VILNIUS UNIVERSITY

Giedrius Gasiūnas

Mechanism of DNA interference by Type II CRISPR/Cas systems

Doctoral dissertation

Physical science, biochemistry (04 P)

Vilnius, 2012

The work presented in this doctoral dissertation has been carried out at the Institute of Biotechnology, Vilnius University during 2007-2011.

Supervisor

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

VILNIAUS UNIVERSITETAS

Giedrius Gasiūnas

DNR interferencijos mechanizmas II tipo CRISPR/Cas sistemose

Daktaro disertacija

Fiziniai mokslai, biochemija (04 P)

Vilnius, 2012

Disertacija rengta 2007-2011 m. Vilniaus universiteto Biotechnologijos institute.

Mokslinis vadovas

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

CONTENTS

LIST OF ABBREVIATIONS

aa	amino acid
Abi	Abortive infection
AHT	anhydrotetracycline
Ap	ampycillin
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
ds	double-stranded
DTT	1,4-dithiothreitol
EDTA	ethylenedinitrilotetraacetic acid
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
kb	kilobase pair
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PAA	polyacrylamide
PAM	proto-spacer adjacent motif
PCR	polymerase chain reaction
PDB	protein data bank
RAMP	repeat associated mysterious protein
RM	restriction - modification
RNP	ribonucleoprotein
SAXS	small-angle X-ray scattering
SDS	sodium dodecyl sulfate
Sie	superinfection exclusion
ss	single-stranded
TALENs	transcription activator-like effector nucleases
TEMED	N,N,N’,N’-tetramethylethylenediamine
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
ZNFs	zinc finger nucleases
WT	wild type

INTRODUCTION

Bacteria and archaea rely on a diversity of defense systems that allow them to survive exposure to foreign genetic elements such as bacterial viruses (bacteriophages). In their natural habitats, bacteria have evolved a battery of defense mechanisms to prevent bacteriophage infection, including prevention of adsorption, blocking of DNA entry, or degradation of foreign nucleic acids (Labrie *et al*, 2010; Sturino & Klaenhammer, 2006). Recently, an adaptive prokaryotic immune system based on Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) was identified that provides acquired immunity against viruses and plasmids (Barrangou *et al*, 2007). CRISPR consists of arrays of short conserved repeat sequences interspaced by unique DNA sequences of similar size called spacers, which often originate from phage or plasmid DNA (Barrangou *et al*, 2007; Bolotin *et al*, 2005; Mojica *et al*, 2005). CRISPR arrays, together with *cas* (CRISPR-associated) genes form the CRISPR/Cas adaptive immune system.

CRISPR are widespread in the genomes of many bacteria and almost all archaea. There is strong evidence suggesting that CRISPR/Cas systems can move between distinct species in bacteria and archaea via horizontal gene transfer, however the mobility of CRISPR/Cas systems is yet to be demonstrated.

The CRISPR/Cas system functions by acquiring short pieces of foreign DNA (spacers) which are inserted into CRISPR region and provide immunity against subsequent exposures to phages and plasmids that carry matching sequences (Barrangou *et al*, 2007; Brouns *et al*, 2008). CRISPR repeat-spacer arrays are transcribed into long primary transcripts that are further processed into a set of short CRISPR RNAs (crRNAs) containing a conserved repeat fragment and a variable spacer sequence (a guide) complementary to the invading nucleic acid (Carte *et al*, 2008; Brouns *et al*, 2008; Hale *et al*, 2009).

crRNAs combine with Cas proteins to form an effector complex which recognizes the target sequence in the invasive nucleic acid by base pairing to the complementary strand (Jore *et al*, 2011) and induces sequence-specific cleavage, thereby preventing proliferation and propagation of foreign genetic elements.

CRISPR/Cas systems have been categorized into three main Types, based on core elements content and sequences (Makarova *et al*, 2011b). The structural organization and function of effector ribonucleoprotein (RNP) complexes involved in crRNA-mediated silencing of foreign nucleic acids differ between distinct CRISPR/Cas Types (Wiedenheft *et al*, 2012). In the Type I-E system, as exemplified by *Escherichia coli*, crRNAs are incorporated into a multisubunit effector complex called Cascade (CRISPR-associated complex for antiviral defence) (Brouns *et al*, 2008), which binds to the target DNA and triggers degradation by the signature Cas3 protein (Sinkunas *et al*, 2011; Beloglazova *et al*, 2011). In Type III CRISPR/Cas systems of *Sulfolobus solfataricus* and *Pyrococcus furiosus*, Cas RAMP module (Cmr) and crRNA complex recognize and cleave synthetic RNA *in vitro* (Hale *et al*, 2012; Zhang *et al*, 2012) while the CRISPR/Cas system of *Staphylococcus epidermidis* targets DNA *in vivo* (Marraffini & Sontheimer, 2008). The structure and composition of ribonucleoprotein complexes involved in DNA silencing by Type II CRISPR/Cas systems remain to be established.

In this study we have focused on the understanding the mechanism of Type II CRISPR/Cas system from *Streptococcus thermophilus*. The CRISPR3/Cas system of *S. thermophilus* DGCC7710 strain (Horvath & Barrangou, 2010) consists of four *cas* genes: *cas9*, *cas1*, *cas2*, and *csn2*, that are located upstream of 12 repeat-spacer units. *cas9* is the signature gene for Type II systems (Makarova *et al*, 2011b). In the closely related *S. thermophilus* CRISPR1/Cas system, disruption of *cas9* abolishes crRNA-mediated DNA interference (Barrangou *et al*, 2007).

The specific aims of this study were:

(i). To perform genetic and biochemical analyses of the *S. thermophilus* CRISPR3/Cas system cloned and expressed in the heterologous *E. coli* host.

(ii). To identify the effector complex of Type II CRISPR3/Cas system and determine the mechanism of DNA interference.

Scientific novelty. We show here for the first time that the CRISPR3/Cas system of the Gram-positive *S. thermophilus* species can be cloned into a plasmid vector and transferred to a Gram-negative *E. coli* host to provide protection against plasmid transformation and phage infection. We further demonstrate that *cas9* is the sole *cas* gene necessary for CRISPR3-encoded interference. We provide first experimental evidence that the Cas9-crRNA complex *in vitro* introduces a double-stranded break at the specific site in DNA containing a sequence complementary to crRNA. We demonstrate that DNA cleavage is executed by Cas9, which uses two distinct active sites, RuvC and HNH, to generate site-specific nicks on opposite DNA strands. Our results show that the Cas9-crRNA complex functions as an RNA-guided endonuclease with RNA-directed target sequence recognition and protein-mediated DNA cleavage. These findings pave the way for the engineering of universal, programmable RNA-guided DNA endonucleases.

Practical value. We demonstrate that functionally active CRISPR/Cas systems can be transferred across distant genera and provide heterologous interference against invasive nucleic acids. This finding opens novel possibilities for practical applications, notably the transfer of active CRISPR/Cas systems between species in order to “vaccinate” bacteria against viruses or plasmids. This can be used to develop strains more robust against phage attack, and safer organisms less likely to uptake and disseminate plasmid-encoded undesirable genetic elements.

In addition, the simple modular organization of the Cas9-crRNA complex, where specificity for DNA targets is encoded by a small crRNA and the cleavage machinery consists of a single, multidomain Cas9 protein, provides a versatile platform for the engineering of universal RNA-guided DNA endonucleases. Indeed, by altering the RNA sequence within the Cas9 crRNA complex, programmable endonucleases can be designed both for *in vitro* and *in vivo* applications, and in this thesis we provide a proof of concept for this novel and promising application. For example, the Cas9 endonuclease system can be readily re-programmed using customized RNAs to specifically cleave dsDNA, thus expanding the genome editing enzyme repertoire beyond zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs).

The major findings presented for defence in this thesis:

- CRISPR3/Cas system from Gram-positive *S. thermophilus* bacteria provides protection against plasmid transformation and phage infection in the Gram-negative *E. coli* host.
- Cas9 is the sole Cas protein required for the immunity step in Type II CRISPR/Cas systems.
- Cas9 protein forms a ribonucleoprotein complex with 42 nt crRNA.
- Cas9 protein guided by crRNA cleaves double-stranded DNA at the specific position in a proto-spacer adjacent motif dependent manner.
- Cas9 uses two distinct active sites, RuvC and HNH, to generate site-specific nicks on opposite DNA strands.

1. LITERATURE OVERVIEW

1.1. Bacterial antiviral defense mechanisms

Bacteriophages (phages) are the most abundant biological entity on earth and outnumbers their hosts (bacteria) more than ten times (Brüssow & Hendrix, 2002). To protect themselves against bacteriophages and to survive in a hostile environment, bacteria have developed multiple defense mechanisms. They target almost every step of virus proliferation: adsorption of bacteriophage, genomic DNA injection, genomic DNA replication, transcription and translation, phage assembly and cell lysis. Major bacterial defence mechanisms include prevention of phage adsorption, blocking of phage DNA injection, degradation of the virus nucleic acid and abortive infection (Figure 1) (Sturino & Klaenhammer, 2006; Labrie *et al*, 2010).

Adsorption of phages on host receptors is the initial step of infection crucial for a host recognition by incoming virus (Labrie *et al*, 2010). Prevention of phage particle adsorption on the cell receptors can be accomplished by several ways. First, bacteria block phage receptors or alter their conformation to impede phage binding. For example, *Staphylococcus aureus* produces a protein A, which masks phage receptors and interferes with phage infection (Nordström & Forsgren, 1974). The outer-membrane lipoprotein TraT, encoded by the *E. coli* F^+ strains, masks or modifies the conformation of outer-membrane protein A (OmpA), which is a receptor for many T-even-like *E. coli* phages (Riede & Eschbach, 1986). Second, in some cases the polymers of the extracellular matrix can provide a physical barrier between phages and their receptors. For example, production of exopolysaccharides, alginates, increases phage resistance of *Azobacter* spp. cells (Hanlon *et al*, 2001). Finally, bacteria can produce competitive inhibitors, which specifically bind to bacteriophage receptors, making these receptors unavailable for phages. For example, some *E. coli* strains produce a small peptide, microcin J25. This peptide specifically

interacts with the FhuA iron transporter, which is also a receptor for T5 and other phages (Destoumieux-Garzón *et al*, 2005). Microcin J25 is produced under conditions of nutrient depletion, to inhibit the growth of phylogenetically related strains and protects the bacterium from phage infection (Destoumieux-Garzón *et al*, 2005).

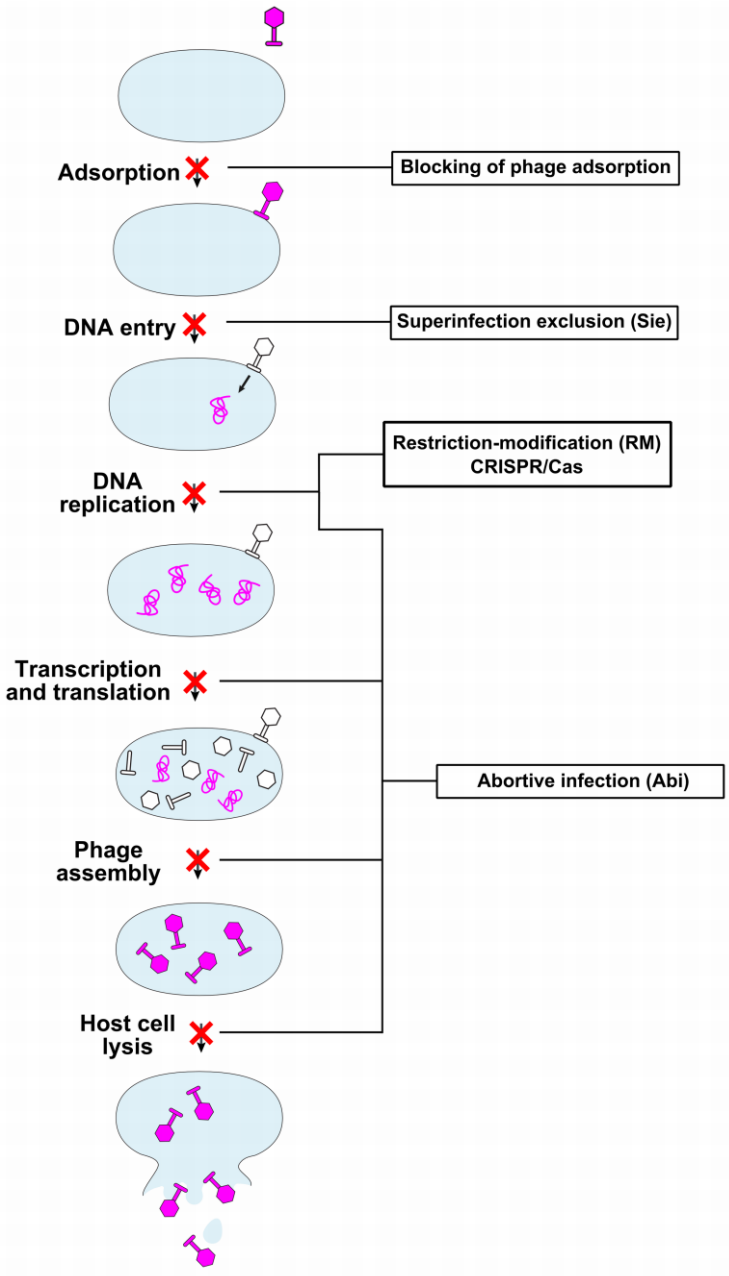

Figure 1. Schematic representation of lytic bacteriophage life cycle steps targeted by specific phage-resistance mechanisms. Modified from (Sturino & Klaenhammer, 2006; Labrie *et al*, 2010).

After adsorption, the phage injects its genomic DNA into the bacterial cell. Superinfection exclusion (Sie) systems are membrane anchored or membrane associated proteins, which block the entry of phage DNA. However, Sie protein encoding genes are often found in prophages, suggesting that originally Sie systems were involved in phage-phage, but not phage-bacteria interactions (Labrie *et al*, 2010). The two best characterized Sie systems are encoded by the *E. coli* T4 phage *imm* and *sp* genes. The *E. coli* cells already infected with T4 phage could not be infected further with T-even-type phages (Lu & Henning, 1994). These systems prevent DNA injection performed by other T-even phages, but act independently and have different mechanisms of action. The Imm protein, which is located in the membrane, changes the conformation of a receptor to prevent phage DNA transfer (Lu & Henning, 1994). On the other hand, the membrane protein Sp prevents degradation of peptidoglycan and subsequent entry of phage DNA by inhibiting T4 lysozyme. This protein is located at the end of the phage tail and facilitates DNA injection by creating the hole in the bacterial cell wall (Lu & Henning, 1994). Some Sie systems also were found in Gram-positive bacteria *Lactoccocus lactis* and *S. thermophilus.* These systems are not characterized very well, but it is predicted to be localized in the bacterial membrane (Sturino & Klaenhammer, 2006; Labrie *et al*, 2010). Interestingly, Sie-like system (*ltp* gene encodes signal peptide bearing lipoprotein) found in *S. thermophilus* prophage is active and confers resistance to some lactococci phages when transferred into *L. lactis* (Sun *et al*, 2006).

Once phage DNA is injected into the bacterial cell, it could be destroyed by restriction – modification (RM) or CRISPR/Cas systems. The restriction-modification systems are found in almost all bacteria and archaea, and one cell usually harbors multiple RM systems. The main function of RM systems is to defend cells against foreign double-stranded DNA, including bacteriophage genomes. The methyltransferases modify host DNA and protect it from cleavage, while incoming DNA lacks modification and is recognized as foreign and destroyed by restriction endonucleases (Pingoud *et al*, 2005; Pingoud,

2004). The activities of RM systems are encoded by several heterogeneous genes and considering subunit composition and cofactor requirement have been classified into at least four types (type I–type IV) (Roberts, 2003).

The last, but not least bacterial defense system line is abortive infection (Abi) systems (Labrie *et al*, 2010). The Abi systems kill infected bacterial cells, thereby blocking the multiplication of the infecting phages and their spread to other nearby cells (Snyder, 1995). In most cases very heterogeneous Abi systems target important phage multiplication steps: replication, translation or transcription (Labrie *et al*, 2010).

The best characterized Abi system is the Rex system from phage λ-lysogenic *E. coli* strains, which requires two proteins, RexA and RexB. The RexA protein is activated by phage DNA replication or recombination intermediate. The intracellular sensor RexA activates the membrane ion channel, RexB, that allows the passage of monovalent cations, destroying the membrane potential and killing the cell (Snyder, 1995). The other two *E. coli* Abi systems, Lit and anticodon nuclease PrrC, target the translation step and arrest phage development in this stage (Snyder, 1995). The Gol peptide presented in the major capsid protein of T4 phage, activates the Lit zinc metalloproteinase, which cleaves an elongation factor – Tu and inhibits protein synthesis in *E. coli* cells (Snyder, 1995; Bingham *et al*, 2000). On the other hand, PrrC cleaves tRNA-Lys in the anticodon loop. PrcC activity is neutralized by a restriction endonuclease from the *Eco*prrI restriction-modification system (Snyder, 1995). The small T4 phage encoded peptide Stp alters the interaction between *Eco*prrI and PrrC, releasing the PrrC enzyme and causing cell death (Snyder, 1995).

1.2. CRISPR/Cas systems

Recently, an adaptive microbial immune system, named Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs), has been identified that provides acquired immunity against viruses and plasmids. It

consists of an array of highly conserved short DNA repeat sequences (typically 21 to 48 bp length), which are interspaced by stretches of variable sequence called spacers. The spacer sequences generally originate from phage or plasmid DNA (Hols *et al*, 2005; Mojica *et al*, 2005) (Figure 2). A set of *cas* (CRISPR-associated) genes is typically located in the vicinity of the repeat-spacer array (Jansen *et al*, 2002; Makarova *et al*, 2006). CRISPRs, in combination with the associated Cas proteins, form the CRISPR/Cas systems (Figure 2). CRISPR/Cas systems are widespread, and are found in 48 % of bacteria (in 916 species from 1911 sequenced) and more than 85 % of archaea (in 114 species from 132 sequenced) genomes (Grissa *et al*, 2007) (2012.09.04). One bacteria cell might contain more than one CRISPR/Cas system. In contrast to other bacterial antiphage mechanisms, the CRISPR/Cas systems are invader-specific, adaptive and heritable.

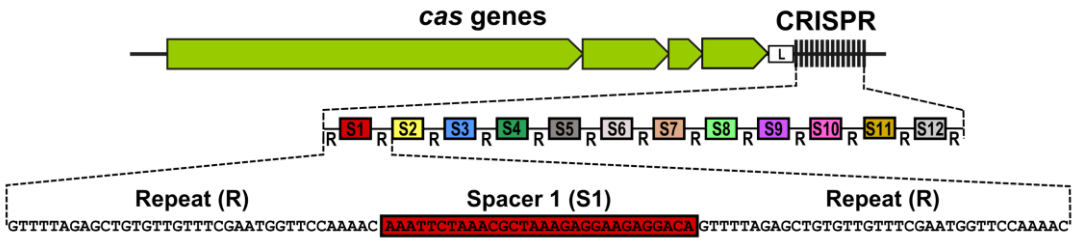

Figure 2. Schematic representation of CRISPR/Cas system. CRISPR array consists of short partially palindromic repeats (R) interspaced by short unique spacer sequences (S1-12). Genes of Cas proteins (green) are usually encoded in the vicinity of CRISPR region. A/T-rich sequence, located immediately upstream of the first repeat and containing putative promoter, is termed leader sequence (L).

The CRISPR/Cas mechanism might be divided into two major steps: immunization (adaptation) and immunity (Figure 3). In the immunization step (Figure 3 A), the Cas proteins or Cas protein complex recognizes foreign DNA (bacteriophage genomic DNA or plasmid) and integrates short pieces of it into the CRISPR region as a new spacer (Barrangou *et al*, 2007; Garneau *et al*, 2010; Yosef *et al*, 2012; Swarts *et al*, 2012). The new spacer is always integrated in the leader end of the repeat region (Barrangou *et al*, 2007; Yosef *et al*, 2012) and the first repeat serves as the template for the newly inserted

repeat (Yosef *et al*, 2012). The detailed mechanism of the initial foreign DNA recognition and new spacer acquisition is still unknown, however it seems that the large majority of bacteria die upon virulent phage infection, and only a small proportion of the population survives by acquisition of phage-derived spacers (Barrangou *et al*, 2007). Cas1 and Cas2 proteins, conserved in all known CRISPR/Cas systems, are involved in the spacer acquisition step (Yosef *et al*, 2012).

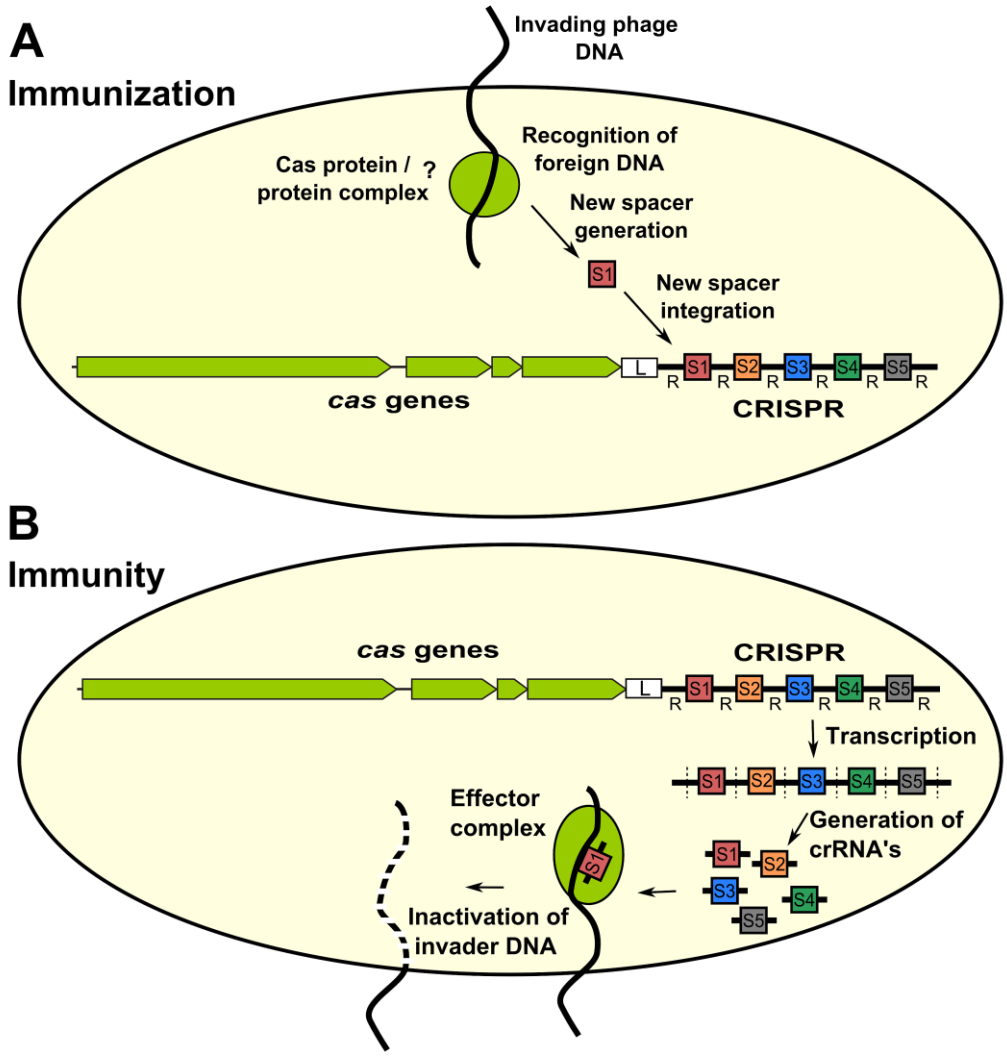

Figure 3. Schematic representation of CRISPR/Cas systems mechanism. (A) Immunization (adaptation) step. Cas proteins/protein complex recognizes foreign DNA and integrates a novel repeat-spacer unit at the leader end of the CRISPR locus. (**B)** Immunity step. The CRISPR repeat-spacer array is transcribed into a long RNA molecule, which is further processed into mature crRNAs. crRNAs are bound by the effector complex and used as a guide for invading nucleic acid recognition and inactivation. Adapted from (Horvath & Barrangou, 2010).

The immunity step might be divided into two parts. Firstly, the CRISPR repeat-spacer array is transcribed into a long primary RNA transcript that is further processed into a set of small CRISPR RNAs (crRNAs), containing a conserved repeat fragment and a variable spacer sequence (guide) complementary to the invading nucleic acid (Carte *et al*, 2008; Brouns *et al*, 2008; Hale *et al*, 2009). Next, crRNAs combine with Cas proteins to form an effector complex, which 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*, 2012), and induces sequence-specific cleavage (Garneau *et al*, 2010), thereby preventing proliferation and propagation of foreign genetic elements.

1.3. Diversity and nomenclature of CRISPR/Cas systems and Cas proteins

There is a huge diversity between CRISPR/Cas systems found in eubacteria and archaea. The sequence and length of repeat and spacer units are conserved within the system, but greatly diverge between systems. The length of repeats may vary from 24 to 47 bp, while spacer length might be from 26 to 72 bp (Grissa *et al*, 2007; Sorek *et al*, 2008). The number of repeat-spacer units in one CRISPR loci varies from 2 to 249 (Kunin *et al*, 2007; Sorek *et al*, 2008). More than one CRISPR/Cas system might be found in one genome, the record holder is *Methanocaldococcus jannaschii*, which contains 18 CRISPR loci (Sorek *et al*, 2008) .

Initially, 4 distinct genes encoding conserved Cas proteins were identified and named *cas1-4* (Jansen *et al*, 2002), but subsequent bioinformatic analyses have shown that Cas proteins are more diverse. Now it is recognized that *cas* genes encode 65 sets of orthologous Cas proteins, which can be classified into 25 or 45 different families (Haft *et al*, 2005; Makarova *et al*, 2006, 2011b). Only the *cas1* and *cas2* genes seem to be universal and found in the majority of

CRISPR/Cas systems. The phylogenetic analysis of Cas1 sequences indicated the existence of several versions of CRISPR/Cas system.

Recently, a new classification of CRISPR/Cas systems has been suggested, where 10 families of CRISPR/Cas systems with different *cas* gene arrangements were divided into three major Types (Figure 4) (Makarova *et al*, 2011b). More than one CRISPR/Cas system Type is usually found in one organism, suggesting that these systems are compatible and could share functional components (Makarova *et al*, 2011b; Wiedenheft *et al*, 2012).

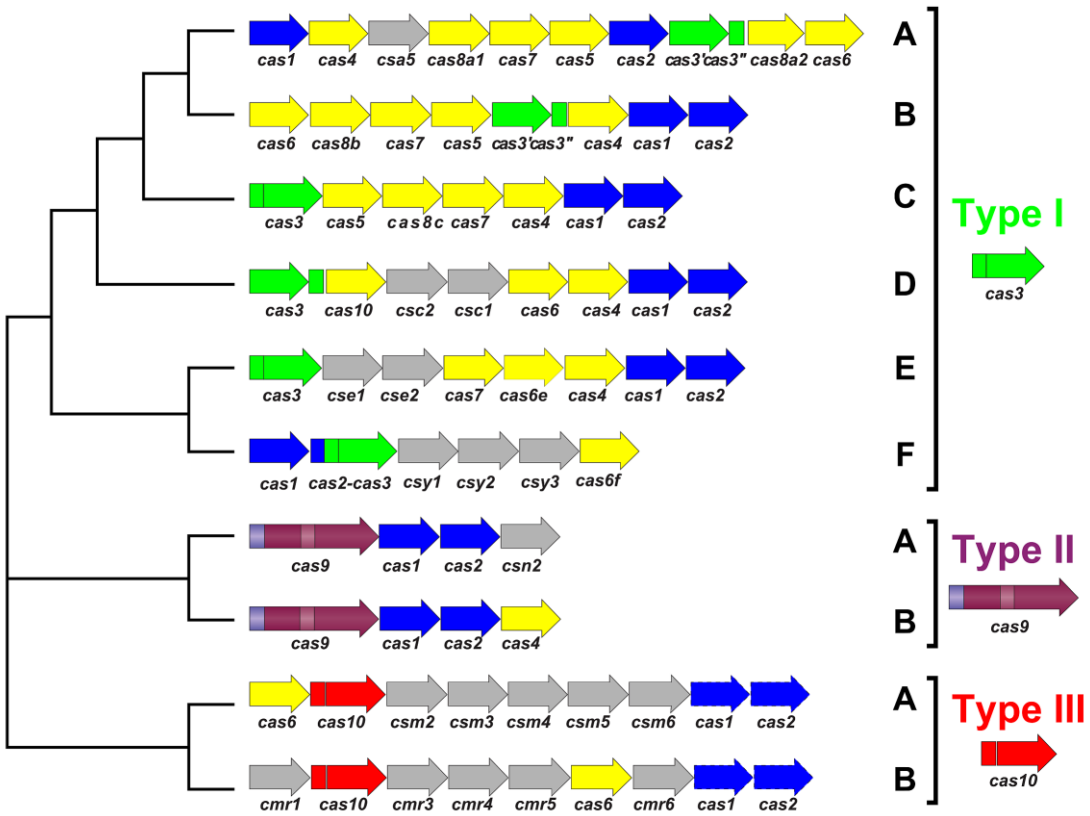

Figure 4. Diversity and nomenclature of CRISPR/Cas systems. Ten subtypes of CRISPR/Cas systems belong to three major Types. Type signature genes are color coded (*cas3* – green, *cas9* – purple, *cas10* – red). The genes (*cas1*, *cas2*) conserved in all CRISPR/Cas systems are colored in blue. The remaining *cas* genes (*cas4*, *cas5*, *cas6*, *cas7* and *cas8*) are colored in yellow, while subtype specific genes - grey. The hypothetical evolutionary relationships between subtypes are represented in the left. Adapted from (Makarova *et al*, 2011a).

All Type I CRISPR/Cas systems encode the signature *cas3* gene. These systems differ in number and composition of *cas* genes and are further divided

into the 6 subtypes (I-A, I-B, I-C, I-D, I-E and I-F). Type II CRISPR/Cas systems are found only in bacteria, but not archaea. They encode just 4 *cas* genes, one of which is always *cas9* (previously known as *cas5* or *csn1*) and are divided into two subtypes (II-A and II-B).

Type III systems are more common in archaea, but also found in bacteria. They have a signature *cas10* gene and are divided into two subtypes. The differences of two subtypes are illustrated by two systems from *Staphylococcus epidermidis* and *Pyrococcus furiosus*. *S. epidermidis* CRISPR/Cas system (III-A subtype) targets plasmid DNA *in vivo* (Marraffini & Sontheimer, 2008), while the purified components of subtype III-B immune system from *P. furiosus* cleaves only single-stranded RNA *in vitro* (Hale *et al*, 2009, 2012).

There is a huge diversity of Cas proteins, for this reason their nomenclature and naming have changed a few times. For example, the signature protein of Type II systems, Cas9 (Makarova *et al*, 2011a), firstly had been named Cas5 (Barrangou *et al*, 2007) and later Csn1 (Horvath & Barrangou, 2010). The Cas proteins might be divided into three groups (Figure 4). Cas1 and Cas2 proteins are ubiquitous and their genes are found in almost all CRISPR/Cas systems. Other 8 Cas proteins (Cas3-10) are found more than in one subtype of CRISPR/Cas systems, the signature genes *cas3, cas9* and *cas10* also belong in this group. Some proteins are found in two Types (Cas6 is present in Type I and Type III), while others (Cas5, Cas8) are found only in few subtypes from Type I. Properties of the 10 major families of Cas proteins are summarized in Table 1.

The most heterogeneous group of Cas proteins are subtype specific proteins. All CRISPR/Cas subtypes, except subtype II-B, encode specific proteins. Subtype specific proteins are named according subtype or microorganism in which they were discovered for the first time. For example, subtype I-E specific proteins are named Cs<u>e</u>1 and Cs<u>e</u>2 since they firstly were discovered in *E. coli*. Subtype IF proteins named Cs<u>y</u>1, Cs<u>y</u>2 and Cs<u>y</u>3 were

discovered in *Yersinia pestis.* Type III specific proteins are named Csm (III-A) and Cmr (III-B), accordingly.

A large part of Cas proteins contains an RNA recognition motif (RRM, or ferredoxin-fold) and because of this feature are assigned to the RAMP (Repeats associated mysterious proteins) superfamily. Cas5, Cas6, Cas7, most of Csm and Cmr proteins belong to the RAMP superfamily (Makarova *et al*, 2006, 2011b).

Table 1. Major families of Cas proteins (Cas1-10). Adapted from (Bhaya *et al*, 2011)

Cas protein	Distribution	Process	Function
Cas1	Universal	Spacer acquisition	DNase
Cas2	Universal	Spacer acquisition	Small RNase specific to U-rich regions
Cas3	Type I (Signature)	Immunity	HD nuclease and DNA helicase
Cas4	Type I and II	Spacer acquisition (?)	Hypothetical RecB-like nuclease
Cas5	Type I	Immunity	RAMP. Part of effector complex
Cas6	Type I and III	Immunity	RAMP. Endoribonuclease involved in crRNA biogenesis, might be part of effector complex
Cas7	Type I	Immunity	RAMP. Part of effector complex
Cas8	Type I	Immunity	Part of effector complex
Cas9	Type II (Signature)	Immunity	Large protein, HNH- and RuvC-like domains
Cas10	Type III (Signature)	Immunity	Part of effector complex, hypothetical HD and polymerase domains

1.4. Mechanisms of CRISPR/Cas systems

1.4.1. Immunization step

The immunization (spacer acquisition) step seems to be conserved between CRISPR/Cas systems. Although metagenomic studies of environmental samples suggest that CRISPR loci evolve rapidly in dynamic equilibrium together with neighboring phage populations (Wiedenheft *et al*, 2012), direct spacer acquisition from phage or plasmid has been demonstrated only in Type II systems form *S. thermophilus* (Barrangou *et al*, 2007; Deveau *et al*, 2008; Garneau *et al*, 2010; Horvath *et al*, 2008) and Type I-E CRISPR/Cas system from *E. coli* (Yosef *et al*, 2012; Swarts *et al*, 2012). Phage challenge experiments in *S. thermophilus* bacteria showed that a small proportion of bacteria integrated new spacers, derived from phage genomic sequences, at the leader end of the CRISPR locus (Barrangou *et al*, 2007; Deveau *et al*, 2008; Horvath *et al*, 2008). Each integration event is accompanied by the duplication of the leader end repeat sequence, thus a new repeat spacer unit is created, maintaining the architecture of CRISPR loci (leader-repeat-spacer-repeat…) (Yosef *et al*, 2012). When overexpressed, Cas1 and Cas2 proteins, are the only Cas proteins required for spacer acquisition in *E. coli* CRISPR/Cas system (Yosef *et al*, 2012). In addition, when multiple spacers were integrated in a CRISPR loci of a full length *E. coli* system, all spacers targeted the same strand of DNA, implying that CRISPR interference caused by the first integrated spacer directs subsequent spacer acquisition events in a strand specific manner (Swarts *et al*, 2012). However, it has been demonstrated that the Cas1 protein from *E. coli* interacts with repair and recombination proteins (Babu *et al*, 2011), indicating that recombination machinery might be involved in spacer acquisition step.

