Skolkovo Institute of Science and Technology

# INVESTIGATION OF DNA-BINDING SPECIFICITY OF THE CAS1-CAS2 CRISPR ADAPTATION COMPLEX IN *ESCHERICHIA COLI*

*Doctoral Thesis*

by

OLGA MUSHAROVA

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisors
Professor Konstantin Severinov
Doctor Ekaterina Savitskaya

Moscow - 2017

## Abstract

It was long believed that bacteria have only non-specific defense systems, while the ability to "learn" to recognize specific pathogens by means of an adaptive immune system was an exclusive ability of higher eukaryotes. The discovery of the adaptive CRISPR-Cas immune system in prokaryotes therefore came as a surprise. The CRISPR-Cas system consists of a CRISPR DNA array of short repeated sequences separated by unique spacers derived from foreign DNA and CRISPR-associated (*cas*) genes that encode Cas proteins essential to the immune response. A complex of Cas proteins and crRNA (the product of CRISPR array transcription and processing) specifically recognizes "protospacers", which are DNA fragments complementary to the crRNA spacer. In one of the most studied CRISPR-Cas systems, found in *Escherichia coli*, the CRISPR machinery consists of crRNA in a complex with Cascade and Cas3 nuclease/helicase. After protospacer recognition the foreign DNA is destroyed in a process called "CRISPR interference". New spacers are acquired from the foreign DNA in a process referred to as "CRISPR adaptation". With every new spacer, an additional copy of the CRIPSR repeat is generated. Cas1 and Cas2, the most conserved proteins in CRISPR-Cas systems, are responsible for that activity. The presence of a preexisting spacer against foreign DNA in the CRISPR array strongly increases the efficiency of new spacer acquisition from the same DNA. This phenomenon is called "primed adaptation". Apart from Cas1-Cas2 complex, primed adaptation also requires Cas proteins involved in CRISPR interference. Details of the mechanism of primed CRISPR adaptation are not yet fully understood. In this study the DNA binding specificity of the *E. coli* Cas1-Cas2 adaptation complex during primed CRISPR adaptation was investigated *in vivo* using chromatin immunoprecipitation.

## Publications

The results were presented at scientific conferences and published in international peer-reviewed journals.

List of publications:

1. Musharova O, Klimuk E, Datsenko KA, Metlitskaya A, Logacheva M, Semenova E, Severinov K, Savitskaya E. Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res.* 2017 Apr 7;45(6):3297-3307.
2. Semenova E, Savitskaya E, Musharova O, Strotskaya A, Vorontsova D, Datsenko KA, Logacheva MD, Severinov K. Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc Natl Acad Sci U.S.A.* 2016 Jul 5;113(27):7626-7631.
3. Savitskaya EE, Musharova OS, Severinov KV. Diversity of CRISPR-Cas-mediated mechanisms of adaptive immunity in prokaryotes and their application in biotechnology. *Biochemistry (Mosc).* 2016 Jul;81(7):653-661.
4. Xue C, Seetharam AS, Musharova O, Severinov K, Brouns SJ, Severin AJ, Sashital D.G. CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.* 2015 Dec 15;43(22):10831-10847.

Conferences:

1. Musharova O., Medvedeva S., Severinov K., Savitskaya E. “CRISPR array spacers selection during primed adaptation in *Escherichia coli*”, Skoltech-MIT joint conference, Shaping the Future (Moscow, Russia, April 2017).
2. Olga Musharova “Non-double stranded spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation”, 1st Annual Workshop of Skoltech/MIT Next Generation Program 2016 (Moscow, Russia, October 2016).
3. O. Musharova, E. Savitskaya, K. Severinov “DNA-binding by the Cas1-Cas2 complex in *Escherichia coli* cells undergoing during primed adaptation”, CRISPR-2016 (Rehovot, Israel, May 2016).
4. Musharova O. S., Savitskaya E. E., Severinov K. V. “Investigation of DNA-binding specificity of Cas1-Cas2 CRISPR-adaptation complex in *Escherichia coli*”, SkoltechOn 2015 (Moscow, Russia, October 2015).

**Acknowledgements**

The author would like to express her thanks to Konstantin Severinov, Ekaterina Savitskaya, Evgeny Klimuk, Ekaterina Semenova, Kirill A. Datsenko, Maria Logacheva Anastasia Metlitskaya, Sofia Medvedeva and Ekaterina Rubtsova.

## Table of Contents

## List of Symbols, Abbreviations

Abi – Abortive Infection
ATP - Adenosine Triphosphate
BSA - Bovine Serum Albumin
ChIP – Chromatin Immunoprecipitation
CRISPR - Clustered Regular Interspaced Short Palindromic Repeats
CRISPR-Cas - CRISPR Associated
crRNA - CRISPR RNA
CS – Cold Spot
DBS – Double Strand Break
DNA - Deoxyribonucleic Acid
DND – DNA phosphorothioation system
dNTP - Deoxynucleotide
EDTA - Ethylenediaminetetraacetic acid
GTP - Guanosine Triphosphate
HDR – Homologous Recombination
HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEPN - Higher Eukaryotes and Prokaryotes Nucleotide
HS – Hot Spot
HTS - High Throughput Sequencing
IPTG - Isopropyl β-D-1-thiogalactopyranoside
LB - Luria-Bertani medium
MOPS -3-(N-morpholino)propanesulfonic acid
NHEJ - Non-Homologous Ends Joining
OD – Optical Density
PAM – Protospacer Adjacent Motif
PBS - Phosphate-buffered Saline
PCR – Polymerase Chain Reaction
qPCR – quantitative PCR
PMSF - Phenylmethylsulfonyl Fluoride
PNK- Polynucleotide Kinase
PolI – DNA polymerase I
R-M – Restriction-Modification
RNA - Ribonucleic Acid
SDS – Sodium dodecyl sulfate
sgRNA - single guide RNA
T-A – Toxin-Antitoxin
TAE - Tris-Acetate-EDTA buffer
TE - Tris- EDTA buffer
tracrRNA - trans-acting CRISPR RNA
Tris – Trisaminomethane

bp - Base Pairs
Da – Dalton
mM – Millimole
nm – Nanometere
ng – Nanogram
pM – Picomole
µl – Microlitre
µg – Microgramme
U - Enzyme unit

## List of Figures

## List of Tables

## Novelty And Practical Use

In this work, the DNA binding specificity of the Cas1-Cas2 complex during primed CRISPR adaptation by the *E. coli* type I-E CRISPR-Cas system was investigated *in vivo*. It was shown that the Cas1-Cas2 adaptation complex is localized in the leader region of the CRISPR array in cells undergoing primed adaptation. Also spacer-sized DNA fragments corresponding to “hot” protospacers that preferentially serve as a source of new spacers were found to be bound by the Cas1-Cas2 adaptation complex. These DNA fragments were shown to be not in the standard double-stranded form. It was shown that such fragments are cut from one strand of foreign DNA in a Cas1- and Cas3-dependent manner. It is suggested that they are the intermediates of CRISPR adaptation on their way to CRISPR array insertion. On this basis, a comprehensive model of the earliest stages of primed CRISPR adaptation is proposed.

## Personal Contribution

The greater part of this research was performed by the author. The Cas1-Cas2 adaptation complex *E. coli* was purified and used to obtain an anti – Cas1 polyclonal antibody. After purification, the anti-Cas1 antibody was used for chromatin immunoprecipitation. Precipitated DNA fragments were subjected to quantitative PCR. The localization of the Cas1-Cas2 adaptation complex during primed CRISPR adaptation was thus investigated. The structure and source of fragments associated with the Cas1-Cas2 complex were analyzed.

# Chapter 1. Review of Literature

## 1.1 Prokaryotic Defense Systems

Prokaryotes have been constantly threatened by genetic invaders such as bacteriophages, plasmids, and transposons, and yet they survive and flourish. An ongoing arms race between prokaryotic hosts and invaders is one of the key drivers of evolution (Koonin and Zhang, 2017). A variety of defense strategies is the result of this long-term stand-off (Fig. 1).

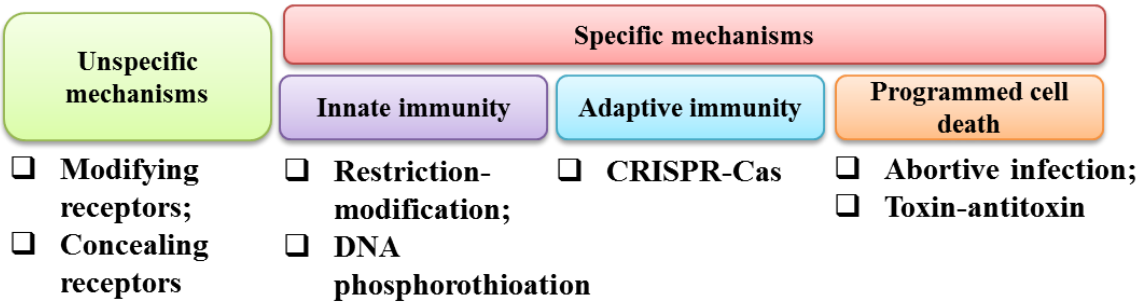



**Fig.1. Overview of prokaryotic defense system mechanisms**

Unspecific mechanisms prevent phage adsorption by the modification of receptor structure through mutation and/or concealing of receptors by physical barriers. Besides unspecific protection mechanisms, prokaryotes have evolved a variety of specific defense mechanisms, including the mechanisms of cell immunity and programmed cell death.

Prokaryotic immune mechanisms are able to distinguish “self” from “non-self” DNA molecules, and can be divided into innate and adaptive immunity systems. Systems of innate immunity in prokaryotes are represented by DNA modification systems, which are subdivided into two types: restriction-modification (R-M) and DNA phosphorothioation (DND). Innate immunity utilizes two types of enzymes, one of which modifies host DNA, while the other physically destroys non-modified invader DNA. Adaptive immunity in prokaryotes is represented by CRISPR-Cas systems, which are able to “memorize” invader agents and specifically destroy them upon re-infection.

All these strategies may allow individual cells to survive attack by an invader. If they fail, the cell can activate programmed cell death – an altruistic suicide that prevents invader expansion and thus protects the rest of the population. Programmed

cell death can be mediated by abortive infection systems (Abi) and toxin-antitoxin systems (T-A). Abi systems are mechanistically diverse, and are often encoded by mobile genetic elements, including prophages and plasmids (Samson *et al.*, 2013). For *Lactococcus lactis* the number of Abis exceeds 20 (Seed, 2015). Abi systems act at all stages of viral infection. Some of them, such as AbiP, act at an early stage of phage replication (Domingues *et al.*, 2004), while others, such as AbiZ, induce premature lysis of infected cells (Durmaz and Klaenhammer, 2007). Toxin-antitoxin systems are genetic elements found on plasmids or chromosomes of various bacteria and archaea. Under normal growth conditions, the activity of the toxin protein or its translation is counteracted by an antitoxin protein or noncoding RNA, but in certain circumstances the amount of the antitoxin is reduced, leading to cell death from the toxin action.

Specific defense mechanisms will be described in detail below.

### *1.1.1 DNA Modification Systems And Innate Immunity*

Innate immunity in prokaryotes functions on the self–non-self discrimination principle, where DNA is the object of the discriminatory recognition. The DNA molecule in the host cell undergoes modifications, such that the cell defense system subsequently recognizes it as "self". Foreign DNA does not contain modifications, and therefore triggers the immune response. Two types of defense systems based on two different types of DNA modifications represent innate immunity in prokaryotes: the restriction-modification (R-M) system and the DNA phosphorothioation (DND) system.

The restriction-modification system for destroying invader DNA is one of the most widespread initial immunity strategies in prokaryotes. Restriction-modification systems are believed to function in prokaryotes in a way that is formally similar to eukaryotic innate immune systems fucntion, which also distinguish self- from non-self agents.

Restriction-modification systems are widespread among both eubacteria and archaea. They are formed by two types of enzymes: a restriction endonuclease and a methyl-transferase. The restriction endonuclease recognizes and cleaves the unmethylated DNA sequence at a specific motif. The methyltransferase activity provides discrimination between self and non-self DNA by transferring methyl groups to the same DNA motif within the host genome. Endonuclease cleavage occurs at phosphodiester bonds and as a result 5′ or 3′ overhangs or blunt ends are generated. Methyltransferase transfers the methyl group from S-adenosyl methionine to the C-5 carbon or the N4 amino group of cytosine or to the N6 amino group of adenine (Vasu and Nagaraja, 2013). R-M systems are highly variable and, to date, nearly 4000 systems have been described. Restriction endonucleases are mainly responsible for this diversity (Vasu and Nagaraja, 2013).

The following factors are taken into consideration in classification of R-M systems: 1. subunit composition; 2. recognition site; 3. mode of cleavage; 4. involvement of co-factors. All known R-M systems can be divided into four types (Roberts *et al.*, 2003). Systems of Type I operate as a single enzyme consisting of five subunits. Restriction endonuclease cleaves a phosphodiester bond in the DNA molecule and has helicase activity. A restriction site is represented by a bipartite sequence: 3-5 specific nucleotides at the 5'-end, a 6-8 nucleotide stretch of random nucleotides

followed by a specific sequence of 4-5 nucleotides at the 3’-end. Cleavage of the DNA molecule occurs at considerable distances (up to thousand bp) from the recognition site, and ATP hydrolysis is needed for this process. Methyltransferase methylates adenine residues in restriction sites. The type I R-M systems are divided into four families (I A-D) (Murray, 2000).

In type II R-M systems, methyltransferase and endonuclease activities reside in independent proteins. Endonuclease cuts DNA at strictly defined positions within the recognition site. DNA ends resulting from cleavage have 3'-hydroxyl and 5'-phosphate ends. Because of these properties, restriction endonucleases of type II systems are widely used in genetic engineering. Type II restriction endonucleases do not require ATP or any other energy source for target cleavage. Type II methyltransferases methylate adenine as well as cytosine residues of restriction sites. The R-M systems of type II are subdivided into several families (Pingoud *et al.*, 2005).

The R-M systems of type III include two subunits (Res and Mod) with endonuclease and methyltransferase activities combined in a heterotetramer (Res2Mod2). A Res-subunit has helicase activity and requires ATP hydrolysis in order to function. The interaction of two heterotetrameric enzymes is required for hydrolysis of DNA. Each heterotetramer recognizes identical restriction sites located in opposite orientations. Unlike most other known R-M systems, only one strand of DNA in the recognition site is methylated (Dryden al., 2001).

The R-M systems of type IV cleave only modified DNA containing methylated, glycosyl-hydroxymethylated or hydroxymethylated bases in the recognition sites. GTP hydrolysis is required for DNA cleavage. Methyltransferase activity is absent from such systems. The endonuclease activity is positively affected by S-adenosyl-methionine, but ATP has no influence on activity of the enzymes. The asymmetric restriction site is composed of two separate parts. DNA cutting occurs at one of the sites (Roberts *et al.*, 2003).

Restriction was first demonstrated in 1952 by Salvatore Luria and Mary Human (Luria and Human, 1952) in experiments with phage. The phage λ (propagated in *E. coli* B) was found to grow poorly on *E. coli* K-12. Immunity is achieved by cleavage of foreign (phage) DNA, which is unmethylated, while the genome of the host remains protected due to methylation by the cognate methyltransferases. The

effectiveness of R-M systems in host protection against phage infection has been demonstrated in various studies to vary from 10 to $10^8$-fold (Tock and Dryden, 2005). The defense function of R-M systems is also proved by the fact that bacteriophages and other mobile genetic elements have evolved a variety of strategies to avoid restriction (Vasu and Nagaraja, 2013). The most studied anti R-M strategies are as follows:

- Bacteriophages may encode their own methyltransferases. Phage methyltransferases have broad specificity and can simultaneously protect the bacteriophage genome from the action of several host endonucleases. For instance, phages of *Bacillus subtilis* φ3T, δl1, SPR and SPβ may avoid restriction by methylation of phage DNA at various sequences (Warren, 1980);
- The bacteriophage masks the restriction site by the addition of bulky groups to nitrogenous bases, e.g., hydroxymethylation, glycosylation, addition of acetamide group. The modified DNA molecule becomes resistant to the action of restriction endonucleases. The *mom* gene of bacteriophage Mu encodes a protein that catalytically transfers an acetamide group to the N-6 position of adenine in the sequence context 5′-G/C-A-G/C-N-C/T-3′. This modification provides resistance against a host restriction endonucleases (Hattman, 1999);
- Phages and plasmids may contain proteins that block restriction enzymes, such as the OCR ("overcome classical restriction") protein of T7 group phages (Walkinshaw *et al.*, 2002);
- Bacteriophages undergo an internal selection for reduction of the number of restriction sites in their genomes. The distribution of the EcoP1I restriction sites in bacteriophage T7 genome is one example of such a strategy. For efficient cleavage of the DNA endonuclease EcoP1I needs two copies of its motif oriented head to head. In the genome of bacteriophage T7 all EcoP1I motifs are located in the same orientation, thus preventing cleavage.

The site-specific DNA backbone S-modification and cleavage of unmodified DNA were first discovered in *Streptomyces lividans* 1326. The S-modification of DNA is represented by substitution of one oxygen atom in the DNA sugar-phosphate backbone with sulfur. DNA phosphorothiolation occurs naturally in a variety of bacteria and archaebacteria and results from activity of Dnd proteins. The DND system is represented in the bacterial genome by a cluster containing eight genes (*dndB-H*) (He *et*

*al.*, 2015). Products of *dndB-E* genes are responsible for modifications of DNA molecules, and *dndF-H* genes code proteins, which cleave and destroy unmodified DNA. Despite progress in defining the biochemistry of S-modification, its function *in vivo* remains obscure.

*1.1.2 Programmed Cell Death*

The scenario of programmed cell death can be realized in different ways, including the abortive infection (Abi) mechanism and T-A systems. Abi includes very diverse mechanisms and can be mediated by T-A systems.

Abi, which is most characteristic of lactic acid bacteria, prevents virus infection at various stages by causing the death of the infected cell (Labrie *et al.*, 2010). This altruistic suicide spares non-infected bacterial cells from phage infection, thus preserving stability of the population. Abi systems typically consist of a single protein or protein complex and are often encoded by mobile genetic elements, such as prophages and plasmids, which probably favors their dissemination. Initially, Abi systems were found in *Lactococcus lactis*, and based on diverse phenotypic and molecular effects were classified from A to K (Chopin *et al.*, 2005). Later the existence of Abi was shown in *E. coli* (Labrie *et al.*, 2010). Their exact mode of action is unknown in most cases, although phages that by-pass these systems have been characterized. In particular, mutant escape phages were found for the AbiD1, AbiK and AbiV systems (Bidnenko *et al.*; 2009, Wang *et al.*, 2011; Haaber *et al.*, 2009; Labrie *et al.*, 2012).

According to bioinformatics analyses, the proteins AbiD, AbiF, AbiJ, AbiU2, AbiV and the C-terminal domain of AbiA belong to the HEPN endoribonucleases and are predicted to target the translation system (Makarova *et al.*, 2013). AbiI was predicted as ribonuclease and is apparently involved in translation regulation (Makarova *et al.*, 2013). Ability to interact with R-M systems has been shown for AbiU1, AbiL and AbiR (Makarova *et al.*, 2011; Makarova *et al.*, 2013). The AbiA and AbiK proteins contain the reverse transcriptase domain (Kojima and Kanehisa, 2008; Makarova *et al.*, 2013). As was shown, AbiK catalyses non-template DNA synthesis. The DNA strand remains covalently attached to the protein and contributes to Abi (Makarova *et al.*, 2013).

The abortive infection mechanisms are variable and act at all stages of phage development, leading to decline in the generation of phage progeny. For example, the two-component RexA-B system, which is found in λ phage, induces a loss of membrane potential followed by the decreasing of ATP level, thus protecting infected cells from multi-infection (Snyder, 1995). Another example is AbiP, which acts at an initial stage

of bacteriophages replication, and involves in the disruption of phage DNA replication process as well as temporal transcription (Domingues *et al.*, 2004). The AbiZ activation leads to premature lysis of infected cells thus preventing bacteriophage assembly and the release (Durmaz and Klaenhammer, 2007).

The well-studied Lit system from *E. coli* belongs to Abi and is encoded by the defective prophage e14. Activation of this system leads to a general inhibition of translation during phage T4 infection (Snyder, 1995). The peptide Gol, derived from the capsid of T4 and bound with the ribosomal elongation factor Tu (EF-Tu), activates the Lit system protease (Bergsland *et al.*, 1990; Bingham *et al.*, 2000). Lit protease cleaves EF-Tu leading to cell death and abortion of infection. Derivatives of T4 carrying a mutation in the *gol* gene cannot activate the Lit system and are capable of replication in Lit-positive *E. coli* strains (Samson *et al.*, 2013).

Some Abi mechanisms mediate their activity by T-A system principles (Fineran *et al.*, 2009; Makarova *et al.*, 2013). It is common to all known T-A systems that they encode both a stable toxin protein that can cause cell death or cell arrest and also a RNA or protein of an unstable antitoxin that can preclude toxicity (Gerdes *et al.*, 2005). T-A systems of bacteria and archaea are represented by a set of two or more genes, which are usually localized on plasmids (Gerdes *et al.*, 1986; Ogura and Hiraga, 1983), but chromosomal T-A systems have also been described (Gerdes *et al.*, 2005). For example, the AbiE system, which is encoded by bicistronic operons, can prevent phage proliferation by inducing bacteriostasis via a type IV T-A mechanism (Dy *et al.*, 2014; Makarova *et al.*, 2013).

The permanent arms race between bacteriophages and their hosts is the main evolutional driver of these groups of organisms. Despite very sophisticated bacterial defense systems such as R-M, T-A or Abi, phages have developed many means of escape, making even very complex bacterial systems ineffective. The broad diversity of defense mechanisms has in turn evolved in response to this phage mutability. The only systems that represent an inherited ability to memorize a particular invader and effectively destroy it in case of re-infection and, what is more, to adjust to its escape mutations are CRISPR-Cas systems. They will be discussed in the following part of this literature review.

## Chapter 2. CRISPR-Cas Systems And Adaptive Immunity Of Prokaryotes

The existence of previously unknown defense mechanisms, christened CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated genes), has been discovered in recent decades. The mechanism of their action is specific and adaptive to particular invaders. The CRISPR-Cas system is represented in the bacterial genome by a CRISPR array (the number of arrays in the bacterial genome may vary from one to several dozens) and associated *cas* genes. A CRISPR array is a cluster of short repeated genomic DNA fragments separated by unique spacer sequences, some of which originate from foreign DNA. An AT-rich leader region is located prior to the CRISPR array (Jansen *et al*., 2002). The *cas* genes encode protein components of the CRISPR-Cas mechanism. CRISPR-Cas systems are responsible for two different cellular processes: CRISPR adaptation and CRISPR interference. The process of new spacer integration into CRISPR array is called CRISPR adaptation. During CRISPR adaptation the CRISPR array is elongated by one new spacer and one repeat. The array is then transcribed with the formation of pre-crRNA that is processed into short protective crRNAs, so that each crRNA contains a spacer flanked by partial repeats. crRNA that has been bound by Cas proteins forms a CRISPR interference (also called “effector”) complex, which enables specific recognition of a protospacer i.e., a target DNA sequence complementary to the spacer of the crRNA, followed by degradation of the target DNA molecule (Fig. 2). This process is called CRISPR interference.

The CRISPR-Cas system displays the following major features shared with eukaryotic immunity: i) the ability to discriminate between its own and foreign components; ii) specificity (CRISPR-Cas systems are able to recognize specific features of pathogens, like T and B cells in mammals); iii) memory (CRISPR-Cas systems are able to memorize invaders, which have been presented in the past).

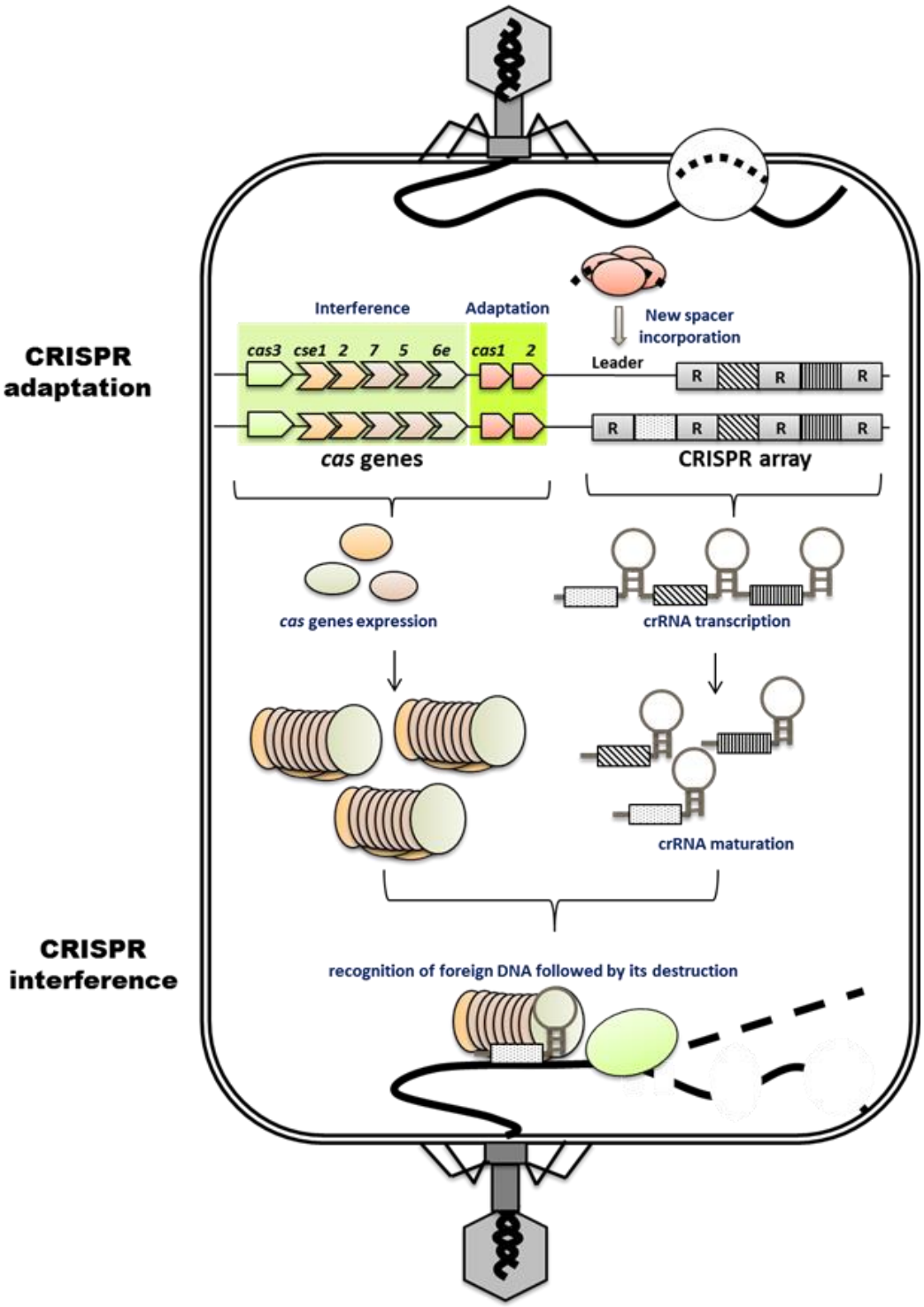



**Fig. 2. Mechanism of CRISPR-Cas adaptive immunity. After entering a bacterial cell, fragments of foreign DNA integrate into a CRISPR array in the process of CRISPR adaptation. A CRISPR array is elongated by one new spacer and one repeat. The CRISPR array is then transcribed with the formation of pre-crRNA that is processed into short crRNAs, so that each crRNA contains a spacer flanked**

**by partial repeats. The *cas* genes code for protein components of the CRISPR interference and CRISPR adaptation complexes. A CRISPR interference complex that includes crRNA and Cas proteins interacts with a protospacer, i.e. a target DNA sequence complementary to the sequence of the crRNA spacer, and this interaction leads to the degradation of the target DNA molecule.**

An unusual *E. coli* genomic locus containing alternating short repeats separated by unique sequences was first described in 1987 (Ishino *et al.*, 1987). Subsequently the presence of a CRISPR array in *Tubercle bacillus* (Van Soolingen *et al.*, 1993) and halophilic archaea (Mojica *et al.*, 2005) was demonstrated by means of bioinformatics analysis. Thus, by 2000 CRISPR arrays had been found in the genomes of many bacteria and archaea. Genes specifically associated with CRISPR arrays (*cas* genes) were described in 2002 by Jansen (Jansen *et al.*, 2002).

