are subdivided into eleven subtypes. Each type is characterized by a hallmark gene: *cas3* (red), *cas9* (yellow) and *cas10* (blue) are signature genes of type I, II and III, respectively. The most conserved genes throughout all CRISPR-Cas systems are *cas1* and *cas2* (green). Genes (*cas4*, *cas5*, *cas6*, *cas7* and *cas8*) coloured in orange are less conserved and distributed between several subtypes while subtype specific genes are coloured in grey. Modified form (Makarova et al., 2011b).

genes that are associated with one subtype only (e.g., *cse1* and *cse2* are specific for the type I-E while *csy1*, *csy2* and *csy3* are found only in the type I-F) (Makarova et al., 2011b).

Type I CRISPR-Cas systems are widely distributed among archaea and bacteria. Type I is divided into 6 subtypes (from type I-A to I-F). All Type I systems encode Cas3 protein, which has an HD phosphohydrolase domain and a DExH helicase domain (Jackson et al., 2014b; Makarova et al., 2002; Makarova et al., 2011b). Sometimes (as exemplified in subtypes A, B and D)

those domains are found as separate proteins (HD-protein termed Cas3′ and helicase – Cas3′′) (Figure 3). Another crucial element of type I systems is a multioligomeric RNP complex named CRISPR-associated complex for antiviral defense (Cascade). Cascade complexes are assembled by subtype specific Cas proteins and/or core proteins (Cas7, Cas5) on the crRNA carcass (Jore et al., 2011; Makarova et al., 2011b). Both Cas3 and Cascade are required to confer resistance to DNA invader (Brouns et al., 2008).

Type II CRISPR-Cas systems have been found only in bacteria. Fraction of these systems consists of only three genes: *cas9*, *cas1* and *cas2* genes (type II-C) (Chylinski et al., 2014). Other type II systems has additional forth gene either *csn2* (Type II-A) or *cas4* (Type II-B) (Makarova et al., 2011b). *cas9* [previously known as *csn1* (Haft et al., 2005) or *cas5* (Barrangou et al., 2007)] is a hallmark gene of type II systems (Figure 3) that encodes a large multidomain protein. Cas9 in association with two additional RNA molecules forms effector complex, which destroys invading DNA (Gasiunas et al., 2012; Jinek et al., 2012).

Two type III CRISPR-Cas systems have been identified (type III-A and III-B) that are more common for archaea. Type III-B systems are usually found in conjugation with other CRISPR-Cas types. Cas10 is a signature protein of Type III (Figure 3), however, its function remains to be delineated (Makarova et al., 2011b). Proteins of these systems form Cascade-like effector complexes termed Csm in type III-A and Cmr in type III-B systems. It was shown that type III systems target RNA (Hale et al., 2009; Staals et al., 2013; Tamulaitis et al., 2014; Zhang et al., 2012b).

1.4. Mechanism of CRISPR-Cas systems

The CRISPR–Cas mechanism is arbitrarily divided into three main stages: (1) adaptation (spacer acquisition), (2) expression and processing (crRNA generation), and (3) interference (silencing) (Figure 4). During adaptation, the Cas proteins recognize invasive MGE and integrate short pieces of the foreign

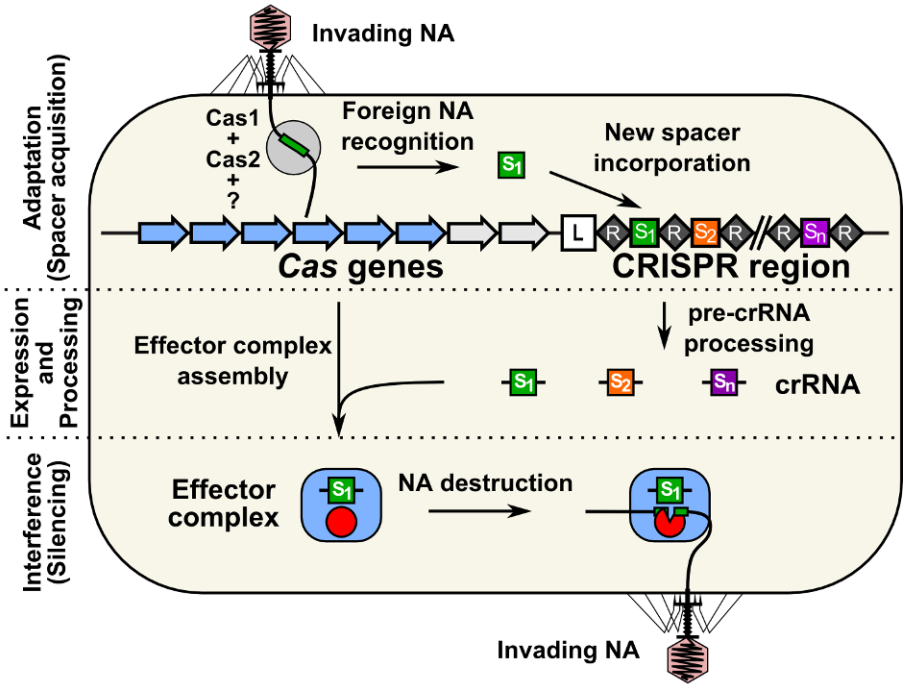

Figure 4. Schematic representation of CRISPR-Cas mechanism. The CRISPR–Cas mechanism is arbitrarily divided into three main stages: (1) adaptation or spacer acquisition, (2) expression and processing (crRNA generation), and (3) interference or silencing. During adaptation, Cas proteins recognize invasive nucleic acid (NA) and integrate short pieces of foreign DNA into the CRISPR region as new spacers (S_1). Spacers are inserted at the leader (L) proximal end followed by duplication of the repeat (R). In the expression and processing stage, the CRISPR repeat-spacer array is transcribed into a long primary RNA transcript (pre-crRNA) that is further processed into a set of small crRNAs, containing a conserved repeat fragment and a variable spacer sequence (guide) complementary to the invading nucleic acid. crRNAs further combine with Cas proteins into an effector complex. In the interference or silencing stage, the effector complex recognizes the target sequence in the invasive nucleic acid by base pairing and induces sequence-specific cleavage, thereby preventing proliferation and propagation of foreign genetic elements.

DNA into the CRISPR region as new spacers (Barrangou et al., 2007; Datsenko et al., 2012). Next, the CRISPR array is transcribed into a long primary RNA transcript that is further processed into a set of small CRISPR RNAs (crRNAs), containing a repeat fragment and a spacer sequence (guide), complementary to the invading nucleic acid (Carte et al., 2008; Deltcheva et al., 2011; Hale et al., 2008; Haurwitz et al., 2010). Lastly, crRNAs combine with Cas proteins into an effector complex which recognizes the target sequence in the invasive nucleic acid by base pairing to the complementary strand of double-stranded (ds) DNA (Gasiunas et al., 2012; Jinek et al., 2012;

Jore et al., 2011) or single-stranded (ss) RNA (Hale et al., 2009; Zhang et al., 2012b), and induces sequence-specific cleavage (Garneau et al., 2010), thereby preventing proliferation and propagation of foreign genetic elements.

CRISPR-Cas system is a powerful weapon protecting prokaryotic cell from phages or other MGEs. However, phages evolve rapidly and they have developed some strategies to overcome CRISPR-Cas immunity. Firstly, phages escape CRISPR-targeting simply by mutating protospacer or PAM sequences in its genome (Barrangou et al., 2007; Semenova et al., 2011). On the other hand, CRISPR-Cas systems can overcome these mutants by new spacer acquisition; furthermore, in some CRISPR-Cas systems adaptation is stimulated by mutant protospacer, which serves as a priming site for new spacer acquisitions (Datsenko et al., 2012; Swarts et al., 2012). Next, some phages encode proteins, termed anti-CRISPR, that inhibit CRIPR interference stage (Bondy-Denomy et al., 2013; Pawluk et al., 2014). Even more, there are phages that embezzle CRISPR-Cas systems and use them to evade the host (Seed et al., 2013).

1.4.1. Spacer acquisition stage

When a microbe, that carries the CRISPR-Cas system, encounters extracellular invader, such as phage, for the first time, CRISPR-Cas system must create a record about the invader in order to overcome the attack. This record is made in form of a new spacer during spacer acquisition (adaptation) stage of CRISPR-Cas action mechanism (Fineran & Charpentier, 2012; Westra et al., 2012a). Spacer is a small fragment of foreign DNA that is incorporated into CRISPR region at the leader sequence proximal end mediating Cas proteins (Figure 4). Although increasing experimental evidences draw the contours of this stage, the mechanistic details remain shaded.

1.4.1.1. Two pathways of spacer acquisition

For the first time, the spacer acquisition was observed in the CRISPR region of the *S. thermophilus* DGCC7710 strain (type II-A) in the phage

challenge experiments (Barrangou et al., 2007; Deveau et al., 2008). Spacer acquisition has been later detected under laboratory conditions in the *Haloarcula hispanica* type I-B (Li et al., 2014), *E. coli* type I-E (Datsenko et al., 2012; Yosef et al., 2012; Swarts et al., 2012), *Pseudomonas aeruginosa* (Cady et al., 2012) and *P. atrosepticum* (Richter et al., 2014) type I-F, *Streptococcus agalactiae* type II-A (Lopez-Sanchez et al., 2012), and *Sulfolobus solfataricus* type I and III-B (Erdmann & Garrett, 2012) CRISPR-Cas systems. *E. coli* type I-E CRISPR-Cas system remains the major model system for analysis of the spacer acquisition stage.

Overexpression of Cas1 and Cas2 proteins in *E. coli* strain lacking the intrinsic *cas* genes led to expansion of the CRISPR region (Yosef et al., 2012). This result proved that Cas1 and Cas2 are the sole Cas proteins essential for the spacer acquisition. The same study showed (Yosef et al., 2012) that a small portion of the leader sequence adjacent to the CRISPR array is essential for the acquisition step. Furthermore, repeat in the vicinity to the leader sequence is duplicated upon acquisition of a new spacer and only one repeat is sufficient for this event (Yosef et al., 2012). Recent study (Arslan et al., 2014) has identified intermediate states of the spacer integration into a genomic DNA of *E. coli* CRISPR array. It was shown that the insertion of the new spacer occurs by site-specific nicking at both strands of the leader-proximal repeat in a staggered way, and is accompanied by joining of the resulting 5′-ends of the repeat strands with the 3′-ends of the incoming spacer (Arslan et al., 2014) (Figure 5). Therefore, expansion of the CRISPR locus is polarized, integrating new spacers at the leader-proximal end of the CRISPR array, while the oldest spacers reside at the leader-distal end. This empowers genotyping of microbial strains by using spacer sequences of the CRISPRs (Comas et al., 2009; Cui et al., 2008). However, some strains of prokaryotes acquire spacers very rapidly and may delete part of CRISPR array, thus CRISPR arrays in different colonies are too diverse to use CRISPRs for genotyping (Pleckaityte et al., 2012). In some systems, spacer acquisitions are not polarized, for example, spacer

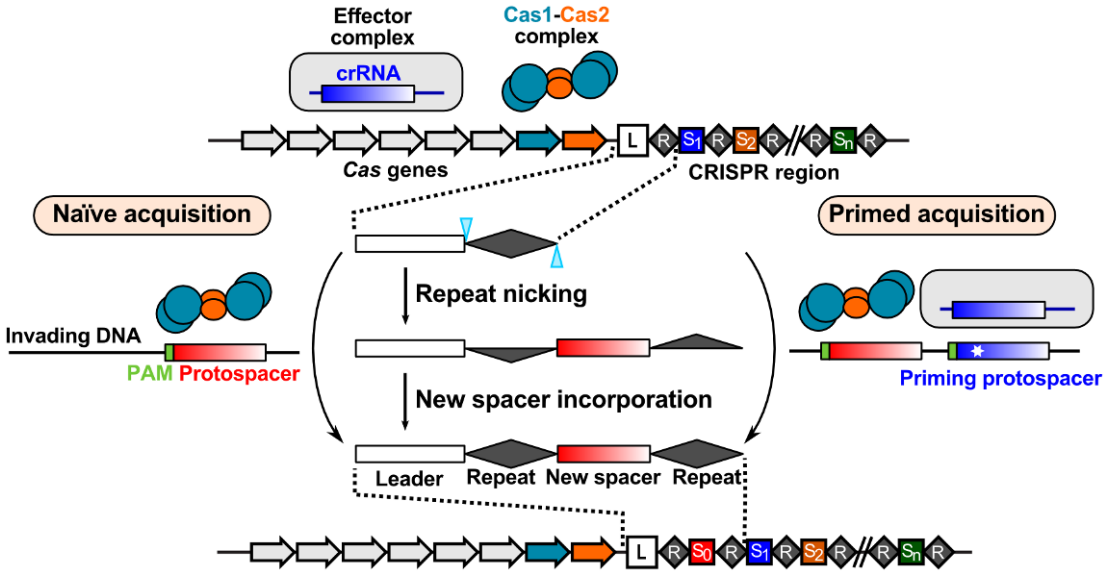

Figure 5. Schematic representation of the spacer acquisition stage. New spacers are incorporated at the leader-proximal end of a CRISPR region. Spacer acquisition begins with site-specific nicking (cyan triangles) at both strands of the leader (white rectangle)-proximal repeat (dark grey diamond) in a staggered way. The resulting 5′-ends of the repeat strands (dark grey triangles) are joined with the 3′-ends of the incoming spacer (red-white rectangle) and accompanied with complementary strand synthesis of single-stranded repeat regions. New spacer is polarized (depicted as a red-white colour gradient) according to the leader sequence in the same direction as the protospacer is polarized according to the PAM (green rectangle). If CRISPR region has no records from the invading DNA, spacers are acquired through naïve acquisition pathway mediating only Cas1-Cas2 protein complex. New spacer confers resistance against DNA invader; nevertheless, DNA invader can overcome CRISPR interference by introducing mutations (white star) at the PAM or protospacer sequences. This triggers primed spacer acquisition pathway where mutant protospacer (blue-white gradient-coloured rectangle with white star) serves as a priming site for both the Cas1-Cas2 complex and effector complex-mediated new spacer selection and incorporation.

acquisition events were detected throughout the CRISPR array in the *S. solfataricus* (Erdmann & Garrett, 2012).

Spacers are selected non-randomly; furthermore, there are hot-spots in the invading DNA with increased frequency of protospacer selection (Paez-Espino et al., 2013; Savitskaya et al., 2013). Major motive for the discrimination of foreign DNA is a protospacer adjacent motif (PAM) sequence. This 2-5 bp sequence is adjacent to the protospacer in the invading DNA but is not present in the vicinity of the spacer in the CRISPR array of the host. PAM sequences vary between CRISPR-Cas systems, and their position in respect to the

protospacer depends on the CRISPR-Cas type (Mojica et al., 2009). For example, *E. coli* type I-E CRISPR-Cas system recognizes 5′-AWG-3′ (Yosef et al., 2012) PAM sequence, while PAM sequence for *P. artrosepticum* type I-F CRISPR-Cas system is 5′-CC-3′ (Vercoe et al., 2013), and both PAMs are localized upstream of the protospacer. On the other hand, type II systems recognize PAM sequences 5′-AGAAW-3′ and 5′-NGGNG-3′, respectively for *S. themophilus* type II-A CRISPR1-Cas and CRISPR3-Cas (Deveau et al., 2008) that are positioned downstream of the protospacer. When a correct PAM is present in a vicinity of the protospacer, the spacer is polarized upon integration according to the leader sequence in the same direction as the protospacer is polarized according to the PAM (Mojica et al., 2009) (Figure 5). However, protospacers lacking correct PAM sequence are incorporated in both directions (Shmakov et al., 2014); thus, PAM orients spacer during acquisition. Contrarily to the type I and II systems, targets of type III systems do not contain PAM sequences; therefore, spacer selection should differ from type I and II systems (Shah et al., 2013). Recently, additional sequence motive was associated with increased spacer uptake into *E. coli* type I-E CRISPR locus. This acquisition affecting motif is 5′-AA-3′ sequence that is localized at the PAM-distal terminus of the protospacer (Yosef et al., 2013).

At least in the type I systems, adaptation can be dissected into two different pathways, termed naïve and primed (Figure 5). Naïve adaptation takes place when CRISPR-Cas encounters invading MGE for the first time. It relies on action of Cas1 and Cas2 proteins although in natural systems spacer acquisition is a rare event (Yosef et al., 2012). When MGE sequence has protospacer with correct PAM and perfectly matching spacer, CRISPR-Cas immunity destroys the invader (Barrangou et al., 2007). MGEs may escape CRISPR-Cas immunity by mutating target sequences; therefore, CRISPR-Cas systems developed a positive feedback loop, termed primed acquisition, which boosts up spacer acquisition events (Datsenko et al., 2012). All Cas proteins and the CRISPR array of the CRISPR-Cas system are prerequisite of this

pathway (Figure 5). It was shown that priming in type I-E of *E. coli* results in a spacer selection strand bias, whereby new spacers target the same strand as original priming spacer (Datsenko et al., 2012; Fineran et al., 2014; Swarts et al., 2012). However, in type I-F of *P. atrosepticum*, spacers were primed with a bias from both strands (Richter et al., 2014), while primed spacers were incorporated from both strands without detectable bias in type I-B of *H. hispanica* (Li et al., 2014). These results suggest that priming mechanisms may vary in distinct CRISPR-Cas systems.

1.4.1.2. Proteins associated with spacer acquisition stage

Cas1 and Cas2 are the most conserved proteins, which are found in all CRISPR-Cas systems (Makarova et al., 2011b). They are key players in the naïve spacer acquisition stage (Yosef et al., 2012) (Figure 5), although some additional proteins might be involved. Deletion of *csn2* gene in the type II-A systems of *S. thermophilus* resulted in vanished spacer acquisition events (Barrangou et al., 2007). Cas4 interacts with Cas1-Cas2 fusion protein in the type I-A of *Thermoproteus tenax*, suggesting that this protein participate in the acquisition process (Plagens et al., 2012). In some genomes, such as *Myxococcus xanthus*, the *cas1* and *cas4* genes are fused, further emphasizing their likely functional interaction (Makarova et al., 2013; Viswanathan et al., 2007).

Cas1. Cas1 protein is associated with wide range of biochemical activities. Three studies demonstrated that Cas1 proteins can hydrolyze nucleic acids in a divalent-metal-dependant and sequence non-specific manner. Cas1 protein from *Archaeoglobus fulgidus* (type I-A) hydrolyzes ssRNA and dsDNA molecules (Kim et al., 2013). Cas1 from *P. aeruginosa* (type I-F) degrades both ssDNAs and dsDNAs (Wiedenheft et al., 2009), while Cas1 protein from *E. coli* (type I-E), in addition to the ss and ds DNAs, cleaved branched DNAs, Holliday junctions and other intermediates of DNA repair and recombination (Babu et al., 2011). Moreover, in the presence of Cas2, it interacts with the CRISPR array (Nunez et al., 2014). On the other hand, Cas1 from *S.*

solfataricus (type I-A) binds DNA and RNA with high affinity in a sequence non-specific manner, while no detectable nuclease activity is found (Han et al., 2009).

The 3-D structures of Cas1 proteins from different CRISPR-Cas types are almost identical (Babu et al., 2011; Kim et al., 2013; Wiedenheft et al., 2009). A Cas1 monomer is arranged of the N-terminal domain (NTD) comprised as β-sandwich with a short α-helix and the mainly α-helical C-terminal domain (CTD). These globular domains are connected by a loop. Two monomers assemble to form a butterfly-like dimeric structure (Figure 6A). Dimerization

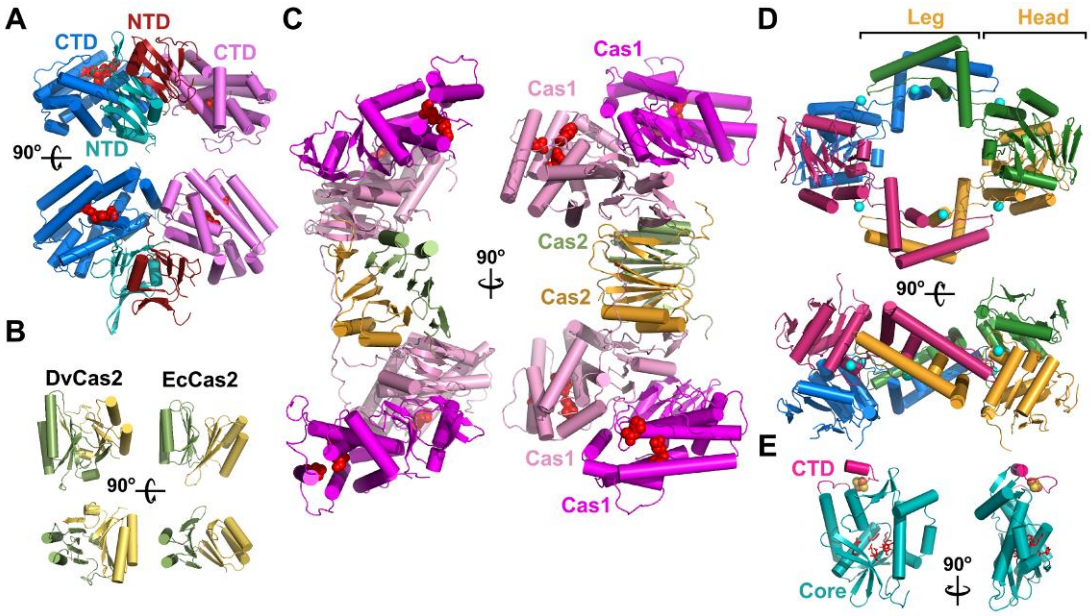

Figure 6. Crystal structures of Cas1, Cas2, Csn2 and Cas4 proteins involved in the spacer acquisition stage. (**A**) Cas1 is composed of an N-terminal (NTD) and a C-terminal (CTD) domain (coloured differently). Two NTDs interact with each other forming Cas1 dimer (*P. aeruginosa*; PDB: 3GOD). Active site amino acid residues (red spheres) are localised at the CTD. (**B**) Cas2 proteins adopt a ferredoxin-like fold. Arrangement of subunits in Cas2 dimer may differ, as exemplified in Cas2 dimers of *D. vulgaris* (Dv; PDB: 3OQ2) and *E .coli* (Ec; PDB: 4MAK). (**C**) Two Cas1 dimers sandwich one Cas2 dimer upon Cas1-Cas2 complex (*E. coli*; PDB: 4P6I) formation. Cas1 active site residues are represented as red sphares. (**D**) Csn2 is comprised of globular "head" and extended α-helical "leg" domains. "Leg" domain is stabilized by Ca^{2+} ions (cyan spheres). Four subunits (coloured differently) of Csn2 form a ring-shaped tetramereric structure (*E. faecalis*; PDB: 3S5U). (**E**) The Fe-S cluster is located at the C-terminal domain (CTD) of Cas4 protein (*P. calidifontis*; PDB: 4ONB) while exonuclease active site (red sticks) is positioned at the core domain. All crystal structures are represented in two orientations that are related by a 90° rotation.

interface is localized mainly between the NTDs (Babu et al., 2011; Kim et al., 2013; Wiedenheft et al., 2009). Conserved amino acid triad (glutamate-histidine-aspartate) coordinates metal ion, thus forming a catalytic site within the CTD in a positively charged pocket. Disruption of the active site by mutations leads to abolished nuclease activity *in vitro* and vanished spacer acquisition events *in vivo* (Yosef et al., 2012; Nunez et al., 2014; Wiedenheft et al., 2009).

Cas2. The biochemical activity of Cas2 proteins is ill-defined. Mg^{2+}-dependent ssRNAse activity has been reported for six different Cas2 proteins from type I-A and I-B CRISPR-Cas systems (Beloglazova et al., 2008). On the other hand, Cas2 from *Bacillus halodurans* (type I-C) shows DNase activity (Nam et al., 2012a), whereas protein from *Desulfovibrio vulgaris* (type I-C) neither cleaves nor binds nucleic acids at all (Samai et al., 2010).

A protomer of Cas2 homodimer adopts a ferredoxin-like fold (with βαββαβ topology of secondary structure) composed of two α-helices packing against antiparallel β-sheet (Beloglazova et al., 2008; Nam et al., 2012a; Samai et al., 2010). The C-terminal β-strand following ferredoxin fold may interact either with adjoining (as in the type I-C of *D. vulgaris*) or parental (as in the type I-E of *E. coli*) protomer by extending antiparallel β-sheet (Nunez et al., 2014; Samai et al., 2010) (Figure 6B). Aspartate residue localized near the N-terminus of the first β-strand was shown to be responsible for nuclease activities (Beloglazova et al., 2008; Nam et al., 2012a). However, resent study have demonstrated that mutation of the active site amino acid does not interfere with spacer acquisition *in vivo* (Nunez et al., 2014). This suggests that Cas2 activities are residual, or Cas2 participates in other cellular functions. Cas2 proteins are homologous to VapD toxin and may act as an RNase toxin in a putative toxin-antitoxin system (Makarova et al., 2012).

Cas1-Cas2 complex. A collective action of Cas1 and Cas2 proteins in spacer acquisition is rationalized by a structure of *E. coli* (type I-E) Cas1-Cas2

complex. It is demonstrated that two dimers of Cas1 sandwich one dimer of Cas2 leading to $Cas1_4$:$Cas2_2$ overall stoichiometry of the complex. Both subunits of Cas2 dimer interacts with only one subunit in each Cas1 dimer (Figure 6C). Furthermore, mutations of interaction interface abrogate spacer acquisition *in vivo* (Nunez et al., 2014).

Csn2. Crystal structures of *S. agalactiae* (Ellinger et al., 2012), *Streptococcus pyogenes* (Koo et al., 2012), *Enterococcus faecalis* (Nam et al., 2011) and *S. thermophilus* (Lee et al., 2012) Csn2 proteins are solved to date. A monomer contains a globular α/β domain (also termed "head" domain) and extended α helical domain (or "leg" domain) inserted in the middle of α/β domain. Some Csn2 proteins have an additional domain, composed of α helices, at the C-terminus (Lee et al., 2012). The α/β domain has a helicase fold; however, conservative regions responsible for ATPase activity are degenerated (Ellinger et al., 2012). Ca^{2+} ions are coordinated by conservative amino acids of the extended α helical domain and are structurally important (Nam et al., 2011). Csn2 dimerizes by extensive interactions of the α/β domains. Further, two dimers interact by the extended α helical domains forming a diamond-shaped tetrameric ring (Ellinger et al., 2012; Koo et al., 2012; Lee et al., 2012; Nam et al., 2011). The tetrameric structure contains a large positively charged central hole (Figure 6D), which is large enough to encircle the dsDNA. It was shown that Csn2 binds to the ends of linear dsDNA and diffuses inward, probably through rotation-coupled translocation (Arslan et al., 2013).

Cas4. Cas4 proteins contain a RecB-like nuclease motif and four conserved Cys residues, which coordinate an [Fe-S] cluster. *S. solfataricus* and *Pyrobaculum calidifontis* Cas4 proteins are exonucleases (like most of RecB-like nucleases) that degrade linear ssDNA with a 5′→3′ directionality starting from the 5′-end (Lemak et al., 2013; Lemak et al., 2015; Zhang et al., 2012a).

Cas4 proteins are composed of a core αβα domain, capped by a small, mostly α -helical subdomain. The core domain contains a V-shaped β -sheet,

which is surrounded by five α–helices, creating a large elongated cavity with wide entrance on one side and smaller exit on the other. The small subdomain is formed by three helices and a long loop where four conserved Cys residues coordinate the [4Fe-4S] or [2Fe-2S] cluster (Lemak et al., 2013; Lemak et al., 2015) (Figure 6E). Ten protomers of *S. solfataricus* Cas4 assemble into a toroidal structure of two pentameric rings with large central channel, while *P. calidifontis* Cas4 is a monomer (Lemak et al., 2013; Lemak et al., 2015).

1.4.2. Expression and processing stage

CRISPR-Cas systems use crRNA molecules as a guide for invading nucleic acid detection and destruction (Brouns et al., 2008; Marraffini & Sontheimer, 2008). These molecules originate from a long single transcript of the CRISPR array termed precursor crRNA (pre-crRNA) (Lillestol et al., 2009; Tang et al., 2002; Tang et al., 2005), which is processed into small crRNA molecules (Brouns et al., 2008; Carte et al., 2014; Hale et al., 2008; Richter et al., 2012). Depending on the system, mature crRNA molecules are generated in one or two steps. Primary maturation step occurs when the pre-crRNA is cleaved by endoribonuclease within a repeat sequences to yield a spacer sequence surrounded by repeat sequence fragments (Brouns et al., 2008; Deltcheva et al., 2011; Hatoum-Aslan et al., 2011; Haurwitz et al., 2010; Nam et al., 2012b). Further, crRNA molecule is trimmed either from 5′- or 3′-end during secondary maturation step (Deltcheva et al., 2011; Hale et al., 2009; Hatoum-Aslan et al., 2011; Richter et al., 2012), with the exception of the type I-C, I-E and I-F systems, which lack the second step (Brouns et al., 2008; Haurwitz et al., 2010; Nam et al., 2012b) (Figure7).

1.4.2.1. Pathways of crRNA maturation

Pathway of pre-crRNA processing varies between CRISPR-Cas systems. Most of the type I and III systems use endoribonucleases of Cas6 family for pre-crRNA processing (Brouns et al., 2008; Lintner et al., 2011; Przybilski et al., 2011) with the exception of type I-C system that uses Cas5c (also named

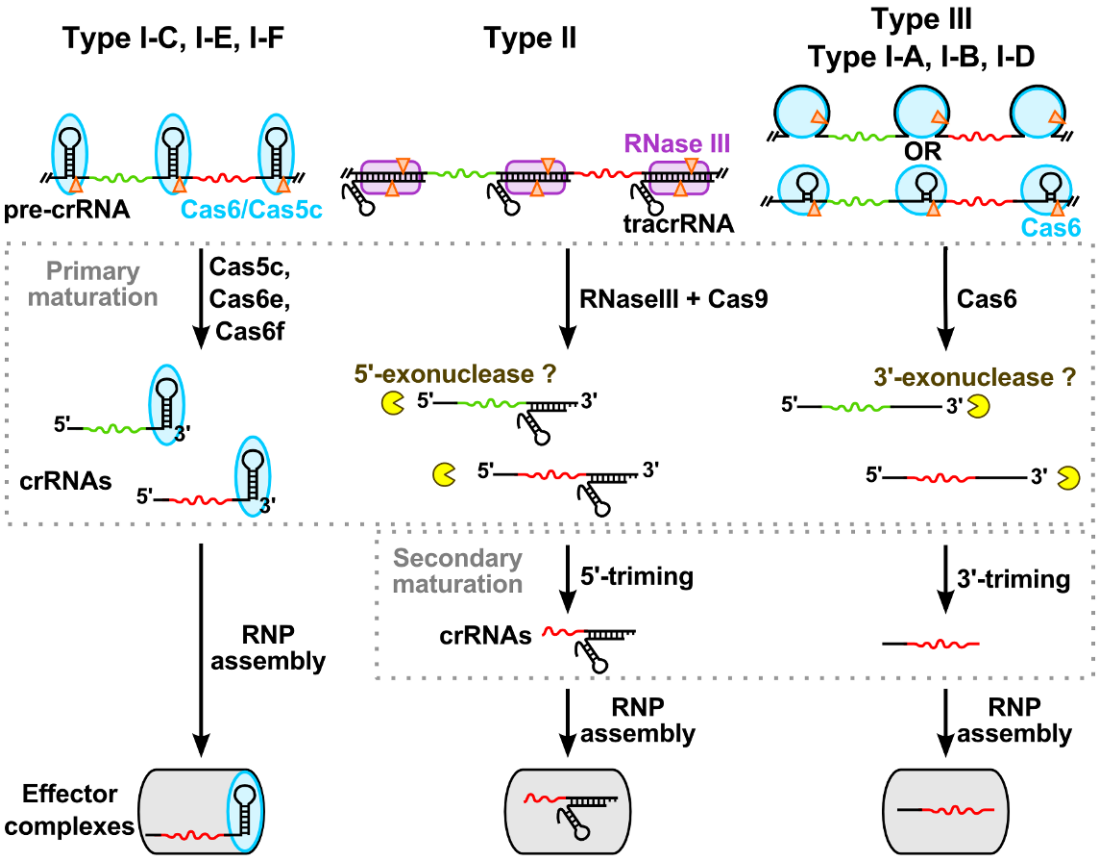

Figure 7. Schematic mechanisms of expression and processing stages. Transcription of a CRISPR locus results in a long transcript termed pre-crRNA which is processed into small crRNA molecules during one or two step process (framed in dotted rectangles). Endoribonuclease cleavage positions in the pre-crRNA are indicated by orange triangles. Spacer and repeat sequence of crRNAs are shown as red or green wavy lines and black solid lines, respectively. Type I and III utilises Cas6 or Cas5d endoribonucleases in the primary maturation step while type II uses RNase III that cleaves within repeat-tracrRNA duplex in a Cas9-dependent fashion. During secondary maturation step crRNA molecules are trimmed by unknown nucleases (yellow Pac-Man shape) from the 5′-terminus in the type II or 3′-terminus in the type III, I-A, I-B and I-D systems. crRNA in the type I-C, I-E and I-F CRISPR-Cas systems is not subjected to a second maturation step. Mature crRNAs assemble together with Cas proteins into ribonucleoprotein (RNP) effector complexes that provide the interference.

Cas5d) nuclease instead (Garside et al., 2012; Koo et al., 2013; Nam et al., 2012b).

Cas6 homologs specifically cleave phosphodiester bond within a repeat sequence in a metal-independent manner. The resulting crRNAs typically have repeat-derived 5′-handle of 8 nt followed by the intact spacer and repeat-derived 3′-handle of variable length that forms a hairpin structure in some

systems (Carte et al., 2014; Haurwitz et al., 2010; Jore et al., 2011). Cas5c endoribonuclease produces a 5′-handle of 11 nt and a 21-26 nt 3′-handle (Garside et al., 2012; Nam et al., 2012b). Multiple turn-over Cas6 proteins of type I-A, I-B, I-D, III-A and III-B do not associate with the product in an effector complex of these systems; therefore, crRNAs are further processed by yet unidentified nucleases that trim 3′-repeat sequence (Hale et al., 2008; Hatoum-Aslan et al., 2011; Nickel et al., 2013; Richter et al., 2012). Furthermore, other Cas proteins besides Cas6 may be implicated in the crRNA maturation processes (Brendel et al., 2014; Hatoum-Aslan et al., 2011; Hatoum-Aslan et al., 2013; Li et al., 2013b). In contrast, crRNAs produced by Cas5c (type I-C), Cas6e (type I-E) and Cas6f (type I-F) proteins persist further processing and assemble with these and other Cas proteins into a large RNP Cascade complex (Jore et al., 2011; Nam et al., 2012b; Wiedenheft et al., 2011b) (Figure 7).

The pathaway for crRNA maturation is different for type II systems. In addition to *cas* operon and CRISPR array, CRISPR-Cas systems of this type encode a *trans*-activating crRNA (tracrRNA) which has a sequence partialy complementary to the repeat region of the pre-crRNA (Chylinski et al., 2013). RNA duplex of the pre-crRNA and tracrRNA is recognized and cleaved by a housekeeping ribonuclease RNase III in the presence of Cas9 protein (Deltcheva et al., 2011). The 5′-end of the crRNA is further trimmed by an unidentified nuclease to produce a mature crRNA with a 20 nt spacer fragment and a 3′-handle of 19-22 nt which remains associated with the tracrRNA and Cas9 in an effector complex of CRISPR interference (Deltcheva et al., 2011; Gasiunas et al., 2012; Jinek et al., 2012; Karvelis et al., 2013a) (Figure 7). However, RNase III maturation step could be bypassed by separate promoters within each CRISPR repeat as exemplified by the type II-C of *Neisseria meningitidis* (Zhang et al., 2013), or artificially reducing CRISPR array to a single repeat-spacer-repeat unit (Karvelis et al., 2013a).

1.4.2.2. Proteins of primary maturation step

Cas6 and Cas5 both belong to the repeat-associated mysterious protein (RAMP) superfamily (Makarova et al., 2011a). Up to five conserved motives can be identified in sequence of RAMP proteins, of which glycine (G)-loop (motif V) is the most prominent (Ebihara et al., 2006). In general, RAMP proteins have two domains of ferredoxin fold [also known as RNA recognition motif (RRM)] linked in tandem (Wang & Li, 2012). The N-terminal domain (NTD) of all Cas6 proteins resembles the classic ferredoxin fold, which harbours active site residues of a ribonuclease. The C-terminal domain (CTD) is often interrupted by insertions or deviates in the conformation of the fold (e.g. CTD of Cas6f shows a degraded ferredoxin fold); nonetheless, it embodies conserved structural elements, such as G-loop and β-hairpin, that are critical for repeat binding (Reeks et al., 2013b) (Figure 8 A and B). Usually, Cas6 is a monomer (Carte et al., 2008; Haurwitz et al., 2010; Sashital et al., 2011), however, some homologs are dimers (Niewoehner et al., 2014; Shao & Li, 2013), or can dimerize upon repeat binding (Wang et al., 2012).