Mutational analysis of *S. thermophilus cas* genes showed that the *csn2* gene is required for new spacer acquisition, but not for the interference step (Barrangou *et al*, 2007). This gene is not conserved in other Types of CRISPR/Cas systems, suggesting that either the acquisition step in Type II

systems is distinct from other CRISPR/Cas systems, or other systems contain functional homologs of Csn2 protein (Wiedenheft *et al*, 2012)

The integration machinery should distinguish foreign phage or plasmid DNA from host genomic DNA (Sternberg *et al*, 2012). How this function is fulfilled still is unknown, but analysis of newly acquired spacers suggested that spacer sequences are selected not randomly (Horvath *et al*, 2008; Swarts *et al*, 2012; Yosef *et al*, 2012). The short 2-5 bp length motifs are found in the vicinity of spacer matching sequences (proto-spacers) in the phage or plasmid DNA. Such motifs are named PAM (proto-spacer adjacent motif). PAM length and sequence varies between CRISPR/Cas systems, usually depending on the system Type (Mojica *et al*, 2009). For example, PAMs of two subtype II-A CRISPR1/Cas and CRISPR3/Cas systems from *S. thermophilus* (5'-AGAAW-3' and 5'-NGGNG-3') are located upstream the proto-spacer (Deveau *et al*, 2008; Horvath *et al*, 2009). On the other hand, the PAM 5'-AWG-3'of subtype I-E system from *E. coli* is located downstream the proto-spacer sequence (Swarts *et al*, 2012; Yosef *et al*, 2012). The differences between PAM sequences and location in respect to the proto-spacer suggest that each CRISPR/Cas system contains Cas protein or proteins involved in PAM recognition. However the mechanism of PAM recognition is still unknown (Wiedenheft *et al*, 2012). It is worth to note, that subtype III-B systems, working on RNA, are predicted to be PAM independent (Hale *et al*, 2009; Zhang *et al*, 2012).

1.4.2. Immunity step

Although the mechanism of a new spacer acquisition seems to be conserved in all CRISPR/Cas systems, the immunity step varies greatly between different Types. The immunity step usually is divided into crRNA maturation (RNA transcription and cleavage) and interference (foreign nucleic acid degradation) (Figure 5).

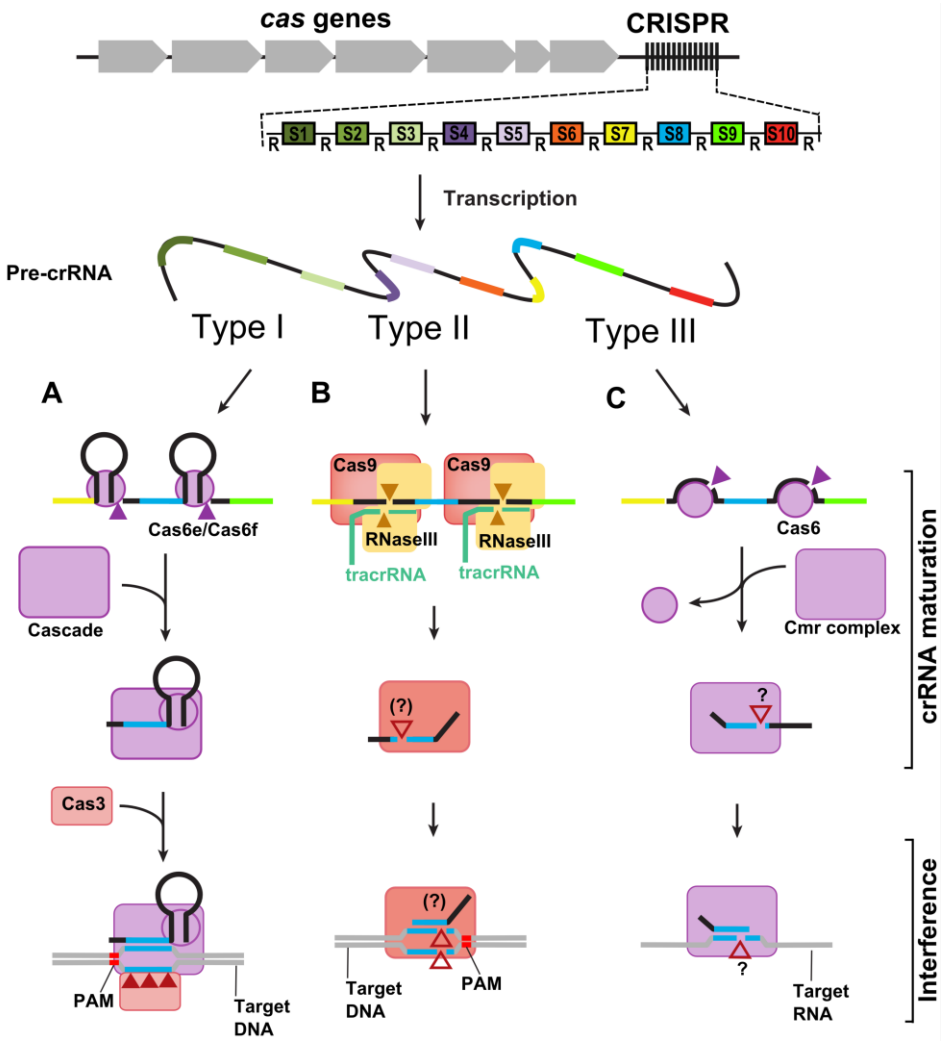

Figure 5. Comparison of interference step in different Types of CRISPR/Cas systems. (**A)** Type I system. Repeats of Type I CRISPR/Cas systems form a hairpin structure, which is recognized and cleaved by a Cas6 family member (Cas6e or Cas6f enzymes). Cas6e/Cas6f stays bound to the crRNA and serves as platform for the effector complex (Cascade and Csy complex in subtypes I-E and I-F accordingly) assembly. The crRNA contains a full spacer sequence (blue), an 8 nt length repeat handle in the 5' terminus and the remaining part of the repeat forming a hairpin structure at the 3' terminus. Effector complex and crRNA binds to the dsDNA target, forms an R-loop which is recognized and degraded by Cas3 protein. PAM sequence is required. (**B)** Type II system. A trans-encoded small RNA (tracrRNA) in the presence of Cas9 protein base-pairs with the repeat region of pre-crRNA and is cleaved by host RNase III. The crRNA is further trimmed at the 5' end by an unknown mechanism. The dsDNA targets are cleaved within the proto-spacer, PAM sequence (red) is required. (**C)** Type III system. Processing of crRNA requires Cas6 protein, but crRNAs are transferred to the effector complex (Cmr in subtype III-B). The 3' end of crRNA is further trimmed by an unknown mechanism. Subtype III-B effector complex targets RNA, while III-A systems seems to be active against DNA. PAM is not required for interference by Type III systems. Filled triangles represent experimentally demonstrated nuclease activity, while activity of open triangles has not yet been identified. Modified from (Makarova *et al*, 2011b; Bhaya *et al*, 2011).

1.4.2.1. Generation of crRNA

Firstly, the long primary transcript of CRISPR array (pre-crRNA) is transcribed by cellular RNA polymerase. Next, the pre-crRNA is cleaved in the repeat sequence, generating short crRNAs, which contain one spacer and one repeat sequence. In Type I and Type III systems these reactions are catalyzed by endoribonucleases from the Cas6 superfamily (see 1.5.2.1), which operates either as a subunit of a larger effector complex (i. e. Cas6e from CRISPR-associated complex for antiviral defence (Cascade) in *E. coli*) or as a single enzyme (i. e. Cas6 from archaea *Pyrococcus furiosus*). Recently, it has been demonstrated in *Streptococcus pyogenes*, that Type II CRISPR/Cas systems contain an additional trans-encoded small RNA (tracrRNA), which acts as a guide for the processing of pre-crRNA. A part of the tracrRNA is complementary to the repeat and hybridizes with pre-crRNA. The double-stranded RNA in the repeat is recognized and cleaved by bacteria encoded RNase III enzyme in the presence of Cas9 protein (Deltcheva *et al*, 2011). The crRNA in Type I systems usually are the same length as a repeat-spacer unit, while in other systems crRNA seems to be shorter. It has been shown that both in Type II and Type III systems, crRNA is processed further, and is shorter because of secondary cleavage at the fixed position within the spacer region (Figure 5). Which proteins cleave crRNA within the spacer is still unknown (Deltcheva *et al*, 2011; Hale *et al*, 2009; Hatoum-Aslan *et al*, 2011).

1.4.2.2. crRNA guided interference

During the interference step, foreign DNA or RNA is targeted and cleaved within the proto-spacer sequence. The crRNAs associated with Cas proteins form ribonucleoprotein effector complexes, which recognize invading nucleic acids. Targets on foreign nucleic acids are identified by base-pairing interactions between crRNA spacer sequence and complementary strand in the intruder (proto-spacer) (Wiedenheft *et al*, 2012). The interference mechanism and effector complex structure varies between the Types (Figure 5). In Type I systems from *E. coli*, crRNA together with Cas6e and other Cas proteins,

forms a multisubunit complex (see 1.5.3.1) (Brouns *et al*, 2008). It has been shown that Cascade identifies targets by base-pairing the spacer region of crRNA to the complementary strand of double-stranded DNA, while the other, non-complementary, strand is displaced to form an R-loop (Jore *et al*, 2011). The R-loop is recognized and foreign DNA is degraded by an HD nuclease domain of the Cas3 protein (see 1.5.3.2) (Sinkunas *et al*, 2011; Westra *et al*, 2012).

In Type II systems, the proteins involved in foreign DNA degradation are yet unknown, but the pathway recently has been demonstrated in *S. thermophilus* (Garneau *et al*, 2010). It is clear that PAM plays an important part in the target recognition and degradation processes (Garneau *et al*, 2010). Furthermore, DNA sequencing and southern blot analysis revealed that both phage and plasmid DNA strands were cleaved by CRISPR1/Cas system within the proto-spacer sequence near PAM *in vivo* (Garneau *et al*, 2010).

The interference mechanism of Type III CRISPR/Cas systems seems to be the most complex. In subtype III-B CRISPR/Cas systems of *Sulfolobus solfataricus* and *Pyrococcus furiosus*, Cmr protein and crRNA complex recognize and cleave synthetic RNA *in vitro* (Hale *et al*, 2009; Zhang *et al*, 2012), while the CRISPR/Cas system of *Staphylococcus epidermidis* targets DNA *in vivo* (Marraffini & Sontheimer, 2008).

1.5. Biochemistry of Cas proteins

To understand mechanisms of CRISPR/Cas systems, the biochemical properties and the structural architecture of Cas proteins and protein complexes are intensively investigated. Three dimensional structures of Cas proteins known to-date are summarized in Table 2. The following part of literature review covers structural and biochemical properties of Cas proteins and effector complexes.

Table 2. Available structures of Cas proteins.

Protein	Function	Process	Organism	Subtype	PDB ID	Ref.
Cas1	Metal-dependent DNA nuclease	Spacer acquisition	*Pseudomonas aeruginosa*	I-Y	3god	(Wiedenheft *et al*, 2009)
Cas1	Metal-dependent DNA nuclease	Spacer acquisition	*Thermogota maritima*	I-B	3lfx	
Cas1	Metal-dependent DNA nuclease	Spacer acquisition	*Aquiflex aeolicus*	I-B	2yzs	-
Cas1	Metal-dependent DNA nuclease	Spacer acquisition	*Pyrococcus horikoshii*	I-B	3pv9	-
Cas1	Metal-dependent DNA nuclease	Spacer acquisition	*Escherichia coli*	I-E	3nke	(Babu *et al*, 2011)
Cas2	ssRNA nuclease	Spacer acquisition	*Sulfolobus solfataricus*	I-A	2i8e, 2ivy	(Beloglazova *et al*, 2008)
Cas2	ssRNA nuclease	Spacer acquisition	*Desulfovibrio vulgaris*	I-A	3oq2	(Beloglazova *et al*, 2008)
Cas2	ssRNA nuclease	Spacer acquisition	*Pyrococcus furiosus*	III-B	2i0x	(Beloglazova *et al*, 2008)
Cas2	ssRNA nuclease	Spacer acquisition	*Thermus thermophilus*	I-E	1zpw	(Beloglazova *et al*, 2008)
Cas3	Endonuclease (HD domain)	Interference	*Thermus thermophilus*	I-E	3sk9	(Mulepati & Bailey, 2011)
Cas3	Endonuclease (HD domain)	Interference	*Methanocal-dococcus jannaschii*	I-A	3s4l	(Beloglazova *et al*, 2011)
Cse1	Cascade component	Interference	*Thermus thermophilus*	I-E	2zop	(Sashital *et al*, 2012)
Cse2	Cascade component	Interference	*Thermus thermophilus*	I-E	2zca	-
Cas6	Cascade component	crRNA generation	*Pyrococcus furiosus*	III-B	3i4h, 3pkm	(Carte *et al*, 2008)
Cas6e	Cascade component	crRNA generation	*Thermus thermophilus*	I-E	1wj9 2xli,	(Sashital *et al*, 2011; Gesner *et al*, 2011)
Cas6f	Cascade component	crRNA generation	*Pseudomonas aeruginosa*	I-Y	2xli	(Haurwitz *et al*, 2012)
Cas7	Cascade component	Interference	*Sulfolobus solfataricus*	I-A	3ps0	(Lintner *et al*, 2011)
Cas10	Cmr complex component	Interference	*Pyrococcus furiosus*	III-B	3ung, 4doz	(Cocozaki *et al*, 2012; Zhu & Ye, 2012).
Cmr5	Cmr complex component	Interference	*Thermus thermophilus*	III-B	2zop	(Sakamoto *et al*, 2009)
Csn2	dsDNA binding	Spacer acquisition	*Streptococcus agalactiae*	II-A	3qhq	(Ellinger *et al*, 2012)
Csn2	dsDNA binding	Spacer acquisition	*Enterococcus faecalis*	II-A	3s5u	(Nam *et al*, 2011)
Csn2	dsDNA binding	Spacer acquisition	*Streptococcus pyogenes*	II-A	3v7f	(Koo *et al*, 2012)

1.5.1. Proteins involved in spacer acquisition

1.5.1.1. Cas1 protein

The exact function of Cas1 protein in CRISPR/Cas mechanism is not clear yet, but recently it has been shown that Cas1 is required for the new spacer acquisition step (Yosef *et al*, 2012). Cas1 proteins from three different sources (*E. coli*, *P. aeruginosa* and *S. solfataricus*) have been purified and characterized biochemically. The Cas1 protein from *S. solfataricus* (subtype I-A) binds double-stranded and single-stranded DNA or RNA with high affinity (K_d values ~ 18 – 50 nM) but in a sequence non-specific manner. Nuclease activity of this protein was not detected (Han *et al*, 2009). However, Cas1 proteins from *P. aeruginosa* (subtype I-Y) and *E. coli* (subtype I-E) act as divalent metal dependent nucleases. Cas1 from *P. aeruginosa* cleaves single and double-stranded DNA at the presence of Mn^{2+} ions as a cofactor (Wiedenheft *et al*, 2009). Cas1 protein from *E. coli* exhibits a wider nuclease activity. It cleaves single-stranded and double stranded DNA, single-stranded RNA and branched DNA substrates including Holliday junctions, and other intermediates of DNA repair and recombination (Babu *et al*, 2011). In both cases Cas1 protein does not show any sequence specificity for nucleic acid.

In addition, proteomic studies showed that the Cas1 protein from *E. coli* binds double-stranded DNA with a preference for mismatched or abasic substrates (Chen *et al*, 2008). This is in agreement with data that Cas1 physically and genetically interacts with multiple *E. coli* proteins involved in DNA recombination and repair, including recB, recC and ruvB. *E. coli* strain lacking Cas1 shows defects in DNA repair and chromosome segregation (Babu *et al*, 2011).

Five crystal structures of Cas1 protein are known to date (Table 2) (Babu *et al*, 2011; Wiedenheft *et al*, 2009). Despite of the huge diversity in sequence level (proteins share less than 15 % identical amino acids), all five proteins have a similar structure. Cas1 protein is divided into two structurally distinct domains; small N-terminal β-sheet domain connected with larger C-terminal α-

helical domain via flexible linker (Figure 6). The C-terminal domain of Cas1 contains a conserved divalent metal-ion binding site, and point mutations of the metal coordinating residues abolish DNA degradation activity. The active site of Cas1 is surrounded by basic residues which create a potential nucleic acid binding site in proximity to the metal ion. The Cas1 proteins are homodimers in crystal and in solution. The dimer is formed through interactions between N-terminal β-strand domains.

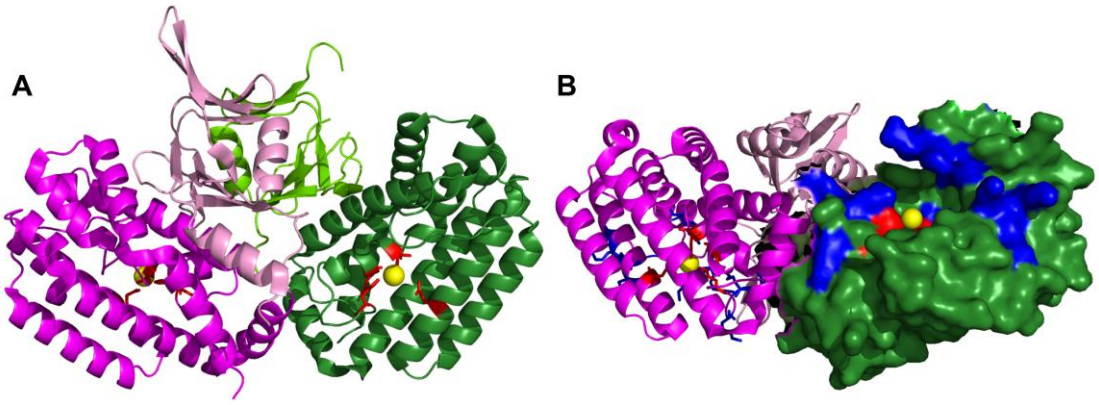

Figure 6. Crystal structure of Cas1 protein from *P. furiosus* (PDB ID: 3GOD). (**A)** Homodimeric organization of Cas1. Monomers of Cas1 (green and magenta) are composed of N-terminal β-strand domain (light color) and the C-terminal α-helical (dark color). At the active site, acidic amino acids (red) chelate Mn^{2+} cation (yellow). (**B)** Potential DNA binding surface. The subunit of Cas1 protein is demonstrated as surface. A cluster of basic residues (blue) around the acidic metal-binding pocket (red) and metal ion (yellow sphere) may form a potential DNA binding site.

1.5.1.2. Cas2 protein

Cas2 protein from *E. coli*, together with Cas1, is required for a new spacer acquisition (Yosef *et al*, 2012) and its homologs are present in almost every CRISPR/Cas system. Six Cas2 proteins from 5 different bacteria species were characterized. All of them are endoribonucleases cleaving single-stranded ssRNAs within U-rich regions and produces 3'-OH and 5'-phosphate ends in the presence of Mg^{2+} or Mn^{2+} cations (Beloglazova *et al*, 2008). The crystal structure of Cas2 protein from *S. solfataricus* (PDB ID: 2I8E) showed a homodimer containing a ferredoxin-like fold (Beloglazova *et al*, 2008). The ferredoxin-like fold is present in numerous RNA binding proteins including

RNA-binding domain, the anticodon-binding domain of PheRS, ribosomal proteins S6 and S10. The catalytically essential aspartate 10 is situated at the dimeric interface and two symmetry-related Asp-10 residues are 6.5 Å apart, suggesting one active site formation (Figure 7 A) (Beloglazova *et al*, 2008).

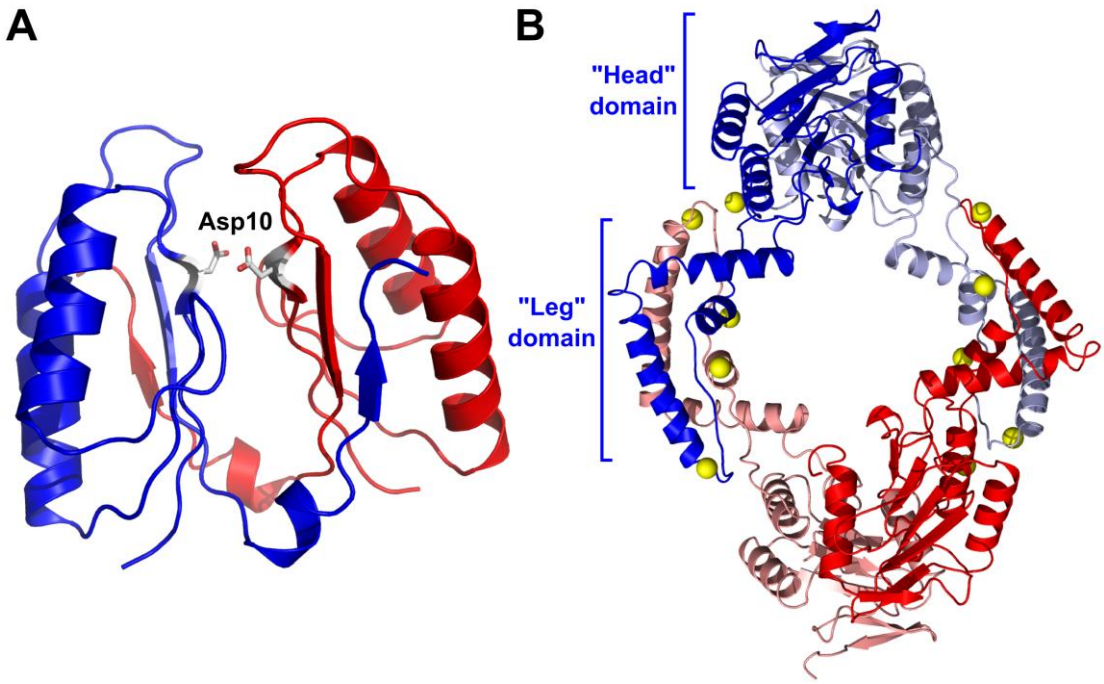

Figure 7. The structures of Cas2 and Csn2 proteins. (A) Cas2 (PDB ID: 2i8e) is a homodimer containing the ferredoxin-like fold. Subunits are colored in red and blue. The active site is located at the dimer interface with the conserved and catalytically essential Asp shown as sticks and labeled. **(B)** Csn2 protein (PDB ID: 3qhq) is a homotetramer stabilized by Ca^{2+} ions (yellow spheres). Monomers of Csn2 form an inner ring which could accommodate double-stranded DNA.

1.5.1.3. Csn2 protein

Csn2 is a subtype IIA specific protein, which functions in the process of new spacer acquisition (Makarova *et al*, 2011b). Deletion of Csn2-like protein from *S. thermophilus* CRISPR1/Cas system inhibited the acquisition of new spacers (Barrangou *et al*, 2007).

Three Csn2 proteins were purified and characterized structurally (Nam *et al*, 2011; Ellinger *et al*, 2012; Koo *et al*, 2012). Although biochemical activity of Csn2 proteins was not detected, all three proteins bind double-stranded DNA. Csn2 proteins are tetramers in solution, and co-purify with Ca^{2+} ions, which stabilize an oligomeric structure (Nam *et al*, 2011; Ellinger *et al*, 2012;

Koo *et al*, 2012). Crystal structures of Csn2 confirmed tetrameric state/form of the protein. Monomer of Csn2 contains two domains: N-terminal globular α/β "head" domain is connected with C-terminal extended α-helical "leg" domain through a flexible hinge (Figure 7B). The dimer is formed through interactions of two "head" domains, while hydrophobic interactions between two dimers at the "leg" domains form a tetrameric ring. The positively charged inner ~ 26 Å wide ring might accommodate double-stranded DNA (Nam *et al*, 2011). However, recent molecular dynamic experiments suggested two possible binding modes: double-stranded DNA could pass through the ring of the tetramer or could be bound in a groove outside the ring (Ellinger *et al*, 2012).

1.5.2. Cas proteins involved in crRNA maturation

1.5.2.1. Cas6 protein family

Members of the Cas6 protein family cleave CRISPR RNA transcripts in the repeat region to form crRNAs in Type I and Type III CRISPR/Cas systems (Makarova *et al*, 2011b; Wiedenheft *et al*, 2012). Cas6 protein family members from three different organisms and systems have been characterized structurally and biochemically: Cas6e from *T. thermophilus* (subtype I-E), Cas6f from *P. aeruginosa* (subtype I-E) and Cas6 from *P. furiosus* (subtype III-B) (Table 2). All Cas6 proteins cleave pre-RNA transcript in the repeat sequence 8 nt upstream of the spacer sequence independently from divalent metal ions (Brouns *et al*, 2008; Carte *et al*, 2008; Haurwitz *et al*, 2010). The Cas6, Cas6e and Cas6f proteins share little sequence homology, however, their three dimensional structures have the ferredoxin-like fold (Figure 8), suggesting a similar mechanism for RNA cleavage. (Wiedenheft *et al*, 2012). The structural arrangement of Cas6 protein substrates – repeats of the CRISPR region, differs. In the case of Type I systems, the repeat sequence forms an internal stable stem-loop structure and RNA is cleaved in the base of the hairpin (Brouns *et al*, 2008; Haurwitz *et al*, 2010), while the repeats of subtype III-B are predicted to be unstructured (Kunin *et al*, 2007) (Figure 8).

Although both Cas6e and Cas6f proteins recognize the sequence and the shape of the repeat hairpin structure, interacting with double-stranded RNA via major groove and single-stranded/double-stranded RNA junction, overall structures and elements required for substrate recognition of these proteins differ (Wiedenheft *et al*, 2012). Cas6e protein is composed of two ferredoxin-like domains (Figure 8 A) (Ebihara *et al*, 2006). The positively charged β-hairpin from the C-terminal domain of Cas6e interacts with the major groove of the RNA A-form helix, which positions the phosphate backbone of the 3′ strand of the RNA stem-loop into a positively charged cleft along the protein (Sashital *et al*, 2011; Gesner *et al*, 2011). RNA binding induces conformational changes in the Cas6e-substrate complex: the bottom base pair of the stem is disrupted; protein forms several base specific contacts with bases located upstream and downstream of the scissile bond, positioning RNA in the active site of the N-terminal domain for specific RNA cleavage (Sashital *et al*, 2011). Cas6f protein has an N-terminal ferredoxin-like domain similar to that of the Cas6e protein, while the C-terminal domain, which is responsible for RNA repeat recognition, has a different fold (Haurwitz *et al*, 2010) (Figure 8A and B). The positively charged arginine rich α helix from the C-terminal domain is inserted in the major grove of double-stranded RNA and two bottom base pairs of the stem-loop are recognized via specific hydrogen bonds. These interactions position the scissile phosphate in the active site of Cas6f (Haurwitz *et al*, 2010). However, Cas6f protein active site, compared to other metal ion-independent RNases (like RNase A), lacks a general acid to protonate the leaving group and positively charged amino acid residues to stabilize transition state of the reaction, resulting in 3-4 orders of magnitude lower cleavage rates (Haurwitz *et al*, 2012).

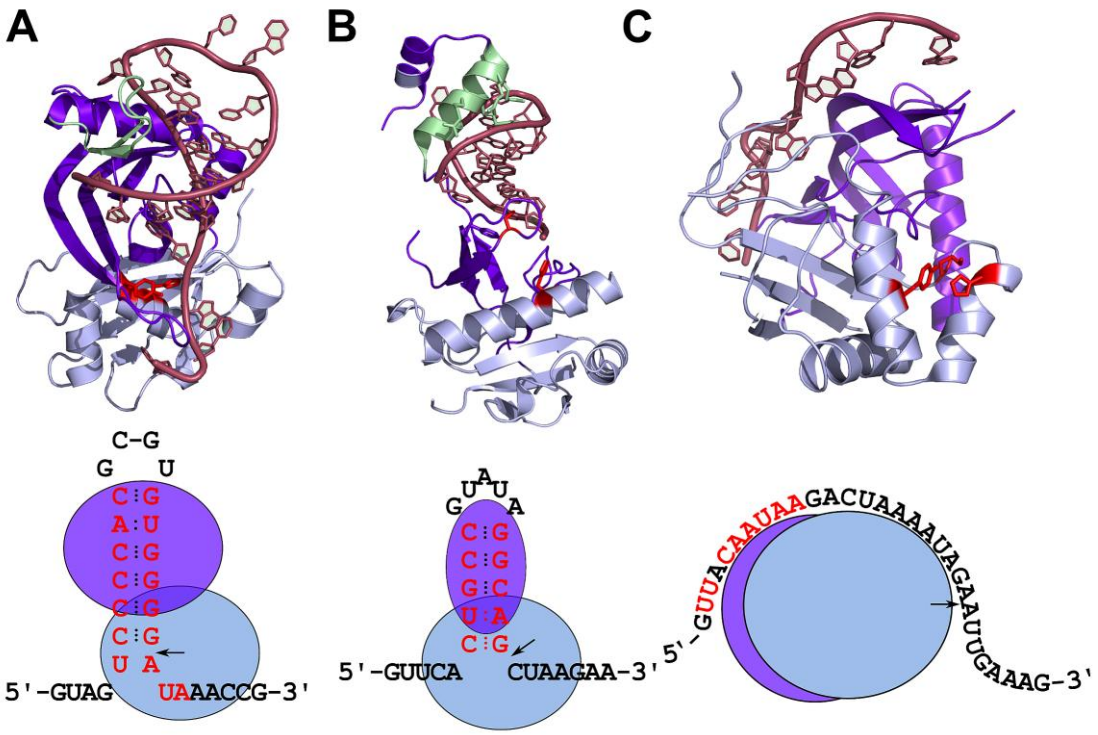

Figure 8. Comparison of Cas6 proteins structures with RNA (upper panel) and RNA repeat recognition mechanisms (lower panel). Protein structures represented in ribbon diagram. N-terminal domains are colored in light blue, C- terminal domains - in purple blue. Catalytically important active site residues are shown as sticks and colored in red. In schemes, ferredoxin-like domains are represented as circle and RNA bases important for recognition are colored in red. Cas6e (**A**) and Cas6f (**B**) proteins (PDB ID: 2y8w and 2xlk) recognize crRNA stem-loop through interactions with the major groove using a positively charged β-hairpin and an α-helix (colored green) accordingly. Cleavage (position indicated by arrow) occurs at the bottom of the hairpin. In Cas6 (**C**) from subtype III-B (PDB ID: 3pkm) RNA recognition and cleavage sites are located at different sides of the protein. Cas6 captures and specifically recognizes single-stranded RNA at the 5' end of repeat, wraps the RNA around the protein and cleaves RNA at the 3' end of repeat.

Both Cas6e and Cas6f are single turnover enzymes and stay tightly bound to the stem of the reaction products (Sashital *et al*, 2011; Sternberg *et al*, 2012). Cas6e or Cas6f complex with crRNA most probably serve as assembly points for CRISPR effector Cascade complexes (Cascade or Csy complex) (discussed in 1.5.3.1 and 1.5.3.3). Csy complex is not formed when active site mutant of Cas6f is co-expressed instead of wild type (WT) protein with other Csy proteins (Haurwitz *et al*, 2012). Distinctly from Cas6e and Cas6f, Cas6 protein from *P. furiosus* is not incorporated into the effector complex (Hale *et al*, 2009).

Repeats of subtype III-B are unstructured and linear (Kunin *et al*, 2007), so Cas6 protein interacts only with single-stranded RNA. Two ferredoxin-like domains are arranged back-to-back and form a central cleft where single-stranded RNA can be bound (Figure 8 C). The recognition and cleavage of the repeat sequence occurs at physically distant sites at different sides of the protein (Carte *et al*, 2008). Cas6 specifically interacts with 8 nucleotides of 5' part of repeat in the "front" surface. Single-stranded repeat RNA is wrapped around protein surface to position scissile phosphate in the active site located in the "back" surface of protein (Wang *et al*, 2011). RNA cleavage occurs at the different repeat side, 8 nt from the 3' terminus (Carte *et al*, 2008, 2010) (Figure 8 C).

1.5.3. Complexes and Cas proteins involved in foreign nucleic acid recognition and degradation

1.5.3.1. Subtype II-E effector complex: Cascade

The crRNA associated with Cas proteins form a large ribonucleoprotein complex that recognizes invading nucleic acids. The existence of an effector complex - Cascade (CRISPR-associated complex for antiviral defense) for the first time was reported in subtype I-E system from *E. coli* (Brouns *et al*, 2008). The *E. coli* CRISPR/Cas system encodes 8 Cas proteins downstream the CRISPR region (Figure 9 A). *E. coli* Cascade is a 405-kDa size ribonucleoprotein complex comprising of an unequal stoichiometry five Cas proteins (Cse1, 2 copies of Cse2, 6 copies of Cas7, Cas5, Cas6e) and a 61 nt length crRNA (Figure 9) (Brouns *et al*, 2008; Jore *et al*, 2011; Wiedenheft *et al*, 2011b). Cascade recognizes double-stranded DNA targets by forming base pairs between the spacer region of crRNA and the complementary strand of double-stranded DNA, while the other, non-complementary, strand is displaced to form an R-loop (Jore *et al*, 2011). It has been shown that binding by Cascade leads to DNA bending (Westra *et al*, 2012). This process does not require ATP, but is more efficient when target DNA is negatively supercoiled

(Westra *et al*, 2012). In addition, Cascade binds only double-stranded DNA targets, which contain PAM near a proto-spacer sequence (Semenova *et al*, 2011; Westra *et al*, 2012). After the R-loop formation by Cascade binding, a Cas3 nuclease/helicase (see 1.5.3.2) is recruited by Cse1 subunit of the Cascade for target DNA degradation (Westra *et al*, 2012).