A function of CRISPR-Cas systems for prokaryotes was first proposed in 2005, when researchers noted that some of the spacers in the CRISPR array corresponded to nucleotide fragments from plasmids and bacteriophage genomes (Mojica *et al.*, 2005; Pourcel *et al.*, 2005; Bolotin *et al.*, 2005). These observations implied the involvement of CRISPR-Cas systems in prokaryotic defense mechanisms. In 2007 the process of CRISPR adaptation was discovered by Barrangou and co-workers. *Streptococcus thermophilus* cells infected with a bacteriophage were shown to acquire resistance to repeated infection by the same bacteriophage due to the insertion of new spacers complementary to fragments of the phage genome into their CRISPR arrays (Barrangou et al, 2007). In 2008 it was shown that the presence of *cas* genes and a spacer of the CRISPR array, which is complementary to the plasmid protospacer, prevented plasmid conjugation (Marraffini and Sontheimer, 2008). The target DNA was destroyed by RNA–protein complexes composed of the *cas* gene products and short crRNAs formed after CRISPR array transcript processing (Brouns *et al.*, 2008). Thus the phenomenon of CRISPR interference was discovered.

*2.1 CRISPR Interference Mechanisms*

CRISPR-Cas systems have been found in more than 40% of eubacteria and in most of archaea (Sorek *et al.*, 2008; Makarova *et al.*, 2011). The growing list of discovered CRISPR-Cas systems enabled researchers to compare and classify them. Since the proteins responsible for CRISPR adaptation are homologous in all CRISPR-Cas systems, this classification is based mostly on the protein composition of the complexes involved in CRISPR interference. According to the latest data, CRISPR-Cas systems can be subdivided into two classes, six types, and 19 subtypes (Makarova *et al.*, 2015; Shmakov *et al.*, 2015). The two classes are distinguished based on composition of the interference complexes: they consist of multiple subunits in Class 1 (types I, III, IV), while they contain only one subunit in Class 2 (types II, V and VI) (Table 1). The mechanisms of crRNA maturation and the proteins involved in CRISPR interference vary substantially. No possible evolutionary relationship between components of the CRISPR interference of Class 1 and Class 2 has been found (Makarova *et al.*, 2015). Koonin and Krupovic suggested that modular composition of CRISPR-Cas systems allowed to combine different unrelated interference modules with an adaptation module as a plausible way of CRISPR-Cas evolution (Koonin and Krupovic, 2015). Classes are subdivided into types, each characterized by a specific interference complex and mechanism. Specific organization of the interference modules of different CRISPR-Cas systems will be discussed below.

**Table 1. Classification of CRISPR-Cas systems**

| | Type | Typical protein | Pre-crRNA processing | Target | Self vs. non-self discrimination | Effectors |
|---|---|---|---|---|---|---|
| **Class I** | Type I | Cas3 | Cas6 | DNA | PAM | Cascade crRNA Cas3 |
| | Type III | Cas10 | Cas6+unknown factor | DNA RNA | CRISPR repeat | Cmr/Csm crRNA Cas10 |
| | Type IV | Csf1 | ? | ? | ? | ? |
| **Class II** | Type II | Cas9 | Rnase III | DNA | PAM | Cas9 crRNA tracrRNA |
| | Type V | Cpf1 | Cpf1 | DNA | PAM | Cpf1 crRNA tracrRNA (in some cases) |
| | Type VI | C2c2 | ? | RNA | ? | C2c2 crRNA |

*2.1.1 Class 1 CRISPR-Cas systems*

*Type I.* Type I CRISPR-Cas systems are characterized by the presence of a Cas3 protein that contains the core SF2 family helicase domain with RecA motifs and the N-terminal HD nuclease domain (Jackson *et al.*, 2014) (Table 1). They also contain the multi-subunit crRNA-binding Cascade complex (Makarova *et al.*, 2015) responsible for target recognition during CRISPR interference. Type I includes the most studied CRISPR-Cas I-E subtype system from *E. coli*. The 405 kDa Cascade complex from *E. coli* is composed of the Cse1, Cse2, Cas5, Cas7, and Cas6e proteins in 1:2:1:6:1 ratio (Brouns *et al.*, 2008; Jore *et al.*, 2011). The crRNA is formed by processing of the CRISPR array primary transcript (pre-crRNA) by ribonuclease activity of the Cas6e protein (Brouns *et al.*, 2008). The Cas6e protein is dispensable if another source of mature crRNAs exists (Semenova *et al.*, 2015). In the the subtype I-A systems the Cas6 is not a stable constituent of the Cascade complex, and this fact causes considerable variations in the length of the crRNA, whose termini remain unprotected (Lintner *et al.*, 2011). The crRNA-containing Cascade complex recognizes the protospacer. Complementarity between the crRNA spacer and protospacer is more critical in the several nucleotides called the “seed” region (Semenova *et al.*, 2011), while many

mismatches are tolerated in the other nucleotide positions (Fineran *et al.*, 2014). DNA of the protospacer then unwinds with the formation of an R-loop, i.e. the heteroduplex between the crRNA spacer and the complementary target DNA strand and the displaced single-stranded "non-target" DNA strands The next stage requires the executore Cas3 protein (Westra *et al.*, 2012; Hochstrasser *et al.*, 2014), which introduces breaks into the displaced DNA strand of the R-loop, thereby initiating target degradation (Mulepati and Bailey, 2013; Sinkunas *et al.*, 2011). The 3D structures of some Cas3 proteins and their complexes with short single-stranded DNA fragments, presumably DNA degradation products, have been determined (Huo *et al.*, 2014; Zhao *et al.*, 2014; Gong *et al.*, 2014). The detailed mechanism of target degradation by Type I CRISPR-Cas systems is not fully understood.

In type I systems, prevention of an autoimmune response, i.e., discrimination between the protospacer in the target DNA and the CRISPR array spacer, occurs by the following mechanism. Cse1, one of the Cascade complex components, recognizes a short motif PAM that precedes the protospacer (Sashital *et al.*, 2012). Such recognition is an obligatory condition for the interaction with the target DNA (Westra *et al.*, 2012; Sashital *et al.*, 2012). It allows to prevent autoimmune response against the CRISPR array spacer, because the repeat sequence differs from the PAM sequence. On the other hand, the PAM requirement narrows the possibility for target recognition. Thus, mutations in PAMs adjacent to protospacers completely matching crRNA spacers allow bacteriophages to avoid CRISPR interference (Andersson and Banfield, 2008; Deveau *et al.*, 2008; Paez-espino *et al.*, 2015). Mutations in the PAM or seed sequence of a protospacer (Semenova *et al.*, 2011; Deveau *et al.*, 2008; Fineran *et al.*, 2014) reduce the binding affinity of Cascade (Westra *et al.*, 2012; Semenova *et al.*, 2011) and affect the recruitment of the Cas3 nuclease and its cleavage activity (Hochstrasser *et al.*, 2014) in type I-E systems, and resulting escape of interference.

*Type III.* Type III systems are characterized by the presence of the Cas10 protein. Cas10 contains a Palm domain that is similar to the RNA-recognizing domains of polymerases (Table 1). There are two families of multi-subunit interference complexes – Cas10–Csm and Cas10–Cmr – that are typical for Subtype III-A and III-B systems, respectively (Makarova *et al.*, 2015). Type III CRISPR-Cas systems are active against both DNA (Marraffini and Sontheimer, 2008; Garrett *et al.*, 2015; Goldberg *et*

*al.*, 2014; Samai *et al.*, 2015) and RNA (Samai *et al.*, 2015; Hale *et al.*, 2009; Staals *et al.*, 2013; Tamulaitis *et al.*, 2014; Zebec *et al.*, 2014; Zhang *et al.*, 2012): they first recognize and cleave a protospacer-containing RNA transcript and then the corresponding template DNA (Samai *et al.*, 2015). Recognition of RNA is mediated by the Cas7 family proteins, such as Csm3 in subtype III-A and Cmr4 in subtype III-B (Kazlauskiene *et al.*, 2016). Cas10 is responsible for the DNase activity (Samai *et al.*, 2015), while Csm3 and Cmr4 display ribonuclease activity (Samai *et al.* 2015; Staals *et al.*, 2013; Tamulaitis *et al.*, 2014). Active transcription of the target sequence is an obligatory condition for DNA CRISPR interference mediated by type III systems (Goldberg *et al.*, 2014; Deng *et al.*, 2013). This requirement allows to avoid CRISPR interference against prophages that can result in host genome degradation and cell death (Goldberg *et al.*, 2014). The crRNA maturation in type III systems is not completely understood. It is known that the Cas6 ribonuclease, which is not a component of the interference complex, cleaves the CRISPR array primary transcript into intermediates composed of individual spacers flanked by repeat fragments on both ends. The last eight nucleotides of the 5'-flanking repeat were named a "crRNA tag" (Carte *et al.*, 2008). Partially matured crRNA is incorporated by an unknown mechanism into the Csm or Cmr complexes, where its 3'-terminus undergoes additional processing (Hatoum-Aslan *et al.*, 2013).

Unlike type I systems, in which avoidance of autoimmune response depends on PAM recognition, avoidance of autoimmunity in type III systems is determined by complementary interactions between the crRNA tag and the target sequence. Full complementarity of the crRNA tag and a fragment of the CRISPR array repeat prevents CRISPR interference. In all other cases, CRISPR interference is possible (Zebec *et al.*, 2014; Marraffini and Sontheimer, 2010).

*Type IV.* The existence of Type IV CRISPR-Cas systems was predicted through bioinformatics searches (Makarova *et al.*, 2015). Their characteristic feature is the presence of the *csf1* gene (See Table 1) usually associated with *cas5* and *cas7* genes. The presence of the *cas5* gene encoding for Csf3 protein was shown for *Acidithiobacillus ferrooxidans* ATCC 23270, *Azoarcus sp.* (strain EbN1), and *Rhodoferax ferrireducens* (strain DSM 15236/ATCC BAA-621/T118). In the latter two species, the CRISPR-Cas locus was found on a plasmid. The occurrence of *cas* genes

without CRISPR arrays suggests that they have functions other than adaptive immunity. Clearly, the mechanisms of action and the functions of Type IV CRISPR-Cas systems require further study and experimental verification.

### *2.1.2 Class 2 CRISPR-Cas systems*

*Type II.* The interference complex of type II CRISPR-Cas systems includes the Cas9 protein and two RNAs – crRNA and trans-encoded small RNA (tracrRNA), the latter containing a sequence complementary to the CRISPR repeat (Deltcheva *et al.*, 2011) (Table 1). The tracrRNA pairs with the pre-crRNA and directs its cleavage by RNase III to produce mature crRNA (Deltcheva *et al.*, 2011). Similarly to type I systems, the interaction of crRNA with the target protospacers requires the presence of a PAM (Anders *et al.*, 2014; Szczelkun *et al.*, 2014; Sternberg *et al.*, 2014). Recognition of the PAM and the protospacer, local unwinding of the DNA target, R-loop formation, and introduction of breaks into target DNA are provided by the activity of Cas9 in a complex with the two above-mentioned RNAs. Cas9 has two domains – HNH and RuvCm (Makarova *et al.*, 2006). Each of these domains cleaves one of the protospacer strands in the R-loop. The breaks in both chains are located across from each other, so that the cleavage produces a DNA molecule with blunt ends (Barrangou *et al.*, 2007; Deltcheva *et al.*, 2011; Garneau *et al.*, 2010; Sapranauskas *et al.*, 2011; Gasiunas *et al.*, 2012; Jinek *et al.*, 2012). The simplicity of the CRISPR interference mechanism in the type II systems and the possibility of combining crRNA and tracrRNA into a single guiding RNA offered a basis for rapid development of Cas9-mediated techniques for eukaryotic genome editing (Jiang and Marraffini, 2015).

*Type V.* The typical feature of type V CRISPR-Cas systems is the presence of the Cpf1 protein and its homologs (Table 1). Similar to the Cas9 protein from the type II system, Cpf1 contains a RuvC-like nuclease domain. However, Cpf1 does not contain a HNH nuclease domain (Makarova *et al.*, 2015; Schunder *et al.*, 2013). The functional activity of some type V CRISPR-Cas systems has only recently been experimentally confirmed (Zetsche *et al.*, 2015; Shmakov *et al.*, 2015). It appears that, similarly to type II systems, type V systems require a PAM for target recognition. In some type V systems, crRNA maturation does not require tracrRNA and RNase III and is presumably catalyzed by the Cpf1 protein itself. Cleavage of two target DNA strands occurs with a

shift, which results in the formation of “sticky” single-stranded ends that are 4-5 nucleotides long (Zetsche *et al.*, 2015). The possibility of crRNA-guided genome editing has been demonstrated for at least two Cpf1 proteins – from *Acidaminococcus sp.* BV3L6 and *Lachnospiraceae bacterium* ND2006 (Zetsche *et al.*, 2015).

*Type VI.* The existence of the type VI CRISPR-Cas system was predicted in 2015 as a result of bioinformatics searches (Shmakov *et al.*, 2015). The effector protein C2c2 from subtype VI-A CRISPR-Cas system was described in 2016. Similar to other CRISPR-Cas systems from class II, the CRISPR locus in the type VI system has a simplified structure. In fact, the CRISPR locus of the type VI system from *Leptotrichia shahii* contains just *cas1*, *cas2*, *c2c2* genes and a CRISPR array. The effector complex contains crRNA bound to C2c2 nuclease, and is able to cleave single-stranded RNA molecules. In contrast to all known CRISPR nucleases, C2c2 nuclease mediates RNA cleavage by a HEPN domain. Mutation in the catalytic center of the NEPN domain leads to inactivation of the effector complex, but RNA-binding activity of the resulting protein is retained (Abudayyeh *et al.*, 2016). Because of its ability to bind RNA molecules in a predetermined manner, C2c2 nuclease can be used as an effective tool for RNA editing and regulation of gene expression.

## 2.2 CRISPR Adaptation And Immune Memory

New spacers are introduced into CRISPR arrays in a process termed “CRISPR adaptation”. CRISPR adaptation was first demonstrated in 2007 in experiments with a II-A CRISPR-Cas system in *Streptococcus thermophilus* (Barrangou *et al.*, 2007). The term was suggested two years later by Van der Oost (Van der Oost *et al.*, 2009). During the last 10 years, the *de novo* spacer acquisition process was observed in different bacteria and archaea: *Pseudomonas aeruginosa* type I-F, *Escherichia coli* type I-E, *Sulfolobus solfataricus* type I-A and III-B (Erdmann and Garrett, 2012), *Haloarcula hispanica* type I-B (Li *et al.*, 2014), *Streptococcus agalactiae* type II-A.

Spacer acquisition requires Cas1 and Cas2, which are the most conserved protein components of all CRISPR-Cas systems (Koonin and Makarova, 2013). Many specific features have been described for different CRISPR-Cas systems, including the process of primed adaptation of type I CRISPR-Cas systems which allows fine adjustment to follow and counter infections by escape phage.

### *2.2.1 Spacer Acquisition Mechanism*

The acquisition of new spacers predominantly occurs at the promoter-proximal side of CRISPR array (Wei *et al.*, 2015; Yosef *et al.*, 2012). Thus, spacers are incorporated into the CRISPR array in a chronological order, and spacers most distant from the leader region can be considered to be the most ancient, while the set of spacers present in a given bacterium can be regarded as a record of encounters of its ancestors with genetic invaders. Because the length of the CRISPR array is limited, most of the earlier acquired spacers tend to be lost, for example, by recombination between repeats (Andersson and Banfield, 2008).

Incorporation of a new spacer is accommodated by duplication of the CRISPR repeat (Andersson and Banfield, 2008; Barrangou *et al.*, 2007; Horvath *et al.*, 2008; Tyson and Banfield, 2008; Pride *et al.*, 2012; Yosef *et al.*, 2012). In type I-E systems the first repeat in the CRISPR array is used as template for the new repeat (29 nt long) amplification. It has been proved by doing experiments with repeat sequences containing point mutation. Being incorporated into the first repeat, point mutation nad been replicated in newly repeat. Since mutations of the last nucleotide of the repeat are not passed on to new repeats, thus only bases 1 through 28 of the repeat in the array are

used as a template for new one. The last 29th base of newly apmlified repeat is originated from the protospacer and includes the last nucleotide of the PAM sequence (Datsenko *et al.*, 2012; Goren *et al.*, 2012). Recently, new details of spacer integration into CRISPR arrays have been revealed in experiments performed *in vivo* (Arslan *et al.*, 2014) and *in vitro* (Nuñez *et al.*, 2014; Nuñez *et al.*, 2015). It was found that the Cas1–Cas2 complex introduces a single-strand break exactly at the leader–repeat junction and catalyzes the nucleophilic attack by the 3'-OH end of the incoming spacer onto the 5'-end of the first repeat. Similarly, the other strand is nicked at the first repeat–spacer junction and the 5'-end of the repeat strand is joined to the 3'-end of the new spacer. As a result, the incorporated spacer is flanked by the single-stranded repeat sequences that are filled later. Similar intermediates are known for transposase-mediated mobile element integration, suggesting that spacer acquisition and transposon integration reactions are mechanistically similar (Mizuuchi and Adzuma, 1991; Rollie *et al.*, 2015).

The Cas1 and Cas2 proteins are the key participants in the CRISPR adaption process, and they are also the most evolutionarily conserved components of all CRISPR-Cas systems (Koonin and Makarova, 2013). As a rule, *cas1* and *cas2* genes are co-located, and the encoded proteins form a stable complex (Nuñez *et al.*, 2014; Nuñez *et al.*, 2015). Deletion of *cas1* and *cas2* genes has no effect on crRNA maturation and interference and in systems of type I (Brouns *et al.*, 2008) and type II (Deltcheva *et al.*, 2011; Sapranauskas *et al.*, 2011) and type III (Hatoum-Aslan *et al.*, 2014). Cas1 is an endonuclease (Wiedenheft *et al.*, 2009; Babu *et al.*, 2011). Cas1 has shown to be able for solving Holliday junctions with subsequent promotion of DNA integration and recombination *in vitro* (Beloglazova *et al.*, 2015). Cas2 displays nuclease activity toward both RNA and DNA *in vitro* (Nam *et al.*, 2012; Samai *et al.*, 2010). However, CRISPR adaptation *in vivo* requires only Cas1 nuclease activity (Nuñez *et al.*, 2014). Cas1 nuclease activity is not sufficient for spacer acquisition: ability to assemble with Cas2 is also essential. Mutations disrupting complex formation *in vitro* interfere with spacer acquisition *in vivo* (Nuñez *et al.*, 2014). Moreover, this functional requirement is conserved across divergent CRISPR-Cas systems. Recent experiments have provided evidence for Cas1–Cas2 containing complexes in the type I-A CRISPR system in *Thermoproteus tenax*, where Cas1 and Cas2 exist as a fusion protein, and in the Type I-F system in the plant pathogen *Pectobacterium atrosepticum* (Plagens *et al.*, 2012).

Interestingly, in type I-F system the Cas2 is fused to the Cas3 nuclease in the trans-acting nuclease that is used for target degradation (Rollins *et al.,* 2017).

Cas1 and Cas2 proteins of the *E. coli* type I-E CRISPR-Cas system form a stable complex containing two asymmetrical Cas1 homodimers (a-b and c-d) and one symmetrical Cas2 homodimer. Each Cas1 monomer has an N-terminal β-sheet domain and C-terminal α-helical domain. The Cas2 protein contains a core ferredoxin fold domain. Cas1a and Cas1c make contacts with the Cas2 dimer, with no contacts between Cas1b or Cas1d and the Cas2 dimer. Similar contacts are present between Cas1a and Cas1c with Cas2 on opposite sides, creating an overall symmetrical complex. Each of the four Cas1 monomers in the Cas1-Cas2 complex contains a PAM-recognition domain. The presence of a PAM in the active site of Cas1 mediates the appropriate position of the substrate and PAM relative to the cleavage site (Wang *et al.*, 2015). Since only PAM-containing spacers will be effectively targeted by interference machinery in type I CRISPR-Cas systems, PAM recognition should occur in a coordinated manner by both CRISPR adaptation and CRISPR interference machineries.

Crystallized Cas1-Cas2 complex was bound to partially double-stranded splayed DNA fragments that may correspond to physiologically relevant fragments of foreign DNA on their way of becoming spacers (Nuñez *et al.*, 2015; Wang *et al.*, 2015). The ends of the dsDNA spacer precursor are splayed by tyrosine resides in each Cas1 dimer, which fixes in place a core 23–base pair dsDNA region. The 3′ single-stranded ends of the precursor extend into active subunits of each corresponding Cas1 dimer (Wang *et al.*, 2015; Nuñez *et al.*, 2015). The length of spacer precursor depends on fixed distances between the two Cas1 active sites. Many CRISPR-Cas systems have highly consistent, yet system-specific spacer lengths. These should be determined by the Cas1-Cas2 “molecular rulers”, which exist in these systems (Wang *et al.*, 2015, Nuñez *et al.*, 2015). However, in some systems, such as those of type III, the length of spacers found within CRISPR arrays appears more variable, though there is a lack of data on Cas1-Cas2 structure and function in these systems.

Although strong affinity of the Cas1-Cas2 adaptation complex to the leader-repeat sequence was shown *in vitro* (Nuñez *et al.*, 2015), the result has not been wholly justified *in vivo*. An integration host factor (IHF) from *E. coli* is proposed to mediate spacer acquisition *in vivo* and integration into linear DNA *in vitro*. IHF binds to the

leader sequence and induces a sharp DNA bend, allowing the Cas1-Cas2 integrase to catalyze the first integration reaction at the leader-repeat border (Nuñez *et al.*, 2016). IHF is absent in many prokaryotes, including archaea, indicating that other leader-proximal integration mechanisms exist. In addition to the Cas1-Cas2 adaptation complex and integration factors, several other proteins, such as RecB, RecC, RecD, PolI, may be involved in spacer incorporation (Levy *et al.*, 2015;Ivančić-Baće *et al.*, 2015).

### *2.2.2 Naïve And Primed Adaptation*

For type I systems two modes of adaptation have been described: naïve and primed (Fineran and Charpentier, 2012; Datsenko *et al.*, 2012; Yosef *et al.*, 2012; Swarts *et al.*, 2012). While naïve adaptation is not very efficient and relies only on Cas1 and Cas2 proteins, much more efficient primed adaptation relies on specific DNA targeting. Primed adaptation requires not only Cas1-Cas2 proteins, but also the Cascade-crRNA complex and the Cas3 nuclease (Datsenko *et al.*, 2012; Swarts *et al.*, 2012; Fineran *et al.*, 2014).

**Naïve adaptation**

In spite of its apparent simplicity, naïve adaptation is somehow biased towards incorporation of spacers from extrachromosomal DNA, rather than from host DNA (Díez-Villaseñor *et al.*, 2013; Levy *et al.*, 2015; Yosef *et al.*, 2012). One possible explanation for this bias was proposed in a study by Levy and co-workers (Levy *et al.*, 2015). They showed that stalled replication forks are the preferential source for spacer acquisition, and components of the RecBCD complex affect acquisition efficiency. Reparation of stalled replication forks occurs through RecBCD-mediated homologous recombination, and resulting DNA fragments might be used for integration into the CRISPR array by the Cas1-Cas2 adaptation complex. It can be assumed that processing of the RecBCD complex stops at octamer sequences called “Crossover Hotspot Instigators” (“Chi”) (Smith, 2012), that are frequent in the *E. coli* genome but quite rare in plasmids and bacteriophages. It has also been proposed that RecBCD recognizes the unprotected double-stranded DNA ends that are present in great numbers in at least some phage genomes during infection, and this may be an additional factor for a bias of naïve adaptation towards foreign DNA (Levy *et al.*, 2015). Naïve adaptation levels are

decreased but not abolished out by RecBCD-complex withdrawal, implying that additional sources and mechanisms of naive spacer acquisition will be discovered.

**Primed adaptation**

Mutations in PAM or the protospacer sequence prevent target degradation by effectors (Deveau *et al.*, 2008; Jinek *et al.*, 2012; Semenova *et al.*, 2011; Fineran *et al.*, 2014; Westra *et al.*, 2013). For avoiding CRISPR immunity phages can mutate the PAM or protospacer sequences (Deveau *et al.*, 2008). So, in order to maintain effective defense, the CRISPR-Cas system needs to update the set of spacers faster than escape phages can arise. To meet this requirement Type I CRISPR-Cas systems have evolved a special mechanism of "primed adaptation". Primed adaptation occurs when preexisting spacers that recognize a target promote the acquisition of additional spacers from the same target. Primed CRISPR adaptation requires Cas1-Cas2 complex, Cascade, Cas3 nuclease, and a "priming" crRNA, which targets a protospacer. In the absence any of these components primed adaptation does not occur (Datsenko *et al.*, 2012).

In previous studies primed adaptation has been supposed to be provoked by Cascade-crRNA targeting of partially but not of fully matching protospacers, and two different structures of the Cascade-crRNA complex for priming and interference were suggested. On this basis, two similar models of primed adaptation, differing in their details, have been proposed. They envision that the effector complex bindings with a partially matching target (priming protospacer) thus mediating Cas3 recruitment. Cas3 slides along the double-stranded DNA from the protospacer, and interaction of Cas3 and the Cas1-Cas2 adaptation complex directs the latter to excision or PAM-associated protospacers. Both models suggest that binding of a fully matching protospacer carrying functional PAM by the effector complex and Cas3 recruitment lead to target degradation without adaptation, while interactions of Cas3 with binding of the effector complex to a priming protospacer does not lead to DNA degradation (Amitai and Sorek, 2016; Wright *et al.*, 2016). These models postulate that effector complexes interacting with fully matching protospacers is qualitatively different from that bound to partially mismatched, adaptation-competent protospacers.