The Cas5c contains an N-terminal ferredoxin-like domain (NTD) and a C-terminal twisted β-sheet domain (CTD). The NTD of Cas5c resembles the CTD of Cas6; it contains the β-hairpin and the G-loop, which sequence differs from the Cas6s' consensus (Nam et al., 2012b) (Figure 8 C).

Cas5c and Cas6 are both metal-independent endoribonucleases that produce products with 5′-hydroxyls and 3′-cyclic phosphates by a general acid-base mechanism when deprotonated 2′-hydroxyl performs a nucleophilic attack on the scissile phosphate (Carte et al., 2008; Garside et al., 2012; Haurwitz et al., 2010; Jore et al., 2011). Active site residues are usually localized at the NTD of Cas6s. Generally, a semiconserved histidine is observed at the active site, however, other amino acids deviate in different homologs (Carte et al., 2008; Haurwitz et al., 2012; Reeks et al., 2013c; Sashital et al., 2011; Shao & Li, 2013). The putative Cas5c active site is possitioned differently to that of Cas6 (Figure 8 B and C), suggesting that the active sites evolved independently of

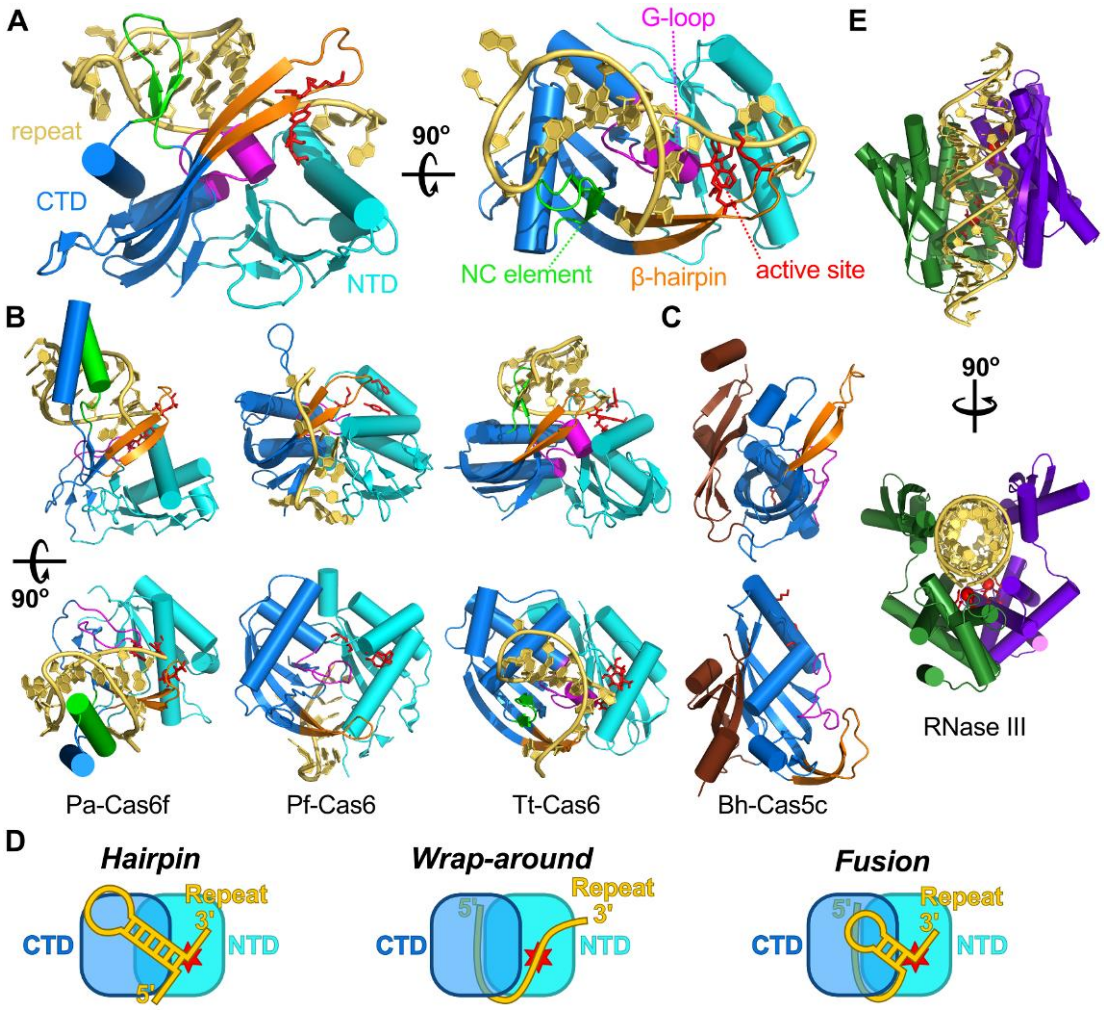

Figure 8. Comparison of pre-crRNA processing endoribonucleases. (**A**) Cas6e from type I-E of *T. thermophilus* (PDB: 2Y8W) is comprised of N-terminal (NTD; cyan) and C-terminal (CTD; marine) ferredoxin-fold domains. It recognize hairpin-structured repeat (yellow) using a non-conserved β-hairpin element (NC element; green) as well as conserved β-hairpin (orange) and G-rich loop (G-loop; magenta). Active site residues are represented as red sticks. (**B**) Cas6f from type I-F of *P. aeruginosa* (Pa-Cas6f; PDB: 4AL5) recognize hairpin-structured repeat while Cas6 from type III-A of *P. furiosus* (Pf-Cas6; PDB: 3PKM) binds unstructured repeats. Cas6 from type I-B of *T. thermophilus* (Tt-Cas6; PDB: 4C8Z) recognize repeats with unstable hairpins, however, 5′-unstructured part of the repeat is necesary for efficient binding. All structural elements are coloured as in (A). (**C**) Cas5c is a repeat endoribonuclease in type I-C systems. The NTD and structural elements of Cas5c from *B. halodurans* (Bh-Cas5c; PDB: 4F3M) are coloured as in structurally homologous CTD of Cas6. The CTD of Cas5c is brownish red. (**D**) Cas6 proteins developed three mechanisms (hairpin, wrap-around and fusion) to bind repeat RNA. Active sites are represented as red stars. Structural examples of these mechanisms are shown in (A) and (B). (**E**) Differently from type I and III, a non-Cas protein, RNase III is employed in the type II systems (PDB: 2NUG) for cleavage of repeat-tracrRNA duplexes. Subunits of the RNase III dimer are coloured in green and purple. Residues of active sites are represented as red sticks; RNA duplex is coloured in yellow. All crystal structures are represented in two orientations related by a 90° rotation.

each other. Moreover, catalytic residues among proteins of Cas5c family are diverse (Garside et al., 2012; Koo et al., 2013; Nam et al., 2012b).

Although overall fold of Cas6 proteins are similar, structural mechanism for the repeat recognition and cleavage is markedly different between Cas6 orthologs (Figure 8 A and B). Knowing diversity of repeat sequences and their secondary structures (Kunin et al., 2007), this could be rationalized by a common evolution of both repeats and Cas6 proteins. Structured and unstructured repeats are recognized by at least two different mechanisms (Figure 8 D). An ssRNA of unstructured repeat sequence in the pre-crRNA is wrapped around Cas6 protein measuring cleavage position in a ruler-like manner. On the other hand, hairpin-structured repeats are recognized as dsRNA and cleaved at the 3′-basis of the hairpin.

Positively charged groove is formed between two ferredoxin-fold domains of Cas6 from *Pyrococcus furiosus* (type I-A/III-B) and *Pyrococcus horikoshii* (type I-A) where 8 nt sequence at the 5′-end of the unstructured repeat is specifically bound; however, active site residues are localized on the opposite side (Carte et al., 2010; Carte et al., 2008; Wang et al., 2012) (Figure 8 B). Therefore, repeat wraps around the protein to reach the active site (Wang et al., 2011).

On the other hand, recognition and cleavage of structured repeats occurs on the same face of Cas6, although residues responsible for these activities are localized in different domains (Figure 8 D). Hairpin of the repeat is specifically recognized by structural elements of the CTD. Cas6e from *Thermus thermophilus* (type I-E) provides specific contacts by inserting a non-conserved β-hairpin into a major groove of the hairpin‘s helix (Gesner et al., 2011; Sashital et al., 2011), while Cas6f from *P. aeruginosa* (type I-F) interacts through arginine-rich α-helix instead (Haurwitz et al., 2010). Furthermore, conserved β-hairpin and G-loop provide additional interaction contacts with the substrate (Gesner et al., 2011; Haurwitz et al., 2010) (Figure 8 A and B). Cas6e forms base-specific contacts with one unpaired nucleotide upstream the

hairpin along with 3 nt downstream (Sashital et al., 2011). In contrast, stem-loop structure is the only determinant for Cas6f binding and sequence or length changes of the stem and loop is not tolerated (Haurwitz et al., 2012; Sternberg et al., 2012).

Some repeats have short (3-4 nt) palindromic sequences, which may form unstable hairpins in solution; nevertheless, Cas6 from *S. solfataricus* (type I-A) stabilizes hairpin structure upon binding (Reeks et al., 2013c; Shao & Li, 2013; Sokolowski et al., 2014). Cas6 from *T. thermophilus* (type I-B) associates with short hairpin structure (Figure 8 B); however, ssRNA sequence at the 5′-end of the repeat is essential for efficient binding (Niewoehner et al., 2014). Therefore, Cas6 proteins might use fusion of the wrap-around and the hairpin recognition mechanisms to discriminate the repeat sequences with unstable hairpin structures (Figure 8 D). Repeat with a long hairpin is a substrate for the Cas5c from *B. halodurans*, however, structural evidences are lacking for exact recognition mechanism (Nam et al., 2012b).

Repeats of type II systems are processed by RNase III. This protein is a Mg^{2+}-dependant, dsRNA-specific endoribonuclease (Macrae et al., 2006). Double-stranded regions result from hybridization of tracrRNAs with the repeats in pre-crRNA (Deltcheva et al., 2011). The enzyme generates 5′-phosphoryl and 3′-hydroxyl ends with a 2 nt overhang in the dsRNA product. RNase III is conserved and widely distributed among bacterial and eukaryotic proteins, such as Drosha and Dicer (Court et al., 2013). It is composed of catalytic domain (RIIID) at the N-terminus and dsRNA binding domain (dsRBD) at the C-terminus. Seven α-helices embody the RIIID, while 2 α-helices stack onto a three-stranded antiparallel β-sheet forming the dsRBD. Two RIIIDs associate with each other forming a catalytic cleft where dsRNA is positioned by clamping it with dsRBDs (Gan et al., 2008) (Figure 8 E).

1.4.3. Interference stage

Mature crRNAs combine with Cas proteins into an effector complex, which guards the cell from foreign genetic elements. The complex recognizes the target sequence in the invasive nucleic acid by base pairing to the complementary strand of double-stranded DNA (Jore et al., 2011) or single-stranded RNA (Hale et al., 2009; Zhang et al., 2012b), and induces sequence-specific cleavage (Garneau et al., 2010). Mechanistic details of interference as well as structural features of the complexes diverge between different CRISPR-Cas types; therefore, interference stages of type I, II, and III will be discussed separately.

1.4.3.1. DNA interference in type I CRISPR-Cas systems

Type I systems are subdivided into six subtypes that differ by the number and arrangement of *cas* genes (Makarova et al., 2011b) (Figure 3). Despite of the differences, all type I systems encode a hallmark Cas3 protein and a large multisubunit RNP complex termed Cascade (CRISPR-associated complex for antiviral defense). These components confer resistance against invading dsDNA. Cascade using crRNA as a guide locates the target DNA in a process that yet has to be defined, and binds to the complementary DNA strand creating an R-loop, if a short PAM (Jore et al., 2011) is present upstream of the matching protospacer. However, Cascade binding to the matching sequence in the invading DNA does not trigger silencing. Degradation of the foreign DNA requires an accessory Cas3 protein, which alongside with Cascade is essential for DNA interference *in vivo* (Brouns et al., 2008) (Figure 9 A).

Cascade surveillance complex. RNP complexes of type I CRISPR-Cas interference are comprised of crRNA and Cas proteins. Low resolution electron microscopy structures have been solved for *E. coli e*Cascade [first letter "*e*" represents type I-E from which Cascade is originated (Reeks et al., 2013b)] apo-structure and in complex with protospacer RNA as well as dsDNA (Hochstrasser et al., 2014; Jore et al., 2011; Wiedenheft et al., 2011a), *B. halodurans c*Cascade (Nam et al., 2012b), *P. aeruginosa f*Cascade

(Wiedenheft et al., 2011b) as well as the core complex of *S. solfataricus* *a*Cascade (Lintner et al., 2011). Although composition of Cas proteins in Cascade complexes differs between different subtypes (Brendel et al., 2014; Brouns et al., 2008; Nam et al., 2012b; Plagens et al., 2012; Wiedenheft et al., 2011b), overall architecture of the complexes are very similar.

Recently high resolution X-ray structures have been solved for *E. coli* *e*Cascade in an apo-form and in a complex with complementary ssDNA (Jackson et al., 2014a; Mulepati et al., 2014; Zhao et al., 2014) (Figure 9 B and C). The complex is comprised of 5 Cas proteins that assemble on a crRNA

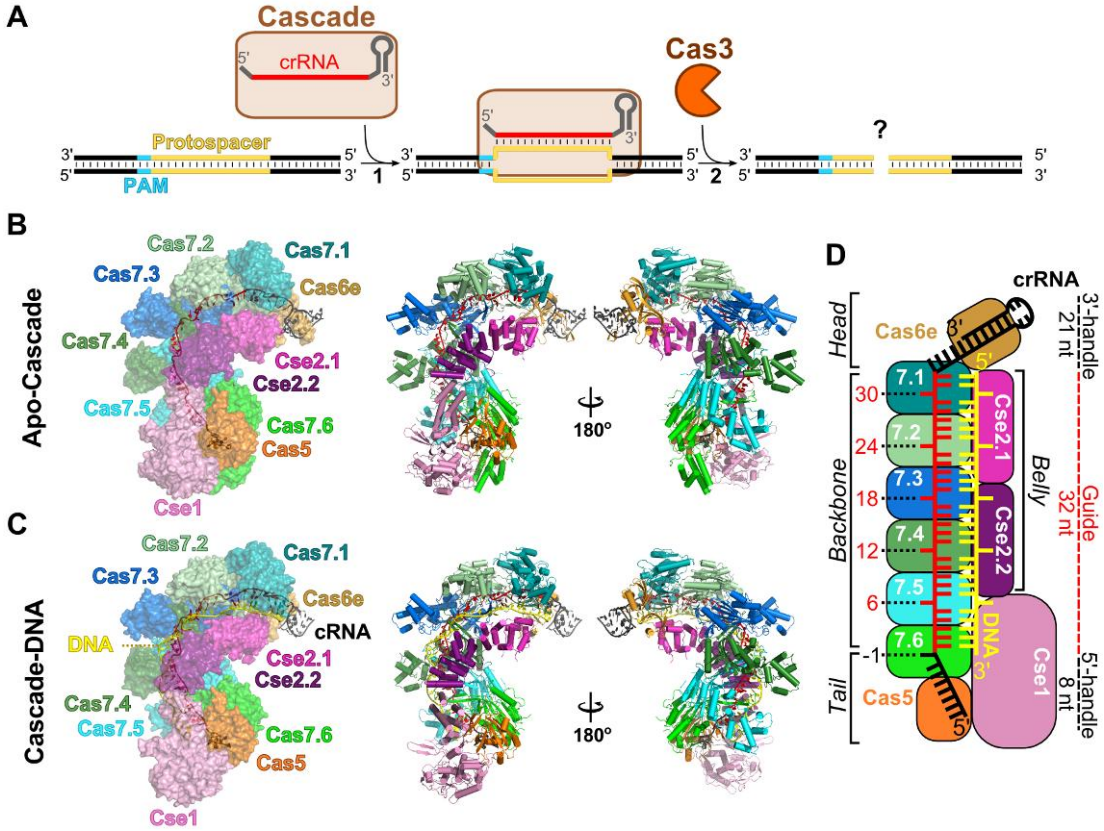

Figure 9. Type I interference mechanism and structure of the Cascade complex. (**A**) Cascade is a surveillance complex of the type I that detects foreign DNA. Cascade targets DNA by binding to the complementary strand of a protospacer in a PAM-dependant manner that leads to an R-loop formation (1). Cascade-DNA complex recruits Cas3 which probably cuts DNA target leading to DNA interference (2). Crystal structure of (**B**) apo-Cascade (PDB: 1VY9) and (**C**) Cascade-DNA (PDB: 4QYZ) complexes from *E. coli* type I-E. Cas protein subunits, crRNA and DNA target are coloured differently. Cartoon representations of Cascade complexes are shown in two different orientations related by a 180° rotation. (**D**) Schematic representation of subunit arrangement in the Cascade. Colouring is the same as in (B) and (C).

carcass with the following stoichiometry: $(Cse1)_1$:$(Cse2)_2$:$(Cas7)_6$:$(Cas5)_1$:$(Cas6e)_1$:$(crRNA)_1$ (van Duijn et al., 2012). The mature 61 nt crRNA (Jore et al., 2011) is comprised of 32 nt spacer flanked by 8 nt 5′- and 21 nt 3′-handles resulting from the pre-crRNA cleavage within repeat stems by the Cas6e endoribonuclease (Gesner et al., 2011; Sashital et al., 2011). The handles are anchored at the opposite ends of the sea-horse-shaped complex. Cas6e binds 3′-handle at the head of the complex, while 5′-handle is sandwiches between Cse1, Cas5 and Cas7.6 subunits in the tail. Helical backbone of six Cas7 subunits (Cas7.1-Cas7.6) embed spacer region of crRNA. Two Cse2 subunits (Cse2.1 and Cse2.2) form a belly, which connects Cse1 and Cas6e proteins (Jackson et al., 2014a; Zhao et al., 2014) in the apo-Cascade, while interaction between Cse2.2 and Cas6e is disrupted upon Cascade-DNA complex formation (Mulepati et al., 2014) (Figure 9 D).

Cas7 of the backbone. The backbone of *e*Cascade is comprised of six Cas7 proteins that oligomerize along the crRNA forming an interwoven architecture (Figure 10 A). The Cas7 protein is shaped like a right hand. The modified ferredoxin-fold (or RRM; belongs to RAMP superfamily) domain forms the palm, a loop inserted in the ferredoxin-fold creates the thumb, and a helical domain takes on the shape of the fingers (Figure 10 B). Spacer region of crRNA is divided into six discrete segments. Five nucleotides of each segment interact with two adjacent Cas7 subunits that order these spacer segments in a pseudo A-form configuration with solvent-exposed bases. Sixth nucleotide of the segment is sandwiched between the thumb from one Cas7 subunit and a helix of the palm from the adjacent subunit (Jackson et al., 2014a; Zhao et al., 2014). Protospacer and crRNA form a DNA-RNA hybrid with every exposed five nucleotide segment, while the thumb interfere with interaction at the 6^{th} nucleotides (Figure 10 B); therefore, mutations at these positions have no impact on the target binding (Fineran et al., 2014; Jackson et al., 2014a; Mulepati et al., 2014). On the other hand, disruption of complementarity at 1-5, 7 and 8 positions of crRNA spacer region and target results in incapability of

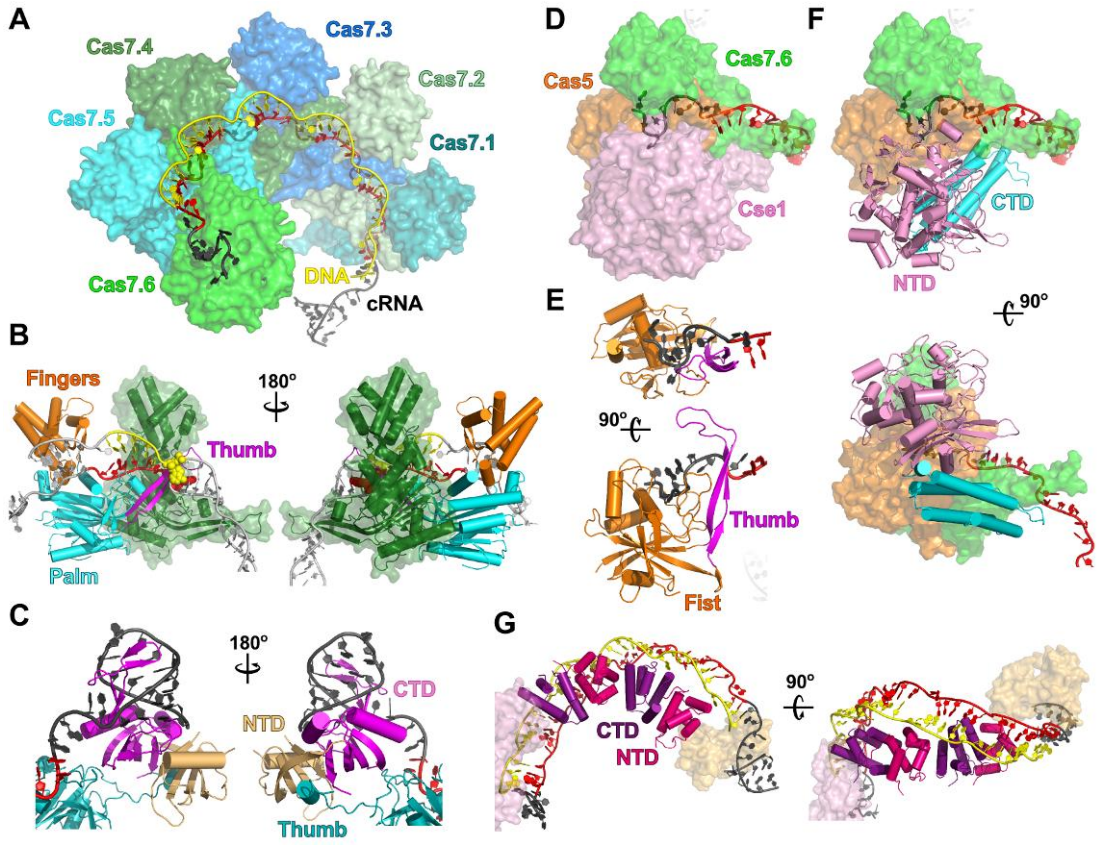

Figure 10. Crystal structures of *e*Cascade subunits. (**A**) Six Cas7 subunits interweave guide of a crRNA forming backbone of the Cascade. (**B**) Right-hand shaped Cas7 is comprised of palm (cyan), thumb (magenta) and fingers (orange) domains. Two adjacent Cas7 subunits separate guide into 6 nt segments (red) where 5 nt base pairs with target DNA (yellow) while 6th nucleotides (shown in sphere representation) of guide and target are kinked by the thumb. (**C**) Cas6e of the head is composed of N- (pale yellow) and C-terminal (magenta) ferredoxin-fold domains (NTD and CTD) that bind hairpin of the crRNA 3′-handle and thumb of the Cas7.1 subunit. (**D**) The 5′-handle of crRNA is anchored by Cas5, Cas7.6, and Cse1 subunits. (**E**) The 5′-handle is clenched by a thumb (magenta) and ferredoxin-fold fist domain (orange) of the Cas5. (**F**) Cse1 is the largest domain of the Cascade composed of N- (pink) and C-terminal (cyan) domains (NTD and CTD). Cse1 is positioned at the tail. (**G**) Head-to-tail dimer of Cse2 forms a belly of the Cascade. Cse2 subunit is composed of α-helical N- (hot pink) and C-terminal (deep purple) domains (NTD and CTD). Crystal structures are represented in two orientations related by a 180° (**B**, **C**) or 90° (**F-G**) rotation. Structures represent 1VY9 and 4QYZ PDB entries.

the *e*Cascade to bind the target (Semenova et al., 2011). This protospacer region is located adjacent to the PAM sequence and named a seed sequence. The importance of the seed sequence was also demonstrated in the *b*Cascade (*Haloferax volcanii*) (Maier et al., 2013) and *f*Cascade (*P. aeruginosa*)

(Wiedenheft et al., 2011b). The PAM and seed may play a role in the initial recognition of the invader DNA.

Overall structure of Cas7 subunits is similar; however, Cas7 subunits that interact with the head and the tail show conformational differences. The thumb of Cas7.1 is rotated about 70° comparing with other Cas7 subunits; moreover, it forms a short helix that fits into the hydrophobic groove of Cas6e, tethering the head and backbone. On the other hand, the finger domain of Cas7.6 is rotated 180˚, providing a platform for the recruitment of Cse1 to the tail (Jackson et al., 2014a; Zhao et al., 2014).

Structure of Cas7 from type I-A (*S. solfataricus*) (Lintner et al., 2011) as well as Csc2 from type I-D (*Thermofilum pendens*) (Hrle et al., 2014) is similar to Cas7 from *e*Cascade (*E. coli*) (Jackson et al., 2014a; Mulepati et al., 2014; Zhao et al., 2014). Furthermore, *a*Cascade subcomplex of Cas7 and Cas5 assembles into helical structures (Lintner et al., 2011). Helical arrangement similar to *e*Cascade was also observed in EM images of *c*Cascade and *f*Cascade (Nam et al., 2012b; Wiedenheft et al., 2011b). Therefore, it is suggested that helical backbone of Cas7 family proteins is a conserved and perhaps a characteristic feature of all Cascade complexes (Makarova et al., 2011a; van der Oost et al., 2014).

Cas6e of the head. Cas6e is a metal-independent endoribonuclease composed of two tandem ferredoxin-fold domains. The protein recognizes hairpin structure of the repeat and cleaves it at the basis of the stem as discussed previously (see section 1.4.2.2.). Cas6e as well as Cas6f remains associated with the hairpin structure, which is positioned at the 3′-handle of crRNA (Haurwitz et al., 2010; Sashital et al., 2011). In contrast, Cas6 from type I-A is not a component of the *a*Cascade (Plagens et al., 2012). The Cas6e is entrenched by interactions with the major groove of crRNA hairpin and with a 3 nt upstream the basis of hairpin as well as with the thumb of Cas7.1 (Jackson et al., 2014a; Zhao et al., 2014) (Figure 10 C).

Cas5 of the tail. The S-shaped 5′-handle is anchored mainly by the Cas5 protein. The Cas5 embodies a right-handed fist-shape structure. The fist is composed of a modified ferredoxin-fold domain with an insertion, which forms a thumb. The thumb of Cas5 folds over the top of the first kinked base, performing similar function as in the Cas7. However, unlike straight thumb on Cas7 subunits, the Cas5 thumb bends over the top of the fist and, interacting with the fingers of Cas7.6 subunit, clenches 5′-handle. The Cas7.6 palm and the Cse1 protein sandwiches the fist of Cas5 forming the tail of the complex (Jackson et al., 2014a; Zhao et al., 2014) (Figure 10 D and E).

Cas5 proteins typically do not show catalytic activity in the *in vitro* assays with the exception of Cas5c from the type I-C, which is a metal-independent endorybonuclease and cleaves within the repeat sequences (Nam et al., 2012b). Cas5c structure is very similar to Cas5 of *e*Cascade, however, it has evolved the endorybonuclease active site within the ferredoxin-fold domain (Nam et al., 2012b).

Cse1 of the tail. The Cse1 protein is the largest subunit of *e*Cascade. It is comprised of an N-terminal globular domain and a C-terminal four-helix bundle domain (Mulepati et al., 2012; Sashital et al., 2012). The globular domain docks at Cas5 side; moreover, it makes specific contacts with three nucleotides of 5′-handle by inserting a short α-helix to a cylindrical pore on Cas5 (Figure 10 F). The four-helix bundle makes contacts with tail components as well as with the belly (Jackson et al., 2014a; Zhao et al., 2014).

Cse1 from *E. coli* was implicated in the PAM sequence (5′-AWG-3′) recognition as well as interaction with the Cas3 protein (Hochstrasser et al., 2014; Sashital et al., 2012; Westra et al., 2012b). However, amino acid residues responsible for these interactions still need to be identified. It was suggested that large subunits Cas8, Cas10d and Csy1 from other type I subtypes could be structural or functional homologs (Makarova et al., 2011b) of Cse1, however, this hypothesis still lacks the experimental evidences.

Cse2 of the belly. The Cse2 protein is the smallest subunit of *e*Cascade composed of two α-helical domains. Two Cse2 subunits (Cse2.1 and Cse2.2) dimerize in a head-to-tail fashion forming the belly (Figure 10 G). The C-terminal domain of Cse2.2 subunit interacts with Cse1 four-helix bundle, while the N-terminal domain of Cse2.1 interacts with the Cas6e in the apo-Cascade complex (Jackson et al., 2014a). Basic amino acid residues of Cse2 stick between Cas7 subunits (Zhao et al., 2014).

Binding of target DNA is accompanied by the rearrangement of Cse1 and Cse2 subunits. Contacts between Cse2.1 and Cas6e are disrupted during this conformational change. Cse2 subunits form contacts with target DNA strand upon the R-loop formation. The belly has an additional basic patch, which probably stabilizes the displaced non-target DNA strand in the R-loop (Mulepati et al., 2014).

There are two type I Cascades that have small subunits (Makarova et al., 2011b), i.e. Cse2 of *e*Cascade and Csa5 of *a*Cascade. Csa5 (*S. solfataricus*) protein has been shown to be structurally similar to the C-terminal domain of the Cse2 (Reeks et al., 2013a). Moreover, *T. tenax* Csa5 was implicated in the R-loop stabilization (Daume et al., 2014); therefore, it seems that Cse2 and Csa5 perform similar functions. Four of six type I subtypes do not have small Cascade subunits (Makarova et al., 2011b), thus these complexes act differently or other Cas proteins fulfil the function of the Cse2 and Csa5.

Cas3 is the accessory protein of the DNA interference. Cascade binding to the matching sequence of the invading DNA does not trigger silencing *in vivo*, which requires an accessory Cas3 protein (Brouns et al., 2008). Cas3 is a signature protein of the type I systems (Makarova et al., 2011b) and typically contains HD phosphohydrolase and Superfamily 2 (SF2) helicase domains, arranged into a single subunit protein; however, HD- and helicase domains sometimes are encoded as individual Cas3′ and Cas3′′ subunits, respectively (Haft et al., 2005; Jackson et al., 2014b; Makarova et al., 2006; Makarova et al., 2011b) (Figure 3). Furthermore, in some CRISPR systems, the single chain

Cas3 or separate Cas3 domains are fused to other Cas proteins (e.g., Cas2-Cas3, Cas3-Cse1) (Makarova et al., 2011b; Westra et al., 2012b). In most cases, Cas3 is expressed as a stand-alone protein and associates with Cascade only in the presence of target DNA; however, Cas3′ and Cas3′′ from type I-A in *T. tenax* are subunits of *a*Cascade complex (Plagens et al., 2012). It is postulated that Cas3 is a nuclease-helicase; nevertheless, it should be demonstrated experimentally.

1.4.3.2. DNA interference in type II CRISPR-Cas systems

All type II CRISPR-Cas systems contain a conserved signature protein, Cas9 (Makarova et al., 2011b), which is the sole Cas protein responsible for the DNA interference in these systems (Garneau et al., 2010; Sapranauskas et al., 2011). Cas9 is a large multidomain protein that associates with a mature crRNA and tracrRNA, forming the effector complex (Deltcheva et al., 2011; Jinek et al., 2012; Karvelis et al., 2013a). There are two major differences between mature crRNAs in type II and I systems. First, crRNA in type II lacks a 5′-handle and contains an extended 22 nt 3′-handle generated by the RNase III cleavage within the repeat region of the pre-crRNA:tracrRNA duplex. Second, the spacer fragment in the type II crRNA is shorter, because the 5′-end of the spacer sequence is trimmed to 20 nt by unknown nuclease(s). Consequently, the spacer in the mature crRNA matches only 20 nt of 30 nt protospacer sequence in the invading nucleic acid. The non-matching fragment in the protospacer is not important for the CRISPR-mediated immunity; however, shortening of the protospacer sequence to 19 nt or more abrogates CRISPR-mediated plasmid interference (Gasiunas et al., 2012; Jiang et al., 2013; Jinek et al., 2012).

Similarly to type I Cascade complex, type II effector complex targets dsDNA, bearing complementary protospacer adjacent to a PAM sequence, throughout a seed-sequence-mediated R-loop formation. Contrary to Cascade, Cas9 recognizes PAM sequence that is localized immediately downstream, but not upstream, of the protospacer. Furthermore, type II systems do not use

accessory proteins, such as Cas3 in type I, for target destruction. Cleavage of the target DNA is performed by two signature nuclease domains (HNH and RuvC) of Cas9 protein. These differences are governed by fundamentally different structural organization of Cas9 in comparison with the Cascade.

Cas9 effector complex. High resolution structural data for Cas9 has been obtained recently. Firstly, the type II-A (*S. pyogenes*) and type II-C (*Actinomyces naeslundii*) Cas9 crystal structures were solved in the absence of nucleic acids (Jinek et al., 2014). A bilobed apo-Cas9 organization was uncovered by these structures (Figure 11). Both HNH and RuvC nuclease domains together with a PAM-interaction (Pi) domain (formed by C-terminal and Topo-like subdomains) assemble into a nuclease (NUC) lobe, which is connected to an α-helical recognition (REC) lobe by an arginine-rich bridge helix (BH). REC lobe is composed of REC1 and REC2 subdomains that are the least conserved regions across the Cas9 families. On the other hand, HNH as well as three RuvC subdomains (RuvC I-III) have the most conserved sequences (Hsu et al., 2014) (Figure 11 A). Apo-Cas9 is in an autoinhibited conformation because HNH interaction with the RuvC domain restricts target DNA access to the active sites (Jinek et al., 2014).

Subsequently, crystal structures were solved for the type II-A *S. pyogenes* Cas9 (SpCas9) in complex with a single-guide RNA (sgRNA; it is an artificial functional fusion of crRNA and tracrRNA) hybridized to an ssDNA or partially dsDNA target, containing a PAM sequence (Anders et al., 2014; Nishimasu et al., 2014) (Figure 11 B). The NUC and REC lobes of Cas9 are reoriented, following sgRNA binding, and form a central channel, which accommodates the target DNA. The REC lobe binds a crRNA duplexed with a tracrRNA anti-repeat sequence and target DNA strand. The 3′-terminus sequence of the tracrRNA forms hairpin structures that are anchored between the RuvC and Pi domains (Nishimasu et al., 2014).

PAM sequence is absolutely required for dsDNA targeting by Cas9 (Gasiunas et al., 2012; Sapranauskas et al., 2011) and is a starting point for the

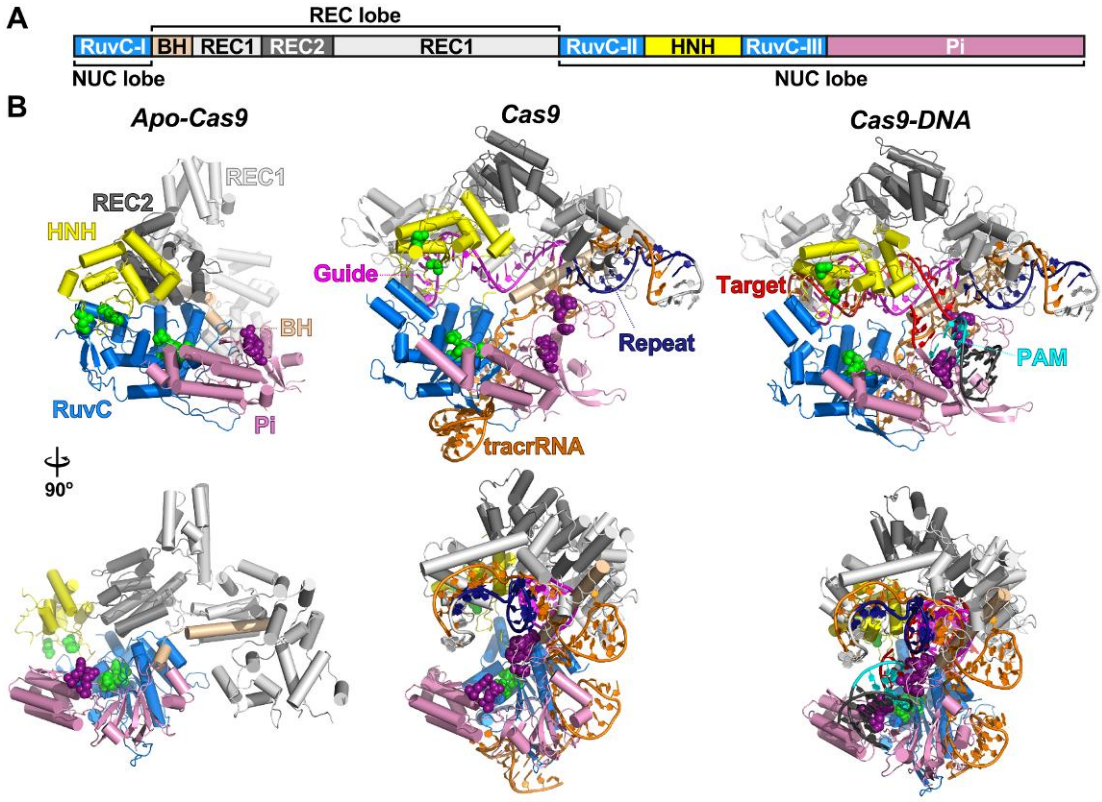

Figure 11. Cas9 is an effector complex of type II DNA interference. (**A**) Arrangement of Cas9 domains in the primary sequence. Domains, comprising nuclease (NUC) and recognition (REC) lobe, are indicated. (**B**) Crystal structure of *S. pyogenes* Cas9 protein in the apo-form (*Apo-Cas9*; PDB: 4CMQ), in complex with the crRNA-tracrRNA hairpin (*Cas9*; PDB: 4OO8), and the target DNA (*Cas9-DNA*; PDB: 4UN3). RuvC, HNH, Pi REC1, REC2 domains and bridge helix (BH) of Cas9 protein are coloured in marine, yellow, pink, grey, dark grey and wheat, respectively. The tracrRNA is orange. Repeat and guide sequences of crRNA are coloured in magenta and dark blue, while PAM and protospacer (target) of target DNA are coloured cyan and red, respectively. Residues of active sites are represented as green spheres, while residues of PAM recognition are shown as deep purple spheres. Crystal structures are represented in two orientations related by a 90° rotation.

