

## CRISPR-Cas: the effective immune systems in the prokaryotes

Yalda Vasebi (Ph.D student) and Reza Khakvar(Ph.D)

*Department of Plant Protection, Faculty of Agriculture, University of Tabriz.*

### ARTICLE INFO

*Article history:*

Received: 13 February 2014

Accepted: 17 April 2014

Available online: 1 June 2014

*Keywords:*

CRISPR,

Cas proteins, Types I, II and III.

### ABSTRACT

Approximately all sequenced archaeal and half of eubacterial genomes have some sort of adaptive immune system, which enables them to target and cleave invading foreign genetic elements by an RNAi-like pathway. CRISPR–Cas (clustered regularly interspaced short palindromic repeats–CRISPR-associated proteins) systems consist of the CRISPR loci with multiple copies of a short repeat sequence separated by variable sequences with similar size that are derived from invaders and cas genes encode proteins involved in RNA binding, endo- and exo-nucleases, helicases, and polymerases activities. There are three main types (I, II and III) of CRISPR/Cas systems. All systems function in three distinct stages: (1) adaptation, (2) crRNA biogenesis, and (3) interference. This review focuses on the features and mechanisms of the CRISPR-Cas systems and current finding about them.

### 1. Introduction

Horizontal gene transfer (HGT) refers to the transfer of genes between organisms. HGT has been shown to be an important factor in the evolution of many organisms such as bacteria (Gyles and Boerlin, 2014). There are several mechanisms for HGT in bacteria such as transformation, transduction and conjugation. Each microbe must balance the need to acquire new beneficial traits by HGT with the need to prevent the entry of genetic elements that impose fitness costs (Levin, 2010).

Bacteria and Archaea have several mechanisms to deal with invading foreign genetic elements, such as plasmids, phages, integrative and conjugative elements (ICEs) and transposons (Richter et al., 2012; Fineran and Charpentier, 2012; Barrangou, 2013). Foreign DNA integration may distort the function of cell genes and so the prokaryotes have several HGT-limiting mechanisms to prevent foreign DNA

either from entering or distributing within the cell. One of the known mechanisms is related to restriction-modification (RM) systems. Modification enzymes methylate the restriction sites in the bacterial genome to prevent their cleavage by restriction nucleases. Another HGT-preventing mechanism is based on mutations in the genes for the cell surface receptors that bacteriophages utilize to enter the cell. Intracellular mechanisms as prokaryotic Abortive infection mechanism (Abi) activate death in infected cells and thereby prevent phages and plasmids from spreading in the population.

Another newly studied mechanism that inhibit DNA uptake by phage infection, plasmid conjugation, and artificial transformation is based on clustered regularly interspaced short palindromic repeats (CRISPRs) (Wiedenheft et al., 2012; Abedon, 2012; Weinberger and Gilmore, 2012). CRISPR/Cas systems are widely distributed among prokaryotes (~50% of

<sup>†</sup>Corresponding author. Dr. Yalda Vasebi  
Tel.: 09377675279  
E-mail address: Yalda\_vasebi@yahoo.com

bacteria and ~99% of archaea) and are present in both pathogenic and commensal organisms (Sampson and Weiss, 2013). Many CRISPR–cas loci belong to the “islands” that contain various “high-mobility” genes such as toxins–antitoxins, transposases and components of other defense systems (Makarova et al., 2009).

The most unexpected feature in CRISPR-mediated resistance to mobile genetic elements is that the cell only becomes protected after foreign DNA has entered it. The acquired protection is inherited through generations (Pougach et al., 2012). This adaptive immunity system, which uses a library of small noncoding RNAs as a potent weapon against fast-evolving viruses, is also used as a regulatory system by the host (Bhaya et al., 2011). This mechanism has two major features; first, the host can specifically incorporate short sequences from invading genetic elements (virus or plasmid) into a region of its genome distinguished by CRISPRs. These repetitive loci serve as molecular vaccination cards. Second, when these sequences are transcribed and precisely processed into small RNAs, they lead a multifunctional protein complex (Cas proteins) to recognize and cleave incoming foreign genetic material (Wiedenheft et al., 2012; Bhaya et al., 2011). The CRISPR-Cas system provides a unique opportunity to observe and model co-evolution between host and virus in natural environments or in controlled settings because acquisition and immunity occur on short time scales and evidence of past genetic aggressions can be deduced in some cases (Bhaya et al., 2011).