Surprisingly, it has recently been shown that primed CRISPR adaptation can be promoted by either fully or partially matched protospacer recognition (Semenova *et al.*, 2016; Severinov *et al.*, 2016). Moreover, fully matched protospacers promote more

efficient spacer acquisition than partially matched ones if the effect of intracellular invader copy number and differential rate of degradation by interference is excluded (Semenova *et al.*, 2016). So a positive feedback loop is formed, providing effective defense against phages even in the absence of escape mutants (Swarts *et al.*, 2012). If the copy number of invaders within the host cell is factored in, adaptation-competent protospacers impact more on spacer acquisition as they persist for longer periods of time (Severinov *et al.*, 2016; Kunne *et al.*, 2016; Semenova *et al.*, 2016).

Thus, primed adaptation implies not only recognition of the protospacer, but also its destruction in the course of Cascade-Cas3 mediated interference, raising the possibility of interaction of the Cas1-Cas2 adaptation complex with the interference machinery. Indeed, in *Pectobacterium atrosepticum*, Cas2 exists as an N-terminal fusion with Cas3 (Richter *et al.*, 2012) and interaction of the Cas3 and Cas1-Cas2 complex in *E. coli* has been shown *in vitro* (Redding *et al.*, 2015). Because priming is initiated by site-specific target recognition, it specifically increases acquisition of spacers located in *cis* with the protospacer (Staals and Jackson 2016; Shmakov *et al.*, 2014; Swarts *et al.*, 2012; Datsenko *et al.*, 2012; Richter *et al.*, 2014). Plasmid targeting of the *E. coli* type I-E system leads to preferential acquisition of spacers from the non-target strand (Datsenko *et al.*, 2012; Savitskaya *et al.*, 2013; Swarts *et al.*, 2012). Targeting the phage genome leads to a gradient of spacer acquisition efficiency decreasing with distance from the target protospacer and opposing strand bias of acquired spacers upstream and downstream of the priming site (Strotskaya *et al.*, 2017). Preferential acquisition of new spacers from the primed protospacer region compared with distant regions was detected for the type I-B CRISPR-Cas system, yet no clear gradient was detected (Jackson *et al.*, 2017). As in the above-mentioned systems, the distribution of newly acquired spacers in the type I-F system shows a gradient around the targeted protospacer, but the direction of strand bias is dependent on the model system used (Richter *et al.*, 2014; Vorontsova *et al.*, 2016). Since the partially mismatched protospacer can be recognized and degraded by interference machinery, it was suggested that products of DNA degradation fuel primed CRISPR adaptation (Kunne *et al.*, 2016).

In conclusion, priming promotes the acquisition of new spacers from the same targets, thus reducing for phage the possibility of avoiding CRISPR immunity, as a

result resistance of bacterial cell is reinforced. Although primed adaptation is much more effective and beneficial, naïve adaptation is essential for unknown foreign agents and seems to be a universal feature of all CRISPR-Cas systems.

### *2.2.3 CRISPR Adaptation in Type II and Type III Systems*

Although the existence of bona fide primed adaptation has been shown exclusively for type I CRISPR-Cas systems, CRISPR-Cas adaptation in type II systems requires not just Cas1-Cas2, but also Cas9 and crRNA (Heler *et al.*, 2015; Wei *et al.*, 2015a). The most recently acquired spacers in the type II CRISPR-Cas system are selected from protospacers with PAMs recognized by Cas9, this implay that Cas9 recognazes PAM sequences during adaptation. To check this hypothesis, different variants of Cas9 with mutation in active centre or PAM-recognition residues were studied. The Cas9 carrying mutation in active center was able to acquire spacers with the correct PAM. In case of the PAM-recognition resides mutation only spacers without PAM were acquired (Wei *et al.*, 2015a; Heler *et al.*, 2015; Sternberg *et al.*, 2016). Some Cas9 variants can also function with non-CRISPR RNAs and tracrRNA (Sampson *et al.*, 2013). This raises a possibility that host or mobile genetic element-derived RNAs might direct promiscuous Cas9 activity, resulting in DNA breaks or replication fork stalling that could potentially result in spacer precursor generation (Jackson *et al.*, 2017).

Spacers acquired by type II-A CRISPR are variable (Lopez-Sanchez *et al.*, 2012; Horvath *et al.*, 2008; Sternberg *et al.*, 2016). As was shown recently, a frequency of usage particular spacer sequences of the phage genome reflects the protective efficiency of a spacer (Sternberg *et al.*, 2016), and unfunctial spacers are eliminated during long-term experiments. No preferential acquisition of new spacers from an invader was detected for spacer acquisition in the *S. thermophilus* type II-A system (Wei *et al.*, 2015a). Since acquisition of a spacer from the host genome is lethal (Paez-Espino *et al.*, 2013; Sternberg *et al.*, 2016), it seems that there must be a specific mechanism of protection from self-immunity that remains to be described.

New spacers acquired by type III systems are produced from from invader RNA (Hale *et al.*, 2009; Goldberg *et al.*, 2014). In some type III systems Cas1 protein is fused to reverse transcriptase domains (RTs), providing a mechanism to integrate

spacers from RNA substrates (Silas *et al.*, 2016). For example, the RT-Cas1 fusion from *Martelella mediterranea* can integrate RNA precursors into an array, which are subsequently reverse-transcribed to generate DNA spacers (Jackson *et al.*, 2017; Silas *et al.*, 2016). Integration of DNA-derived spacers also occurs, indicating that the RNA derived–spacer route is not exclusive (Jackson *et al.*, 2017; Silas *et al.*, 2016). So the combined integrase and reverse transcriptase activity of RT-Cas1–Cas2 enhances CRISPR adaptation against highly transcribed DNA mobile genetic elements and potentially against RNA-based invaders (Jackson *et al.*, 2017).

Despite enormous progress during the last decade in the understanding of adaptive immunity in prokaryotes mediated by CRISPR-Cas systems, the detailed mechanism of CRISPR adaptation is still far from being understood. Unanswered questions include: the mechanisms of primed adaptation; the *in vivo* DNA specificity of the Cas1-Cas2 complex; mechanisms of adaptation intermediates processing from original protospacer to Cas1-Cas2 adaptation complex bound fragments ready for incorporation into array; determinants of spacer choice bias; details of cooperation between CRISPR interference and adaptation machineries during primed adaptation; and many others. Future investigations will doubtless shed light on such mechanisms, revealing the unique mechanisms of adaptive immunity.

## 2.3 Regulation Of CRISPR-Cas Immunity

The host bears substantial “fitness costs” for maintaining its CRISPR-Cas systems. In experiments with *S. thermophilus* type II-A, it was shown that high costs of CRISPR-Cas immunity are mainly due to maintenance of the defense function, while acquisition of new spacers and subsequent increase of immune memory does not entail any major cost (Vale *et al.*, 2015). Many mechanisms for CRISPR-Cas regulated expression have been evolved in order to reduce maintenance expenses. Some of these are described below.

In *E. coli*, H-NS was shown as negative regulator of *cas* genes expression (Pul et al., 2010, Pougach et al., 2010). Although positive regulators LeuO (Westra et al., 2010) as well as RcsB-BglJ (Arslan et al., 2013) were also described, the physiological conditions of *E. coli* type I-E CRISPR-Cas activation need to be determined.

The Csa3a from the *Sulfolobus islandicus* type I-A system has been identified as a positive adaptation regulator (Liu *et al.*, 2015). The *csa3a* gene is located near to the *cas* operon. The Csa3a may was bind distinct two promoters, thus carrying out the regulation of adaptation genes expression. The overexpression of *csa3a* gene leads to activation of naïve adaptation.

The CRP protein has been described as the regulator of *cas* operon expression (Patterson *et al.*, 2015). The influence of the CRP protein on *cas* operon transcription is variable and depends on the type of the CRISPR-Cas system and bacterial species. The CRP regulates transcription by interaction with a consensus motif located upstream of the *cas1* gene. The CRP regulator was shown to increase both adaptation and interference in *Pectobacterium atrosepticum* type I-F system, CRP regulation increases *cas* genes transcription in the *Thermus thermophilus* type I-E and type III-A systems, while in the *E. coli* type I-E system it represses interference. As was proposed recently, distinct regulatory modes depends on the ecological niches occupying by different bacteria (Sternberg *et al.*, 2016).

### *2.3.1 Anti-CRISPR Strategies*

Bacteriophages have evolved different strategies to overcome the CRISPR-Cas defense barrier in the course of the endless arm race between prokaryotes and their invaders. More sophisticated anti-CRISPR mechanisms have been evolved by viruses in

addition to PAM or protospacer sequence mutations mentioned above (van Houte *et al.*, 2016).

Bacteriophages of *Pseudomona aeruginosa* belonging to the Mu-phage family have been found to encode proteins, which can inactivate CRISPR immunity. For type I-F systems five anti-CRISPR families of phage proteins were found, and four families were described for type I-E systems (Bondy-Denomy *et al.*, 2012; Bondy-Denomy *et al.*, 2015; Pawluk *et al.*, 2014). The anti-CRISPR system of type I-F is represented by AcrF1, AcrF2 and AcrF3 proteins (Bondy-Denomy *et al.*, 2015). Interactions of AcrF1 and AcrF2 with the Csy effector complex lead to blocking its DNA-binding activity. The Acr3 directly binds the Cas3 nuclease blocking its recruitment to the Csy interference complex (Maxwell, 2016).

Anti-CRISPR homologs are not limited to phages. As was shown by comparative genomic analysis, non-phage genomic regions of several *Pseudomonas* strains contain anti-CRISPR loci. These loci were co-localized with genes activated during DNA transfer and conjugation (Pawluk *et al.*, 2014; Bondy-Denomy *et al.*, 2012). It is possible that mobile genetic elements use anti-CRISPR strategy to increase their survival.

It is also thought that mobile genetic elements capable of blocking CRISPR immune response contribute to increasing the bacterial strain virulence. For instance, an active pathogenicity genomic island containing anti-CRISPR gene homologs was found in highly virulent clinical isolates of *P. aeruginosa*. Presumably this pathogenic island may be introduced into cells during the conjugation process (Battle *et al.*, 2008). If so, mobile genetic elements “armed” with anti-CRISPR genes may play a critical role in horizontal transfer, allowing invader elements to avoid the immune response of a new host (Shabbir *et al.*, 2016).

# Chapter 3. CRISPR-Cas Systems And Their Biotechnological Applications

Practical interest in CRISPR-Cas systems originated from their ability to specifically recognize almost any unique DNA locus. This recognition is programmed via complementary interactions between crRNA spacers and protospacers in target molecules.

## 3.1 Genome Editing

The development of CRISPR-mediated genome editing has revolutionized biotechnology in the past three years. The indisputable advantage of CRISPR technologies compared to the existing editing methods, based on TALEN and ZFN, arises from easily programmable target DNA recognition by interference complexes charged with appropriate crRNAs. Cas9 from *S. pyogenes* is widely used for genome editing applications, because of its simplicity. The two-component system contains Cas9 and crRNA fused with tracrRNA into a single guide RNA (sgRNA); this is all that is nedded for introducing double-stranded break (DSB) into DNA. If recombination template is absence, then mutations may be introduced by non-homologous end joining (Jiang and Marraffini, 2015). The presence of a complementary template (sister chromosome or a synthetic fragment with appropriate flanking sequences) leads to reparation of DSBs by homologous recombination (HDR). The mutated template will lead to incorporation of the mutations into the target site. The recombinant template must contain additional mutations altering the seed or PAM sequence and preventing Cas9 nuclease activity for avoiding cleavage of repaired targets by Cas9.

Limitation of Cas9 nuclease application consists of a PAM requirement located downstream of the target site. For instance, *S. pyogenes* Cas9 requires an NGG or NAG PAM (Jiang *et al.*, 2013). The GG dinucleotides occur at a frequency of every eight base pairs in genome; so extended regions which do not contain this PAM sequence are rare (Cong *et al.*, 2013). Other Cas9 nucleases that have been found in *S. thermophilus* and *Neisseria meningitidis*, utilize NNAGAAW and NNNNGATT (Cong *et al.*, 2013; Hou *et al.*, 2013) respectively, extended the set of PAMs.

Simultaneous introduction of several sgRNA into the cell enables editing of multiple genomic loci, so ability to multiplex is another key advantage of CRISPR-Cas

mediated editing technology. This approach can be used to mediate large-scale chromosomal rearrangements, such as deletions, translocations and inversions.

For making this system functional in eukaryotic cells, nuclear localization signals are added to Cas9, and supplying the nuclease gene with the appropriate regulation elements required for protein expression and subsequent guides (Jiang and Marraffini, 2015; Mali *et al.*, 2013; Cong *et al.*, 2013). The Cas9-mediated genome editing occurs through transformation of target cell with a plasmid carrying *cas9* gene and a set of guide RNAs. Other ways to introduce Cas9 into cell are loading of the Cas9 protein with the sgRNA followed by injection (Kim *et al.*, 2014) and *cas9* mRNA coinjected with sgRNAs (Zhou *et al.*, 2014; Yang *et al.*, 2013, Niu *et al.*, 2014). Finally, transgenic organisms expressing Cas9 require only an sgRNA introduction for induction NHEJ-mediated indels and mutant generating. This method was successfully applied for mice (Platt *et al.*, 2014), flies (Kondo and Ueda, 2013), and worms (Tzur *et al.*, 2013).

The simplicity and efficiency of Cas9-mediated gene editing technology provide opportunities for rapid development of gene therapy and correction genetic disorders. The effectiveness of genetic therapeutics depends on delivery efficiency into the target cell of both the Cas9 and the recombinant template with sgRNA. For efficient Cas9 delivery into a different cell types adenoviral vectors are usually used for (Wright *et al.*, 2016). The limitation of this vehicle consist of a limited amount of DNA that may be packed into viruses, and this is a problem for *S. pyogenes cas9* gene, which is 4,107 bp length. To overcome this limitation, adenoviral vectors containing smaller alleles of *cas9* were used for efficient genome editing in live mice (Ran *et al.*, 2015). A bioinformatics pipeline has been developed by S. Shmakov and colleagues to search for new singe-subunit Cas effectors, that potentially could be useful in biotechnology (Shmakov *et al.*, 2017). Application of the pipeline led to the discovery of a new type of class II CRISPR-Cas system – type V – with a single multidomain effector Cfp1 (Shmakov *et al.*, 2016), which was later shown to be effective in gene editing.

The Cas9-mediated technology is also applied for gene engineering in microbiology. The integration of the recombinant template at the target site is not induced by cleavage of the bacterial chromosome with Cas9 and causes lethality. That is used for effective selection of cells that have incorporated the mutation by recombination before Cas9 cleavage. For preventing the recombinant target from

recleavage by Cas9, mutations altering the seed or PAM sequences are introduced. This technology has been successfully used for genetic manipulation in different bacteria such as *E. coli*, *Streptomyces coelicolor*, *Lactobacillus reuteri*, *Clostridium beijerinckii*, *Streptococcus pneumoniae*, as well as large number of bacteriophages (Jiang and Marraffini, 2015).

### 3.2 Regulation Of Gene Expression

A Cas9 protein with impaired nuclease activity suppresses transcription of bacterial genes at the initiation or elongation stages, when the protospacer was located in the promoter or coding regions (Qi *et al.*, 2013, Bikard and Marraffini, 2013). The observed suppression might occur by occlusion of the target locus within the promoter region or by physically stopping the transcription elongation complex by Cas9 tightly bound to the transcription template. This type of suppression is used with much success for expression control in eukaryotic systems. The nuclease Cas3-deficient multi-subunit type I Cascade complex from *E. coli* can also be used for transcription suppression (Luo *et al.*, 2015; Rath *et al.*, 2014). The opposite effect, transcription activation, can be achieved using Cas9 with impaired nuclease activity fused with transcription activatation domains. Such a combination provides a reversible increase in the levels of gene expression in bacteria, yeast, mouse, and human cells (Bikard and Marraffini, 2013; Cheng *et al.*, 2013; Gilbert *et al.*, 2013). A catalytically dead version of Cas9 was shown to also be useful for the visualization of specific genomic loci in living cells (Gilbert *et al.*, 2013).

There are numerous prerequisites for creating CRISPR-based tools for gene expression regulation at the post-transcriptional level. In 2013, Sampson *et al.* identified a new activity of the Type II CRISPR-Cas system in *Francisella novicida* that was not realted to defense against foreign DNA (Sampson *et al.*, 2013). They showed that Cas9, tracrRNA, and an additional RNA named scaRNA interacted with a transcript of the lipoprotein gene and caused its degradation. As a result, the cells became highly virulent. The mechanisms of precrRNA processing present considerable interest for post-transcriptional regulation, because the corresponding components of the CRISPR-Cas systems can recognize repeat fragments in the context of an RNA molecule and introduce specific breaks in them. The possibility of RNA stability regulation was

demonstrated for the I-F subtype system protein, Cas6f, also known as Csy4 (Lee *et al.*, 2013, Borchardt *et al.*, 2015). It is possible that molecular tools for post-transcriptional regulation of gene expression will develop further in the near future, since such regulation is especially important for prokaryotic cells, in which efficient RNA interference has not yet been achieved.

### 3.3 Cell Selection

Cas9 nuclease induces chromosomal cleavages in prokaryotes that lead to cell death. This finding suggests using Cas9 nuclease for selective removal of cells containing specific genetic sequences from complicated bacterial populations like the human microbiota (Bikard *et al.*, 2014, Gomaa *et al.*, 2014). Activation of CRISPR interference could be used for directed manipulation of the content of bacterial communities. Bikard and co-workers showed that delivery of crRNA bearing a spacer matching a protospacer of the virulence-providing genes into *Staphylococcus aureus* cells possessing the Type II CRISPR-Cas system selectively inhibited the growth of virulent cells (Bikard *et al.*,2014). In the same work, the delivery of crRNA complementary to the protospacer of an antibiotic resistance-encoding plasmid resulted in the loss of cell resistance to the antibiotic. Similarly, induction of expression of components of the subtype I-E CRISPR interference system in *E. coli* caused selective death of cells containing a protospacer complementary to the crRNA spacer in their genomes (Gomaa *et al.*, 2014; Caliando and Voigt, 2015). The Cas9-mediated antimicrobials could be programmed with guide RNAs matching a specific virulence gene sequence to kill the bacterial pathogen harboring this gene, while sparing the rest of the microbiota (Bikard *et al.*, 2014). Bikard *et al.* have shown that the Cas9 antimicrobials can selectively kill antibiotic-resistant, but not other staphylococci by using a mouse model of staphylococcal skin colonization.

### 3.4 Strain Subtyping

Profiling of CRISPR array spacers by PCR or restriction analysis has been used for years as a way of identifying microorganism strains, and began a long time before elucidation of the functions and mechanisms of CRISPR-Cas systems (Kamerbeek *et al.*, 1997; van Soolingen *et al.*, 1993). The main prerequisite for such

profiling is a relatively stable composition of CRISPR arrays due to low levels of CRISPR adaptation. Since no expression of *cas* genes has been observed in *E. coli* and closely related genera *Yersinia* and *Salmonella*, at least in laboratory conditions (Pougach *et al.*, 2010; Westra *et al.*, 2010), profiling of their spacers is still regarded as useful, especially for subtyping of pathogenic strains (Delannoy *et al.*, 2012; Shariat *et al.*, 2013). At present, profiling is mostly performed by means of high-throughput sequencing (Fabre *et al.*, 2012).

In strains with high levels of CRISPR-Cas immunity, profiling of spacers allows prediction of strain resistance to certain bacteriophages based on the already existing spacers. This might be important for selecting bacteriophages for phage therapy and for tracing interactions of bacteria with bacteriophages (Andersson and Banfield, 2008; Sun *et al.*, 2016; Paez-Espino *et al.*, 2013), particularly in the course of therapy.

### 3.5 Design Of Strains With Required Sets Of Spacers

Creation of industrial microorganisms with resistance to bacteriophage infection determined by their CRISPR arrays is highly important in biotechnology. It is also often necessary to express crRNA with particular properties in bacterial cells, for example, for the study of CRISPR-Cas system interactions with various bacteriophages. A new simple approach for rapid creation of such strains has been suggested based on the phenomenon of primed adaptation (Strotskaya *et al.*, 2015). *E. coli* cells bearing inducible *cas* genes were transformed with a plasmid that contained a fragment of bacteriophage genome, against which new spacers should be obtained, and a protospacer partially complementary to an existing spacer of the bacterial CRISPR array. After induction of *cas* genes, interactions between crRNA spacer and the priming protospacer provided preferential selection of new spacers from the plasmid. Over 50% of the cells acquired the “phage” fragment spacer.

## Concluding Remarks

Despite the very high diversity of prokaryotic defense systems, none of them ensure 100% safety for the host, and most bacteria and archaea combine several types of defense systems to effectively counter invaders. Many examples of co-occurrence and even of cooperation between different defense systems have been described. For example, genomic loci that encode for R-M and DND immunity systems often also include genes that encode toxins implicated in the induction of dormancy or cell death (Makarova *et al.*, 2013) suggesting evolutionary association between immunity-based and programmed cell death strategies. Makarova describes the following four clusters of prokaryotic organisms based on the relative abundance of different defense systems: 1. In the genomes of the first cluster, all defense systems are under-represented (some of these organisms may use defense systems that have not yet been characterized). 2. The total amount of defense system genes is close to the expected level, with prevalence of R-M and Abi over T-A and CRISPR-Cas. 3. The total amount of defense systems genes is close to the expected level, with prevalence of T-A and CRISPR-Cas over R-M and Abi. 4. All defense systems are over-represented in the genome (Makarova *et al.*, 2013). The abundance of defense systems genes strongly correlates with the size of the genome (Makarova *et al.*, 2011). Analysis of defense genes distribution shows enrichment in archaea compared with bacteria and in thermophiles (especially hyperthermophiles) compared with mesophiles and psychrophiles (Makarova *et al.*, 2013). Although information about prokaryotic defense systems has been vastly expanded in recent comprehensive studies (Koonin 2017), discoveries of novel defense mechanisms in the future is very likely. The investigation of defense mechanisms is crucial for understanding the complexity of prokaryotic and invader co-evolution, and some of the newly discovered systems may be of use in biotechnology.

# Chapter 4. Project Objectives

Despite the enormous progress in understanding of CRISPR-Cas system-mediated adaptive immunity in prokaryotes during the last decade, the details of CRISPR adaptation mechanisms are still enigmatic. Although the molecular mechanism of new spacer incorporation is supposed to be similar to transposon integration reaction, and the Cas1-Cas2 adaptation complex was crystallized bound to partially double-stranded splayed DNA fragments that may correspond to physiologically relevant adaptation intermediates, little is known about the earliest stages of adaptation, i.e., the specific selection of fragments that become CRISPR spacers. Some hypotheses implementing the RecBCD complex have been elaborated to explain possible sources of spacers during naïve adaptation but no accepted mechanism of primed adaptation that would explain the strand bias of spacer choice, the existence of preferred sources of spacers and the relationship with target degradation is yet available. This thesis project concerns with studies of DNA specificity of the Cas1-Cas2 complex during primed CRISPR adaptation by the *E. coli* type I-E CRISPR-Cas system. The following specific aims were set:

1. To assess the interactions between the Cas1-Cas2 adaptation complex and genomic DNA during primed CRISPR adaptation;
2. To assess the interactions between the Cas1-Cas2 adaptation complex and targeted plasmid DNA during primed CRISPR adaptation;
3. To define the structure of the Cas1-Cas2 adaptation complex-bound DNA;
4. To evaluate the contribution of CRISPR-Cas system components to the emergence of primed adaptation intermediates;
5. To create a comprehensive model of primed CRISPR adaptation.

## 4.1 Preliminary Results

Results beyond the scope of the current Thesis specific aims but aiming towards the same general goal were obtained in parallel with the main body of work presented in the Thesis and were published in the following articles:

1. Semenova E, Savitskaya E, Musharova O, Strotskaya A, Vorontsova D, Datsenko KA, Logacheva MD, Severinov K. Highly efficient primed spacer acquisition

from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex. *Proc Natl Acad Sci U.S.A.* 2016 Jul 113(27):7626-31.

2. Xue C, Seetharam AS, Musharova O, Severinov K, Brouns SJ, Severin AJ, Sashital DG. CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.* 2015 Dec 43(22):10831-47.

In particular, the following results were obtained by the author.

### *4.1.1 Over-Expression Of Cas1 and Cas2 Increases Primed CRISPR Adaptation From Fully Matching Protospacer Targets*

To test if increasing of the spacer acquisition rate could affect the efficiency of primed adaptation with fully matching targets, *E. coli* KD546 strain (Semenova *et al.*, 2016) containing the *cas3* gene, the *casABCDE* operon and the *cas1cas2* genes, under control of separate inducible promoters and a CRISPR array with a single spacer, was used. As a control, *E. coli* KD263 strain without additional promoter in front of *cas1cas2* genes was used. Cells were transformed with plasmids containing a protospacer fully (pG8) or partially (pG8mut) matching the CRISPR array spacer, and after *cas* genes induction spacer acquisition was monitored. Upon induction, acquisition of new spacers was detected for KD546 cells transformed with either pG8 or pG8mut plasmid (Fig. 3A, lanes 3 and 4). In control KD263 cells acquisition of new spacers was detected only in the presence of pG8mut (lane 2).

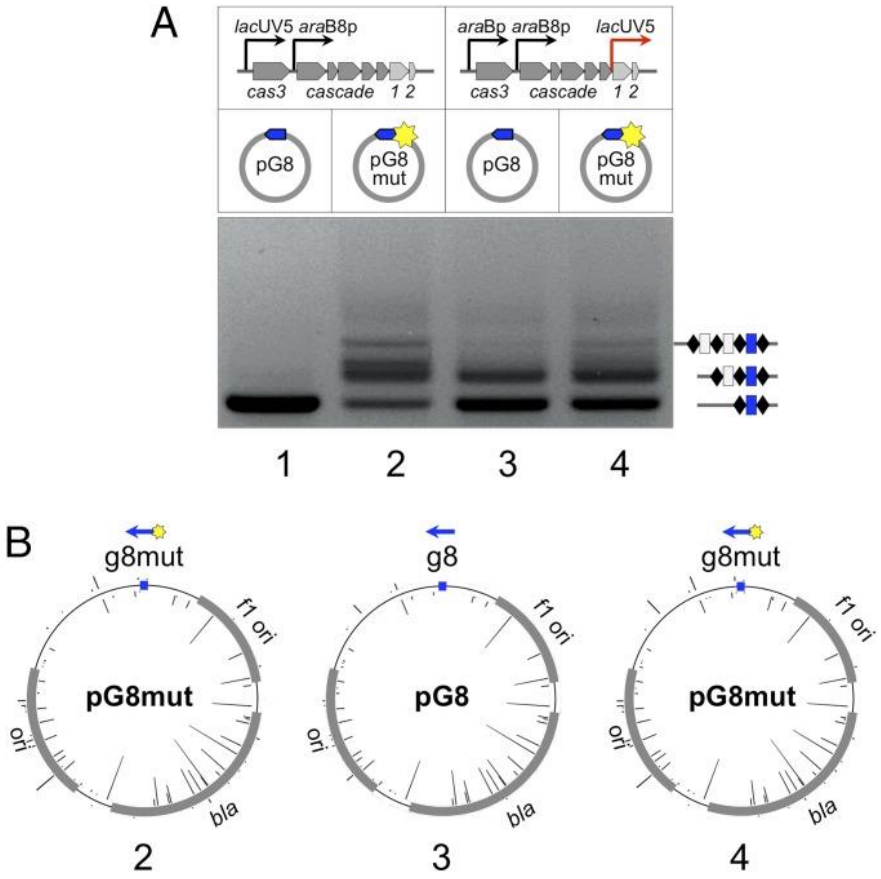



**Fig. 3. Increased expression of *cas1* and *cas2* leads to primed adaptation from fully matching protospacer targets. (A) An agarose gel showing the results of separation of products of PCR amplification of CRISPR arrays from aliquots of KD263 (lanes 1 and 2) and KD546 (lanes 3 and 4) transformed with indicated plasmids. (B) Results of mapping of spacers acquired by KD263 cultures transformed with pG8mut (2) and KD546 cultures transformed by pG8 and pG8mut (3, 4) on donor plasmids. The height of bars corresponding to individual protospacers reflects the frequency of occurrence of the corresponding spacer in expanded arrays, and the position of the bars stands for start positions of corresponding protospacers. The priming protospacer is shown as a blue arrow.**

Acquired spacers were subjected to HTS analysis and mapped on plasmid sequence. Mapping showed that the distribution of donor protospacers along the plasmid backbone, the strand bias, and the AAG PAM preference were typical for primed adaptation (Fig. 3B). Thus, increasing the intracellular concentration of Cas1 and Cas2 is sufficient to cause primed adaptation from targets with spacer-matching

protospacer and functional PAM that are normally subject to CRISPR interference without visible adaptation. This observation was published in Semenova et al. article.