R-loop formation (Sternberg et al., 2014). SpCas9 use two arginine residues of Pi domain to read out PAM motif (5′-NGG-3′) nucleotides from the major groove of dsDNA (Figure 11 B). Furthermore, Pi interaction with the minor groove of the PAM duplex and the phosphodiester group at the first position in the target strand of the protospacer work like a hinge, which starts strand separation immediately upstream of the PAM (Anders et al., 2014). PAM sequences considerably differ between Cas9 proteins. For example, type II-A Cas9 proteins from *S. pyogenes* CRISPR-Cas, *S. thermophilus* CRISPR1-Cas and CRISPR3-Cas systems target dsDNA with 5′-NGG-3′, 5′-NNAAGAW-3′,

and 5′-NGGNG-3′ PAMs, respectively (Deveau et al., 2008; Fonfara et al., 2014). Therefore, PAM recognition mechanism may be different between Cas9 orthologs. Interestingly, SpCas9 PAM specificity can be changed by Pi domain shuffling between similar Cas9 proteins (Nishimasu et al., 2014).

Mutagenesis of the protospacer highlighted mismatch-sensitive 10-12 bp sequence that is proximal to PAM (Cong et al., 2013; Fu et al., 2013; Hsu et al., 2013; Jinek et al., 2012; Mali et al., 2013a; Pattanayak et al., 2013). This sequence, termed seed, is important for initial hybridization with the complementary target DNA strand, leading to an R-loop formation. The main structural framework for seed sequence is the bridge helix, which is inserted in-between of a crRNA-tracrRNA-DNA three-way junction. Conserved arginine residues of the BH interact with nucleotides of the crRNA seed region, exposing nucleotide bases to the solvent, that serve as a nucleation site for the target DNA hybridization. Hybridized DNA strand is further stabilized at the corresponding seed sequence by interactions with the REC lobe (Nishimasu et al., 2014).

Unlike the *e*Cascade, Cas9 anchors only 3′-end of the crRNA; therefore, spacer region of crRNA wraps around target strand during R-loop formation, making full length RNA/DNA hybrid (Nishimasu et al., 2014). R-loop formation leads to target DNA cleavage within protospacer 3 nt upstream of the PAM by Cas9 nuclease domains, leaving blunt-ended products (Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012). In the presence of Mg^{2+} ions, RuvC domain cleaves the displaced non-target DNA strand, while HNH domain hydrolyses DNA target strand that is duplexed with crRNA. Therefore, inactivation of active site of either nuclease results in nicked DNA products (Gasiunas et al., 2012; Jinek et al., 2012). Interestingly, HNH of SpCas9 manages to cleave ssRNA that is complementary to Cas9-associated guide crRNA when PAM sequence is present as RNA-DNA heteroduplex (O'Connell et al., 2014).

It seems that the HNH and RuvC domains determine the location of their cut sites using different mechanisms. The RuvC domain employs a ruler mechanism, measuring cleavage position from the PAM, while the HNH cuts DNA at the fixed position determined by the guide-target duplex (Chen et al., 2014). Furthermore, the active sites of the nuclease domains do not co-localize with DNA cut sites in the crystal structures, thus suggesting that these domains require conformational rearrangement to reach the target (Anders et al., 2014; Nishimasu et al., 2014).

Mechanism of DNA-interference in the type II systems. In summary, genetic, biochemical, and structural studies of the CRISPR-encoded immunity in the type II CRISPR-Cas systems are consistent with a following general mechanism of DNA interference (Figure 12): i) Cas9, mature crRNA, and tracrRNA assemble into a bilobed ternary effector complex; ii) Pi domain of the Cas9 scans for the PAM sequence in dsDNA; iii) PAM serves as a priming site for strand separation and subsequent target strand hybridization to the seed sequence of the guide crRNA; iv) if the guide and the target forms stable duplex at the seed region, the hybridization progresses, forming an R-loop structure within the guide sequence and the target DNA strand; v) in the presence of Mg^{2+} ions, DNA is cleaved in both strands within the protospacer 3 nt upstream of the PAM sequence to generate blunt DNA ends. RuvC-active site cuts the displaced strand of the R-loop, while HNH-active site cleaves the crRNA-duplexed DNA strand. Taken together, data demonstrates that the Cas9-crRNA-tracrRNA complex functions as an RNA-guided endonuclease where sequence specificity is dictated by the crRNA, while Cas9 provides the cleavage machinery. This establishes a molecular basis for CRISPR-mediated immunity in type II systems, which solely rely on the signature Cas9 protein.

The simple modular organization of the Cas9 effector complex paves the way for the engineering of universal RNA-guided DNA endonucleases. While the proof of the principle for re-programmable RNA-guided endonucleases have been provided in the pioneering publications (Gasiunas et al., 2012; Jinek et al., 2012), recent studies demonstrate that Cas9 can be employed for a

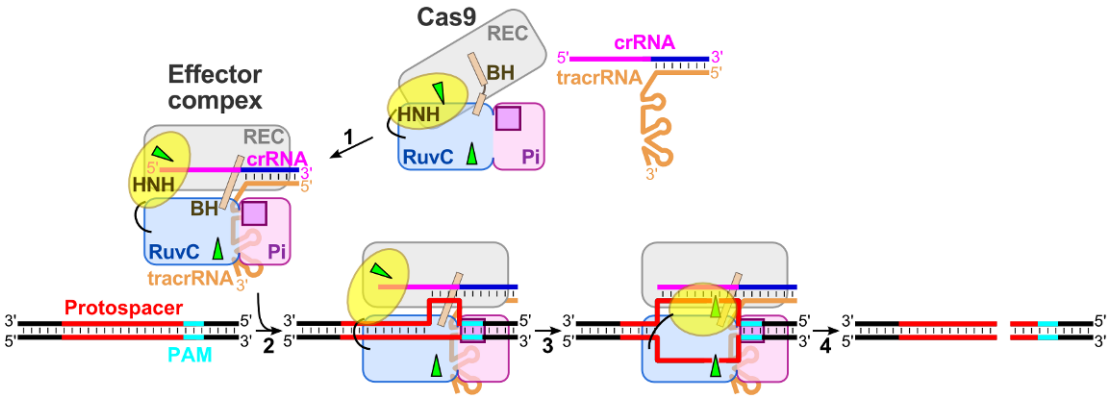

Figure 12. Mechanism of type II DNA interference. Cas9, mature crRNA and tracrRNA assemble into a bilobed ternary effector complex (1). Pi domain detects PAM sequence, which serves as a priming site for strand separation and subsequent target strand hybridization to the seed sequence of the guide crRNA (2). If guide and target forms a stabile duplex at the seed region, the hybridization progresses within the guide sequence and the target DNA strand, forming an R-loop structure, which triggers hydrolysis of the target DNA. The RuvC-active site cuts the displaced strand of the R-loop, while the HNH-active site cleaves the crRNA-duplexed DNA strand within a protospacer 3 nt upstream of the PAM sequence (3) to generate blunt DNA ends (4). Colours of effector compex components are as in (Figure 11). Green triangles represent active sites of nucleases, while PAM recognition site is shown as a purple square.

precise DNA cutting *in vitro* (Karvelis et al., 2013b) or efficient editing of the human (Cho et al., 2013; Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013b), monkey (Niu et al., 2014), mouse (Cong et al., 2013; Wang et al., 2013), plant (Li et al., 2013a), zebrafish (Chang et al., 2013; Hwang et al., 2013), yeast (DiCarlo et al., 2013), and bacteria (Jiang et al., 2013) genomes *in vivo*. Furthermore, a catalytically deficient Cas9 variant or its fusion with transcription factors was used in bacteria and human cells as a tool to specifically regulate transcription through the Cas9 binding (Gilbert et al., 2013; Qi et al., 2013). Moreover, it has been reported that in *Francisella novicida*, a Cas9 variant is involved in regulation of the bacterial gene, contributing to the virulence by triggering proinflammatory innate immune response of the eukaryotic host (Sampson et al., 2013). This finding may open the way for novel Cas9 applications. Compiled lists of applications are provided in (Gasiunas & Siksnys, 2013; Hsu et al., 2014; Sander & Joung, 2014).

1.4.3.3. Interference in type III CRISPR-Cas systems

Type III systems are widespread in archaea that typically encode multiple CRISPR-Cas loci belonging to different subtypes (Makarova et al., 2006; Makarova et al., 2011b). These systems are classified into III-A and III-B subtypes. The type III encodes a signature Cas10 protein that is incorporated into a Csm (Cas subtype of Mtube) or Cmr (Cas RAMP module) effector complex of type III-A or III-B, respectively (Makarova et al., 2011b).

Initially, two different type III systems seemed to target different types of nucleic acids. In type III-B systems of *P. furiosus* (Pf), *T. thermophilus* (Tt), and *S. solfataricus* (Ss), the Cmr complex recognized and cleaved synthetic RNA *in vitro* (Hale et al., 2009; Staals et al., 2013; Zhang et al., 2012b), whereas the type III-A system of *S. epidermidis* (Se) targeted DNA *in vivo* (Marraffini & Sontheimer, 2008). However, recently, Csm complexes from *S. thermophilus* (St) and *T. thermophilus*, homologous to Se-Csm, were identified as RNases *in vitro* (Staals et al., 2014; Tamulaitis et al., 2014).

The St-Csm and Tt-Csm, as well as Pf-Cmr and Tt-Cmr complexes, cut the target RNA within a protospacer at multiple sites with 6 nt intervals, measuring the distance from 5′-end of the crRNA (Benda et al., 2014; Hale et al., 2014; Staals et al., 2013; Staals et al., 2014; Tamulaitis et al., 2014). In contrast to the Pf-Cmr and Tt-Cmr complexes that exploit a molecular ruler mechanism for the target RNA cleavage, the Ss-Cmr complex cleaves target RNA as well as crRNA at UA dinucleotides in a sequence-specific manner (Zhang et al., 2012b). Interestingly, recent *in vivo* experiments of the III-A system from *S. epidermidis* and the III-B system from *Sulfolobus islandicus* have shown a transcription-dependent DNA targeting that relies on the direct protospacer transcription into RNA (Deng et al., 2013; Goldberg et al., 2014). Therefore, it seems that type III systems have bi-layered defense barrier, targeting intruder both at RNA and DNA levels. However, mechanistic details of type III DNA interference level remain to be excavated.

Similarly to type I Cascade, type III encodes a large multisubunit complex that associate with a mature crRNA (Hale et al., 2009; Rouillon et al., 2013; Spilman et al., 2013; Staals et al., 2013; Tamulaitis et al., 2014; Zhang et al., 2012b). This complex recognizes the complementary protospacer sequence in target RNA. Contrary to Cascade, intrinsic nuclease is responsible for target cleavage. Furthermore, differently from type I and II, targets of type III systems do not depend on PAM motifs (Hale et al., 2009; Tamulaitis et al., 2014).

Type III effector complexes. The Csm complexes are composed of six Cas proteins (Cas10, Csm2-Csm5), while Cmr complexes may be composed of six or seven Cas proteins (Cmr1, Cas10, Cmr3-Cmr6 and in some complexes Cmr7). Csm and Cmr complexes both incorporate two crRNA populations that differ by 6 nucleotides. For example, Se-Csm complex assembles with 43 and 37 nt crRNAs (Hatoum-Aslan et al., 2011), while 45 and 39 nt crRNAs are found in Pf-Cmr complex (Hale et al., 2009). Mature crRNAs are the product of Cas6 cut within a repeat sequence and subsequent 3′-end trimming by unknown nuclease (Hale et al., 2008). These molecules consist of a conservative 8 nt 5′-handle, originating from the repeat region, and a spacer region that differ by 6 nt between two crRNA populations (Hatoum-Aslan et al., 2013).

High resolution structural information for type III complexes is still lacking. However, crystal structures are solved for separate subunits of Cmr complex and Csm3 subunits of Csm. Furthermore, low resolution structures of Csm complex from *S. solfataricus* and *T. thermophilus* (Rouillon et al., 2013; Staals et al., 2014) as well as Cmr complexes from *P. furiosus* and *T. thermophilus* (Spilman et al., 2013; Staals et al., 2013) were recently determined by electron microscopy, showing similar extended helical architecture of these complexes (Figure 13). Type III complexes resemble helical assembly of type I Cascade complex, suggesting the common evolutionary origins (Heidrich & Vogel, 2013).

Cas10 and Csm4 (type III-A) or Cmr3 (type III-B) form a "bottom" of the effector complexes, which is attached to two parallel helical strings. A "backbone" string is composed of multiple Csm3 (type III-A) or Cmr4 (type III-B) subunits. A "belly" string is comprised of multiple Csm2 or Cmr3 subunits in type III-A or III-B, respectively. The strings are capped with Csm5 (type III-A) or Cmr1-Cmr6 (type III-B) subunits (Rouillon et al., 2013; Spilman et al., 2013; Staals et al., 2013) (Figure 13 A). Differently from Pf-Cmr and Tt-Cmr complexes, Ss-Cmr complex has an additional protein (Cmr7); furthermore, EM structure of SsCmr markedly deviates from other Cmr and Csm complexes (Zhang et al., 2012b).

The "bottom". Cas10 (also known as Csm1 in type III-A and Cmr2 in type III-B) is a hallmark protein of type III systems (Makarova et al., 2011b). It has a permutated N-terminal HD domain that was postulated as a nuclease; however, the deletion of this domain had no effect on nuclease activity of Cmr complex (Cocozaki et al., 2012). In addition to HD domain, Cas10 (from type III-B) has two ferredoxin-like fold (D1 and D3) and two α-helical (D2 and D4) domains (Figure 13 B). The D1 and D3 domains are arranged side-by-side, resembling a dimeric form of adenylyl cyclase. The ADP and divalent metal ions are bound between these domains. Small α-helical domains are stacked on the D3 surface (Benda et al., 2014; Cocozaki et al., 2012; Zhu & Ye, 2012).

Cas10 is classified as a large subunit of type III effector complexes (Makarova et al., 2011b). Large subunit of type I-E (Cse1) is positioned adjacent to Cas5, which binds the 5′-handle of crRNA (Jackson et al., 2014a; Mulepati et al., 2014; Zhao et al., 2014). Similarly, Cas10 interacts with Cmr3 protein (Csm4 is a homolog in type III-A), which anchors 5′-handle of crRNA (Osawa et al., 2013; Shao et al., 2013; Spilman et al., 2013). Cmr3 is comprised of N- and C-terminal ferredoxin-fold domains (NTD and CTD) with an insertion domain (I). The ferredoxin-fold domains share structural similarity with Cas6 and Cas5 proteins. The CTD and I domains make interactions with D1 domain of Cas10 (Osawa et al., 2013; Shao et al., 2013) (Figure 13 B).

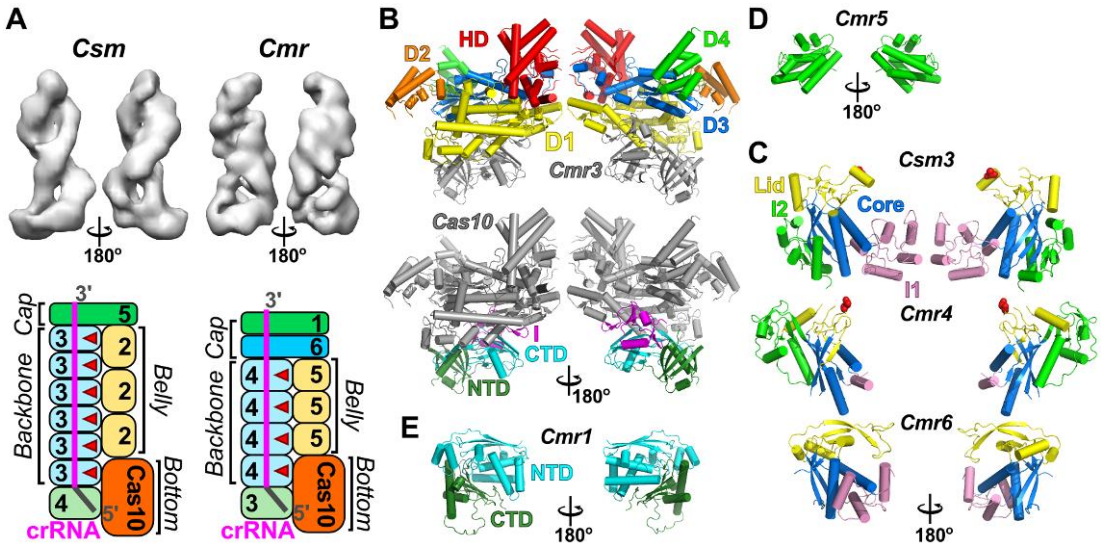

Figure 13. Structures of type III effector complexes. (**A**) Electron microscopy structures of *S. solfataricus* Csm (EMD-2418) and *T. thermophilus* Cmr (EMD-2420) complexes (upper panel), and the schematic representation of subunit organisation in the Csm and Cmr complexes (lower panel). Complexes can be dissected into four arbitrary parts: bottom (Cas10-Csm4 or Cas10-Cmr3), backbone (Csm3 or Cmr4), belly (Csm2 or Cmr5), and cap (Csm5 or Cmr1-Cmr6). Red triangle represents active site of RNases. (**B**) Structure of Cas10 and Cmr3 complex (PDB: 4W8Y and 4H4K). Different domains of these proteins are coloured differently. (**C**) Csm3 (PDB: 4N0L), Cmr4 (PDB: 4W8W) and Cmr6 (PDB: 4W8V) proteins are Cas7-like proteins. They are comprised of conserved core domain (marine), surrounded by lid (yellow), insertion-1 (I1; pink), and insertion-2 (I2; green) domains. (**D**) Cmr5 protein (PDB: 2OEB) is similar to D4 domain of Cas10 (green). (**E**) Structure of Cmr1 (PDB: 4W8Z) is similar to Cmr3. N- and C-terminal ferredoxin-fold domains (NTD and CTD) are coloured in cyan and green, respectively.

Although Cas10 does not share structural similarity with the Cse1, it occupies similar position to Cse1 in the effector complex, proposing the conservative function of the large subunits (van der Oost et al., 2014).

The" backbone" string. Six copies of Csm3 in type III-A and four copies of Cmr4 in type III-B form a helical backbone that is morphologically similar to the Cas7 backbone in the *e*Cascade complex (Benda et al., 2014; Rouillon et al., 2013; Staals et al., 2013). Moreover, Csm3 and Cmr4 are structural homologs of Cas7 and Csc2 proteins (Benda et al., 2014; Hrle et al., 2014). *Methanopyrus kandleri* Csm3 and *P. furiosus* Cmr4 are built on a conservative core of ferredoxin-fold that is surrounded by the lid, insertion-1 (I1) and insertion-2 (I2) domains (Benda et al., 2014; Hrle et al., 2013) (Figure 13 C).

These non-conservative helical domains are inserted into the ferredoxin-fold domain and markedly deviated between the Cas7-like proteins.

As mentioned above, the Cmr and Csm complexes cleave target RNA at multiple positions in 6 nt intervals. The number of cut sites correlates with the number of either Csm3 or Cmr4 subunits. Moreover, mutations of the conserved aspartates that are localized on the lid domain of both the Csm3 and Cmr4 (Figure 13 C) abolish target RNA hydrolysis (Benda et al., 2014; Tamulaitis et al., 2014). Therefore, the backbone of type III effector complexes utilises two functions: (i) anchors guide of the crRNA and (ii) cuts the guide-complementary ssRNA target.

The "belly" string. Three-four Csm2 or Cmr5 subunits assemble in parallel to the backbone of Csm or Cmr complexes, respectively. Together with a Csa5 and Cse2 from type I, the Csm2 and Cmr5 are classified as small subunits of CRISPR-Cas systems (Makarova et al., 2011b). Differently from large subunits, small subunits share structural similarity. Alpha helical structure of Cmr5 (Figure 13 D) resembles N-terminal domain of Cse2 (from *e*Cascade) while Csa5 (from *a*Cascade) is similar to C-terminal domain of Cse2 (Park et al., 2013; Reeks et al., 2013a; Sakamoto et al., 2009). Moreover, Cmr5 structure is homologous to D4 domain of Cas10 (Zhu & Ye, 2012). The Cse2 forms interactions with Cas7 but not with crRNA in the *e*Cascade. Similarly, Cmr5 makes contacts with Cmr3, however, it does not interact with crRNA (Spilman et al., 2013).

The "cap". The strings are capped with Csm5 (type III-A) or Cmr1-Cmr6 (type III-B) subunits (Rouillon et al., 2013; Spilman et al., 2013; Staals et al., 2013). The Cmr1 protein (Figure 13 E) is comprised of two tightly associated N- and C-terminal ferredoxin-fold domains (NTD and CTD) that resemble Cas6; however, the CTD occupy opposite spatial position in comparison to Cas6 (Sun et al., 2014). Moreover, the Cmr1 is structurally similar to the Cmr3 (Benda et al., 2014).

The C-terminal half of *P. furiosus* Cmr6 protein (Figure 13 D) is composed of ferredoxin-fold domain, which resembles the NTD of Cmr1. Moreover, this domain is similar to the Cmr4 (Benda et al., 2014).

Mechanism of nucleic acids interference in the type III systems. In summary, genetic, biochemical, and structural studies of the CRISPR-encoded immunity in the type III CRISPR-Cas systems are consistent with a following general mechanism of interference (Figure 14 A): i) the mature crRNA is incorporated into a Csm or Cmr effector complex, which both embody similar

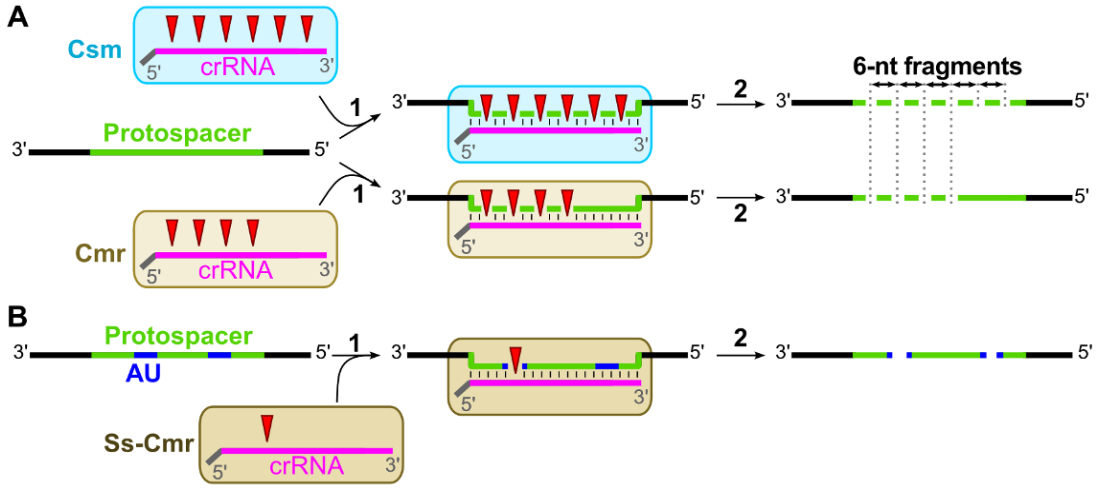

Figure 14. Mechanisms of type III interference. (**A**) Using crRNA as a guide (magenta) Csm and Cmr complexes target complementary protospacer (green) in an ssRNA (1). Complexes have multiple active sites (red triangles) that cleave bound RNA target producing fragments of 6 nt in length (2). (B) Cmr complex of *S. solfataricus* (Ss-Cmr) uses different approach to interfere with ssRNA. It binds complementary ssRNA target (1) and cleaves it in at 5′-UA-3′ sequences (blue) (2). The nuclease of this complex has to be determined.

structural organisation, resembling type I effector complex; ii) the type III effector complexes use the crRNA as a guide to find the complementary target ssRNA in a PAM-independent manner; iii) the backbone subunits of effector complexes cut the target RNA within the protospacer at multiple sites, producing ssRNA fragments of 6 nt. On the other hand, Cmr complex from *S. solfataricus* cleaves the target RNA in a sequence-dependent fashion (Figure 14 B). Moreover, Csm and Cmr complexes interfere with DNA in a transcription-dependent manner *in vivo*, although the mechanism of this process remains to be determined.

MATERIALS AND METHODS

1.5. Materials

1.5.1. Chemicals

All chemicals used in this study were of the highest purity grade; they were purchased from Sigma-Aldrich, Roth, Fluka, and Thermo Scientific.

1.5.2. Enzymes

DreamTaq, TaqI and PfuI DNA polymerases; FastAP thermosensitive alkaline phosphatase; T4 DNA ligase; T4 polynucleotide kinase (PNK); bovine serum albumin (BSA); FastDigest restriction enzymes used in this study were obtained from Thermo Scientific. All these products were used according to the manufacturer's instructions. P1 nuclease was purchased from Sigma.

1.5.3. Kits for molecular biology

"CloneJET PCR Cloning Kit", "Rapid DNA Ligation Kit", "GeneJET Gel Extraction Kit", "GeneJET PCR Purification Kit", "GeneJET Plasmid Miniprep Kit" and "T7 high yield transcription kit" were purchased from Thermo Scientific. "Cycler Reader DNA Sequencing kit" was made by Fermentas. "Malachite green assays kit" was obtained from BioAssay Systems. Kits were used according to the manufacturer's instructions.

1.5.4. Bacterial strains

E. coli strain DH5α [F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–] was used for the cloning procedures.

E. coli strain ER2267 [(F´ $proA^+B^+$ $lacI^q$ Δ(lacZ)M15 zzf::mini-Tn10 (Kan^R)/ Δ(argF-lacZ)U169 glnV44 $e14^-$($McrA^-$) rfbD1? recA1 relA1? endA1

spoT1? thi-1 Δ(mcrC-mrr)114::IS10] (New England Biolabs) was used for Cas3 protein expression.

E. coli strain BL21(DE3) (F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) was used for Cascade complex expression.

1.5.5. DNA

Genomic DNA of *Streptococcus thermophilus* DGCC7710 strain was kindly provided by DuPont (Dangé-Saint-Romain, France).

Plasmid vectors

pBAD24-CHis (Ap^r; kindly provided by dr. R. Sukackaite) expression vectors was used to obtain Cas3, Cse1 or Cas7 proteins fused with C-terminal His_6-tag sequence. Cassette of *cse1-cse2-cas7-cas5-cas6e* genes was cloned into pCDF-Duet1 (Str^r; Novagen), while homogeneous CRISPR region with 6 copies of spacer-1 was inserted into pACYC-Duet1 (Cm^r; Novagen) vector. Casettes of *cas5-cas6e* and *cse2-cas7* genes were inserted following the distinct promoters of modified version of the pCDF-Duet1 plasmid (Novagen), which contains the His-tag sequence instead of the S-tag. (Table 1).

Residues of Cas3 HD and helicase active sites were replaced by alanines using pCas3 as a template for the site-directed mutagenesis (Table 1).

Plasmid substrates

Single-stranded DNA of phage M13mp18 genome (New England Biolabs) and double-stranded supercoiled pUC57 (Thermo Scientific) was used in the ATPase and nuclease assays of Cas3. Partial duplexes of oligonucleotides, listed in Table 2, and DNA of M13mp18 were used in helicase assays of the Cas3.

Oligonucleotide duplexes, containing protospacer-1 or protospcer-3 and distinct PAM sequence, were cloned into pUC19 (Ap^r; Thermo Scientific), and resulting plasmids pSP1-NN (where N represents A, G, C or T nucleotide of the PAM sequence) or pSP3-AA (Table 1) were used as substrates in the

ATPase and nuclease assays of Cas3 (in the presence of Cascade). Moreover, these plasmids were used as the templates for production of DNA constructs for the magnetic tweezers.

1.5.6. Oligonucleotides

All oligonucleotides were purchased from Metabion. Oligonucleotide substrates used in this study are listed in Table 2. Oligonucleotides used in the nuclease, helicase or EMSA assays were 5′-end-labelled using PNK (Thermo Fisher Scientific) and [γ-^{33}P]ATP or [γ-^{32}P]ATP (Hartmann Analytic). Labelled oligonucleotides were annealed to the unlabeled complementary DNA.

Primers for 2.1 kbp PCR fragment, containing the protospacer variants, for the magnetic tweezers experiments with the Cascade are 5′-gcgtaagtctcgagaactagttccgtaagatgcttttctgtgact-3′ and 5′-gcgtaagtgcggccgcttcgttccactgagcgtcaga-3′. Fragment of 1.2 kbp for biotin and digoxigenin handles for ligation to the 2.1 kbp PCR fragments was amplified using 5′-gaccgagatagggttgagtg-3′ and 5′-tttgtgatgctcgtcagggg-3′ primers.

1.5.7. Buffers

NBE buffer: 100 mM H_3BO_3-NaOH, 15 mM sodium acetate, 2 mM EDTA (pH 8.2 at 25°C).

TAE buffer: 40 mM Tris, 20 mM acetic acid, 1 mM EDTA (pH 8.0 at 25°C).

TBE buffer: 100 mM Tris-borate (pH 8.2 at 25°C) and 2 mM EDTA

PE1 buffer: 125 mM Tris-HCl (pH 6.8 at 25°C) and 0.1% SDS (w/v).

PE2 buffer: 375 mM Tris-HCl (pH 8.8 at 25°C) and 0.1% SDS (w/v).

PE3 buffer: 25 mM Tris, 190 mM glycine (pH 8.3 at 25°C) and 0.1% SDS (w/v).

“Acrylamide” solution I: acrylamide/N,N′-methylenebisacrylamide (37.5:1 (w/w)) solution.

“Acrylamide” solution II: acrylamide/N,N′-methylenebisacrylamide (29:1 (w/w)) solution.

Protein loading solution: 100 mM Tris-HCl (pH 6.8 at 25°C), 4% SDS (w/v), 20% (v/v) glycerol, 200 mM DTT, trace of Bromphenol blue.

DNA loading solution: 25 mM EDTA, pH 9.0, 0.01% (w/v) bromphenol blue and 95% (v/v) formamide

Cas3 storage buffer: 10 mM Tris-HCl (pH 7.5), 300 mM KCl, 1 mM EDTA, 1 mM DTT, 50% (v/v) glycerol.

Cascade storage buffer: 20 mM Tris-HCl (pH 8), 500 mM NaCl and 50% (v/v) glycerol.

MS1 buffer: 0.1 M triethylammonium acetate (TEAA) (pH 7.0) (Fluka);

MS2 buffer: buffer MS1 with 25% LC MS grade acetonitrile (v/v) (Fisher).

MS3 buffer: 0.4 M 1,1,1,3,3,3,-hexafluoro-2-propanol (HFIP, Sigma-Aldrich) adjusted with triethylamine (TEA) to pH 7.0 and 0.1 mM TEAA.

MS4 buffer: buffer MS3 with 50% methanol (v/v) (Fisher).

AB1 buffer: 10 mM Tris-HCl (pH 7.5 at 25°C), 30 mM KCl, 5% (v/v) glycerol, 2 mM $MgCl_2$, 0.1 mg/ml BSA, 0.5 mM ATP.

AB2 buffer: 10 mM Tris–HCl (pH 7.5 at 25°C), 75 mM NaCl, 40 mM KCl, 7% (v/v) glycerol, 0.1 mg/ml BSA, 1.5 mM $MgCl_2$, 2 mM ATP.

TLC buffer: 0.325-M KH_2PO_4 (pH 3.5).

HB buffer: 10 mM Tris-HCl (pH 7.5), 25 mM KCl, 15% (v/v) glycerol, 1 mM $MgCl_2$, 2 mM ATP (or AMP-PNP).

“Stop” solution: 67.5 mM EDTA, 27% (v/v) glycerol, 0.3% (w/v) SDS.

Binding buffer: buffer TAE with 150 mM NaCl, 0.1 mg/ml BSA, 10% glycerol.

Footprint buffer: 10 mM Tris-HCl (pH 8), 100 mM NaCl, and 0.1 mg/ml BSA.

NB1 buffer: 10 mM Tris-HCl (pH 7.5), 60 mM KCl, 10% (v/v) glycerol, 10 mM $MgCl_2$.

NB2 buffer: 10 mM Tris–HCl (pH 7.5), 75 mM NaCl, 40 mM KCl, 7% (v/v) glycerol, 1.5 mM $MgCl_2$, 0.1 mM $NiCl_2$, 2 mM ATP.

MT buffer: 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.1 mg/ml BSA.

PCI solution: Phenol-Chloroform-Isoamyl Alcohol [25:24:1 (v/v/v)] saturated with 10 mM Tris (pH 8.0), 1 mM EDTA.

Table 1. Plasmids constructed in this study.

Plasmid	**Description**	**Primer sites**[#]	**Primer sequence (5′→3′)**[#]
pCas3	*cas3* gene was inserted into the pBAD24-CHis via *Nco*I and *Xho*I cloning sites; His_6-tag on C-terminus	*cas3*; ***Eco31*I**, fw	a**ggtctc**acatgaaacatattaatgattatttttgggc
		cas3; ***Xho*I**, rv	a**ctcgag**aaccgactcattcttatccaac
pCas3-D77A	Alanine replacement of D77 in HD domain of Cas3; pCas3 was used as template	fw	gctgttcatg**cg**atcggtaaagcaacaccagc
		rv	gctttaccgat**cg**catgaacagctcctaggaattg
pCas3-D227A	Alanine replacement of D227 in HD domain of Cas3; pCas3 was used as template	fw	ctcataatg**agc**g**ct**tggattgctagtaatgagc
		rv	ctagcaatcca**agc**g**ct**cattatgagtaaacctg
pCas3-Q290A	Alanine replacement of Q290 in helicase domain of Cas3; pCas3 was used as template	fw	gtcca**a**gg**g**aattttt**gc**gctgatactctcacaaac
		rv	gtatcagc**gc**aaaatt**c**ct**t**ggactaaatccaaatc
pCas3-K316A	Alanine replacement of K316 in helicase domain of Cas3; pCas3 was used as template	fw	gggaatcgg**cgcc**acagaggcggctctagcgg
		rv	gccgcctctgt**ggcg**ccgattcccattggcgc
pCas3-D452A	Alanine replacement of D452 in helicase domain of Cas3; pCas3 was used as template	fw	cgttattg**c**tgaagt**g**ca**t**gcttatgatgcttatatg
		rv	cataagc**a**tg**c**acttca**g**caataacgataacttttttac
pCas3-E453A	Alanine replacement of E453 in helicase domain of Cas3; pCas3 was used as template	fw	gatgc**a**gtccacgc**a**tatgatgcttatatgagcc
		rv	gcatcata**t**gcgtggact**g**catcaataacgataac
pCas3-R663A	Alanine replacement of R663 in helicase domain of Cas3; pCas3 was used as template	fw	cgtatcgg**cgcc**ctacatcgtcacaaaatcaaaagg
		rv	gtgacgatgtag**ggcg**ccgatacgttgaatgag
pCas3-R666A	Alanine replacement of R666 in helicase domain of Cas3; pCas3 was used as template	fw	ggacgact**g**cat**gc**tcacaaaatcaaaaggccc
		rv	gattttgtga**gc**atg**c**agtcgtccgatacgttg
pCascade	*cse1-cse2-cas7-cas5-cas6e* cassete was inserted into the pCDF-Duet1 via *Nco*I and *Pac*I cloning sites; no tags	*cse1*; ***Eco31*I**, fw	ccat**ggtctc**acatgagtcggtttaatttacttgatgaaccc
		cse1; ***Pac*I**, rv	**ttaattaa**tcattcctcaagtggtaccactgtcatc

See next page for Table 1 extension

Extension of the Table 1

pCascadeΔ1	*cas5-cas6e* and *cse2-cas7* gene cassettes were inserted into the modified pCDF-Duet1 via *Nco*I/*Not*I and *AatII/XhoI* cloning sites, respectively; His_6-tag on C-terminus of CasC	*cas5*; ***Eco31*I**, fw	**ggtctc**acatggtgaatgctatgaagacgatattg
		cas6e; ***Not*I**, rv	a**gcggccgc**tcattcctcaagtggtaccactg
		cse2; ***Aat*II**, fw	t**gacgtc**tatgtcacaacatactaaacaaacagtttg
		cas7; ***Xho*I**, rv	**ctcgag**aatatattgggctatcatttctgaaaaatc
pCse1	*cse1* gene was inserted into the pBAD24-CHis via *Nco*I and *Xho*I cloning sites; His_6-tag on C-terminus	*cse1*; ***Eco31*I**, fw	**ccatgg**tctcacatgagtcggtttaatttacttgatgaaccc
		cse1; ***Xho*I**, rv	**ctcgag**ttttaactgttgacgaagccaaaaatcg
pCas7	*cas7* gene was inserted into the pBAD24-CHis via *Nco*I and *Xho*I cloning sites; His_6-tag on C-terminus	*cas7*; ***Eco31*I**, fw	**ggtctcac**atgacaactgaacaacgattatttttag
		cas7; ***Xho*I**, rv	**ctcgag**aatatattgggctatcatttctgaaaaatc
pCRh	Homogeneous CRISPR was assembled from oligos and inserted into the pACYC-Duet1 via *Nco*I and *Pac*I cloning sites; no tags	5′-end; ***Nco*I**, fw	tcgag**ccatgg**ttccattgggatcttttagtgtttttcc
		5′-end; ***Nco*I**, rv	cccgcgtgtgcgggaaaaacactaaaagatcccaatggaa**ccatgg**c
		Center, fw	cgcacacgcgggggtgatcctatacctatatcaatggcctcccacgcataagcgtttttcc
		Center, rv	cccgcgtgtgcgggaaaaacgcttatgcgtgggaggccattgatataggtataggatcacc
		3′-end; ***Pac*I**, fw	cgcacacgcgggggtgattctacaaatttaggtcatat**ttaattaa**c
		3′-end; ***Pac*I**, rv	catgg**ttaattaa**atatgacctaaatttgtagaatcacc
pSP(X)-NN or pSP(X)-ΔY*	Oligoduplexes with protospacer and distinct PAM sequence were cloned into pUC19 via *Sma*I.	See table 2	See table 2

Sequence complementary to a gene is underlined; restriction nuclease recognition site is in bold; fw – forward; rv – reverse.