### 1.1. Discovery of the CRISPR

The repeats were first described in 1987 for the bacterium *Escherichia coli* K12 by Ishino et al. during downstream sequencing of *iap* gene. The authors noted a set of 29 nucleotide (nt) repeats separated by unrelated, non-repetitive and similarly short sequences (spacers). In 2000, similar clustered repeats were identified in further bacteria and archaea and were termed as Short Regularly Spaced Repeats (SRSR) (Mojica et al., 2000). In 2002, SRSR was renamed CRISPR by Jansen et al. They

discovered a set of genes, cas genes, associated with CRISPR repeats.

In 2005, three independent research teams founded that many CRISPR spacers are similar in sequence to several phage DNA and extrachromosomal DNA as plasmid. These findings indicated that the CRISPR/cas system could have a role in adaptive immunity in bacteria by conferring bacteriophage resistance and preventing plasmid transformation (Pougach et al., 2012; Harvath and Barrangou, 2010). In 2007, first experimental evidence showed that the CRISPR/Cas system is an antiviral defense system that alters the resistance of *Streptococcus thermophilus* to phage attack with spacer DNA (Barrangou et al., 2007). In recent years CRISPRs were used as a new genome engineering tool in human cell culture (Jinek et al., 2012), baker's yeast (*Saccharomyces cerevisiae*) (Dicarlo et al., 2013), nematodes (*Caenorhabditis elegans*) (Friedland et al., 2013), plants (Jiang et al., 2013a) and mice (Wang et al., 2013).

### 1.2. Architecture of CRISPR/cas system –repeat, spacer, leader sequence and cas genes

The CRISPR array located on the either chromosome or plasmid. A single genome can harbor more than one CRISPR array. Those can vary considerably in size with the largest identified to date, in *Haliangium ochraceum* DSM 14365, containing 587 repeats (Bhaya et al., 2011; Richter et al., 2012). A CRISPR consists of an array of highly conserved short DNA direct repeat(R) sequences (21-48 bp long), which are interspaced by stretches of variable similar length sequence called spacers(S) (26-72 bp). The spacer sequences generally originate from phage or plasmid DNA (Gasiunas et al., 2014) and they represent a “memory of past genetic aggressions” (Stern et al., 2010). The repeat sequences within a CRISPR locus are conserved, but in different CRISPR loci can vary in both sequence and length although there are partially conserved sequences such as a GTTTg/c motif at the 5´ end and a GAAAC motif at the 3´ end (Bhaya et al., 2011; Kunin et al., 2007). In addition, the number of repeat–spacer units in a CRISPR

locus varies widely among organisms (Wiedenheft et al., 2012). The “leader” sequence, which is located upstream of the first repeat in CRISPR array, has about 200–500 bp long with A+T-rich sequence and serves as a promoter element for CRISPR transcription. The leader region is also important for the acquisition of new spacers (Richter et al., 2012; Wiedenheft et al., 2012).

CRISPR loci often have groups of conserved protein-encoding genes, named cas (CRISPR associate) genes, in their neighborhood. Based on computational analyses, Cas proteins were predicted to contain identifiable domains characteristic of helicases, nucleases, polymerases, and RNA-binding proteins, which led to the initial speculation that they may be part of a novel DNA repair system (Makarova et al., 2002). Not all CRISPR loci have adjoining cas genes, it is possible that only the subset of CRISPR loci that have adjacent cas genes are functionally active (Bhaya et al., 2011; Diez et al., 2010; Horvath et al., 2009).

CRISPR/Cas systems are currently classified into type I, II and III, based on the phylogeny and presence of particular Cas proteins. There is further division within each type into subtypes. The Cas proteins are important for the differentiation of major CRISPR/Cas types and the subtypes (Makarova et al., 2011a; Richter et al., 2012). The systems can be complex since some bacteria contain multiple CRISPR/Cas subtypes, each of which can have multiple CRISPR arrays that function with the appropriate Cas cluster (van Belkum et al., 1998) suggesting that these systems are compatible and could share functional components (Gasiunas et al., 2014; Wiedenheft et al., 2012).

### 1.3. The CRISPR/cas mechanism

The CRISPR–Cas mechanism is arbitrarily divided into three main stages: (1) adaptation or new spacer acquisition, (2) CRISPR transcription and processing (crRNA generation), and (3) interference or silencing (Gasiunas et al., 2014; Wilkinson and Wiedenheft, 2014; Lange et al., 2013; Richter et al., 2013). During adaptation, Cas proteins

recognize invasive nucleic acid and integrate short pieces of foreign DNA (protospacer) into the CRISPR region as new spacers. Spacers are inserted at the leader proximal end followed by duplication of the repeat. Functionally, the process of spacer acquisition can be divided into distinct steps involving (a) recognition of the invasive nucleic acid and scanning foreign DNA for potential PAMs (CRISPR motifs) (Mojica et al., 2009; Deveau et al., 2008), conserved short regions (typically only 2 to 5 nt long), (b) the generation of a new repeat-spacer by processing of the nucleic acid, and (c) the integration of the new CRISPR repeat-spacer unit at the leader end of the CRISPR locus (Bhaya et al., 2011).