In vivo experiments showed that only 7 nucleotides proximal to the PAM of the target DNA should perfectly match the crRNA sequence of Cascade, suggesting a similar role to the "seed" sequences in microRNA's. It has been suggested that the crRNA "seed" sequence in *E. coli* plays a role in the initial scanning of invader DNA for a perfect match before base pairing of the full-length spacer can occur (Semenova *et al*, 2011).

Low resolution electron microscopy of Cascade suggested sea-horse shape (Jore *et al*, 2011), which was confirmed by sub-nanometer resolution structures (Wiedenheft *et al*, 2011b) (Figure 9B and C). Six copies of Cas7 protein form a helical elongated backbone of Cascade (spine of the sea horse). The extended crRNA molecule lies in a groove formed by Cas7 proteins helix, while both ends of the crRNA molecule are anchored by specific interactions with other Cas proteins. Hairpin at the 3' terminus of crRNA, stays bound to the Cas6e ribonuclease after pre-crRNA cleavage (see 1.5.2.1), while the 5' end of crRNA forms a hook-like structure in the pocket formed in the surface between the 6^{th} subunit of Cas7 and Cse1. Cas5 protein is positioned near crRNA 5' end binding interface and interacts with 6^{th} subunit of Cas7 and Cse1. Two Cse2 subunits form an elongated dimer positioned along the inner surface of the crRNA-Cas7 spine, connecting the "head" (Cas6e protein) and the "tail" (Cse1 protein) of the complex. First subunit of Cse2 protein interacts with the 1^{st} and 2^{nd}, while second Cse2 subunit interacts with 3^{rd} and 4^{th} subunits of Cas7. Extended conformation of Cse2 dimer creates a deep cleft, which accommodates 3' half of the spacer sequence from crRNA (Wiedenheft *et al*, 2011b). Summarizing overall Cascade structure, crRNA is displayed on the surface formed by Cas7 protein helix, which protects crRNA from degradation

maintaining its availability for base pairing with complementary DNA targets (Wiedenheft *et al*, 2012).

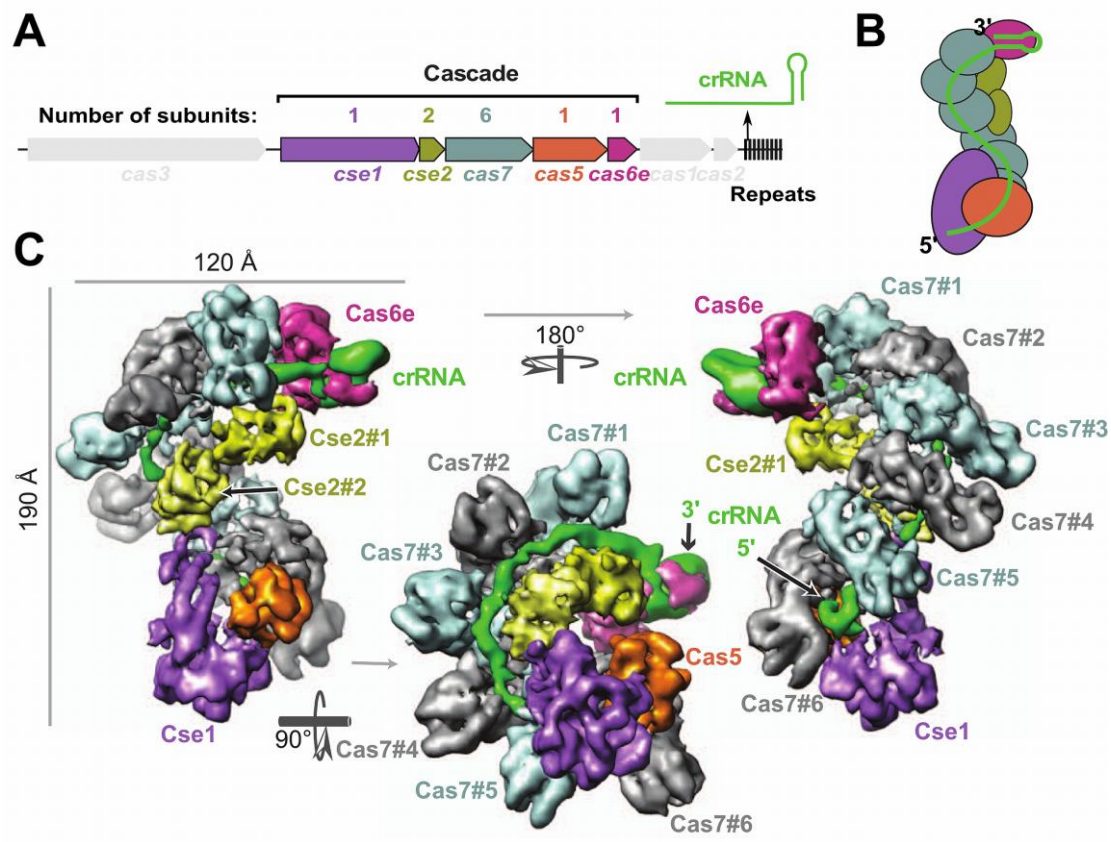

Figure 9. Electron microscopy structure of Cascade complex from *E. coli*. (A) The CRISPR/Cas system from *E. coli* consists of 8 *cas* genes and the CRISPR region. The *cse1*, *cse2*, *cas7*, *cas5* and *cas6* genes encode the Cascade complex. The number of subunits in the complex is indicated above the genes. (**B)** Schematic representation of sea-horse shaped Cascade structure**.** Positions Cas7 subunits are indicated by numbers 1-6. **(C)** Electron microscopy structure of Cascade. Cascade assembles into a sea-horse shape form, where crRNA (green) is positioned along a helical arrangement of six Cas7 subunits. The helical structure is caped at its ends by Cas6e (magenta) and Cse1 (purple). Adapted from (Wiedenheft *et al*, 2011b) .

The structure of Cascade bound to a single-stranded 32 nt length RNA complementary to a crRNA reveals a concerted conformational change in which the Cse1, Cse2, and Cas6e subunits shift and rotate along the Cas7-crRNA backbone. Also, the hook at 5' terminus of crRNA seems to be disrupted (Wiedenheft *et al*, 2011b).

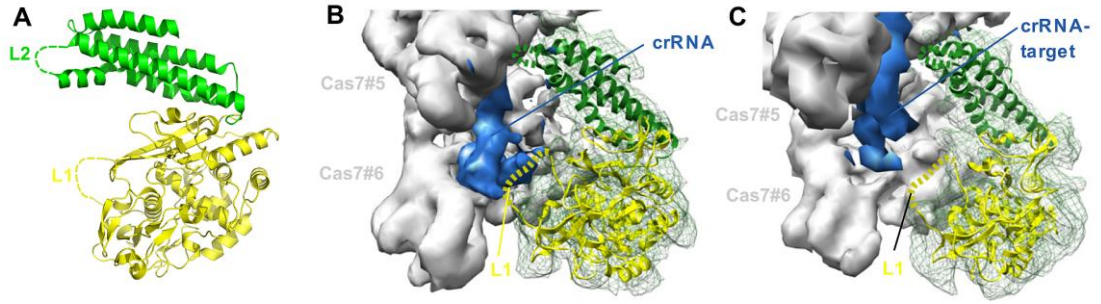

Figure 10. Crystal structure and possible function of Cse1 protein. (A) Cse1 protein consists of two domains: N-terminal domain (yellow) and C-terminal domain (green). Disordered loops L1 and L2 are showed as dashed lines. (**B)** and (**C**) Crystal structure of Cse1 docked into EM structure of *E. coli* Cascade non-bound and bound to the target RNA, respectively. Adapted from (Sashital *et al*, 2012).

Biochemical experiments showed that the largest subunit of Cascade, Cse1 is required for recognition and binding of the target DNA molecule (Mulepati *et al*, 2012; Sashital *et al*, 2012). Crystal structure of Cse1 form *T. thermophilus* showed that Cse1 protein has two domains: N-terminal domain, which adopts new fold and C-terminal four helix bundle domain (Figure 10 A). Each domain has a loop (L1 and L2 accordingly), which is disordered in the crystal (Mulepati *et al*, 2012; Sashital *et al*, 2012). When docked into electron microscopy structure reconstructions of *E. coli* Cascade, the Cse1 structure reveals that a flexible L1 loop from N-terminal domain is located near the crRNA and target RNA (Figure 10 B and C). When target sequence is not bound, L1 loop from N-terminal domain of Cse1 appears to interact with 5' end of crRNA. After the target sequence binding, Cse1 and L1 loop is shifted and might interact with PAM of the target DNA (Mulepati *et al*, 2012; Sashital *et al*, 2012). Mutations in the L1 loop revealed a three amino acid motif (FVN) that might interact with the PAM sequence and is required to accomplish resistance to the phage. In addition, there is evidence that PAM sequence recognition might be linked with DNA destabilization (Sashital *et al*, 2012).

A pool of biochemical and structural data led to suggest a model of how *E. coli* Cascade searches for and binds to the target DNA sequence (Figure 11). Initially, the Cascade searches for possible targets by scanning double stranded

phage DNA for potential PAM sequences through Cse1 L1 loop mediated nonspecific contacts. Once PAM is detected, L1 loop locally destabilizes the double-stranded DNA helix, enabling crRNA access the PAM adjacent DNA sequence in single-stranded form. When Cascade encounters the target sequence, helical distortion at the PAM would be stabilized by base pairing between the seed region of target sequence and crRNA, pausing Cascade at this site and initiating nucleation of full crRNA-target DNA duplex (Sashital *et al*, 2012). This leads to complete non-target strand displacement, which might be a signal for the recruitment of the Cas3 nuclease/helicase (Sinkunas *et al*, 2011; Westra *et al*, 2012).

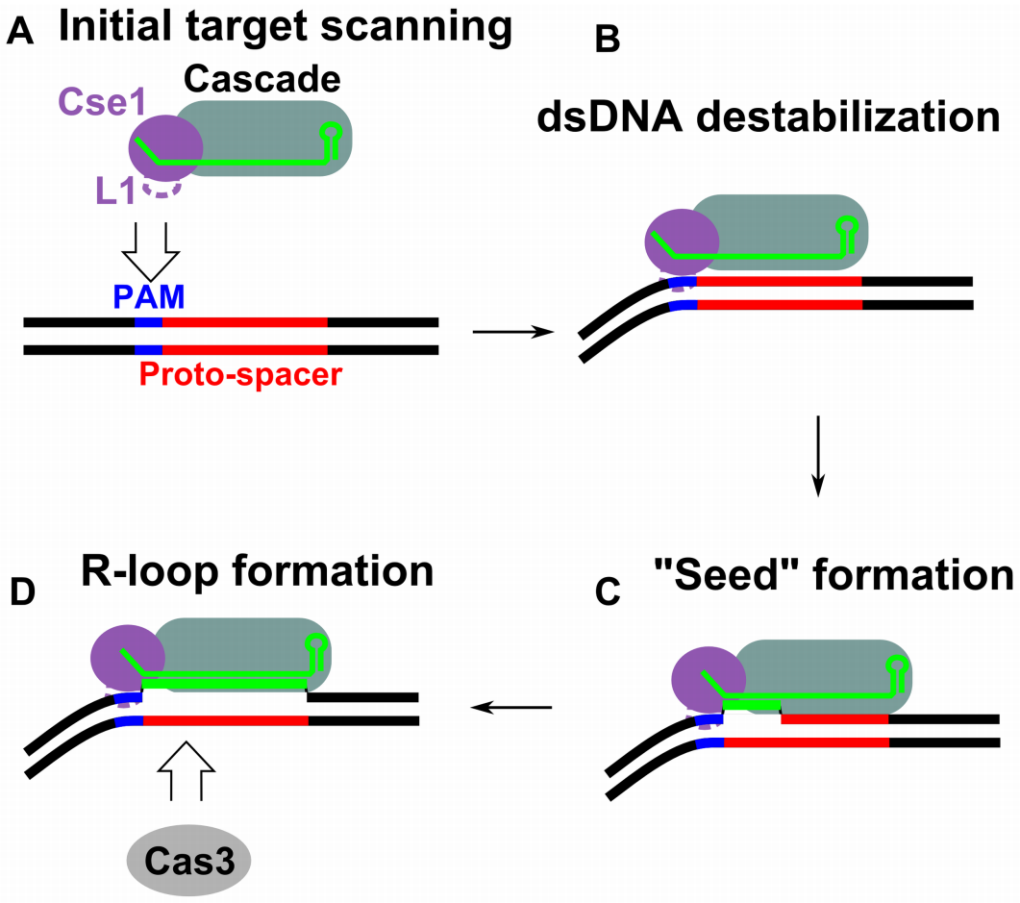

Figure 11. Schematic representation of model for Cascade binding to foreign DNA. (**A**) Cascade scans double-stranded DNA for PAM. (**B**) When PAM is located, Cse1 and loop L1 binding destabilizes the DNA duplex. (**C**) In the presence of proto-spacer, DNA destabilization facilitates strand invasion by "seed" sequence. (**D**) When R-loop is formed, Cas3 protein is recruited for target DNA degradation. Modified from (Sashital *et al*, 2012; Westra *et al*, 2012).

1.5.3.2. Cas3

Cas3 protein is essential for CRISPR/Cas-encoded immunity in *E. coli* cells and is present in all Type I CRISPR/Cas systems (Brouns *et al*, 2008; Makarova *et al*, 2011b). According to sequence analysis, Cas3 proteins usually contain HD phosphohydrolase and Superfamily 2 (SF2) helicase domains, which can be covalently fused (Cas3) or encoded as individual proteins (Cas3') or (Cas3"), respectively (Makarova *et al*, 2006; Haft *et al*, 2005; Makarova *et al*, 2011b). In addition, the Cas3 protein or its domains sometimes are fused with other Cas proteins (Cas2-Cas3, Cas3-Cse1) (Westra *et al*, 2012; Makarova *et al*, 2011b). Cas3 functions together with the Cascade–crRNA complex (see 1.5.3.1) to defend cells against foreign DNA (Brouns *et al*, 2008; Westra *et al*, 2012). In addition, in *P. aeruginosa* Cas3 functions downstream of small crRNA production and activities of HD and helicase domains are required to suppress biofilm formation by phage infected cells (Cady & O'Toole, 2011).

The HD-type nuclease protein (Cas3') from *S. solfataricus* protein has been demonstrated to catalyze hydrolysis of double-stranded DNA and RNA, with the preference to G or C bases (Han & Krauss, 2009). However, all other biochemically characterized Cas3 proteins or HD domains are specific to single-stranded nucleic acids. Structural and biochemical analyses of the Cas3 protein HD domain from *T. thermophilus* HB8 (subtype I-E) have revealed Mg^{2+} or Ni^{2+} cation dependent endonuclease activity on single-stranded DNA (Mulepati & Bailey, 2011). The HD domain protein from *M. jannaschii* (MjCas3') (subtype I-A) has Mg^{2+} dependent endo- and 3'-5' exonuclease activity on single-stranded DNA and RNA, as well as on 3'-flaps, splayed arms, and R-loops which represent the potential intermediates of DNA degradation. The degradation of branched DNA substrates by MjCas3' is stimulated by the helicase domain containing protein (MjCas3") and ATP.

The biochemical characterization of the Cas3 protein from *S. thermophilus* DGCC7710 strain (subtype I-E) confirmed that N-terminal HD-like domain is

responsible for single-stranded DNA cleavage, while C-terminal helicase domain contains single-stranded DNA- stimulated ATPase activity, which is coupled to unwinding of DNA/DNA and RNA/DNA duplexes in the 3' to 5' direction (Sinkunas *et al*, 2011). The unwinding of DNA-RNA duplexes activity was also reported for Cas3 from *E. coli* (subtype I-E) (Howard *et al*, 2011). It has been shown recently that upon DNA target recognition *in vivo* Cas3 protein interacts with the Cascade complex via Cse1 protein. Both helicase and nuclease domains of Cas3 are important for CRISPR encoded immunity (Westra *et al*, 2012). However, due to its poor solubility and aggregation, the *E. coli* Cas3 protein can only be expressed in very limited amounts (Sinkunas *et al*, 2011; Howard *et al*, 2011; Westra *et al*, 2012). For this reason, to obtain a Cas3-Cascade effector complex *in vitro*, Cas3 was fused to one of the Cascade components, Cse1 protein. Fused Cas3-Cascade complex provided resistance *in vivo*, as well as destroyed plasmid containing crRNA matching proto-spacer plasmid *in vitro* (Westra *et al*, 2012).

1.5.3.3. Subtype I-F effector complex: Csy complex

P. aeruginosa CRISPR/Cas system belongs to subtype I-F and contains 6 genes (Figure 12, A). Four Cas (Csy1, Csy2, Csy3 and Cas6f) proteins and crRNA form a 350 kDa size ribonucleoprotein complex (Csy complex). Similarly to *E. coli* Cascade, Csy complex interacts with double-stranded DNA targets containing crRNA matching sequences (Wiedenheft *et al*, 2011a). Isothermal titration calorimetry experiments showed that target recognition by the Csy complex is enthalpically driven and confirmed that the DNA target selection mechanism is governed by a "seed" sequence (Wiedenheft *et al*, 2011a), which is also suggested for *E. coli* Cascade (Semenova *et al*, 2011; Wiedenheft *et al*, 2011b).

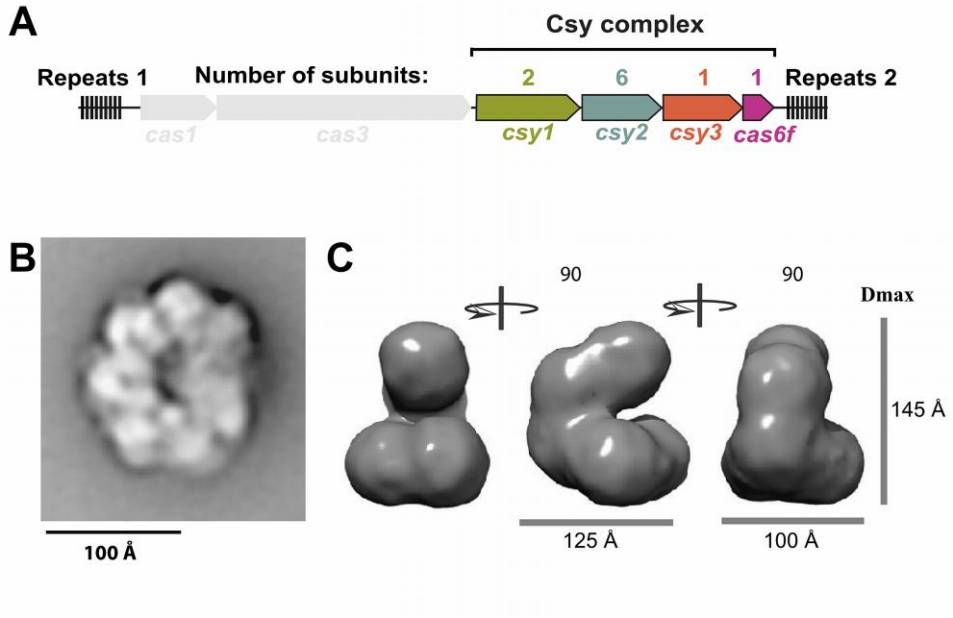

Figure 12. Csy protein complex. (**A)** The CRISPR/Cas system from *P. aeruginosa* consists of 6 *cas* genes and two CRISPR regions. The *csy1, csy2, csy3* and *cas6f* genes encode the Csy complex. The number of subunits in the complex is indicated above the genes. (**B)** Electron microscopy (EM) reconstruction of Csy complex. **C.** Small-angle X-ray scattering (SAXS) reconstruction of Csy complex reveals a crescent architecture. Adapted from (Wiedenheft *et al*, 2011a).

Structural analysis of the Csy complex by native mass spectrometry, electron microscopy and small-angle X-ray scattering showed a particle containing a crescent shape and confirmed an unequal subunit stoichiometry ($Csy1_1$:$Csy2_1$:$Csy3_6$: $Cas6f_1$:$crRNA_1$) (Figure 12). Although comprised of distinct proteins, the shape and stoichiometry of the Csy complex is reminiscent to the *E. coli* Cascade (see 1.5.3.1.). However, the Csy complex lacks the large tail, which is formed by Cse1 protein in *E. coli* Cascade (Wiedenheft *et al*, 2011a). Since Cse1 is suggested to play an important part in target DNA selection (Sashital *et al*, 2012), the mechanism of target localization and binding by Cascade and Csy complexes might differ.

1.5.3.4. Type I-A effector complex: aCascade

CRISPR/Cas systems belonging to subtype I-A are more complex than those belonging to subtypes I-E or I-F. The ribonucleoprotein complex from *S. sulfolobus* I-CRISPR/Cas system has been purified and characterized. One

of the three subtype I-A CRISPR/Cas systems of *S. solfataricus* consists of 13 genes (Figure 13 A). Tagged Cas7 protein expressed in *S. solfataricus* co-purifies together with crRNA and Cas5 protein and the complex was referred to as aCascade. Small amounts of Csa5 and Cas6 proteins also co-purify with aCascade, suggesting weak interactions or the existence of transient complexes. Electron microscopy showed that Cas7 proteins form a helix like backbone similarly to *E. coli* Cascade. In addition, recombinant Cas7 and Cas5 binds crRNA *in vitro* forming a complex which is able to bind complementary single-stranded DNA (Lintner *et al*, 2011).

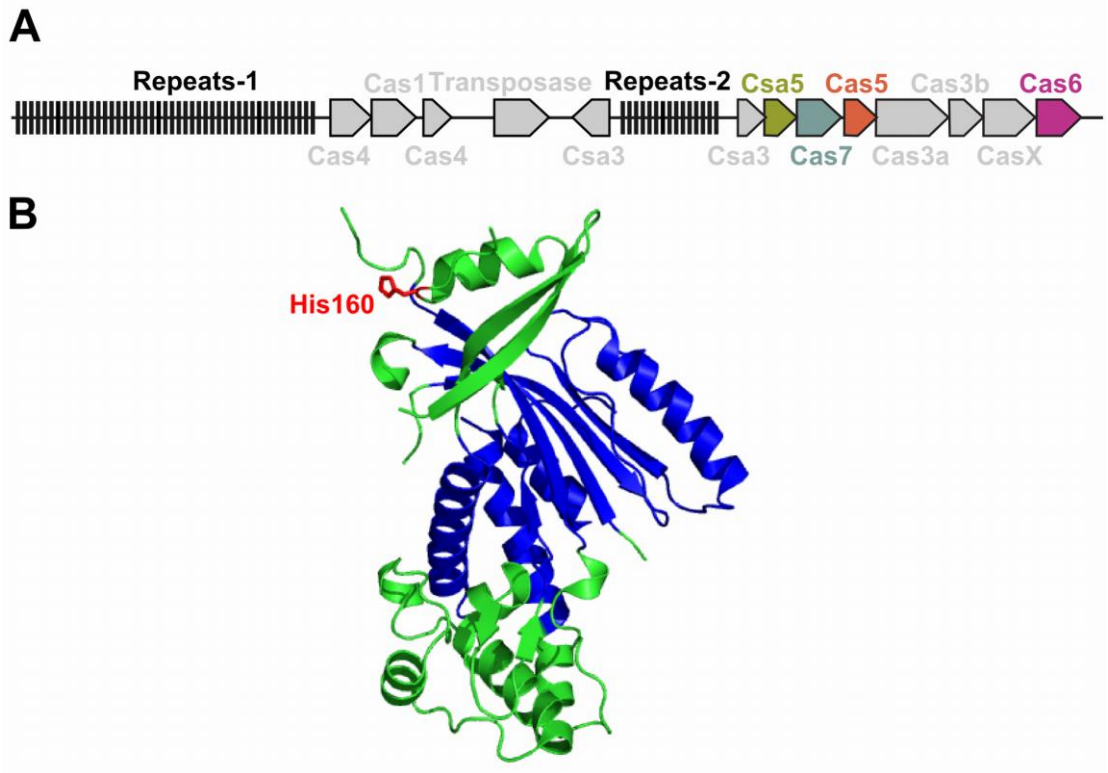

Figure 13. Formation of aCascade. (**A)** The subtype I-A CRISPR/Cas system from *S. solfataricus* consists of 12 *cas* genes and two CRISPR regions. Cas7, Cas5, Csa5 and Cas6 proteins co-purify with crRNA as aCascade. (**B)** Structure of Cas7 protein. Cas7 protein consists of three domains. Two flanking domains (green) are presented as insertions to the central ferredoxin-like domain. Conserved His160 residue, which might be involved in interactions with crRNA shown as sticks and labeled (Lintner *et al*, 2011).

The crystal structure of Cas7 revealed a crescent-shaped three-domain architecture (Figure 13 B). The modified ferredoxin-like domain, which frequently serves for interactions with RNA, is located in the center of the

structure. Two additional domains are presented as insertion in the ferredoxin-like domain (Lintner *et al*, 2011). Although the structure of Cas7 does not explain how proteins interact with each other to form aCascades helical spine, it revealed amino acid motifs which might be responsible for interaction with crRNA. Mutation of a conserved histidine residue to alanine in the possible RNA interaction surface significantly reduced Cas7 protein affinity for the crRNA (Lintner *et al*, 2011).

1.5.3.5. The Cmr complexes

Compared to bacteria, archaea usually have more CRISPR/Cas systems and they are more complex. Cas proteins and CRISPR loci are usually arranged separately. For example, *P. furiosus* contains three modules of Cas proteins (subtypes I-A, I-B and III-B) and seven CRISPR loci (Hale *et al*, 2008, 2009). In addition, *S. solfataricus* contains five Cas protein modules (two belonging to subtype III-B and three – subtype I-A) and 6 CRISPR loci (Zhang *et al*, 2012).

The subtype III-B CRISPR/Cas systems usually encode a Cas10 protein (putative polymerase/adenylate cyclase), and 5 or 6 subtype specific Cmr proteins, belonging to RAMP superfamily. Two subtype III-B effector complexes from *P. furiosus* and *S. solfataricus* (PfCmr and SsCmr complexes) have been purified and biochemically characterized. Differently from other CRISPR/Cas systems characterized today, both III-B effector (Cmr) complexes target single-stranded RNA, but not double-stranded DNA. However, the RNA cleavage mechanism seems to be different for both complexes (Zhang *et al*, 2012; Hale *et al*, 2012, 2009).

Cmr complex purified from *P. furiosus* cells (PfCmr), consists of 6 Cas proteins (Cas10, Cmr1, Cmr3-6) and crRNA (39 or 45 nt length) generated from 7 different CRISPR regions (Figure 14 A). Maturation of crRNA in *P. furiosus* undergoes in two steps (Figure 14 B), for this reason the crRNA in the PfCmr complex is shorter than the repeat-spacer unit (~ 70 bp). Firstly, a long CRISPR RNA transcript is cleaved by the Cas6 ribonuclease (see 1.5.2.1.) within the repeat sequence, leaving 8 nt length 5’ tag from repeat sequence.

Further pre-crRNA is processed by unknown mechanism within the spacer sequence, leaving 31 or 37 nt length 3’ tag. Therefore, *P. furiosus* crRNA always contains identical 8 nt length 3’ tag originated repeat, and unique 5’ tag of two possible lengths (Hale *et al*, 2008; Carte *et al*, 2008).

The *P. furiosus* Cmr complex cleaves complementary target RNA *in vitro* at the fixed 14 nt distance from the 3’ end of the crRNAs. Because the PfCmr complex is exploiting a molecular ruler mechanism and crRNA might be of two different lengths, the target RNA might be cleaved at two positions (Figure 14) (Hale *et al*, 2009). The active PfCmr complex was assembled *in vitro* using 6 recombinant Cas proteins and 39 or 45 nt length crRNA. Analysis showed that all proteins except Cmr5 are required for *in vitro* assembled complex activity, however, the candidate protein catalyzing target RNA cleavage is yet unknown (Hale *et al*, 2009). Although target RNA cleavage reaction by PfCmr complex requires divalent metal cations, the product has a 3’ phosphate or 2’-3’ cyclic phosphate terminus and a 5’ hydroxyl terminus. Such products are typical in metal independent RNA cleavage reactions, where 2’ hydroxyl is used as nucleophile to attack the scissile 3’phosphate (Yang, 2011). The 8 nt length 5’ tag of crRNA is critical for PfCmr complex activity, and synthetic crRNAs bearing 5’ tag can be used to direct cleavage of novel targets (Hale *et al*, 2012). Only recently the evidence for PfCmr complex RNA cleavage activity *in vivo* has been found. Deep sequencing confirmed that the endogenous RNA, complementary to CRISPR loci, is cleaved in *P. furiosus* cells (Hale *et al*, 2012).

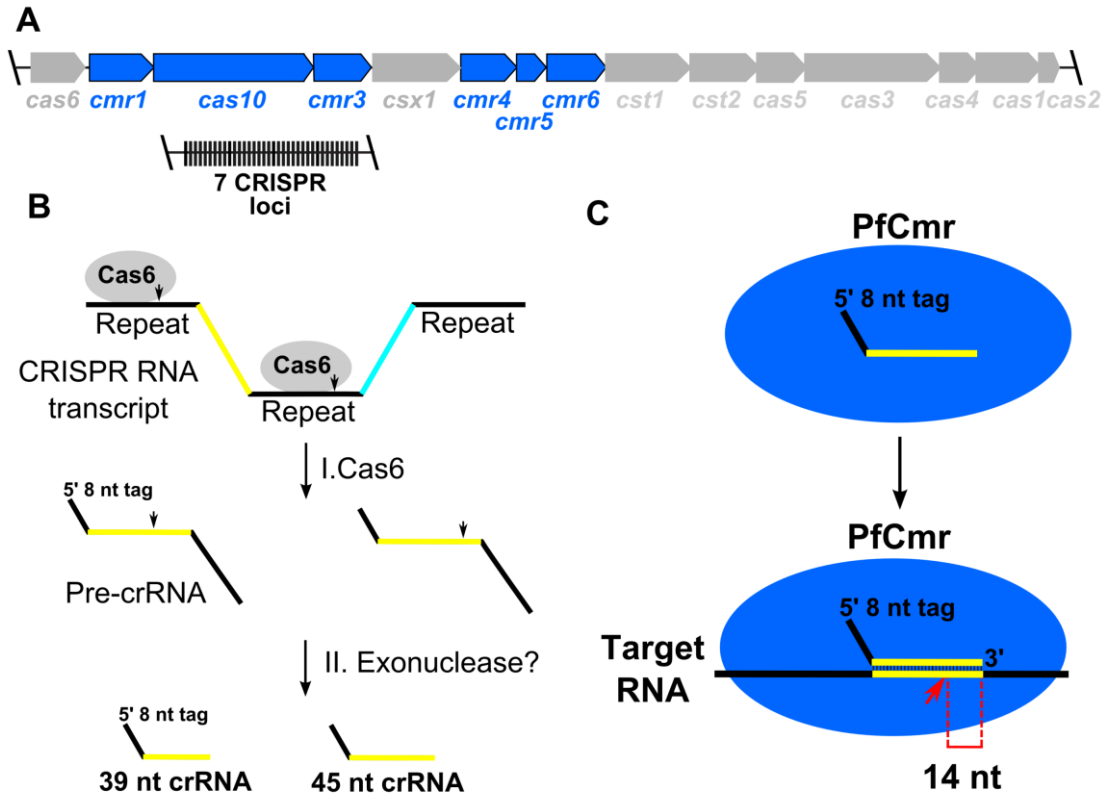

Figure 14. Cmr ribonucleoprotein complex from *P. furiosus*. (**A)** The subtype III-B CRISPR/Cas system from *P. furiosus* encodes 15 *cas* genes. Cmr1, Cas10 and Cmr3-6 proteins together with crRNA from 7 distinct CRISPR loci form a PfCmr complex. (**B)** Schematic representation of the crRNA maturation pathway in *P. furiosus*. The CRISPR RNA transcript initially is cleaved by Cas6 ribonuclease within the repeat, forming an 8 nt tag at the 5' terminus originated from repeat. Next, pre-crRNA is processed further by an unknown mechanism, where 39 nt and 45 nt length crRNA is formed (contains 31 and 37 nt from spacer, accordingly) (Hale *et al*, 2008; Carte *et al*, 2008; Hale *et al*, 2009). (**C)** Target RNA cleavage by Cmr complex. Mature crRNA is incorporated into PfCmr complex and guide recognition of target RNA sequence. Molecular ruler mechanism is exploited, PfCmr cleaved target RNA specifically in one position 14 nt from 3' end of crRNA (Hale *et al*, 2009).

Cas10 protein, the largest subunit of Cmr complex, was predicted to have a permutated HD superfamily nuclease domain at N-terminus, a zinc finger domain and a polymerase-like domain (Makarova *et al*, 2011b). It was shown that the HD nuclease domain of Cas10 is not required for PfCmr complex mediated target RNA cleavage *in vitro* (Cocozaki *et al*, 2012). Crystal structure of the truncated Cas10 protein lacking N-terminal HD domain (Cas10 ΔHD), revealed a flat, triangular, four domain architecture: two ferredoxin-like domains and two small α-helical domains (Figure 15) (Cocozaki *et al*, 2012;

Zhu & Ye, 2012). Two ferredoxin-like domains form a structure similar to adenylyl-cyclase homodimer, instead of previously predicted polymerase. After crystal soaking with ADP, the structure of Cas10ΔHD bound ADP and metal ions has been obtained, however, neither free Cas10 protein, nor Cmr complex does not catalyzed ATP hydrolysis. In addition, mutations of the conserved ATP and divalent metal ion binding residues did not abolish RNA cleavage activity of the PfCmr complex (Cocozaki *et al*, 2012). Thus, Cas10 most probably plays a structural, rather than catalytical, role in the PfCmr complex (Cocozaki *et al*, 2012; Huang, 2012).