* (X) – protospace-1 or protospacer-3; N – A, T, G or C of the PAM; Y – number of deleted nucleotides from the PAM-distal end of the protospacer-1.

Table 2. Oligonucleotide substrates used in this study.

Subsrate	Sequence	Comment
S1-AA or SP1-TAA	5′-GACCACCCTTTTTGATA***TAA***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTAT***ATT***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***TAA PAM***, *
SP3-AA	5′-GACCACCCTTTTTGATAT***AA***CGCAACCCCTCCTTAGACATGGGAACAGTACTAGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***TT***GCGTTGGGGAGGAATCTGTACCCTTGTCATGATCGTCTATGCAAGACTCCCTT-5′	Protospacer-3, ***AA PAM***, *
SP1-CC	5′-GACCACCCTTTTTGATAT***CC***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***GG***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***CC PAM***, *
SP1-AG	5′-GACCACCCTTTTTGATAT***AG***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***TC***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***AG PAM***, *
SP1-AC	5′-GACCACCCTTTTTGATAT***AC***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***TG***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***AC PAM***, *
SP1-AT	5′-GACCACCCTTTTTGATAT***AT***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***TA***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***AT PAM***, *
SP1-GA	5′-GACCACCCTTTTTGATAT***GA***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***CT***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***GA PAM***, *
SP1-GG	5′-GACCACCCTTTTTGATAT***GG***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***CC***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***GG PAM***, *
SP1-GC	5′-GACCACCCTTTTTGATAT***GC***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***CG***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***GC PAM***, *
SP1-GT	5′-GACCACCCTTTTTGATAT***GT***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***CA***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***GT PAM***, *
SP1-CA	5′-GACCACCCTTTTTGATAT***CA***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***GT***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***CA PAM***, *

See next page for Table 2 extension

Extension of the Table 2

SP1-CG	5′-GACCACCCTTTTTGATAT***CG***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***GC***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***CG PAM***,*
SP1-GC	5′-GACCACCCTTTTTGATAT***GC***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***CG***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***GC PAM***,*
SP1-GT	5′-GACCACCCTTTTTGATAT***GT***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***CA***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***GT PAM***,*
SP1-CA	5′-GACCACCCTTTTTGATAT***CA***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***GT***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***CA PAM***,*
SP1-CG	5′-GACCACCCTTTTTGATAT***CG***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***GC***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***CG PAM***,*
SP1-CT	5′-GACCACCCTTTTTGATAT***CT***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***AG***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***CT PAM***,*
SP1-TA	5′-GACCACCCTTTTTGATAT***TA***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***AT***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***TA PAM***,*
SP1-TG	5′-GACCACCCTTTTTGATAT***TG***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***AC***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***TG PAM***,*
SP1-TC	5′-GACCACCCTTTTTGATAT***TC***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***AG***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***TC PAM***,*
SP1-TT	5′-GACCACCCTTTTTGATAT***TT***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***AA***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***TT PAM***,*
SP1-AAA	5′-GACCACCCTTTTTGATA***AAA***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTAT***TTT***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***AAA PAM***,*
SP1-GAA	5′-GACCACCCTTTTTGATA***GAA***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTAT***CTT***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***GAA PAM***,*

See next page for Table 2 extension

Extension of the Table 2

SP1-CAA	5′-GACCACCCTTTTTGATA***CAA***TATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTAT***GTT***ATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***CAA PAM***, *
SP1-Δ2	5′-GACCACCCTTTTTGATAT***AA***TATACCTATATCAATGGCCTCCCACGCATAACGGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***TT***ATATGGATATAGTTACCGGAGGGTGCGTATTGCCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***AA PAM,*** Δ2 nt#, *
SP1-Δ4	5′-GACCACCCTTTTTGATAT***AA***TATACCTATATCAATGGCCTCCCACGCATTTCGGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***TT***ATATGGATATAGTTACCGGAGGGTGCGTAAAGCCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***AA PAM***, Δ4 nt#, *
SP1-Δ6	5′-GACCACCCTTTTTGATAT***AA***TATACCTATATCAATGGCCTCCCACGCTATTCGGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***TT***ATATGGATATAGTTACCGGAGGGTGCGATAAGCCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***AA PAM***, Δ6 nt#, *
SP1-Δ8	5′-GACCACCCTTTTTGATAT***AA***TATACCTATATCAATGGCCTCCCACCGTATTCGGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***TT***ATATGGATATAGTTACCGGAGGGTGGCATAAGCCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***AA PAM***, Δ8 nt#, *
SP1-Δ10	5′-GACCACCCTTTTTGATAT***AA***TATACCTATATCAATGGCCTCCCGCAGATACGTTCTGAGGGAA-3′ 3′-CTGGTGGGAAAAACTATA***TT***ATATGGATATAGTTACCGGAGGGCGTCTATGCAAGACTCCCTT-5′	Protospacer-1, ***AA PAM***, Δ10 nt#, *
H1	5′-CCTGCAGGTCGACTCTAGAG-3′	Complementary for M13mp18
H2	5′-CAUGCCUGCAGGUCGACUCUAG-3′	Complementary for M13mp18
H3	5′-GCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTT-3′	Complementary for M13mp18
H4	5′-CTTTTCACCAGCGAG-3′	Complementary for M13mp18
H5	5′-GGGGGGGGGGTAGTTGAGAA-3′	Complementary for 3’-end of H7
H6	5′-CCCGCGCGCGTCGTCATGCG-3′	Complementary for 5’-end of H7
H7	5′-GGGCGCGCGCAGCAGTACGCTAGTACTGTTCCCATGTCTAAGGAGGGGTTGCGTTCTCAACTACCCCCCCCCC-3′	Complementary for H5 and H6

* ALQ13.2 phage surounding sequences; upper sequence is non-target strand of protospacer while lower sequence is target strand of protospacer.

Guide-complementary sequences of protospacer-1 were truncated from a PAM-distal end.

1.6. Methods

1.6.1. Electrophoresis

1.6.1.1. Denaturing (SDS) polyacrylamide gel electrophoresis

Denaturing SDS-PAGE of proteins was employed to verify the homogeneity of Cas3 and Cascade protein preparations. Protein samples were mixed at 1:1 (v/v) ratios with the protein loading solution and denatured at 95°C for 5 min. Polyacrylamide gels comprised of stacking and separation gel layers were used for protein electrophoresis (Sambrook, 1989). The stacking gel consisted of 4% “acrylamide” solution I in PE1 buffer, while separating gel was 12-15% “acrylamide” solution I in PE2 buffer. Electrophoresis was carried out in PE3 buffer at room temperature for 1-1.5 hours at 25 V/cm. Gels were stained with Page Blue protein staining solution (Thermo Scientific).

1.6.1.2. Non-denaturing agarose gel electrophoresis

Separations of PCR reaction products or plasmid DNA was performed in 0.8% agarose gels in the NBE buffer supplemented with 0.5 g/ml ethidium bromide. DNA samples were mixed with 1/3 volume of “Stop” solution and electrophoresed at 3 V/cm for ~50 min. Separated DNA was visualized with UV light and digital images of the gel were taken.

DNA fragments required for genetic engineering procedures were separated in 1-1.5% agarose gels in the ethidium bromide-free TAE buffer. The gel slices containing required DNA fragments were excised according to the ethidium bromide stained markers. DNA was recovered using the “GeneJET Gel Extraction Kit” according to the manufacturer’s recommendations.

1.6.1.3. Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel electrophoresis was employed in the Cascade binding assay. The gels consisted of 8% “acrylamide” solution II in the TAE buffer; polymerisation was initiated by adding TEMED and ammonium persulphate. The gels were 1 mm thick and ~20 cm length. Prior to gel casting, one of the glass plates was processed with “bind silane” (3-methacryloxypropyltrimethoxysilane) and the other with “repeal silane” (5-7% (v/v) dichlorodimethylsilane in $CHCl_3$). Electrophoresis was run at room temperature for 2-3 hours at ~6 V/cm.

After electrophoresis the glass plate with “repeal silane” was removed and the gel was dried on the glass plate with “bind silane” under a hot air flow. Radiolabeled DNA was detected in the dried gels using BAS-MS image plates (FujiFilm) and FLA-5100 phosphorimager (Fujilm). The amounts of various DNA fragments were quantified with OptiQuant 3.0 software (Perkin Elmer).

1.6.1.4. Denaturing (urea) polyacrylamide gel electrophoresis

Denaturing PAGE was employed in Cas3 nuclease assays as well as in Cascade footprint assay. Gels consisted of 20% “acrylamide” solution II and 7 M urea in the TBE buffer. Prior electrophoresis, samples of radiolabeled DNA were mixed 1:1 (v/v) with DNA loading solution and denatured at 95 °C for 5 min and rapidly chilled in the -20 °C freezer. Glass plates were prepared as in section (2.2.1.3.). Electrophoresis was run for 15 min at 30 V/cm without the samples and for another 2-3 hours with the DNA samples. Gels were dried and analysed as described above (see section 2.2.1.3.).

Denaturing PAGE was also employed for the analysis of crRNA extracts. In this case, “silane”-untreated glass plates were used. The gels consisted of 15% “acrylamide” solution II and 7 M urea in the TBE buffer. Prior electrophoresis, samples of crRNA were mixed 1:1 (v/v) with RNA loading

dye (Thermo Scientific) and denatured at 70 °C for 10 min and rapidly chilled at +4 °C. Electrophoresis was run for 30 min at 30 V/cm. Gel was stained with CybrGold (Invitrogen). The crRNA was visualized at UV light and digital images of the gel were taken.

1.6.2. DNA manipulations

1.6.2.1. Techniques for recombinant DNA isolation

Recombinant plasmids (Table 1) were constructed using standard cloning procedures (Sambrook, 1989). Plasmids were isolated using "GeneJET Plasmid Miniprep Kit". The "GeneJET Gel extraction Kit" was used for isolation of DNA fragments from agarose gels. All enzymes and corresponding buffers used for DNA manipulations were purchased from Thermo Scientific.

1.6.2.2. Cloning of Cas3 and Cascade expression vectors

S. thermophilus DGCC7710 genomic DNA was used as a template in PCR reactions. The *cas3* and *cse1* were cloned into the pBAD24-CHis expression vector via *Nco*I and *Xho*I sites to generate pCas3 and pCse1 plasmids, respectively, which were used for the overexpression of the C-terminal $(His)_6$-tagged Cas3 and Cse1 protein variants (Table 2).

Components of Cascade complex were expressed from three vectors: pCascade, pCas7 and pCRh. The *cse1-cse2-cas7-cas5-cas6e* gene cassette was cloned into pCDF-Duet1 via *Nco*I/*Pac*I sites, while *cas7* gene was inserted into *Nco*I/*Xho*I sites of pBAD24-CHis expression vector. A CRISPR locus containing six copies of the repeat-spacer-1 unit (6 × SP1) of the WT *S. thermophilus* CRISPR4 system was assembled from oligonucleotides and cloned into pACYC-Duet1 vector (Table 2).

To obtain the Cascade that lacks the Cas7 subunit (CascadeΔ1), *cas5-cas6e* and *cse2-cas7* gene cassettes were cloned into the modified version of the pCDF-Duet1 plasmid (Novagen). The *cas5-cas6e* was inserted between *Nco*I

and *Not*I sites, while *cse2-cas7* was cloned via *Aat*II and *Xco*I sites, fusing the 3′-end of *cas7* gene with His_6-tag coding sequence (Table 2).

Full sequencing of cloned DNA fragments confirmed perfect matches to the original sequences.

1.6.2.3. Construction of the plasmid substrates

Oligoduplexes (Table 2) carrying single protospacer and PAM sequence were assembled by annealing complementary oligodeoxiribonucleotides. The oligoduplexes were phosphorylated using PNK and cloned into pUC19 via *Sma*I site. Plasmids with single unidirectional insertions were selected by sequencing.

1.6.2.4. Construction of substrates for magnetic tweezers

All DNA constructs were based on pUC19 plasmids into which single protospacer/PAM elements were inserted (Tables 2). For preparing constructs for the tweezers experiments, a ~2.1 kbp fragment containing a single protospacer/PAM combination was made by PCR from the recombinant plasmids (see section 2.1.6. for primer sequences), digested with *Not*I and *Spe*I and purified. Biotin- or digoxigenin-modified attachment handles were made using 1.2 kbp DNA fragments that were labelled with biotin- or digoxigenin-dUTP by PCR (see section 2.1.6. for primer sequences), and which were digested with either *Not*I or *Spe*I. The protospacer fragment was ligated with the biotin/dig-labelled handles using T4 DNA ligase. The full ligation product was purified from agarose gels preventing any exposure to EtBr and UV light.

1.6.2.5. Cas3 mutagenesis

Conserved amino acids of HD hydrolase and DExD/H domains were identified by multiple alignment of Cas3 homologues using the COBALT tool (Papadopoulos & Agarwala, 2007). The mutants D77A, D227A, Q290A, K316A, D452A, E453A, R663A and R66A (Table 1) were obtained by the site-directed mutagenesis as previously described (Tamulaitis et al., 2007).

Sequencing of the entire gene for each mutant confirmed that only the designed mutation had been introduced.

1.6.3. Expression and purification

1.6.3.1. Expression and purification of Cas3

E. coli ER2267 strain was transformed with pCas3 vector. Cells were grown in LB broth (BD) supplemented with ampicillin (100 μg/ml) and kanamycin (25 μg/ml) at 37°C to OD_{600} of ~0.5, the growth temperature was decreased to 16°C and Cas3 expression induced with 0.2% (w/v) arabinose for 20 h. Harvested cells were disrupted by sonication and cell debris were removed by centrifugation. The supernatant was loaded onto the Ni^{2+}-charged 5 ml HiTrap chelating HP column (GE Healthcare) and eluted with a linear gradient of increasing imidazole. The fractions containing Cas3 were pooled and subsequently loaded onto heparin column eluting protein with linear gradient of increasing NaCl concentration. The fractions containing Cas3 were pooled and dialysed against Cas3 storage buffer and stored at -20°C. The homogeneity of protein preparations was estimated by SDS-PAGE. Concentrations of Cas3 and its mutants were determined by measuring absorbance at 280 nm using an extinction coefficient of 132 700/M/cm (Gill & von Hippel, 1989).

1.6.3.2. Expression and purification of Cascade complex

E. coli BL21 (DE3) strain was co-transformed with pCascade, pCas7 and pCRh vectors. Cells were grown in in LB broth supplemented with ampicillin (25 μg/ml), chloramphenicol (17 μg/ml), and streptomycin (25 μg/ml) at 37°C to OD_{600} of ~0.5 and expression of Cascade complex was induced with 0.2% (w/v) arabinose and 1 mM IPTG for 3 h. Harvested cells were disrupted by sonication and cell debris removed by centrifugation. The Cascade complex was first purified on the Ni^{2+}-charged HiTrap column (GE Healthcare) followed by Superdex 200 (HiLoad 16/60; GE Healthcare) and heparin (GE Healthcare) chromatography steps. The Cascade complex was stored at -20°C

in Cascade storage buffer. The Cascade complex was subsequently analysed by SDS–PAGE and the sequence of the Cse1, Cse2, Cas7, Cas5, and Cas6e proteins was further confirmed by mass spectrometry of tryptic digests. Cascade complex concentration was estimated by Bradford assay (Thermo Scientific) using bovine serum albumin (BSA) as a reference protein. Conversion to molar concentration was performed assuming that the Cascade stoichiometry $Cse1_1$:$Cse2_2$:$Cas7_6$:$Cas5_1$:$Cas6e_1$:$crRNA_1$ is analogous to that of the *E. coli* Cascade (Jore et al., 2011).

1.6.3.3. Expression and purification of CascadeΔA complex and Cse1 protein

To obtain Cascade lacking the Cse1 subunit (CascadeΔ1), *E. coli* BL21 (DE3) cells were co-transformed with the pCascadeΔ1 and pCRh plasmids. Cells were grown in LB broth supplemented with chloramphenicol (17 μg/ml), and streptomycin (25 μg/ml) at 37°C to OD_{600} of ~ 0.5. The CascadeΔA complex expression was induced with 1 mM IPTG for 3 h. Harvested cells were disrupted by sonication and cell debris removed by centrifugation. The CascadeΔ1 complex was purified as described in section (2.2.3.2.).

To obtain the Cse1 protein, *E. coli* BL21 (DE3) cells transformed with pCse1 were grown in LB broth supplemented with ampicillin (50 μg/ml) at 37°C to OD_{600} of ~ 0.7 and Cse1 expression induced with 0.2% (w/v) arabinose for 3 h. Harvested cells were disrupted by sonication and cell debris were removed by centrifugation. The supernatant was loaded onto the Ni^{2+}-charged HiTrap chelating HP column (GE Healthcare) and eluted with a linear gradient of increasing imidazole. The fractions containing Cse1 were pooled and subsequently loaded onto Superdex 200 (HiLoad 16/60; GE Healthcare) column. Both the CascadeΔ1 complex and the Cse1 protein were stored at -20°C in Cascade storage buffer.

1.6.4. Analysis of crRNA

1.6.4.1. Extraction of crRNA

Nucleic acids co-purified with Cascade were isolated by phenol:chloroform:isoamylalcohol (PCI) (25:24:1, v/v/v) extraction. Purified nucleic acids were incubated with DNase I (Thermo Scientific), supplemented with 2.5 mM $MgCl_2$ or RNase A/T1 (Thermo Scientific), for 30 min at 37 °C. Nucleic acids were separated on a denaturing 15% polyacrylamide gel and visualized by SybrGold (Invitrogen) staining.

1.6.4.2. HPLC purification of crRNA

All samples were analysed by ion-pair reversed-phased-HPLC on an Agilent 1100 HPLC with UV260nm detector (Agilent) using a DNAsep column 50 mm × 4.6 mm I.D. (Transgenomic, San Jose, CA). The chromatographic analysis was performed using MS1 and MS2 buffers. The crRNA was obtained by injecting purified Cascade complex at 75°C using a linear gradient starting at 15% buffer MS1 and extending to 60% MS1 in 12.5 min, followed by a linear extension to 100% MS1 over 2 min at a flow rate of 1.0 ml/min. Analysis of the 3′ terminus was performed by incubating the HPLC-purified crRNA in a final concentration of 0.1 M HCl at 4°C for 1 h. The samples were concentrated to 10–20 µl on a vacuum concentrator (Eppendorf) prior to ESI-MS analysis.

1.6.4.3. ESI-MS analysis of crRNA

ESI-MS was performed in negative mode using an UHR TOF mass spectrometer (maXis) (Bruker Daltonics), coupled to an online capillary liquid chromatography system (Ultimate 3000, Dionex, UK). RNA separations were performed using a monolithic (PS-DVB) capillary column (50 mm × 0.2 mm I.D., Dionex). The chromatography was performed using the MS3 and MS4 buffers. RNA analysis was performed at 50°C with 20% buffer of MS4 buffer, extending to 40% MS4 in 5 min followed by a linear extension to 60% MS4 over 8 min at a flow rate of 2 µl/min; 250 ng crRNA was digested with 1 U

RNase A and RNaseT1 (Applied Biosystems). The reaction was incubated at 37 °C for 4 h. The oligoribonucleotide mixture was separated on a PepMap C-18 RP capillary column (150 mm × 0.3 μm I.D., Dionex) at 50°C using gradient conditions starting at 20% buffer MS3 and extending to 35% MS4 in 3 min, followed by a linear extension to 60% MS4 over 40 min at a flow rate of 2 μl/min. The mass spectrometer was set to select a mass range of 250–2000, m/z and the capillary voltage was kept at −3650 V. Oligoribonucleotides with −2 to −4 charge states were selected for tandem mass spectrometry using collision induced dissociation.

1.6.5. ATPase assays

1.6.5.1. ATPase assays of stand-alone Cas3.

ATPase reactions were conducted at 30°C in an AB1 buffer containing 3 nM of ssDNA (M13mp18) or dsDNA (supercoiled form of pUC57 plasmid) and 250 nM of Cas3 or its mutant variants. In the radioactivity assay, reaction mixtures were supplemented with [α^{32}P]ATP (5 Ci/mmol) (Hartmann Analytic). An aliquot (1 μl) was spotted onto a polyethyleneimine-cellulose thin-layer plate (Merck) then ATP was separated from resulting ADP product by chromatography in a TLC buffer and visualized using a FLA-5100 phosphorimager (Fujilm).

Malachite green assay kit (BioAssay Systems) was used to measure ATP hydrolysis through the detection of liberated-free phosphate. Reaction mixtures, in the described above buffer, contained varying amounts of ssDNA (circular form of M13mp18), dsDNA (supercoiled form of pUC57 plasmid), or 2223 nt RNA (obtained by transcription of the control template using a “T7 high yield transcription kit”) together with 250 nM of Cas3 or its mutant variants. The reactions were initiated by adding the enzyme to a mixture of the other reaction components. Aliquots were removed at fixed time intervals and the reaction was stopped by adding EDTA to a final concentration of 27 mM. The relationship between the absorbance and phosphate concentration was

established by using KH_2PO_4 as a standard and used to calculate phosphate concentrations in the ATPase assays. ATPase rates are quoted as the mean of three independent experiments.

1.6.5.2. ATPase assay of Cas3 in the presence of Cascade.

ATPase reactions were conducted at 37°C in the AB2 buffer containing 3 nM supercoiled double-stranded plasmid (Table 1), 12 nM of Cascade complex, and 300 nM of Cas3 or D452A mutant. Reactions were initiated by adding $MgCl_2$ and ATP to a mixture of the other reaction components. Reactions were stopped and rate of ATP hydroysis was calculated as described above.

1.6.6. Helicase assay

Partial DNA or RNA-DNA duplexes (Table 2) were used in the helicase assay. The 5′-ends of oligodeoxyribonucleotide 20-mer H1 and oligoribonucleotide 22-mer H2 were radiolabelled using PNK (Thermo Scientific) and [γ^{32}P]ATP (Hartmann Analytic). Partial duplexes were assembled by mixing M13mp18 ssDNA and labelled complementary oligonucleotide at 1.5:1 molar ratio followed by annealing in 10 mM Tris-HCl (pH 7.5) buffer. The helicase reactions were performed at 30°C for 60 min in an HB buffer supplemented with 0.5 nM of substrates and indicated amounts of Cas3 or its mutants. Reactions were initiated by addition of Cas3 and terminated by addition of 1/3 of “Stop” solution. The products were separated by electrophoresis through 8% (w/v) non-denaturing polyacrylamide gel, and visualized using a FLA-5100 phosphorimager (Fujilm).

To determine Cas3 polarity, two types of substrates were used. The M13mp18-based substrate was constructed by phosphorylating the 5′-ends of H3 and H4 oligonucleotides with [γ^{32}P]ATP or ATP, respectively (or vice versa), followed by annealing to the complementary sequences of M13mp18 and ligation by T4 ligase (Thermo Scientific). The resulting 60 bp partial

duplexes were cut with *Ehe*I to produce linear ssDNA molecules with short duplex regions at both ends.

For oligonucleotide-based helicase substrates, oligonucleotide 20-mers H5 or H6 were labelled at the 5′-end using PNK and [γ^{32}P]ATP, and then annealed to the complementary sequence at the 5′- or 3′-end of the 73-mer H7 to generate partial oligoduplexes, containing 53 nt 3′- or 5′-overhangs. Reactions were performed as described above except that 500 nM of Cas3, 1.5 nM of the substrate and 250 nM of trap DNA (unlabelled oligonucleotide) were used in the reaction.

1.6.7. Cascade binding assay

Synthetic oligoduplexes (Table 2) 73 bp in length were used in Cascade binding experiments. Each oligoduplex contained a 33-bp protospacer sequence corresponding to the first spacer (protospacer-1) of the *S. thermophilus* CRISPR4 locus and various PAM sequences. In control experiments, 73 bp oligoduplex containing a protospacer-3 instead of protospacer-1 was used. An oligodeoxynucleotide corresponding to the target strand was labelled at the 5′-end using PNK (Thermo Scientific) and [γ^{33}P]ATP (Hartmann Analytic) and an oligoduplex assembled by mixing the labelled target and unlabelled non-target strands at a molar ratio of 1:1.5, followed by annealing in 2 mM Tris–HCl buffer (pH 8). Increasing amounts of Cascade were incubated with 0.1 nM of radioactively labelled oligoduplex in the binding buffer for 20 min at 37°C. The samples were subjected to electrophoresis in 8% (w/v) polyacrylamide gel, and visualized using the FLA-5100 phosphorimager (Fujilm). The K_d values for Cascade-DNA complexes were calculated as previously described (Tamulaitis et al., 2006). K_d values represent the average value of three independent experiments.

1.6.8. Cascade footprinting

For probing with P1 nuclease, oligoduplexes SP1-AA, SP1-AG, SP1-CC, and SP3-AA (Table 2) were ^{33}P-5′-end-labelled at either the target or

non-target strand. Labelled oligoduplex at 2 nM concentration was incubated with or without 10 nM of Cascade complex at 37°C for 15 min in 20 µl of footprint buffer. Then, 0.02 U of P1 nuclease (Sigma) in 20 µl of 30 mM sodium acetate buffer (pH 5.3) was added and incubated at 37°C for 10 min. The reactions were stopped by addition of phenol–chloroform followed by sodium acetate/isopropanol precipitation. The cleavage products were separated on a denaturing 20% polyacrylamide gel and visualized by autoradiography. Products of dideoxy sequencing reactions ("Cycler Reader DNA Sequencing kit") of SP1-AA strands were used as size markers.

1.6.9. Nuclease assays

1.6.9.1. Nuclease assay of stand-alone Cas3

Single-stranded M13mp18 DNA and double-stranded supercoiled form of pUC57 plasmid were used as substrates in the nuclease assay. Nucleic acid cleavage reactions were performed at 37°C for 120 min in an NB1 buffer containing 4 nM ssDNA of M13mp18 or dsDNA supercoiled form of pUC57 plasmid, and 500 nM Cas3 or mutant proteins. Reactions were initiated by addition of protein and stopped by addition of phenol-chloroform, followed by chloroform extraction. Aqueous fraction was mixed with 1/3 of "Stop" solution. The products were separated by electrophoresis through 0.8% (w/v) agarose gels and visualized by the ethidium bromide staining.

1.6.9.2. Nuclease assay of Cas3 in the presence of Cascade

Supercoiled or linearized pUC19 or its derivative plasmids (Table 1) were used as substrates in the DNA cleavage assay. Cleavage reactions were performed at 37°C for indicated time intervals in the NB2 buffer. Supercoiled or linearized plasmid DNA at 5 nM concentration was incubated with 20 nM of Cascade complex and 100 nM of Cas3 or its mutants unless otherwise stated. Reactions were initiated by addition of Cas3 and stopped by mixing

with 1/3 of “Stop” solution. Reaction products were analysed by 0.8% (w/v) agarose gels electrophoresis and visualized by ethidium bromide staining.

To monitor oligoduplex (Table 2) cleavage, either the target or non-target strands were ^{33}P-5′-end-labelled and 2 nM of labelled oligoduplex was incubated with 4 nM of Cascade complex and 100 or 500 nM of Cas3 in the presence or absence of ATP, respectively. The cleavage products were separated on a denaturing 20% polyacrylamide gel and visualized by autoradiography. Products of dideoxy sequencing reactions (“Cycler Reader DNA Sequencing kit”) of SP1-AA strands were used as size markers.

1.6.10. Single-Molecule Experiments

1.6.10.1. Single-molecule observation of R-loop formation

Single-molecule experiments with Cascade were carried out as previously described (Revyakin et al., 2005; Seidel et al., 2005) using a home-built magnetic tweezers (Klaue & Seidel, 2009) instrument (equipped with a Pulnix 1067CL CCD camera, image acquisition at 120 Hz). The fluidic cells for the experiments were constructed from a polystyrene-coated and an uncoated 24x60 mm coverslip (Menzel-Gläser No. 1) and a Parafilm spacer. Anti-digoxigenin (Roche) and BSA were adsorbed directly to the glass by incubation for >3 hours at room temperature. Each DNA construct was bound at its biotin-modified end to excess streptavidin-coated magnetic beads (1 μm diameter, MyOne, Invitogen) and added into the fluidic cell to allow the DNA to bind the surface via its digoxigenin-modified end. Non-magnetic particles (3.2 μm tosylactivated polystyrene or 2.0 μm aldehyde/sulphate latex, Invitrogen) were adhered to the glass (in phosphate buffered saline or 50 mM MES, pH 5.5, respectively) to correct for the instrument drift. The three-dimensional position of the magnetic bead and thus the orientation and length of the attached DNA molecule was determined from video images at the camera frame rate (see above) using real-time 3D particle tracking with sub-nm accuracy (Klaue & Seidel, 2009; Lionnet et al., 2012; Otto et al., 2010).

Suitable topologically-constrained DNA was identified from rotations curves and the rotational zero reference set as determined from a rotation curve at 0.3 pN. Experiments with Cascade were carried out in MT buffer. Measurements were performed using 9 nM Cascade (at room temperature), unless otherwise noted. When recording the shift in rotational zero due to R-loop formation, magnets were rotated at 0.5 – 1.0 Hz. For measuring the on/off times as a function of torque, the magnets were turned at 1 Hz. In all time trajectories and rotation curves depicted, raw DNA length data taken at the camera acquisition rate is shown in light colours (light grey, blue or red), while data smoothed with either a 1 or 2 Hz moving average is shown as in dark colours (dark, blue, green, grey and red).

1.6.10.2. Determination of rotational shifts upon R-loop binding

At the typically applied concentration of Cascade of 9 nM, rotational shifts due to the formation of stable R-loops could only be determined from the shift of the right (positive) side of the rotation curve. This is due to the practically instantaneous R-loop formation at low negative twist, such that an "R-loop free" left side of the rotation curve could not be obtained. Also, comparing the left side of a rotation curve with an R-loop to the left side of a rotation curve taken in absence of Cascade is not a reliable measure of the shift due to the observed DNA destabilization by Cascade, which causes a shift in the left side even in the absence of an R-loop. For R-loop formation, the shift of the right side of the rotation curve is determined by fitting a straight line to the linear part of the rotation curve at positive turns before and after R-loop induction and determining the rotational shift of the midpoint of the fit ($-N_{loop}$). For R-loop dissociation, the rotational shift is determined from the magnitude of the sudden DNA length increase upon R-loop dissociation. The length increase divided by the slope of the rotation curve at the corresponding force provides then the rotational shift ($+N_{loop}$). The rotational shifts both upon R-loop

formation and dissociation are in agreement within error. In addition, the dissociation data also provides the rotational shift of the substep.

Using the shift of the right side of the rotation curve only assumes that the curves in the absence and presence of an R-loop exhibit an identical shape. To test this assumption, but also to obtain a method to extract rotational shifts for the DNA substrates with unstable R-loops, we carried out experiments at decreased Cascade concentrations. Under such conditions, R-loop formation becomes limited by Cascade binding and full rotation curves with and without R-loop formation events can be obtained in one experiment. Curves with and without R-loops were grouped and for each type an average rotation curve was obtained. The rotational shift of the curve centre is determined by fitting the peak of each average curve with a parabolic function and calculating the shift of the maximum position of the fit. The curve in the presence of an R-loop appears to be slightly broader and has a slightly lower maximum then the curve in absence of the R-loop. Thus, the shift of the right side of the rotation curve underestimates the shift of the curve centre, while the shift of the left side of the rotation curve overestimates the shift of the curve centre. Rotational shifts given in the case of stable R-loops were determined from the shift of the rotation curve centre at low Cascade concentrations, unless otherwise noted. In the case of unstable R-loops, rotational shifts were determined from the shift of the left side of the rotation curve at low Cascade concentrations, from which 0.15 turns were subtracted to correct for the curve broadening in the presence of the R-loop.

2. RESULTS AND DISCUSSION

2.1. CRISPR4-Cas system of *S. thermophilus*

S. thermophilus is a gram-positive lactic acid bacterium widely used in diary industry (Bolotin et al., 2004). The main environmental threat of these bacteria is bacteriophages (Deveau et al., 2006). Therefore, they have developed several anti-phage barriers: Sie, RM and CRISPR-Cas systems (Labrie et al., 2010). RM and CRISPR-Cas systems are compatible and increase resistance against phages (Dupuis et al., 2013). Furthermore, more than one defense system of the same kind can be detected in one bacterium. For example, *S. thermophilus* DGCC7710 strain carries four distinct CRISPR-Cas systems in its genome: CRISPR1-Cas, CRISPR2-Cas, CRISPR3-Cas and CRISPR4-Cas (Horvath & Barrangou, 2010). CRISPR1-Cas and CRISPR3-Cas systems belong to the type II-A, which encodes Cas9 protein (Makarova et al., 2011b). In these two systems, rapid spacer acquisition has been demonstrated *in vivo*, while DNA interference stage has been reconstituted both *in vivo* and *in vitro* (Barrangou et al., 2007; Deveau et al., 2008; Gasiunas et al., 2012; Karvelis et al., 2013a; Karvelis et al., 2013b; Sapranauskas et al., 2011). Cas protein expression as well as crRNA maturation has been shown for all four systems (Carte et al., 2014; Young et al., 2012). However, neither spacer acquisition nor nucleic acid interference have been investigated experimentally for CRISPR2-Cas (type III-A) and CRISPR4-Cas (type I-E) systems yet.

In this study we analyzed DNA interference stage of CRISPR4-Cas system reconstituted *in vitro*. The system belongs to the type I-E and is comprised of eight *cas* genes arranged similarly to the CRISPR-Cas system of *E. coli* K-12. Twelve 33-bp length spacers are incorporated between 28-bp length repeat sequences in the CRISPR region (Figure 15 A).

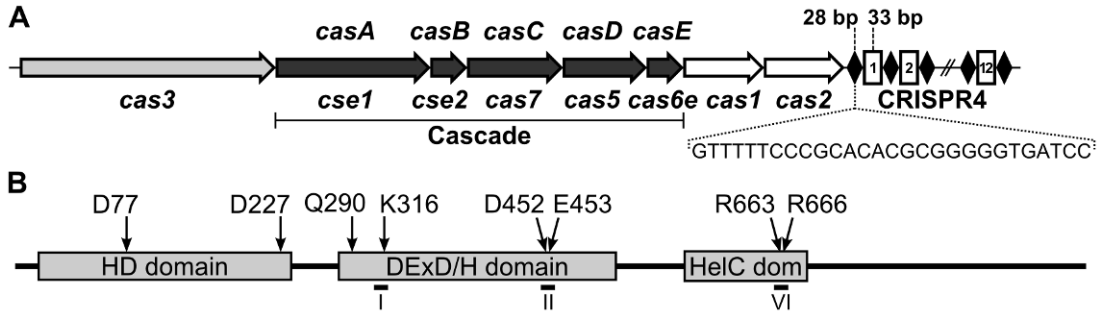

Figure 15. *S. thermophilus* CRISPR4-Cas system and Cas3 protein. (**A**) Schematic representation of the CRISPR4-Cas locus containing eight *cas* genes and twelve repeat-spacer units (conserved 28-bp palindromic repeats are separated from each other by 33-bp spacers of variable sequence). *S. thermophilus* Cascade genes homologous to the *E. coli* Cascade are underlined. Genes names according to (Brouns et al., 2008) and (Makarova et al., 2011b) are indicated, respectively, above and below corresponding genes. (**B**) Domain architecture of the *S. thermophilus* Cas3 protein. Domains identified by *in silico* analysis are shown as grey boxes. HD domain denotes HD-type phosphohydrolase/nuclease domain; DExD/H domain denotes DExD/H-box helicase domain; HelC dom denotes the C-terminal helicase domain. Conserved residues characteristic of the different domains and subject to alanine mutagenesis are indicated above the boxes. Location of the conserved helicase motifs are indicated by numbers I, II and VI (Singleton et al., 2007).