### *4.1.2 A Self-Signal Protects The Host Genome From Autopriming*

To investigate if repeat-derived nucleotides at PAM positions protect CRISPR arrays from suicidal autopriming, a plasmid containing a non-transcribed repeat – spacer – repeat sequence was constructed (Fig. 4). The plasmid was created by cloning a PCR amplicon containing the first two repeats and one spacer from *E. coli* PIM5 CRISPR array (Swarts *et al.,* 2012) into the *PciI* site of plasmid pGFP-Kan (Fineran *et al.,* 2014). This plasmid was used for transformation of *E. coli* cells containing the same spacer in the genome. After transformation, stable maintenance and spacer acquisition in non-selective media were checked. Plasmid loss was assessed on non-selective plates by scoring colony fluorescence (the plasmid encodes GFP and so loss of plasmid leads to fluorescence loss). Non-fluorescent colonies were analyzed by colony PCR for integration of new spacers in CRISPR arrays. No spacers were acquired from cells transformed with a plasmid containing a CRISPR array-like structure, suggesting that the presence of repeat sequences around a protospacer matching crRNA spacer inhibited primed adaptation.


**Fig. 4. Repeat nucleotides at the PAM (i.e., CCG at position -3, -2, -1) prevent self-priming at CRISPR array. The set of plasmids containing a non-transcribed repeat – spacer sequence was analyzed in individual experiments in Δ*hnf* cells with active CRISPR-Cas system. The sequence of the distal repeat is shown in truncated form in the upper two sequences. The sequence of the spacer is shaded in gray in the top sequence. Red color highlights mutated nucleotides. A 24 h plasmid loss was calculated as a percentage of non-fluorescent colonies on non-selective plates. White colonies were subjected to PCR with primers annealing up- and downstream at CRISPR array for obtaining CRISPR array extension due to primed adaptation. Plasmid variants causing priming are marked with an asterisk.**

When the proximal repeat was deleted from the plasmid, the ability to induce primed adaptation was restored (Fig. 4). A series of plasmids with variations in the proximal CRISPR repeat was next constructed (the variants, shown in Fig. 4, included mutations in several consecutive nucleotides and truncations from the 5′ end of the repeat). Assessing the priming behavior or resulting plasmids revealed that only when repeat nucleotides at positions -3 to -1 (i.e. CCG) were altered, primed adaptation was detected. The result thus establishes that the three repeat nucleotides directly adjacent to the spacer are both necessary and sufficient to protect CRISPR arrays from self-

priming. Thus, these nucleotides save cells from autoimmune response that would have been inevitable caused by acquisition of interference-competent spacers from bacterial genome after self-priming on the array. This work was performed in Wageningen University during academic mobility program in the course of author's Ph.D. studies and was published in the Xue et al. article.

# Chapter 5. Materials And Methods

## 5.1 Strains And Plasmids

*E. coli* KD263 (K-12 F+, lacUV5-cas3 araBp8-cse1, CRISPR I: repeat-spacer g8-repeat, ΔCRISPR II) has been described (Shmakov *et al.*, 2014). *E. coli* KD454 is a derivative of KD263 carrying a deletion of the *cas3* gene, and was provided by K. Datsenko (Pardue University). *E. coli* BW40297 (25113 F+ G8+ araB8p-casA::cat uv5-cas3, cas1 H208A, CmR) has been described (Datsenko *et al.*, 2012). *E. coli* AM7-7 (K-12 F+, lacUV5-cas3 araBp8-cse1, CRISPR I: repeat-spacer g8-repeat, ΔCRISPR II, Δ*ihf*) is a derivative of KD263 carrying a deletion of the *ihfA* gene. *E. coli* BL21 DE3 ($F^-$ ompT gal dcm lon $hsdS_B(r_B^-m_B^-)$ λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) $[malB^+]_{K\text{-}12}(\lambda^S)$) was used for protein expression. *E. coli* DH5α (F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK–mK+), λ–) was used for molecular cloning.

Plasmid pEK21 carrying *cas1* and *cas2* genes was constructed by cloning of *E. coli* genomic *cas1* and *cas2* genes amplified with Cas1_for– Cas1_rev and Cas2_for - Cas2_rev primers, respectively (Supplementary Table 1), under the inducible T7 RNA polymerase promoter in the pET28 vector.

Plasmid pEK1 was constructed by cloning of *E. coli* genomic *cas1* amplified with Cas1_for – Cas1_rev primers and cloned into a pET28 vector. Plasmid pT7blue is commercially available (Novagen). Plasmids pG8 and pG8mut have been described previously (Datsenko *et al.*, 2012).

Plasmid pG8mut_CCG is a pG8mut derivative containing CCG instead of AAG PAM in front of hot protospacer 1 (HS1) (5’ GTGCTCATCATTGGAAAACGTTCTTCGGGGCGA 3’). The mutation was introduced by the standard mutagenesis protocol. Plasmid pG8mut was amplified with primers HS1_CCG_for and HS1_CCG_rev (primer sequences are available in Supplemental Table 1) by using iProof™ High-Fidelity DNA Polymerase (Bio-Rad). The PCR reaction contained: 10 µg pG8mut plasmid, 1 × iProof HF buffer, 200 mM dNTPs, 0.2 µM of each primer, 1 U iProof DNA polymerase, and sterile water up to 50 µl. The thermal cycling protocol was as follows: 1. 98 °C - 30 sec, 2. 98 °C - 10 sec, 3. 65 °C - 30 sec, 4. 72 °C - 1 min 30 sec, 5. 72 °C - 5 min, cycles 2-4 were repeated 25

times. The resulting PCR product was purified using a GeneJET PCR Purification Kit (Thermo Scientific), following the manufacturer's instructions, and subsequently treated with 0.05 U/µl *DpnI.* After *DpnI* digestion DNAwas purified using a GeneJET PCR Purification Kit (Thermo Scientific) check at 0.8% agarose gel. Because non-phosphorylated primers were used for amplification, DNA was treated with T4 polynucleotide kinase (PNK) (Thermo Scientific), following the manufacturer's instructions, to phosphorylate the 5'-ends. After PNK treatment DNA was precipitated, solved in 5 µl of sterile water and used for ligation. The ligation reaction was performed in the following conditions: 1 ×T4 DNA ligase buffer, 5 µl linearized plasmid, 0.5 U T4 DNA ligase (Thermo Scientific), sterile water up to 10 µl. The ligation mixture was incubated at room temperature for 60 min, and used for transformation of competent *E.coli* DH5α cells. Plasmids from obtained clones were sequenced using the Sanger protocol to confirm mutation acquisition.

### 5.2 Growth Conditions

*E. coli* cells were grown at 37 °C in LB (5 $g \cdot L^{-1}$ NaCl, 5 $g \cdot L^{-1}$ yeast extract, and 10 $g \cdot L^{-1}$ tryptone) at 225 rpm with the corresponding antibiotic, if needed. Bacterial growth was measured at 600 nm (OD600).

### 5.3 Competent Cell Preparation

The standard rubidium chloride method was used for preparation of *E.coli* competent cells (Promega Protocols and Applications Guide ($3^{rd}$ edition)). A single bacterial colony from an LB plate was inoculated into a 2.5 ml LB medium and incubated overnight at 37 $^{o}$C with aeration. The overnight culture was inoculated into 250 ml of LB containing 20 mM $MgSO_4$ (this results in 1:100 dilution) and was grown until OD reached 0.4-0.6. Cells were harvested by centrifugation at 4000 × g for 5 min at 4 $^{o}$C. After supernatant removal the cell pellet was resuspended in 0.4 volume of ice cold TFI buffer (30 mM potassium acetate, 10 mM CaCl, 50 mM MnCl, 100 mM RbCl, 15% glycerol, pH 5.8, adjusted with 1 M acetic acid) and incubated in ice for 5 min. After incubation cells were harvested by centrifugation as described previously. The cell pellet was resuspended in 0.04 volume of ice-cold TFII buffer (10 mM MOPS, 75 mMCaCl, 10 mM RbCl, 15% glycerol, pH 6.5 adjusted with 1 M KOH) and incubated

in ice for 45 minutes. After incubation the cells were aliquoted at 100 µl into pre-chilled tubes and frozen in liquid nitrogen.

### 5.4 Transformation Of *E. coli* Cells

The aliquot of competent cells was thawed in ice for 15 min. 50-100 ng of plasmid DNA was added to the competent cells, followed by 10-20 min incubation in ice. The cells were then subjected to heat shock by 45 sec incubation at 42 ºC, and immediately cooled in ice for 3-4 min. The chilled cells were treated with 1 ml of LB medium and incubated at 37 ºC with shaking for 30 min. After transformation the cells were spread onto LB-agar plates containing 1.5% (15 g $L^{-1}$) agar supplemented with 100µg/ml ampicillin.

### 5.5 *In vivo* Induction Of CRISPR-Cas System

*E. coli* KD263, AM7-7, KD454 or BW40297 were transformed with pG8mut, pG8mut_CCG or pT7blue (Novagen) plasmids. Transformants were selected on LB agar plates containing 100 µg/ml ampicillin. Overnight cultures obtained from individual colonies were diluted 1:100 and grown in liquid medium until OD reached 0.3-0.4. They were then divided into two parts and one of them was induced with 1 mM IPTG and 1 mM L-arabinose as described (Semenova *et al.*, 2016). 3 hours post-induction, the aliquots of induced and uninduced cultures were processed for ChIP or total DNA purification.

### 5.6 Plasmid And Total DNA Purifications

Plasmid DNA purification was performed using a GeneJET Plasmid Miniprep Kit (Thermo Scientific), following the manufacturer's instructions.

Total DNA was purified from 2 ml of cell cultures with a Genomic DNA Purification Kit (Thermo Scientific), following the manufacturer's instructions and adding glycogen (Thermo Scientific) during precipitation steps to promote recovery of short and singe-stranded DNA fragments. Total DNA (~5 µg) was dissolved in 25 µl of deionized water.

### 5.7 Generation Of Model Substrate

20 pM of oligonucleotides HS1_full/cmp_for and HS1_full/cmp_rev (or HS1_part/cmp_for and HS1_part/cmp_rev) were subjected to annealing (5 min at 95 ºC followed by slow temperature reduction in 1 × annealing buffer (20 mM Tris–HCl pH 7.5, 50 mM NaCl, 20 mM $MgCl_2$) in 100 µl reaction volume. The annealed substrate was precipitated with glycogen (Thermo Scientific) and used in downstream experiments. A model substrate 210-bp in length was prepared by PCR with appropriate primer pairs (Supplementary Table 1) using pG8mut plasmid as a template for amplification. The PCR-products were purified using a GeneJET PCR Purification Kit (Thermo Scientific).

### 5.8 Antibody Preparation And Purification

A pEK21 plasmid encoding N-terminally 6-His tagged Cas1 and untagged Cas2 was used to obtain antibody for ChIP experiments. Plasmid-borne *cas* genes were expressed in the *E. coli* BL21 (DE3) strain in LB medium containing 30 µg/ml kanamycin. The cells were grown at 37 ºC until OD600 reached 0.6, followed by induction with 1 mM IPTG and further growth for 2 h. The cells were harvested by centrifugation for 20 min at 5000 × g at 4 ºC and frozen at −80 ºC. Cell pellets were resuspended in buffer A (20 mM Tris-HCl pH 8.0, 0.5 M NaCl) containing 1 mg/ml lysozyme. The cells were then disrupted by sonication and cell lysate was clarified by centrifugation at 16 000 × g for 1 h and filtering through a 0.45 µm filter. The extract was loaded to a 1 ml Chelating HP column (GE Healthcare) loaded with $Ni^{2+}$ and equilibrated with buffer A. The column was washed with buffer A containing 20 mM and 50 mM imidazole and bound proteins were eluted with 300 mM imidazole in buffer A. The total amount of protein was ranged 4-5 mg.

Immunization was performed in a commercial facility at the Institute of Protein Research of the Russian Academy of Sciences (Pushchino). Five rats were treated with purified Cas1 protein in a mixture with Freund adjuvant for antibody production. Treatment was performed at 0, 7, 14, 35, 42 days, with antisera obtained on day 51 after treatment. Following immunization the antisera were tested by Western blot analysis with recombinant proteins, and highly reactive antiserum was used for further antibody purification.

Antibody purification was performed using an affinity column containing recombinant Cas1. Recombinant Cas1 plasmid pEK1 containing the *E. coli cas1* gene under T7 RNA polymerase promoter was used for purification. Recombinant Cas1 contained a hexahistidine tag at the N terminus. Purification of Cas1 protein was performed as described above for the Cas1-Cas2 complex.

Purified Cas1 was immobilized on cyanogen bromide-activated sepharose 4B (Sigma-Aldrich) and antibody purification was performed, following the manufacturer's instructions. Cas1 protein in the amount of 5 mg per 1 ml of cyanogen bromide-activated sepharose 4B was dialyzed overnight against coupling buffer containing 100 mM $Na_2CO_3$, 100 mM $NaHCO_3$, 500 mM NaCl, pH 8.3. The sepharose was activated with ice cold 1 mM HCl and washed three times with ice cold coupling buffer. For immobilization activated beads were liganded in 10 ml of coupling buffer and were incubated overnight at 4 $^{o}$C in a tumbler. On the next day beads were washed three times with coupling buffer and incubated with 15 ml of 100 mM Tris-HCl, pH 8.0 during 2 h at room temperature in a tumbler. After that three wash cycles at 4 °C with 100 mM TRIS buffer pH 8.0 (100 mM Tris-HCl, 500 mM NaCl, pH 8.0 adjusted with HCl) followed by 100 mM acetate buffer pH 4.0 (100 mM sodium acetate, 500 mM NaCl, pH 4.0 adjusted with acetic acid) were performed. The resulting beads were washed with ice-cold 1 × PBS (2.7 mM KCl, 137 mM NaCl, 1.76 mM potassium phosphate, 10 mM sodium phosphate, pH 7.4 adjusted with HCl) and subjected to binding of antiserum. For binding of antiserum beads were resuspended with 10ml of 1 × PBS, and 10 ml of antiserum were added. The resulting mixture was incubated overnight at 4 °C in a tumbler. Elution of antibody was performed with 200 mM glycine pH 2.8. The resulting antibody was dialysed overnight at 4 °C against 1 × PBS in dialysis tubing with a cut off of 12000-14000 Da.

**5.9 Western Blotting**

Pellets of cells with or without *cas* genes induction were resuspended in 0.5 ml of ice-cold IP buffer (50 mM HEPES–KOH H 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS and 1 mM PMSF), and sonicated using a Vibra-Cell VCX130 machine (Sonics) at 80% power for 1.5 min. The resulting

samples were separated by centrifugation at 130000 × at 4 ºC. Supernatants containing total proteins were used for future analyses.

Total proteins were separated by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (25 μg of proteins per lane) and then transferred to polyvinylidene difluoride (PVDF) membranes (Millipore). After blocking with 5% nonfat milk at 37 °C for 1 hours, the membranes were incubated overnight at 4 °C with anti-Cas1 antibody (rat polyclonal antibody, 1:5000 diluted) or anti-RNAPα antibody (mouse monoclonal antibody (Abcam), 1:1000 diluted), followed by incubation with alkaline phosphatase-linked secondary antibodies. The immunoreactive bands were visualized using an Immun-Star AP Substrate (Bio-Rad).

### 5.10 In-Gel Digestion For Mass-Spectrometric Characterization of Proteins

For the purpose of mass spectrometry analysis, protein samples were extracted from 10% sodium dodecyl sulfate-polyacrylamide gel after Coomassie staining, as has been described (Granvogl *et al.*, 2007). The protein band was excised from the gel and washed for 20 min in 100 μl solution of 100 mM ammonium bicarbonate (Sigma) in 40% acetonitrile at 37 ºC. The procedure was repeated twice, discarding the liquid after each cycle. After washing the samples were covered with 100% acetonitrile and incubated for 5 min at room temperature with subsequent discarding of the acetonitrile. The samples were treated with 4 μl of trypsin solution (trypsin: protein ratio of 1:20-1:100, w/w) in 50 mM ammonium bicarbonate for 4 h at 37 ºC. After incubation the digestion reaction was stopped with 7 μl of 0.5% solution of trifluoroacetic acid (Merck) in 10% acetonitrile. The resulting samples were used for mass spectrometric analysis, which was performed at the A.N. Belozersky Research Institute of Physico-Chemical Biology (Moscow State University).

### 5.11 Chromatin Immunoprecipitation

A chromatin immunoprecipitation procedure was performed as described with minimal modifications (Kahramanoglou *et al.*, 2011). Three biological replicates of every immunoprecipitation experiment were performed. Summarizing the procedure: formaldehyde was added to twenty milliliters of cultures with induced or uninduced *cas* genes to a final concentration of 1% for 20 min at room temperature with rotation. The

reaction was quenched by adding glycine (0.5 M final concentration) and incubated under the same conditions for 5 min. The cells were pelleted by centrifugation and washed three times with 1 × TBS (10 mM Tris-HClpH 7.5, 150 mM NaCl). One milliliter of lysis buffer (10 mM Tris-HClpH 8.0, 20% sucrose, 50 mM NaCl, 10 mM EDTA, 20 mg/ml lysozyme) was added and the samples were incubated at 37 ºC for 30 min. After adding of 4 ml of IP buffer (50 mM HEPES–KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS,1 mM PMSF) the samples were subjected to sonication on a Vibra-Cell VCX130 machine (Sonics) at 80% power for 5 min yielding DNA fragments with length of 200–300 bp. This and later steps were performed on ice. After centrifugation, 800 µl of supernatant was preincubated with 20 µl of Protein A/G Sepharose beads (Thermo Scientific) to pull down proteins unspecifically interacting with the resin, and unbound fraction was combined with 30 µl of BSA-blocked Protein A/G Sepharose and 7 µl of anti-Cas1 antibody and incubated overnight on a rotary platform. Standard washing with IP buffer, high salt IP buffer (the IP buffer supplied 500 mM NaCl), wash buffer (10 mM Tris-HCl pH=8.0, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% sodium deoxycholate), TE buffer (10 mM Tris-HCl pH=8.0, 1 mM EDTA) and elution steps were performed as described (Kahramanoglou *et al.*, 2011) in elution buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% SDS). The immunoprecipitated samples and sheared DNA samples before IP (input) were de-crosslinked in 0.5 × elution buffer containing 0.8 mg/ml Proteinase K at 42 ºC for 2 h followed by 65 ºC for 6 h. DNA was precipitated with glycogen and dissolved in 20 µl of MilliQ water. The typical yield of DNA was 40–60 ng.

### 5.12 Real-Time PCR Quantification

Each qPCR reaction was carried out in triplicate (three technical repeats) in a 20 µl reaction volume with 0.8 U of HS Taq DNA polymerase (Evrogen) and 0.01 µl of Syto13 intercalating dye (LifeTechnology) using a DTlite4 (DNA-Technology) amplifier. For each reaction, melting curves were analyzed to ensure amplicon quality and exclude primer dimer formation during amplification. Additionally, amplicons from qPCR reactions were cloned and, for each amplicon, several randomly chosen recombinant plasmids were sequenced. In each case the size and sequence of the cloned

inserts matched the expectation (Supplementary Table S2). The enrichment ratio ΔΔCt = ΔCt ind (mean Ct IP – mean Ct input) − ΔCt unind (mean Ct IP – mean Ct input) was determined. A $2^{-\Delta\Delta Ct}$ value was determined in order to convert ΔΔCt values to relative differences in DNA fragment concentrations.

10 µl aliquots of ChIP material were treated with 0.5 U of *TaiI*/*FaiI* restriction endonucleases (Thermo Scientific), following the manufacturer's instructions. After precipitation, qPCR was conducted as described above. The fold enrichment between treated and untreated DNA was next calculated as $2^{-(\Delta\Delta Ct[treated] - \Delta\Delta Ct[untreated])}$.

10 µl aliquots of total DNA obtained from KD263, AM7-7, KD454, BW40297 cells containing pG8mut or pG8mut_CCG plasmid were treated either with 0.02 U of S1 nuclease (Thermo Scientific) or 0.5 U of *Tai*I/*Fai*I restriction endonucleases (Thermo Scientific), following the manufacturer's instructions. After precipitation, qPCR was conducted as described above. The ratio between treated and untreated samples was calculated as $2^{-Ct[treated]\_[untreated])}$. For S1 or *TaiI*/*FaiI* treatment 1 pmole of model fragments was used, followed by qPCR reactions, which were constructed as described above. The ratio between treated and untreated samples was calculated as $2^{-Ct[treated]\_[untreated])}$.

### 5.13 Spacer Acquisition Analysis

Aliquots of induced and uninduced total DNA samples were subjected to PCR with primers P4518 and P4581 annealing at both sides of the CRISPR array. The results were analyzed by agarose gel electrophoresis. For electrophoresis in agarose gels, the DNA samples were combined with 0.5 volumes of buffer containing 50% glycerol and 0.025% bromophenol blue, loaded into a gel (1.8% agarose) buffered with 1 × TAE (40 mM Tris-HCl, 20 mM acetic acid, 1 mM EDTA, pH 8.4), and separated at 100 V. Bands corresponding to the expanded CRISPR array were excised from the gel and subjected to purification. DNA fragment purification from agarose gel was performed using a GeneJET Gel Extraction Kit (Thermo Scientific), following the manufacturer's instructions.

Obtained DNA fragments corresponding to the expanded CRISPR array were sequenced using Miseq Illumina in pair-end 250-bp long-read mode, according to the manufacturer's protocols. Raw sequencing data were analyzed using ShortRead and

BioStrings packages (Pages *et al.*, 2012). Illumina-sequencing reads were filtered for quality scores of ≥20 and reads containing two repeats (with up to two mismatches) were selected. Analysis was performed as described earlier (Shmakov *et al.*, 2014, 2016). Reads that contained 33-bp sequences between two CRISPR repeats were next selected. The 33-bp segments were considered as spacers. For reads corresponding to multiple spacer acquisitions, only spacers that were acquired first were selected for further analysis to avoid biases in spacer composition caused by secondary priming. Spacers were next mapped onto the pG8mut or pG8mut_CCG plasmids with no mismatches allowed. Read mapping and spacer statistics analysis was performed with R (Shmakov *et al.*, 2014). Graphical representation was carried out using the EasyVisio tool developed by Ekaterina Rubtsova.

The Pearson coefficient, which is a measure of the linear dependence between two variables, was used to compare the pattern of spacer choice specificity between different experiments. A Pearson coefficient of 1 indicates total positive linear correlation, 0 indicates no linear correlation and −1 indicates total negative linear correlation.

### 5.14 Primer Extension Analysis

Oligonucleotides HS1_for pr/ext, HS1_rev pr/ext, HS2_for pr/ext or HS2_rev pr/ext were radiolabeled with γ-[$^{32}$P] ATP at the 5'-end with T4 PNK. Extension reactions (10 µl) were performed with 200 ng of total DNA as a template using 40 thermal cycles of 15 sec at 95 ºC, 30 sec at 50 ºC and 30 sec at 72 ºC. The reactions contained 1 pM of labeled primer, 0.2 mM dNTP, l × Taq-buffer, 0.8 U of Taq polymerase. As a marker, sequencing reactions with the same primer were set up on the purified pG8mut plasmid as a template using a Thermo Cycle Sequencing Kit (Affymetrix), following the manufacturer's instructions. The products were separated by denaturing (urea) 6% polyacrylamide gel and visualized using a Phosphorimager.

# Chapter 6. Results

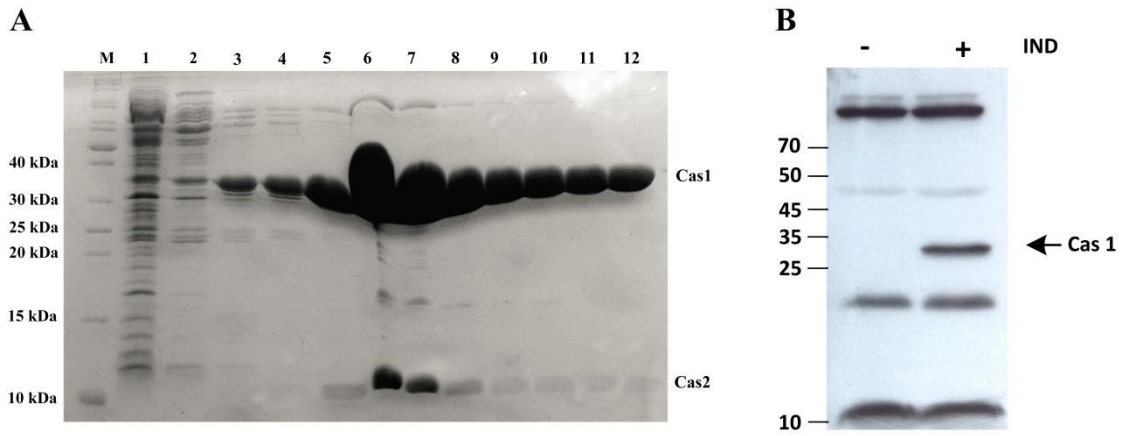
## PART 1

### 1.1 Cas1-Antibody Purification And Characterization

For assessment of adaptation complex DNA binding specificity, Cas1 and Cas2 proteins were purified to be used for subsequent immunization. For this purpose, the pET28-based plasmid pEK21 was constructed and used for protein expression in *E. coli* BL21(DE3) cells, and the protein complex was purified as described in "Materials and Methods". The result of SDS-PAGE analysis of the Cas1 and Cas2 proteins in eluted fractions is shown in Fig. 5A. Fractions 6 and 7 containing highest concentrations of Cas1 and Cas2 proteins were pooled and used to immunize rats. The antisera obtained were tested by Western blotting and purified in an affinity column containing immobilized recombinant Cas1. Antibodies specific for Cas2 could have been lost/depleted during this stage.