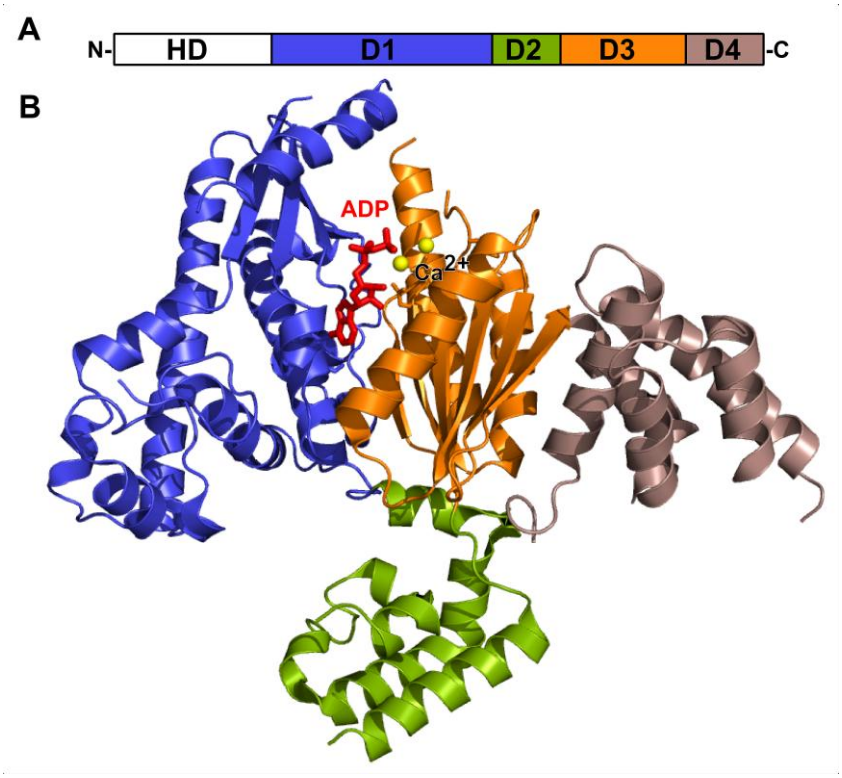

Figure 15. Structure of Cas10 protein. (A) Schematic representation of Cas10 protein from *P. furiosus* domain organization. N-terminal HD domain is not important for Cmr complex cleavage activity *in vitro*. (**B**) Crystal structure of Cas10 protein lacking the N-terminal HD domain bound ADP (red) and Ca^{2+} (yellow) ions. Ferredoxin like domains (D1, blue; D3, orange) forms a structure similar to adenylyl-cyclase homodimer. ADP and metal ions are bound between D1 and D3 domains. D2 and D4 are small helical domains (Cocozaki *et al*, 2012; Zhu & Ye, 2012).

The Cmr complex purified from *S. sulfolobus* (SsCmr) consists of 7 Cas proteins (Cmr1, Cas10 and Cmr3-7) and various length crRNA, containing 8 nt 3' tag (Figure 16 A). Interestingly, the Cmr7 protein is presented only in *S.*

solfataricus species. SsCmr complex cleaves RNA at UA dinucleotide and, differently from the PfCmr complex, does not exploit a molecular ruler mechanism. This complex can cleave both target and guide RNA (crRNA) sequences *in vitro*, although a single crRNA molecule can support the degradation of multiple RNA targets (Zhang *et al*, 2012). The electron microscopy structure of full Cmr complex and subcomplex (only Cas10, Cmr3 and Cmr7 proteins) devoid of crRNA was obtained. The overall form of the structure is different from *E. coli* Cascade, and is similar to a clamp containing a deep cleft, which could accommodate double-stranded RNA (Figure 16 B) (Zhang *et al*, 2012).

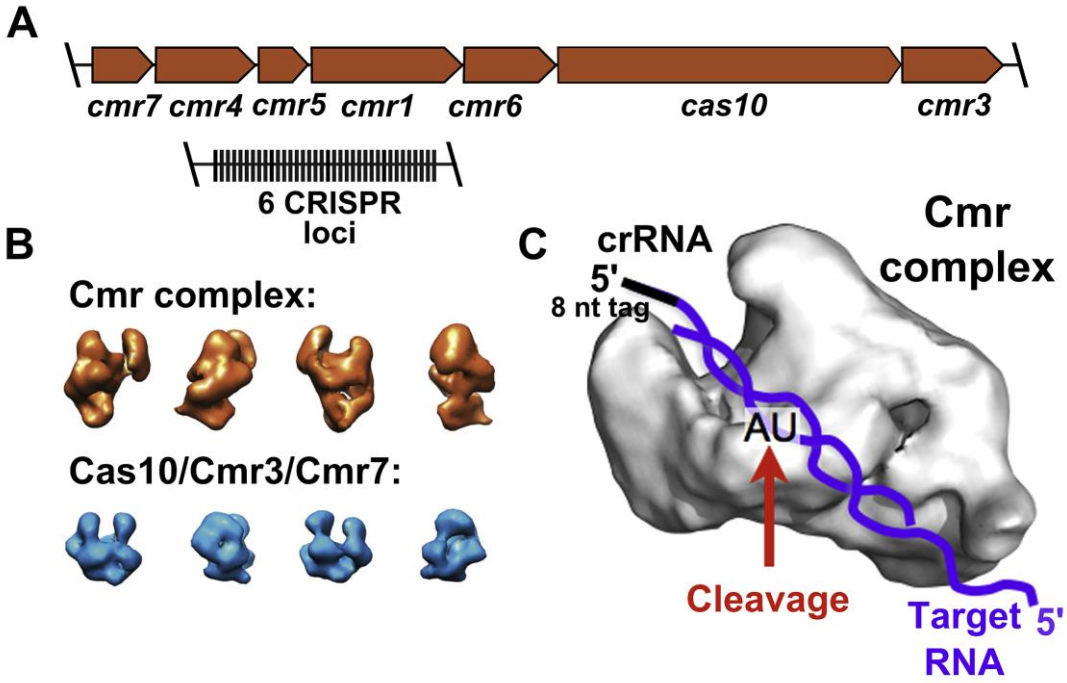

Figure 16. Cmr complex from *S. solfataricus* (SsCmr). (A) 7 Cas proteins (Cmr1, Cas10, Cmr3-7) and crRNA forms SsCmr complex. crRNA copurified with Cmr proteins might be derived form 6 distinct CRISPR loci. (**B)** Surface representation of electron microscopy structures of full SsCmr complex (7 proteins and crRNA) and subcomplex devoid of crRNA (Cas10/Cmr3/Cmr3). Adapted from (Zhang *et al*, 2012). **C**. Model of SsCmr complex with bound target RNA. SsCmr complex forms clamp like structure. crRNA together with bound target RNA most probably are located in the cleft. Both target RNA, and crRNA is cleaved at AU dinucleotides *in vitro*, but SsCmr complex also could cleave multiple targets. Adapted from (Zhang *et al*, 2012).

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Chemicals

Chemicals used in this study were brought from “Sigma-Aldrich”, “Roth”, “Fluka” or “Thermo Scientific”, and used without further purification.

2.1.2. Enzymes

T4 polynucleotide kinase; DreamTaq, TaqI and PfuI DNA polymerases; FastAP thermosensitive alkaline phosphatase; T4 DNA ligase; T4 DNA polymerase; bovine serum albumin; FastDigest restriction enzymes used in this study were obtained from Thermo Fisher Scientific (former “Fermentas”). All these products were used according to the manufacturer’s instructions.

2.1.3. Kits for molecular biology

TranscriptAid T7 High Yield Transcription Kit, CloneJET PCR Cloning Kit, Rapid DNA Ligation Kit, GeneJET Gel Extraction Kit, GeneJET PCR Purification Kit, GeneJET Plasmid Miniprep Kit were purchased from Thermo Fisher Scientific. miRNeasy Mini kit was purchased from Qiagen. Kits were used according to the manufacturer’s instructions.

2.1.4. Bacterial and viral strains

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) used in the cloning and Cas9-Chis protein expression. *E. coli* strain RR1 (F^{-} mcrB mrr hsdS20(r_B^{-} m_B^{-}) leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(Sm^{R}) glnV44 λ^{-}) (Bolivar *et al*, 1977) were used for transformation experiments and Cas9-crRNA complex expression. *E. coli* cells were grown in LB medium

supplemented with ampicillin (Ap) (50 μg/ml) and/or chloramphenicol (Cm) (10 μg/ml) when necessary.

Virulent lambda phage (λ_{vir}) was obtained from dr. R. Sapranauskas.

2.1.5. Plasmids and DNA

Plasmid vectors pACYC184 (Cm^r) (Chang & Cohen, 1978) and pUC18 (Ap^r) (Yanisch-Perron *et al*, 1985) were used for cloning and subcloning procedures. pASK-IBA3plus (Ap^r) (IBA) and pBAD24-CHis (Ap^r) (Sinkunas *et al*, 2011) expression vectors were used to obtain Cas9 protein fused with C-terminal Strep-tag and His-tag sequences respectively.

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

2.1.6. Oligonucleotide substrates

All oligonucleotide substrates used in this study are listed in Table 3. Oligodeoxyribonucleotides were purchased from Metabion. The 5′-ends of oligonucleotides were radioactively labeled using PNK (Thermo Fisher Scientific) and [γ-^{33}P]ATP (Hartmann Analytic). Duplexes were made by annealing two oligonucleotides with complementary sequences (SP1, SP1-Δp, SP2). A radioactive label was introduced at the 5′ end of individual DNA strands prior to the annealing with unlabeled strands.

2.1.7. Buffers

Reaction buffer - 10 mM Tris-HCl pH=7.5, 10 mM NaCl, 10 mM $MgCl_2$, 0.1 mg/ml BSA.

Binding buffer - 40 mM Tris–acetate, pH 8.3 at 25 C, 0.1 EDTA, 0.1 mg/ml BSA, 10% v/v glycerol.

Loading dye solution - 0.01 % bromphenol blue and 75 mM EDTA in 50 % v/v glycerol.

Table 3. Oligonucleotide substrates. Proto-spacer sequence is underlined, PAM is on bold.

Oligonucleotide	Sequence	Specification
SP1	5′-GCTCGAATTGAAATTCTAAACGCTAAAGAGGAAGAGGACAT**GG**T**G**AATTCGTAAT-3′ 3′-CGAGCTTAACTTTAAGATTTGCGATTTCTCCTTCTCCTGTA**CC**A**C**TTAAGCATTA-5′	55 bp oligoduplex substrate containing proto-spacer1 and PAM
SP1-pΔ	5′-GCTCGAATTGAAATTCTAAACGCTAAAGAGGAAGAGGACAAATTCGTAAT-3′ 3′-CGAGCTTAACTTTAAGATTTGCGATTTCTCCTTCTCCTGTTTAAGCATTA-5′	50 bp oligoduplex substrate containing proto-spacer2
SP2	5′-GCTCGAATTGTACTGCTGTATTAGCTTGGTTGTTGGTTTGT**GG**T**G**AATTCGTAAT-3′ 3′-CGAGCTTAACATGACGACATAATCGAACCAACAACCAAACA**CC**A**C**TTAAGCATTA-5′	50 bp oligoduplex substrate containing proto-spacer2 and PAM (oligoduplex without proto-spacer1)
s(+)SP1	5′-ATTACGAATT**C**A**CC**ATGTCCTCTTCCTCTTTAGCGTTTAGAATTTCAATTCGAGC-3′	55 nt ssDNA oligonucleotide substrate, (+) strand of SP1 oligoduplex
s(+)SP1-pΔ	5′-ATTACGAATTTGTCCTCTTCCTCTTTAGCGTTTAGAATTTCAATTCGAGC-3′	50 nt ssDNA oligonucleotide substrate, (+) strand of SP1-pΔ oligoduplex
s(+) SP2	5′-ATTACGAATT**C**A**CC**ACAAACCAACAACCAAGCTAATACAGCAGTACAATTCGAGC-3′	55 nt ssDNA oligonucleotide substrate, (+) strand of SP2 oligoduplex
s(-)SP1	5′-GCTCGAATTGAAATTCTAAACGCTAAAGAGGAAGAGGACAT**GG**T**G**AATTCGTAAT-3′	55 nt ssDNA oligonucleotide substrate, (-) strand of SP1 oligoduplex
SP1-20	5′-GCTCGAATTGCGCTAAAGAGGAAGAGGACAT**GG**T**G**AATTCGTAAT-3′ 3′-CGAGCTTAACGCGATTTCTCCTTCTCCTGTA**CC**A**C**TTAAGCATTA-5′	45 nt oligoduplex substrate containing 20 nt of proto-spacer1 and PAM
SPN	5′-GCTCGAATTGCCACCCAGCAAAATTCGGTTTTCTGGCTGAT**GG**T**G**AATTCGTAAT-3′ 3′- CGAGCTTAACGGTGGGTCGTTTTAAGCCAAAAGACCGACTA**CC**A**C**TTAAGCATTA-5′	55 bp oligoduplex substrate containing proto-spacerN and PAM

2.2. Methods

2.2.1. Electrophoresis

2.2.1.1. Non-denaturing electrophoresis through agarose

Separation of supercoiled, open circular and linear forms of plasmid DNA was performed in 0.8 % agarose gels in the electrophoretic buffer containing 100 mM H_3BO_3-NaOH, 15 mM CH_3COONa, 2 mM EDTA (pH 8.2 at room temperature) and 0.5 μg/ml ethidium bromide. DNA samples were mixed with 1/3 volume of loading dye solution (75 mM EDTA, pH 9.0, 0.01 % bromphenol blue and 50 % (v/v) glycerol) and electrophoresed at 3 V/cm until the bromphenol blue dye migrated for 3 cm. Digital images of the gels were taken with BioDocAnalyze (Biometra) gel documenting systems.

DNA fragments required for genetic engineering procedures were separated in 0.8-1.5 % agarose gels in the ethidium bromide-free electrophoretic buffer containing 40 mM Tris-acetate and 0.1 mM EDTA (pH 8.3 at room temperature). The gel slices containing required DNA fragments were excised according to the ethidium bromide stained markers. DNA was recovered using GeneJET Gel Extraction Kit.

2.2.1.2. Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel electrophoresis (PAGE) was employed in nucleic acid binding experiments (see 2.2.11.). The electrophoretic buffer consisted of 40 mM Tris-acetate (pH 8.3 at room temperature) and 0.1 mM EDTA. The gels consisted of 8 % acrylamide/N,N'-methylenebisacrylamide (29:1 w/w) in the electrophoretic buffer, polymerisation was initiated by adding TEMED and ammonium persulfate. The gels were 1 mm thick and ~20 cm in length. Prior to gel casting, one of the glass plates was processed with “bind silane” (3-methacryloxypropyltrimethoxysilane) and the other with “repeal silane” (5 % v/v dichlorodimethylsilane in $CHCl_3$).

The binding buffer consisted of 40 mM Tris-acetate (pH 8.3 at room temperature), 0.1 mM EDTA, 0.1 mg/ml BSA and 10 % v/v glycerol. Samples of radiolabeled DNA, binding buffer and protein were mixed in the wells of ELISA plate, left for 15 min at room temperature and then loaded on the gel. Electrophoresis was run at room temperature for 3 hours at ~5 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 Cyclone™ phosphorimager (Perkin Elmer). The amounts of various DNA fragments were quantified with OptiQuant 3.0 software (Perkin Elmer).

2.2.1.3. Denaturing polyacrylamide gel electrophoresis

Denaturing PAGE was employed in oligonucleotide DNA cleavage studies (see 2.2.8.). The electrophoretic buffer consisted of 100 mM Tris-borate (pH 8.2 at room temperature) and 2 mM EDTA. The gels consisted of 20 % acrylamide/ N,N'-methylenebisacrylamide (29:1 w/w) and 8.5 M urea in the electrophoretic buffer.

Samples of radiolabeled DNA were mixed 1:1 (v/v) with loading dye solution (25 mM EDTA, pH 9.0, 0.01 % w/v bromphenol blue and 95 % v/v formamide), placed for 4 min into a bath with 95 °C water and chilled on ice. Electrophoresis was run for 30 min at 30 V/cm without the samples and for another 4 hours with the DNA samples. Gels were dried and analysed as described above (2.2.1.2).

2.2.1.4. Denaturing (SDS) polyacrylamide gel electrophoresis of proteins

Denaturing SDS-PAGE of proteins was employed to verify homogeneity of protein preparations and to determine amount of Cas9-crRNA complex. The electrophoretic buffer consisted of 25 mM Tris, 190 mM glycine (pH 8.3 at room temperature) and 0.1 % SDS (w/v). The stacking and separating gels

consisted of 4 % (in 125 mM Tris-HCl (pH 6.8 at room temperature) and 0.1 % SDS (w/v)) and 12 % (in 375 mM Tris-HCl (pH 8.8 at room temperature) and 0.1 % SDS (w/v)) acrylamide/N,N'-methylenebisacrylamide (37.5:1 (w/w)), respectively (Sambrook *et al*, 1989). Protein samples were mixed 1:1 (v/v) with sample buffer (100 mM Tris-HCl (pH 6.8 at room temperature), 4 % SDS (w/v), 200 mM DTT, 20 % (v/v) glycerol) and placed for 4 min into a bath with 95 °C water before loading. Electrophoresis was run at room temperature for 1-1.5 hours at ~30 V/cm. Gels were stained with Page Blue protein staining solution (Thermo Fisher Scientific). Amounts of protein were quantified by densitometric analysis using ImageJ software (Schneider *et al*, 2012).

2.2.2. DNA manipulations

2.2.2.1. Isolation of DNA and recombinant DNA techniques

Plasmids were isolated using GeneJET Plasmid Miniprep Kit. Standard procedures (Sambrook *et al*, 1989) were used for generation of recombinant plasmid. The GeneJET Gel extraction Kit was used for isolation of DNA fragments from agarose gels. DNA sequencing was carried out on an ABI PRISM 377 sequencer by using The BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) in the Sequencing center of the VU Institute of Biotechnology. All enzymes used for DNA manipulations and corresponding buffers were obtained from Thermo Fisher Scientific.

2.2.2.2. Construction of plasmid substrates

For construction of plasmids used in the transformation assay and plasmid cleavage assay, the pUC18 vector was cut with EcoRI and dephosphorylated with FastAP alkaline phosphatase, oligoduplexes (Table 4) containing sticky EcoRI ends were assembled by annealing complementary oligonucleotides (Metabion), phosphorylated with T4 polynucleotide kinase and ligated using T4 ligase.

2.2.2.3. Cloning of full CRISPR3/Cas system

Three separate DNA fragments I, II and III covering the CRISPR3/Cas module in *S. thermophilus* DGCC7710 and carrying SapI sites at the ends were obtained by PCR with the following primer pairs: 5'-GCTCTTCTGTCGGTCATTGAGATCTTGGATGAGG-3' and 5'-GCTCTTCAGACTTACCTTGATTGGTATATTG-3' (fragment I), 5'-GCTCTTCAGTCTAATTCCCAACAACGCTTG-3' and 5'-GCTCTTCCGACAACCTCACGTGCAAAC-3' (fragment II), and 5'-GCTCTTCTGTCGTCTGTGGTTGTATGAC-3' and 5'-GCTGGATATTCGTATAACATGTC-3' (fragment III). Amplified DNA containing SapI sites at the ends was cloned into pACYC184 vector, pre-cleaved with EcoRV to generate recombinant plasmids pCR1, pCR2 and pCR3 carrying fragments I, II and III, respectively (Figure 17). The resulting plasmids were purified and their inserts sequenced. The 7.6-kb DNA fragment containing the full-length CRISPR3/Cas system was reassembled during subsequent sub-cloning steps. First, insert from the pCR2 plasmid was cut with SapI, purified and ligated into pCR3 pre-cleaved with SapI to yield plasmid pCR2+3. Subsequently, an insert was cut from pCR1 with SapI, purified and ligated into pCR2+3 cut with SapI to generate the pCRISPR3 plasmid, which contains the full-length CRISPR3/Cas system. The nucleotide sequence of the final insert was confirmed by sequencing. The system was cloned by dr. R. Sapranauskas.

2.2.2.4. Frameshift mutations/deletions of *cas* genes

To inactivate *cas* genes, frameshift mutations or small deletions were created. The pCas9(-) plasmid was constructed by cutting pCRISPR3 with Eco105I-Bpu1102I. The inactivation of *cas1*, *cas2* and *csn2* genes was performed in two steps. First, the blunt-ended Eco31I-NheI DNA fragment from pCRISPR3, which encodes all three genes, was subcloned into HindIII-EcoRI pre-cleaved and blunt-ended pUC18 vector. The resulting plasmid pUC-CRISPR3del was used for subsequent inactivation of *cas* genes. For *cas1*

inactivation, pUC-CRISPR3del was cleaved with Eco72I-Eco47III and re-ligated with T4 DNA ligase to generate the recombinant pUCΔ*cas1* plasmid which has a 41 bp deletion. For *cas2* gene inactivation, pUC-CRISPR3del was cleaved with XagI, blunt-ended and re-ligated to generate the recombinant plasmid pUCΔ*cas2* which has a 1-bp insertion leading to a frameshift mutation in *cas2*. For the *csn2* gene inactivation, the pUC-CRISPR3 plasmid was cut with EcoRI, blunt-ended and re-ligated to generate the pUCΔ*csn2* plasmid which has a 4-bp insertion in *csn2* gene. Mutated genes in all three cases were transferred into the pCRISPR3 plasmid by subcloning BstXI-Eco147I DNA fragments from pUCΔ*cas1*, pUCΔ*cas2*, and pUCΔ*csn2* plasmids. Recombinant plasmids were called pCas1(-), pCas2(-) and pCsn2(-), respectively. To obtain p Cas9 plasmid lacking *cas1*, *cas2* and *csn2* genes, first the Eco47III-BcuI deletion in pUC-CRISPR3 was obtained and then the BstXI-Eco147I DNA fragment from pUCΔ[*cas1-csn2*] plasmid was subcloned into pCRISPR3 plasmid to yield pCas9 plasmid.

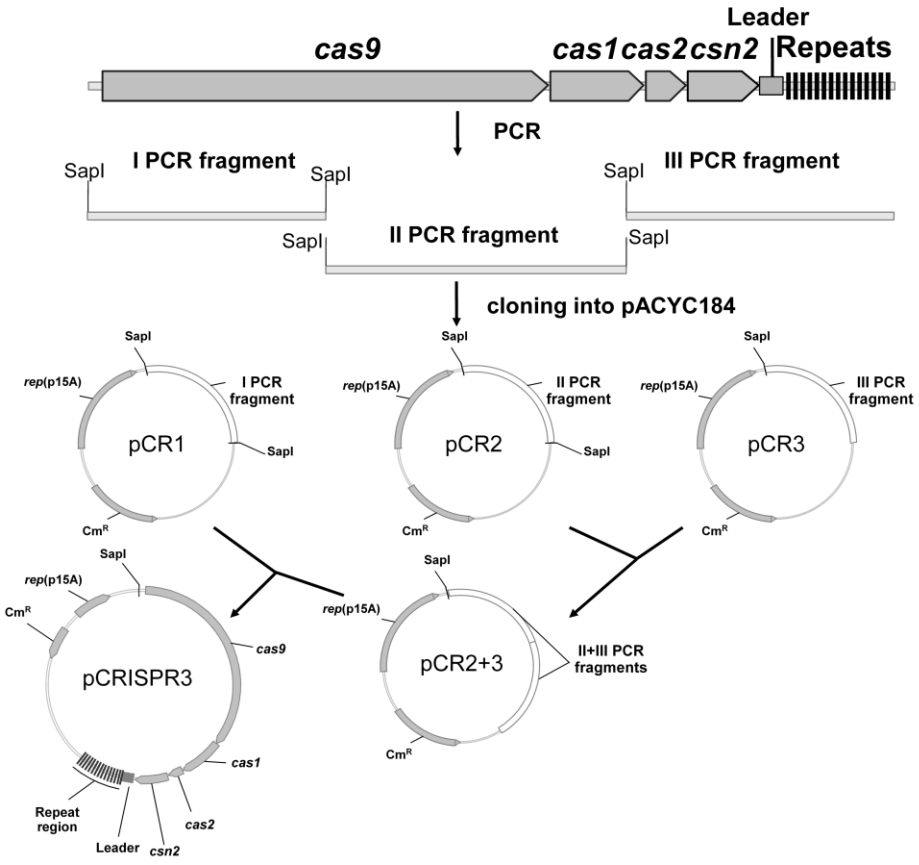

Figure 17. **Cloning of the CRISPR3/Cas system.**

Table 4. Plasmid names and inserts' sequences.

Recombinant plasmid	Inserted oligoduplexes*
pSP1	5'-AATTG**AAATTCTAAACGCTAAAGAGGAAGAGGACA**TGGTG-3' 3'-C**TTTAAGATTTGCGATTTCTCCTTCTCCTGT**ACCACTTAA-5'
pSP2	5'-AATT**TACTGCTGTATTAGCTTGGTTGTTGGTTTG**TGGTG-3' 3'-**ATGACGACATAATCGAACCAACAACCAAAC**ACCACTTAA-5'
pSP1-sA2G	5'-AATTG**AGATTCTAAACGCTAAAGAGGAAGAGGACA**TGGTG-3' 3'-C**TCTAAGATTTGCGATTTCTCCTTCTCCTGT**ACCACTTAA-5'
pSP1-sC11T	5'-AATTG**AAATTCTAAATGCTAAAGAGGAAGAGGACA**TGGTG-3' 3'-C**TTTAAGATTTACGATTTCTCCTTCTCCTGT**ACCACTTAA-5'
pSP1-sG18C	5'-AATTG**AAATTCTAAACGCTAAACAGGAAGAGGACA**TGGTG-3' 3'-C**TTTAAGATTTGCGATTTGTCCTTCTCCTGT**ACCACTTAA-5'
pSP1-sA23T	5'-AATTG**AAATTCTAAACGCTAAAGAGGATGAGGACA**TGGTG-3' 3'-C**TTTAAGATTTGCGATTTCTCCTACTCCTGT**ACCACTTAA-5'
pSP1-sA25T	5'-AATTG**AAATTCTAAACGCTAAAGAGGAAGTGGACA**TGGTG-3' 3'-C**TTTAAGATTTGCGATTTCTCCTTCACCTGT**ACCACTTAA-5'
pSP1-sA28T	5'-AATTG**AAATTCTAAACGCTAAAGAGGAAGAGGTCA**TGGTG-3' 3'-C**TTTAAGATTTGCGATTTCTCCTTCTCCAGT**ACCACTTAA-5'
pSP1-pΔ	5'-AATTG**AAATTCTAAACGCTAAAGAGGAAGAGGACA**-3' 3'-C**TTTAAGATTTGCGATTTCTCCTTCTCCTGT**TTAA-5'
pSP1-pG1C	5'-AATTG**AAATTCTAAACGCTAAAGAGGAAGAGGACA**TCGTG-3' 3'-C**TTTAAGATTTGCGATTTCTCCTTCTCCTGT**AGCACTTAA-5'
pSP1-pG2C	5'-AATTG**AAATTCTAAACGCTAAAGAGGAAGAGGACA**TGCTG-3' 3'-C**TTTAAGATTTGCGATTTCTCCTTCTCCTGT**ACGACTTAA-5'
pSP1-pG4C	5'-AATTG**AAATTCTAAACGCTAAAGAGGAAGAGGACA**TGGTC-3' 3'-C**TTTAAGATTTGCGATTTCTCCTTCTCCTGT**ACCAGTTAA-5'
pSP1-27	5'-AATTG**TTCTAAACGCTAAAGAGGAAGAGGACA**TGGTG-3' 3'-C**AAGATTTGCGATTTCTCCTTCTCCTGT**ACCACTTAA-5'
pSP1-23	5'-AATTG**AAACGCTAAAGAGGAAGAGGACA**TGGTG-3' 3'-C**TTTGCGATTTCTCCTTCTCCTGT**ACCACTTAA-5'
pSP1-19	5'-AATTG**GCTAAAGAGGAAGAGGACA**TGGTG-3' 3'-C**CGATTTCTCCTTCTCCTGT**ACCACTTAA-5'
pSP1-15	5'-AATTG**AAGAGGAAGAGGACA**TGGTG-3' 3'-C**TTCTCCTTCTCCTGT**ACCACTTAA-5'
pSP1-11	5'-AATTG**GGAAGAGGACA**TGGTG-3' 3'-C**CCTTCTCCTGT**ACCACTTAA-5'

* - spacer sequences are shown in bold, PAM nucleotides are underlined and positions of mutations are in grey shading.

2.2.2.5. Cloning of *cas9* gene

Genomic DNA of *S. thermophilus* DGCC7710 was used as a template in PCR reactions to clone *cas9.* To generate pASKIBA3-Cas9, which was used

for the expression of the C-terminal Strep-tagged Cas9 protein variant, a PCR fragment was amplified with primers: 5′-ACGTCTCAAATGTTGTTTAATAAGTGTATAATAATTTC-3′ and 5′-ACGTCTCCGCGCTACCCTCTCCTAGTTTG-3′. The product was subsequently cloned into the pASK-IBA3 expression vector using Esp3I sites. To generate the pBAD-Cas9 plasmid which was used for the expression of the C-terminal $(His)_6$-tagged Cas9 protein variant, PCR amplification was performed using primers: 5-ACGTCTCACATGACTAAGCCATACTCAATT GGAC-3′ and 5′-ACTCGAGACCCTCTCCTAGTTTGGCAA-3′. This product was subsequently cloned into the pBAD24-Chis expression vector using NcoI and XhoI sites. DNA sequencing of *cas9* in pASKIBA3-Cas9 and pBAD-Cas9 plasmids revealed no difference with the wild-type chromosomal *cas9* sequence.

2.2.2.6. Manipulations with CRISPR region

To obtain plasmids pCas9(-)SP1 (Figure 27B) and pCRISPR3-SP1 (Figure 28A) bearing a single spacer1 (5′-AAATTCTAAACGCTAAAG AGGAAGAGGACA-3′) and two repeats, a PCR fragment amplified from pCRISPR3 plasmid with the following primers 5′-GACCACTTATTGAGGTAAATGAG-3′ and 5′-CAAACCAGGATCC AAGCTAATACAGCAG-3′ (BamHI site is underlined) was cloned into pCas9(-) and pCRISPR3 plasmids over Eco31I and BamHI, respectively.

Anti-λ-spacer was designed by selecting a sequence within the phage lambda *o* gene important for the phage genome replication. The proto-spacer sequence in the phage genome was preceded by the PAM sequence (TGGTG) identical to that used in the plasmid transformation assay. Insertion of the anti-λ-spacer into the pCRISPR3 plasmid was performed in several steps (Figure 18).

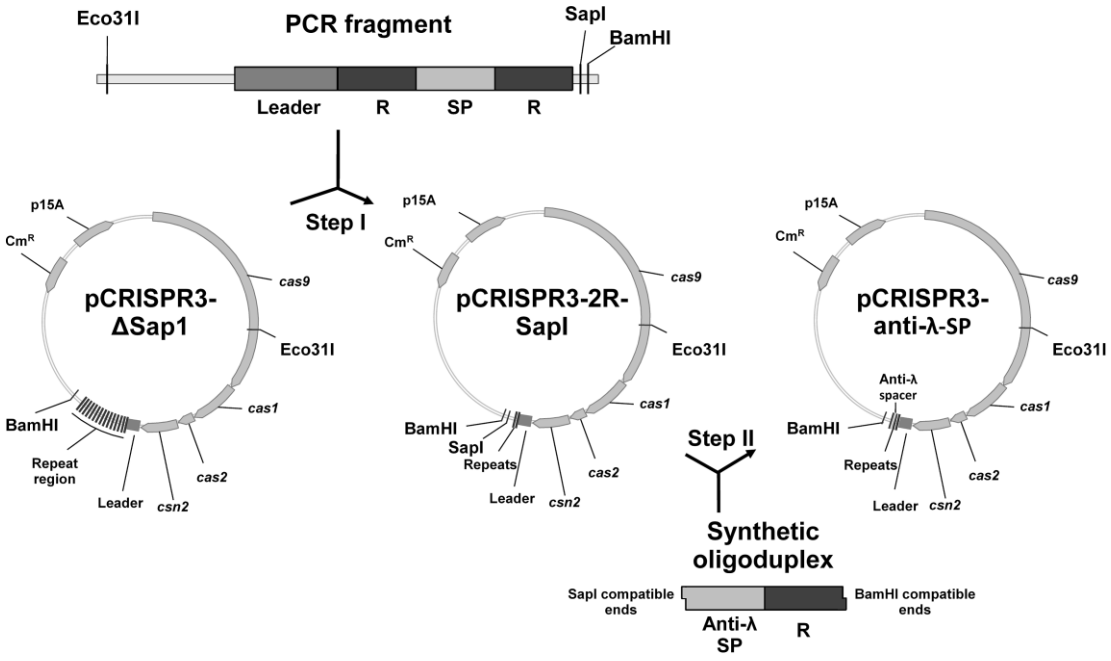

Figure 18. Insertion of the anti-λ-spacer into the pCRISPR3 plasmid.

First, the unique SapI restriction endonuclease site was removed from pCRISPR3 plasmid by XbaI-SapI cleavage followed by re-ligation with T4 DNA ligase to generate pCRISPR3-ΔSapI plasmid. The anti-λ spacer insertion required two additional cloning steps (I and II). In the step I, PCR fragment amplified from pCRISPR3 plasmid with following primer pair: 5'-GACCACTTATTGAGGTAAATGAG-3'/5'-AACCAGGATCCAAGCTA ATAGCTCTTCAGTTTTGGAACCATTC-3' (SapI (GCTCTTC) and BamHI (GGATCC) sites are underlined) was cloned into pCRISPR-ΔSapI and pCas9-ΔSapI plasmid pre-cleaved with Eco31I and BamHI to introduce the SapI site after the first or the second repeat and remove remaining part of CRISPR region. In the second cloning step II, the synthetic oligoduplex containing spacer and repeat unit and sticky ends for the SapI and BamHI restriction endonucleases was assembled (5'-AACCCACCCAGCAAAATTCGGTTTTC TGGCTGAGTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACG-3'/ 5'-GATCCGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACTCA GCCAGAAAACCGAATTTTGCTGGGTGG-3', the anti-λ spacer sequence are underlined) and cloned into pCRISPR3-2R-SapI plasmid through SapI and

BamHI sites to yield pCRISPR3-anti-λ-SP plasmid containing the anti-λ spacer in second position.

The pCas(-)SPN, pCas(-)SP-A and pCas(-)SP-B plasmids encoding corresponding crRNAs were constructed by inserting spacer - repeat unit encoding oligoduplexes, containing spacer sequences (5'-CCACCCAGCAA AATTCGGTTTTCTGGCTG-3', 5'-TTAATGTCATGATAATAATGGTTTC TTAGACGTC-3' and 5'-ACGAGCCGGAAGCATAAAGTGTAAAGCCTG-3', respectively) into pCas(-)-R plasmid over BamHI and SapI sites. pCas(-)-R plasmid were obtained by inactivating *cas9* gene and leaving 1 repeat in the CRISPR region as described above.