Cascade complex together with Cas3 protein are responsible for DNA interference in *E. coli* CRISPR-Cas system (Brouns et al., 2008). In *E.coli* Cse1, Cse2, Cas7, Cas5 and Cas6e (also termed as CasA, CasB, CasC, CasD and CasE, respectively) proteins together with crRNA assemble into the Cascade complex (Jore et al., 2011) [see section (1.4.3.1) for details]. Homologous proteins are encoded in *S. thermophilus* DGCC7710 CRISPR4-Cas system; therefore, analogous *S. thermophilus* Cascade complex should be formed (Figure 15 A). The *cas3* (GenBank: HQ453272) gene of the CRISPR4-Cas system encodes a protein of 926 amino acids with a predicted molecular mass of ~106 kDa. *In silico* analysis of Cas3 protein reveals a multidomain architecture (Figure 15 B). The N-terminal part of Cas3 shows conserved residues characteristic of the HD family of metal-dependent phosphohydrolases (HD domain) (Aravind & Koonin, 1998), whereas the C-terminal part of Cas3 has motifs characteristic of the Superfamily 2 (SF2) helicases (Singleton et al., 2007).

In order to elucidate mechanism of *S. thermophilus* CRISPR4-Cas DNA interference, first, we characterized individual biochemical activities of Cas3 protein and Cascade complex. Then, we reconstituted DNA interference *in vitro* combining these components together.

2.2. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase

2.2.1. Expression and purification of Cas3 protein

The cas3 gene from *S. thermophilus* DGCC7710 was cloned into the pBAD24-CHis plasmid to yield a construct encoding a fusion protein containing a C-terminal His_6-tag. The recombinant plasmid was expressed in the *E. coli* strain ER2267 and the Cas3 protein purified from the crude cell extracts. The purified Cas3 protein and its mutants were nearly homogeneous as evaluated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie Blue staining (Figure 16 A).

2.2.2. Cas3 shows nuclease activity located in the HD domain

The nuclease activity of Cas3 was analysed using circular M13mp18 ssDNA or pUC57 supercoiled double-stranded plasmid DNA (Figure 16 B). Cas3 degraded the M13mp18 ssDNA in a concentration- and time-dependent manner (Figure 16 B and D). In contrast, virtually no hydrolysis of the dsDNA occurred during the 2-h incubation. Mg^{2+} ions were required for the ssDNA hydrolysis.

Conserved amino-acid residues H27, H76, D77 and D276 located in the N-terminal HD-like domain (Aravind & Koonin, 1998) of Cas3 (Figure 15B) were identified as being part of the putative active site responsible for divalent metal binding and ssDNA hydrolysis. Alanine replacement mutants D77A and D227A were constructed by site-directed mutagenesis, mutant proteins were purified and their ability to degrade DNA was analysed (Figure 16 A and C). Experimental data indicate that while ssDNA was fully degraded in 2 h by the

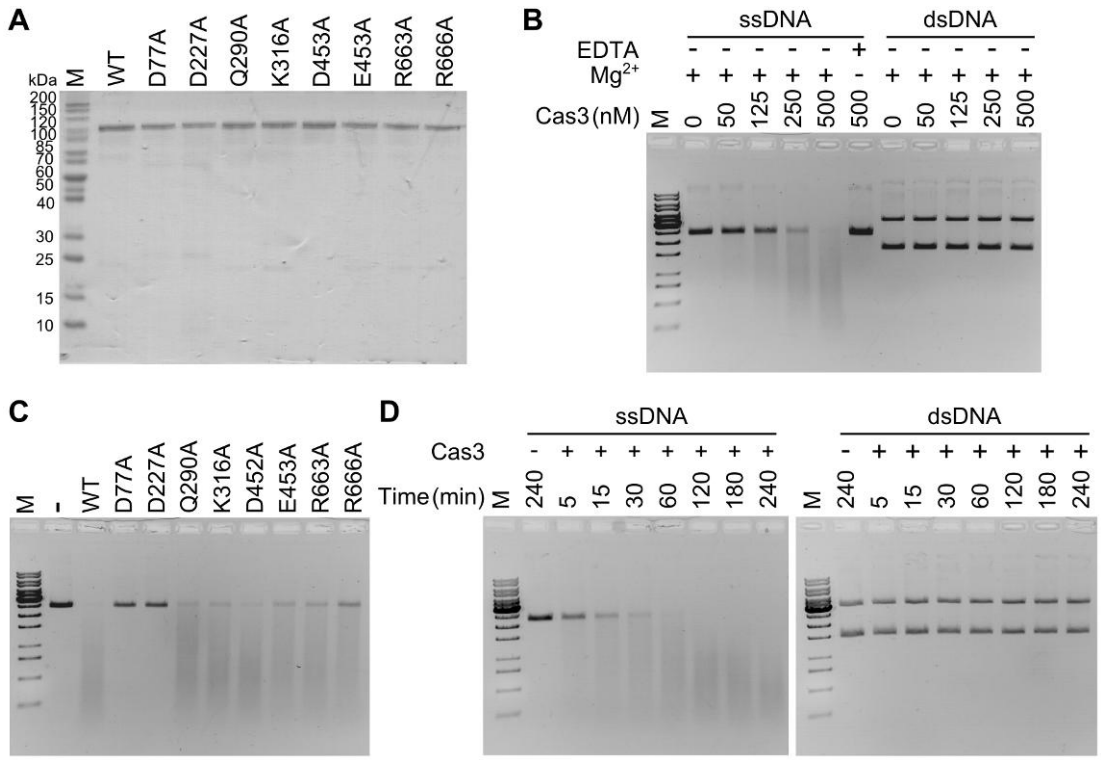

Figure 16. Cas3 nuclease activity. (**A**) SDS-PAGE of the purified Cas3 proteins and its mutants. Samples corresponding to 750 ng of each protein were loaded on the gel. (**B**) Degradation of the ssDNA and dsDNA. Various amounts of Cas3 were incubated in the presence of 4 nM of M13mp18 ssDNA or pUC57 dsDNA at 37°C for 2 h in the presence (+) or absence (−) of 10 mM $MgCl_2$ or 10 mM EDTA. (**C**) Effect of mutations on Cas3 nuclease activity. In all, 500 nM of protein was incubated in the presence of 4 nM of M13mp18 ssDNA at 37°C for 2 h. (**D**) Time courses of nuclease activity of Cas3. 500 nM of Cas3 were incubated in the presence of 4 nM of M13mp18 ssDNA or pUC57 dsDNA at 37°C. Time points of the reactions indicated above the figures. M – protein or DNA marker.

wild type (WT) Cas3, ssDNA hydrolysis was significantly reduced for D77A and D227A. Mutations in the helicase domain had much weaker effects on the associated nuclease activity (Figure 16 C).

2.2.3. Cas3 shows an ssDNA-stimulated ATPase activity

The presence of the characteristic helicase motifs (I, II and VI) responsible for ATP binding and hydrolysis in the primary sequence of Cas3 (Figure 15 B) predicts that the protein would be an ATPase. The ability of Cas3 to hydrolyze ATP was first examined by monitoring the hydrolysis of radioactively labelled [α^{32}P]ATP (Figure 17A). In the presence of the Cas3 protein alone, only minimal ATPase activity was detected. However, the ATP hydrolysis rate

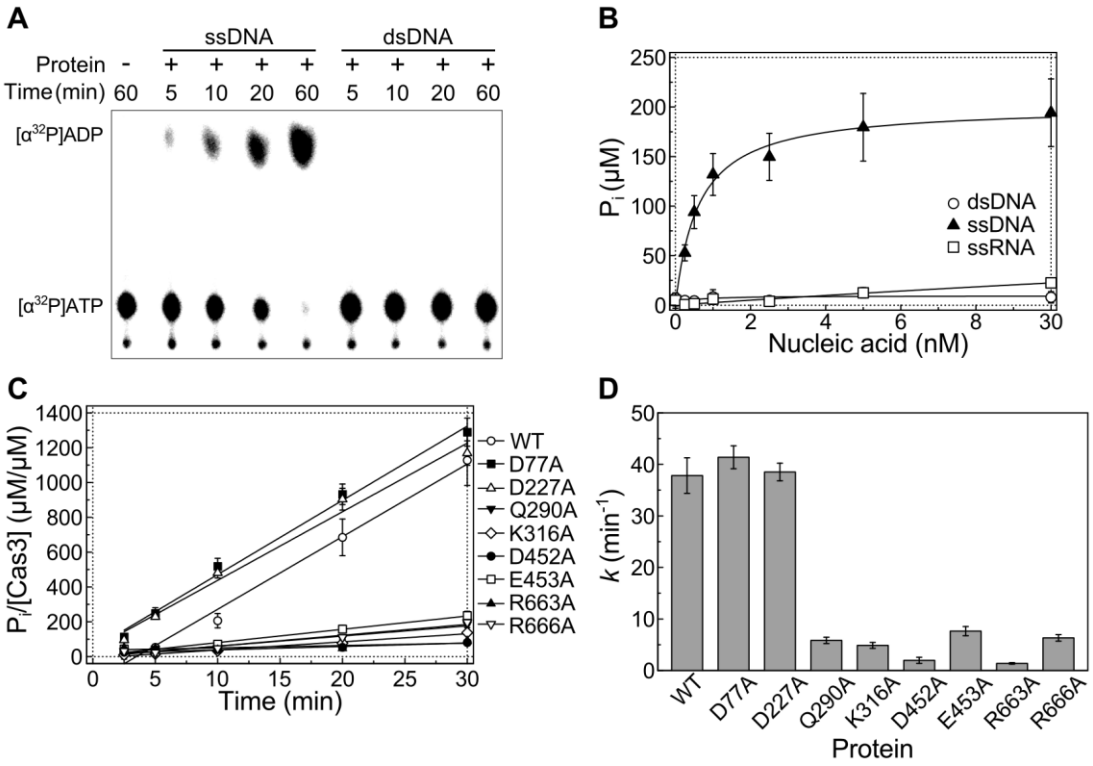

Figure 17. Cas3 ATPase activity. (**A**) Radioactive ATPase assay. ATPase reactions were conducted at 30°C in an AB1 buffer containing 2.5 nM ssDNA (M13mp18) or dsDNA (supercoiled form of pUC57 plasmid), and 250 nM Cas3. Reaction mixtures were supplemented with [α^{32}P]ATP (5 Ci/mmol), spotted onto a polyethyleneimine-cellulose thin-layer plate and separated by chromatography followed by phosphorimager visualization. (**B**) ATP hydrolysis dependence of nucleic acids. Malachite green assay was used to measure ATP hydrolysis through the detection of liberated-free phosphate from ATP. Reaction mixtures in the buffer described above contained varying amounts of ssDNA (M13mp18), dsDNA (supercoiled form of pUC57 plasmid) or 2223 nt RNA. (**C**) Time courses of ATP hydrolysis. Reaction mixtures contained 3 nM ssDNA (M13mp18). Malachite green assay was used to measure ATP hydrolysis through the detection of liberated-free phosphate from ATP. (**D**) ATP hydrolysis rates. Reaction rate constant k (min^{-1}) calculated from slopes of times courses shown in (**C**). Error bars indicate the ±standard deviation for the rate constant k value determined in three separate experiments.

increased significantly in the presence of M13 circular ssDNA. We further investigated the ATP hydrolysis by Cas3 in the presence of ssDNA, dsDNA and RNA by measuring the concentration of accumulated phosphate product using a colorimetric assay (Hyun et al., 2008). Data analysis revealed that ATPase activity of Cas3 increased significantly in the presence of ssDNA but was not stimulated by the dsDNA or by RNA (Figure 17 B). The linear accumulation of inorganic phosphate as a function of time (representative trace

shown in Figure 17 C) allowed us to calculate the rate of ATP hydrolysis at 0.5 mM of ATP as ~38 min^{-1}. Taken together, these results indicate that Cas3 possesses an ssDNA-dependent ATPase activity.

To test whether the conserved residues of the helicase domain are important for the Cas3 ATPase activity, we generated a set of alanine-substitution mutants. We mutated amino-acid residues Q290A [located in Q-motif (Tanner et al., 2003) important for ATP binding], K316A (located in motif I involved in ATP binding), D452A and E453A (located in motif II involved in Mg^{2+} coordination at the ATPase active site), R663A and R666A [located in motif VI (Tanner & Linder, 2001) involved in ATP binding]. The ATPase activity of the mutants decreased significantly (Figure 17 D), indicating the importance of the conserved amino-acid residues for the ATP binding/hydrolysis. Conversely, mutations in the N-terminal HD domain had no effect on the ATPase activity of Cas3.

In order to test the nucleotide specificity of Cas3, we compared the catalytic activity using ribo- and deoxyribonucleotide cofactors (Figure 18 A). Cas3 exhibited a strong preference towards ATP or dATP. We also found that GTP

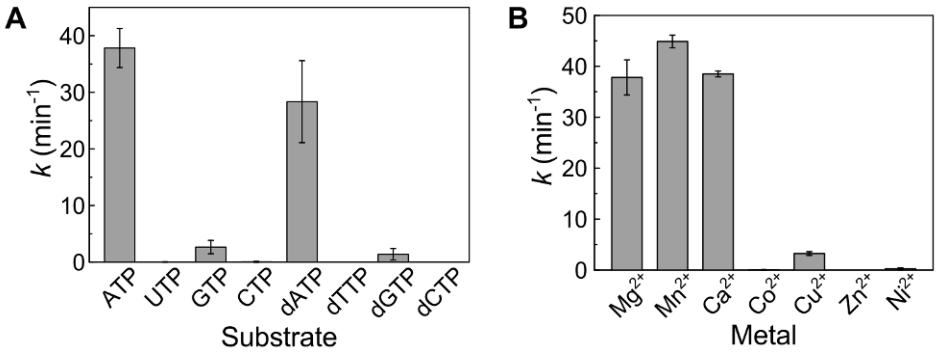

Figure 18. Cas3 nucleotide and metal specificity. (**A**) Hydrolysis of ribonucleotides and deoxyribonucleotides by the Cas3. Reactions were conducted in the reaction buffer containing various ribonucleotides and deoxyribonucleotides. (**B**) Cas3 ATPase activity dependence on divalent metal ions. Reactions were conducted in the reaction buffer containing various divalent ions. Malachite green assay was used to measure nucleotide hydrolysis through the detection of liberated free phosphate from nucleotide. Reaction rate constant *k* was calculated from the linear slopes of respective time courses. Error bars indicate the ±standard deviation for the rate constant *k* value determined in three separate experiments.

and dGTP could be hydrolyzed in the presence of ssDNA, but with specific activities that were approximately 10-fold lower than observed with ATP. No significant hydrolytic activity was observed with other nucleotides. Study of the divalent metal ion dependence of the Cas3 ATPase activity revealed that the fastest rate was observed with Mn^{2+}, followed closely by Mg^{2+} or Ca^{2+}. Cu^{2+} supported hydrolysis at a much lower rate. No significant ATPase activity above background was observed using Co^{2+}, Zn^{2+} or Ni^{2+} ions (Figure 18 B).

2.2.4. Cas3 shows DNA unwinding activity

An ssDNA-dependent ATPase activity of Cas3 suggests a possible translocase/helicase function. To analyse whether the purified Cas3 protein possesses true helicase activity (i.e. DNA strand separation), we determined its capacity to unwind DNA duplex substrates (Figure 19). Since many helicases require a stretch of ssDNA in order to load onto the substrate and since Cas3 ATPase activity is stimulated by ssDNA, the substrate was constructed by hybridizing a 20-nt oligodeoxynucleotide, labelled with ^{32}P at its 5′-terminus, to a circular M13mp18 ssDNA that contained a complementary sequence (Matson, 1986) (Figure 19 A). This substrate was incubated with Cas3 and the reaction products were separated on a polyacrylamide gel. The labelled oligonucleotide was then detected by autoradiography. The experimental results indicate (Figure 19 B) that Cas3 possesses a DNA unwinding activity that is dependent upon the presence of both Mg^{2+} and ATP but is not supported in the presence of the non-hydrolysable ATP analogue 5′-adenylyl-β,γ-imidodiphosphate (AMP-PNP). Furthermore, the Cas3 nuclease-deficient mutant D77A unwound the DNA similarly to the WT protein, while the D452A mutant, which has compromised ATPase activity, showed no unwinding activity.

To test whether the Cas3 protein unwinds an RNA/DNA heteroduplex, a substrate was constructed by hybridizing a 22-nt oligoribonucleotide to M13mp18 ssDNA as mentioned above. The experimental results indicate (Figure 19 C) that Cas3 displaces the 22-nt oligoribonucleotide in the presence

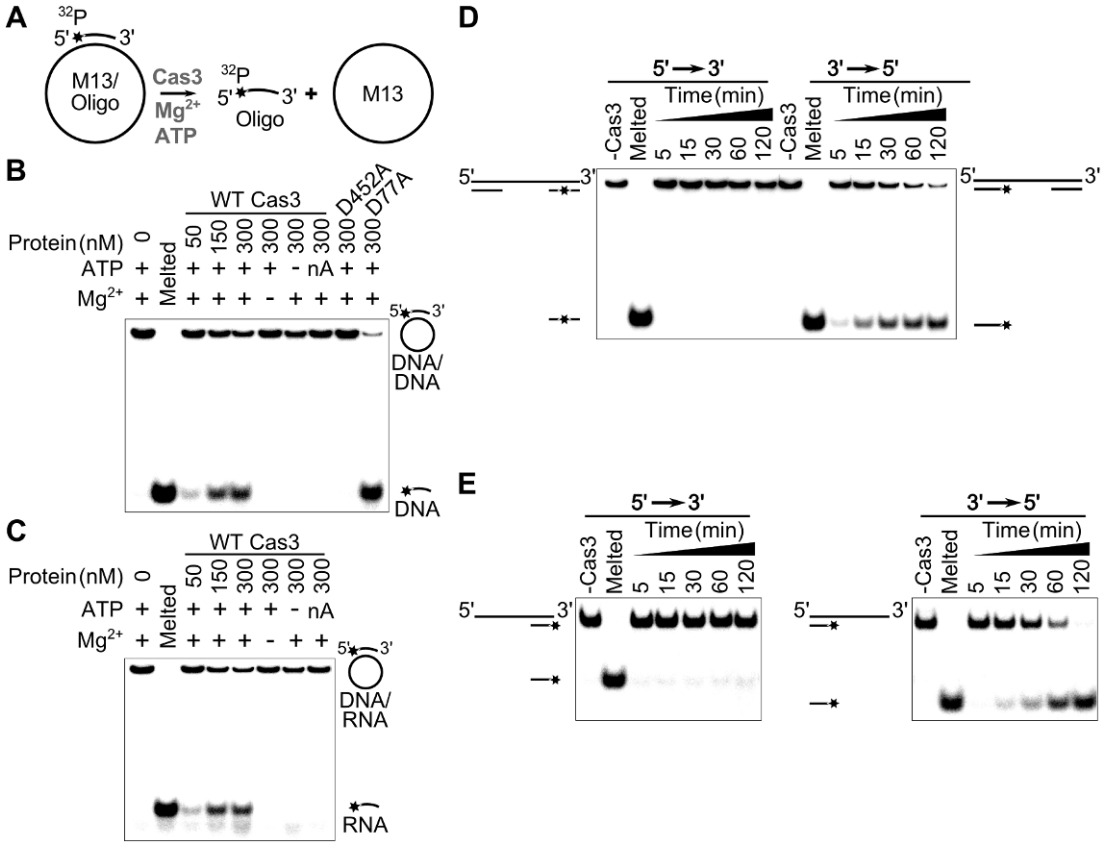

Figure 19. Cas3 helicase activity and polarity. (**A**). Schematic representation of duplex unwinding assay. (**B**, **C**) DNA–DNA and RNA–DNA duplex unwinding by Cas3. Cas3 displacement of a ^{32}P-labelled 20 nt oligodeoxynucleotide (**B**) or 22 nt oligoribonucleotide (**C**) annealed to an ssM13mp18 DNA is monitored in the polyacrylamide gel. Reactions were performed at 30°C for 1 h in the HB buffer containing 0.5 nM substrate and various amounts of protein. nA denotes the ATP analogue AMP-PNP. D452A and D77A are ATPase and nuclease domain mutants, respectively. (**D**) Cas3 polarity assay I. Cas3 displacement of 30 nt double-stranded fragments at the ends of the linear M13mp18 DNA. Partial duplex DNA was prepared by *Ehe*I cleavage of the labelled 60 nt oligodeoxynucleotide annealed to the M13pm18 DNA. Duplex regions are separated by a single-stranded region of few thousand nucleotides. Reaction mixture contained 500 nM of Cas3. Reactions were stopped at defined time intervals. (**E**) Cas3 polarity assay II. Cas3 displacement of the oligonucleotide-based 73 nt substrates containing 53 nt 3′- or 5′-overhangs. Reactions were performed as in (**B** and **C**) except that with the oligonucleotide-based substrates 500 nM of Cas3, 1.5 nM of substrate and 250 nM trap DNA (unlabelled oligonucleotide) were used in the reaction.

of ATP and Mg^{2+} ions. Thus, Cas3 can be classified as both a DNA-DNA and DNA-RNA helicase.

With partially duplex substrates, each DNA helicase is thought to bind first to an ssDNA region and then to approach and unwind duplex DNA in a particular direction. To determine the directionality of Cas3, we first prepared

a pair of labelled helicase substrates (Figure 19 D) that contained duplex DNA at both ends of a long linear ssDNA molecule. Since the substrates comprise duplex regions at both ends of a long linear molecule, Cas3 must first bind to the internal single-stranded regions of these substrates. If the enzyme subsequently moves into a 3′-5′ direction along the ssDNA segment, it would displace the 5′-labelled fragment from the substrate. In contrast, the 3′-labelled fragment would be displaced if the enzyme migrates in a 5′-3′ direction. Experimental data indicate that Cas3 moves primarily 3′ to 5′ along the ssDNA segment (Figure 19 D).

We also performed an alternative unwinding assay using radiolabelled DNA substrates of different structures. These 73 nt partial oligoduplexes contained 53 nt 3′- or 5′-single-stranded overhangs in addition to a 20-bp duplex region (Figure 19 E). Cas3 could only unwind the substrates containing a 3′-overhang, confirming the 3′-5′ polarity seen above. The unwinding activity of Cas3 was observed in the presence of ATP and was not detected in the absence of ATP or in the presence of the non-hydrolyzable ATP analogue (AMP-PNP), suggesting that ATP hydrolysis is required for helicase function.

2.2.5. Cas3 degrades ssDNA and unwinds DNA in the presence of ATP

All type I CRISPR-Cas systems encode a signature Cas3 protein, which participates in DNA interference (Brouns et al., 2008; Makarova et al., 2011b). The Cas3 protein from type I-E of *S. thermophilus* is arranged as a polypeptide comprised of an N-terminal HD-type phosphohydrolase/nuclease domain and a C-terminal SF2 helicase domain (Haft et al., 2005; Makarova et al., 2006). This architecture is characteristic for most Cas3 proteins but it is not absolutely conserved. For example, in the Cas3 protein of type I-A and I-B, HD and helicase domains are found as separate proteins Cas3′ and Cas3′′, respectively (Makarova et al., 2011b).

The HD-type phosphohydrolase/nuclease domain is found in a superfamily of enzymes with either a predicted or known phosphohydrolase activity

(Aravind & Koonin, 1998). According to Prosite database (de Castro et al., 2006), bacterial HD domains are found in combination of 49 different partner domains, which presumably modulate protein function. HD-domain-containing proteins appear to be involved in nucleic acid metabolism and signal transduction, as well as other unknown functions (Aravind & Koonin, 1998). For example, the HD domain of the *E. coli* tRNA nucleotidyltransferase exhibits 2,3-cyclic phosphodiesterase, 2-nucleotidase and phosphatase activities (Yakunin et al., 2004).

The N-terminal domain of Cas3 contains a characteristic signature of the HD-type phosphohydrolase/nuclease domain. The conserved residues (H…HD…D) predicted to be involved in the coordination of the divalent metal (Aravind & Koonin, 1998) are conserved in the N-terminal domain of *S. thermophilus* Cas3 and correspond to the residues H27, H76, D77 and D227. We show here that in the presence of magnesium ions, Cas3 has a nuclease activity that degrades ssDNA. It does not act on dsDNA. Mutation of the key metal-coordinating residues D77 and D227 compromised the ability of Cas3 to degrade ssDNA, but did not affect the ATPase activity, consistent with *in silico* predictions.

Recently, crystal structures of Cas3′ from type I-A (*Methanocaldococcus jannaschii*) and HD domain of Cas3 from type I-E (*T. thermophilus*) have been solved showing α-helical architecture of the HD domain (Beloglazova et al., 2011; Mulepati & Bailey, 2011). Based on the structures, conserved residues of the HD domain are clustered on surface of the domain were they co-ordinate metal ions. These structural data supports our biochemical evidences that conserved residues of the HD form the active site of the nuclease.

The C-terminal fragment of Cas3 protein carries signature motifs characteristic of the SF2 helicases of the DExD/H subgroup (Jackson et al., 2014b; Singleton et al., 2007). Helicases use ATP to unwind and translocate nucleic acids or remodel nucleic acids or nucleic acid–protein complexes (Cordin et al., 2006; Fairman-Williams et al., 2010; Singleton et al., 2007). We

show here that the Cas3 protein exhibits ssDNA-dependent ATPase activity, which absolutely requires Mg^{2+} ions. Nine conserved domains Q, I, Ia, Ib and II–VI have been identified in the SF2 group of DExD/H-type helicases (Cordin et al., 2006; Fairman-Williams et al., 2010; Singleton et al., 2007). The highest level of sequence conservation in the SF1 and SF2 helicase families is seen in the residues that coordinate binding and hydrolysis of the triphosphate (motifs I, II and VI). Mutations of the amino-acid residues Q290, K316, D452, E453, R663 and R666, in the conserved motifs I, II and VI, abolished or significantly compromised ATP hydrolysis. However, none of the helicase mutants showed any change in the ssDNA-degrading activity. Among different deoxy- and ribonucleotides tested, only ATP or dATP were hydrolysed significantly. The maximum rate of ATP hydrolysis was ~38 molecules per minute. This value is also close to the value reported for the bacterial XPB helicase (Biswas et al., 2009) and NS3 helicase from hepatitis C virus (Kyono et al., 2003). The inefficient ATP hydrolysis suggests that Cas3 alone is not very processive or that it is an intrinsically slow ATPase involved, for example, in the local remodelling of protein-crRNA complex.

In parallel to the ATPase activity, Cas3 protein is able to unwind oligonucleotide-M13 DNA complex. The unwinding activity of Cas3 is dependent on the protein concentration and presence of ATP and Mg^{2+} ions. Using a linear M13 molecule with a large internal region of ssDNA where a helicase can assemble, and two ^{32}P-labelled duplex regions of different lengths, we show that Cas3 helicase has $3' \rightarrow 5'$ directionality. It is likely that Cas3 functions as DNA translocase. Cas3 contains a long C-terminal extension that follows the conserved helicase domains of the SF2 superfamily. Terminal domains of helicases have been demonstrated to direct recruitment of partner proteins/complexes, to promote interactions with other proteins, or to facilitate recognition of specific nucleic acid regions (Karow & Klostermeier, 2010). One cannot exclude that the C-terminal domain of Cas3 performs similar functions, e.g., interaction with Cascade complex.

Recently, crystal structure of full-length Cas3 protein from type I-E (*Thermobifida fusca*) in complex with a ssDNA have been solved, showing the 3′-terminus of ssDNA is bound by HD domain, while the 5′-terminus is threaded through a pore formed by the helicase and C-terminal domains (Huo et al., 2014). These structural data are in agreement with our biochemical evidences for 3′ → 5′ directional translocation of Cas3 that is accompanied by ssDNA degradation.

2.3. Target DNA recognition by Cascade complex

Bioinformatic analysis of PAM sequences was performed at DuPont by dr. Philippe Horvath.

Mass-spectrometry analyses of crRNA and proteins of S. *thermophilus* Cascade complex were done at University of Sheffield by dr. Sakharam P. Waghmare and dr. Mark J. Dickman.

Single-molecule experiments with the Cascade on magnetic tweezers were done at University of Münster by Maria S. Tikhomirova and prof. dr. Ralf Seidel.

2.3.1. Cascade complex targets dsDNA via a PAM-mediated R-loop formation

2.3.1.1. Cloning, expression, and isolation of St-Cascade

In *E. coli*, Cse1, Cse2, Cas7, Cas5, and Cas6e proteins and crRNA form a Cascade complex (Brouns et al., 2008), which, together with Cas3, provide interference against invading foreign DNA. We tested the hypothesis that homologous *S. thermophilus* Cas proteins (Figure 15 A) may assemble into a similar Cascade complex, and designed the following strategy for complex isolation. First, three compatible heterologous plasmids containing, respectively, a *cse1-cse2-cas7-cas5-cas6e* cassette, the C-terminal His-tagged variant of *cas7*, and six copies of the repeat-spacer-1 unit (6 × SP1) of the *S. thermophilus* CRISPR4 region, were engineered. Next, all three plasmids were

co-expressed in *E. coli* BL21 (DE3) strain and the Cascade complex was purified by subsequent Ni^{2+}-chelating, size exclusion and heparin affinity chromatography steps. SDS-PAGE analysis (Figure 20 A) of the isolated complex revealed five bands that matched to individual Cas proteins, suggesting that Cse1, Cse2, Cas7, Cas5, and Cas6e proteins assemble into a Cascade complex similar to that of *E. coli*. The identity of all Cas proteins in Cascade was confirmed by mass spectrometry analysis (Data not shown). The stoichiometry of the protein complex was not directly determined; however, the band intensity in the SDS-PAGE (Figure 20 A) in conjunction with the mass spectrometry analysis of the Cascade tryptic digest suggests that the Cas7 protein is the most abundant protein present in Cascade similar to the *E. coli* Cascade (Jore et al., 2011). Denaturing PAGE analysis revealed that small RNA co-purifies with the Cascade complex (Figure 20 B).

2.3.1.2. Characterization of *S. thermophilus* CRISPR4-Cas crRNA

Next, we used denaturing RNA chromatography in conjunction with electrospray ionization mass spectrometry (ESI-MS) to characterize the mature crRNAs isolated directly from the Cascade complex. Denaturing ion pair reverse phase chromatography was used to purify the crRNA directly from the Cascade complex (Dickman & Hornby, 2006; Waghmare et al., 2009). The RNA isolated from this complex consisted of a single mature crRNA with a retention time consistent with an approximate length of 60-nt (Figure 20 C). Purified mature crRNA was further analysed using ESI-MS to obtain the accurate intact mass. A molecular weight of 19482 Da was obtained (Figure 20 D). In addition, ESI-MS/MS was used to analyse the oligoribonucleotide fragments generated from RNase T1 and RNase A digestion of the mature crRNA (Data not shown). In conjunction with the intact mass analysis and denaturing PAGE, these indicate that processing of *S. thermophilus* CRISPR4-Cas crRNAs is similar to that of *E. coli* CRISPR-Cas crRNAs, generating a 61-nt crRNA (consisting of a 7-nt 5′ handle, a 33-nt spacer, and a 21-nt 3′ handle) with 5′-OH and 3′-P_i (MW 19481.5 Da) (Figure 20 E). Further verification of

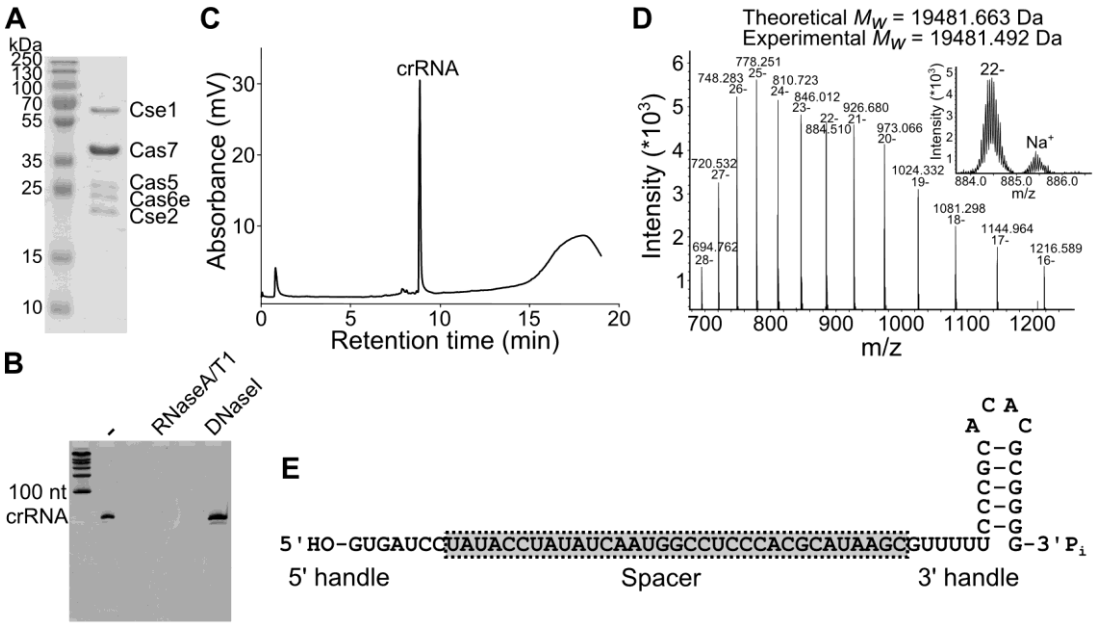

Figure 20. *S. thermophilus* Cascade complex. (**A**) Coomassie blue-stained SDS-polyacrylamide gel of Cascade complex proteins isolated using the Cas7-His_6 protein as bait. (**B**) Cascade contains small RNA. Nucleic acids were isolated from the Cascade complex and treated with RNaseA/T1 or DNaseI. (**C**) IP RP HPLC analysis of mature crRNA. (**D**) LC ESI-MS analysis of purified *S. thermophilus* crRNA. Inset shows an enhanced view of the 22-charge state. (**E**) Architecture of crRNA co-purifying with the Cascade protein complex.

the 3′-P_i termini was obtained upon acid treatment of the crRNA where no change in mass was observed using ESI-MS.

2.3.1.3. PAM sequence analysis of the *S. thermophilus* CRISPR4-Cas system

The PAM located in the vicinity of a protospacer is absolutely required for silencing of invading DNA by type I and type II CRISPR-Cas systems (Deveau et al., 2008; Horvath et al., 2008; Sapranauskas et al., 2011; Semenova et al., 2011). In the *E. coli* type I-E system, the PAM corresponds to the 5′-AWG-3′ sequence located immediately upstream of a protospacer (Mojica et al., 2009) and is essential for *E. coli* Cascade binding and subsequent DNA interference (Semenova et al., 2011). On the other hand, experimental analysis of CRISPR repeat boundaries in *E. coli* suggests a dinucleotide 5′-AW-3′ as PAM (Goren et al., 2012). To determine the putative PAM sequence of the CRISPR4-Cas system, we analysed all currently available CRISPR4 spacer sequences found

in *S. thermophilus* strains. A CRISPR4 locus is present in DGCC7710 (Horvath & Barrangou, 2010) and three other strains from the DuPont culture collection. In DGCC7710, the CRISPR4 locus contains 12 unique spacers, and 26 more unique spacers were identified in the three other CRISPR4-positive strains. Sequence similarity searches, both in public and proprietary sequence databases, showed that most (26 out of 38) of these CRISPR4 spacer sequences have matches (protospacers) in *S. thermophilus* phage sequences. Only perfect matches (100% identity over the complete spacer sequence) between spacer and protospacer were considered, providing a set of 106 matching sequences. The sequences located immediately upstream and downstream of these protospacers were examined for the presence of a possible PAM. After removal of redundant alleles, a Weblogo representation (Crooks et al., 2004) was used to depict sequence conservation over a 15-nt segment of 28 (upstream) and 21 (downstream) unique sequences (Figure 21 A). A 2-base pair (bp) conserved motif 5′-AA-3′ could be identified immediately upstream of the protospacers.