Reactivity of the antibody (1:5000 dilution) on a Western blot against proteins from whole cell extracts of KD263 *E. coli* cells capable of expression of full set of *cas* genes (see below) is shown in Fig. 5B, and highly specific labeling of Cas1 is seen.

Purified anti-Cas1 polyclonal antibody was used to precipitate proteins from *E. coli* cells with induced and uninduced expression of *cas* genes. The precipitated material was analyzed by SDS PAGE and gel slices containing proteins with apparent molecular mass of Cas2 (~11 kDa) were subjected to in-gel tryptic digestion and MALDI-TOF mass-spectrometry (performed by Dr. M. Serebrykova at the A.N. Belozersky Research Institute of Physico-Chemical Biology, Moscow State University).



**C**

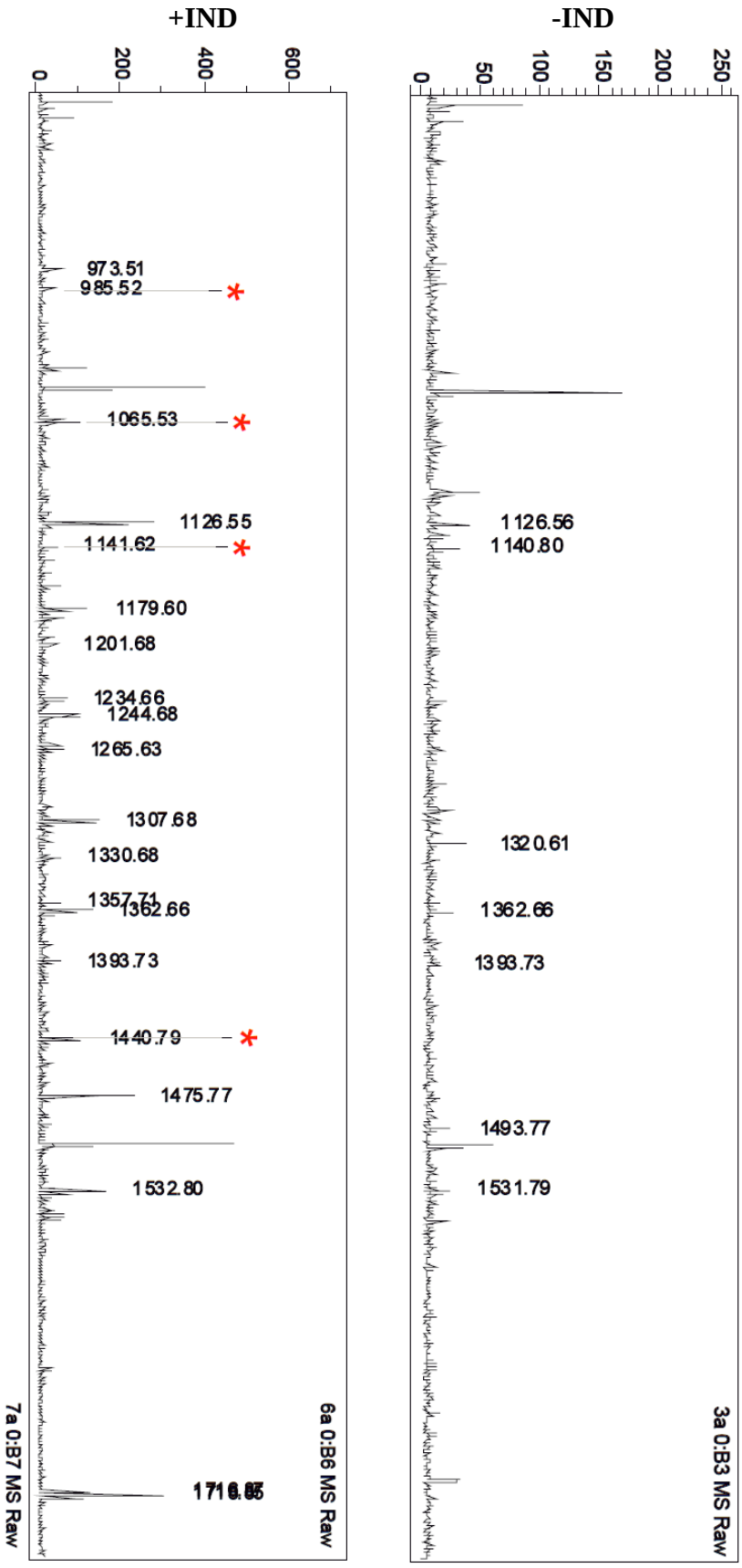

+IND
-IND
0
200
400
600
50
100
150
200
250
973.51
985.52
1065.53
1126.55
1141.62
1179.60
1201.68
1234.66
1244.68
1265.63
1307.68
1330.68
1357.71
1362.66
1393.73
1440.79
1475.77
1532.80
1126.56
1140.80
1320.61
1362.66
1393.73
1493.77
1531.79
6a 0:B6 MS Raw
7a 0:B7 MS Raw
3a 0:B3 MS Raw

**Fig. 5. Obtaining and characterizing anti-Cas1 antibody. (A) SDS gel showing final stages of affinity purification of hexahistidine-tagged Cas1 and untagged Cas2 from lysates of co-overexpressing cells. Fractions 6 and 7 were pooled and used for rat immunization. (B) Western blotting of proteins from *E. coli* KD263 lysates obtained in the presence (+) or in the absence (-) of *cas* gene expression with purified anti-Cas1 polyclonal antibody. (C) Mass spectrometric analysis of proteins precipitated from induced or uninduced KD263 cells. Mass peaks labeled with red asterisks match tryptic fragments of Cas2. The results of MS-MS analysis of one such peak (m/z=1440.7) are shown below.**

As can be seen from the mass-spectrometric data presented in Fig. 5C, pulled-down proteins from induced, but not from uninduced cells revealed mass-peaks matching tryptic fragments of Cas1, Cas2 and tandem MS analysis confirmed this identification. So the purified Cas1-antibody can pull down both Cas1 and Cas2 from induced *E. coli* cells.

## 1.2 The Cas1-Cas2 Complex Interacts With The CRISPR Array Leader Region In Cells Undergoing Primed Adaptation

The *E. coli* KD263 strain was used to study DNA specificity of the Cas1-Cas2 adaptation complex (Fig. 6A). The strain contains the *cas3* gene under the control of the *lacUV5* promoter and the *casABCDE12* operon under the *araBp8* promoter. The KD263 cells harbor a single genetically modified CRISPR array with two repeats and a single G8 spacer (Shmakov *et al.*, 2014). Induction of *cas* genes expression allows to control CRISPR adaptation and interference in KD263.

For investigation of the DNA specificity of the Cas1-Cas2 adaptation complex, uninduced KD263 cells were transformed with a pG8mut plasmid, which is a derivative of the pT7blue cloning vector containing the G8 protospacer (Datsenko *et al.*, 2012). The G8mut protospacer sequence harbored a C to T change at the +1 position, which decreases CRISPR interference mediated by the G8 spacer-containing crRNA and enables effectient primed spacer acquisition (Semenova *et al.*, 2011, Semenova *et al.*, 2016). As a control KD263 cells were transformed with empty pT7blue vector. Cells harboring these plasmids were grown without the antibiotic required for plasmid maintenance to allow plasmid loss in the process of CRISPR interference. As expected, there was expansion of the CRISPR array in cultures induced to *cas* genes expression, which harbored the protospacer plasmid but not the control vector (Fig. 6B). It has recently been reported that IHF, an architectural DNA binding protein that interacts with the AT-rich leader, is required for spacer acquisition (Nunez *et al.*, 2016). Indeed, no new spacers were acquired in cultures of pG8mut carrying AM7-7, a derivative of KD263 strain harboring the disrupted *ihfA* gene (Fig. 6B, lane 6).

The cultures analyzed in Fig. 6B were subjected to formaldehyde crosslinking followed by immunoprecipitation with polyclonal antibody raised against Cas1. The precipitated material was subjected to qPCR with a pairs of primers that amplified a 138-bp fragment spanning the CRISPR leader and a portion of the first repeat or a shorter, 34-bp, leader fragment (Fig. 6C). The fold enrichment between induced and uninduced cultures harboring each plasmid was determined. The results obtained with both 34-bp and 138-bp DNA amplified leader fragments were similar. There was no enrichment of leader DNA in the antibody-associated fraction from induced cells harboring pT7blue. By contrast, at least 8-fold enrichment was observed in cells

undergoing primed adaptation (Fig. 6C). The results proved highly reproducible in three biological replicates of the ChIP experiment (Fig. 7A). There was no enrichment of leader DNA in Cas1-antibody-associated fraction from *ihf* mutant cells (Fig. 4C).

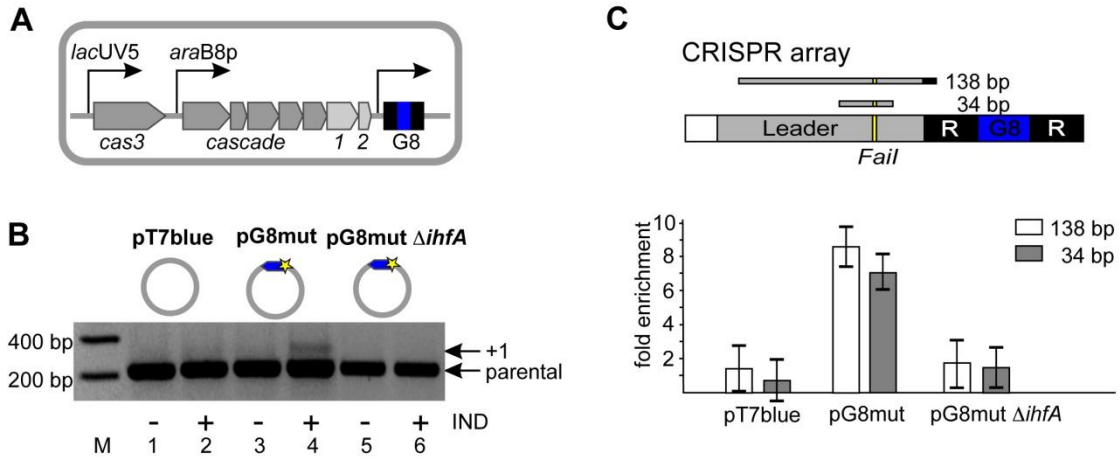



**Fig. 6. The Cas1-Cas2 complex is associated with the CRISPR array leader region during primed CRISPR adaptation. (A) The KD263 cell encoded for inducible *cas* genes and a single CRISPR array containing G8 spacer (blue) are schematically shown. (B) PCR analysis of CRISPR arrays from KD263 strain and AM7-7 strain carrying a deletion of the *ihfA* gene (*ΔihfA*) transformed with pT7blue vector or pG8mut plasmid. The G8 protospacer (blue) is shown by a blue arrow. C1T mismatch between the G8 crRNA spacer and G8 protospacer is shown by a yellow star. The additional band in line 4 corresponds to the expanded CRISPR array. (C) The KD263 CRISPR array and upstream leader region is schematically shown at the top. 138 bp and 34 bp DNA products of PCR amplification are shown. The internal *FaiI* recognition site is indicated. The results of Cas1 ChIP-qPCR analysis of cells shown in panel B are presented below. The qPCR data represent fold enrichment values corresponding to ratios of fragment concentrations in induced and uninduced cells. Mean values and standard deviations in triple technical replicates are shown.**

Taken together, the results suggest that the Cas1-Cas2 adaptation complex binds the CRISPR array only in cells undergoing CRISPR adaptation.

Observed enrichment of leader DNA precipitated from induced pG8mut harboring cells was not due to the effect of a different amount of proteins, because, as can be seen from Fig. 7B, the amount of Cas1 protein in induced KD263 cultures was

similar and did not depend on the presence of the protospacer plasmids (compare lanes 2 and 4).

To test the possibility of interaction of the Cas1-Cas2 adaptation complex with the leader region of the CRISPR array in the absence of primed adaptation, ChIP-qPCR experiments were performed with the *E. coli* KD27 strain (where only *cas1*, *cas2* and the CRISPR array are present) harboring the pG8mut plasmid (Fig. 7C). No adaptation was detected and there was no enrichment for the leader fragment in precipitated material (Fig. 7D). To exclude leader enrichment being determined by DNA interaction of free Cas1, the ChIP-qPCR procedure was performed with KD1 cells lacking *cas2* and harboring the pG8mut plasmid. There was no enrichment in this strain (Fig. 7D), despite the fact that the level of *cas1* expression was the same is in induced KD263 (Fig. 7C). Thus, the results obtained with the Cas1 antibody do indeed report on nucleic acids associated with the Cas1-Cas2 complex, not free Cas1.

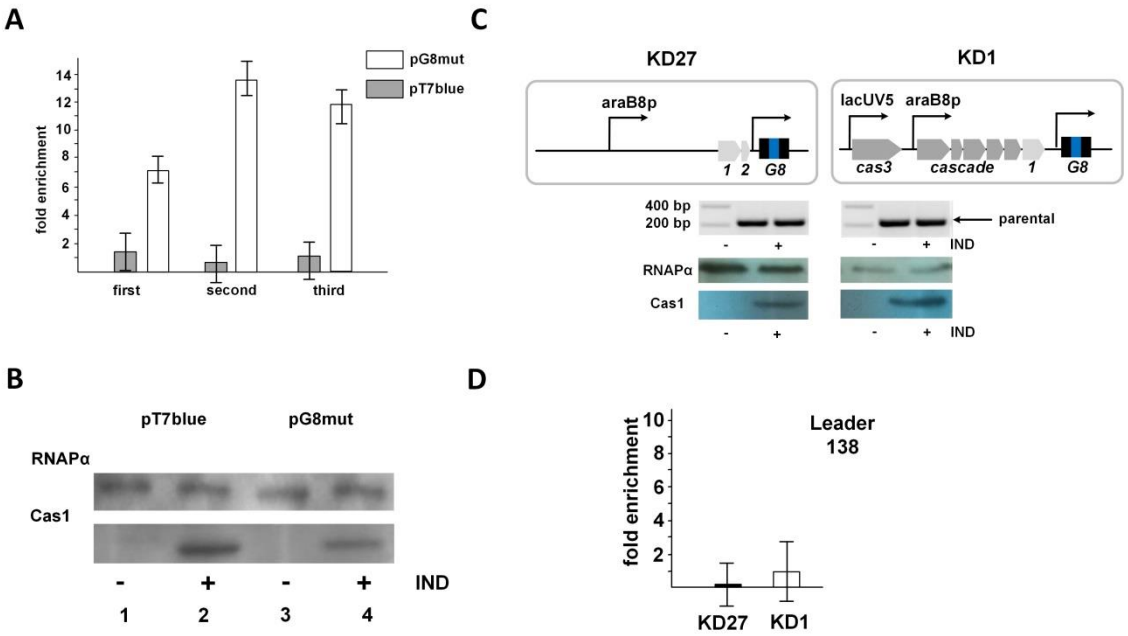



**Fig. 7. Verification of association of the Cas1-Cas2 complex with the CRISPR array leader region during primed CRISPR adaptation. (A) The results of three biological replicates of ChIP-qPCR experiments in KD263 cells transformed with pT7blue or pG8mut plasmid are consistent. (B) Western blotting of KD263 cultures transformed with indicated plasmids and grown in the presence or in the absence of *cas* gene inducers. Staining was performed with anti-Cas1 polyclonal antibody used for ChIP experiments (below) or with anti-RNA polymerase α**

**subunit antibody (loading control, above). (C) KD27 cells (lack *cas* genes, except *cas1* and *cas2*) and KD1 cells (lack *cas2*) are schematically shown. Transformation of these cells with plasmid pG8mut followed by *cas* genes induction does not lead to primed CRISPR adaptation and CRISPR array expansion. Western blotting of KD27 or KD1 carrying pG8mut plasmid and grown in the presence or in the absence of *cas* gene inducers shows Cas1 expression at the normal level. (D) ChIP-qPCR experiments with KD27 and KD1 *E. coli* cells containing pG8mut plasmid do not reveal enrichment of the leader fragment.**

Taken together, these data prove that the Cas1-Cas2 adaptation complex interacts with the leader region of the CRISPR array only in cells undergoing primed adaptation.

### 1.3 The Cas1-Cas2 Adaptation Complex Is Associated With Fragments Corresponding To Efficiently Used Protospacers

To identify the possibility of interaction of the Cas1-Cas2 adaptation complex with spacer precursor fragments, we first determined what spacers are preferentially acquired by KD263 cells harboring pG8mut plasmid during primed CRISPR adaptation. For this purpose we purified a DNA band corresponding to extended arrays from gel (Fig. 6B, lane 4) and subjected it to high-throughput sequencing (HTS) on an Illumina MiSeq platform.

Analysis of Illumina reads obtained showed that most of the newly acquired spacers corresponded to plasmid protospacers and matched the non-targeted strand of the plasmid (Table 2). That is consistent with spacer bias previously observed in primed adaptation experiments by the *E. coli* type I-E CRISPR-Cas system (Datsenko *et al.*, 2012; Savitskaya *et al.*, 2013; Semenova *et al.*, 2016). In total 811545 plasmid spacers were found. Most of the newly acquired spacers contained an AAG PAM (97%), and matched the reverse strand of the plasmid (88%) (Table 2).

Although protospacers with consensus PAM AAG gave rise to 97% of spacers, some of them behaved as adaptation “hot” spots and were preferentially used as a source of new spacers (Fig. 8A). No interdependence of functional plasmid parts, such as *ori*, f1 origin and Amp resistance gene, and the frequency of protospacer usage was detected. When the consensus AAG PAM of one such “hot” spot, Hot Spot 1 (HS1), was changed to CCG (plasmid pG8mut_CCG, Fig. 8B) no spacers corresponding to HS1 were acquired, while efficiency of other protospacers use as spacer donors was unaffected (Pearson correlation co-efficient of 0.95, p-value <2.2e-16). High correlation of spacer usage efficiency between pG8mut and pG8mut_CCG plasmids confirms that there exist some as yet unknown determinants of spacer choice that are unrelated to PAM preferences.

**Table 2. Characterization of spacers acquired in KD26 cells from pG8mut and pG8mut_CCG plasmids**

| | **Total amount of reads** | **Mapping on plasmid** | **Unique plasmid spacers** | **Mapping on forward strand** | **Mapping on reverse strand** | **AAG content** |
|---|---|---|---|---|---|---|
| **pG8mut** | 912835 | 811545 (89%) | 767 | 95487 (12%) | 716058 (88%) | 785743 (97%) |
| **pG8mut_CCG** | 58208 | 19134 (32%) | 433 | 2587 (14%) | 16547 (86%) | 18291 (96%) |

Five “hot” (HS1 through HS5) protospacers with different ability to produce spacers and one “cold” (CS) protospacer with a non-consensus ATG PAM were chosen to test specificity of the Cas1-Cas2 adaptation complex. Cas1-antibody precipitated DNA from *cas* induced and uninduced cultures harboring pG8mut or pT7blue plasmids was analyzed by qPCR with primer pairs that amplified 33 nucleotide-long sequences of each chosen protospacer. Strong enrichments were detected for HS amplicons (Fig. 8C). Values of fold enrichments observed for HS-containing amplicons ranged from ~17 for HS1 to ~6 for HS5, apparently reflecting the relative usage frequency of these protospacers as spacer donors. No significant enrichment was observed for the CS amplicon. Plasmid enrichment of HS1 amplicon was not observed in precipitated DNA obtained from KD263 cells containing pG8mut_CCG, while enrichment of HS2 amplicon was unaffected (Fig. 8C).

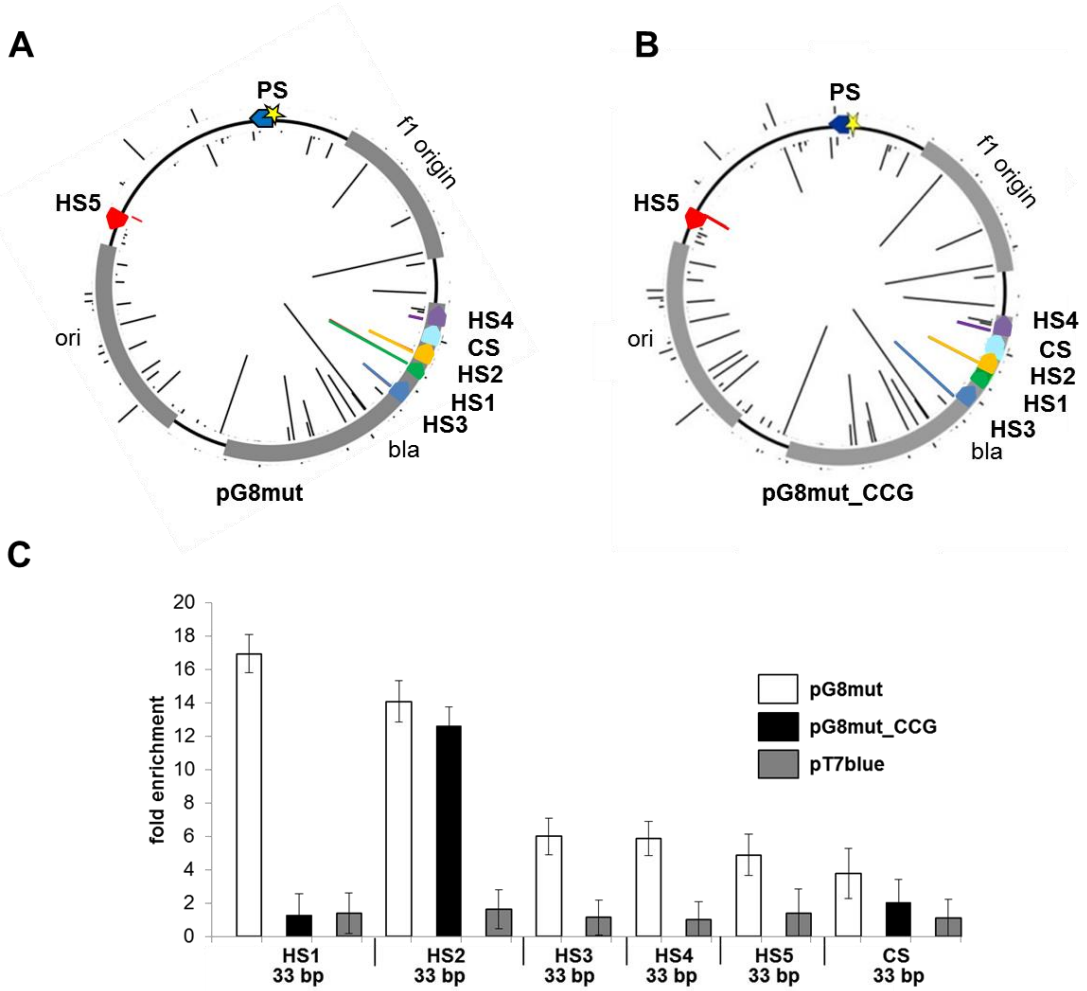



**Fig. 8. The Cas1-Cas2 adaptation complex is preferentially associated with fragments corresponding to efficiently acquired protospacers. Plasmids pG8mut (A) and pG8mut CCG (B) are shown. The height of bars corresponding to individual protospacers reflects the frequency of occurrence of the corresponding spacer in expanded arrays, and the position of the bars stands for start positions of corresponding protospacers. The priming G8 protospacer is shown as a blue arrow with an asterisk indicating a mismatch with the G8 crRNA spacer. CS and HS1-HS5 are, respectively, cold and hot protospacers. They are differently colored in the diagrams. In A, the bar indicating the efficiency of spacer acquisition from the HS1 protospacer is highlighted in green. No spacers from HS1 were acquired from pG8mut CCG (B). Grey rectangles marked "bla", "*ori*" and "f1 origin" show ampicillin resistance gene, plasmid and f1 replication origins, respectively. (C) Results of Cas1 ChIP analysis with primer pairs amplifying 33-bp fragments corresponding to HS1-5 and CS.**

A derivative of KD263 cells that lacks functional IHF and is unable to acquire new spacers but is functional for CRISPR interference was also tested. While enrichment of the leader fragment was not significant in this case (Fig. 6C), the level of enrichment for HS1 was even higher than in KD263 cells with functional IHF (~24 versus ~16 fold, Fig. 9A, bar 2). Thus, the Cas1-Cas2 adaptation complex is associated with protospacers and the degree of this association correlates with the efficiency of protospacers use as spacer donors.

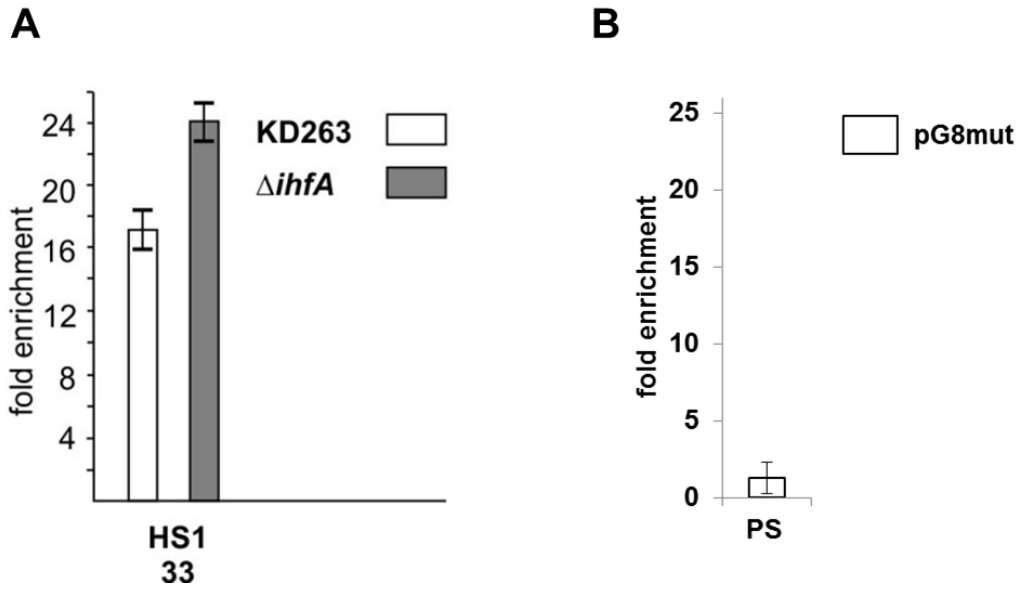



**Fig. 9. (A) IHF deficient cells show higher enrichment of the 33-bp HS1 fragment for AM7-7 cells than for the original KD263 strain. (B) Results of Cas1 ChIP-qPCR analysis with primer pairs amplifying 33-bp priming protospacer (PS) in KD 263 cells transformed with pG8mut plasmid.**

Recently, data have been presented suggesting that Cas1 and Cas2 may stimulate Cas3 recruitment to the priming site (Redding *et al.*, 2015). The ability of the Cas1-Cas2 adaptation complex to interact with the priming G8 protospacer in the pG8mut plasmid was therefore analyzed. However, no enrichment was observed when primers amplifying the 33-bp fragment of the priming protospacer were used (Fig. 9B), indicating that Cas1-Cas2 are not localized at the priming protspacer in our system.