2.2.2.7. Site-directed mutagenesis

Megaprimer method (Barik, 1996) was used to obtain plasmids carrying a mutated *cas9* gene in pCas9 plasmid. To obtain pD31A plasmid encoding D31A mutant, the direct/ reverse primer pair 5'-CTAACGCAGTCAGGCACCGTG-3'/5'-CTATTCGTTCCAATAGCAAG TCCAATTGAG-3 and pCas9 plasmid template were used in the PCR reaction to obtain a megaprimer, which was purified, and used in the next PCR round together with a 5'-CTGATATGGTATCGAACCATTGTC-3' primer. The amplified DNA was cut with BglII and Eco105I and ligated into pCas9 plasmid precut with BglII and Eco105I. To obtain pH868A plasmid encoding H868A mutant, the direct/ reverse primer pair 5'-GTAATTATGATATTGATGCTATTATTCCTCAAGC-3'/5'-CTCATTAA TTGTGTTACTTATTCTC-3 and pCas9 plasmid template were used in the PCR reaction to obtain a megaprimer, which was purified, and used in the next PCR round together with a 5'-GATAGAGAGATGATAAAACAACGT-3' primer. The amplified DNA was cut with Eco31I and Bpu1102I and ligated into pCas9 precut with Eco31I and Bpu1102I. The same strategy was used for construction of pN882A and pN891A plasmids, encoding N882A and N891A mutants, respectively, except that different direct primers, 5'-GATAATTCTATTGACGCTAAAGTACTTGTTTC-3' for pN882A and

5'-GTTTCATCTGCTAGTGCCCGTGGTAAATCAG-3' for pN891A, were used.

The mutants D31A and N891A in pASKIBA3-Cas9 plasmid were obtained by modified QuickChange Mutagenesis Protocol (Zheng *et al*, 2004)as previously described (Tamulaitis *et al*, 2007). Sequencing of the entire gene for each mutant confirmed that the sole difference consisted of the designed mutation.

2.2.3. Preparation of RNA substrates

RNA fragments used for analysis were generated by in vitro transcription (TranscriptAid T7 High Yield Transcription Kit, Thermo Fisher Scientific) from PCR fragments with inserted T7 promoter at the front end of RNA coding sequence. PCR fragments coding (+) and (-) RNA strands were obtained from pSP1 plasmid with following primer pairs accordingly: 5'-TAATACGACTCACTATA**G**GGTACCGAGCTCGAATTG-3'/5'-GGGAAACAGCTATGACCATGATTACGAATTC-3' and 5'-GGGTACCGAGCTCGAATTGAAATTCTAAACG-3'/5'-TAATACGACTCACTATA**G**GGAAACAGCTATGACCATGATTACG-3' (T7 RNA polymerase promoter underlined, transcription start on bold).

2.2.4. Plasmid transformation

Transformations were performed using the $CaCl_2$ heat-shock procedure (Sambrook *et al*, 1989). *E. coli* RR1 strain, carrying pCRISPR3 plasmid or its derivatives, was used as a recipient strain. Cells were grown in LB medium at 37ºC until OD_{600}=0.5. One ml aliquot of bacterial culture was used for each transformation experiment which was performed at 0ºC. Cells were recovered by centrifugation, washed using 0.5 ml of "Na solution" (5 mM Tris-HCl (pH 8.0), 100 mM NaCl and 5 mM $MgCl_2$) and resuspended into 0.5 ml "Ca solution" (5 mM Tris-HCl (pH 8.0), 100 mM $CaCl_2$, 5 mM $MgCl_2$) and incubated for 20 min. After incubation, cells were centrifugated, and pellets were resuspended into 50 µl "Ca solution" at 0ºC. One ng of plasmid DNA was

added to competent *E. coli* cells, incubated for 20 min at 4°C followed by 2 min incubation at 42°C. Then 450 µl of LB medium were added to the transformation mix and incubated at 37ºC for 1 hr. Finally, transformants were plated on LB agar with appropriate antibiotics. All transformation experiments were repeated at least three times. Bars in the graphs are presented as mean values from ≥3 independent experiments ±1SD. Part of experiments were performed by dr. R. Sapranauskas.

2.2.5. Phage plaque assay

Phage sensitivity of *E. coli* RR1 strain bearing pCRISPR3 or pCRISPR3-anti-λ-SP plasmids was analyzed using a virulent lambda phage (λ_{vir}). *E. coli* cells were grown in LB media overnight and subjected to the plaque assay following a standard procedure (Sambrook *et al*, 1989). In brief, 200 µl of overnight culture was mixed with the pre-cooled to 55°C top-layer agarose (0.4% agarose, 0.5% NaCl, 0.5% $MgCl_2$) in the presence of LB and chloramphenicol (10 µg/ml). Ten fold serial dilutions of λ_{vir} phage were prepared in 10 mM $MgSO_4$ solution. The host sensitivity to the phage infection was calculated as the efficiency of plaguing (Barrangou *et al*, 2007) ,which is the plaque count ratio of a strain containing an anti-λ_{vir} spacer in the CRISPR region to that of a strain containing a CRISPR with native spacers. Plaque assay experiments were repeated at least three times. Bars in the graph are presented as mean values from ≥3 independent experiments ±1SD.

2.2.6. Expression and purification of Cas9 protein and Cas9-crRNA complex

A $(His)_6$-tagged version of Cas9 was expressed and purified using a scheme described for the Cas3 protein of the *S. thermophilus* CRISPR4/Cas system (Sinkunas *et al*, 2011).

For purification of the Cas9-crRNA complex, Strep-tagged version of the Cas9 protein was expressed in *E. coli* RR1, bearing pCas9(-)SP1 plasmid (Figure 27B). LB broth was supplemented with ampicillin (Ap) (100 µg/ml)

and chloramphenicol (Cm) (10 μg/ml). *E. coli* cells for the Cas9-crRNA complex isolation were grown in two steps. First, 4 ml of cell culture were incubated at 37°C to an OD_{600} of ~0.5, and expression was induced by adding 0.2 μg/ml of anhydrotetracycline (AHT) (Sigma). After 4 h, $1/400^{th}$ of the pre-induced culture was inoculated into fresh LB medium supplemented with Ap (100 μg/ml), Cm (12 μg/ml) and AHT (0.2 μg/ml) and grown at 37°C during 17 hours. Harvested cells were disrupted by sonication and cell debris removed by centrifugation. The supernatant was loaded onto the 1 ml StrepTrap HP column (GE Healthcare) and eluted with 2.5 mM of desthiobiotin. Approximately 1.5 μg of the Cas9 protein was obtained in a single run from 1 L of *E. coli* culture. The fractions containing Cas9 were stored at + 4°C. The homogeneity of protein preparations was estimated by SDS-PAGE (Figure 19). Protein concentrations in the Cas9-crRNA complexes were determined by densitometric analysis of SDS-PAGE gels containing samples of Strep-Tactin purified Cas9 proteins along with known amounts of His-tagged Cas9 protein.

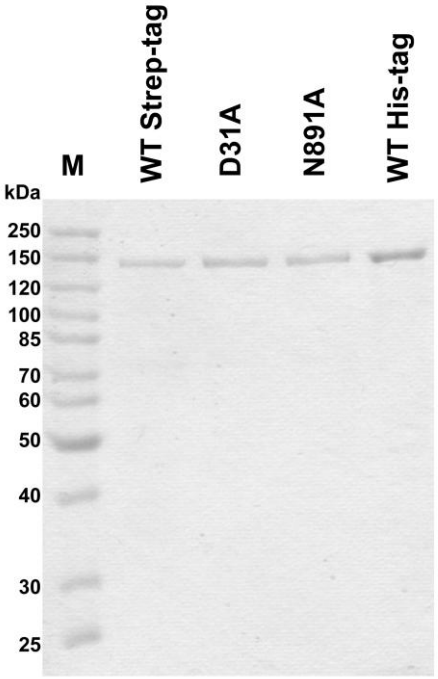

Figure 19. ***SDS-PAGE of the purified Cas9 proteins and mutant variants.*** M – protein molecular marker "PageRuler Unstained Protein Ladder" (Thermo Fisher Scientific). The numbers shown on the left indicate markers' size in kDa. 50 ng of His-tagged and 30 ng of Strep-tag Cas9 proteins were loaded on the gel.

2.2.7. Northern blot analysis

Cas9-bound RNA was isolated from Strep-Tactin purified Cas9, co-expressed with pCas9(-)SP1 plasmid using the miRNeasy Mini kit (Qiagen). Northern blots were performed by running RNA on a 10 % polyacrylamide gel with 7 M urea in 20 mM MOPS/NaOH (pH=8) buffer. The RNA was transferred to a SensiBlot Plus Nylon Membrane (Thermo Fisher Scientific) by semi-dry blotting using a Trans-blot SD (Bio-Rad). RNA was cross-linked to the membrane with solution of 0.16 M l-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) (Pierce) and 0.13 M 1-methylimidazole (Sigma) (pH=8) at 60°C for 1 h. The membrane was pre-hybridized with 2× SSC buffer containing 1% SDS and 0.1 mg/ml denatured DNA from fish testes (Ambion) for 1 h at 40°C. Blots were probed for 12 h with a ^{32}P-5′-labeled 42-nt anti-crRNA DNA oligonucleotide containing 20 nt of spacer1 and 22 nt of the repeat sequence (5′-TCGAAACAACACAGCTCTAAAACTGTCCTCTTCCT CTTTAGC-3′). The blots were washed 3× for 15 min with 0.2× SSC buffer containing 0.2% SDS, and were visualized using phosphorimaging. A 42-nt synthetic oligoribonucleotide (5′-CGCUAAAGAGGAAGAGGACAGUUUU AGAGCUGUGUUGUUUCG-3′) and 84-nt DNA oligonucleotide (5′-AAC**AAATTCTAAACGCTAAAGAGGAAGAGG**ACAGTTTTAGAG CTGTGTTGTTTCGAATGGTTCCAAAACTGAAGAGCTATTAG-3′, repeat sequence underlined, spacer1 sequence in bold) was used as size markers.

2.2.8. Reactions with oligonucleotide substrates

Reactions were typically carried out by adding 2 nM of Cas9-crRNA complex to 1 nM labeled oligonucleotide in 10 mM Tris-HCl (pH 7.5 at 37°C), 10 mM NaCl, 0.1 mg/ml BSA and 10 mM $MgCl_2$ at 37°C. Aliquots were removed at timed intervals and quenched with loading dye (95 % v/v formamide, 0.01 % bromphenol blue, 25 mM EDTA, pH 9.0) and subjected to denaturing gel electrophoresis (2.2.1.3).

2.2.9. Reactions with plasmid substrates

Reactions on pUC18 plasmid and its derivatives (Table 4) were conducted at 37ºC in the buffer used for reactions on oligonucleotide substrates. Reaction mixtures typically contained 2.5 nM supercoiled plasmid and 2 nM of Cas9-crRNA complex. The reactions were initiated by adding protein to the mixture of the other components. Aliquots were removed at timed intervals and quenched with phenol/chloroform. The aqueous phase was mixed with loading dye solution (0.01 % bromphenol blue and 75 mM EDTA in 50 % v/v glycerol) and analyzed by agarose electrophoresis.

2.2.10. Plasmid cleavage position determination

To achieve complete plasmid cleavage, 8 nM of Cas9-crRNA complex was incubated with 2.5 nM of supercoiled pSP1 plasmid in the reaction buffer at 37°C for 10 min. Reaction products were purified and concentrated using the GeneJET PCR Purification Kit. Spacer1 regions in Cas9 linearized and nicked pSP1 plasmids were directly sequenced with the following primers: 5′-CCGCATCAGGCGCCATTCGCC-3′ (sequencing of (+)strand) and 5′-GCGAGGAAGCGGAAGAGCGCCC-3′ (sequencing of (-)strand), for the (+) and (-) strands, respectively.

2.2.11. Binding assay

Increasing amounts of Cas9-crRNA complex were mixed with 0.5 nM of ^{33}P-labeled double-stranded and single-stranded DNA substrates (Table 3) in the binding buffer (40 mM Tris–acetate, pH 8.3 at 25°C, 0.1 EDTA, 0.1 mg/ml BSA, 10% v/v glycerol) and incubated for 15 min at 22°C. Free DNA and protein–DNA complexes were separated on the non-denaturing 8% polyacrylamide gel (2.2.1.2).

2.2.12. Sequence analysis

The theoretical pI, molecular mass and extinction coefficient of Cas9 were determined with the ProtParam tool (http://www.expasy.org/tools/protparam.html) (Gasteiger *et al*, 2005).

To identify amino acids belonging to RuvC and HNH active sites protein alignments were assembled using "COBALT"(Papadopoulos & Agarwala, 2007) tool and "SMART 6" database (Letunic *et al*, 2009), respectively. Alignments were visualized with "BioEdit 7.0.5" (Hall, 1999) program. Multiple alignments of the protein or DNA sequences were assembled using ClustalX program (Larkin *et al*, 2007).

2.2.13. Gel filtration

Gel filtration experiments were carried out at room temperature using Superdex 200 10/300 GL column (GE healthcare) pre-equilibrated with 10 mM sodium phosphate (pH 7.4) buffer containing 500 mM NaCl .The apparent M_w of His-tagged Cas9 were calculated by interpolation from the standard curve obtained using a set of proteins of known M_w (Bio-Rad Gel Filtration Standards).

3. RESULTS

3.1. CRISPR/Cas systems of *S. thermophilus*

A few model systems have been established in the study of CRISPR/Cas functionality, notably in *Escherichia coli* (Brouns *et al*, 2008; Pul *et al*, 2010), *Staphylococcus aureus* (Marraffini & Sontheimer, 2008), *Pyrococcus furiosus* (Hale *et al*, 2009) and *Streptococcus thermophilus* (Barrangou *et al*, 2007; Garneau *et al*, 2010). The *S. thermophilus* DGCC7710 model organism, for which CRISPR/Cas interference has been demonstrated against phages (Barrangou *et al*, 2007; Deveau *et al*, 2008) and plasmids (Garneau et al., 2010) contains four distinct CRISPR/Cas systems: CRISPR1, CRISPR2, CRISPR3 and CRISPR4 (Figure 20). Direct spacer acquisition activity has been demonstrated for the CRISPR1 and CRISPR3 systems, with the former being more active in this strain (Barrangou *et al*, 2007; Garneau *et al*, 2010; Deveau *et al*, 2008). CRISPR1 and CRISPR3, which both belong to Type II CRISPR/Cas systems, share a similar architecture, with 4 *cas* genes located upstream of the CRISPR spacer array. Both *cas1* and *cas2* are universal, whereas *cas9* (formerly named *cas5* and *csn1*) is the signature gene of the Type II system. It has been shown that *cas9* and *csn2* are involved in interference and spacer acquisition, respectively (Barrangou *et al*, 2007; Garneau *et al*, 2010). *In silico* analysis of phage sequences adjacent to CRISPR1 and CRISPR3 proto-spacers (nucleotide sequences in the target DNA corresponding to the spacers) revealed the presence of conserved PAM (Proto-spacer Adjacent Motif) sequences, NNAGAAW and NGGNG respectively (Deveau *et al*, 2008; Horvath *et al*, 2008; Mojica *et al*, 2009), that are involved in interference. Single point mutations in the proto-spacer or the PAM allow the phages to circumvent CRISPR-mediated immunity (Deveau et al., 2008).

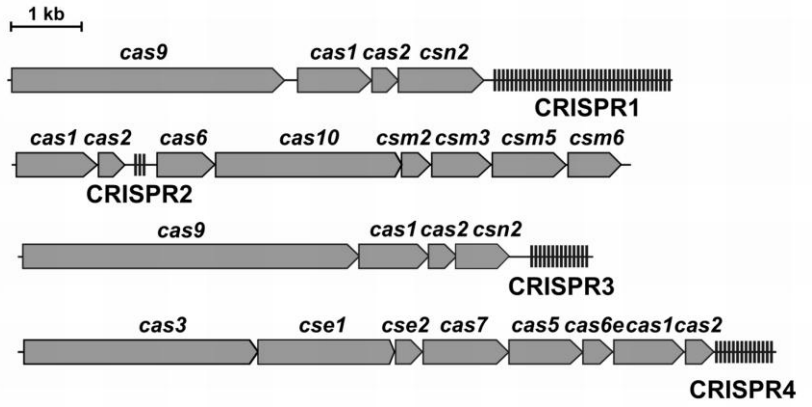

Figure 20. **CRISPR/Cas systems of *S. thermophilus* DGCC7710**. Cas proteins of the CRISPR1 and CRISPR3 systems belong to Type II whilst CRISPR2 and CRISPR4 belong to Type III and Type I, respectively

The CRISPR2 and CRISPR4 systems present in the *S. thermophilus* DGCC7710 genome belong to the Type III-A and Type I-E, respectively (Horvath & Barrangou, 2010; Makarova *et al*, 2011b). Differences between types can be observed in terms of repeat, spacer and *cas* gene content and sequence. The multiplicity of CRISPR/Cas systems in *S. thermophilus* is explained by their susceptibility to horizontal gene transfer, and phage selective pressure.

3.2. Characterization CRISPR3/Cas system in *E. coli* cells

3.2.1. Cloning of the CRISPR3/Cas system in *E. coli*

The *S. thermophilus* DGCC7710 CRISPR3/Cas (SthCRISPR3) locus (GenBank HQ712120) spans ~ 7.6 kb and consists of four *cas* genes: *cas9*, *cas1*, *cas2* and *csn2*, followed by an A+T rich 382-bp leader sequence and an array of 13 repeat-spacer units (Figure 21). In order to clone the CRISPR3/Cas locus in *E. coli*, we used 3 different primer pairs to generate three DNA fragments covering the ~7.6 kb CRISPR3/Cas locus ("Materials and methods", Figure 17). These fragments were cloned separately into the pACYC184

plasmid, and reassembled into a full-length CRISPR3/Cas locus using two subcloning steps.

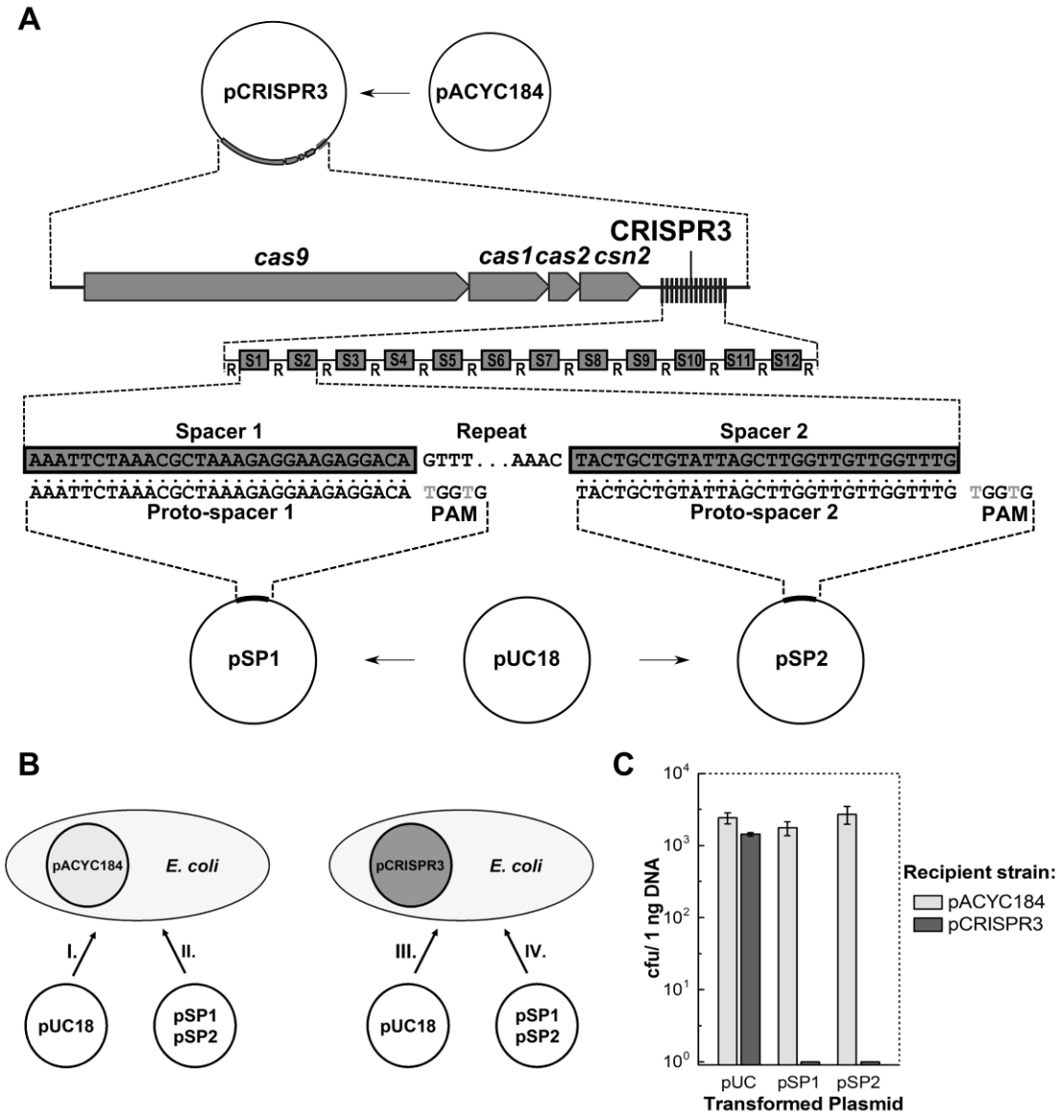

Figure 21. SthCRISPR3/Cas system provides immunity against plasmid transformation in *E. coli* cells. (A) Schematic representation of SthCRISPR3/Cas system cloning and construction of the plasmids for interference assay. SthCRISPR3/Cas system was cloned into *E. coli* plasmid pACYC184. Plasmids for interference assays were obtained by inserting a proto-spacer and PAM into pUC18 plasmid. (**B)** Schematic representation of the plasmid transformation interference assay. *E. coli* RR1 recipient strains carrying plasmids pCRISPR3 and pACYC184 with and without the SthCRISPR3/Cas system, respectively, were transformed with plasmids pSP1 and pSP2 carrying proto-spacers and PAMs or pUC18. **C.** Interference of plasmid transformation by SthCRISPR3/Cas system in *E. coli* cells. Transformation efficiency is expressed as cfu per ng of plasmid DNA (mean ± SD).

In the final pCRISPR3 plasmid (Figure 21), the ~ 7.6 kb fragment containing the natural *S. thermophilus* promoter and other possible regulatory elements located upstream of Cas9, is inserted 487 bp downstream of the constitutive *tet* gene promoter (P_{tet}). It remains to be established whether the components of the recombinant interference system are expressed from P_{tet} or from the native CRISPR/Cas promoters.

3.2.2. The CRISPR3/Cas system prevents plasmid transformation in *E. coli*

To test the functional activity of the CRISPR3 system in the heterologous *E. coli* host, we used a plasmid DNA transformation assay (see 2.2.4.). Using pUC18, which is compatible with pACYC184 in *E. coli*, we engineered pSP1 and pSP2 (Figure 21A), which contained proto-spacer sequences identical to spacer1 and spacer2 in the CRISPR3 array, together with the corresponding proto-spacer adjacent motif (PAM) 5'-TGGTG-3' downstream of the proto-spacer sequence (Table 4), and tested the plasmid transformation efficiency in recipient *E. coli* cells carrying either pCRISPR3 or pACYC184 (Figure 21B). Typically, 1 ng of plasmid DNA per ml of *E. coli* bacterial culture in LB medium (OD_{600} = 0.4) was used for transformation. When the recipient strain carrying pACYC184 plasmid was transformed using pSP1, pSP2, or pUC18, the number of transformants exceeded 10^3 colony forming units (cfu) (Figure 21C). However, when the recipient strain carrying the pCRISPR3 plasmid was transformed with the same set of plasmids, no colony was obtained for the pSP1 and pSP2 plasmids (Figure 21C), but the control plasmid (pUC18), which lacks a proto-spacer, yielded 10^3 cfu. This is consistent with the presence of CRISPR-encoded specific immunity against spacers SP1 and SP2 in the pCRISPR3 plasmid. Interestingly, pUC18 contains 95 additional 5'-GGTG-3' sequences corresponding to the PAM of the CRISPR3 system but no corresponding proto-spacer sequences. This is consistent with the necessity to have both a proto-spacer and an adjoining PAM in order for CRISPR-encoded immunity to occur. Altogether these results indicate that the heterologous

plasmid pCRISPR3 interferes with transformation of pSP1 and pSP2 plasmids, both bearing corresponding proto-spacer and PAMs, but not with the control pUC18 plasmid. When a recipient strain carrying plasmid pACYC184 is used, no interference with plasmid transformation is observed.

To estimate the efficiency of plasmid transformation, we performed similar experiments using 100 ng of plasmids. Again, in the recipient strain carrying pCRISPR3, no colony was obtained but more than 10^5 cfu of transformants were obtained with the control vector (pUC18). The same number of transformants was obtained when the recipient strain containing pACYC184 was transformed with 100 ng of pSP1. This means that SthCRISPR3/Cas system in *E. coli* reduces the efficiency of plasmid transformation with more than five orders of magnitude. This is quantitatively comparable to the level of phage interference previously established in *S. thermophilus*, where CRISPR-based immunity reduces phage-sensitivity and efficiency of plaquing by 4-5 orders of magnitude per spacer (Barrangou *et al*, 2007; Deveau *et al*, 2008). These results indicate that CRISPR/Cas systems can be transferred between distant bacterial genera and used heterologously to provide immunity against plasmids.

3.2.3. The CRISPR3/Cas system provides resistance against phage lambda in *E. coli*

In order to investigate the interference scope of the heterologous *S. thermophilus* CRISPR3/Cas system in *E. coli*, we assessed whether it could provide resistance against phages in addition to interference against plasmids. We inserted *in vitro* a spacer (anti-λ_{vir}) targeting a lambda phage sequence adjacent to the corresponding PAM (Horvath *et al*, 2008) into the heterologous CRISPR3 locus on the pCRISPR3 plasmid (Figure 18) and subsequently tested whether the engineered CRISPR system could confer resistance against phage lambda. Results showed that *E. coli* cells containing the SthCRISPR3/Cas system with the anti-λ_{vir} spacer were less sensitive to phage lambda by approximately three orders of magnitude, in comparison to the control cells

that carried a wild-type CRISPR3 locus without the anti-λ_{vir} spacer (Figure 22). These preliminary results indicate that the heterologous CRISPR3/Cas system can also provide immunity against phages in *E. coli*, and sets the stage for further studies that will assess the interference potential of heterologous CRISPR/Cas systems.

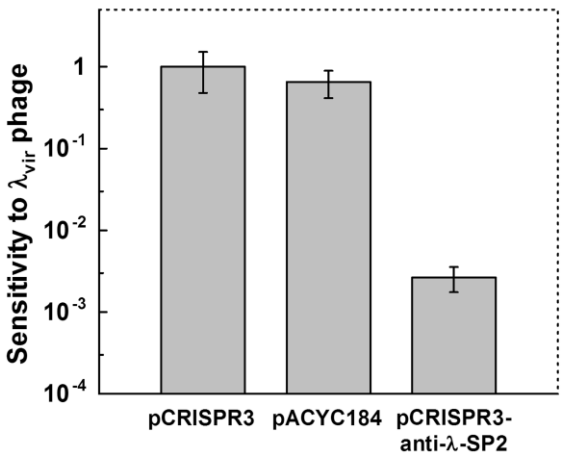

Figure 22. Sensitivity to lambda phage (λ_{vir}) of *E. coli* cells. *E. coli* cells bearing different recombinant plasmids were used as hosts in the phage challenge assay. pCRISPR3 plasmid contains full length CRISPR region with native spacers; pACYC184 plasmid lacks SthCRISPR3/Cas system; pCRISPR3-anti-λ-SP plasmid contains a full set of *cas* genes and engineered anti-λ_{vir} spacer. Sensitivity to λ_{vir} phage is the plaque count ratio of an analyzed strain (e.g. containing anti- λ_{vir} spacer) to that of a strain containing pCRISPR3 plasmid (no anti- λ_{vir} spacer).

3.2.4. Position-dependent tolerance for mismatches in CRISPR-based plasmid interference

Phage challenge assays in *S. thermophilus* DGCC7710 have revealed that to ensure phage interference, 100 % identity between the spacer sequence from the CRISPR array and the corresponding proto-spacer sequence in the phage DNA is required (Barrangou *et al*, 2007; Deveau *et al*, 2008). To assess whether sequence identity between a spacer and the target proto-spacer sequence is important for the efficiency of plasmid transformation, we engineered plasmids pSP1-sA2G, pSP1-sC11T, pSP1-sG18C, pSP1-sA23T, pSP1-sA25T, and pSP1-sA28T (Table 4) which carried single mutations at distinct positions across the proto-spacer sequence (A2G, C11T, G18C, A23T, A25T, and A28T, respectively), and analyzed transformation efficiency of the recipient strain containing pCRISPR3 (Figure 23A). Consistent with the phage challenge assay, single point mutations A25T and A28T in the proto-spacer abolished recipient strain ability to prevent plasmid transformation. Unexpectedly, single mismatches at the proto-spacer positions 2, 11, 18, 23

had no effect on the recipient strain ability to interfere with plasmid transformation (Figure 23A). Taken together our data suggest that CRISPR3/Cas system tolerates single nucleotide mismatches between spacer and proto-spacer at certain positions. In light of recent results indicating that cleavage of invading plasmid and phage DNA occurs 3 bp upstream of the PAM sequence in CRISPR1, and given the relatedness of CRISPR1 and CRISPR3 systems, it is likely that the mutations at positions 25 and 28 impact cleavage, whereas distal mutations would not.

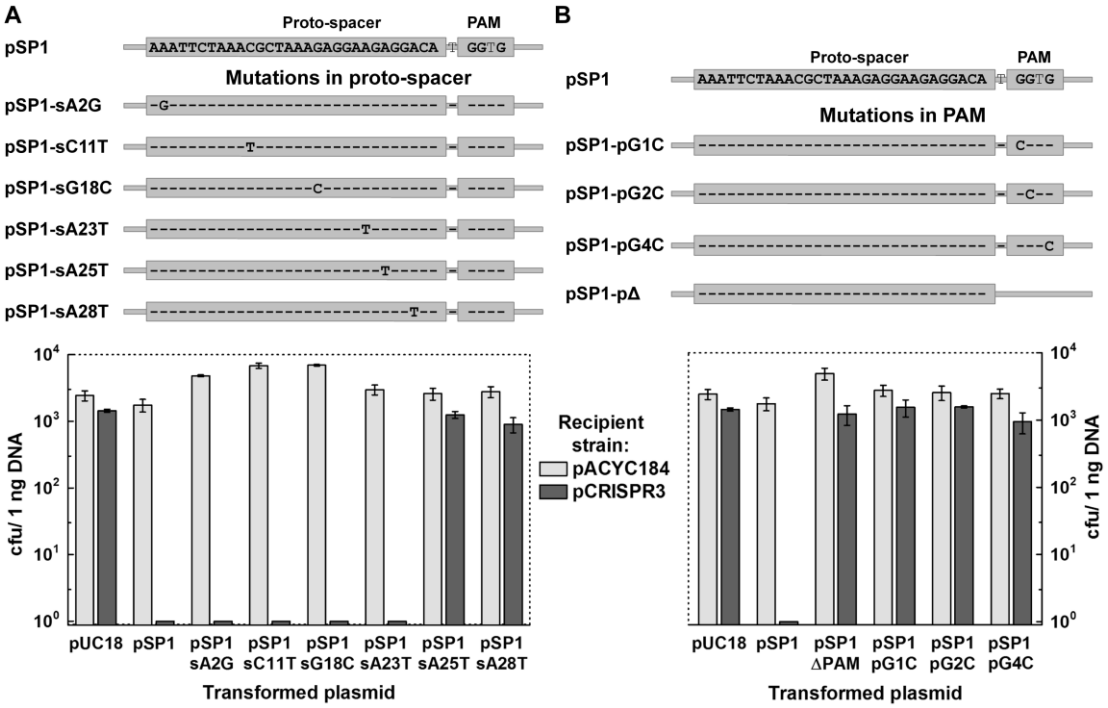

Figure 23. Impact of proto-spacer and PAM mutations on CRISPR-encoded plasmid immunity. (A) Effect of mutations in the proto-spacer region on the plasmid transformation efficiency. Mutations are shown schematically above the figure. (**B**) Effect of mutations in the PAM region on the plasmid transformation efficiency.

3.2.5. The PAM sequence is important for plasmid interference

PAM sequences are of crucial importance for the phage interference by the CRISPR1 system of *S. thermophilus* (Deveau *et al*, 2008; Horvath *et al*, 2008). To test whether the PAM sequence is important in prevention of plasmid DNA transformation by the CRISPR3/Cas system, we engineered pSP1-pΔ plasmid so as to carry proto-spacer1 without its adjacent PAM sequence.

Transformation efficiency of *E. coli* recipient cells with or without CRISPR3 system by the pSP1-pΔ plasmid was similar (Figure 23B). This indicates that the PAM sequence is required for CRISPR3 interference of plasmid transformation, and that the sole presence of a proto-spacer is not sufficient. To test whether all three conserved residues of the predicted PAM sequence 5'-NGGNG-3' are equally important, we constructed plasmid variants pSP1-pG1C, pSP1-pG2C and pSP1-pG4C (Table 4) were three guanine residues in the PAM sequence 5'-tGGtG-3' were replaced by cytosine residues. Transformation experiments revealed that *E. coli* cells containing pCRISPR3 or pACYCY184 plasmids were efficiently transformed (10^3 cfu) by the plasmids containing single mutations in the PAM region (Figure 23B). These data clearly demonstrate that while CRISPR3/Cas system tolerates single mutations at certain positions in the proto-spacer region, a conserved PAM sequence is required for the plasmid DNA interference in *E. coli*.

3.2.6. Only Cas9 is required for plasmid interference in *E. coli*

It was previously shown that inactivation of *cas9* from the CRISPR1 system in *S. thermophilus* resulted in the loss of phage resistance, while inactivation of *csn2 (cas7)* did not alter phage resistance but impaired the ability to incorporate new spacers (Barrangou *et al*, 2007). To determine the role of individual Cas proteins in the interference step, we engineered frameshift mutants of individual *cas* genes, generating plasmids pCas9(-), pCas1(-), pCas2(-), pCsn2(-) (Figure 24A) and monitored the efficiency of the pSP1 plasmid transformation into corresponding recombinant *E. coli* recipient cells. Mutants lacking *cas1*, *cas2* or *csn2* retained the ability to interfere with plasmid transformation (Figure 24B). However, *E. coli* cells carrying pCas9(-) plasmid with mutated *cas9*, were efficiently transformed both by pSP1 and pUC18 plasmids. This is consistent with previous results indicating that *cas9* and *csn2* are involved in interference and spacer acquisition, respectively (Barrangou *et al*, 2007).