2.3.1.4. PAM is required for Cascade binding to the protospacer

To determine whether the predicted PAM sequence is important for protospacer recognition, we analysed Cascade binding to a set of synthetic 73-bp oligoduplexes containing the protospacer-1 sequence and variable nucleotides at positions -2 and -1 (Figure 21 B). Oligoduplexes were radiolabelled at the 5′-end of the target strand, and the Cascade binding affinity was evaluated by electrophoretic mobility shift assay (EMSA). Binding analysis revealed that oligoduplexes fall into three categories with regards to Cascade binding. Oligoduplexes containing N(-2)A(-1) nucleotides in the predicted PAM display high binding affinity with K_d ~0.2 nM, oligoduplexes containing N(-2)T(-1) nucleotides show binding with K_d <10 nM, while all other oligoduplexes except A(-2)G(-1) bind with the same affinity as the non-specific oligoduplex containing protospacer-3 instead of protospacer-1. Thus, these results suggest that a single nucleotide PAM, A or T (W) at the -1

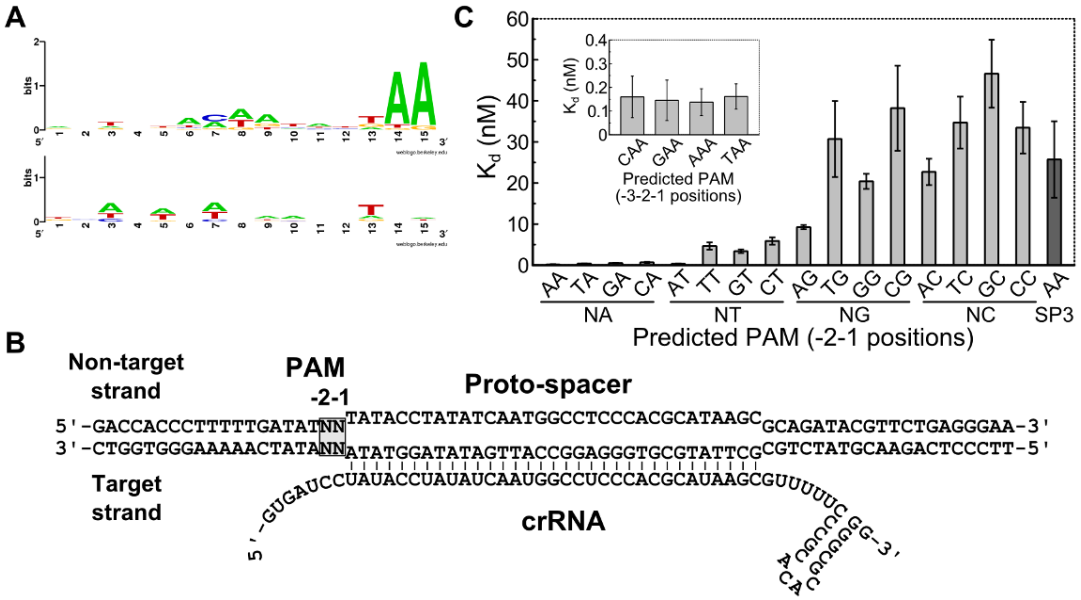

Figure 21. PAM-dependent Cascade binding. (**A**) Predicted PAM for the CRISPR4-Cas system. Weblogo representation (Crooks et al., 2004) of 15-nt sequences found immediately upstream (top) and downstream (bottom) of phage protospacers that match known CRISPR4 spacers. A conserved, 2-nt PAM (5′-AA-3′) is located immediately upstream of the proto-spacers. (**B**) A schematic representation of a putative R-loop structure resulting from the Cascade binding to the 73-bp oligodeoxynucleotide. Nucleotides NN at −1 and −2 positions of predicted PAM were varied. In the R-loop structure, a target strand bound to the crRNA is engaged into a heteroduplex while the non-target strand is displaced as a single-stranded DNA. (**C**) PAM sequence dependence of a protospacer-1 binding by Cascade. Bar diagram shows dissociation constant K_d values obtained by EMSA. Error bars represent standard deviations of average K_d value determined in three separate experiments. Oligoduplex containing a non-matching protospacer-3 sequence was used as a non-specific DNA control. Inset shows K_d values of oligodeoxynucleotides with varied -3 position in the vicinity of the predicted PAM.

position upstream of the protospacer is required for the CRISPR4-Cas system. The G and C nucleotides are not tolerated at this position except for the A(-2)G(-1) dinucleotide (Figure 21 C).

To test whether a non-conserved nucleotide at the -3 position in the vicinity of the predicted PAM is important for spacer recognition, we analysed Cascade binding to a set of oligoduplexes containing a conserved A(-2)A(-1) dinucleotide and any nucleotide at the -3 position (SP1-TAA, SP1-AAA, SP1-GAA, and SP1-CAA, respectively). EMSA analysis revealed that Cascade bound all oligoduplexes with a variable N(-3) nt with the same affinity (Figure

21 C inset), confirming that the -3 position is not important for Cascade binding.

In type I CRISPR systems, as exemplified by *E. coli* and *P. aeruginosa*, target recognition is governed by the crRNA seed sequence located at the 5′-end of the spacer region (Semenova et al., 2011; Wiedenheft et al., 2011b) and results in the formation of an R-loop where the target strand of the protospacer is engaged into a heteroduplex, while the non-target strand is displaced as single-stranded DNA (ssDNA). To demonstrate the formation of the R-loop upon *S. thermophilus* Cascade binding to a protospacer, we used the P1 nuclease that specifically cleaves ssDNA regions (Jore et al., 2011) (Figure 22). In the oligoduplexes SP1-AA and SP1-AG that contain correct PAMs, the non-target strand is susceptible to endonuclease P1 cleavage, while the target strand is resistant to P1 nuclease treatment. On the other hand, in the oligoduplex SP1-CC, which lacks a correct PAM, or in the oligoduplex SP3-AA which contains a PAM but lacks a matching protospacer sequence, both DNA strands were resistant to nuclease P1 cleavage. Thus, nuclease P1 assay confirms that an R-loop is formed only when both the correct PAM and a matching protospacer sequence are present in the oligoduplex.

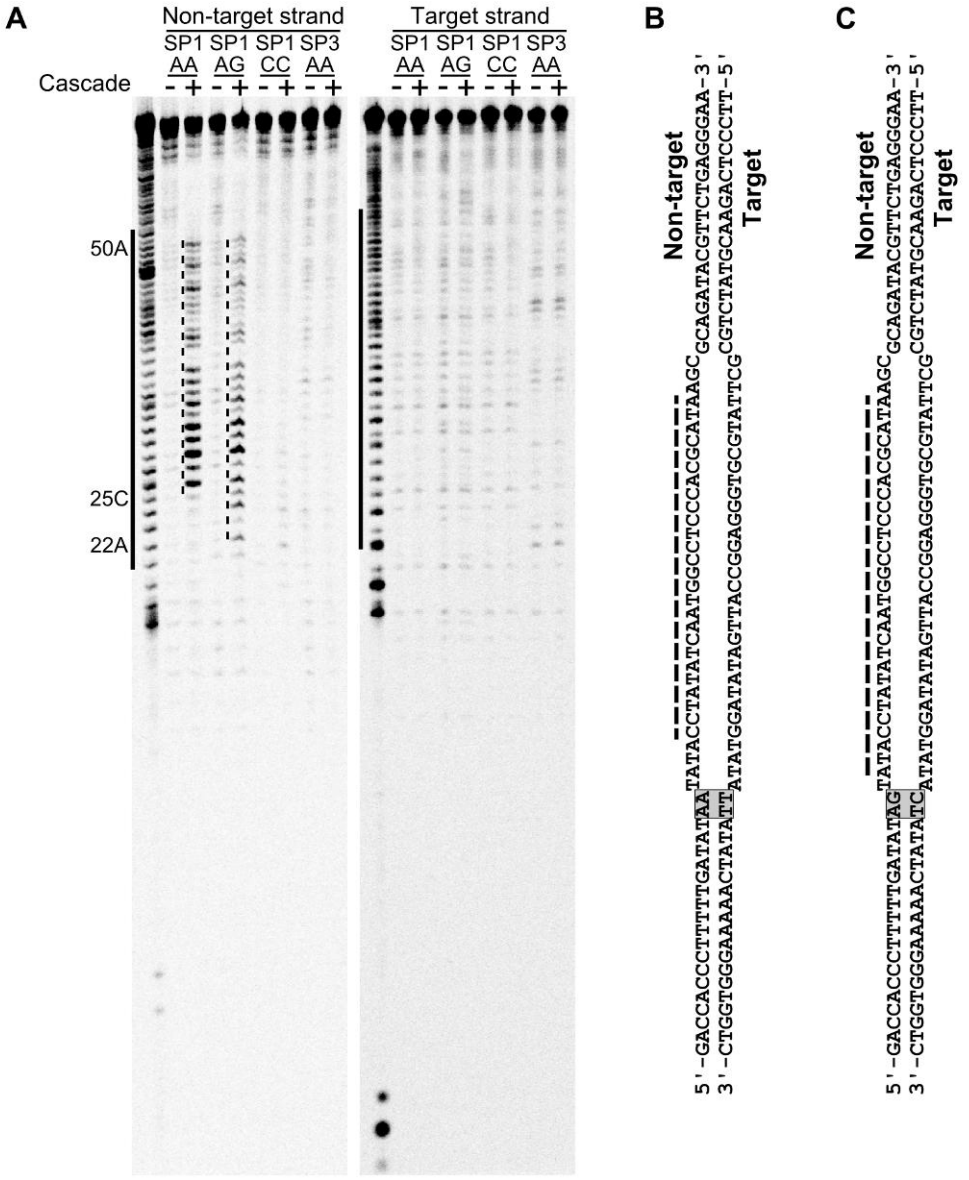

Figure 22. Mapping of ssDNA regions in the Cascade–target DNA complex using nuclease P1. (**A**) Denaturating polyacrylamide gels of oligoduplex footprints. Sensitive regions are indicated by marginal nucleotides and dashed lines. The protospacer regions are indicated by solid lines according to the sequencing lanes of each strand. (**B**). SP1-AA oligoduplex sequence. (**C**). SP1-AG oligoduplex sequence. The protospacer regions are shown as loops. PAM sequences are marked by gray squares. P1-sensitive regions of the non-target strands are indicated by dashed lines.

2.3.1.5. Target DNA recognition by Cascade

Cascade isolated from the heterologous *E. coli* host carries a 61-nt crRNA with a 5′-OH and 3′-phosphate, which guides Cascade binding to the protospacer sequence in the target DNA. In the *S. thermophilus* CRISPR4 array, the first 28-nt of the repeat are strictly conserved, while the 29th nt is degenerated (C or T). Therefore, we postulate that in the St-CRISPR4 array the repeat and spacer sequences have a length of 28 and 33-nt, respectively. In the orthologous *E. coli* CRISPR-Cas system, 29-nt repeat and 32-nt spacer sequences were initially proposed (Ishino et al., 1987; Jansen et al., 2002). However, the analysis of sequences of newly inserted repeats in an *E. coli* CRISPR array *in vivo* showed that a base previously thought to belong to the repeat is actually derived from the protospacer (Goren et al., 2012; Swarts et al., 2012). Therefore, the conserved repeat sequence [‘duplicon’, (Goren et al., 2012)] in the *E. coli* CRISPR array was proposed to be 28-nt, delimiting 33-nt spacers. Both *E. coli* and *S. thermophilus* processed crRNAs are composed of 61-nt, suggesting that in the precursor crRNA the cleavage position by Cas6e endoribonucleases is conserved and located at the 21st nt within the repeat sequence. In this case, the mature crRNAs are made of a 7-nt 5′ handle, a 33-nt spacer, and a 21-nt 3′ handle.

Similarly to other type I systems, Cascade binding to the oligoduplexes containing a matching protospacer sequence requires a PAM sequence located in the vicinity of the protospacer. In the *S. thermophilus* CRISPR4 system, the PAM predicted by *in silico* analysis of the matching protospacer sequences in *S. thermophilus* phages is an AA dinucleotide located immediately upstream of the protospacer. Surprisingly, according to EMSA experiments, the PAM required for Cascade binding to the protospacer sequence is limited to a single A(-1) or T(-1) nucleotide. Nucleotide replacement at the −1 position with a G or C abrogates Cascade binding to the protospacer, with the only exception of an A(−2)G(−1) dinucleotide that still functions as a PAM and promotes Cascade binding, albeit less efficiently than A(−1) or T(−1) variants. In the presence of the correct PAM, Cascade binds to the DNA containing a

protospacer sequence with a sub-nanomolar K_d. We suggest that PAM recognition is a key step that triggers subsequent *S. thermophilus* crRNA binding to the matching DNA strand, presumably through the seed sequence in the immediate vicinity of the PAM (Semenova et al., 2011; Wiedenheft et al., 2011b).

The discrepancy between the predicted and experimentally determined PAM may be due to several reasons. First, because only a small part of available phage sequence space is explored, it is possible that not all PAM variants were identified in the shallow subset of investigated phage genomes. Alternatively, the requirements for the PAM stringency may differ for the spacer acquisition and interference steps (Swarts et al., 2012). To escape CRISPR interference, bacteriophages often mutate PAM or protospacer sequences (Deveau et al., 2008), therefore a short and promiscuous PAM [such as A(-1) or T(-1)] identified for the *S. thermophilus* CRISPR4-Cas system may be advantageous for interference. The PAM identified by *in silico* analysis of the matching protospacer sequences in *S. thermophilus* phages may reflect the more stringent PAM requirement for the spacer acquisition step, executed by Cas1 and Cas2 proteins (Yosef et al., 2012).

2.3.2. Direct observation of R-loop formation by single Cascade complex

2.3.2.1. Direct observation of R-Loop formation in real time

In bulk, the R-loop is formed upon the PAM-mediated Cascade complex binding to the DNA target (Figure 23 A). To observe R-loops in single DNA molecule supercoiling experiments, we used magnetic tweezers (Brutzer et al., 2010; Mosconi et al., 2009) (Figure 23 B). A 2.1-kbp DNA, containing a single protospacer and PAM, was attached at one end to a magnetic bead and at the other end to the bottom of a fluidic cell. A pair of magnets above the cell was used to stretch the DNA and to supercoil it by rotating the magnets. Simultaneously the DNA length was measured (Klaue & Seidel, 2009). Upon

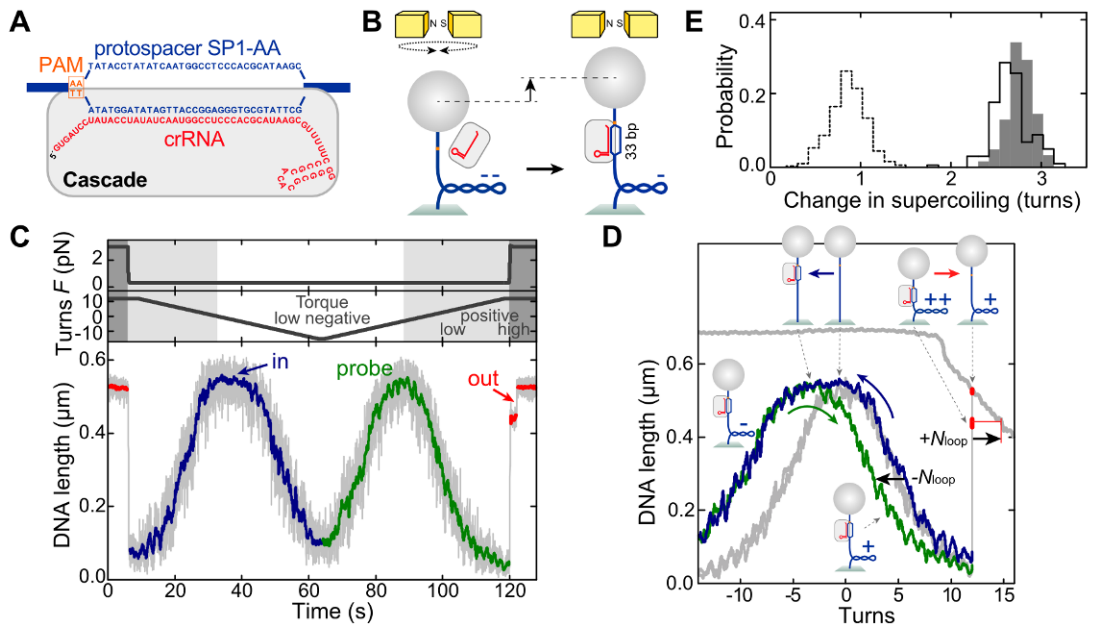

Figure 23. R-loop formation and dissociation by Cascade observed in single-molecule twisting experiments. (**A**) Schematics of the anticipated R-loops formed by Cascade (33 bp). (**B**) Magnetic tweezers-based twisting assay. R-loop formation on supercoiled DNA molecules at fixed rotation causes local DNA untwisting. Compensatory overtwisting of the DNA changes the supercoiling, resulting in a DNA length change. (**C** and **D**) R-loop cycle experiment in the presence of 10 nM Cascade. DNA with matching protospacer/PAM (A) is negatively supercoiled at 0.31 pN to induce R-loop formation (blue area of trace), followed by positive supercoiling to probe: the presence of the R-loop (green area of trace); and R-loop dissociation at an increased force of 3.0 pN (red area of trace). Blue and red arrows indicate the positions of R-loop formation and dissociation, respectively. In (**D**), the lower and upper gray supercoiling curves were taken on the same DNA molecule at 0.31 and 3.0 pN, respectively, before Cascade addition. (**G**) Cascade-induced shift of the supercoiling curve. Shifts of the right part of the supercoiling curve after R-loop formation (gray bars; $-N_{\mathrm{loop}}$ in D) and after full R-loop dissociation (bars with solid black outline; $+N_{\mathrm{loop}}$ in D) are shown. Bars with dashed black outline show the shift of the first R-loop dissociation substep for Cascade (Figure 24 A).

supercoiling DNA at constant force, its length stays initially constant. Once a critical torque in the molecule is reached, its length starts to decrease because of formation of a plectonemic superhelix resulting in a characteristic rotation curve and an associated torque profile (Figure 23 B) (Forth et al., 2008; Kauert et al., 2011; Mosconi et al., 2009; Oberstrass et al., 2012). Enzyme-dependent local DNA unwinding (e.g., attributable to R-loop formation) changes the DNA twist and can be seen as a shift of the whole rotation curve or as a DNA length change (Howan et al., 2012) (Figure 23 B). To detect R-loop formation, we carried out "R-loop cycles" with Cascade on DNA with matching

protospacer and 5′-AA-3′ PAM sequence (Figure 23 A). First, we slightly untwisted the DNA (producing negative supercoiling) at low force to help R-loop formation (blue curves in Figure 23 C and D). Subsequently, we probed R-loop dissociation by rewinding the DNA to produce positive supercoiling (green curves in Figure 23 C and D). We observed efficient R-loop formation (100% of all cases; n = 89), which was seen as a shift of the left side of the probe curve toward negative turns compared with the curve in absence of the proteins (Figure 23 D). Cascade formed an R-loop instantaneously at low negative supercoiling (-1 to -2 turns). The R-loop was stable at positive turns and low force, seen as a stable shift of the right side of the probe curve (Figure 23 D). R-loop dissociation could be observed at elevated force (corresponding to elevated positive torque) as an abrupt length jump (Figure 23 C and D). The right part of the supercoiling curve shifted by −2.62 ± 0.04 turns (formation) and +2.67 ± 0.03 (dissociation), whereas the center of the curve moved by 2.81 ± 0.07 turns (Figure 23 E). These values are slightly smaller than anticipated (3.1 turns considering a DNA helical pitch of 10.5 bp), possibly because of compensatory writhe from DNA bending induced by Cascade binding (Westra et al., 2012b).

2.3.2.2. Torque dependence of R-Loop formation and dissociation

To obtain insight into the previously proposed regulation of R-loop formation by supercoiling (Westra et al., 2012b) and into the energetics of the R-loop structure, we quantified the R-loop formation and dissociation kinetics as a function of the applied torque from repeated R-loop cycles (Figure 24 A). For Cascade, both R-loop formation and dissociation were torque-dependent (Figure 24 B). R-loop dissociation required an ~fourfold higher absolute torque value than formation. Most dissociation events (76%; n = 482) displayed a short-lived and torque-dependent intermediate state corresponding to 0.9 ± 0.1 turns (Figures 23 E and 24 A and B). By fitting the torque dependence of the kinetics to an Arrhenius-like model (Figure 24 B), we could calculate the transition state distances (ΔN_{in} = 1.5 ± 0.2 turns for R-loop formation; $\Delta N_{out}{}^{1}$ =

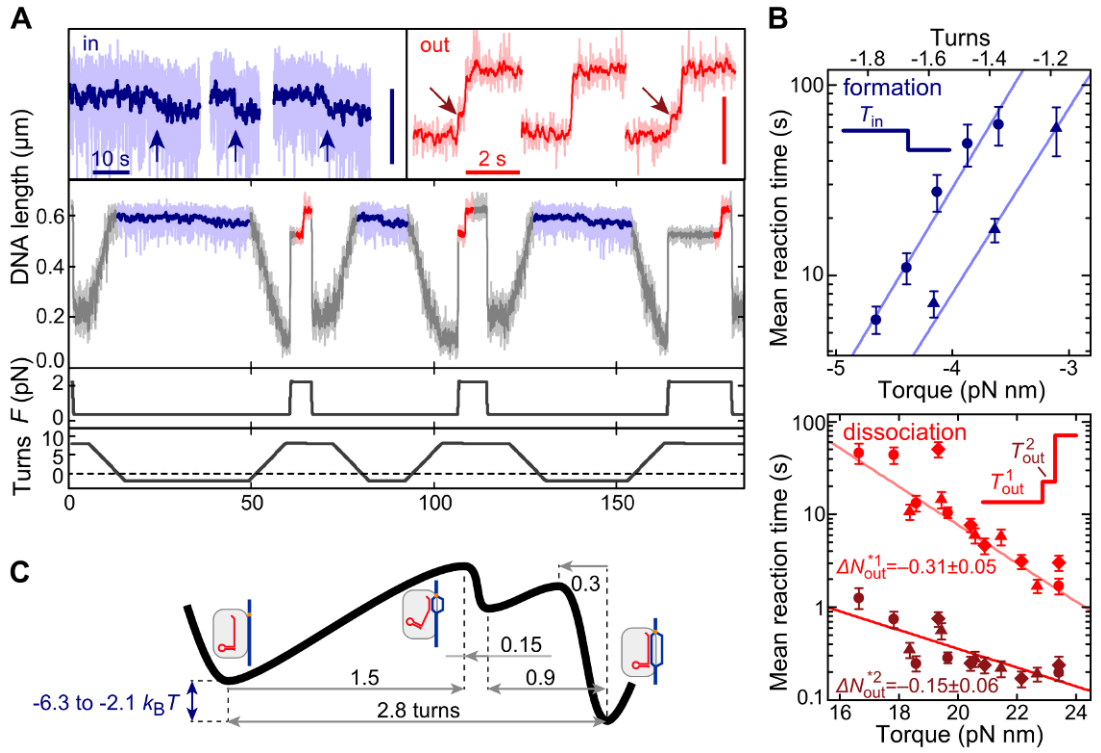

Figure 24. Torque dependence of R-loop formation and dissociation by Cascade. (**A**) Repetitive cycles of R-loop formation (at −1.6 turn, 0.36 pN) and R-loop dissociation (at +8 turns, 2.2 pN) by Cascade. R-loop formation and dissociation is seen as a DNA length decrease or increase, respectively (enlarged views on top). (Vertical scale bars: 100 nm). Blue arrows indicate R-loop formation, and brown arrows indicate a short-lived intermediate state upon R-loop dissociation. (**B**) Mean R-loop formation times (blue scatter plots) and dissociation times (red and brown scatter plots for the first and second dissociation step, respectively) as a function of torque. Different symbols indicate measurements on different DNAs to show experimental variation. Solid lines are exponential fits to the data. Numbers indicate distances to the transitions states as obtained from the fits. (**C**) Energy landscape of the R-loop formation process by Cascade based on the torque dependency.

0.31 ± 0.05 and $\Delta N_{out}{}^2$ = 0.15 ± 0.06 turns for the two dissociation intermediates). From the mean transition times in the absence of torque, a free energy gain for R-loop formation of 2.1–6.3 k_BT was determined. A simplified energy landscape for R-loop formation by Cascade is suggested (Figure 24 C), using the transition state distances combined with the rotational shifts for full and intermediate R-loop states.

2.3.2.3. PAM mutations hinder R-Loop formation but not its stability

To clarify whether the PAM regulates R-loop formation by kinetic inhibition or altered R-loop stability, we measured the dynamics of R-loop

formation and dissociation on substrates with matching protospacers but mutated PAMs. For Cascade, we compared four PAMs with affinities following the order AA > TT > AG >> CC. Although we observed efficient R–loop formation for the TT and AG PAMs, higher negative turns/torque were required compared with the AA PAM (Figure 25 A and B). In contrast, the torque-dependent mean dissociation times of both substeps were unchanged within error, suggesting that the R-loops are equally stable once formed. Following the same protocol using the CC PAM, we were unable to observe any R-loops. However, R-loops could be induced with moderate efficiencies at

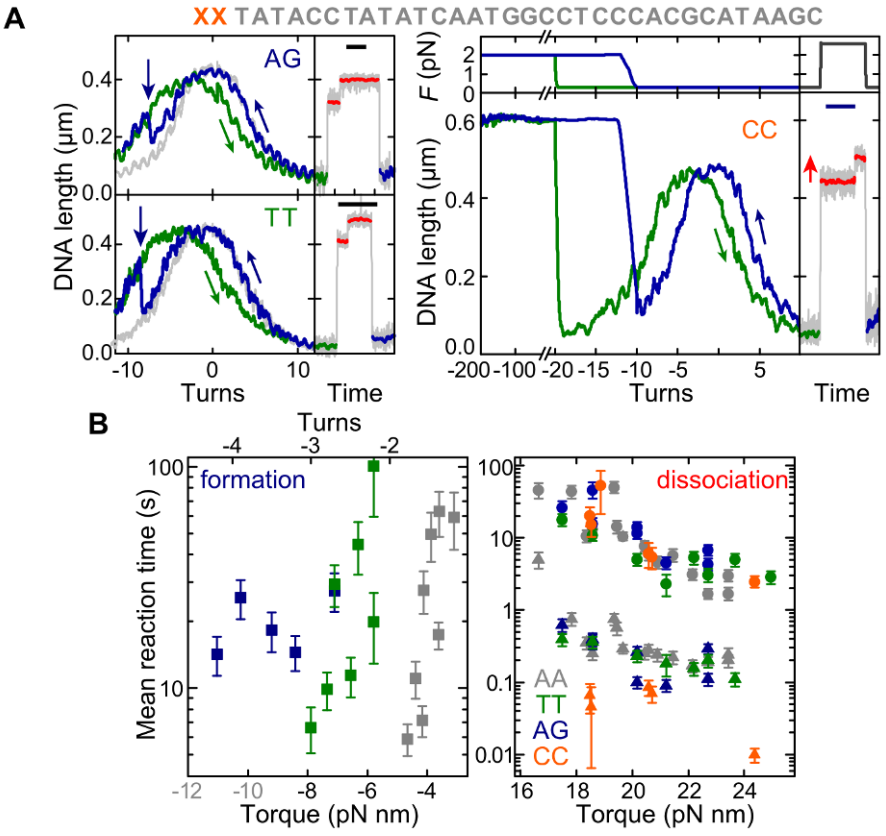

Figure 25. PAM mutations affect primarily R-loop formation and not dissociation. (**A**) Repetitive R-loop cycles in the presence of Cascade on DNA with the matching protospacer S1 but modified dinucleotide PAMs (see sketch above and labels in the graphs). Curve coloring and experimental conditions are as in Figure 23 C and D. For the CC PAM, R-loop formation requires −200 turns (at 2.0 pN; blue curve). The shift in the probe curve (green) and the dissociation step (red curve) reveals the presence of the R-loop. (Scale bars: 10 s.) (**B**) Mean R-loop formation and dissociation times as a function of torque for the different PAMs (colors given in the key). Circles and triangles indicate the first and second dissociation step, respectively. Gray torque values indicate the phase where the torque is no longer proportional to turns.

elevated forces and high negative twist [35% of all attempts for −200 turns (n = 51); 19% for −100 turns (n = 27); 0% at −10 turns (n = 78)], conditions that mechanically cause extensive DNA denaturation. R-loops for the CC PAM were also stable. While the torque-dependent mean time for the first dissociation step of the CC DNA was indistinguishable from the other PAMs, the mean time for the second step was significantly reduced (Figure 25 B). Thus, PAM regulates R-loop formation by kinetic inhibition.

2.3.2.4. DNA-helix destabilization by Cascade

To further explore the role of the PAM during target site search and priming of the R-loop, we carried out supercoiling experiments on DNA without a matching protospacer (but with multiple orphan PAMs; Figure 26). Interaction with PAMs may result in changes in DNA structure as the CRISPR enzymes probe the adjacent protospacer for complementarity. In line with this, Cascade globally destabilized the DNA helical structure, an effect that required the bound Cse1 subunit (Figure 27). The observed behavior agrees only with DNA destabilization (i.e., lowering of the melting temperature) but not with active DNA helix distortion, as seen for DNA intercalators, where stable supercoiling changes are observed (Gunther et al., 2010; Lipfert et al., 2010).

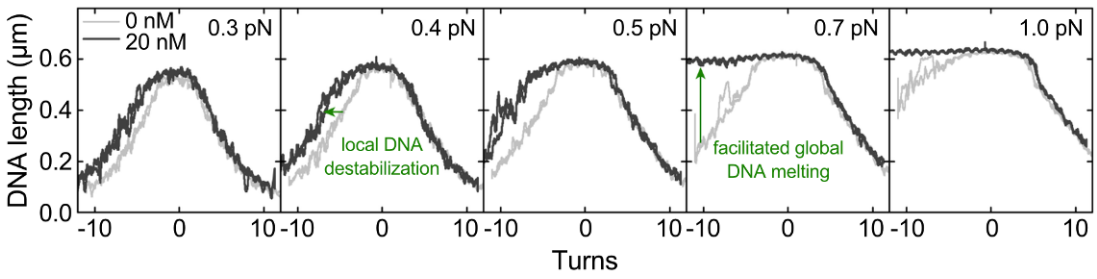

Figure 26. DNA helix destabilization by Cascade. Supercoiling curves at different forces using DNA with a non-matching protospacer in the presence or absence of the Cascade. Rotation curves in the absence of Cascade are shown in light grey. In the presence of Cascade (black curves) the negative side of the rotation curve is shifted to lower turns and the force for B-form DNA melting is lowered in a concentration dependent manner. This behaviour is in agreement with a destabilization of the DNA structure due to Cascade.

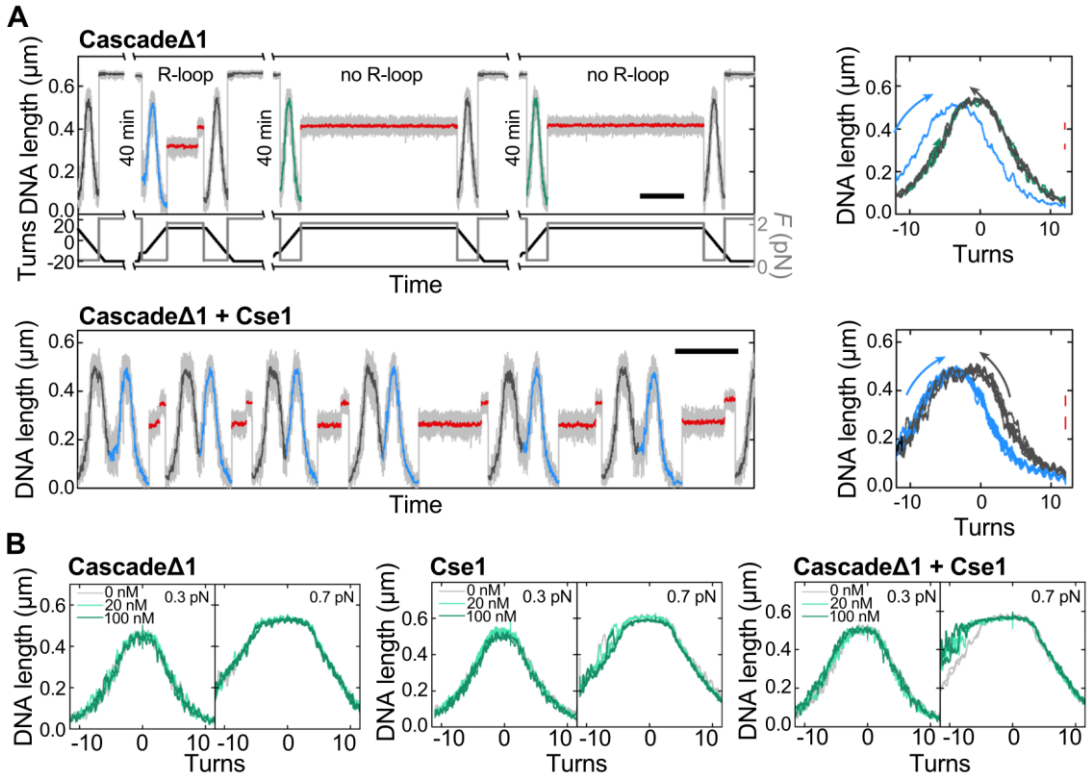

Figure 27. Characterization of Cascade lacking the Cse1 subunit. (**A**) R-loop cycle experiments using DNA with matching protospacer in the presence of 9 nM CascadeΔ1 (upper) and 9 nM CascadeΔ1 + 9 nM Cse1 (lower). R-loop induction traces are shown in dark grey, probe traces with and without R-loops having been formed are shown in blue and green respectively, while dissociation traces are shown in red. CascadeΔ1 can form R-loops with 25% efficiency (N = 44) with high negative torsion (>2 pN force, -20 to -100 turns, no clear influence of turn number observed) and elevated waiting times (40 min). When adding CascadeΔ1 mixed with an equal amount of Cse1, efficient R-loop formation similar to WT protein was restored. This provides evidence for the integrity of the CascadeΔ1 and Cse1 complex. (**B**) Supercoiling curves at 0.3 and 0.7 pN forces using DNA with a non-matching protospacer in the presence of different concentrations of CascadeΔ1, Cse1 and reconstituted Cascade. For CascadeΔ1 and Cse1 alone, no curve broadening indicative of DNA helix destabilization was observed, even in the presence of 100 nM protein. The reconstituted complex exhibits a facilitated DNA denaturation at 0.7 pN. This suggests that Cse1 plays an important role in helix destabilization. Nonetheless the remainder of the Cascade complex is also required for helix destabilization. Possibly, the latter ensures sufficient non-specific DNA-binding by the full complex.

Insertion of aromatic amino acid residues of the Cse1 subunit into the DNA helix (Sashital et al., 2012) may be responsible for the helix destabilization but will be a transient rather than a stably bound state. The destabilization may also cause the PAM dependence of the second dissociation step for Cascade (Figure 25 B).

2.3.2.5. Protospacer end truncations alter R-Loop stability and reveal a conformational lock for Cascade

The PAM seems to exclusively control R-loop formation, while not affecting R-loop stability once formed. To further test a unidirectional R-loop formation mechanism, we determined the effect of protospacer truncations at the end opposite from the PAM (Figure 28 A) on the dynamics of R-loop formation. For Cascade, R-loops formed efficiently at low negative twist on all substrates truncated by up to 10 bp (as seen by the shift of the left side of the supercoiling curve; Figure 28 B). The size of the shifts decreased with increasing truncation, in line with a decrease in R-loop size correlated with complementarity (Figure 28 B). For truncations of 6 bp and above, R-loops were formed but were unstable and dissociated as soon as low positive turns at

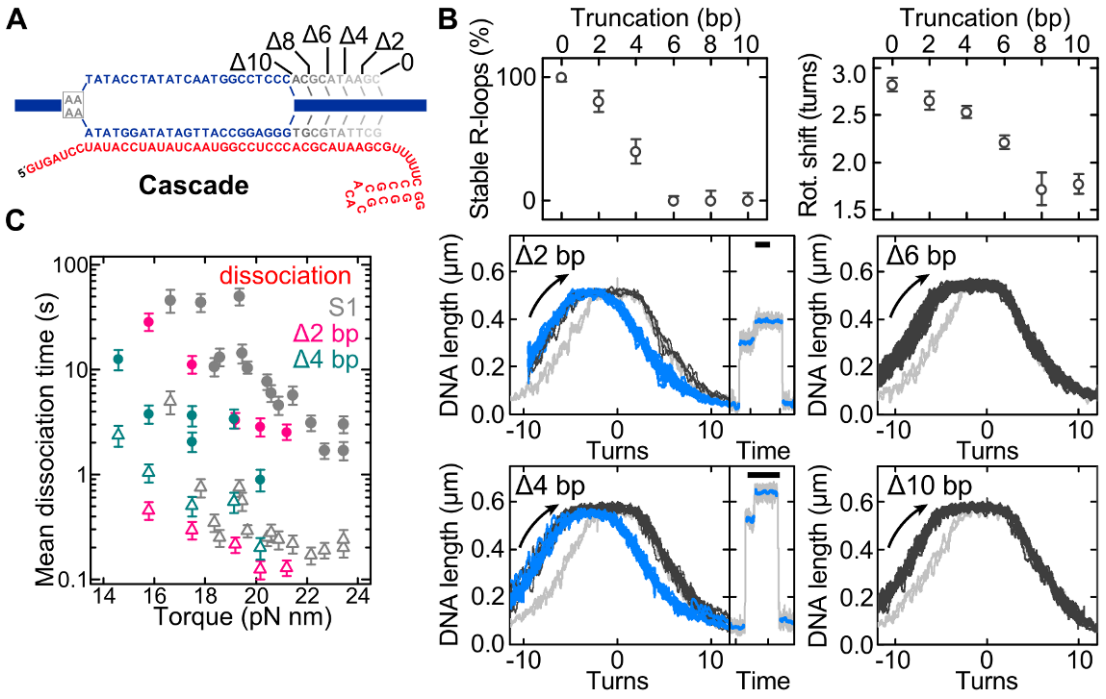

Figure 28. Protospacer end truncations affect the R-loop stability and reveal a locking mechanism for Cascade. (**A**) The protospacer-1 was truncated 2, 4, 6, 8 or 10 bp from the PAM-distal end. (**B**) R-loop probe and dissociation curves in the presence of Cascade for protospacer-1 truncated by 2, 4, 6, and 10 bp, revealing stable R-loops (blue) and unstable R-loops (dark gray). Percentages of stable R-loops and the sizes of the rotational shifts are shown in the upper graphs. (**C**) Mean dissociation times as function of torque (from experiments in B) for the stable R-loops of 0-, 2-, and 4-bp truncations. Filled circles and open triangles indicate the first and second dissociation step, respectively.

low force were reached; thus, dissociation rates could not be measured (Figure 28 B). For truncations of 2 and 4 bp, both stable and unstable R-loops were observed (Figure 28 B). By increasing the experimental waiting time at negative twist during R-loop induction for the 4-bp truncation, the proportion of stable R-loops increased when subsequently probed (Data not shown), suggesting that stable R-loops originate from an unstable R-loop intermediate. The transition from the unstable to stable R-loop thus locks the R-loop. For the stable R-loops, the stability relative to the full protospacer was only minimally reduced for the 2-bp truncation but was more significantly reduced for the 4-bp truncation, in particular for the first dissociation step (Figure 28 C).