To exclude the possibility that the observed results are due to the binding of free Cas1 instead of the Cas1-Cas2 complex, the ChIP-qPCR procedure was also performed with KD1 cells (lack *cas2*) containing the pG8mut plasmid. No enrichment was detected for the HS1 fragment (Fig. 10A, bar 1). Also, no enrichment of the HS1

fragment was observed for KD27 cells, which do not have Cascade and are incapable of primed adaptation and interference (Fig. 10A, bar 2).

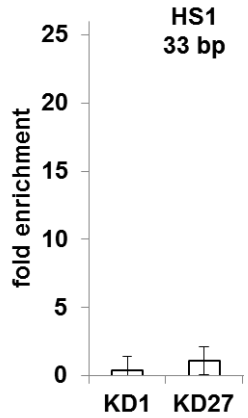



**Fig. 10. Results of ChIP-qPCR experiments with cells of the *E. coli* strains KD1 (lack *cas2*) and KD27 (only *cas1* and *cas2* genes) showed no enrichment of the HS1 fragment in precipitated DNA.**

Taken together, our data indicate that the Cas1-Cas2 complex is preferentially associated with those protospacers that are efficiently used for spacer incorporation during primed adaptation.

### 1.4 Protospacer Corresponding Fragments Bound By The Cas1-Cas2 Adaptation Complex Are Spacer-Sized Intermediates Of Primed Adaptation

In subsequent analyses we focused on target DNA fragments associated with the Cas1-Cas2 adaptation complex. To find the exact length of such fragments, precipitated DNA was probed with primer pairs amplifying longer HS1-containing plasmid fragments (Fig. 11A). The result showed that fold enrichment decreased when primer pairs amplifying 47-bp DNA fragments extended at either side of the HS1 protospacer were used. A longer, 61-bp fragment that contained HS1 in its center was not enriched in the precipitated fraction (Fig. 11B). A similar result was obtained with HS2 (Fig. 11C,D). The preferential amplification of short-length HS fragments from precipitated material suggests that they either represent intermediates of primed adaptation associated with the Cas1-Cas2 complex or they correspond to newly incorporated spacers of the CRISPR array.

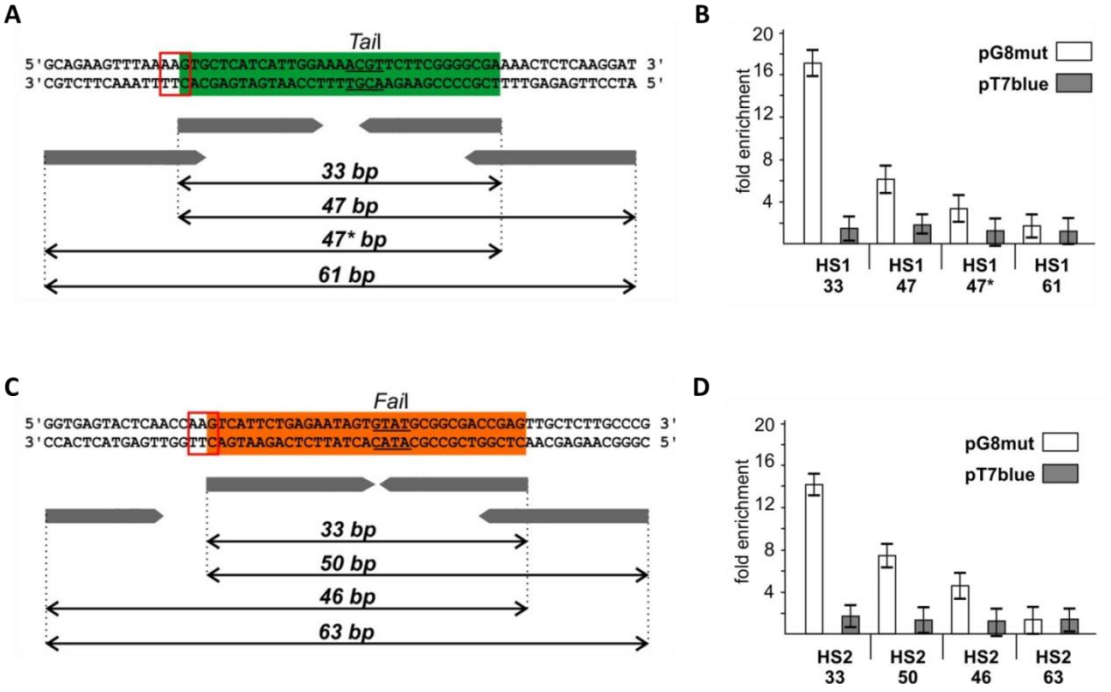



**Fig.11. The Cas1-Cas2 adaptation complex is associated with spacer-sized fragments. (A), (C) The HS1 and HS2 protospacers (green and orange respectively) in plasmid DNA context are shown. The AAG PAM is highlighted with a red box. The *Tai* or *FaiI* recognition sites are underlined. Primers used to amplify HS-containing fragments of different lengths are shown below. (B), (D)**

**Results of ChIP-qPCR analysis with primer pairs amplifying HS1 or HS2 fragments of different lengths.**

PCR with a HS specific primer and a primer annealing downstream of the CRISPR array was performed in order to test the possibility that enrichment of spacer-sized fragments in the Cas1-Cas2 bound DNA fraction arose due to amplification of spacers that have been already incorporated into the CRISPR array (Fig. 12A). Amplification of the leader region was used as control. A 138-bp leader amplicon was detected in induced cultures before and after precipitation with antibodies, as expected. Product containing the HS1 spacer and CRISPR repeat was detected in input material before precipitation with antibodies, but was absent even after 40 amplification cycles of DNA associated with the Cas1-Cas2 complex (Fig. 12B). The same result was obtained with HS2 (Fig. 12B).

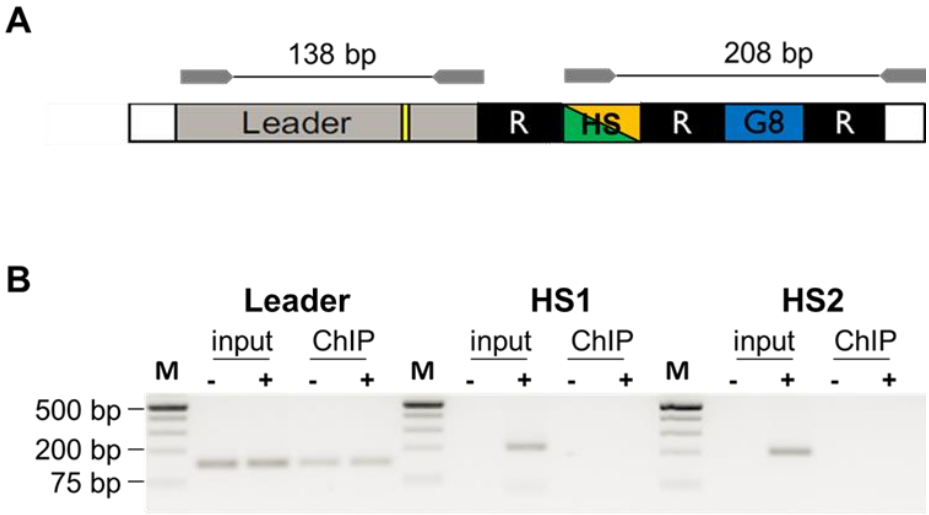



**Fig. 12. The Cas1-Cas2 adaptation complex is not associated with newly acquired spacers. (A) Schematic representation of the CRISPR array containing a newly acquired spacer. Primers amplifying the 138-bp fragment were used for detection of the leader region. Primers complementary to new spacers and a region downstream of the CRISPR array were used for detection of spacer integration. The resulting PCR products were 208-bp long. (B) Results of PCR amplification (40-cycles) of pre- and post- immunoprecipitation material from KD263 cells transformed with pG8mut plasmid.**

Thus, during the adaptation process the Cas1-Cas2 adaptation complex does not bind to spacers incorporated in the array and must therefore interact with spacer-sized fragments originating from target plasmid DNA.

### 1.5 Cas1-Cas2 Complex-Associated Protospacer-Corresponding DNA Fragments Are Not In A Standard Double-Stranded Form

The presence of restriction endonuclease recognition sites was used to assess the state of Cas1-Cas2 complex-associated DNA. Restriction endonucleases digests only double-stranded DNA, so this feature can be used to probe DNA structure. Restriction endonucleases *TaiI* and *FaiI* were chosen, as their recognition sites are present in HS1 (*TaiI*), HS2 (*FaiI*) and leader (*FaiI*) amplicons (Fig. 11A, Fig. 11C). The activity of restriction endonucleases was proven with appropriately sized model substrates (Fig. 13A). Fully double-stranded model 33-bp HS1 or HS2 double-stranded oligonucleotides were efficiently destroyed by, respectively, *TaiI* or *FaiI*, while corresponding single-stranded DNA oligonucleotides of the same size were fully resistant (Fig. 13A). As an additional control, input DNA was treated by *TaiI* (after reversal of cross-linking but before antibody precipitation) and then the effect of this digestion on a randomly selected 33-bp amplicon of *E. coli* genomic DNA containing a *TaiI* site was determined. The *TaiI* treatment prevented fragment amplification (Fig. 13A).

After immunoprecipitation, aliquots of Cas1-Cas2 complex–associated DNA were treated with *FaiI* or *TaiI*. qPCR analysis was performed with primer pairs amplifying the HS1 protospacer (33 nucleotide), the HS2 protospacer (33 nucleotide) and 34-nucleotide CRISPR array leader fragment. The treatment had no effect on enrichment of the 33-nucleotide HS1 fragment (Fig. 13B). Likewise, treatment with *FaiI* had no effect on enrichment of a 33-nucleotide DNA fragment for HS2, which contains an internal *FaiI* site (Fig. 13B). However, *FaiI* treatment abolished enrichment of a Cas1-Cas2 complex-associated 34-nucleotide leader amplicon (Fig. 13B), an expected result for double-stranded chromosomal DNA.

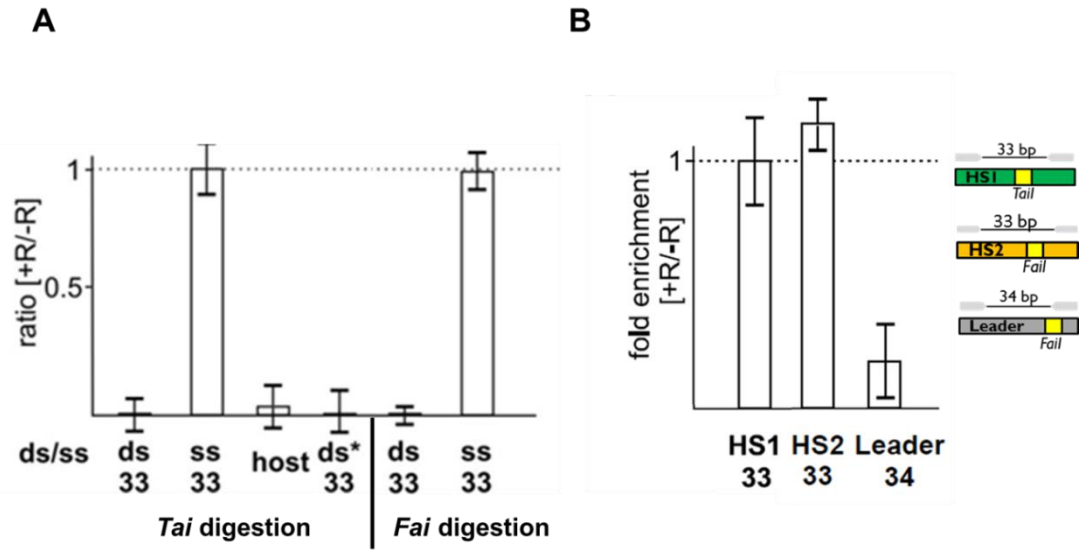



**Fig. 13. Cas1-Cas2 complex-associated protospacer-corresponding DNA fragments are not in a standard double-stranded form. (A) Samples were treated with restriction endonucleases and the ratios of apparent concentrations (calculated from Ct values) with and without treatment were calculated. Mean values and standard deviations for triple replicates are shown. Model double-stranded (ds), single-stranded (ss), partially double stranded HS1 (ds*) substrates or total *E. coli* DNA (host) were treated with *TaiI* or *FaiI* followed by qPCR analysis. For the bars labeled 'ds', 'ss', and 'ds*' qPCR analysis was performed with a primer pair amplifying the HS1 or HS2 protospacer when appreciated. Analysis with a primer pair amplifying a 33 bp genomic fragment containing an internal *TaiI* site was performed for the bar labeled 'host'. (B) KD263 cells transformed with pG8mut were processed for Cas1 ChIP-qPCR analysis. The ratios of fold enrichment values between *TaiI*-(for HS1) or *FaiI*- (forHS2 and leader) treated ('+R') and corresponding untreated ('−R') samples are presented.**

Taken together, these data show that target DNA fragments associated with the Cas1-Cas2 complex during primed adaptation are resistant to restriction endonuclease digestion and are therefore not in a standard double-stranded form. Structure of the *E. coli* Cas1-Cas2 adaptation complex bound to a partially double-stranded model substrate has recently been determined (Nuñez *et al.*, 2015b; Wang *et al.*, 2015). Similar substrates based on the HS1 protospacer were fully sensitive to *TaiI* digestion (Fig.13A). Thus, complex-associated fragments from cells undergoing primed

adaptation must be different from substrates bound to the adaptation complex in published structures.

## Summary 1

1. The Cas1-Cas2 complex interacts with the leader region of the CRISPR array during primed adaptation;
2. The Cas1-Cas2 complex is associated with spacer-sized fragments originated from a plasmid during primed adaptation and extent of association correlates with efficiency of adaptation of such fragments into CRISPR array;
3. Cas1-Cas2 complex-associated DNA plasmid fragments are not double-stranded.

## PART 2

### 2.1 Target DNA In Cells Undergoing CRISPR Interference And Primed Adaptation Acquires A Structure Sensitive To Single-Strand Specific Nuclease Digestion

To assess the state of target plasmid DNA in cells undergoing primed adaptation, S1 nuclease analysis was applied, since this enzyme is able to cleave single-stranded DNA fragments (Desai and Shankar, 2003). As shown by qPCR analysis, treatment with S1 nuclease in pilot experiments with model double-stranded DNA fragments of different lengths resulted in disappearance of short (up to 120 bp) double-stranded susbtrates but had no effect on longer fragments (Fig. 14). This suggests that the S1 nuclease treatment may be used for assessment of the state of DNA fragments exceeding 120 bp in length.

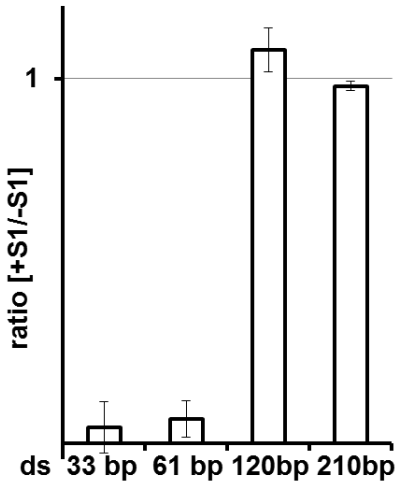



**Fig. 14. DNA fragments exceeding 120 bp in length might be used for S1 nuclease analysis. Results of qPCR analysis (Ct values) of model substrates treated or untreated with S1 nuclease. Ratios between treated and untreated samples are shown.**

Total DNA was purified from *cas* gene induced and uninduced KD263 cells carrying the pG8mut plasmid, which are capable of both primed adaptation and interference. Half of each sample was treated with S1 nuclease. qPCR was performed with primer pairs amplifying a 210 nucleotide plasmid fragment or a 200 nucleotide genomic fragment. As was shown above, fragments longer that 120 bp are fully resistant to S1 treatment if double-stranded (Fig. 14). The ratio of apparent

concentrations (calculated from Ct values) with and without S1 treatment was calculated for each substrate.

S1 treatment of DNA from uninduced cultures had no effect on the DNA amount measured by qPCR for both genomic and plasmid fragments (Fig. 15D). S1 treatment of DNA of induced KD263 cultures revealed sensitivity of the plasmid fragment to S1 digestion (Fig. 15D), while there was no effect on amplification of the 200 bp genome fragment (Fig.15D). This result suggests that a significant portion of plasmid originated DNA is present in single-stranded form in cells undergoing primed adaptation.

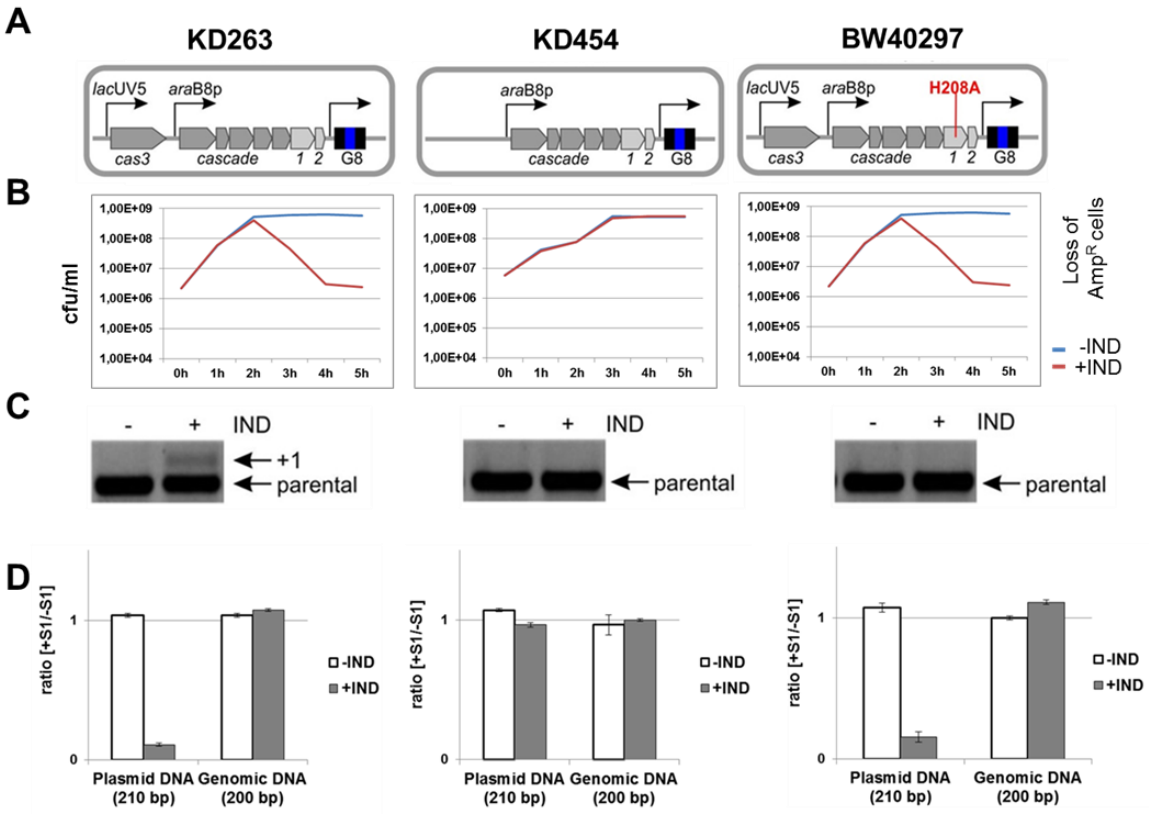



**Fig. 15. Target DNA in cells undergoing CRISPR interference and primed adaptation is sensitive to single-strand specific nuclease digestion. (A) The structures of CRISPR locuses in strains KD263, KD454, BW40297. (B) Growth curves reflecting the loss of ampicillin resistance in cells upon *cas* gene induction. (C) CRISPR adaptation in KD263, KD454, or BW40297 cells. DNA from cells transformed with pG8mut plasmid and grown with or without induction was purified. PCR was performed with oligonucleotide primers annealing upstream and downstream of the CRISPR array. "Parental" marks a band corresponding to**

**an unexpanded, parental CRISPR array. The "+1" band corresponds to CRISPR arrays expanded by one spacer-repeat unit. (D) Ratios between treated and untreated samples estimated by qPCR for genomic and plasmid corresponding locuses.**

In order to evaluate the contribution of CRISPR-Cas system components to the emergence of adaptation intermediates, KD454 (KD263 derivative lacking *cas3*) and BW40297 (no functional *cas1*) cells were transformed with the pG8mut plasmid and then *cas* gene expression was induced. In agreement with previous observations (Datsenko *et al.*, 2012; Semenova *et al.*, 2016), BW40297 cells were incapable of adaptation but proficient at interference, while cells of the KD454 strain showed impairment of both adaptation and interference (Fig. 15B,C). DNA from KD454 and BW40297 cultures was tested for S1 sensitivity. S1 treatment of DNA from KD454 cells had no effect on the ratio of the amount of 210 nt pG8mut amplicon (Fig. 15D). By contrast, results obtained with BW40297 cells DNA were similar to those obtained with KD263 cells (Fig. 15D). It therefore follows that extended-length S1-sensitive fragments of target DNA require functional Cas3 and do not require catalytically active Cas1.

These data imply that, during CRISPR interference or/and primed CRISPR adaptation, target DNA undergoes Cas3-mediated degradation accompanied by the production of long S1-sensitive DNA fragments.

## 2.2 Spacer-Sized Fragments Are Excised From One Strand Of Target DNA

As shown in previous sections, the Cas1-Cas2 adaptation complex is associated with short non-double-stranded DNA fragments that originate from target DNA, while S1-sensitive fragments generated by Cas3 nuclease are considerably longer. To determine how shorter fragments might be generated, total DNA from KD263 cells carrying the pG8mut plasmid was subjected to primer extension analysis with primers annealing upstream and downstream of the HS1 protospacer (Fig. 16A,B). No primer extension products were detected with a primer annealing upstream of HS1. By contrast, two distinct primer extension products were detected at the boundaries of the HS1 protospacer (and including the last G of the AAG PAM) with the downstream primer (anneals 42 bp away from the HS1 to a strand targeted by the priming protospacer) (Fig. 16B). A similar result was obtained for the HS2 protospacer (Fig. 16E,F). The primer extension products at protospacer boundaries were only detected in the presence of functional Cas3 and Cas1 (Fig. 16C). When primer extension reaction was conducted with DNA prepared from cultures harboring pG8mut CCG, no cleavage at the HS1 PAM was detected (Fig. 16D).

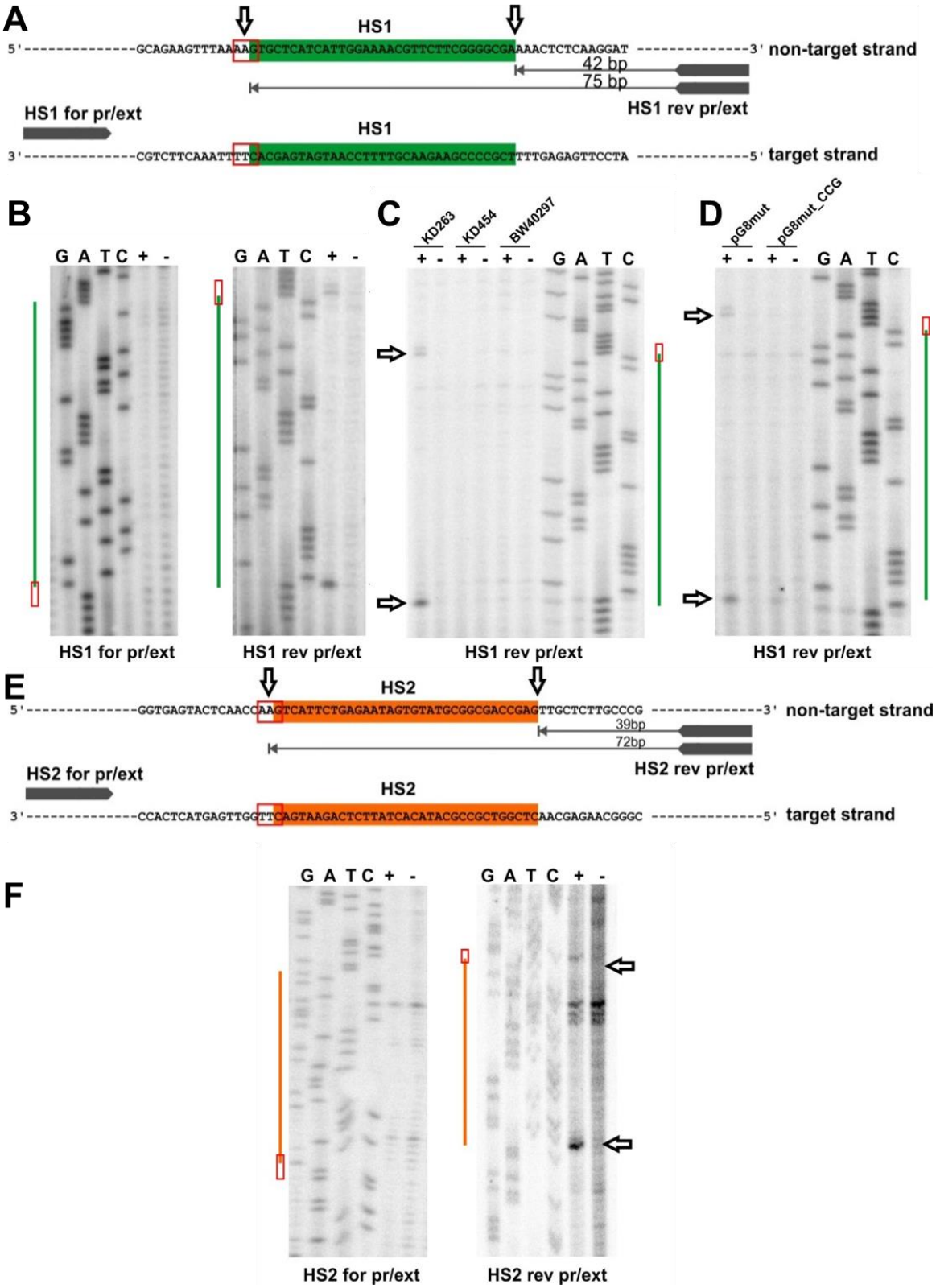



**Fig. 16. Single-stranded breaks are detected in target DNA. (A), (E) fragments of plasmid DNA containing the HS1 (green) or HS2 protospacer (orange) are shown and positions of annealing of primers used for primer extension reactions are**

**indicated. For HS1 rev pr/ext and HS2 rev pr/ext primer, left-oriented arrows indicate detection of primer extension products. (B), (F) Results of the primer extension experiment using total DNA purified from KD263 cells transformed with pG8mut with ('+') and without ('−') *cas* gene induction. The position of the HS1 protospacer (green) or HS2 protospacer (orange) and PAM (red box, see also A) is shown at the side of each autoradiogram. Horizontal arrows indicate the positions of migration of primer extension products on the gel. (C) As in B but using DNA prepared from indicated cells for primer extension with the HS1 rev pr/ext primer. (D) Primer extension reactions were performed with DNA prepared from KD263 cells transformed with pG8mut or pG8mut CCG plasmids.**

Interestingly, primer extension product corresponding to the downstream cleavage of HS1 protospacer was not strongly affected in the pG8mut_CCG. We propose that primer extension products mark the boundaries of protospacers excised by the Cas1-Cas2 adaptation complex from DNA intermediates generated by Cas3 and channeled for incorporation into the CRISPR array.