To further confirm that the Cas9 alone can provide interference with plasmid transformation, we deleted the three other *cas* genes (*cas1*, *cas2*, *csn2*) in pCRISPR3 to generate a pCas9 plasmid which contains only *cas9*, the leader sequence, and a repeat-spacer region (Figure 24A). The recipient strain containing pCas9 plasmid retained the ability to interfere with pSP1 plasmid transformation (Figure 24B), indicating that Cas9 is the only Cas protein required to provide resistance against foreign DNA. This suggests that Cas1, Cas2 and Csn2 might rather be involved in novel spacer acquisition.

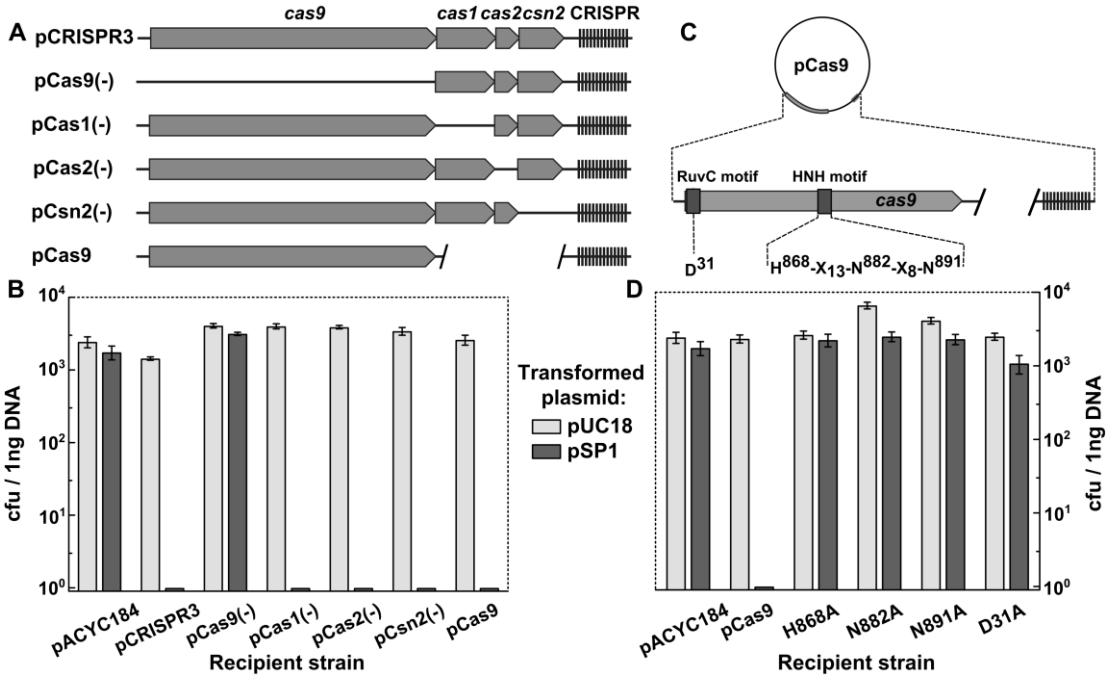

Figure 24. Mutational analysis of *cas* genes. (A) Schematic representation of plasmids carrying mutant variants of *cas* genes. Individual genes were disrupted by frameshift mutations or small deletions. Three *cas* genes (c*as1, cas2, csn2)* were removed by deletion. (**B)** *cas9* gene alone prevents plasmid DNA transformation. (**C)** Mutagenesis of Cas9 protein. Conserved domains and mutated amino acids are indicated. (**D)** Mutations in the conserved RuvC and HNH domains inactivate *cas9*.

3.2.7. Mutational analysis of *cas9*

Cas9 proteins are predicted to contain RuvC and HNH signature motifs (Makarova *et al*, 2006). The HNH motif is characteristic of many nucleases that act on double stranded DNA including colicins (Kleanthous *et al*, 1999; Ko *et al*, 1999), restriction enzymes (Saravanan *et al*, 2004) and homing

endonucleases (Shen *et al*, 2004). RuvC fold includes proteins that show wide spectra of nucleolytic functions, acting both on RNA and DNA (RNase H, RuvC, DNA transposases and retroviral integrases, and PIWI domain of Argonaut proteins) (Nowotny *et al*, 2005). To test whether the conserved amino acid residues D31 (RuvC motif) (Figure 25), H865, N882, and N891 (HNH motif) (Figure 26) are important for Cas9 function, alanine replacement mutants were constructed in the pCas9 plasmid by site-directed mutagenesis to generate recombinant plasmids pD31A, pH868A, pN882A and pN891A, respectively. Plasmid transformation assays revealed that all three mutations in the HNH motif and a single mutation in RuvC motif abolish the Cas9-dependent plasmid interference (Figure 24D). Accordingly, the amino acid residues of the HNH motif and the conserved aspartate in the N-terminal part of protein play an essential role in the plasmid DNA interference by Cas9 *in vivo*.

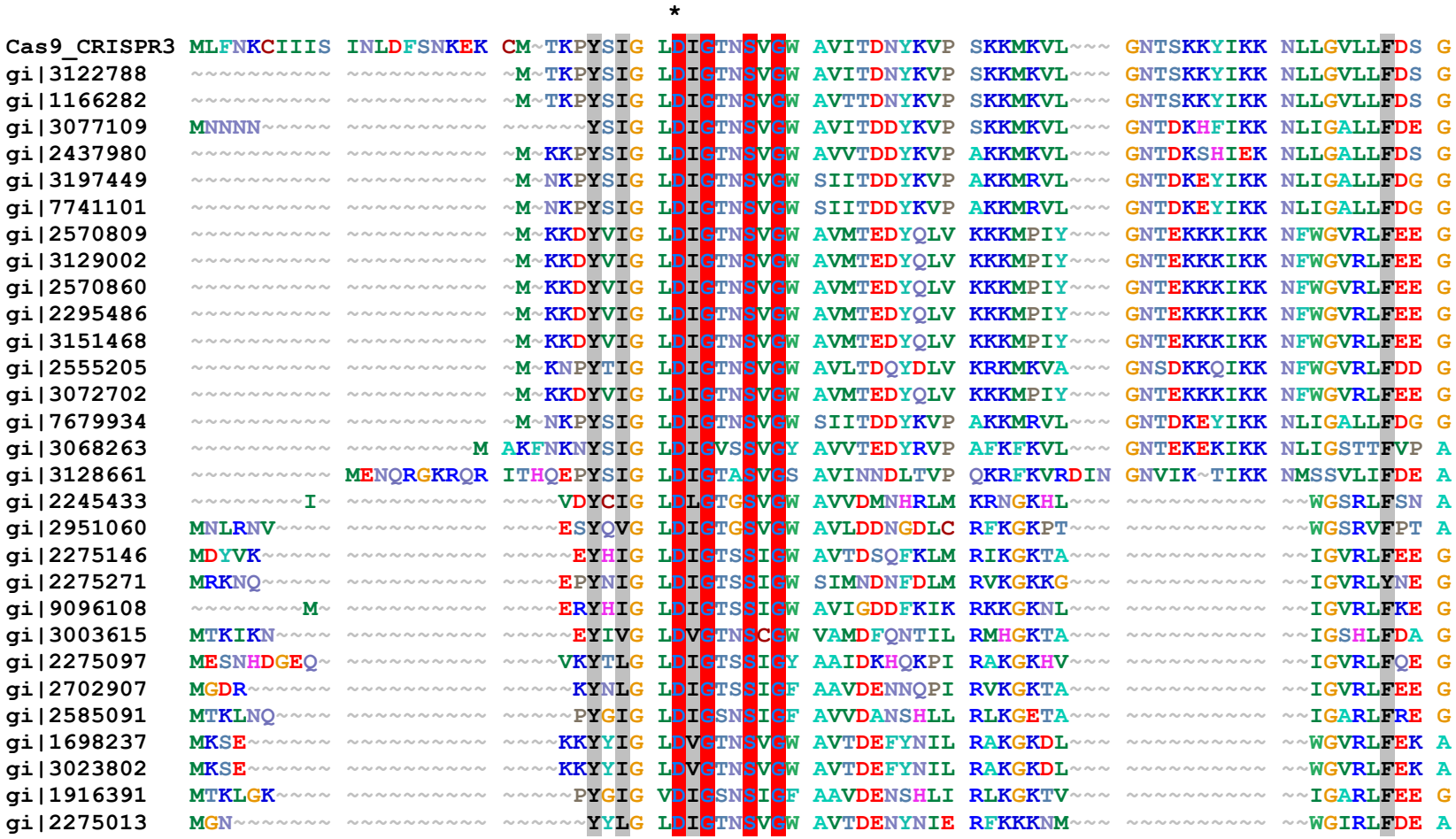

Figure 25. Alignment of N-terminal domain of Cas9 from SthCRISPR3/Cas system. Identical amino acids are shaded red, D31 amino acid residue subjected to mutagenesis is marked by an asterisk. Homologous sequences with different similarity to Cas9 protein from *S. thermophilus* CRISPR3 system were used for construction of alignment. GeneBank identification numbers of proteins are given in the left.

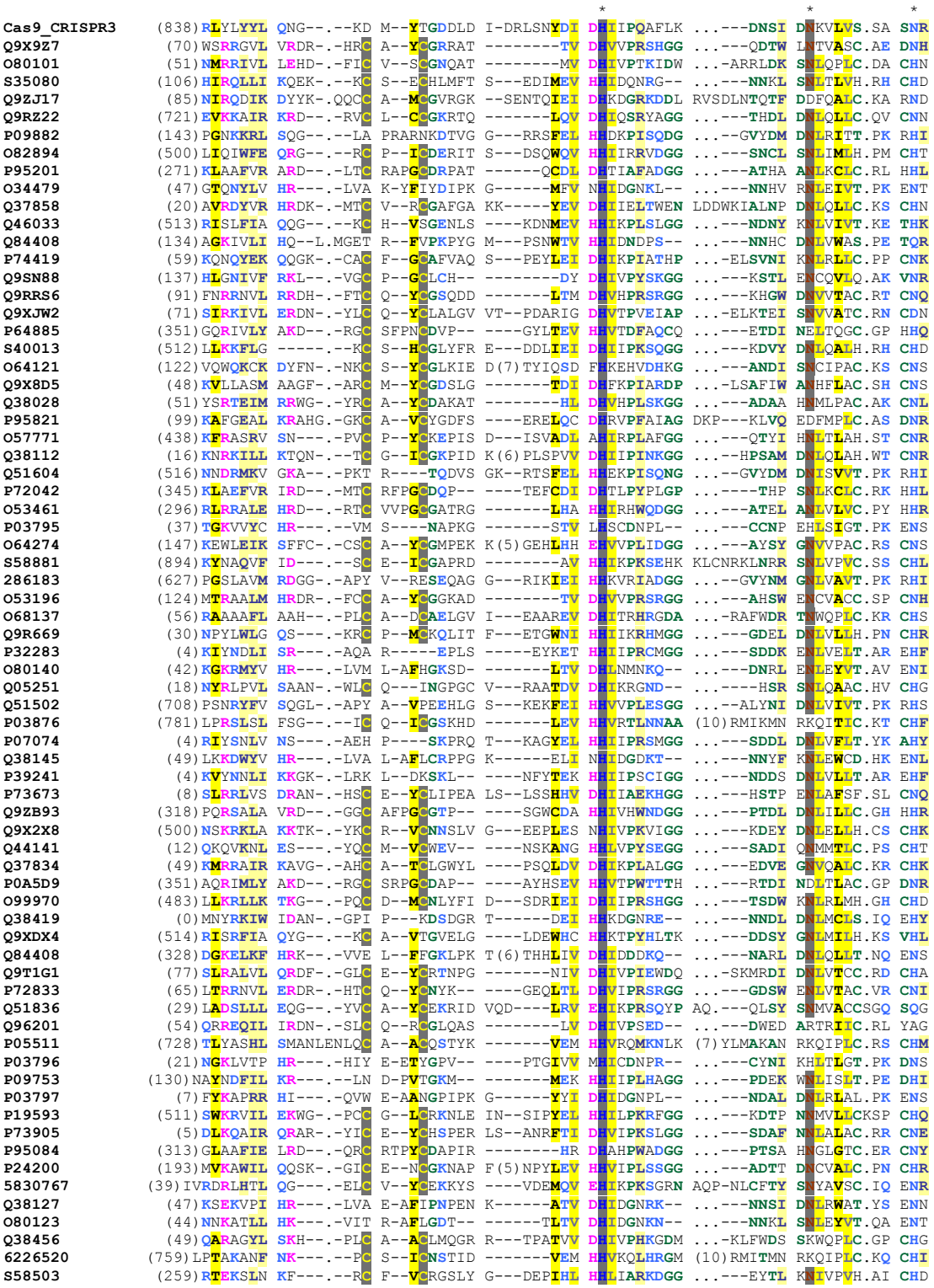

Figure 26. Alignment of HNH domain of Cas9 from SthCRISPR3. Amino acid residues subjected to mutagenesis are marked by an asterisk. Homologous sequences with different similarity to Cas9 protein from *Streptococcus thermophilus* DGCC7710 CRISPR3 system were used for construction of alignment. UniProt accession numbers of proteins are given in the left.

3.2.8. The role of Cas proteins in the defense mechanism

CRISPR/Cas loci are highly diverse and fall within different categories depending on the type, number of *cas* genes and architecture of the *cas* operon (Haft *et al*, 2005; Makarova *et al*, 2006, 2011b). The diversity of CRISPR/Cas systems implies mechanistic differences which remain to be established. Meanwhile, two different pathways which target invading genetic elements via the crRNAs and act on DNA (Brouns *et al*, 2008; Marraffini & Sontheimer, 2008) or RNA (Hale *et al*, 2009) targets are emerging. It was shown recently that the *S. thermophilus* CRISPR1/Cas system specifically cleaves plasmid and bacteriophage double-stranded DNA within the proto-spacer *in vivo*, at specific sites (Garneau *et al*, 2010). This endonuclease activity seems to require Cas9 but it remains to be established whether other Cas proteins of the CRISPR1/Cas system contribute to the cleavage. The CRISPR3/Cas system of *S. thermophilus* DGC7710 belongs to the same subtype II-A, and contains a similar set of *cas* genes (Figure 20), suggesting that both systems may be mechanistically similar. We show here that CRISPR3/Cas module cloned into *E. coli* is functionally active and provides host cell with interference against plasmid and phage.

The *S. thermophilus* CRISPR3/Cas Cas9 is a large protein comprised of 1,388 amino acid residues. *In silico* analysis identified a HNH-nuclease motif and a RuvC -like nuclease signature in the Cas9 protein sequence (Haft *et al*, 2005; Makarova *et al*, 2006). We provide experimental evidence that the alanine replacement of the conserved D31 (RuvC-motif), and H865, N882, and N891 (HNH-motif) residues abolishes Cas9-mediated plasmid interference in *E. coli*. Since HNH-domains are often found in nucleases that act on double stranded DNA, we suggest that the HNH-domain of Cas9 is involved in the DNA degradation step. In the CRISPR1/Cas system, cleavage of target DNA occurs at both DNA strands within the proto-spacer sequence (Garneau *et al*, 2010). If Cas9 of the CRISPR3/Cas system is a monomer in solution and contains a single HNH-motif, it has to dimerise or employ a second active site to cleave both DNA strands. It is possible that HNH- and RuvC -like catalytic

sites may act on different DNA strands to achieve a double strand break; a similar strategy is exploited by some restriction endonucleases (Armalyte *et al*, 2005). The RuvC-like nuclease domain is identified in proteins that act on both DNA and RNA, including RNAse H, RuvC, and PIWI domain of Argonaut proteins (Nowotny *et al*, 2005). Therefore, we cannot exclude that the N-terminal RuvC domain in Cas9 might be involved in crRNA maturation. Studies of Cas9 nuclease and ribonuclease functions both *in vivo* and *in vitro* are currently on-going.

Our studies of the *S. thermophilus* CRISPR3/Cas system demonstrate that different CRISPR/Cas systems follow different strategies to achieve cleavage of invading DNA. Indeed, in contrast to a single *S. thermophilus* Cas9 protein which is potentially involved in both crRNA maturation and DNA cleavage steps, in *E. coli* a large nucleic acid-protein complex which include crRNA, Cascade, and Cas3 is thought to be involved in the degradation of foreign DNA (Brouns *et al*, 2008).

3.2.9. "Vaccination" of *E. coli* against plasmids and phages by the heterologous CRISPR3/Cas system

There is strong evidence suggesting that CRISPR/Cas systems can move between distinct species in bacteria and archaea via horizontal gene transfer (Godde & Bickerton, 2006; Horvath *et al*, 2009; Portillo & Gonzalez, 2009). Firstly, CRISPR/Cas content of closely related species or strains might differ, while on the other hand, evolutionary distant species (even across Bacteria and Archaea) might harbor similar CRISPR systems (Haft *et al*, 2005; Godde & Bickerton, 2006; Makarova *et al*, 2002). For example, genome sequencing of the three closely related *S. thermophilus* strains revealed that CNRZ1066 and LMG 18311 possess two CRISPR/Cas systems (CRISPR1/Cas and CRISPR2/Cas) (Horvath *et al*, 2008), whereas the LMD-9 strain has an additional Type II system, CRISPR3/Cas (Horvath *et al*, 2008). Furthermore, *S. thermophilus* DGCC7710 contains a fourth CRISPR/Cas system (Figure 20) which belongs to Type I (Horvath & Barrangou, 2010). It is likely that in *S.*

thermophilus the CRISPR4/Cas system has been acquired recently by horizontal gene transfer, especially since this species is naturally competent for transformation (Fontaine *et al*, 2010). Secondly, there is often a codon bias and/or a marked GC-content difference between CRISPR/Cas loci and the rest of the chromosome (Horvath *et al*, 2009). In addition, mobile genetic elements such as insertion sequences are usually located in the vicinity of CRISPR/Cas systems, while some CRISPR/Cas systems are located on large plasmids (> 40 kbp) (Godde & Bickerton, 2006) and even prophages (Sebaihia *et al*, 2006).

Here we provide the first experimental evidence showing that the CRISPR3/Cas system of the Gram-positive *S. thermophilus* species can be cloned into a plasmid and transferred to a Gram-negative *E. coli* host. Furthermore we show that the heterologous system provides resistance against incoming plasmids and phages that carry matching proto-spacer sequences and PAMs. This finding illustrates that CRISPR/Cas systems may function as mobile gene cassettes that overcome barriers between distant species such as incompatibility of promoters and other regulation signals. This successful transfer of a functional CRISPR/Cas system into a phylogenetically distant host opens novel possibilities for practical applications, notably the transfer of active CRISPR/Cas systems between species in order to "vaccinate" bacteria against viruses or plasmids. In light of recent results indicating that CRISPR-encoded immunity can target antibiotic resistance genes (Garneau *et al*, 2010), there is great interest in transferring active CRISPR/Cas systems in order to strengthen the immunity of select species or strains against the uptake and dissemination of antibiotic resistance genes.

3.3. The effector complex of Type II systems

3.3.1. Isolation and analysis of the Cas9-crRNA complex

The CRISPR3 system of *S. thermophilus* DGCC7710 is comprised of four *cas* genes that are located upstream of 12 repeat-spacer units (Figure 27A). For isolation of the Cas9-crRNA complex from the heterologous *E. coli* host, we

first cloned the *cas9* gene into the pASK-IBA3 vector to generate a construct encoding a single Cas9 protein with a C-terminal Strep(II)-tag (Figure 27B).

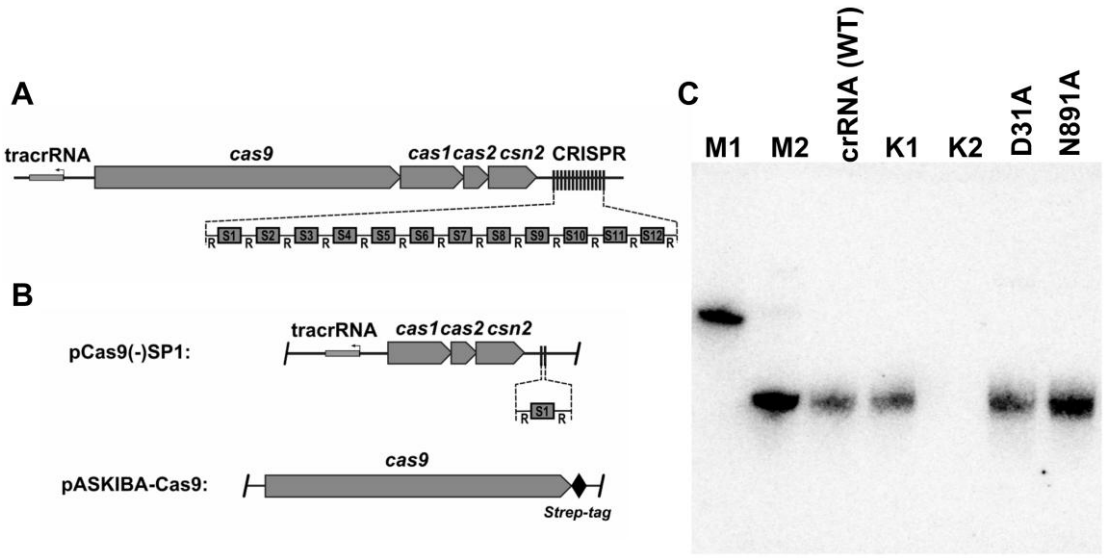

Figure 27. The Cas9 protein co-purifies with crRNA. (A) Schematic representation of the CRISPR3/Cas system of *S. thermophilus* DGCC7710. Four *cas* genes (*cas9, cas1, cas2, csn2*) are located upstream of the CRISPR repeat-spacer array, consisting of 13 repeat (R) sequences and 12 unique spacers (S1-S12). The tracrRNA, required for crRNA maturation in Type II CRISPR/Cas systems (Deltcheva *et al,* 2011), is located upstream the *cas9* gene and encoded on the opposite DNA strand (shown by an arrow) with respect to the other elements of this system. **(B)** Schematic representation of heterologous loci in two plasmids used for the co-expression of the Cas9-crRNA complex. *E. coli* RR1 contained pCas9(-)SP1 (encoding Cas1, Cas2, Csn2, SP1 and tracrRNA) and pASKIBA-Cas9 (encoding Strep-tagged version of Cas9) plasmids. **(C)** Northern analysis of Cas9-crRNA complexes using anti-crDNA oligonucleotide as a probe. M1 – 84-nt oligodeoxynucleotide corresponding to the spacer1-repeat unit; M2 – 42-nt synthetic oligoribonucleotide corresponding to the predicted SthCRISPR3 crRNA (Figure 29); crRNA (WT) – crRNA isolated from the WT Cas9-crRNA complex; K1 - crRNA (WT) treated with DNase I for 15 min; K2 - crRNA (WT) treated with RNase I for 15 min, D31A - crRNA purified from the Cas9 D31A mutant complex; N891A - crRNA purified from the Cas9 N891A mutant complex.

To obtain a homogeneous Cas9-crRNA complex, we next engineered a pCas9(-)SP1 plasmid with a single spacer (spacer1) inserted between two CRISPR repeats (Figure 27B). A plasmid interference assay confirmed that the minimized CRISPR array carrying only spacer1 provides interference against plasmid pSP1 transformation in *E. coli,* similar to that of the SthCRISPR3/Cas system carrying a complete 12 spacer array (Figure 28B). To achieve simultaneous transcription of all target genes, we obtained *cas9* gene

expression in two steps. First, we induced Cas9 expression in a small volume of *E. coli* culture and after 4 h transferred an aliquot of pre-induced culture into a larger volume of fresh LB medium already containing the inductor and incubated during 17 hours. The Cas9-crRNA complex was isolated using Strep-Tactin Sepharose column, and crRNA bound to Cas9 protein was analyzed.

The CRISPR3/Cas system of *S. thermophilus* belongs to Type IIA (formerly Nmeni) (Makarova *et al*, 2011b). In the homologous Type IIA CRISPR/Cas system of *S. pyogenes*, trans-encoded small RNA (tracrRNA) and bacterial RNase III are involved in the generation of the 42-nt crRNAs that carry the 22-nt "3′-handle" comprised of the repeat sequence and a 20-nt spacer fragment (Deltcheva *et al*, 2011). crRNAs of similar length are also generated in the *S. thermophilus* LMD-9 CRISPR3/Cas system (Deltcheva *et al*, 2011), which is nearly identical to the CRISPR3/Cas system of DGCC7710 (Figures 29A and B). We assumed that crRNAs present in the Cas9-crRNA complex isolated from the heterologous *E. coli* strain may have the same length (Figure 29C).

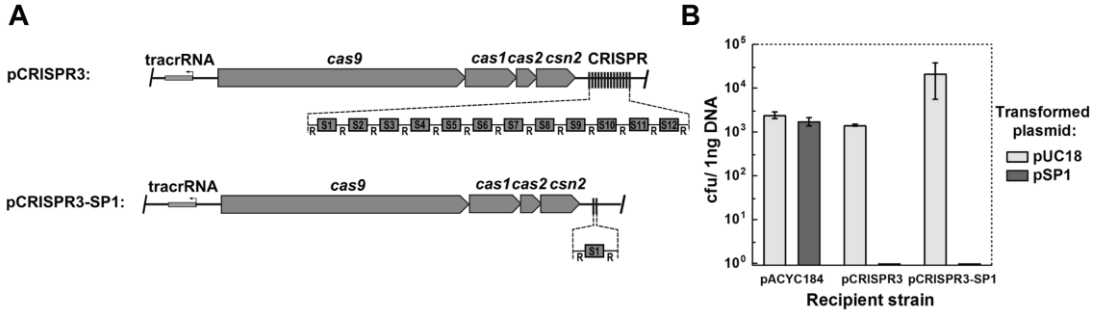

Figure 28. Immunity against plasmid transformation in *E. coli* cells provided by the SthCRISPR3/Cas system. (A) Schematic representation of CRISPR/Cas locus of recombinant plasmid pCRISPR3 carrying indigenous 12 spacer-repeat array of SthCRISPR3/Cas system and engineered pCRISPR3-SP1 plasmid carrying 1 spacer-repeat unit. **(B)** Interference of plasmid transformation by *Sth*CRISPR3/Cas system in *E. coli* cells. *Escherichia coli* RR1 recipient strains carrying plasmids pACYC184, pCRISPR3 or pCRISPR3-SP1, were transformed with plasmid pSP1 carrying proto-spacers and PAM or pUC18. Transformation efficiency is expressed as cfu per nanogram of plasmid DNA (mean ± SD).

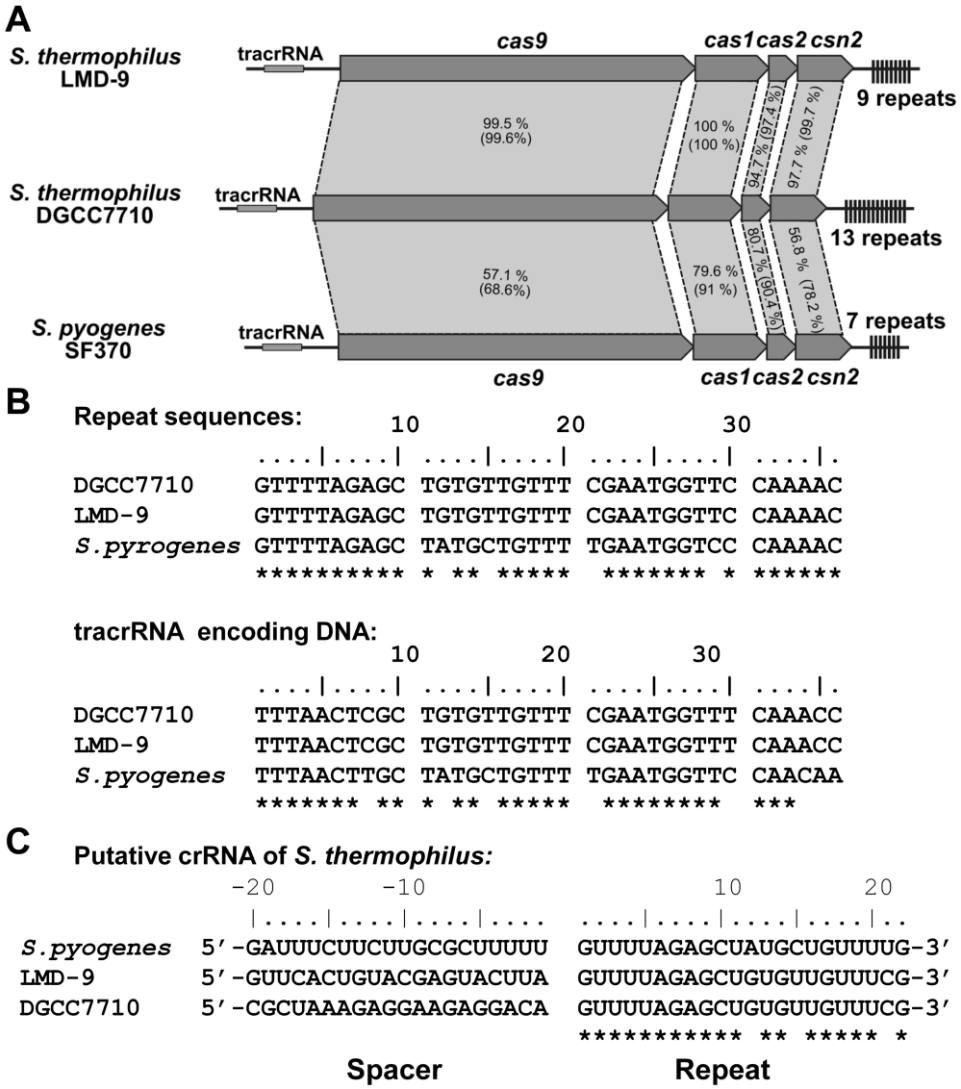

Figure 29. Comparison of Type IIA CRISPR/Cas systems from *S. thermophilus* DGCC7710, LMD-9 and *S. pyogenes* SF370 strains. (A) Schematic organization of the CRISPR/Cas systems. Nucleotide sequences corresponding to the tracrRNA required for the crRNA maturation in *S. pyogenes* (Deltcheva *et al*, 2011) are present in LMD-9 and DGCC7710. Percentage of identical and similar (in parenthesis) residues between corresponding protein sequences that are connected by dashed lines. **(B).** Alignment of the conserved repeat sequences and tracrRNA. Corresponding sequences from DGCC7710 and LMD-9 are identical. Nucleotide positions identical in all three strains are labeled with an asterisk below aligned sequences. **(C)** Comparison of crRNA sequences. The sequence and length of *S. pyogenes* crRNA was determined by deep sequencing analysis. The approximate length of crRNA from *S. thermophilus* LMD-9 (Deltcheva *et al*, 2011) and DGCC7710 (this work) strains were determined by the northern blot analysis.

Therefore, to probe nucleic acids extracted from the Strep-Tactin-purified Cas9 complex, we used a 42-nt anti-crRNA ssDNA oligonucleotide comprised of 22 nt corresponding to the 3′-end of the repeat sequence and 20 nt at the 5′-

end of SP1 fragment. Nucleic acids present in the Cas9-crRNA complex hybridized with anti-crRNA oligonucleotide, and were sensitive to RNase but not DNase treatment (Figure 27C). The size of extracted RNAs was identical to the 42-nt synthetic oligoribonucleotide corresponding to the putative crRNA of the CRISPR3 system of *S. thermophilus* DGCC7710 (Figure 29C and 30 A). Taken together, these data confirm that Cas9 Strep-tag protein co-purifies with the 42-nt crRNAs that are derived from the CRISPR3 array.

3.3.2. The Cas9-crRNA complex cleaves double-stranded DNA within the proto-spacer

To test the *in vitro* activity of the purified Cas9-crRNA complex, we first used the SP1 oligoduplex (Table 3) containing a proto-spacer sequence (proto-spacer1) identical to spacer1 in the CRISPR3 array, the PAM sequence 5′-TGGTG-3′ (conserved nucleotides are underlined) downstream of the proto-spacer, and 10-nt flanking sequences present in pSP1 plasmid (Figure 30A). The oligoduplex strand complementary to crRNA is named (+)strand, while the opposite duplex strand is called the (-)strand. To monitor the cleavage reaction, either (+) or (–)strand of the SP1 oligoduplex was P^{33}-labeled at the 5′-terminus. Analysis of the reaction products by PAGE (Figure 30B) revealed that the Cas9-crRNA complex cleaves both strands of the SP1 oligoduplex at a fixed position. Mapping of the cleavage position using synthetic oligonucleotides as size markers revealed that cleavage occurs within the proto-spacer, 3 nt away from the terminal end of the proto-spacer adjacent to the PAM, leaving blunt ends (Figure 30B). Control experiments have shown that crRNA-guide is absolutely necessary for the SP1 oligoduplex cleavage, since no cleavage was produced by Cas9 in the absence of crRNA (Figure 31A).