2.3.2.6. Model for R-Loop formation and dissociation by Cascade.

Our data presented above show that the PAM controls tightly the R-loop formation kinetics but leaves the R-loop stability practically unchanged. Thus, the PAM provides a kinetic rather than a thermodynamic control of R-loop formation. Distal protospacer mutations affected the R-loop stability but hardly altered the formation kinetics. This reveals a unidirectionality in the R-loop formation and dissociation cycle, which is additionally supported by the differential dependence of the first and second dissociation steps of Cascade on protospacer truncations and PAM mutations, respectively.

Firstly, the Cascade uses DNA distortion, guided by PAM binding, to accomplish homology search. Matching hydrogen bonding between the crRNA and the protospacer then leads to propagation of R-loop formation over the adjacent base pairs. Under unfavorable energetic conditions (high positive torque or mismatches between the protospacer and crRNA), R-loop dissociation occurs in a PAM-independent manner. The high torsional stability of Cascade is not attributable to a large energetic bias but rather to a ratchet-like asymmetry in the energy landscape. This increased stability seems to be achieved by an extra “locking” step after most of the R-loop has been formed, which was revealed by the bistability of R-loops for a 4-bp truncated protospacer and the two-step process for R-loop dissociation. We suggest that

the locking is attributable to domain reorganization within Cascade, such as movement of Cse2 subunits stabilizing the free, non-target DNA strand (Mulepati et al., 2014; Wiedenheft et al., 2011a). The locking step represents an additional proofreading that ensures that complementarity between the crRNA and protospacer leads to complete R-loop expansion.

2.4. Cas3 degrades Cascade-targeted DNA

2.4.1. Cascade binding to the protospacer triggers Cas3 ATPase activity

Cascade complex binding to the protospacer creates an R-loop (Figure 21 B) where a non-target strand is displaced as ssDNA and may function as a docking site for Cas3. We used a colorimetric assay to monitor Cas3 ATPase activity in the presence of Cascade and DNA. The Cas3 protein was mixed with the Cascade complex and pUC19 plasmid variants that either contain or lack protospacer-1 in the context of the correct or mutated PAM (Table 2), and ATPase reactions were initiated by addition of ATP and Mg^{2+} ions. The assay revealed that ATPase activity of Cas3 is triggered only in the presence of the pSP1-AA plasmid, containing a matching protospacer-1 and a correct PAM (Figure 29 A). Plasmids lacking protospacer-1 and PAM or containing a different protospacer (pSP3-AA) or incorrect CC PAM (pSP1-CC) did not stimulate Cas3 ATPase activity in the presence of Cascade complex. No ATPase activity was detected for the Cas3 ATPase-deficient mutant D452A with an impaired Walker B motif. Taken together, these results suggest a link between the ATPase activity and Cas3 docking on ssDNA formed upon Cascade binding to the matching protospacer flanked by a correct PAM.

2.4.2. Cascade binding to the protospacer triggers Cas3-mediated plasmid degradation

Consistent with published data on the *T. thermophilus* Cas3 protein (Mulepati & Bailey, 2011), *S. thermophilus* Cas3 degraded ss M13mp18 DNA

more rapidly in the presence of Ni^{2+} compared to Mg^{2+} ions (Figure 29 B). Since the ATPase activity of Cas3 is not supported by Ni^{2+} ions, a mixture of Mg^{2+} and Ni^{2+} ions that supports both ATPase/helicase and nuclease activities of Cas3 was used in further experiments.

To determine whether Cas3 docking on the ssDNA formed upon Cascade binding triggers nuclease activity, we analysed Cas3-mediated cleavage of plasmid DNA. pSP1-AA plasmid was pre-incubated with Cascade and an ATP

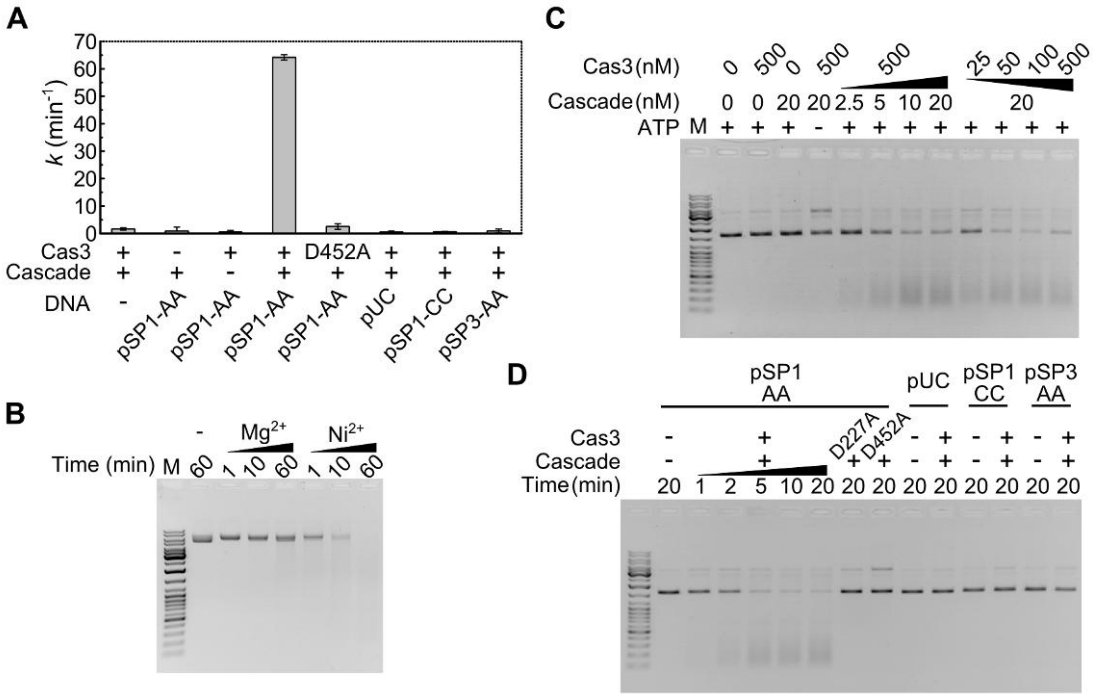

Figure 29. ATPase and nuclease activities of Cas3 induced by Cascade binding to the protospacer dsDNA. (**A**) ATP hydrolysis rates. Malachite green assay was used to measure ATP hydrolysis through the detection of free phosphate liberated from ATP. Reaction rate constant *k* values (min^{-1}) were calculated from linear slopes of time courses of phosphate liberation per Cas3 amount added. ATPase reactions were conducted at 37°C in the AT2 buffer supplemented with 3 nM supercoiled plasmids, 12 nM Cascade and 300 nM Cas3 or the ATPase-deficient mutant D452A. Error bars indicate the ±standard deviation for the rate constant *k* value determined in three separate experiments. (**B**) Nickel ions are better cofactor for Cas3 HD-domain than magnesium ions. Nuclease reactions were conducted at 37°C for indicated time intervals in the NB1 buffers containing 7 nM ss M13 DNA, 100 nM Cas3 and 5mM Mg^{2+} or Ni^{2+} ions. (**C**) dsDNA degradation requires Cas3, Cascade, and ATP. Nuclease reactions were conducted at 37°C for 10 min in the NB2 buffer supplemented with 5 nM pSP1-AA and indicated amounts of Cas3 and Cascade. (**D**) PAM and a protospacer are essential for DNA degradation. Nuclease reactions were conducted at 37°C for indicated time intervals in the NB2 buffer supplemented with 100 nM Cas3, 20 nM Cascade, and 5 nM of respective supercoiled plasmids.

solution containing Mg^{2+} and Ni^{2+} ions, followed by addition of Cas3. Under these conditions, the pSP1-AA plasmid was degraded in a Cascade and Cas3 concentration- and time-dependent manner (Figure 29 C and D). On the other hand, the pUC19 plasmid, or plasmids containing a non-matching protospacer (pSP3-AA) or a defective PAM (pSP1-CC), was resistant to Cas3 cleavage (Figure 29 D). ATP hydrolysis was required for pSP1-AA plasmid degradation. In the absence of ATP, the supercoiled pSP1-AA plasmid was converted into a nicked form but not degraded (Figure 29 C). The identical cleavage pattern was observed for the Cas3 ATPase-deficient mutant D452A in the presence of ATP (Figure 29 D). In contrast, the D227A replacement in the nuclease active site abolished DNA cleavage activity. Taken together, these data suggest that both ATPase/helicase and nuclease activities of Cas3 are required for pSP1-AA plasmid degradation in the presence of Cascade.

2.4.3. PAM and stable R-loop is required for Cas3-mediated plasmid degradation

DNA binding studies revealed that Cascade binding to a matching protospacer sequence requires a correct PAM sequence (Figures 21 C and 25 A). To check whether plasmid DNA cleavage in the *in vitro* reconstituted interference system follows the same dependence on PAM, we engineered plasmid substrates containing all possible combinations of base pairs at the −2 and −1 positions relative to the protospacer (within the predicted PAM), and monitored Cas3-mediated cleavage in the presence of Cascade and ATP (Figure 30 A). Consistent with previous binding assays, we found that plasmids containing A(−1), T(−1), or A(−2)G(−1) nucleotides upstream of protospacer-1 were efficiently degraded, while plasmids with B(−2)G(−1) (where B=T or C or G) or C(−1) sequences were resistant to cleavage. Altogether, DNA binding and cleavage experiments demonstrate that protospacer recognition by Cascade is PAM dependent, and that subsequent R-loop formation triggers dsDNA degradation by Cas3.

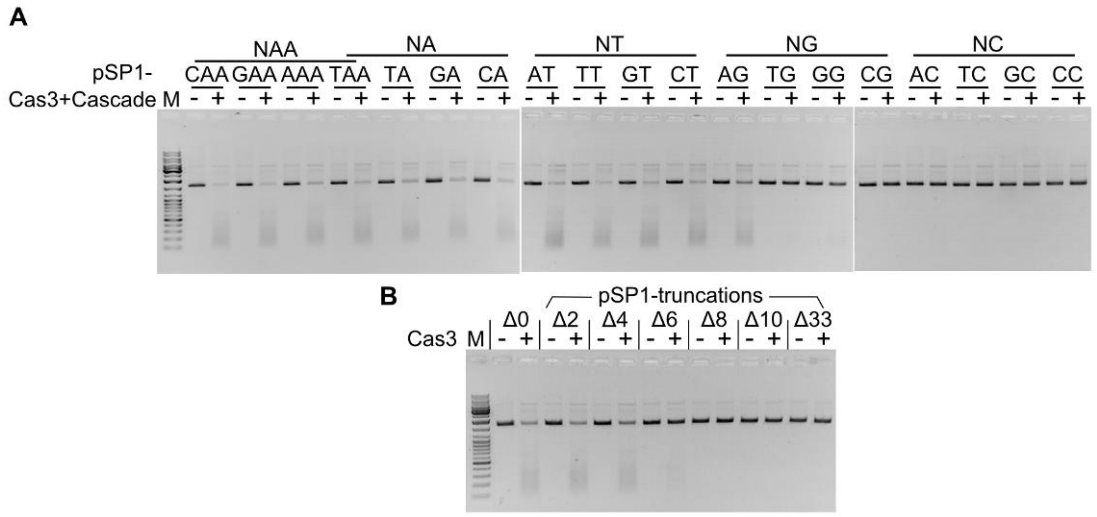

Figure 30. DNA degradation assay in the presence of Cascade and Cas3 using different PAM variants and protospacer truncations. (**A**) DNA degradation assay in the presence of Cascade and Cas3 using different PAM variant. Plasmid substrates containing A(-1), T(-1) or A(-2)G(-1) PAM were degraded by Cas3 in the presence of Cascade. This is in agreement with Cascade binding experiments (see Figure 21 C). Nuclease reactions were conducted at 37°C for 10 min in the NB2 buffer containing 5 nM supercoiled plasmid, in the absence (-) or presence (+) of 20 nM Cascade and 100 nM Cas3. (**B**) DNA degradation assay in the presence of Cascade and Cas3 using different end-truncations of the protospacer-1. Δ0 and Δ33 indicate, respectively, matching and non-matching protospacers for Cascade. Significant DNA degradation was only seen with the full-length protospacer and with the protospacers end-truncated by 2 or 4 bp. Stable R-loops were only detected in the single-molecule experiments using these substrates (see Figure 28 C). Minor DNA degradation is observed for the protospacer end-truncated by 6 bp, whereas measurable DNA degradation was not detected for larger end-truncations. Using these substrates only, unstable R-loops were detected (see Figure 28 B).

Magnetic tweezers experiments showed that Cascde forms stable R-loops when the guide completely basepairs with the protospacer (Figure 28). We examined whether truncations at a PAM-distal end of the protospacer-1 influence degradation of plasmid substrates bearing these protospacer modifications. Plasmid with 2 bp protospacer truncation was degraded similarly as plasmid with a full length protospacer. Truncations of 4 and 6 bp markedly reduced cleavage rates, while substrates with more than 6 bp truncated protospacers were not degraded (Figure 30 B). In agreement with single molecule experiments, a locked-and-stable R-loop is the prerequisite for Cas3 recruitment and subsequent DNA degradation.

2.4.4. Cas3 cleaves DNA within the protospacer and upstream of the PAM

To map Cas3 cleavage sites, oligoduplexes SP1-AA, SP1-CC, and SP3-AA were ^{33}P-5′-end-labelled either on the target or non-target strand, and Cas3-induced cleavage was assessed on each strand of each duplex in the absence or presence of ATP. In the absence of ATP, only a non-target DNA strand of the SP1-AA oligoduplex is cut in the protospacer region, while the target DNA strand is resistant to cleavage (Figure 31 A and C). In contrast, in the presence of ATP, both target and non-target strands of the SP1-AA oligoduplex are cleaved at multiple positions (Figure 31 B and D). The non-target strand is extensively cut within the protospacer and upstream of the PAM at the 5′-end proximal region. The target strand is extensively cleaved within the protospacer with minor cuts occurring at both the 5′- and 3′-proximal termini (Figure 31 D). Consistent with plasmid DNA cleavage data, no Cas3-mediated cleavage was observed for the oligoduplex lacking a protospacer (SP3-AA) or with a mutated PAM sequence (SP1-CC), neither in the presence or absence of ATP. Furthermore, the nuclease-deficient mutant D227A did not cleave the SP1-AA oligoduplex. In contrast, ATPase-deficient D452A mutant cleaved only the non-target strand, in both the absence and presence of ATP (Figure 31 A and B). The cleavage pattern of the SP1-AA oligoduplex explains why the nicked DNA form is a major product during plasmid DNA cleavage by the D452 mutant or WT Cas3 in the absence of ATP. Interestingly, almost all cleavage sites are located at the 3′-end of pyrimidine (T and C) bases (Figure 31 C and D). Preference for pyrimidine bases is also characteristic for the Cas3 cleavage of single-stranded oligodeoxynucleotides (data not shown). The non-target strand cleavage at the 5′-end proximal side of the oligoduplex suggests subsequent Cas3 translocation in the 3′→5′ direction, followed by DNA degradation.

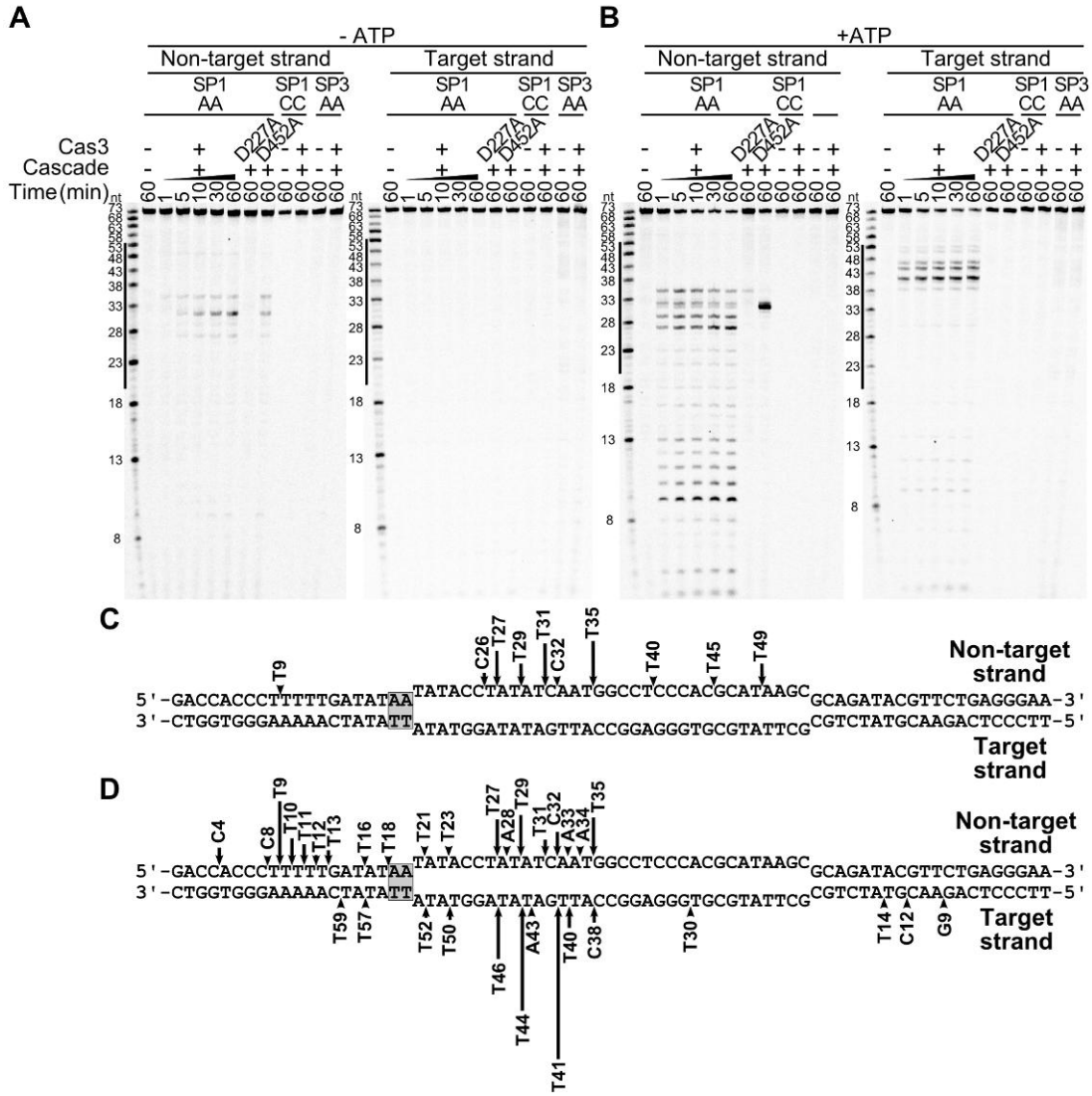

Figure 31. Cas3 cleavage of Cascade bound to target dsDNA. Oligoduplexes ^{33}P-labelled in either the non-target or target strand were pre-incubated with Cascade without (**A**) or with (**B**) ATP and reaction products analysed in denaturating polyacrylamide gels and mapped on the SP1-AA oligoduplex sequence (**C, D**), respectively. Cleavage reactions were conducted at 37°C in the NB2 buffer containing 0 mM (A) or 2 mM (B) ATP, 8 nM Cascade and 100 nM (B) or 500 nM (A) Cas3 or D227A and D452A mutants supplemented with 2 nM SP1-AA, SP1-CC, or SP3-AA ^{33}P-labelled oligoduplexes. Solid lines designate protospacer boundaries. Arrows indicate cleavage positions, height of the arrow correlates with a relative amount of cleavage product after 10 min incubation.

2.4.5. DNA degradation by Cas3 in the St-CRISPR4-Cas system is unidirectional

To determine whether the Cas3-mediated DNA cleavage is directional, we linearized pSP1-AA plasmid using four different restriction endonucleases (*Xap*I, *Bam*HI, *Pdm*I, *Alw*NI) to generate a set of linear dsDNA molecules of identical length that have a protospacer sequence located at different distances with respect to DNA termini (Figure 32 A). In the pSP1-AA-*Xap*I DNA, the protospacer is located almost at the 5′-end, while in the pSP1-AA-*Pdm*I DNA

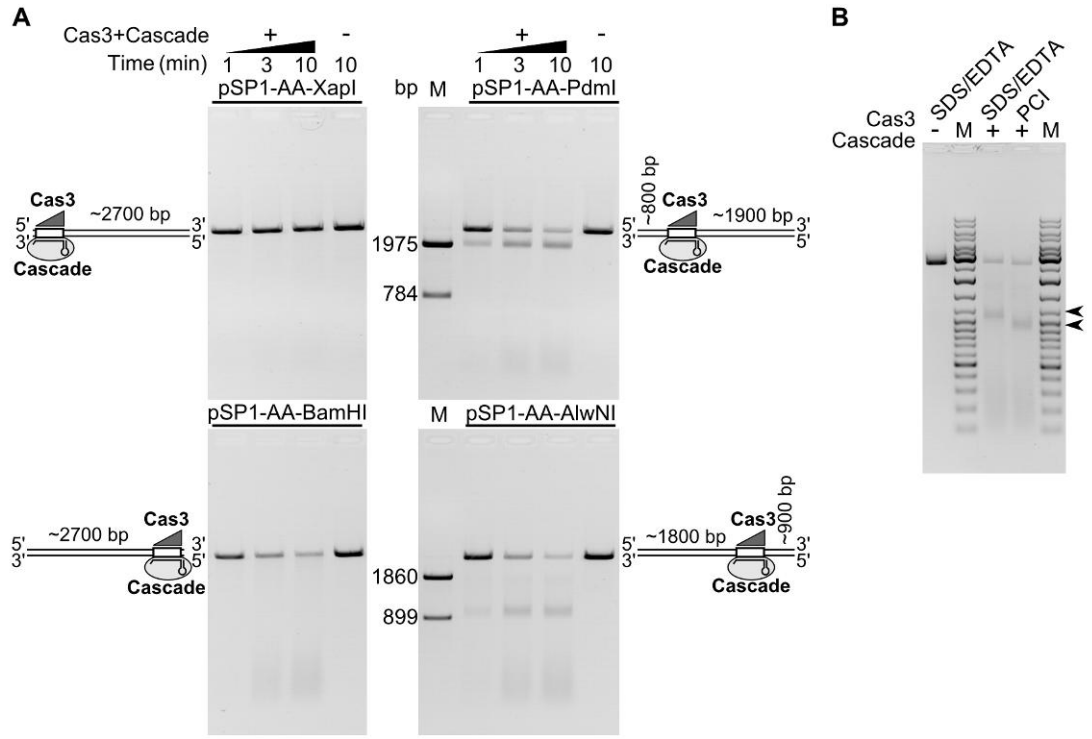

Figure 32. Cas3-mediated DNA degradation is unidirectional. (**A**) Four linear 2759 bp DNA fragments pSP1-AA-*Xap*I, pSP1-AA-*Pdm*I, pSP1-AA-*Bam*HI or pSP1-AA-*Alw*NI that contain a protospacer sequence located at different distance in respect to DNA ends were generated by a restriction endonuclease cleavage of pSP1-AA plasmid. Nuclease reactions were initiated by addition of 20 nM Cascade and 100 nM Cas3 into a NB2 buffer containing 5 nM of respective DNA. Linear DNA molecules with Cas3 and Cascade bound to a proto-spacer are schematically depicted at respective gel picture. M—DNA markers to map cleavage products were obtained by cleaving pSP1-AA with *Pdm*I, *Xap*I and *Alw*NI, *Xap*I restriction endonucleases, respectively. (**B**) CRISPR interference complex remains associated with proto spacer after reaction. Nuclease reactions were conducted as in (A) using pSP1-AA-*Alw*NI linearised plasmid DNA in the presence (+) or absence (-) of Cascade and Cas3. After 10 minutes reactions were stopped with "Stop" solution (SDS/EDTA) or PCI solution (PCI; see section 2.1.7.) that strips off proteins from the DNA. Black arrows indicate bands with altered migration. M – molecular marker.

it is located ~800-bp away from the 5′-terminus of the non-target strand. In two other linear DNA substrates, pSP1-AA-*Bam*HI and pSP1-AA-*Alw*NI, the protospacer sequence is located at the 3′-end or ~900-bp away from the 3′-end, respectively. Analysis of reaction products resulting from Cas3 cleavage in the *in vitro* reconstituted interference system revealed that the linear DNA pSP1-AA-*Xap*I remained nearly intact, while the pSP1-AA-*Bam*HI substrate was degraded in a similar fashion to the circular plasmid DNA. Furthermore, Cas3 acting on the pSP1-AA-*Pdm*I and pSP1-AA-*Alw*NI substrates produced defined ~1.9- and ~0.9-kb products, respectively, while the remaining DNA fragments were degraded into smaller fragments (Figure 32 A). Interestingly, the reduced mobility of the ~0.9-kb product resulting from the pSP1-AA-*Alw*NI cleavage (Figure 32 B) suggests that the Cascade (or Cacade-Cas3) complex remains bound to the cleaved DNA. Taken together, these data are consistent with a model in which Cas3 first makes a double-stranded break in a protospacer region (or in its immediate vicinity), and subsequently degrades DNA upstream of the protospacer in the 3′→5′ direction in respect to the non-target strand. Therefore, DNA downstream of the protospacer remains intact.

2.4.6. DNA cleavage by Cas3 in the Cascade-target DNA complex

Upon Cascade-mediated R-loop formation the crRNA and the complementary target DNA strand are engaged into a heteroduplex, and the non-target strand is displaced as an ssDNA. The R-loop formation is a key pre-requisite for the binding of Cas3 protein, which is an ssDNA nuclease that displays an ATPase/helicase activity stimulated by ssDNA. Cas3 alone does not interact with a dsDNA and therefore first requires Cascade binding to the dsDNA to generate the R-loop where the non-target strand of the protospacer is displaced as ssDNA and serves as a platform for the Cas3 loading. Indeed, the Cas3 ATPase and nuclease activities are triggered only when stable R-loop is formed by Cascade binding.

Our data are consistent with a mechanism proposed recently for the *E. coli* system where the Cse1 (CasA) subunit of *E. coli* Cascade recruits Cas3, which

subsequently catalyses nicking of target DNA through its HD-nuclease domain (Westra et al., 2012b). In the absence of ATP, the oligoduplex substrate is only weakly cleaved by Cas3 and cleavage is limited to the non-target strand, which is displaced as an ssDNA. Consistent with the oligoduplex cleavage pattern without ATP, plasmid DNA under these conditions is converted into a nicked form. In the presence of ATP, the cleavage pattern of both the oligoduplex and plasmid DNA is radically changed. Cas3 extensively cuts both DNA strands in the protospacer region of the oligoduplex and upstream of the PAM. The plasmid DNA in the presence of ATP is subsequently degraded by Cas3.

Our results also reveal that DNA degradation by the Cas3 nuclease is directional. After cleaving both DNA strands in the protospacer region, Cas3 further degrades DNA upstream of the protospacer in the 3′→5′ direction in respect to the non-target strand, while DNA downstream of the protospacer remains intact. Cas3 cleavage of the non-target strand in the 3′→5′ direction generates stretches of ssDNA on the target strand that can serve as a loading platform for the same or other Cas3 molecules, promoting further degradation. It is tempting to speculate that the Cas3 cleavage directionality may contribute to the mechanism of adaptive spacer acquisition proposed recently for the *E. coli* K12 CRISPR system (Datsenko et al., 2012; Swarts et al., 2012). According to this mechanism, the DNA strand, from which new spacers are selected, is largely determined by the priming protospacer orientation. We show here that Cascade binding to the matching protospacer sequence determines, which strand will be extruded into the R-loop and subjected to degradation in the 3′→5′ direction. In this way, the unidirectional DNA degradation by Cas3 may contribute to the selection of a specific DNA strand, from which new spacers are subsequently acquired (Datsenko et al., 2012; Swarts et al., 2012).

2.5. The mechanism of type I-E CRISPR-Cas interference

CRISPR-Cas systems have been categorized into three main types that differ by the structural organization and function of RNP complexes involved in crRNA-mediated silencing of foreign nucleic acids (Bhaya et al., 2011; Sorek et al., 2013; van der Oost et al., 2014; Westra et al., 2012a; Wiedenheft et al., 2012). In the type I-E CRISPR-Cas system from *E. coli*, a multisubunit RNP complex called Cascade binds to the target DNA and triggers a Cas3-mediated DNA interference (Brouns et al., 2008). The CRISPR4-Cas system of *S. thermophilus* DGCC7710 (Horvath & Barrangou, 2010) displays a similar structural organization to that of *E. coli*. To show that orthologous *S. thermophilus* Cas proteins assemble into a similar Cascade complex, we have isolated the Cascade and demonstrate that it recruits Cas3 to form a functional effector complex, which cleaves target DNA *in vitro*. Moreover, we show that mechanistically the process of DNA interference (Figure 33), provided by the CRISPR4-Cas system, can be dissected into target site recognition and cleavage stages, which are executed by Cascade and Cas3, respectively.

In the target recognition stage, foreign DNA is detected by Cascade surveillance complex (Jore et al., 2011). Two determinants define DNA as Cascade target: (i) a protospacer sequence that is complementary to a guide sequence of the crRNA and (ii) PAM motive that is a few bp sequence located adjacent to the protospacer. PAM functions as an indicator for non-self DNA sensing, and the absence of PAM determinant defines DNA as self (e.g. CRISPR region posesses guide-complementary spacers; however, these spacers are not targeted due to the lack of PAM sequence). In most cases, PAM sequences vary between different complexes (Mojica et al., 2009). It was shown that Cse1 protein of *E. coli* Cascade is involved in PAM sequence (5′-AWG-3′) recognition (Sashital et al., 2012). Here, we determined the PAM of *S. thermophilus* Cascade to be a promiscuous A(-1), T(-1) or A(-2)G(-1) sequence. Invaders (e.g. phages) mutate PAM sequences to overcome

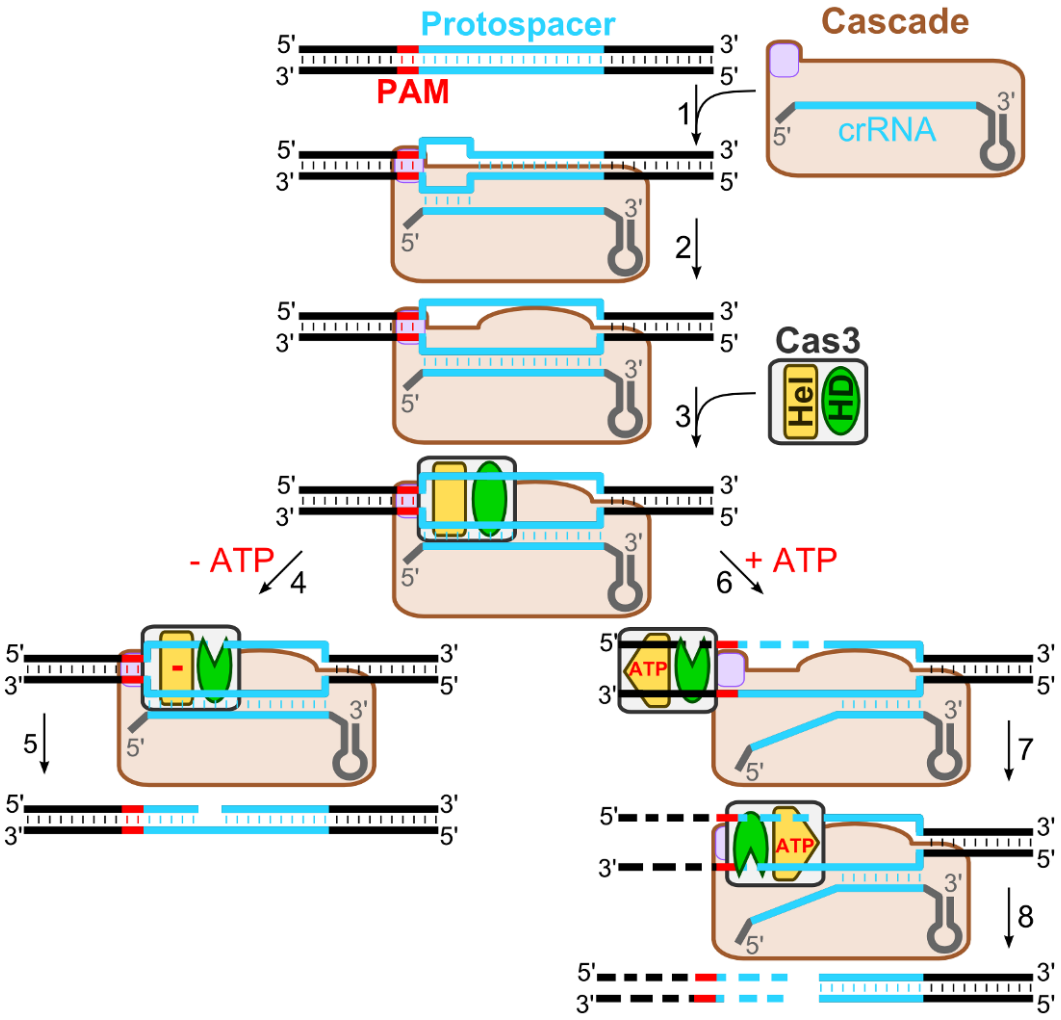

Figure 33. DNA-interference in the type I-E CRISPR-Cas systems. Cascade scans DNA for a protospacer sequence and PAM. Once the correct PAM and a short primary hybridization sequence ("seed") is identified (1), the guide of crRNA basepairs with a complementary DNA strand forming an R-loop, which is stabilized (locked) if a PAM-distal end of the protospacer is hybridized with the guide (2). Displaced DNA strand of the R-loop serves as a landing site for the Cas3 (3). In the absence of ATP, the Cas3 nuclease domain (HD) cleaves a displaced non-target strand within a protospacer (4) producing a nicked DNA (5). In the presence of ATP, Cas3 remodels the Cascade-DNA complex making both target and non-target strands available for the Cas3 cleavage within a protospacer sequence (6). Cas3 further translocates in the 3′→5′ direction powered by a helicase domain (Hel), whereas the HD domain degrades DNA (6; 7) in a unidirectional manner (8).

CRISPR-Cas interference (Samson et al., 2013); therefore, such short and promiscuous PAM sequence could be advantageous to adapt mutations of the invader.

We have shown that correct PAM sequence serves as a starting point for subsequent protospacer recognition through base-pairing with the guide

sequence, forming an R-loop. For R-loop formation by *E. coli* Cascade, a PAM-proximal 8-nt sequence, termed seed, is important (Semenova et al., 2011). Most likely, the seed is a starting point for DNA strand separation. If seed sequence base-pairs with the protospacer without mismatches, hybridization proceeds till the end of the guide where the protospacer is locked, forming stable R-loop. However, if more than 6 nucleotides at a PAM-distal end of the protospacer are mismatched then R-loop is unstable and prone to dissociate. Therefore, this additional checkpoint verifies whether the guide is completely hybridized with the protospacer. Structural data of *E. coli* Cascade (Mulepati et al., 2014) suggest that Cse2 protein may be responsible for locking of the R-loop.