## Summary 2

1. Cas3 nuclease produces long single-stranded DNA fragments of target plasmid DNA during CRISPR interference and primed CRISPR adaptation;
2. Single-stranded breaks flanking hot protospacers occur in a non-target strand of plasmid from which primed adaptation occurs and require both Cas3 and the Cas1-Cas2 adaption complex.

## Chapter 5. Discussion

The process of CRISPR adaptation must consist of multiple steps. A protospacer in foreign DNA with a functional PAM must be selected, a spacer-sized fragment of foreign DNA must be generated, and, finally, the reaction of spacer incorporation in the leader-proximal end of the CRISPR array must occur. There has recently been significant progress in the understanding of late events of the spacer adaptation pathway (Amitai and Sorek, 2016; Sternberg *et al.*, 2016). However, the early events of the pathway remain poorly understood. During primed adaptation, protospacers located in *cis* to the priming protospacer bound by the effector complex must be selectively recognized and a strand bias in spacer acquisition must be somehow maintained while target DNA is destructed. In the work described here, we find evidence that the Cas1-Cas2 adaptation complex in *E. coli* cells undergoing primed adaptation is associated with spacer-sized fragments of plasmid DNA. These fragments are not in the standard double-stranded DNA form, as they are resistant to restriction endonuclease digestion. The abundance of Cas1-Cas2-associated fragments is correlated with efficiency of protospacer use as spacer donors. We propose that these fragments correspond to *in vivo* intermediates of the CRISPR adaptation pathway on their way to incorporation in the array. The Cas1-Cas2 adaptation complex can also be detected on the CRISPR array leader, but only at conditions of ongoing primed CRISPR adaptation.

The Cas1–Cas2 adaptation complex can either (i) itself generate spacer-sized DNA fragments from plasmids containing priming protospacers; (ii) rely on upstream interference machinery, specifically the Cas3 nuclease/helicase, to generate fragments ready for incorporation into the CRISPR array; or (iii) use the products of target DNA degradation by the interference machinery to generate the adaptation substrates.

Earlier data suggested that complex effector interactions with fully matching protospacers lead to CRISPR interference, while interactions with partially matching protospacers that abolish interference lead to primed adaptation (Datsenko *et al.*, 2012; Fineran *et al.*, 2014; Li *et al.*, 2014; Richter *et al.*, 2014; Semenova *et al.*, 2011). Accordingly, it was postulated that two structurally distinct types of complexes, (i) capable of Cas3 recruitment and interference and (ii) capable of Cas3, Cas1 and Cas2 recruitment and adaptation, are formed on, respectively, fully matching and partially matching protospacers (Redding *et al.*, 2015; Blosser *et al.*, 2015). Recent experiments

show, however, that complex Cascade effector interaction with mismatched priming protospacer targets causes interference, albeit at rates slower than those seen for mismatched targets (Xue *et al.*, 2015). Moreover, when the rates of degradation of matched and mismatched protospacer-carrying DNA are made equal, the former is actually much more efficient in promoting primed adaptation (Semenova *et al.*, 2016). The apparent lack of adaptation with matched targets could thus be a trivial consequence of their rapid destruction, which also eliminates Cas3 degradation products from which spacers are selected by Cas1–Cas2. In our work we analyzed primed adaptation in the presence of a partially mismatched protospacer and we demonstrated that the target plasmid undergoes degradation (Fig. 14B). Moreover, in the presence of active Cas3, a significant portion of plasmid DNA carrying the priming protospacer is present as extended (at least 200 nt) fragments that are sensitive to S1 nuclease digestion. These fragments must be generated by Cas3 after the Cascade effector complex recognizes the priming protospacer. Thus, our data are consistent with the suggestion that the products of target DNA degradation by the interference machinery are used to generate the adaptation substrates.

Primer extension analysis reveals nicks in the strand of plasmid DNA not-targeted by effector bound to the priming spacer that produce spacer-sized DNA fragments. We hypothesize that these fragments are the same as those associated with Cas1 in ChIP experiments. Detection of two primer extension products at each side of the protospacer in our experiments indicates that the first cut is introduced at the PAM (which is distal from the primer) followed by the second cut at another end of the protospacer. This is an expected scenario, since Cas1–Cas2 must first recognize a PAM (AAG in case of primed adaptation in *E. coli* type I-E system (Savitskaya *et al.*, 2013, Shmakov *et al.*, 2014) and then use a ruler-like mechanism to introduce another cut further downstream. The efficiency of selection of spacers from sequences associated with AAG PAMs varies by several orders or magnitude, with some protospacers behaving as “hot spots” (Savitskaya *et al.*, 2013). The reasons for such preferential use are not known but, clearly, cannot be determined by the PAM alone. It has been suggested that additional sequences at the other end of the protospacer can contribute to adaptation efficiency (Yosef *et al.*, 2013). Our analysis of a hot spot with mutated PAM

is consistent with this notion, since the downstream cut is maintained in the hot protospacer with PAM mutation, though spacer acquisition is abolished by the mutation.

A hypothetical mechanism of spacer acquisition that is consistent with our data is presented in Fig. 17. It is based on known properties of the effector complex (binding to mismatched protospacers and R-loop formation), preferential cleavage by the *E. coli* Cas3 of the non-targeted strand in the R-loop and its 3'-5' helicase activity, and the data obtained in this work.

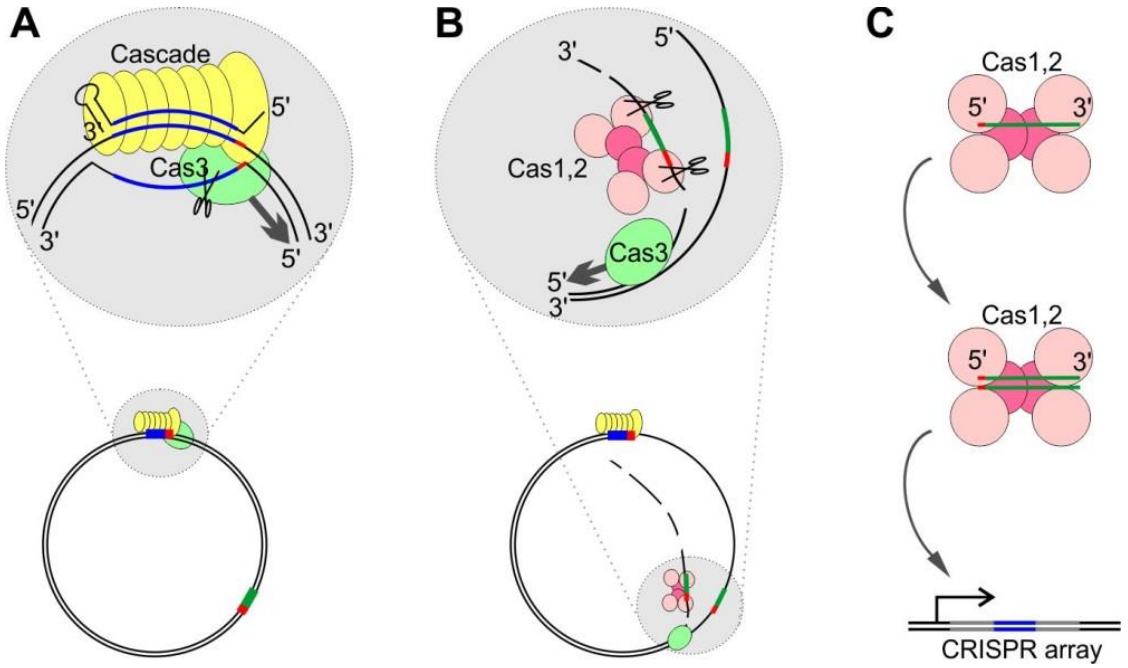



**Fig. 17. A model of primed adaptation by the *E. coli* type I-E CRISPR–Cas system. (A) The priming protospacer is recognized by the effector complex, leading to R-loop formation. (B) Cas3 processively unwinds and degrades the DNA strand that was not targeted by the priming complex, leading to formation of extended single-stranded fragments. The Cas1–Cas2 complex excises protospacers from fragments generated by Cas3. The Cas1–Cas2 complex may (as suggested by the *in vitro* findings of Redding *et al.*, 2015) or may not associate with Cas3. (C) The Cas1–Cas2 complex, first bound to non-double-stranded protospacers, is converted to an adaptation-proficient complex with fully or partially double-stranded protospacers that are inserted in the CRISPR array.**

The model posits that Cas3 processively unwinds target DNA moving along the non-targeted strand in the 3' to 5' direction. The Cas3 nuclease then generates extended-length S1 nuclease sensitive fragments from the non-target strand, from which

Cas1–Cas2 excise protospacers. These fragments are then channeled for integration into the CRISPR array. The major features of the model are consistent with recent results obtained in a reconstituted *in vitro* primed adaptation system, which showed that Cas3 generated partially single-stranded fragments fuel primed adaptation (Kunne *et al.*, 2016), though the S1-sensitivity of Cas3 generated fragments was not assessed in that work.

Published data convincingly show that *in vitro* Cas1–Cas2 use fully or partially double stranded substrates for incorporation in a CRISPR array (Nuñez *et al.*, 2015; Wang *et al.*, 2015; Kunne *et al.*, 2016). To make our data consistent with these observations, it is necessary to postulate that fragments of the target strand must reassociate with Cas1–Cas2 bound single-stranded protospacer DNA or that a special mechanism, which creates the second strand of Cas1–Cas2 bound protospacers, must exist. It has been suggested that the RecBCD nuclease-helicase is responsible for generating material for spacer acquisition during naïve adaptation (Levy *et al.*, 2015). Available data indicate that RecBCD generated fragments are also single-stranded (Dillingham and Kowalczykowski, 2008). Thus, whatever the mechanism responsible for generation of intermediates used for incorporation into a CRISPR array may be, it should be operative in both naïve and primed adaptation.

## References


1. Abudayyeh,O.O., Gootenberg,J.S., Konermann,S., Joung,J., Slaymaker,I.M., Cox,D.B.T., Shmakov,S., Makarova,K.S., Semenova,E., Minakhin,L., et al. (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science (80-. )., 353, aaf5573.
2. Amitai,G. and Sorek,R. (2016) CRISPR–Cas adaptation: insights into the mechanism of action. Nat. Rev. Microbiol., 14, 67–76.
3. Anders,C., Niewoehner,O., Duerst,A. and Jinek,M. (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature, 513, 569–573.
4. Andersson,A.F. and Banfield,J.F. (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. Science, 320, 1047–1050.
5. Arslan,Z., Stratmann,T., Wurm,R., Wagner,R., Schnetz,K. and Pul,Ü. (2013) RcsB-BglJ-mediated activation of Cascade operon does not induce the maturation of CRISPR RNAs in E. coli K12. RNA Biol., 10, 708–715.
6. Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, U. (2014) Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system, Nucleic Acids Res., 42, 7884-7893.
7. 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, E. V., Savchenko, A., Emili, A., Greenblatt, J., Edwards, A. M., and Yakunin, A. F. (2011) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair, Mol. Microbiol., 79, 484-502.
8. Battle,S.E., Meyer,F., Rello,J., Kung,V.L. and Hauser,A.R. (2008) Hybrid pathogenicity island PAGI-5 contributes to the highly virulent phenotype of a Pseudomonas aeruginosa isolate in mammals. J Bacteriol., 190, 7130-7140.
9. Barrangou,R., Fremaux,C., Deveau,H., Richards,M., Boyaval,P., Moineau,S., Romero,D.A. and Horvath,P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science, 315, 1709–1712.

10. Beloglazova,N., Lemak,S., Flick,R. and Yakunin,A.F. (2015) Analysis of nuclease activity of Cas1 proteins against complex DNA substrates. Methods Mol. Biol., 1311, 251–264.
11. Bergsland,K.,J., Kao,C., Yu,Y.,T., Gulati,R., Snyder,L. (1990) A site in the T4 bacteriophage major head protein gene that can promote the inhibition of all translation in Escherichia coli. J Mol Biol., 213, 477-94.
12. Bidnenko,E., Chopin,A., Ehrlich,S.,D., Chopin,M.,C. (2009) Activation of mRNA translation by phage protein and low temperature: the case of Lactococcus lactis abortive infection system AbiD1. BMC Mol Biol., 27,10-14.
13. Bingham,R., Ekunwe,S.,I., 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.
14. Bikard,D., Euler,C.W., Jiang,W., Nussenzweig,P.M., Goldberg,G.W., Duportet,X., Fischetti,V.A. and Marraffini,L.A. (2014) Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nat. Biotechnol., 32, 1146–1150.
15. Bikard,D. and Marraffini,L.A. (2013) Control of gene expression by CRISPR-Cas systems. F1000Prime Rep., 5.
16. Blosser,T.R., Loeff,L., Westra,E.R., Vlot,M., Künne,T., Sobota,M., Dekker,C., Brouns,S.J.J. and Joo,C. (2015) Two distinct DNA binding modes guide dual roles of a CRISPR-cas protein complex. Mol. Cell, 58, 60–70.
17. Bolotin,A., Quinquis,B., Sorokin,A. and Dusko Ehrlich,S. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology, 151, 2551–2561.
18. Bondy-Denomy,J., Garcia,B., Strum,S., Du,M., Rollins,M.F., Hidalgo-Reyes,Y., Wiedenheft,B., Maxwell,K.L. and Davidson,A.R. (2015) Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. Nature, 526, 136–139.
19. Bondy-Denomy,J., Pawluk,A., Maxwell,K.L. and Davidson,A.R. (2012) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature, 493, 429–432.

20. Borchardt,E.K., Vandoros,L.A., Huang,M., Lackey,P.E., Marzluff,W.F. and Asokan,A. (2015) Controlling mRNA stability and translation with the CRISPR endoribonuclease Csy4. RNA, 21, 1921–30.
21. Brouns,S.J.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J.H., Snijders,A.P.L., Dickman,M.J., Makarova,K.S., Koonin,E. V and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science, 321, 960–4.
22. Cheng,A.W., Wang,H., Yang,H., Shi,L., Katz,Y., Theunissen,T.W., Rangarajan,S., Shivalila,C.S., Dadon,D.B. and Jaenisch,R. (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res., 23, 1163–1171.
23. Cady,K.C., Bondy-Denomy,J., Heussler,G.E., Davidson,A.R. and O'Toole,G.A. (2012) The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. J. Bacteriol., 194, 5728–5738.
24. Caliando,B.J. and Voigt,C.A. (2015) Targeted DNA degradation using a CRISPR device stably carried in the host genome. Nat. Commun., 6, 6989.
25. Carte,J., Wang,R., Li,H., Terns,R.M. and Terns,M.P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev., 22, 3489–3496.
26. Chopin,M.,C., Chopin,A., Bidnenko,E. (2005) Phage abortive infection in lactococci: variations on a theme. Curr Opin Microbiol.,8,473–479.
27. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A., et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. Science, 339, 819–23.
28. Datsenko,K.A., Pougach,K., Tikhonov,A., Wanner,B.L., Severinov,K. and Semenova,E. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun., 3, 945.
29. Delannoy,C.M.J., Zadoks,R.N., Lainson,F.A., Ferguson,H.W., Crumlish,M., Turnbull,J.F. and Fontaine,M.C. (2012) Draft genome sequence of a nonhemolytic fish-pathogenic streptococcus agalactiae strain. J. Bacteriol., 194, 6341–6342.

30. Deltcheva,E., Chylinski,K., Sharma,C.M., Gonzales,K., Chao,Y., Pirzada,Z.A., Eckert,M.R., Vogel,J. and Charpentier,E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature, 471, 602–7.
31. Deng,L., Garrett,R.A., Shah,S.A., Peng,X. and She,Q. (2013) A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. Mol. Microbiol., 87, 1088–1099.
32. Desai,N.A. and Shankar,V. (2003) Single-strand-specific nucleases. FEMS Microbiol. Rev., 26, 457–491.
33. Deveau,H., Barrangou,R., Garneau,J.E., Labonté,J., Fremaux,C., Boyaval,P., Romero,D.A., Horvath,P. and Moineau,S. (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J. Bacteriol., 190, 1390–1400.
34. Díez-Villaseñor,C., Guzmán,N.M., Almendros,C., García-Martínez,J. and Mojica,F.J.M. (2013) CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli . RNA Biol., 10, 792–802.
35. Dillingham,M.S. and Kowalczykowski,S.C. (2008) RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. Microbiol. Mol. Biol. Rev., 72, 642–671.
36. Domingues,S., Chopin,A., Ehrlich,S.D. and Chopin,M.C. (2004) A Phage Protein Confers Resistance to the Lactococcal Abortive Infection Mechanism AbiP. J. Bacteriol., 186, 3278–3281.
37. Durmaz,E. and Klaenhammer,T.R. (2007) Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected Lactococcus lactis. In Journal of Bacteriology.Vol. 189, pp. 1417–1425.
38. Dryden,D.,T.,F., Murray,N.,E., Rao,D.,N. (2001) Nucleoside triphosphate-dependent restriction enzymes. Nucleic Acids Res., 29, 3728–3741.
39. Dy,R.,L., Przybilski,R., Semeijn,K., Salmond,G.,P., Fineran,P.,C. (2014) A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. Nucleic Acids Res. 2014 Apr;42(7):4590-605.

40. Erdmann,S. and Garrett,R.,A. (2012) Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol Microbiol., 85, 1044-56.
41. Fabre,L., Zhang,J., Guigon,G., Le Hello,S., Guibert,V., Accou-Demartin,M., de Romans,S., Lim,C., Roux,C., Passet,V., et al. (2012) CRISPR typing and subtyping for improved Laboratory surveillance of Salmonella infections. PLoS One, 7.
42. Fineran,P.C., Blower,T.R., Foulds,I.J., Humphreys,D.P., Lilley,K.S. and Salmond,G.P.C. (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc. Natl. Acad. Sci., 106, 894–899.
43. Fineran,P.C. and Charpentier,E. (2012) Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information. Virology, 434, 202–209.
44. Fineran,P.C., Gerritzen,M.J.H., Suarez-Diez,M., Kunne,T., Boekhorst,J., van Hijum,S.A.F.T., Staals,R.H.J. and Brouns,S.J.J. (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. Proc. Natl. Acad. Sci., 111, E1629–E1638.
45. Gasiunas,G., Barrangou,R., Horvath,P. and Siksnys,V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci., 109, E2579–E2586.
46. Garrett,R., Shah,S., Erdmann,S., Liu,G., Mousaei,M., León-Sobrino,C., Peng,W., Gudbergsdottir,S., Deng,L., Vestergaard,G., et al. (2015) CRISPR-Cas Adaptive Immune Systems of the Sulfolobales: Unravelling Their Complexity and Diversity. Life, 5, 783–817.
47. Garneau,J.E., Dupuis,M.-È., Villion,M., Romero,D.A., Barrangou,R., Boyaval,P., Fremaux,C., Horvath,P., Magadán,A.H. and Moineau,S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature, 468, 67–71.
48. Gerdes,K., Rasmussen,P.,B., Molin,S. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells, Proc. Natl Acad. Sci. USA, 1986, 83, 3116-3120.

49. Gerdes,K., Christensen,S.K. and Løbner-Olesen,A. (2005) Prokaryotic toxin-antitoxin stress response loci. Nat. Rev. Microbiol., 3, 371–382.

50. Gilbert,L.A, Larson,M.H., Morsut,L., Liu,Z., Gloria,A., Torres,S.E., Stern-ginossar,N., Brandman,O., Whitehead,H., Doudna,J. a, et al. (2013) CRISPR-Mediated Modular RNA-Guided Regualtion of Transcription in Eukaryotes. Cell, 154, 442–451.

51. Gomaa, A. A., Klumpe, H. E., Luo, M. L., Selle, K., Barrangou, R., and Beisel, C. L. (2014) Programmable removal of bacterial strains by use of genometargeting CRISPRCas systems, MBio, 5, e0092813.

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

53. Gong,B., Shin,M., Sun,J., Jung,C.-H., Bolt,E.L., van der Oost,J. and Kim,J.-S. (2014) Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. Proc. Natl. Acad. Sci., 111, 16359–16364.

54. Granvogl,B., Plöscher,M. and Eichacker,L.A. (2007) Sample preparation by in-gel digestion for mass spectrometry-based proteomics. Anal. Bioanal. Chem., 389, 991–1002.

55. Goren,M.G., Yosef,I., Auster,O. and Qimron,U. (2012) Experimental definition of a Clustered Regularly Interspaced Short Palindromic duplicon In Escherichia coli. J. Mol. Biol., 423, 14–16.

56. Haaber,J., Moineau,S. and Hammer,K. (2009) Activation and transfer of the chromosomal phage resistance mechanism AbiV in Lactococcus lactis. Appl. Environ. Microbiol., 75, 3358–3361.

57. Hale,C.R., Zhao,P., Olson,S., Duff,M.O., Graveley,B.R., Wells,L., Terns,R.M. and Terns,M.P. (2009) RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. Cell, 139, 945–956.

58. Hattman, S. (1999) Unusual transcriptional and translational regulation of the bacteriophage Mu mom operon. Pharmacol. Ther., 84, 367–388.

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

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

61. He,W., Huang,T., Tang,Y., Liu,Y., Wu,X., Chen,S., Chan,W., Wang,Y., Liu,X., Chen,S., et al. (2015) Regulation of DNA phosphorothioate modification in Salmonella enterica by DndB. Sci. Rep., 5, 12368.

62. Heler,R., Samai,P., Modell,J.W., Weiner,C., Goldberg,G.W., Bikard,D. and Marraffini,L.A. (2015) Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature, 519, 199-202.

63. Hochstrasser,M.L., Taylor,D.W., Bhat,P., Guegler,C.K., Sternberg,S.H., Nogales,E. and Doudna,J.A. (2014) CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc. Natl. Acad. Sci., 111, 6618–6623.

64. van Houte,S., Ekroth,A.K., Broniewski,J.M., Chabas,H., Ashby,B., Bondy-Denomy,J., Gandon,S., Boots,M., Paterson,S., Buckling,A. and Westra,E.R. (2016) The diversity-generating benefits of a prokaryotic adaptive immune system. Nature, 532, 385-8.

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

66. Hou,Z., Zhang,Y., Propson,N.E., Howden,S.E., Chu,L., Sontheimer,E.J. and Thomson,J.A. (2013) Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc. Natl. Acad. Sci. USA, 110, 15644–15649.

67. Huo,Y., Nam,K. H., Ding,F., Lee,H., Wu,L., Xiao,Y., Farchione,M. D., Zhou,S., Rajashankar,K., Kurinov,I., Zhang, R., and Ke, A. (2014) Structures of CRISPR Cas3 offer mechanistic insights into Cascadeactivated DNA unwinding and degradation, Nat. Struct. Mol. Biol., 21, 771-777.

68. Ishino Y., Shinagawa H., Makino K., Amemura M.,N.A. (1988) Nucleotide-sequence of the iap gene, responsible for alkaline-phosphatase isozyme

conversion in Escherichia coli, and identification of the gene-product. J. Bacteriol., 169 (12), 5429–33.

69. Ivančić-Bace,I., Cass,S.D., Wearne,S.J. and Bolt,E.L. (2015) Different genome stability proteins underpin primed and naïve adaptation in E. coli CRISPR-Cas immunity. Nucleic Acids Res., 43, 10821–10830.
70. Jackson,R.N., Lavin,M., Carter,J. and Wiedenheft,B. (2014) Fitting CRISPR-associated Cas3 into the Helicase Family Tree. Curr. Opin. Struct. Biol., 24, 106–114.
71. Jackson,S.A., McKenzie,R.E., Fagerlund,R.D., Kieper,S.N., Fineran,P.C. and Brouns,S.J. (2017) CRISPR-Cas: Adapting to change. Science, 356, 6333.
72. Jansen,R., Van Embden,J.D.A., Gaastra,W. and Schouls,L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol., 43, 1565–1575.
73. Jiang,W., Zhou,H., Bi,H., Fromm,M., Yang,B. and Weeks,D.P. (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res., 41.
74. Jiang,W. and Marraffini,L.A. (2015) CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. Annu. Rev. Microbiol., 69, 209–228.
75. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A Programmable Dual-RNA – Guided. Science (80-. )., 337, 816–822.
76. Jore,M.M., Lundgren,M., van Duijn,E., Bultema,J.B., Westra,E.R., Waghmare,S.P., Wiedenheft,B., Pul,U., Wurm,R., Wagner,R., et al. (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol., 18, 529–536.
77. Kahramanoglou,C., Seshasayee,A.S.N., Prieto,A.I., Ibberson,D., Schmidt,S., Zimmermann,J., Benes,V., Fraser,G.M. and Luscombe,N.M. (2010) Direct and indirect effects of H-NS and Fis on global gene expression control in Escherichia coli. Nucleic Acids Res., 10.1093/nar/gkq934.
78. Kamerbeek,J., Schouls,L., Kolk,A., van Agterveld,M., van Soolingen,D., Kuijper,S., Bunschoten,A., Molhuizen,H., Shaw,R., Goyal,M., et al. (1997)

Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J. Clin. Microbiol., 35, 907–14.

79. Kazlauskiene,M., Tamulaitis,G., Kostiuk,G., Venclovas,Č., Siksnys,V. (2016) Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. Mol Cell., 62, 295-306.
80. Kim,H., Ishidate,T., Ghanta,K.S., Seth,M., Conte,D.Jr., Shirayama,M. and Mello,C.C. (2014) A co-CRISPR strategy for efficient genome editing in Caenorhabditis elegans. Genetics, 197-1069-80.
81. Kojima,K.,K. and Kanehisa,M. (2008) Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture, Mol. Biol. Evol., 25, 1395-1404.
82. Kondo,S. and Ueda,R. (2013) Highly Improved gene targeting by germline-specific Cas9 expression in Drosophila. Genetics, 195, 715–721.
83. Koonin,E. V and Makarova,K.S. (2013) CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. RNA Biol., 10, 679–86.
84. Koonin,E. V. and Krupovic,M. (2015) Evolution of adaptive immunity from transposable elements combined with innate immune systems. Nat. Rev. Genet., 16, 184–192.
85. Koonin,E.V. (2017) Evolution of RNA- and DNA-guided antivirus defense systems in prokaryotes and eukaryotes: common ancestry vs convergence. Biol Direct., 12, 5.
86. Koonin,E.V. and Zhang,F. (2017) Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. Bioessays., 39, 1-9.
87. Kunne,T., Kieper,S.N., Bannenberg,J.W., Vogel,A.I.M., Miellet,W.R., Klein,M., Depken,M., Suarez-Diez,M. and Brouns,S.J.J. (2016) Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. Mol. Cell, 63, 852–864.
88. Labrie,S.,J., Samson,J.,E., Moineau,S. (2010) Bacteriophage resistance mechanisms. Nat Rev Microbiol.,8,317–327.
89. Labrie,S.,J., Tremblay,D.,M., Moisan,M., Villion,M., Magadán,A.,H., Campanacci,V., Cambillau,C., Moineau,S. (2012) Involvement of the major capsid protein and two early-expressed phage genes in the activity of the

lactococcal abortive infection mechanism AbiT. Appl Environ Microbiol., 2012, 78, 6890-9.