In the expression and processing stage, the CRISPR repeats-spacer array is transcribed into a long primary RNA transcript (precrRNA) that is further processed by endonucleolytic cleavage into a set of small crRNAs, containing a conserved repeat fragment and a variable spacer sequence (guide) complementary to the invading nucleic acid. The cleavage of pre-crRNA occurs at the base of the hairpin formed by the palindromic CRISPR repeats, typically yielding a crRNA with an 8-nt tag or handle at the 5' end and a large part of the next repeat including the stem-loop termed the 3' handle (Brouns et al., 2008; Carte et al., 2008).

In the interference or silencing stage the short CRISPR-derived RNAs (crRNAs) assemble with Cas proteins into large surveillance complexes which recognizes the target sequence in the invasive nucleic acid by base pairing to the complementary strand of double-stranded DNA (Jore et al., 2011) or single-stranded RNA (Hale et al., 2009; Zhang et al., 2012), and induces sequence-specific cleavage (Garneau et al., 2010), thereby preventing proliferation and propagation of foreign genetic elements. The unique occurrence of the PAM sequence on the invading foreign DNA is likely to play a dual role: first, in spacer selection and acquisition and second, in the interference process for discrimination of self-versus nonself, which highlights its importance. Indeed, it has been demonstrated that despite perfect matches between spacer and protospacer sequences, mutations in the PAM can circumvent CRISPR

encoded immunity (Garneau et al., 2010; Sapranaukas et al., 2011).

#### 1.4. Type I CRISPR-Cas System

Of the three systems, Type I, thus far, is the most diverse with six different subtypes (subtypes I-A to I-F) and are found in both bacteria and archaea (Makarova et al., 2011b). The essential and significantly conserved marker protein in the interference reaction is Cas3, which contains a HD phosphohydrolase domain and a DExH-like helicase domain (Makarova et al., 2011a). Both domains are also found to be encoded separately by two discrete genes. These two domains have been shown to unwind dsDNA (helicase domain) and cleave ssDNA (HD nuclease domain), depending on ATP and Mg<sup>2+</sup> ions (Mulepati and Bailey, 2011; Sinkunas et al., 2011). Cas3 interacts with a complex of different Cas proteins that bind and deliver the crRNA. This complex is termed Cascade (CRISPR-associated complex for antiviral defense) and the structure of that reveals an unusual seahorse- shape (Jore, 2010). Type I CRISPR-mediated mechanisms of adaptive immunity have been explored for the six model organisms. Two of them (*E.coli* and *S.thermophilus*) belong to the subtype I-E, while the other four are of I-A (*Sulfolobus solfataricus*), I-B (*Haloferax volcanii*), I-C (*Bacillus halodurans*), and I-F (*Pseudomonas aeruginosa*) subtypes, respectively (Gasiunas et al., 2014).

In the Type I systems, repeat-spacer arrays are transcribed into a precursor crRNA (precrRNA) where a palindromic sequence of the repeat forms a hairpin, which is recognized and processed by Cas6 or Cas5d endoribonucleases (Carte et al., 2008; Nam et al., 2012a) to generate a mature crRNA. crRNA is then incorporated into a large multisubunit RNP complex, which together with Cas3 protein induce silencing of invasive DNA (Wiedenheft et al., 2012; Sorek et al., 2013). In Type I, the multisubunit Cascade binds pre-crRNA, which is cleaved by Cas6e in subtype I-E or by Cas6f in subtype I-F, to create crRNAs with a typical 8-nt extension or handle at the 5' end, followed by the spacer and part of the repeat region,

which can form a hairpin structure at the 3' end. The first 6–12 nt of the crRNA spacer are most important for target binding and are termed the “seed sequence” (Wiedenheft et al., 2011a; Wiedenheft et al., 2011).

To avoid that the Cascade: Cas3 complex degrades the host genomic encoded CRISPR cluster, it has to be ensured that the 5' terminal tag of the crRNA and the PAM sequence located upstream of the viral protospacer do not form base pairs. The PAM sequence for type I systems is typically 2–3 bases long and can differ between different subtypes and even organisms. The I-E Cascade complex has a size of 405 kDa and is composed of the five subunits, Cse1 (A), Cse2 (B), Cas7 (C), Cas5 (D) and Cas6e (casE). Cas7 and Cas5 tightly bind and protect the crRNA from degradation (Bronus et al., 2008; Jore et al., 2011), whereas Cse1 and Cse2 were shown to be nucleic acid-binding proteins that preferentially interact with the DNA target (Nam et al., 2012b; Mulepati et al., 2012).