In the target cleavage stage, the R-loop triggers Cas3 for target DNA destruction. Cas3 has the nuclease and ATPase active sites localized to HD and helicase domains, respectively. Stand-alone Cas3 protein cleaves ssDNA and hydrolyzes ATPase, while translocating on ssDNA in the 3′→5′ direction; however, dsDNA is not targeted by Cas3. In the Cascade-formed R-loop structure, target strand of the protospacer is duplexed with the guide, while the non-target strand is displaced as an ssDNA, which could be targeted by Cas3. Indeed, in the absence of ATP, the Cas3 manages to cut non-target strand of the dsDNA, bound to Cascade within the protospacer region, while both DNA strands are cleaved in the presence of ATP. Moreover, the cleavage is unidirectional. In agreement with our results, Cas3 of type I-E CRISPR-Cas system from *E. coli* docks on Cse1 subunit and hydrolyzes target DNA in a similar fashion to *S. thermophilus* CRISPR4-Cas (Hochstrasser et al., 2014; Mulepati & Bailey, 2013; Westra et al., 2012b). Furthermore, similar degradation pattern is seen in type I-A CRISPR-Cas system from *T. tenax* (Plagens et al., 2014).

In summary, Cascade recruits Cas3 to form a functional effector complex, which degrades target DNA *in vitro*. This establishes a molecular basis for CRISPR-based immunity in St-CRISPR4-Cas and other type I systems (Figure

33). Cascade, guided by the crRNA, locates the target DNA site and, if the correct PAM sequence is present, binds to the matching DNA strand, creating an R-loop. The PAM initiates formation of the R-loop, which is stabilized by locking the PAM-distal end of the protospacer. The R-loop serves as a loading site for the Cas3. Cas3 binding to the ssDNA triggers the ATPase/helicase activity that presumably contributes to Cascade remodelling, making both DNA strands in the protospacer region available for Cas3 cleavage. After cleaving both DNA strands within the protospacer, Cas3 translocates on the non-target strand in the 3′→5′ direction in an ATP-dependent manner and cleaves the translocating strand using its HD-nuclease domain.

This sets the stage for molecular exploitation of the Cas machinery for interference and DNA cleavage. Indeed, it was shown that type I-E CRISPR-Cas systems can be used in a programmable gene repression (Luo et al., 2014), removal of bacterial strains (Gomaa et al., 2014) or even engineering of a bacteriophage genome (Kiro et al., 2014).

CONCLUSIONS

1. *S. thermophilus* Cas3 protein is arranged of two domains that perform different functions: the HD domain cuts single-stranded DNA in a sequence-non-specific manner, while the helicase domain hydrolyzes ATP to unwind DNA duplex in a 3′→5′ direction.
2. *S. thermophilus* Cascade complex is comprised of Cse1, Cse2, Cas7, Cas5 and Cas6e proteins that associate with a 61-nt length crRNA molecule, containing a 33-nt spacer sequence flanked by 7-nt 5′-handle and 21-nt 3′-handle of the repeat sequence.
3. *S. thermophilus* Cascade complex functions as a surveillance complex that binds to the target DNA in a PAM-dependent manner.
4. Promiscuous double stranded A(-1), T(-1) or A(-2)G(-1) PAM sequence serves as an initiation point for the *S. thermophilus* Cascade binding and crRNA hybridization to the target strand that leads to a unidirectional formation of an R-loop structure.
5. Base pairing of crRNA with the protospacer sequence at the PAM-distal end locks the R-loop.
6. The R-loop structure recruits the Cas3 protein that degrades Cascade-bound DNA in the presence of ATP in a unidirectional manner starting from the protospacer.

LIST OF PUBLICATIONS

The thesis is based on the following original publications:

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

2. **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

3. Gasiunas G, **Sinkunas T**, Siksnys V (2014) Molecular mechanisms of CRISPR-mediated microbial immunity. Cell Mol Life Sci 71: 449-465

4. 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. Proc Natl Acad Sci U S A 111: 9798-9803

Book chapter:

Sinkunas T, Gasiunas G and Siksnys V. Cas3 nuclease-helicase activity assays. Methods in Molecular Biology: CRISPR, Methods and Protocols. [*In Press*]

CONFERENCE PRESENTATIONS

1. Seidel R, Tikhomirova M, **Sinkunas T**, Rutkauskas M, Szczelkun MD, Siksnys V. Direct observation of R-loop formation by single RNA-guided Cascade effector complex. Berlin-Brandenburg Academy of Sciences and Humanities. CRISPR 2014 - The Prokaryotic Immune System CRISPR/Cas, Berlin, Germany, 2014.05.14 – 16.

2. **Sinkunas T**, Gasiunas G, Waghmare SP, Dickman MJ, Barrangou R, Horvath P, Siksnys V. Mechanism of Type I-E CRISPR immunity in *Streptococcus thermophilus*. University of St Andrews. CRISPR: evolution, mechanisms and infection. St Andrews, United Kingdom, 2013.06.17-19.

3. Waghmare SP, Nwokeoji AO, Wiedenheft B, **Sinkunas T**, Brouns SJ, van der Oost J, Doudna JA, Siksnys V, Dickman MJ. Analysis of CRISPR RNA processing using RNA mass spectrometry. University of St Andrews. CRISPR: evolution, mechanisms and infection. St Andrews, United Kingdom, 2013.06.17-19.

4. **Šinkūnas T**. Bakterijų imuninės sistemos CRISPR/Cas baltymas Cas3: du viename. Lietuvos jaunųjų mokslininkų konferencija. Bioateitis: gamtos ir gyvybės mokslų perspektyvos. Lietuvos Mokslų Akademija, Vilnius 2011-12-07.

5. Siksnys V, **Sinkunas T**, Gasiunas G, Horvath P. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. FASEB Summer Research Conference. Helicases & Nucleic Acid Translocases: Stucture, Mechanism, Function, & Roles in Human Diseases. Steamboat Springs, Colorado, USA. 2011.07.31-08.05.

6. **Sinkunas T**, Gasiunas G, Horvath P, and Siksnys V. Cas3 of the *Streptococcus thermophilus* CRISPR4 system (Ecoli-subtype) is a multifunctional protein possessing nuclease and translocase/helicase activities. Wageningen University. CRISPR meeting: CRISPR Mechanisms & Applications, Wageningen, Netherlands, 2010.10.21-22.

FINANCIAL SUPPORT

The work was supported by the European Social Fund under the Global Grant measure (project R100). Scholarships for academic achievement were recieved from the Research Council of Lithuania.

ACKNOWLEDGEMENTS

I am grateful to my supervisor prof. dr. Virginijus Šikšnys for the oportunity to work in this interesting and dynamic field and also for invaluable discussions, suggestions and help with the preparation of this doctoral thesis.

I would like to thank dr Aušra Ražanskienė and prof. dr. Kęstutis Sasnauskas for teaching me the “alphabet” of work in the laboratory.

I thank to dr. Phillipe Horvath (DuPont) for genomic DNA of *S. thermophilus* DGCC7710 and discussions, Maria S. Tikhomirova and prof. dr. Ralf Seidel (University of Münster) for the magnetic tweezers experiments, dr. Sakharam P. Waghmare and dr. Mark J. Dickman (University of Sheffield) for the mass spectroscopy analyses and dr. Mark D. Szczelkun (University of Bristol) for the introduction with stopped flow experiments.

I am greatful to all my collegues in the Department of Protein - DNA Interactions, especially dr. Giedrius Gasiūnas, dr. Gintautas Tamulaitis, dr. Giedrius Sasnauskas, dr. Mindaugas Zaremba, dr. Giedrė Tamulaitienė and dr Arūnas Šilanskas for disscusions, advise and help with experiments.

I also thank to dr. Milda Plečkaitytė and dr. Jaunius Urbonavičius for the reviews for this doctoral thesis.

My greatest gratitude belongs to my parents and Jonita for understanding and support for reaching my goals.

REFERENCES

1. Achtman M, Kennedy N, Skurray R (1977) Cell--cell interactions in conjugating *Escherichia coli*: role of traT protein in surface exclusion. *Proc Natl Acad Sci U S A* 74**:** 5104-5108

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

3. Aravind L, Koonin EV (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem Sci* 23**:** 469-472

4. 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 Res* 42**:** 7884-7893

5. Arslan Z, Wurm R, Brener O, Ellinger P, Nagel-Steger L, Oesterhelt F, Schmitt L, Willbold D, Wagner R, Gohlke H, Smits SH, Pul U (2013) Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2. *Nucleic Acids Res* 41**:** 6347-6359

6. Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, Gagarinova A, Pogoutse O, Brown G, Binkowski A, Phanse S, Joachimiak A, Koonin EV, Savchenko A, Emili A, Greenblatt J, Edwards AM, Yakunin AF (2011) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol Microbiol* 79**:** 484-502

7. 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

8. Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M, Kochinyan S, Wang S, Chruszcz M, Minor W, Koonin EV, Edwards AM, Savchenko A, Yakunin AF (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J Biol Chem* 283**:** 20361-20371

9. 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. *EMBO J* 30**:** 4616-4627

10. 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. *Mol Cell* 56**:** 43-54

11. Bergh O, Borsheim KY, Bratbak G, Heldal M (1989) High abundance of viruses found in aquatic environments. *Nature* 340**:** 467-468

12. Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* 45**:** 273-297

13. Biswas T, Pero JM, Joseph CG, Tsodikov OV (2009) DNA-dependent ATPase activity of bacterial XPB helicases. *Biochemistry* 48**:** 2839-2848

14. Blower TR, Evans TJ, Przybilski R, Fineran PC, Salmond GP (2012) Viral evasion of a bacterial suicide system by RNA-based molecular mimicry enables infectious altruism. *PLoS Genet* 8**:** e1003023

15. Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD, Kulakauskas S, Lapidus A, Goltsman E, Mazur M, Pusch GD, Fonstein M, Overbeek R, Kyprides N, Purnelle B, Prozzi D, Ngui K, Masuy D, Hancy F, Burteau S, Boutry M, Delcour J, Goffeau A, Hols P (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* 22**:** 1554-1558

16. 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

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

18. Brendel J, Stoll B, Lange SJ, Sharma K, Lenz C, Stachler AE, Maier LK, Richter H, Nickel L, Schmitz RA, Randau L, Allers T, Urlaub H, Backofen R, Marchfelder A (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

19. 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

20. Brutzer H, Luzzietti N, Klaue D, Seidel R (2010) Energetics at the DNA supercoiling transition. *Biophys J* 98**:** 1267-1276

21. 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. *J Bacteriol* 194**:** 5728-5738

22. 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* 93**:** 98-112

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

24. 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

25. Chang N, Sun C, Gao L, Zhu D, Xu X, Zhu X, Xiong JW, Xi JJ (2013) Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res* 23**:** 465-472

26. Chen H, Choi J, Bailey S (2014) Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. *J Biol Chem* 289**:** 13284-13294

27. Chylinski K, Le Rhun A, Charpentier E (2013) The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol* 10**:** 726-737

28. Chylinski K, Makarova KS, Charpentier E, Koonin EV (2014) Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res* 42**:** 6091-6105

29. Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31**:** 230-232

30. Cocozaki AI, Ramia NF, Shao Y, Hale CR, Terns RM, Terns MP, Li H (2012) Structure of the Cmr2 subunit of the CRISPR-Cas RNA silencing complex. *Structure* 20**:** 545-553

31. Comas I, Homolka S, Niemann S, Gagneux S (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS One* 4**:** e7815

32. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339**:** 819-823

33. Cordin O, Banroques J, Tanner NK, Linder P (2006) The DEAD-box protein family of RNA helicases. *Gene* 367**:** 17-37

34. Court DL, Gan J, Liang YH, Shaw GX, Tropea JE, Costantino N, Waugh DS, Ji X (2013) RNase III: Genetics and function; structure and mechanism. *Annu Rev Genet* 47**:** 405-431

35. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14**:** 1188-1190

36. Cui Y, Li Y, Gorge O, Platonov ME, Yan Y, Guo Z, Pourcel C, Dentovskaya SV, Balakhonov SV, Wang X, Song Y, Anisimov AP, Vergnaud G, Yang R (2008) Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS One* 3**:** e2652

37. 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. *Nat Commun* 3**:** 945

38. Daume M, Plagens A, Randau L (2014) DNA binding properties of the small cascade subunit Csa5. *PLoS One* 9**:** e105716

39. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* 34**:** W362-365

40. 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

41. Deng L, Garrett RA, Shah SA, Peng X, She Q (2013) A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Mol Microbiol* 87**:** 1088-1099

42. Destoumieux-Garzon D, Duquesne S, Peduzzi J, Goulard C, Desmadril M, Letellier L, Rebuffat S, Boulanger P (2005) The iron-siderophore transporter FhuA is the receptor for the antimicrobial peptide microcin J25: role of the microcin Val11-Pro16 beta-hairpin region in the recognition mechanism. *Biochem J* 389**:** 869-876

43. 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*. *J Bacteriol* 190**:** 1390-1400

44. Deveau H, Labrie SJ, Chopin MC, Moineau S (2006) Biodiversity and classification of lactococcal phages. *Appl Environ Microbiol* 72**:** 4338-4346

45. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* 41**:** 4336-4343

46. Dickman MJ, Hornby DP (2006) Enrichment and analysis of RNA centered on ion pair reverse phase methodology. *RNA* 12**:** 691-696

47. Dy RL, Przybilski R, Semeijn K, Salmond GP, Fineran PC (2014a) A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic Acids Res* 42**:** 4590-4605

48. Dy RL, Richter C, Salmond GPC, Fineran PC (2014b) Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annual Review of Virology* 1**:** 307-331

49. Drozdz M, Piekarowicz A, Bujnicki JM, Radlinska M (2012) Novel non-specific DNA adenine methyltransferases. *Nucleic Acids Res* 40**:** 2119-2130

50. Dupuis ME, Villion M, Magadan AH, Moineau S (2013) CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat Commun* 4**:** 2087

51. Ebihara A, Yao M, Masui R, Tanaka I, Yokoyama S, 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 Sci* 15**:** 1494-1499

52. Ellinger P, Arslan Z, Wurm R, Tschapek B, MacKenzie C, Pfeffer K, Panjikar S, Wagner R, Schmitt L, Gohlke H, Pul U, Smits SH (2012) The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*. *J Struct Biol* 178**:** 350-362

53. Erdmann S, Garrett RA (2012) Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol Microbiol* 85**:** 1044-1056

54. Fairman-Williams ME, Guenther UP, Jankowsky E (2010) SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* 20**:** 313-324

55. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GP (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A* 106**:** 894-899

56. Fineran PC, Charpentier E (2012) Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* 434**:** 202-209

57. 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. *Proc Natl Acad Sci U S A* 111**:** E1629-1638

58. Fonfara I, Le Rhun A, Chylinski K, Makarova KS, Lecrivain AL, Bzdrenga J, Koonin EV, Charpentier E (2014) Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res* 42**:** 2577-2590

59. Forth S, Deufel C, Sheinin MY, Daniels B, Sethna JP, Wang MD (2008) Abrupt buckling transition observed during the plectoneme formation of individual DNA molecules. *Phys Rev Lett* 100**:** 148301

60. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 31**:** 822-826

61. Gan J, Shaw G, Tropea JE, Waugh DS, Court DL, Ji X (2008) A stepwise model for double-stranded RNA processing by ribonuclease III. *Mol Microbiol* 67**:** 143-154

62. 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

63. 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

64. 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

65. Gasiunas G, Siksnys V (2013) RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol* 21**:** 562-567

66. Georgiou T, Yu YN, Ekunwe S, Buttner MJ, Zuurmond A, Kraal B, Kleanthous C, Snyder L (1998) Specific peptide-activated proteolytic cleavage of *Escherichia coli* elongation factor Tu. *Proc Natl Acad Sci U S A* 95**:** 2891-2895

67. 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

68. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154**:** 442-451

69. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182**:** 319-326

70. Goldberg GW, Jiang W, Bikard D, Marraffini LA (2014) Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature*

71. Gomaa AA, Klumpe HE, Luo ML, Selle K, Barrangou R, Beisel CL (2014) Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *MBio* 5**:** e00928-00913

72. Goren MG, Yosef I, Auster O, Qimron U (2012) Experimental definition of a clustered regularly interspaced short palindromic duplicon in *Escherichia coli*. *J Mol Biol* 423**:** 14-16

73. 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

74. Gunther K, Mertig M, Seidel R (2010) Mechanical and structural properties of YOYO-1 complexed DNA. *Nucleic Acids Res* 38**:** 6526-6532

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

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

77. 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

78. Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover CV, 3rd, Graveley BR, Terns RM, Terns MP (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

79. 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

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

81. Hanlon GW, Denyer SP, Olliff CJ, Ibrahim LJ (2001) Reduction in exopolysaccharide viscosity as an aid to bacteriophage penetration through *Pseudomonas aeruginosa* biofilms. *Appl Environ Microbiol* 67**:** 2746-2753

82. 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

83. 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

84. 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

85. Haurwitz RE, Sternberg SH, Doudna JA (2012) Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *EMBO J* 31**:** 2824-2832

86. Heidrich N, Vogel J (2013) Same same but different: new structural insight into CRISPR-Cas complexes. *Mol Cell* 52**:** 4-7

87. Hynes WL, Hancock L, Ferretti JJ (1995) Analysis of a second bacteriophage hyaluronidase gene from *Streptococcus pyogenes*: evidence for a third hyaluronidase involved in extracellular enzymatic activity. *Infect Immun* 63**:** 3015-3020

88. Hyun M, Bohr VA, Ahn B (2008) Biochemical characterization of the WRN-1 RecQ helicase of *Caenorhabditis elegans*. *Biochemistry* 47**:** 7583-7593

89. 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

90. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327**:** 167-170

91. Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* 190**:** 1401-1412

92. Hoskisson PA, Smith MC (2007) Hypervariation and phase variation in the bacteriophage 'resistome'. *Curr Opin Microbiol* 10**:** 396-400

93. Howan K, Smith AJ, Westblade LF, Joly N, Grange W, Zorman S, Darst SA, Savery NJ, Strick TR (2012) Initiation of transcription-coupled repair characterized at single-molecule resolution. *Nature* 490**:** 431-434

94. 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 Biol* 11

95. 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 Biol* 10**:** 1670-1678

96. Hsu PD, Lander ES, Zhang F (2014) Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* 157**:** 1262-1278

97. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G,

Zhang F (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31**:** 827-832

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

99. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 31**:** 227-229

100. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* 169**:** 5429-5433

101. Yakunin AF, Proudfoot M, Kuznetsova E, Savchenko A, Brown G, Arrowsmith CH, Edwards AM (2004) The HD domain of the *Escherichia coli* tRNA nucleotidyltransferase has 2',3'-cyclic phosphodiesterase, 2'-nucleotidase, and phosphatase activities. *J Biol Chem* 279**:** 36819-36827

102. Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* 40**:** 5569-5576

103. Yosef I, Shitrit D, Goren MG, Burstein D, Pupko T, Qimron U (2013) DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array. *Proc Natl Acad Sci U S A* 110**:** 14396-14401

104. Young JC, Dill BD, Pan C, Hettich RL, Banfield JF, Shah M, Fremaux C, Horvath P, Barrangou R, Verberkmoes NC (2012) Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus*. *PLoS One* 7**:** e38077

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

106. Jackson RN, Lavin M, Carter J, Wiedenheft B (2014b) Fitting CRISPR-associated Cas3 into the Helicase Family Tree. *Curr Opin Struct Biol* 24**:** 106-114

107. 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

108. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31**:** 233-239

109. 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

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

111. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343**:** 1247997

112. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* 18**:** 529-536

113. Karow AR, Klostermeier D (2010) A structural model for the DEAD box helicase YxiN in solution: localization of the RNA binding domain. *J Mol Biol* 402**:** 629-637

114. Karvelis T, Gasiunas G, Miksys A, Barrangou R, Horvath P, Siksnys V (2013a) crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol* 10**:** 841-851

115. Karvelis T, Gasiunas G, Siksnys V (2013b) Programmable DNA cleavage *in vitro* by Cas9. *Biochem Soc Trans* 41**:** 1401-1406

116. Kauert DJ, Kurth T, Liedl T, Seidel R (2011) Direct mechanical measurements reveal the material properties of three-dimensional DNA origami. *Nano Lett* 11**:** 5558-5563

117. Kim TY, Shin M, Huynh Thi Yen L, Kim JS (2013) Crystal structure of Cas1 from *Archaeoglobus fulgidus* and characterization of its nucleolytic activity. *Biochem Biophys Res Commun* 441**:** 720-725

118. Kiro R, Shitrit D, Qimron U (2014) Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. *RNA Biol* 11**:** 42-44

119. Kyono K, Miyashiro M, Taguchi I (2003) Characterization of ATPase activity of a hepatitis C virus NS3 helicase domain, and analysis involving mercuric reagents. *J Biochem* 134**:** 505-511

120. Klaue D, Seidel R (2009) Torsional stiffness of single superparamagnetic microspheres in an external magnetic field. *Phys Rev Lett* 102**:** 028302

121. Koebnik R (1999) Structural and functional roles of the surface-exposed loops of the beta-barrel membrane protein OmpA from *Escherichia coli*. *J Bacteriol* 181**:** 3688-3694

122. 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

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

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

125. Labrie SJ, Samson JE, Moineau S (2010) Bacteriophage resistance mechanisms. *Nat Rev Microbiol* 8**:** 317-327

126. 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

127. 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*. *J Am Chem Soc* 135**:** 17476-17487

128. Lemak S, Nocek B, Beloglazova N, Skarina T, Flick R, Brown G, Joachimiak A, Savchenko A, Yakunin AF (2015) The CRISPR-associated Cas4 protein Pcal_0546 from *Pyrobaculum calidifontis* contains a [2Fe-2S] cluster: crystal structure and nuclease activity. *Nucleic Acids Res* 42**:** 11144-11155

129. Levitz R, Chapman D, Amitsur M, Green R, Snyder L, Kaufmann G (1990) The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease. *EMBO J* 9**:** 1383-1389

130. Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J (2013a) Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat Biotechnol* 31**:** 688-691

131. Li M, Liu H, Han J, Liu J, Wang R, Zhao D, Zhou J, Xiang H (2013b) Characterization of CRISPR RNA biogenesis and Cas6 cleavage-mediated inhibition of a provirus in the haloarchaeon *Haloferax mediterranei*. *J Bacteriol* 195**:** 867-875

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

133. 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. *Mol Microbiol* 72**:** 259-272

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

135. Lionnet T, Allemand JF, Revyakin A, Strick TR, Saleh OA, Bensimon D, Croquette V (2012) Magnetic trap construction. *Cold Spring Harb Protoc* 2012**:** 133-138

136. Lipfert J, Klijnhout S, Dekker NH (2010) Torsional sensing of small-molecule binding using magnetic tweezers. *Nucleic Acids Res* 38**:** 7122-7132

137. Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A, Maskell DJ, Simons RW, Cotter PA, Parkhill J, Miller JF (2002) Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. *Science* 295**:** 2091-2094

138. Loenen WA, Dryden DT, Raleigh EA, Wilson GG (2014a) Type I restriction enzymes and their relatives. *Nucleic Acids Res* 42**:** 20-44

139. Loenen WA, Dryden DT, Raleigh EA, Wilson GG, Murray NE (2014b) Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res* 42**:** 3-19

140. Loenen WA, Murray NE (1986) Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *J Mol Biol* 190**:** 11-22

141. Loenen WA, Raleigh EA (2014) The other face of restriction: modification-dependent enzymes. *Nucleic Acids Res* 42**:** 56-69

142. Lopez-Sanchez MJ, Sauvage E, Da Cunha V, Clermont D, Ratsima Hariniaina E, Gonzalez-Zorn B, Poyart C, Rosinski-Chupin I, Glaser P (2012) The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Mol Microbiol* 85**:** 1057-1071

143. Lu MJ, Stierhof YD, Henning U (1993) Location and unusual membrane topology of the immunity protein of the Escherichia coli phage T4. *J Virol* 67**:** 4905-4913

144. Luo ML, Mullis AS, Leenay RT, Beisel CL (2014) Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res*

145. Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA (2006) Structural basis for double-stranded RNA processing by Dicer. *Science* 311**:** 195-198

146. 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

147. Makarova KS, Anantharaman V, Aravind L, Koonin EV (2012) Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. *Biol Direct* 7**:** 40

148. Makarova KS, Aravind L, Grishin NV, Rogozin IB, Koonin EV (2002) A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res* 30**:** 482-496

149. 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

150. 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

151. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV (2011b) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9**:** 467-477

152. Makarova KS, Wolf YI, Koonin EV (2013) The basic building blocks and evolution of CRISPR-cas systems. *Biochem Soc Trans* 41**:** 1392-1400

153. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM (2013a) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* 31**:** 833-838

154. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013b) RNA-guided human genome engineering via Cas9. *Science* 339**:** 823-826

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

156. Matson SW (1986) Escherichia coli helicase II (urvD gene product) translocates unidirectionally in a 3' to 5' direction. *J Biol Chem* 261**:** 10169-10175

157. Meisel A, Bickle TA, Kruger DH, Schroeder C (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* 355**:** 467-469

158. 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

159. 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

160. Mosconi F, Allemand JF, Bensimon D, Croquette V (2009) Measurement of the torque on a single stretched and twisted DNA using magnetic tweezers. *Phys Rev Lett* 102**:** 078301

161. Mruk I, Kobayashi I (2014) To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res* 42**:** 70-86

162. Mulepati S, Bailey S (2011) Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). *J Biol Chem* 286**:** 31896-31903

163. 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

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

165. 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

166. 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. *J Biol Chem* 287**:** 35943-35952

167. 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

168. Nam KH, Kurinov I, Ke A (2011) Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca^{2+}-dependent double-stranded DNA binding activity. *J Biol Chem* 286**:** 30759-30768

169. 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

170. Niewoehner O, Jinek M, Doudna JA (2014) Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Res* 42**:** 1341-1353

171. 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

172. Niu Y, Shen B, Cui Y, Chen Y, Wang J, Wang L, Kang Y, Zhao X, Si W, Li W, Xiang AP, Zhou J, Guo X, Bi Y, Si C, Hu B, Dong G, Wang H, Zhou Z, Li T, Tan T, Pu X, Wang F, Ji S, Zhou Q, Huang X, Ji W, Sha J (2014) Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* 156**:** 836-843

173. 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. *Nat Struct Mol Biol* 21**:** 528-534

174. O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA (2014) Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature*

175. Oberstrass FC, Fernandes LE, Bryant Z (2012) Torque measurements reveal sequence-specific cooperative transitions in supercoiled DNA. *Proc Natl Acad Sci U S A* 109**:** 6106-6111

176. Osawa T, Inanaga H, Numata T (2013) Crystal structure of the Cmr2-Cmr3 subcomplex in the CRISPR-Cas RNA silencing effector complex. *J Mol Biol* 425**:** 3811-3823

177. Otto O, Czerwinski F, Gornall JL, Stober G, Oddershede LB, Seidel R, Keyser UF (2010) Real-time particle tracking at 10,000 fps using optical fiber illumination. *Opt Express* 18**:** 22722-22733

178. Paez-Espino D, Morovic W, Sun CL, Thomas BC, Ueda K, Stahl B, Barrangou R, Banfield JF (2013) Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* 4**:** 1430

179. Papadopoulos JS, Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23**:** 1073-1079

180. 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 Lett* 587**:** 562-568

181. Parma DH, Snyder M, Sobolevski S, Nawroz M, Brody E, Gold L (1992) The Rex system of bacteriophage lambda: tolerance and altruistic cell death. *Genes Dev* 6**:** 497-510

182. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 31**:** 839-843

183. 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

184. Pingoud A, Wilson GG, Wende W (2014) Type II restriction endonucleases - a historical perspective and more. *Nucleic Acids Res* 42**:** 7489-7527

185. 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

186. 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* 42**:** 5125-5138

187. Pleckaityte M, Zilnyte M, Zvirbliene A (2012) Insights into the CRISPR/Cas system of *Gardnerella vaginalis*. *BMC Microbiol* 12**:** 301

188. Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K (2010) Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol Microbiol* 77**:** 1367-1379

189. 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

190. Przybilski R, Richter C, Gristwood T, Clulow JS, Vercoe RB, Fineran PC (2011) Csy4 is responsible for CRISPR RNA processing in *Pectobacterium atrosepticum*. *RNA Biol* 8**:** 517-528

191. 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

192. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152**:** 1173-1183

193. 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 Biol* 10**:** 762-769

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

195. 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. *Biochem J* 452**:** 223-230

196. Revyakin A, Ebright RH, Strick TR (2005) Single-molecule DNA nanomanipulation: improved resolution through use of shorter DNA fragments. *Nat Methods* 2**:** 127-138

197. 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 Res* 42**:** 8516-8526

198. 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 Res* 40**:** 9887-9896

199. Riede I, Eschbach ML (1986) Evidence that TraT interacts with OmpA of *Escherichia coli*. *FEBS Lett* 205**:** 241-245

200. 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. *Mol Cell* 52**:** 124-134

201. Sakamoto K, Agari Y, Agari K, Yokoyama S, Kuramitsu S, Shinkai A (2009) X-ray crystal structure of a CRISPR-associated RAMP superfamily protein, Cmr5, from *Thermus thermophilus* HB8. *Proteins* 75**:** 528-532

202. Samai P, Smith P, Shuman S (2010) Structure of a CRISPR-associated protein Cas2 from *Desulfovibrio vulgaris*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 66**:** 1552-1556

203. Sambrook J (1989) Molecular Cloning: A Laboratory Manual. *Cold Spring Harbor Laboratory Press*.

204. Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS (2013) A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* 497**:** 254-257

205. Samson JE, Magadan AH, Sabri M, Moineau S (2013) Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* 11**:** 675-687

206. Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 32**:** 347-355

207. Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* 39**:** 9275-9282

208. Sashital DG, Jinek M, Doudna JA (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat Struct Mol Biol* 18**:** 680-687

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

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

211. Schmitt CK, Kemp P, Molineux IJ (1991) Genes 1.2 and 10 of bacteriophages T3 and T7 determine the permeability lesions observed in infected cells of *Escherichia coli* expressing the F plasmid gene pifA. *J Bacteriol* 173**:** 6507-6514

212. Seed KD, Lazinski DW, Calderwood SB, Camilli A (2013) A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* 494**:** 489-491

213. Seidel R, Bloom JG, van Noort J, Dutta CF, Dekker NH, Firman K, Szczelkun MD, Dekker C (2005) Dynamics of initiation, termination and reinitiation of DNA translocation by the motor protein EcoR124I. *EMBO J* 24**:** 4188-4197

214. 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. *Proc Natl Acad Sci U S A* 108**:** 10098-10103

215. Shah SA, Erdmann S, Mojica FJ, Garrett RA (2013) Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol* 10**:** 891-899

216. Shao Y, Cocozaki AI, Ramia NF, Terns RM, Terns MP, Li H (2013) Structure of the Cmr2-Cmr3 subcomplex of the Cmr RNA silencing complex. *Structure* 21**:** 376-384

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

218. Sheng G, Zhao H, Wang J, Rao Y, Tian W, Swarts DC, van der Oost J, Patel DJ, Wang Y (2014) Structure-based cleavage mechanism of *Thermus thermophilus* Argonaute DNA guide strand-mediated DNA target cleavage. *Proc Natl Acad Sci U S A* 111**:** 652-657

219. Shmakov S, Savitskaya E, Semenova E, Logacheva MD, Datsenko KA, Severinov K (2014) Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res* 42**:** 5907-5916

220. Singleton MR, Dillingham MS, Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* 76**:** 23-50

221. Sokolowski RD, Graham S, White MF (2014) Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system. *Nucleic Acids Res* 42**:** 6532-6541

222. Sorek R, Lawrence CM, Wiedenheft B (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem* 82**:** 237-266

223. 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

224. Staals RH, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K, Masuda A, Dohmae N, Schaap PJ, Doudna JA, Heck AJ, Yonekura K, van der Oost J, Shinkai A (2013) Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell* 52**:** 135-145

225. Staals Raymond HJ, Zhu Y, Taylor David W, Kornfeld Jack E, Sharma K, Barendregt A, Koehorst Jasper J, Vlot M, Neupane N, Varossieau K, Sakamoto K, Suzuki T, Dohmae N, Yokoyama S, Schaap Peter J, Urlaub H, Heck Albert JR, Nogales E, Doudna Jennifer A, Shinkai A, van der Oost J (2014) RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of *Thermus thermophilus*. *Mol Cell* 56**:** 518-530

226. Sternberg SH, Haurwitz RE, Doudna JA (2012) Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* 18**:** 661-672

227. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507**:** 62-67

228. Stibitz S, Aaronson W, Monack D, Falkow S (1989) Phase variation in *Bordetella pertussis* by frameshift mutation in a gene for a novel two-component system. *Nature* 338**:** 266-269

229. Streiff MB, Iida S, Bickle TA (1987) Expression and proteolytic processing of the darA antirestriction gene product of bacteriophage P1. *Virology* 157**:** 167-171

230. Sturino JM, Klaenhammer TR (2006) Engineered bacteriophage-defence systems in bioprocessing. *Nat Rev Microbiol* 4**:** 395-404

231. Sun J, Jeon JH, Shin M, Shin HC, Oh BH, Kim JS (2014) Crystal structure and CRISPR RNA-binding site of the Cmr1 subunit of the Cmr interference complex. *Acta Crystallogr D Biol Crystallogr* 70**:** 535-543

232. Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, Snijders AP, Wang Y, Patel DJ, Berenguer J, Brouns SJ, van der Oost J (2014a) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507**:** 258-261

233. Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, van der Oost J (2014b) The evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol* 21**:** 743-753

234. Swarts DC, Mosterd C, van Passel MW, Brouns SJ (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One* 7**:** e35888

235. Tamulaitis G, Kazlauskiene M, Manakova E, Venclovas Č, Nwokeoji Alison O, Dickman Mark J, Horvath P, Siksnys V (2014) Programmable RNA Shredding by the Type III-A CRISPR-Cas System of *Streptococcus thermophilus*. *Mol Cell* 56**:** 506-517

236. Tamulaitis G, Mucke M, Siksnys V (2006) Biochemical and mutational analysis of EcoRII functional domains reveals evolutionary links between restriction enzymes. *FEBS Lett* 580**:** 1665-1671

237. Tamulaitis G, Zaremba M, Szczepanowski RH, Bochtler M, Siksnys V (2007) Nucleotide flipping by restriction enzymes analyzed by 2-aminopurine steady-state fluorescence. *Nucleic Acids Res* 35**:** 4792-4799

238. Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Huttenhofer A (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc Natl Acad Sci U S A* 99**:** 7536-7541

239. 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*. *Mol Microbiol* 55**:** 469-481

240. Tanner NK, Cordin O, Banroques J, Doere M, Linder P (2003) The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell* 11**:** 127-138

241. Tanner NK, Linder P (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell* 8**:** 251-262

242. 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

243. van Duijn E, Barbu IM, Barendregt A, Jore MM, Wiedenheft B, Lundgren M, Westra ER, Brouns SJ, Doudna JA, van der Oost J, Heck AJ (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*. *Mol Cell Proteomics* 11**:** 1430-1441

244. Vercoe RB, Chang JT, Dy RL, Taylor C, Gristwood T, Clulow JS, Richter C, Przybilski R, Pitman AR, Fineran PC (2013) Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial

genomes and expel or remodel pathogenicity islands. *PLoS Genet* 9**:** e1003454

245. Viswanathan P, Murphy K, Julien B, Garza AG, Kroos L (2007) Regulation of dev, an operon that includes genes essential for *Myxococcus xanthus* development and CRISPR-associated genes and repeats. *J Bacteriol* 189**:** 3738-3750

246. Waghmare SP, Pousinis P, Hornby DP, Dickman MJ (2009) Studying the mechanism of RNA separations using RNA chromatography and its application in the analysis of ribosomal RNA and RNA:RNA interactions. *J Chromatogr A* 1216**:** 1377-1382

247. Walkinshaw MD, Taylor P, Sturrock SS, Atanasiu C, Berge T, Henderson RM, Edwardson JM, Dryden DT (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol Cell* 9**:** 187-194

248. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153**:** 910-918

249. Wang R, Li H (2012) The mysterious RAMP proteins and their roles in small RNA-based immunity. *Protein Sci* 21**:** 463-470

250. 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

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

252. Westra ER, Swarts DC, Staals RH, Jore MM, Brouns SJ, van der Oost J (2012a) The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu Rev Genet* 46**:** 311-339

253. Westra ER, van Erp PB, Kunne T, Wong SP, Staals RH, Seegers CL, Bollen S, Jore MM, Semenova E, Severinov K, de Vos WM, Dame RT, de Vries R, Brouns SJ, van der Oost J (2012b) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* 46**:** 595-605

254. 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

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

256. Wiedenheft B, van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A, Westphal W, Heck AJ, Boekema EJ, Dickman MJ, Doudna JA (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

257. 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

258. Zhang Y, Heidrich N, Ampattu BJ, Gunderson CW, Seifert HS, Schoen C, Vogel J, Sontheimer EJ (2013) Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol Cell* 50**:** 488-503

259. 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

260. Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, Graham S, Reimann J, Cannone G, Liu H, Albers SV, Naismith JH, Spagnolo L, White MF (2012b) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* 45**:** 303-313

261. 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* 515**:** 147-150

262. Zhu X, Ye K (2012) Crystal structure of Cmr2 suggests a nucleotide cyclase-related enzyme in type III CRISPR-Cas systems. *FEBS Lett* 586**:** 939-945