90. Lee,H.Y., Haurwitz,R.E., Apffel,A., Zhou,K., Smart,B., Wenger,C.D., Laderman,S., Bruhn,L. and Doudna,J.A. (2013) RNA-protein analysis using a conditional CRISPR nuclease. Proc. Natl. Acad. Sci., 110, 5416–5421.
91. Levy,A., Goren,M.G., Yosef,I., Auster,O., Manor,M., Amitai,G., Edgar,R., Qimron,U. and Sorek,R. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature, 520, 505–510.
92. Li,M., Wang,R., Zhao,D., Xiang,H. (2014) Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process. Nucleic Acids Res., 42, 2483-92.
93. Lintner,N.G., Kerou,M., Brumfield,S.K., Graham,S., Liu,H., Naismith,J.H., Sdano,M., Peng,N., She,Q., Copié,V., et al. (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). J. Biol. Chem., 286, 21643–21656.
94. Liu,T., Li,Y., Wang,X., Ye,Q., Li,H., Liang,Y., She,Q. and Peng,N. (2015) Transcriptional regulator-mediated activation of adaptation genes triggers CRISPR de novo spacer acquisition. Nucleic Acids Res., 43, 1044–1055.
95. Lopez-Sanchez,M.J., Sauvage,E., Da Cunha,V., Clermont,D., Ratsima Hariniaina,E., Gonzalez-Zorn,B., Poyart,C., Rosinski-Chupin,I. and Glaser,P. (2012) The highly dynamic CRISPR1 system of Streptococcus agalactiae controls the diversity of its mobilome. Mol. Microbiol., 85, 1057–1071.
96. Luo,M.L., Mullis,A.S., Leenay,R.T. and Beisel,C.L. (2015) Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. Nucleic Acids Res., 43, 674–681.
97. Luria,S.,E. and Human,M.,L. (1952) A nonhereditary, host-induced variation of bacterial viruses. J Bacteriol, 64, 557-69.
98. Makarova,K.,S., Grishin,N.,V., Shabalina,S.,A., Wolf,Y.,I. and Koonin,E.,V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional

analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct., 16, 1-7.

99. Makarova,K.S., Haft,D.H., Barrangou,R., Brouns,S.J.J., Charpentier,E., Horvath,P., Moineau,S., Mojica,F.J.M., Wolf,Y.I., Yakunin,A.F., et al. (2011) Evolution and Classification of the CRISPR-Cas Systems. Nat. Rev. Microbiol., 9, 467–77.
100. Makarova,K.S., Wolf,Y.I. and Koonin,E. V. (2013) Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res., 41, 4360–4377.
101. Makarova,K.S., Wolf,Y.I., Alkhnbashi,O.S., Costa,F., Shah,S.A., Saunders,S.J., Barrangou,R., Brouns,S.J.J., Charpentier,E., Haft,D.H., et al. (2015) An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol., 13, 722–736.
102. Mali,P., Aach,J., Stranges,P.B., Esvelt,K.M., Moosburner,M., Kosuri,S., Yang,L. and Church,G.M. (2013) Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol., 31, 833–838.
103. Marraffini,L.A. and Sontheimer,E.J. (2008) CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. Science (80-. )., 322, 1843–1845.
104. Marraffini,L.A. and Sontheimer,E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature, 463, 568–571.
105. Maxwell,K.L. (2016) Phages Fight Back: Inactivation of the CRISPR-Cas Bacterial Immune System by Anti-CRISPR Proteins. PLoS Pathog., 12, e1005282.
106. Mojica,F.J.M., Díez-Villaseñor,C., García-Martínez,J. and Soria,E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol., 60, 174–182.
107. Mizuuchi,K. and Adzuma,K. (1991) Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: evidence for a one-step transesterification mechanism. Cell, 66, 129–140.

108. Mulepati,S. and Bailey,S. (2013) In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA Target. J. Biol. Chem., 288, 22184–22192.

109. Nam,K.,H., Ding,F., Haitjema,C., Huang,Q., DeLisa,M.,P. and Ke,A. (2012) Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. J Biol Chem., 287, 35943-35952.

110. Nuñez,J.K., Kranzusch,P.J., Noeske,J., Wright,A. V, Davies,C.W. and Doudna,J.A. (2014) Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. Nat. Struct. Mol. Biol., 21, 528–534.

111. Nuñez,J.K., Lee,A.S.Y., Engelman,A. and Doudna,J.A. (2015) Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. Nature, 519, 193–198.

112. Nuñez,J.K., Harrington,L.B., Kranzusch,P.J., Engelman,A.N. and Doudna,J.A. (2015). Foreign DNA capture during CRISPR–Cas adaptive immunity. Nature, 527, 535–538.

113. Nuñez,J.K., Bai,L., Harrington,L.B., Hinder,T.L. and Doudna,J.A. (2016) CRISPR Immunological Memory Requires a Host Factor for Specificity. Mol. Cell, 62, 824–833.

114. Ogura,T. and Hiraga,S. (1983) Mini-F plasmid genes that couple host cell division to plasmid proliferation, Proc. Natl Acad. Sci. USA,80, 4784-4788.

115. van der Oost,J., Jore,M.M., Westra,E.R., Lundgren,M. and Brouns,S.J.J. (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem. Sci., 34, 401–407.

116. Paez-Espino,D., Morovic,W., Sun,C.L., Thomas,B.C., Ueda,K., Stahl,B., Barrangou,R. and Banfield,J.F. (2013) Strong bias in the bacterial CRISPR elements that confer immunity to phage. Nat. Commun., 4, 1430.

117. Paez-Espino,D., Sharon,I., Morovic,W., Stahl,B., Thomas,B.C., Barrangou,R. and Banfielda,J.F. (2015) CRISPR immunity drives rapid phage genome evolution in streptococcus thermophilus. MBio, 6, 1–9.

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

119. Pawluk,A., Bondy-Denomy,J., Cheung,V.H.W., Maxwell,K.L. and Davidson,A.R. (2014) A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of pseudomonas aeruginosa. MBio, 5.

120. Pingoud,A., Fuxreiter,M., Pingoud,V., Wende,W. (2005) Type II restriction endonucleases: structure and mechanism. Cell Mol Life Sci., 62, 685-707.

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

122. Pougach,K., Semenova,E., Bogdanova,E., Datsenko,K.A., Djordjevic,M., Wanner,B.L. and Severinov,K. (2010) Transcription, processing and function of CRISPR cassettes in Escherichia coli. Mol. Microbiol., 77, 1367–1379.

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

124. Pride,D.,T., Salzman,J., Relman,D.,A. (2012) Comparisons of clustered regularly interspaced short palindromic repeats and viromes in human saliva reveal bacterial adaptations to salivary viruses. Environ Microbiol., 9, 2564-76.

125. Pul,Ü., Wurm,R., Arslan,Z., Geißen,R., Hofmann,N. and Wagner,R. (2010) Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Mol. Microbiol., 75, 1495–1512.

126. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P. and Lim,W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell, 152, 1173–1183.

127. Ran,F.A., Cong,L., Yan,W.X., Scott,D.A., Gootenberg,J.S., Kriz,A.J., Zetsche,B., Shalem,O., Wu,X., Makarova,K.S., et al. (2015) In vivo genome editing using Staphylococcus aureus Cas9. Nature, 520, 186–191.

128. Rath,D., Amlinger,L., Hoekzema,M., Devulapally,P.R. and Lundgren,M. (2015) Efficient programmable gene silencing by Cascade. Nucleic Acids Res., 43, 237–246.

129. Redding,S., Sternberg,S.H., Wiedenheft,B., Jennifer,A., Greene,E.C., Redding,S., Sternberg,S.H., Marshall,M., Gibb,B., Bhat,P., et al. (2015) Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System Article

130. Richter,C., Gristwood,T., Clulow,J.S. and Fineran,P.C. (2012) In Vivo Protein Interactions and Complex Formation in the Pectobacterium atrosepticum Subtype I-F CRISPR/Cas System. PLoS One, 7.

131. Richter,C., Dy,R.L., McKenzie,R.E., Watson,B.N.J., Taylor,C., Chang,J.T., McNeil,M.B., Staals,R.H.J. and Fineran,P.C. (2014) Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res., 42, 8516–8526.

132. Roberts,R.,J., Vincze,T., Posfai,J., Macelis,D. (2003) REBASE: restriction enzymes and methyltransferases. Nucleic Acids Res., 31, 418-20.

133. Rollie,C., Schneider,S., Brinkmann,A.S., Bolt,E.L. and White,M.F. (2015) Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. Elife, 4.

134. Rollins,M.,F., Chowdhury,S., Carter,J., Golden,S.,M., Wilkinson,R.,A., Bondy-Denomy,J., Lander,G.,C. and Wiedenheft,B. (2017) Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. Proc. Natl. Acad. Sci. U. S. A., doi: 10.1073/pnas.1616395114.

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

136. Samai,P., Pyenson,N., Jiang,W., Goldberg,G.W., Hatoum-Aslan,A. and Marraffini,L.A. (2015) Co-transcriptional DNA and RNA cleavage during type III CRISPR-cas immunity. Cell, 161, 1164–1174.

137. Samson,J.E., Magadán,A.H., Sabri,M. and Moineau,S. (2013) Revenge of the phages: defeating bacterial defences. Nat. Rev. Microbiol., 11, 675–687.

138. Sampson,T.R., Saroj,S.D., Llewellyn,A.C., Tzeng,Y.-L. and Weiss,D.S. (2013) Corrigendum: A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature, 501, 262–262.

139. 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, 9275–9282.

140. Sashital,D.G., Wiedenheft,B. and Doudna,J.A. (2012) Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. Mol. Cell, 46, 606–615.

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

142. Schunder,E., Rydzewski,K., Grunow,R. and Heuner,K. (2013) First indication for a functional CRISPR/Cas system in Francisella tularensis. Int. J. Med. Microbiol., 303, 51–60.

143. Semenova,E., Jore,M.M., Datsenko,K. A, Semenova,A., Westra,E.R., Wanner,B., van der Oost,J., Brouns,S.J.J. and Severinov,K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc. Natl. Acad. Sci. U. S. A., 108, 10098–10103.

144. Seed,K.,D. (2015) Battling Phages: How Bacteria Defend against Viral Attack. PLoS Pathog., 11, e1004847.

145. Semenova,E., Kuznedelov,K., Datsenko,K.A., Boudry,P.M., Savitskaya,E.E., Medvedeva,S., Beloglazova,N., Logacheva,M., Yakunin,A.F. and Severinov,K. (2015) The Cas6e ribonuclease is not required for interference and adaptation by the E. Colitype I-E CRISPR-Cas system. Nucleic Acids Res., 43, 6049–6061.

146. Semenova,E., Savitskaya,E., Musharova,O., Strotskaya,A., Vorontsova,D., Datsenko,K.A., Logacheva,M.D. and Severinov,K. (2016) Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex. Proc. Natl. Acad. Sci., 113, 7626–7631.

147. Severinov,K., Ispolatov,I. and Semenova,E. (2016) The Influence of Copy-Number of Targeted Extrachromosomal Genetic Elements on the Outcome of CRISPR-Cas Defense. Front. Mol. Biosci., 3.

148. Shabbir,M.A.B., Hao,H., Shabbir,M.Z., Wu,Q., Sattar,A. and Yuan,Z. (2016) Bacteria vs. bacteriophages: Parallel evolution of immune arsenals. Front. Microbiol., 7.

149. Shariat,N., Kirchner,M.K., Sandt,C.H., Trees,E., Barrangou,R. and Dudley,E.G. (2013) Subtyping of Salmonella enterica serovar newport outbreak isolates by CRISPR-MVLST and determination of the relationship between CRISPR-MVLST and PFGE results. J. Clin. Microbiol., 51, 2328–2336.

150. Shmakov,S., Savitskaya,E., Semenova,E., Logacheva,M.D., Datsenko,K.A. and Severinov,K. (2014) Pervasive generation of oppositely oriented spacers during CRISPR adaptation. Nucleic Acids Res., 42, 5907–5916.

151. Shmakov,S., Abudayyeh,O.O., Makarova,K.S., Wolf,Y.I., Gootenberg,J.S., Semenova,E., Minakhin,L., Joung,J., Konermann,S., Severinov,K., et al. (2015) Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol. Cell, 60, 385–397.

152. Shmakov,S., Smargon,A., Scott,D., Cox,D., Pyzocha,N., Yan,W., Abudayyeh,O.O., Gootenberg,J.S., Makarova,K.S., Wolf,Y.I., et al. (2017) Diversity and evolution of class 2 CRISPR–Cas systems. Nat. Rev. Microbiol., 10.1038/nrmicro.2016.184.

153. Silas,S., Mohr,G., Sidote,D.J., Markham,L.M., Sanchez-Amat,A., Bhaya,D., Lambowitz,A.M. and Fire,A.Z. (2016) Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. Science (80-. )., 351, aad4234-aad4234.

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

155. van Soolingen,D., De Haas,P.E.W., Hermans,P.W.M., Groenen,P.M.A. and Van Embden,J.D.A. (1993) Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of Mycobacterium tuberculosis. J. Clin. Microbiol., 31, 1987–1995.

156. Sorek,R., Kunin,V., and Hugenholtz, P. (2008) CRISPRa widespread system that provides acquired resistance against phages in bacteria and archaea, Nat. Rev. Microbiol., 6, 181-186.

157. Smith,G.,R. (2012) How RecBCD Enzyme and Chi Promote DNA Break Repair and Recombination: a Molecular Biologist's View. Microbiol Mol Biol Rev., 76, 217-228.

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

159. Staals,R.H.J., Agari,Y., Maki-Yonekura,S., Zhu,Y., Taylor,D.W., vanDuijn,E., Barendregt,A., Vlot,M., Koehorst,J.J., Sakamoto,K., et al. (2013) Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. Mol. Cell, 52, 135–145.

160. Staals,R.H. and Jackson,S.A. (2016) Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. Nat Commun., 7,12853.

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

162. Sternberg,S.H., Richter,H., Charpentier,E. and Qimron,U. (2016) Adaptation in CRISPR-Cas Systems. Mol. Cell, 61, 797–808.

163. Strotksaya,A., Semenova,E., Savitskaya,E. and Severinov,K. (2015) Rapid multiplex creation of Escherichia coli strains capable of interfering with phage infection through CRISPR. Methods Mol. Biol., 1311, 147–159.

164. Strotskaya,A., Savitskaya,E., Metlitskaya,A., Morozova,N., Datsenko,K.A., Semenova,E. and Severinov,K. (2017) The action of Escherichia coli CRISPR–Cas system on lytic bacteriophages with different lifestyles and development strategies. Nucleic Acids Res., 10.1093/nar/gkx042.

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

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

167. Szczelkun,M.D., Tikhomirova,M.S., Sinkunas,T., Gasiunas,G., Karvelis,T., Pschera,P., Siksnys,V. and Seidel,R. (2014) Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc. Natl. Acad. Sci., 111, 9798–9803.

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

169. Tyson,G.W. and Banfield,J.F. (2008) Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ. Microbiol., 10, 200–207.

170. Tzur,Y.B., Friedland,A.E., Nadarajan,S., Church,G.M., Calarco,J.A. and Colaiácovo,M.P. (2013) Heritable custom genomic modifications in Caenorhabditis elegans via a CRISPR-Cas9 system. Genetics, 195, 1181–1185.

171. Vale,P.F., Lafforgue,G., Gatchitch,F., Gardan,R., Moineau,S. and Gandon,S. (2015) Costs of CRISPR-Cas-mediated resistance in Streptococcus thermophilus. Proc. R. Soc. B Biol. Sci., 282, 20151270.

172. Vasu,K. and Nagaraja,V. (2013) Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. Microbiol. Mol. Biol. Rev., 77, 53–72.

173. Vorontsova,D., Datsenko,K.A., Medvedeva,S., Bondy-Denomy,J., Savitskaya,E.E., Pougach,K., Logacheva,M., Wiedenheft,B., Davidson,A.R., Severinov,K., et al. (2015) Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. Nucleic Acids Res., 10.1093/nar/gkv1261.

174. Walkinshaw,M.,D., Taylor,P., Sturrock,S.,S., Atanasiu,C., Berge,T., Henderson,R.,M., Edwardson,J.,M., Dryden, D.,T.,F. (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. Mol. Cell, 9,187–194.

175. Wang,C., Villion,M., Semper,C., Coros,C., Moineau,S., Zimmerly,S. (2011) A reverse transcriptase-related protein mediates phage resistance and polymerizes untemplated DNA in vitro. Nucleic Acids Res., 39, 7620-9.

176. Wang,J., Li,J., Zhao,H., Sheng,G., Wang,M., Yin,M. and Wang,Y. (2015) Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. Cell, 163, 840–853.

177. Warren,R.,A. (1980) Modified bases in bacteriophage DNAs. Annu. Rev. Microbiol., 34, 137–158.

178. Wei,Y., Terns,R.M. and Terns,M.P. (2015) Cas9 function and host genome sampling in type II-A CRISPR-Cas adaptation. Genes Dev., 29, 356–361.

179. Wei,Y., Chesne,M.T., Terns,R.M. and Terns,M.P. (2015) Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. Nucleic Acids Res., 43, 1749–1758.

180. Westra,E.,R., Pul,U., Heidrich,N., Jore,M.,M., Lundgren,M., Stratmann,T., Wurm,R., Raine,A., Mescher,M., van Heereveld,L., Mastop,M., Wagner,E.,G., Schnetz,K., van Der Oost,J., Wagner,R., Brouns,S.,J. (2010) H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. Mol Microbiol.,77,1380-1393.

181. Westra,E.R., van Erp,P.B.G., Künne,T., Wong,S.P., Staals,R.H.J., Seegers,C.L.C., Bollen,S., Jore,M.M., Semenova,E., Severinov,K., et al. (2012) CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. Mol. Cell, 46, 595–605.

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

183. Wiedenheft,B., Zhou,K., Jinek,M., Coyle,S.M., Ma,W. and Doudna,J.A. (2009) Structural Basis for DNase Activity of a Conserved Protein Implicated in CRISPR-Mediated Genome Defense. Structure, 17, 904–912.

184. Wright,A.,V., Nuñez,J.,K. and Doudna,J.,A. (2016) Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell, 164, 29-44.

185. Xue,C., Seetharam,A.S., Musharova,O., Severinov,K., Brouns,S.J.J., Severin,A.J. and Sashital,D.G. (2015) CRISPR interference and priming varies with individual spacer sequences. Nucleic Acids Res., 43, 10831–10847.

186. Yang,H., Wang,H., Shivalila,C.S., Cheng,A.W., Shi,L. and Jaenisch,R. (2013) XOne-step generation of mice carrying reporter and conditional alleles by CRISPR/cas-mediated genome engineering. Cell, 154.

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

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

189. Zhang,J., Rouillon,C., Kerou,M., Reeks,J., Brugger,K., Graham,S., Reimann,J., Cannone,G., Liu,H., Albers,S.V., et al. (2012) Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity. Mol. Cell, 45, 303–313.

190. Zhao,H., Sheng,G., Wang,J., Wang,M., Bunkoczi,G., Gong,W., Wei,Z. and Wang,Y. (2014) Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. Nature, 515, 147–150.

191. Zhou,J., Shen,B., Zhang,W., Wang,J., Yang,J., Chen,L., Zhang,N., Zhu,K., Xu,J., Hu,B., Leng,Q. and Huang,X. (2014) One-step generation of different immunodeficient mice with multiple gene modifications by CRISPR/Cas9 mediated genome engineering. Int J Biochem Cell Biol., 46,49-55.

192. Zetsche,B., Gootenberg,J.S., Abudayyeh,O.O., Slaymaker,I.M., Makarova,K.S., Essletzbichler,P., Volz,S.E., Joung,J., van der Oost,J., Regev,A., et al. (2015) Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell, 163, 759–771.

## Supplementary

**Table S1. Oligonucleotides used in the study**

| Name | Sequence (5'-3') | Description/site of annealing |
|---|---|---|
| **PCR** | | |
| **P4518** | AAGGTTGGTGTCTTTTTTAC | CRISPR locus in KD263, KD454 |
| **P4581** | GTCGCTGCCGTGACGTTATG | CRISPR locus in KD263, KD454 |
| **Ec_II for** | AACATAATGGATGTGTTGTTTGTG | CRISPR II locus in BW40297 |
| **Ec_II rev** | GAAATGCTGGTGAGCGTTAATG | CRISPR II locus in BW40297 |
| **HS1-check for** | CATCATTGGAAAACGTTCTTC | used for amplification of HS1 incorporated spacer insight CRISPR array |
| **HS2 for** | GTCATTCTGAGAATAGT | used for amplification of HS2 incorporated spacer insight CRISPR array |
| **qPCR** | | |
| **Host for** | TCGGTTCGGGTGGTGCA | a genomic locus in KD263 |
| **Host rev** | TGACGTCCAGATACAC | a genomic locus in KD263 |
| **Leader for** | GGAACTCTCTAAAAGTATACATTTGTTC | amplifies 138-bp CRISPR leader region, also, together with primer P4581, was used to check for CRISPR array HS1 and HS2 incorporated spacers (235 bp amplicons) |
| **Leader rev** | GGTTGGTGTCTTTTTTACCTGTTTG | amplifies a 138-bp region of CRISPR leader |
| **CS for** | GTAACCCACTCGTGCA | amplifies 33-bp CS |
| **Leader_FaiI for** | GCTTTAAGAACAAATGT | amplifies 34-bp CRISPR leader region |
| **Leader_FaiI rev** | GGAACTCTCTAAAAGTA | amplifies 34-bp CRISPR leader region |
| **CS rev** | GCTGAAGATCAGTTGGG | amplifies 33-bp CS |
| **PS for** | GTTGTCTTTCGCTGCTG | amplifies 33-bp G8 protospacer |

| | | |
|---|---|---|
| **PS rev** | GCGGGATCGTCACC | amplifies 33-bp G8 protospacer |
| **HS1 for** | GTGCTCATCATTGGAAAA | amplifies 33-bp HS1 |
| **HS1 rev** | TCGCCCCGAAGAACG | amplifies 33-bp HS1 |
| **HS1_long for** | GCAGAACTTTAAAAGTGC | amplifies a 61-bp fragment containing HS1; also used as Fr 210 rev |
| **HS1_long rev** | ATCCTTGAGAGTTTTCG | amplifies a 61-bp fragment containing HS1 |
| **Fr 210 for** | TCGCCCTTATTCCCTTT | amplifies a 210-bp fragment of pG8mut or pG8mut_CCG plasmids |
| **HS2 for** | GTCATTCTGAGAATAGT | amplifies a 33-bp fragment containing HS2 |
| **HS2 rev** | CTCGGTCGCCGCATAC | amplifies a 33-bp fragment containing HS2 |
| **HS2_long for** | GGTGAGTACTCAACCAA | amplifies a 68-bp fragment containing HS2 |
| **HS2_long rev** | CGGGCAAGAGCAACT | amplifies a 68-bp fragment containing HS2 |
| **200 for** | CGGTCAACATTGAGGAAGAGC | amplifies a 200-bp genomic fragment which was analyzed after S1 treatment |
| **200 rev** | TACGTCACCAACGACACGG | amplifies a 200-bp genomic fragment which was analyzed after S1 treatment |
| **generation of model substrates** | | |
| **HS1_full/cmp for** | GTGCTCATCATTGGAAAACGTTCTTCGGGGCGA | double-stranded 33-bp HS1 |
| **HS1_full/cmp rev** | TCGCCCCGAAGAACGTTTTCCAATGATGAGCAC | double-stranded 33-bp HS1 |
| **HS1_part/cmp for** | CATCATTGGAAAACGTTCTTCGGGGCGA | splayed HS1 containing 23-bp internal duplex |
| **HS1_part/cmp rev** | CCGAAGAACGTTTTCCAATGATGAGCAC | splayed HS1 containing 23-bp internal duplex |

| | | |
|---|---|---|
| **HS2_full/cmp for** | CTCGGTCGCCGCATACACTATTCTCAGAATGAC | double-stranded 33-bp HS2 |
| **HS2_full/cmp rev** | GTCATTCTGAGAATAGTGTATGCGGCGACCGAG | double-stranded 33-bp HS2 |
| **cloning** | | |
| **HS1_CCG for** | CCAATGATGAGCACGGTTAAAGTTCTGCTAT | used to generate altered PAM in HS1 of pG8mut_CCG |
| **HS1_CCG rev** | ATAGCAGAACTTTAACCGTGCTCATCATTGG | used to generate altered PAM in HS1 of pG8mut_CCG |
| **Cas1 for** | GGTGTAATCATATGACCTGGCTTCCC | used for amplification *cas1* gene of *E.coli* |
| **Cas1 rev** | GGTTCGCATATGACAGCA | used for amplification *cas1* gene of *E.coli* |
| **Cas2 for** | TGCAGCATATGAGTATGTTGGT | used for amplification *cas2* gene of *E.coli* |
| **Cas2 rev** | GATCAGCGGATCCTTTGTTTTCAAACAGG | used for amplification *cas2* gene of *E.coli* |
| **oligonucleotides used for primer extension** | | |
| **HS1 for pr/ext** | GTCATTCTGAGAATAGT | HS1 |
| **HS1 rev pr/ext** | ACATCGAACTGGATCTC | HS1 |
| **HS2 for pr/ext** | CATCCGTAAGATGCTTTTC | HS2 |
| **HS2 rev pr/ext** | TGGCGCGGTATTATCC | HS2 |

**Table S2. Amplicons produced in qPCR reactions**

| Amplicon | Oligonucleotide pair used for amplification | Length, bp | Confirmed by Sanger sequencing YES/NO |
|---|---|---|---|
| HS1 33bp | HS1 for-HS1 rev | 33 | YES |
| HS1 47bp | HS1 for-HS1 long_rev | 47 | YES |
| HS1 47bp* | HS1 long_for-HS1 rev | 47 | YES |
| HS1 61bp | HS1 long_for-HS1 long_rev | 61 | YES |
| HS2 33bp | HS2 for-HS2 rev | 33 | YES |
| HS2 46bp | HS2 long_for– HS2 rev | 46 | YES |
| HS2 50bp | HS2 for – HS2 long_rev | 50 | YES |
| HS2 61bp | HS2 long_for– HS2 long_rev | 61 | YES |
| Leader 34 bp | Leader FaiI for-Leader FaiI rev | 34 | YES |
| Leader 138 bp | Leader for - Leader rev | 138 | YES |
| CS | CS for - CS rev | 33 | YES |
| PS | PS for – PS rev | 33 | YES |
| Host | Host for – Host rev | 33 | YES |
| 210 bp fragment | Fr 210 for – HS1_long for | 210 | YES |
| 200 bp fragment | Fr200_for – Fr200_rev | 200 | YES |