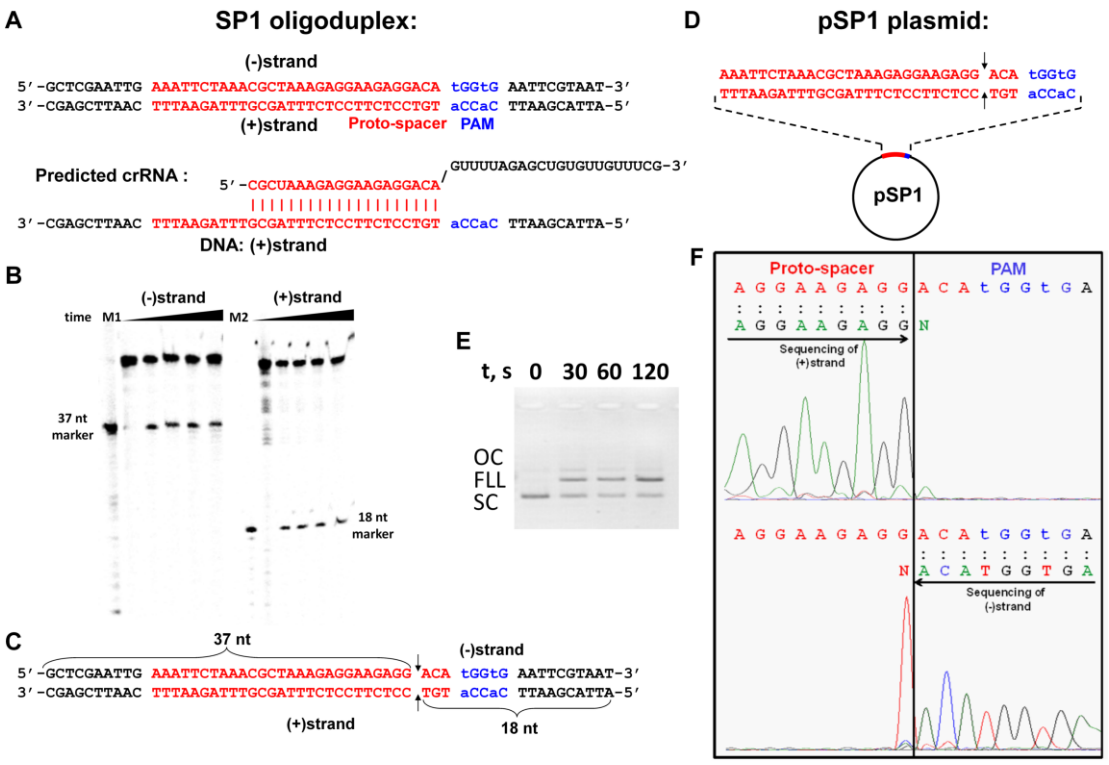

Figure 30. The Cas9-crRNA complex cleaves dsDNA within the proto-spacer. (A) Oligoduplex substrate used in the cleavage assay. 55-nt oligoduplex SP1 contains the proto-spacer1 (red letters), PAM (blue letters) and 10-nt flanking sequences on both sides identical to those in pSP1 plasmid. In the SP1 oligoduplex, the DNA strand complementary to the 5′-terminal fragment of crRNA (red letters) is named (+)strand, and opposite DNA strand is named (-)strand. **(B)** Oligoduplex SP1 cleavage. 2.5 nM of Cas9-crRNA complex and 1 nM SP1 oligoduplex labeled with ^{33}P at the 5′-end of either (+) or (–)strand were incubated in the reaction buffer (Materials and methods) at 37°C for varied time intervals (30 s to 10 min) and reaction products analyzed by 20 % PAA gel electrophoresis. Lanes M1 and M2 contain chemically synthesized 5′-end ^{33}P-labeled 37-nt and 18-nt oligonucleotides corresponding to the cleavage products of (-) and (+) DNA strands, respectively. **(C)**. SP1 cleavage positions are indicated by arrows. **(D)** Schematic representation of pSP1 plasmid used in the plasmid cleavage assay. **(E)** Agarose gel of pSP1 cleavage products. SC – super-coiled plasmid DNA, OC – open circular DNA nicked at one of the strands, FLL – full length linear DNA cut at both strands. Final reaction mixtures at 37°C contained 2.5 nM of pSP1 plasmid and 2 nM of Cas9-crRNA complex in the reaction buffer. Direct sequencing electropherograms (right panel) of (+) (upper part) and (–) (lower part) strands of pSP1 plasmid cleavage product.

To test whether the Cas9-crRNA complex can recognize the proto-spacer and cut DNA *in vitro* in long DNA substrates, mimicking *in vivo* invading foreign DNA, we analyzed cleavage of pSP1(Figure 30C) carrying proto-spacer1 and the accompanying PAM. In the presence of Cas9-crRNA complex, the supercoiled form of pSP1 plasmid was converted into a linear form (Figure

30D), indicating DNA cleavage, while pUC18 plasmid lacking proto-spacer1 was not cleave. This means that both strands of the pSP1 plasmid were cleaved specifically within the proto-spacer region. We used direct sequencing to determine the ends of the linear DNA molecule generated by Cas9-crRNA cleavage. Sequencing results confirmed that plasmid DNA cleavage occurred 3 nt away from the terminal end of the spacer adjacent to the PAM sequence, as seen in the SP1 oligoduplex cleavage (Figure 30D). The cleavage position identified in the *in vitro* experiments (Figure 30) for the CRISPR3/Cas system of *S. thermophilus* is identical to that determined *in vivo* for the CRISPR1/Cas system of *S. thermophilus* (Garneau *et al*, 2010).

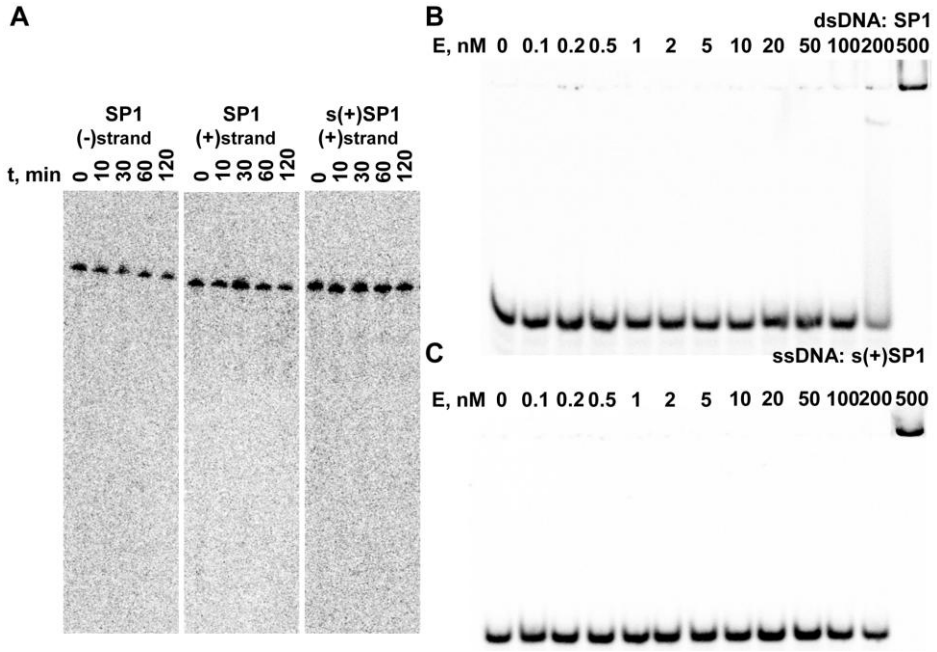

Figure 31. DNA cleavage and binding analysis of Cas9-Chis protein lacking crRNA. (A) Oligonucleotide cleavage assay. 5 nM of Cas9-Chis protein was incubated in the reaction buffer at 37°C with 1 nM oligonucleotide. SP1 oligoduplex was labeled with ^{33}P at the 5'-end of the (+) or (–) strand. Single stranded oligonucleotide s(+)SP1 was labeled with ^{33}P at the 5'-end.EMSA of Cas9-Chis protein binding to **(B)** the double stranded SP1 oligoduplex and **(C)** the single stranded s(+)SP1 oligonucleotide. Electrophoretic mobility shift experiments were performed in the binding buffer. The reactions contained 0.5 nM of the ^{33}P-labeled oligoduplex, and the protein at concentrations as indicated above each lane.

3.3.3. Cas9-crRNA cleavage specificity is directed by the crRNA sequence

To demonstrate directly that Cas9-crRNA complex specificity is pre-programmed by the crRNA guide sequence, we engineered a novel spacer sequence (spacerSN) into the CRISPR locus, purified the Cas9-crRNA complex and monitored cleavage of a plasmid containing a proto-spacerSN complementary to the engineered spacer. More specifically, we first inserted spacerSN instead of spacer1 in SthCRISPR3, generating the pCas(-)SN plasmid containing only a minimal CRISPR array and the tracrRNA-encoding sequence (Figure 32A).

We then co-expressed this plasmid together with pASKIBA-Cas9, and purified the Cas9-crRNA complex on Strep-Tactin Sepharose. The cleavage specificity of the Cas9-crRNA complex was analyzed using pSP1 and pSP1+SPN plasmids. Only the latter plasmid, containing proto-spacerSN matching spacerSN in the CRISPR array, was linearized by the Cas9-crRNA complex, while the pSP1 plasmid, which lacks the complementary sequence, remained intact (Figure 32B). Similar cleavage experiments using the SPN oligoduplex, containing proto-spacerSN and the accompanying PAM (Figure 32D) revealed that cleavage occurred within proto-spacerSN 3 nt upstream of the PAM, identically to other Cas9-crRNA complexes (Figure 32C and D).

Furthermore, to provide direct evidence that Cas9 can be programmed with cognate crRNAs to recognize any DNA target, we show that Cas9-crRNA complexes targeting two different proto-spacers in pUC18 plasmid cleave DNA at positions specified by the complimentary crRNA (Figure 33B and C).

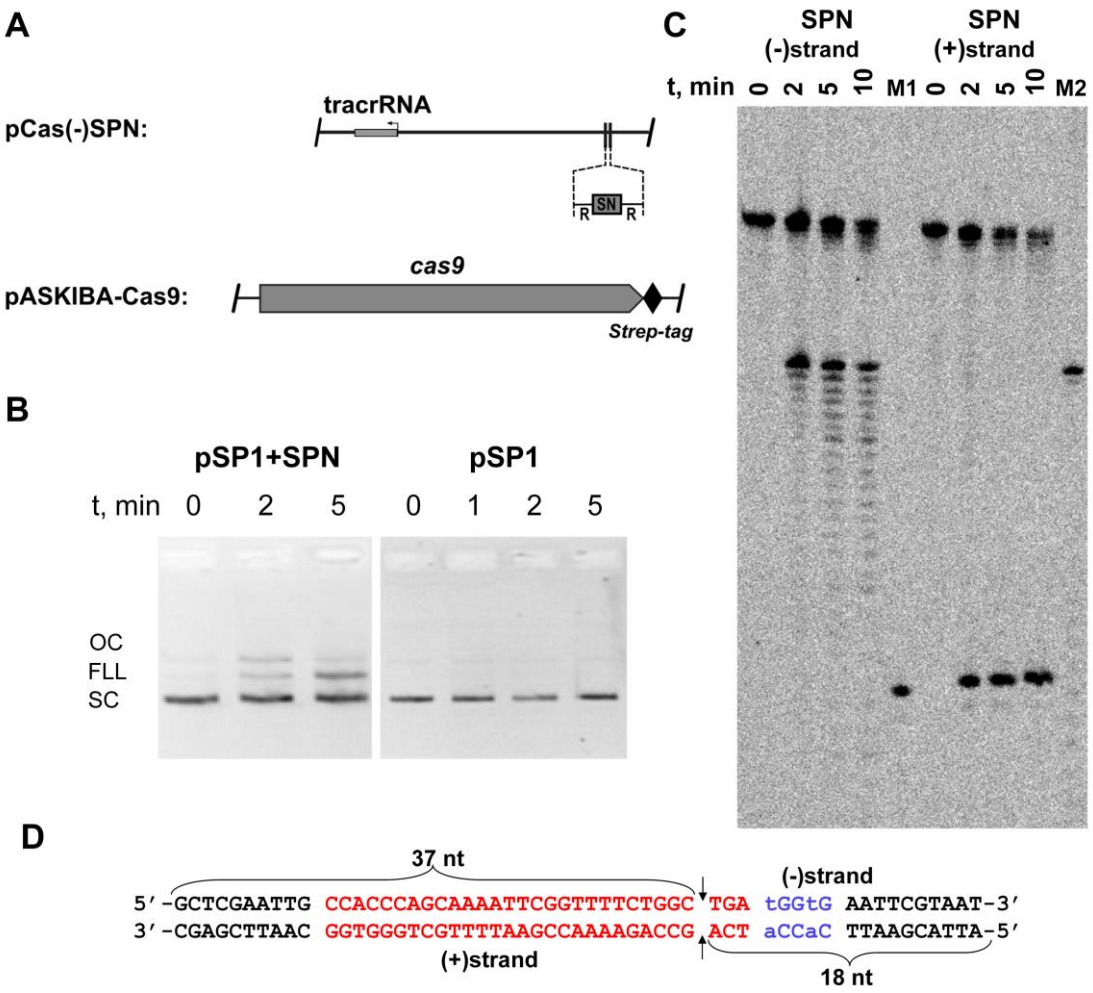

Figure 32. Reprograming of Cas9-crRNA complex. (A) Schematic representation of heterologous loci in two plasmids used for reprogramming of Cas9-crRNA complex. **(B)** Agarose gel analysis of plasmid DNA cleavage products. pSP1 and pSP1+SPN (pSP1 plasmid with inserted new proto-spacer and PAM over AatII site) were incubated at 2.5 nM concentration with 2 nM of Cas9-crRNA complex in the reaction buffer for varied time intervals and reaction products analyzed in the agarose gel. SC – super-coiled plasmid DNA, OC – open circular DNA nicked at one of DNA strands, FLL – full length linear DNA cut at both strands. **(C)** Oligoduplex SPN cleavage. 2.5 nM of Cas9-crRNA complex and 1 nM SPN oligoduplex (Table 3) labeled at the 5'-end of either (+) or (–)strand were incubated in the reaction buffer at 37°C. Lanes M1 and M2 contain chemically synthesized 5'-end ^{33}P-labeled 18 nt and 37 nt oligodeoxynucleotides corresponding to the cleavage products of (+) and (-) DNA strands, respectively. **(D)** Schematic representation of SPN oligoduplex substrate and cleavage products. SPN oligoduplex contains the new proto-spacer (red letters), PAM (blue letters). Cleavage positions are designated by arrows.

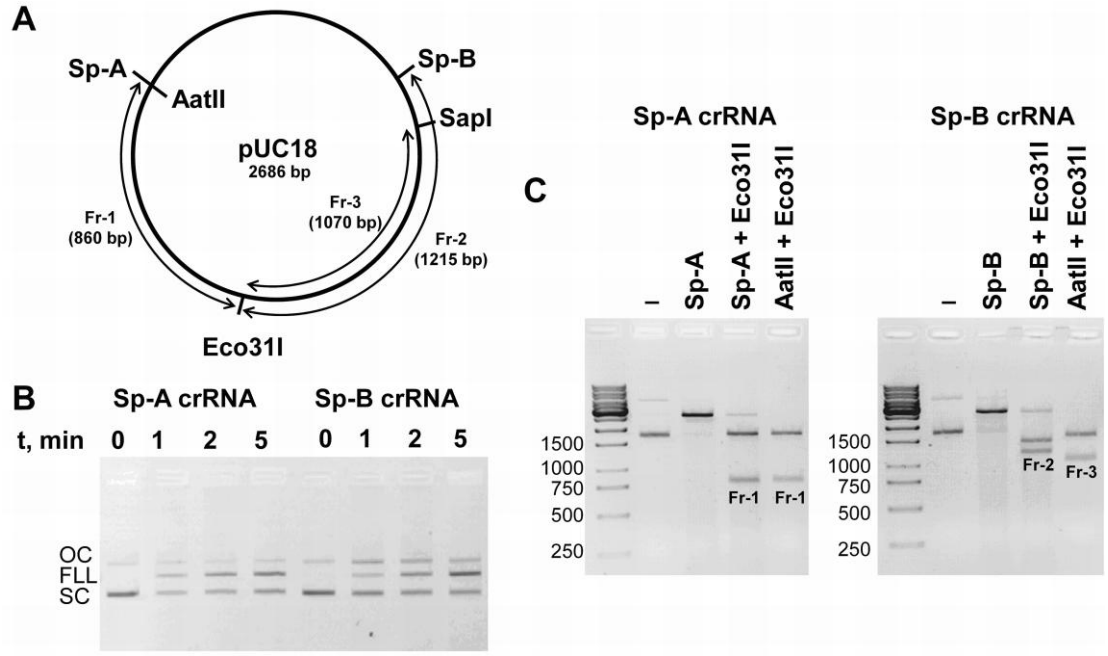

Figure 33. pUC18 cleavage by reprogrammed Cas9-crRNA complexes. (A) Schematic representation of target sites and cleavage fragments length. Sp-A sequence overlaps with AatII site, Sp-B sequence is located 145 bp away of the SapI site. **(B)** Agarose gel analysis of pUC18 cleavage products. pUC18 was incubated at 2.5 nM concentration with 2 nM of complex containing Sp-A or Sp-S crRNA in the reaction buffer at 37°C. SC – super-coiled plasmid DNA, OC – open circular DNA nicked at one of DNA strands, FLL – full length linear DNA cut at both strands. **(C)** pUC18 cleavage by combined action of Cas9-crRNA complex and a restriction enzyme. 2.5 nM of pUC18 was incubated with 5 nM of Cas9-crRNA complex for 15 min at 37°C (Sp-A and Sp-B lanes). Reactions were stopped by heating at 65°C for 5 min. The Eco31I restriction enzyme was added to the mix and incubated at the 37°C for 15 min to yield fragment Fr-1 (Sp-A + Eco31I) and Fr-2 (Sp-B + Eco31I). The DNA fragments, obtained by cleaving pUC18 plasmid with AatII + Eco31I (Fr-1) and AatII + Eco31I (Fr-3) restriction enzymes were used as size markers. Traces of linear pUC18 DNA plasmid in lanes Sp-A + Eco31I and Sp-B + Eco31I are due to the incomplete Cas9-crRNA cleavage.

The typical length of *S. thermophilus* CRISPR3 spacers is 30 nt (Horvath *et al*, 2008). According to the data provided in Figure 27C, the mature crRNA that co-purified with the Cas9 protein has a length of 42 nt. Considering that 22 nt derive from the 3′ of the repeat sequence, this means that only 20 nt of the crRNA are complementary to the (+)strand of the proto-spacer. To assess whether the 5′-end of proto-spacers is important for plasmid interference, we engineered plasmids pSP1-27, pSP1-23, pSP1-19, pSP1-15, pSP1-11 with a 5′-

truncated proto-spacer1, and analyzed transformation efficiency of the recipient strain containing pCRISPR3 (Figure 34B).

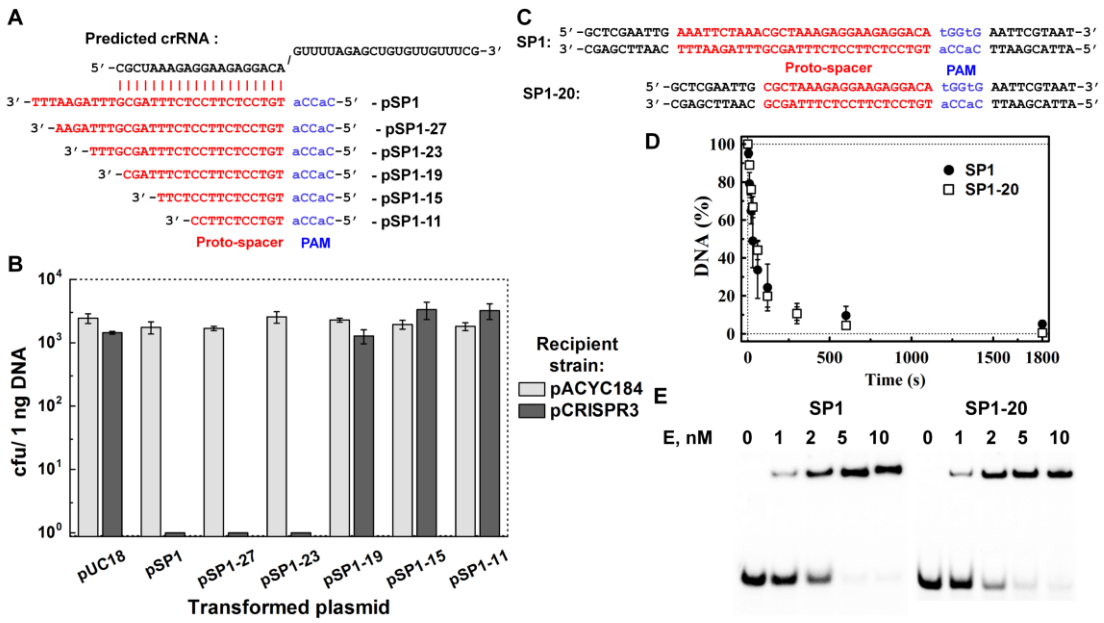

Figure 34. Impact of spacer length on CRISPR-encoded immunity. (A) Schematic representation of shortened versions of proto-spacers inserted in the transformed plasmids**. (B)** Effect of proto-spacer length on the plasmid transformation efficiency. Transformation efficiency is expressed as cfu per ng of plasmid DNA (mean ± SD). **(C).** Schematic representation of oligoduplexes used in the *in vitro* cleavage and binding experiments. **(D)** Time courses of the 30 bp oligoduplex (full length proto-spacer SP1, filled circles) and the 20 bp oligoduplex (truncated proto-spacer SP1-20, square) cleavage by the Cas9-crRNA complex. **(E)** Electrophoretic mobility shift assay of SP1 and SP1-20 oligoduplex binding by the Cas9-crRNA complex.

Plasmids containing 4 or 7 bp truncations at the 5′ end of proto-spacer1 had no effect on the recipient strain ability to interfere with plasmid transformation. Shorter versions of proto-spacer A abolished the ability of the recipient strain to prevent plasmid transformation. These data show that the 10 nt at the 5′ end of proto-spacers, that have no complementarity to mature crRNAs, are not necessary for CRISPR3/Cas-mediated interference. In full support to the *in vivo* experiments, the SP1-20 oligoduplex containing only 20 nt of proto-spacer1 is efficiently cleaved by Cas9-crRNA (Figure 34D and E).

3.3.4. The PAM is required for DNA binding and cleavage by the Cas9-crRNA complex

Plasmids carrying a proto-spacer but no PAM (pSP1-pΔ) or multiple PAMs but no accompanying proto-spacer (pUC18) are resistant to Cas9-crRNA cleavage (Figure 35A). Hence, in accordance with *in vivo* data, both PAM and proto-spacer are required for double-stranded DNA cleavage by Cas9-crRNA complex. To assess whether the PAM is recognized in the context of a double-stranded or a single-stranded DNA, we analyzed Cas9-crRNA binding to and cleavage of oligonucleotides i) SP1 (contains both proto-spacer and PAM), ii) SP1-Δp (contains only proto-spacer), and iii) SP2 (contains only PAM). The (+)strands of these oligodeoxynucleotides were used as single-stranded DNA substrates (s(+)SP1, s(+)SP1-Δp, s(+)SP2, accordingly) (Table 3).

Consistent with the plasmid cleavage experiments, oligoduplexes which have only proto-spacer but no associated PAM are not cut by Cas9-crRNA (Figure 35B). On the other hand, (+)strand in the single-stranded form is cut at a similar rate independently whether it is associated with a PAM, or not (Figure 35B). These data clearly show that a PAM is required only for double-stranded DNA cleavage.

To test whether the PAM is necessary for DNA binding by the Cas9-crRNA complex, electrophoretic mobility shift assays were performed. To avoid cleavage, binding experiments were performed in the absence of Mg^{2+} ions, which are necessary for cleavage. Cas9-crRNA showed different binding patterns for ds- and ss-oligonucleotides. In the case of the SP1 oligoduplex a low mobility complex is observed already at 1 nM concentration (Figure 35C). On the other hand, no binding is observed under the same experimental conditions for oligoduplexes without PAM (SP1-Δp) or without proto-spacer (SP2). Moreover, no low mobility complex is observed in the case of Cas9 protein without crRNA (at high Cas9 concentration (500 nM) the DNA remains in gel wells due to non-specific interactions)(Figure 31B), confirming that crRNA is necessary for complex formation. Thus, binding experiments

altogether show that Cas9-interaction with dsDNA requires both PAM and crRNA complementarity to the target.

On the other hand, single-stranded oligonucleotides ((+)strand) are bound by Cas9-crRNA with the same affinity independent of PAM presence (Figure 35D). Again, no binding was observed for ssDNA oligonucleotide corresponding to the (-)strand (Figure 35D), or for Cas9 protein lacking crRNA (Figure 31C).

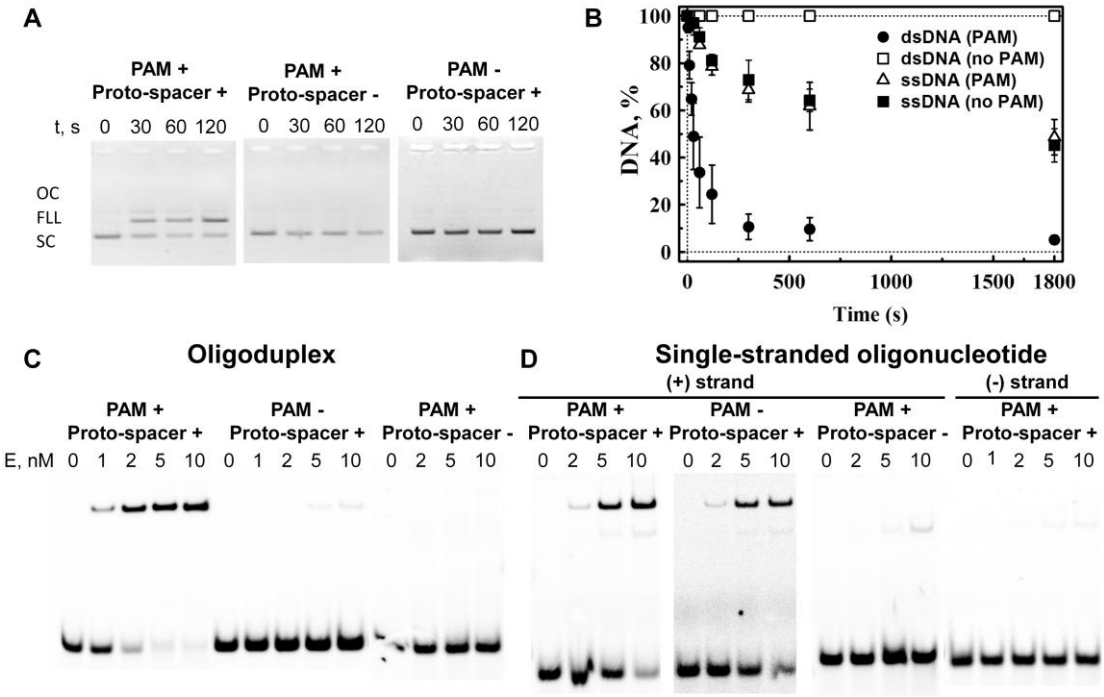

Figure 35. A PAM is required for *in vitro* DNA binding and cleavage by the Cas9-crRNA complex. (A) Agarose gel analysis of plasmid DNA cleavage products. Three different plasmids, PAM+Proto-spacer+ (pSP1 containing both the proto-spacer and PAM), PAM-Proto-spacer- (pUC18 containing multiple PAMs but no proto-spacer), and PAM-Proto-spacer+ (pSP1-pΔ containing a proto-spacer without PAM) were incubated at 2.5 nM concentration with 2 nM of Cas9-crRNA complex in the reaction buffer. SC – super-coiled plasmid DNA, OC – open circular DNA nicked at one of DNA strands, FLL – full length linear DNA cut at both strands. **(B)** Time courses of (+)strand hydrolysis in the single-stranded and double-stranded oligodeoxynucleotides. Reactions containing 2 nM Cas9-crRNA and 1 nM of oligodeoxynucleotide were conducted at 37°C in the reaction buffer. SP1 (filled circles) and SP1-pΔ (open squares) oligoduplexes were used as dsDNA. s(+)SP1 (open triangles) and s(+)SP1-pΔ (filled squares) oligonucleotides were used as ssDNA. **(C)** and **(D)** dsDNA and ssDNA (+)strand) binding by Cas9-crRNA complex. The reactions contained 0.5 nM of the ^{33}P-labeled ssDNA or dsDNA oligonucleotide, and the protein at concentrations as indicated above each lane.

Since some Type III CRISPR systems provide RNA rather than DNA interference, we have studied RNA binding and cleavage by the Cas9-crRNA complex. Cas9-crRNA did not cleave specifically either single-stranded RNA, or double-stranded RNA bearing a proto-spacer and PAM (Figure 36B). This finding is consistent with DNA as the primary target for Type II system-mediated interference. The Cas9-crRNA complex binds a complementary RNA containing a proto-spacer, but bound RNA is not cleaved specifically by Cas9 within the proto-spacer.

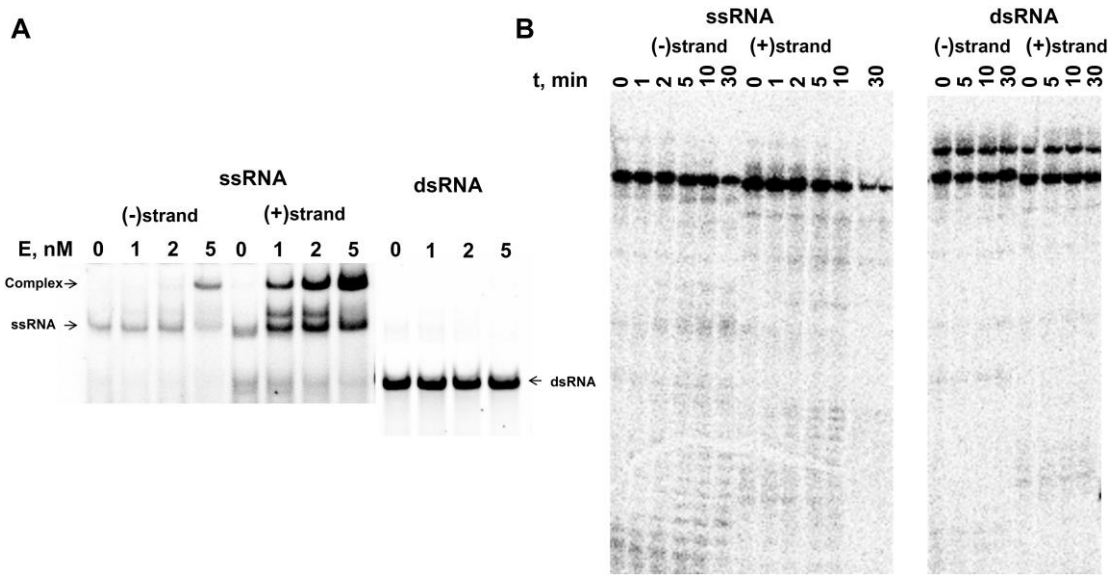

Figure 36. RNA binding and cleavage analysis of Cas9-crRNA complex. (A) Electrophoretic mobility shift analysis (EMSA) of Cas9-crRNA complex binding to 84 nt RNA fragment containing proto-spacer1, PAM and 24 nt flanking sequences on both sides. Left panel: RNA (–) strand; center panel: RNA (+) strand; right panel: double stranded RNA The reactions contained 1 nM of the ^{33}P-labeled RNA fragment, and the protein at concentrations as indicated above each lane. After 15 min at room temperature, the samples were subjected to PAGE for 2 h and analyzed as described in 'Materials and Methods'. **(B)** RNA cleavage assay. 2.5 nM of Cas9-crRNA complex was incubated in the reaction at 37°C in the presence of 1 nM (+) and (-) RNA strands(left panel) or double stranded RNA labeled on (+) or (–) strand (right panel). Reaction products were analyzed on denaturing PAA gel.

3.3.5. Mutagenesis of RuvC and HNH motifs in Cas9

Previous plasmid transformation experiments have revealed that RuvC and HNH motifs (Figure 37) are important for Cas9 function. To test whether these motifs are involved in target DNA cleavage, we constructed, expressed, and purified D31A and N891A Cas9 mutants. Both mutants co-purified with

crRNA that appeared identical to the crRNA in the WT Cas9 complex (Figure 27C). To determine whether mutant proteins retained cleavage activity, we monitored pSP1 plasmid cleavage by mutant Cas9-crRNA complexes. Surprisingly, both mutants generated nicked DNA (Figure 37B), indicating that both active site mutants cleave only one DNA strand of the plasmid substrate.

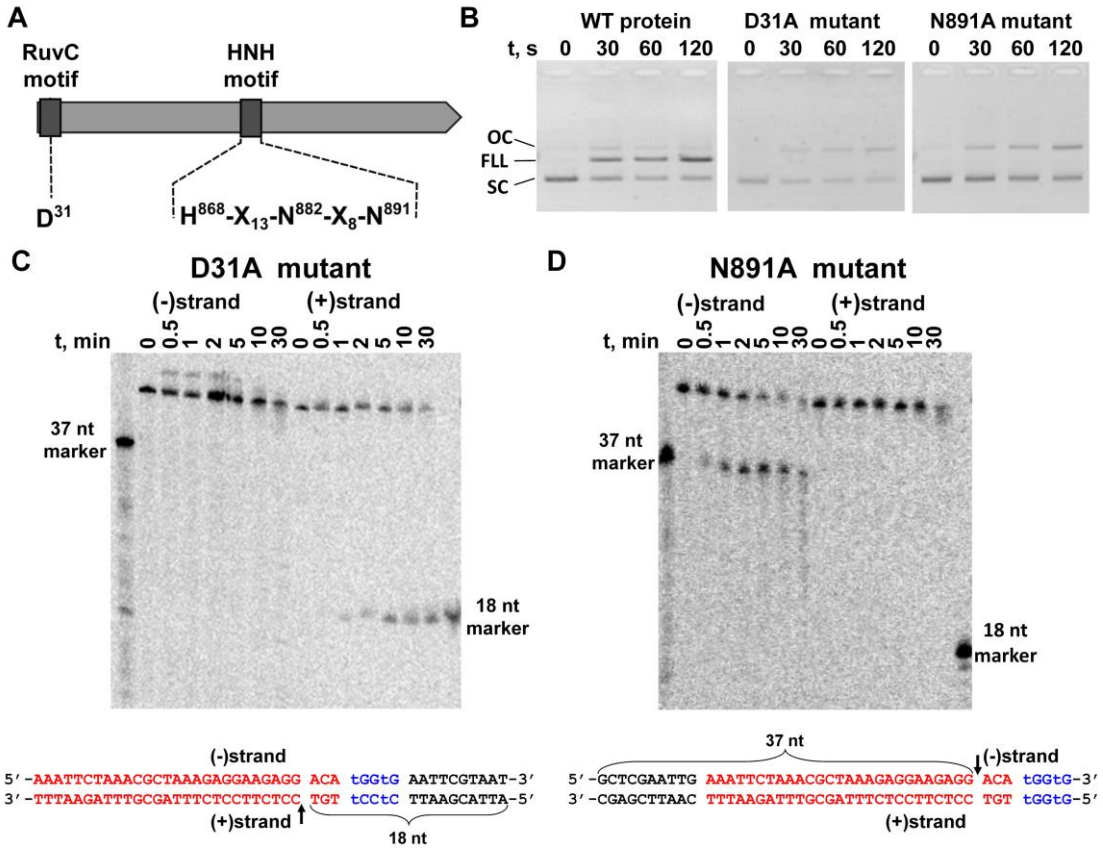

Figure 37. RuvC and HNH active site motifs of Cas9 contribute to the cleavage of opposite DNA strands. (A) Localization of the conserved active site motifs within Cas9 protein. Amino acid residues identified as crucial for Cas9 *in vivo* activity are indicated. **(B)** Agarose gel analysis of pSP1 plasmid cleavage by Cas9 and mutant proteins. Reactions were performed as described in 'Materials and Methods'. **(C)** Strand preference of D31A mutant. Reactions were performed as described in 'Materials and Methods'. D31 mutant cleaves only (+)strand of SP1 oligoduplex. **(D)** Strand preference of N891A mutant. N891A mutant cleaves only (-)strand of SP1 oligoduplex. Cleavage positions are indicated by arrows.

To determine whether mutant proteins exhibit a strand preference, we analyzed D31A and N891A mutant cleavage of the SP1 oligoduplex. The RuvC mutant (D31A) cut (+)strand of oligoduplex at the same position as WT Cas9-crRNA protein, whereas the (-)strand stayed intact (Figure 37C).

Conversely, the HNH mutant (N891A) cleaved only (-)strand, but not (+)strand of the SP1 oligoduplex (Figure 37D), indicating that each active site acts on opposite DNA strands to generate a double-strand break.

To find out whether the same cleavage pattern is conserved during plasmid DNA cleavage, we sequenced the proto-spacer regions of nicked plasmids. Run-off sequence data confirmed that RuvC active site is implicated in DNA (-)strand cleavage, while HNH/McrA is involved in (+)strand cleavage (Figure 38A and B). Furthermore, we found that the RuvC mutant cleaved (+)strand of a single-stranded DNA but no such cleavage was detected for the HNH mutant (Figure 38D). To test whether mutations altered DNA-binding affinity of protein-crRNA complexes, DNA binding was studied using the electrophoretic mobility shift assay. Both mutant protein-crRNA complexes bound oligoduplex SP1 with the same affinity as wild type (Figure 38C). Thus, mutations in the putative active sites of Cas9 have no significant effect on double-stranded DNA-binding properties of the Cas9-crRNA complex. Since 42-nt crRNA was present in the mutant protein complexes (Figure 27C), we conclude that mutant Cas9-crRNA complexes lost the ability to cut one of the target DNA strands. Since the Cas9-HisTag protein is a monomer in solution (Figure 39), it is likely that Cas9 is functional as a monomer and uses two active sites for the cleavage of opposite DNA strands. A similar strategy is implemented by some restriction endonucleases (Armalyte *et al*, 2005)

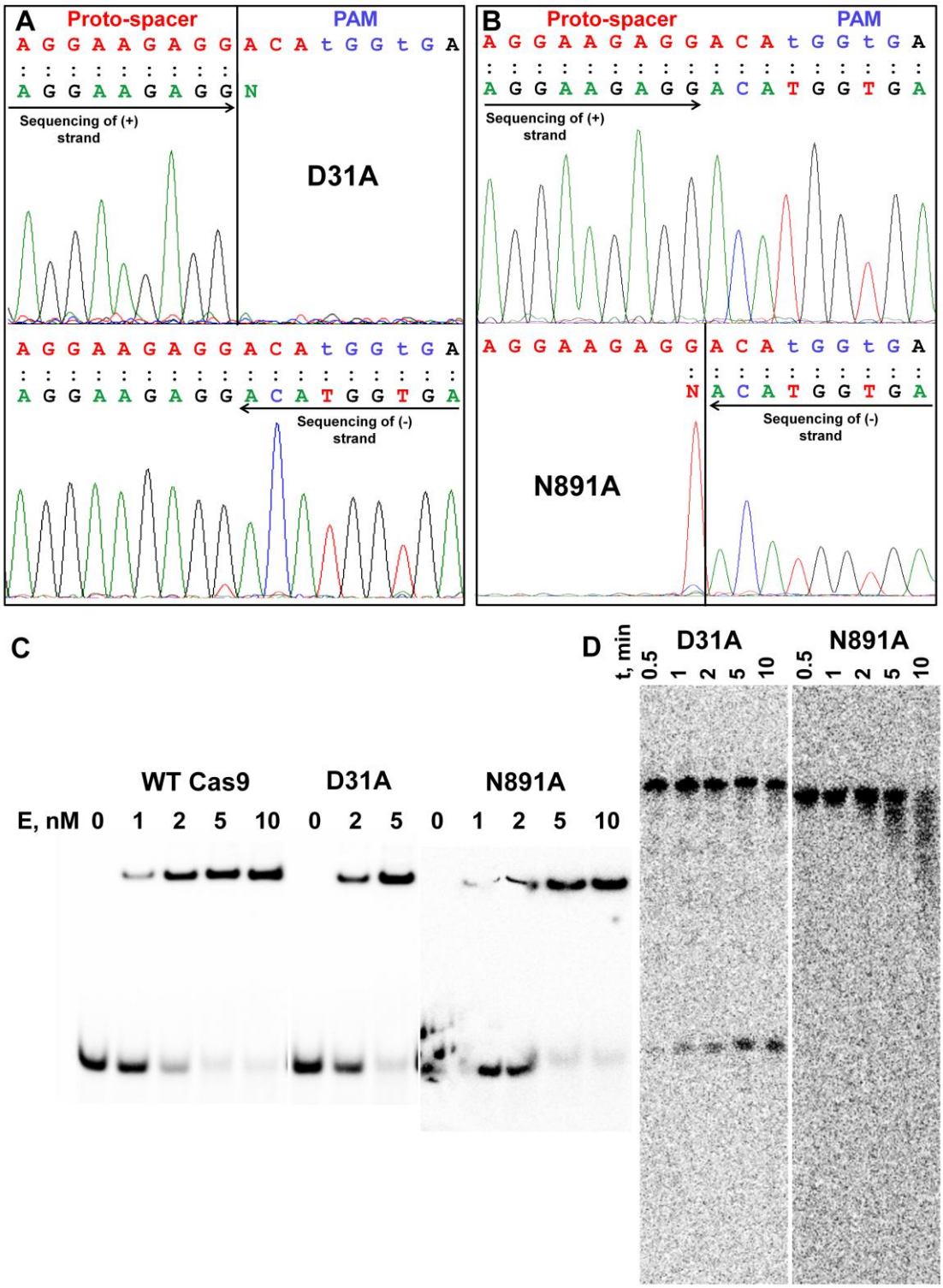

Figure 38. Properties of Cas9 active site mutant-crRNA complexes. (A) Direct sequencing of reaction products obtained with Cas9 mutant D31A (RuvC-like active site motif). **(B)** Direct sequencing of reaction products obtained with Cas9 N891A mutant (HNH-like active site motif). **(C)** SP1 oligoduplex binding by the wt Cas9-crRNA and active site mutant complexes. **(D)** Cleavage of single stranded (+)SP1 oligonucleotide by Cas9-crRNA mutant complexes.

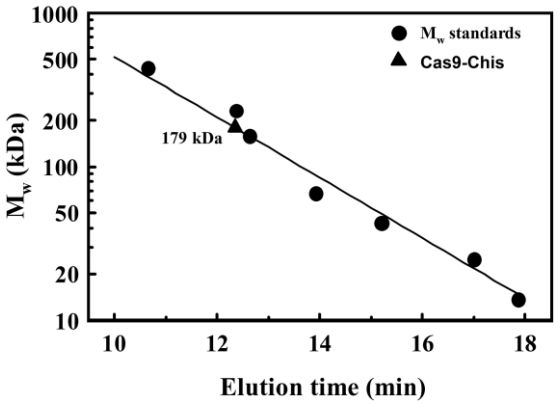

Figure 39. Molecular mass of the WT Cas9-Chis protein. The apparent M_w of Cas9 (black triangle) were calculated by interpolation from the standard curve obtained using a set of proteins of known M_w (black circles (Bio-Rad Gel Filtration Standards).

3.3.6. The Cas9-crRNA complex guides dsDNA cleavage

This work demonstrates that the Cas9-crRNA complex of Type II systems is a crRNA-loaded endonuclease which cuts both DNA strands within the proto-spacer in the presence of Mg^{2+} ions, 3 nt upstream of the PAM sequence, to produce blunt-end cleavage products. Sequence specificity of the Cas9-crRNA complex is dictated by the 42-nt crRNA which includes a 20-nt fragment complementary to the proto-spacer sequence in the target DNA. Therefore, crRNA present in the Cas9-crRNA complex of the *S. thermophilus* CRISPR3/Cas system is complementary only to the part of the proto-spacer sequence adjacent to the PAM. Not surprisingly, truncation of the distal part of the proto-spacer sequence by 10 nt has no effect on Cas9-crRNA cleavage of synthetic oligoduplexes or plasmid DNA (Figure 34).

The cleavage machinery of Cas9-crRNA complex resides in the Cas9 protein which provides two active sites for the phosphodiester bond cleavage. The RuvC- and HNH-like active sites of Cas9 are located on different domains and act independently on individual DNA strands. Alanine replacement of the active site residues in the RuvC- and HNH-motifs transforms Cas9-crRNA complex into a strand-specific nicking endonuclease, similar to nicking enzymes (Chan *et al*, 2011). Consistent with *in vivo* studies, the functional activity of the Cas9-crRNA complex *in vitro* is strictly dependent on the presence of the PAM 5'-NGGNG-3', in the immediate vicinity of the proto-spacer sequence. Data also show that the PAM is required for Cas9-crRNA

binding to the dsDNA. If the PAM sequence is missing in dsDNA, the Cas9-crRNA complex does not bind such DNA, even if it contains a complementary proto-spacer sequence. On the other hand, Cas9-crRNA does not bind DNA containing PAM (or multiple PAMs) if the proto-spacer sequence is absent. Thus, consistent with the *in vivo* data, both PAM and proto-spacer sequences are necessary prerequisite for dsDNA binding and subsequent cleavage. Contrary to dsDNA, the PAM is not absolutely required for single-stranded DNA binding and cleavage: a single-stranded oligodeoxynucleotide containing proto-spacer with or without a PAM sequence is bound equally well but with lower affinity than dsDNA. Moreover, in the presence of Mg^{2+} ions, Cas9 cuts single-stranded DNA bound to the crRNA using its HNH-active site independently of PAM.

3.3.7. Mechanism of DNA interference in Type II systems

Our results shed light on the mechanism of double-stranded DNA cleavage by the Cas9-crRNA complex in the *S. thermophilus* CRISPR3/Cas system, a model Type II CRISPR/Cas system (Figure 40).

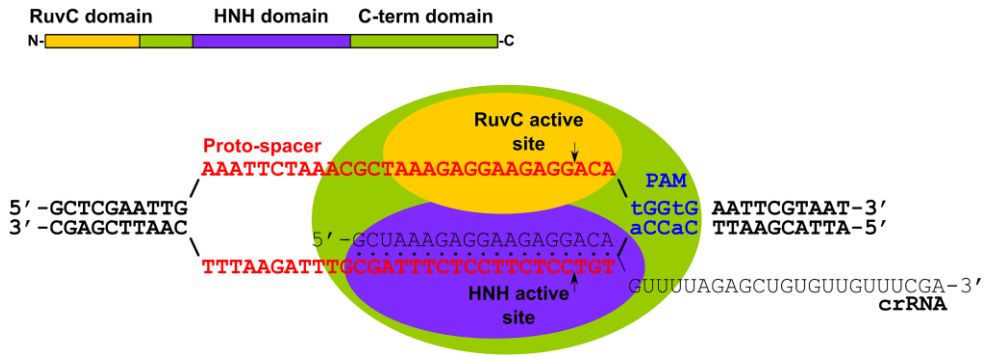

Figure 40. Schematic arrangement and mechanism of crRNA-directed DNA cleavage by the Cas9-crRNA complex. Domain architecture of Cas9 is shown schematically on the top. Cas9-crRNA complex binds to the dsDNA containing a PAM. crRNA binds to the complementary (+)strand, resulting in DNA strand separation and the R-loop formation. In the ternary complex, the RuvC active site of Cas9 is positioned at the scissile phosphate on the unpaired (-)strand, while the HNH active site is located at the scissile phosphate on the DNA (+)strand bound to crRNA. Coordinated action of both active sites results in the double strand break 3 nt upstream the PAM, generating blunt end DNA.

Cas9-crRNA complexes, using a mechanism that yet has to be defined, locate and bind to a proto-spacer sequence within the double-stranded DNA in a PAM-dependent process. The absolute requirement of PAM for dsDNA binding by the Cas9-crRNA complex implies that PAM serves as a priming site for strand separation or is essential for stabilization of the R-loop structure because dsDNA lacking PAM is not bound.

Despite limited mechanistic details, our data clearly demonstrate that the PAM is recognized by Cas9-crRNA in the context of double-stranded DNA. The Cas9-crRNA binding to the target sequence in the dsDNA presumably results in a R-loop structure, where the (-)strand is displaced and the complementary (+) DNA strand is paired with the crRNA. In the presence of Mg^{2+} ions phosphodiester bond cleavage occurs on both strands 3 nt upstream of the PAM sequence to generate blunt DNA ends. Further DNA degradation can be accomplished by host nucleases. DNA cleavage analysis by the RuvC- and HNH-motif mutants demonstrate that RuvC- and HNH-like active sites of Cas9 act on the (-) and (+)strands, respectively. Therefore, in the catalytically active Cas9-crRNA complex, the N-terminal domain containing the catalytic D31 residue of the RuvC motif is positioned at the displaced (–) DNA strand, while the central part of Cas9 containing the HNH motif is located in the vicinity of the scissile phosphodiester bond of (+) DNA strand paired to loaded crRNA.

Interestingly, after DNA cleavage, Cas9-crRNA remains bound to the reaction products (Figure 41). Altogether, data presented here provide a molecular basis for DNA interference by Type II CRISPR/Cas systems. Since *cas9* is a signature gene (Makarova *et al*, 2011b) for Type II-A and Type II-B systems, the cleavage mechanism proposed here is likely to be conserved in other Type II-A and Type II-B systems.

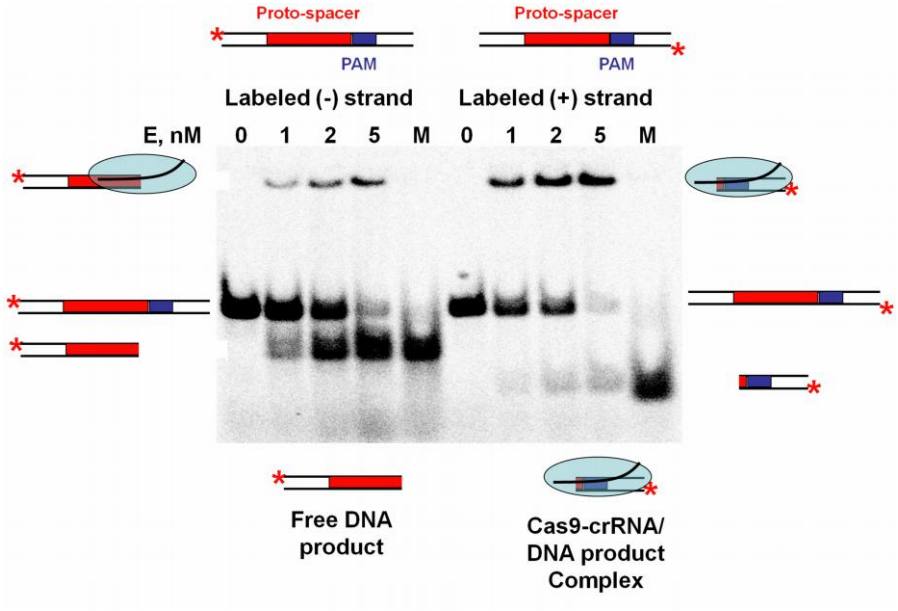

Figure 41. Native electrophoresis of Cas9-crRNA and cleavage products. The protein at concentrations as indicated above each lane, where incubated in the reaction buffer at 37°C for 30 min in the presence of 0.5 nM SP1 oligoduplex. Samples was mixed with loading dye solution and analyzed by non-denaturing PAGE. The gel lanes marked M – melted form of cleavage reactions products. The cartoons in each side of the gel illustrate protein-DNA complexes and DNA that correspond to each band, while cartoons below the gel illustrate major substrate form after reaction.

3.3.8. Comparison to other effector complexes

The mechanism proposed here for the double-stranded DNA cleavage by the Cas9-crRNA complex differs significantly from that for the Type I-E (former Ecoli) system (Jore *et al*, 2011). In the *E. coli* Type I-E system crRNA and Cas proteins assemble into a large ribonucleoprotein complex named Cascade that facilitates target recognition by enhancing sequence-specific hybridization between the crRNA and complementary target sequences (Jore *et al*, 2011). Target recognition is dependent on PAM and governed by the "seed" crRNA sequence located at the 5′-end of the spacer region (Semenova *et al*, 2011). However, while Cascade-crRNA complex alone is able to bind double-stranded DNA containing PAM and proto-spacer, it requires an accessory Cas3 protein for DNA cleavage. Cas3 is a single-stranded DNA nuclease and helicase which is able to cleave single-stranded DNA, producing multiple cuts

(Sinkunas *et al*, 2011). It has been demonstrated recently that Cas3 degrades plasmid DNA *in vitro* in *E. coli*, in the presence of Cascade-crRNA complex (Westra *et al*, 2012). Thus, current data clearly show that mechanistic details of the interference step for the Type I-E system differ from that of Type II systems, both by the catalytic machinery involved and the nature of the molecular mechanisms.

In Type III-B CRISPR/Cas systems, present in many archaea and some bacteria, Cas module RAMP (Cmr) proteins and crRNA assemble into an effector complex that targets RNA (Hale *et al*, 2012, 2009). In *P. furiosus* the RNA silencing complex, comprised of six proteins (Cmr1 to Cmr6) and crRNA, binds to the target RNA and cleaves it at fixed distance from the 3′end. The cleavage activity depends on Mg^{2+} ions, however individual Cmr proteins responsible for target RNA cleavage have yet to be identified. The effector complex of *S. solfataricus*, comprised of seven proteins and crRNA, cuts invading RNA in an endonucleolytic reaction at UA dinucleotides (Zhang *et al*, 2012). Importantly, these two archaeal Cmr-crRNA complexes perform RNA cleavage in a PAM-independent manner.

Overall, we have shown that the Cas9-crRNA complex in Type II CRISPR/Cas systems is a functional homolog of Cascade in Type I systems and so far represents a minimal DNA interference complex. The simple modular organization of the Cas9-crRNA complex, where specificity for DNA targets is encoded by crRNAs and the cleavage enzymatic machinery is brought by a single, multidomain Cas protein, provides a versatile platform for the engineering of universal RNA-guided DNA endonucleases. Indeed, by altering the RNA sequence within the Cas9-crRNA complex, programmable endonucleases can be designed both for *in vitro* and *in vivo* applications. In our effort to provide proof of principle of such a strategy, we engineered *de novo* into a CRISPR locus a spacer targeted to a specific sequence on a plasmid and demonstrated that such a plasmid is cleaved by the Cas9-crRNA complex at a sequence specified by the designed crRNA. Experimental demonstration that RuvC and HNH active site mutants of Cas9 are functional as strand-specific nicking enzymes opens the possibility to generate programmed DNA single strand

breaks *de novo*. Taken together, these findings pave the way for development of novel molecular tools for RNA-directed DNA surgery.

CONCLUSIONS

1. CRISPR3/Cas system from Gram-positive *S. thermophilus* bacteria provides protection against plasmid transformation and phage infection in the Gram-negative *E. coli* host.
2. Cas9 is the sole Cas protein required for immunity step in Type II CRISPR/Cas systems.
3. Cas9 protein forms a ribonucleoprotein complex with 42 nt crRNA.
4. Cas9 protein guided by crRNA cleaves double-stranded DNA at the specific position in a proto-spacer adjacent motif dependent manner.
5. Cas9 uses two distinct active sites, RuvC and HNH, to generate site-specific nicks on opposite DNA strands.

LIST OF PUBLICATIONS

The thesis is based on the following original publications:

1. Sapranauskas, R., **Gasiunas, G*.**, Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res*. *39*, 9275–82.

* - joint first author together with R. Sapranauskas

2. **Gasiunas, G.**, Barrangou, R., Horvath, P., ir Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl AcadSci USA. 109(39)*, E2579-86.

Other publications:

3. Sinkunas, T., **Gasiunas, G.,** Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J*. *30*, 1335–42.

4. **Gasiunas, G.,** Sasnauskas, G., Tamulaitis, G., Urbanke, C., Razaniene, D., and Siksnys, V. (2008). Tetrameric restriction enzymes: expansion to the GIY-YIG nuclease family. *Nucleic Acids Res*. *36*, 938–49.

5. Armalyte, E., Bujnicki, J.M., Giedriene, J., **Gasiunas, G.,** Kosiński, J., and Lubys, A. (2005). Mva1269I: a monomeric type IIS restriction endonuclease from Micrococcus varians with two EcoRI- and FokI-like catalytic domains. *J Biol Chem*. *280*, 41584–94.

Book chapters:

6. Horvath P, Gasiunas G, Siksnys V and Barrangou R. (2013) Applications of the Versatile CRISPR-Cas Systems. „CRISPR-Cas Systems“. R. Barrangou and J. van der Oost (eds.), Springer-Verlag Berlin Heidelberg (in press).

Patent application:

7. Siksnys, V., **Gasiunas, G.,** Karvelis, T. (2012) RNA-directed DNA cleavage by the Cas9-crRNA complex from CRISPR3/Cas immune system of *Streptococcus thermophilus*. USA provisional patent application Nr.: 61/613,373.

CONFERENCE PRESENTATIONS

1. **G. Gasiunas,** R. Sapranauskas, P. Horvath, V. Siksnys. Cas protein complexes. Production and characterization of macromolecular complexes, IGBMC, Illkirch, Strasbourg, France; 2012.03.19-21.

2. **G. Gasiunas,** R. Sapranauskas, G. Sasnauskas, P. Horvath, V Siksnys. Quest for S. thermophilus Cas protein function. CRISPR: Mechanism and application. Wageningen University, Wageningen, Netherlands; 2010.10.21-22.

3. R. Sapranauskas, **G. Gasiunas,** P. Horvath, V. Siksnys. Cloning of the S. thermophilus CRISPR3 system in heterologous host. CRISPR: Mechanism and application. Wageningen university, Wageningen, Netherlands; 2010.10.21-22.

4. **G. Gasiunas,** R. Sapranauskas, G. Sasnauskas, P. Horvath, V Siksnys. Quest for S. thermophilus Cas protein function. 6th New England Biolabs Meeting on Restriction/Modification, Jacobs University, Bremen, Germany; 2010.08.01-06.

FINANCIAL SUPPORT

The work was funded by the European Social Fund under the Global Grant measure (project R100) and Programme Gilibert between the Lithuanian Ministry of Education and Science and by French Ministry of Foreign and European affairs (MAEE); French Ministry of Higher education and Research (MESR).

ACKNOWLEDGEMENT

I wish to thank my supervisor Prof. Virginijus Šikšnys for educational discussions, suggestions and help with preparation of this dissertation.

I thank to my former supervisors Elena Armalytė, dr. Giedrius Sasnauskas and dr. Gintautas Tamulaitis for teaching me to work in the laboratory.

I also would like to thank dr. Rimantas Šapranauskas for cloning and mutagenesis of CRISPR3/Cas system, and discusions.

I am grateful to dr. Philippe Horvath from DuPont (former Danisco) for providing genomic DNA of *S. thermophilus* DGCC7710 strain.

My thanks to all the colleagues in the Department of Protein-DNA Interactions, especially to dr. Giedrius Sasnauskas, dr. Giedrė Tamulaitienė, dr. Gintautas Tamulaitis, dr. Mindaugas Zaremba and dr. Arūnas Šilanskas, for advices, discusions and help with experiments.

Many thanks to my family and friends for their concern and support. Particular thanks to my fiancee Agnė for immense support, patience and understanding.

REFERENCES

1. Armalyte E, Bujnicki JM, Giedriene J, Gasiunas G, Kosiński J & Lubys A (2005) Mva1269I: a monomeric type IIS restriction endonuclease from *Micrococcus varians* with two EcoRI- and FokI-like catalytic domains. *J Biol Chem* **280:** 41584–94
2. 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
3. Barik S (1996) Site-Directed Mutagenesis In Vitro by Megaprimer PCR. *Methods Mol Biol* **57:** 203–215
4. 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:** 1709–12
5. 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:** 20361–71
6. 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:** 4616–27
7. 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:** 273–97
8. Bingham R, Ekunwe SI, Falk S, Snyder L & Kleanthous C (2000) The major head protein of bacteriophage T4 binds specifically to elongation factor Tu. *J Biol Chem* **275:** 23219–26
9. Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Heyneker HL, Boyer HW, Crosa JH & Falkow S (1977) Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2:** 95–113
10. Bolotin A, Quinquis B, Sorokin A & Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151:** 2551–61
11. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin EV & van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321:** 960–4
12. Brüssow H & Hendrix RW (2002) Phage genomics: small is beautiful. *Cell* **108:** 13–6

13. Cady KC & O'Toole G a (2011) Non-Identity-Mediated CRISPR-Bacteriophage Interaction Mediated via the Csy and Cas3 Proteins. *J Bacteriol* **193:** 3433–45

14. Carte J, Pfister NT, Compton MM, Terns RM & Terns MP (2010) Binding and cleavage of CRISPR RNA by Cas6. *RNA* **16:** 2181–8

15. 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:** 3489–96

16. Chan S-H, Stoddard BL & Xu S-Y (2011) Natural and engineered nicking endonucleases--from cleavage mechanism to engineering of strand-specificity. *Nucleic Acids Res* **39:** 1–18

17. Chang AC & Cohen SN (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* **134:** 1141–56

18. Chen C-S, Korobkova E, Chen H, Zhu J, Jian X, Tao S, He C & Zhu H (2008) A proteome chip approach reveals new DNA damage recognition activities in *Escherichia coli*. *Nat Methods* **5:** 69–74

19. 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

20. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada Z a, Eckert MR, Vogel J & Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471:** 602–7

21. Destoumieux-Garzón 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:** 869–76

22. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P & Moineau S (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* **190:** 1390–400

23. Ebihara A, Yao M, Masui R & Tanaka I (2006) Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein* **15:** 1494–1499.

24. Ellinger P, Arslan Z, Wurm R, Tschapek B, MacKenzie C, Pfeffer K, Panjikar S, Wagner R, Schmitt L, Gohlke H, Pul Ü & Smits SHJ (2012) The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*. *J Struct Biol* **178:** 350–62

25. Fontaine L, Dandoy D, Boutry C, Delplace B, de Frahan MH, Fremaux C, Horvath P, Boyaval P & Hols P (2010) Development of a versatile procedure based on natural transformation for marker-free targeted genetic modification in *Streptococcus thermophilus*. *Appl Environ Microbiol* **76:** 7870–7

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

27. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD & Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker JM (ed) pp 571–607. Humana Press.

28. 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:** 688–92

29. Godde JS & Bickerton A (2006) The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol* **62:** 718–29

30. 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

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

32. Hale C, Kleppe K, Terns RM & Terns MP (2008) Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* **14:** 2572–9

33. Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover CVC, 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

34. 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:** 945–56

35. Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41:** 95–8.

36. Han D & Krauss G (2009) Characterization of the endonuclease SSO2001 from Sulfolobus solfataricus P2. *FEBS Lett* **583:** 771–6

37. 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:** 1928–32

38. 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

39. 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 USA* **108:** 21218–22

40. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K & Doudna JA (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329:** 1355–8

41. Haurwitz RE, Sternberg SH & Doudna JA (2012) Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *EMBO J* **31:** 2824–32

42. Hols P, Hancy F, Fontaine L, Grossiord B, Prozzi D, Leblond-Bourget N, Decaris B, Bolotin A, Delorme C, Dusko Ehrlich S, Guédon E, Monnet V, Renault P & Kleerebezem M (2005) New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. *FEMS Microbiol Rev* **29:** 435–63

43. Horvath P & Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327:** 167–70

44. Horvath P, Coûté-Monvoisin A-C, Romero D a D, Boyaval P, Fremaux C & Barrangou R (2009) Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int J Food Microbiol* **131:** 62–70

45. Horvath P, Romero DA, Coûté-Monvoisin A-C, 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:** 1401–12

46. Howard JAL, Delmas S, Ivančić-Baće I & Bolt EL (2011) Helicase dissociation and annealing of RNA-DNA hybrids by *Escherichia coli* Cas3 protein. *Biochem J* **439:** 85–95

47. Huang RH (2012) Cas Protein Cmr2 Full of Surprises. *Structure* **20:** 389–390.

48. Jansen R, Van Embden JDA, Gaastra W & Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43:** 1565–75

49. 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 APL, Dickman MJ, Doudna JA, Boekema EJ, Heck AJR, van der Oost J & Brouns SJJ (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **18:** 529–36

50. Kleanthous C, Kühlmann UC, Pommer a J, Ferguson N, Radford SE, Moore GR, James R & Hemmings a M (1999) Structural and mechanistic basis of immunity toward endonuclease colicins. *Nat Struct Biol* **6:** 243–52

51. Ko TP, Liao CC, Ku WY, Chak KF & Yuan HS (1999) The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure* **7:** 91–102

52. Koo Y, Jung D-K & Bae E (2012) Crystal Structure of Streptococcus pyogenes Csn2 Reveals Calcium-Dependent Conformational Changes in Its Tertiary and Quaternary Structure. *PLoS One* **7:** e33401

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

54. Labrie SJ, Samson JE & Moineau S (2010) Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **8:** 317–27

55. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ & Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947–8

56. Letunic I, Doerks T & Bork P (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res* **37:** D229–32

57. Lintner NG, Kerou M, Brumfield SK, Graham S, Liu H, Naismith JH, Sdano M, Peng N, She Q, Copié 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:** 21643–56

58. Lu M & Henning U (1994) Superinfection exclusion by. *Trends Microbiol* **2:** 137–9

59. 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

60. 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

61. 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

62. Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, Moineau S, Mojica FJM, Wolf YI, Yakunin AF, van der Oost J & Koonin EV (2011b) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9:** 467–77.

63. Marraffini LA & Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322:** 1843–5

64. Mojica FJM, Díez-Villaseñor C, García-Martínez J & Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155:** 733–40

65. Mojica FJM, Díez-Villaseñor C, García-Martínez J & Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60:** 174–82

66. Mulepati S & Bailey S (2011) Structural and biochemical analysis of the nuclease domain of the clustered regularly interspaced short palindromic repeat (CRISPR) associated protein 3(CAS3). *J Biol Chem* **3:** 1–18.

67. 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:** 22445–9

68. Nam KH, Kurinov I & Ke A (2011) Crystal Structure of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated Csn2 Protein

Revealed Ca2+-dependent Double-stranded DNA Binding Activity. *J Biol Chem* **286:** 30759–68

69. Nordström K & Forsgren A (1974) Effect of protein A on adsorption of bacteriophages to *Staphylococcus aureus*. *J Virol* **14:** 198–202
70. Nowotny M, Gaidamakov S a, Crouch RJ & Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121:** 1005–16
71. Papadopoulos JS & Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23:** 1073–9
72. Pingoud A, Fuxreiter M, Pingoud V & Wende W (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci* **62:** 685–707
73. Pingoud AM ed. (2004) Restriction Endonucleases. Springer, Berlin.
74. Portillo MC & Gonzalez JM (2009) CRISPR elements in the Thermococcales: evidence for associated horizontal gene transfer in *Pyrococcus furiosus*. *J Appl Genet* **50:** 421–30
75. 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:** 1495–512
76. Riede I & Eschbach ML (1986) Evidence that TraT interacts with OmpA of Escherichia coli. *FEBS lett* **205:** 241–5
77. Roberts RJ (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* **31:** 1805–1812
78. Sakamoto K, Agari Y, Agari K, Yokoyama S, Kuramitsu S & Shinkai A (2009) X-ray crystal structure of a CRISPR-associated RAMP module Cmr5 protein from Thermus thermophilus HB8. *Proteins* **75:** 528–32
79. Sambrook J, Fritsch E & Maniatis T (1989) Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor Lab Press, Cold Spring Harbor, NY.
80. Saravanan M, Bujnicki JM, Cymerman I a, Rao DN & Nagaraja V (2004) Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily. *Nucleic Acids Res* **32:** 6129–35
81. 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:** 680–7
82. Sashital DG, Wiedenheft B & Doudna JA (2012) Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol Cell* **46:** 606–15
83. Schneider CA, Rasband WS & Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9:** 671–675
84. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeño-Tárraga AM, Wang H, Holden MTG, Wright A, Churcher C, Quail M a, Baker S, Bason N, Brooks K, Chillingworth T, Cronin A, Davis P, *et al* (2006) The multidrug-resistant

human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nat Genet* **38:** 779–86.

85. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJJ & Severinov K (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci USA* **108:** 10098–103

86. Shen BW, Landthaler M, Shub D a & Stoddard BL (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J Mol Biol* **342:** 43–56

87. 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:** 1335–42

88. Snyder L (1995) Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol Microbiol* **15:** 415–20

89. Sorek R, Kunin V & Hugenholtz P (2008) CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6:** 181–6

90. Sternberg SH, Haurwitz RE & Doudna JA (2012) Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* **18:** 661–72

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

92. Sun X, Göhler A, Heller KJ & Neve H (2006) The ltp gene of temperate Streptococcus thermophilus phage TP-J34 confers superinfection exclusion to Streptococcus thermophilus and Lactococcus lactis. *Virology* **350:** 146–57

93. Swarts DC, Mosterd C, van Passel MWJ & Brouns SJJ (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7:** e35888

94. 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:** 4792–9

95. Wang R, Preamplume G, Terns MP, Terns RM & Li H (2011) Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure* **19:** 257–64

96. Westra ER, van Erp PBG, Künne T, Wong SP, Staals RHJ, Seegers CLC, Bollen S, Jore MM, Semenova E, Severinov K, de Vos WM, Dame RT, de Vries R, Brouns SJJ & van der Oost J (2012) CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Mol Cell* **46:** 595–605

97. Wiedenheft B, van Duijn E, Bultema J, Waghmare S, Zhou K, Barendregt A, Westphal W, Heck A, Boekema E, Dickman M & Doudna JA (2011a) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci USA* **108:** 10092–7

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

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

100. 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:** 904–12

101. Yang W (2011) Nucleases: diversity of structure, function and mechanism. *Q Rev Biophys* **44:** 1–93

102. Yanisch-Perron C, Vieira J & Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33:** 103–19

103. Yosef I, Goren MG & Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res* **40:** 5569–76

104. Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, Graham S, Reimann J, Cannone G, Liu H, Albers S-V, Naismith JH, Spagnolo L & White MF (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* **45:** 303–13

105. Zheng L, Baumann U & Reymond J-L (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* **32:** e115

106. Zhu X & Ye K (2012) Crystal structure of Cmr2 suggests a nucleotide cyclase-related enzyme in type III CRISPR-Cas systems. *FEBS Lett* **586:** 939–45