For DNA interference, in DNA-interferencing in the Type I CRISPR–Cas systems, crRNA is incorporated into a Cascade. The Cascade complex scans DNA for a protospacer sequence and PAM. Once the correct PAM and a short primary hybridization sequence “seed” are identified, the crRNA basepairs with a complementary DNA strand forming R-loop that serves as a loading site for the Cas3 protein. Cas3 binding to the ssDNA triggers ATPase/helicase activity. In the presence of ATP, Cas3 remodels the Cascade–DNA complex making both target and non-target strands available for the Cas3 cleavage within a protospacer sequence. In the absence of ATP, the Cas3 nuclease domain (HD) cleaves a displaced non-target strand within a protospacer producing a nicked DNA. Cas3 further translocates in the 3' to 5' direction powered by a helicase domain (Hel) whereas the HD domain degrades DNA in a unidirectional manner (Gasiunas et al., 2014) (Fig. 1).

#### 1.5. Type II CRISPR-Cas System

Type II systems have only been found in bacterial genomes and are characterized by a

distinct minimal set of cas genes (Makarova et al., 2011a; Makarova et al., 2011b). Type II is the simplest of the three CRISPR-Cas types, with only four genes (cas9, cas1, cas2, and either cas4 or csn2) that compose the operon. There are two subtypes, Type IIA (or CASS4 that includes csn2) and Type IIB (or CASS4a that includes cas4) (Bhaya et al., 2011). In these systems, the large multifunctional signature protein, Cas9, is involved in both the generation and maturation of crRNAs as well as target phase and plasmid DNA for degradation and in the subsequent interference reaction (Sapranaukas et al., 2011; Garneau et al., 2010). Type II-A systems have been explored for two model organisms *Streptococcus pyogenes* and *S.thermophilus* DGCC7710.

Cas9 appears to contain two nuclease domains, one at the N-terminus (RuvC-like nuclease) and an HNH (McrA-like) nuclease domain in the middle section (which might be involved in target cleavage based on its endonuclease activity) (Bhaya et al., 2011). The processing of crRNAs is dependent on a trans-activating crRNA (tracrRNA) encoded in the vicinity of CRISPR loci and containing a 25 nt long stretch that is complementary to the crRNA repeat sequence (Deltcheva et al., 2011). The comparison of several tracrRNA molecules did not identify any highly conserved sequence or structure elements other than the anti-repeat sequence (Chylinski et al., 2013). Cas9 facilitates the base pairing of tracrRNA and pre-crRNA, which form a RNA duplex that is then targeted by the host endonuclease, RNase III. Cleavage of this duplex by RNase III generates mature crRNAs with 20 nt spacer-derived 5'-tags and 19–22 nt repeat-derived 3'-tags. The type II repeats do not form stem-loops and has been suggested this deficiency is overcome by pairing with the tracrRNA (Deltcheva et al., 2011).

In the interference step, the cleavage of target dsDNA requires not only crRNA and Cas9, but also the presence of tracrRNA. A ternary Cas9–crRNA–tracrRNA complex, using a mechanism that yet has to be defined, locates and binds to a protospacer sequence within the double-stranded DNA in a PAM-dependent process. The absolute requirement of PAM for dsDNA binding by the Cas9t complex implies

that PAM serves as a priming site for strand separation or is essential for stabilization of the R-loop structure because dsDNA lacking PAM is not bound. The Cas9t binding to the target sequence in the dsDNA presumably results in an R-loop structure, where one DNA strand is displaced and the complementary strand is paired with the crRNA. PAM is located downstream of the protospacer and differs between different systems. For *S.pyogenes* it is NGG, and for *S. thermophilus* CRISPR1 and CRISPR3 systems, NAAGW and NGGNG, respectively. The PAM is required only for a double-stranded but not a single-stranded DNA binding and cleavage by Cas9t (Gasiunas et al., 2012; Jinek et al., 2012; Mojica et al., 2009).

Cas9 cleaves the DNA strand complementary to the crRNA with a McrA/HNH nuclease domain and the non-complementary strand with a RuvC-like (RNase H fold) domain in the presence of Mg<sup>2+</sup> ions (Jinek et al., 2012). The precise DNA cleavage site was identified 3nt upstream of the PAM for the complementary strand, whereas the non-complementary DNA strand is cleaved at additional sites within three to eight base pairs upstream of the PAM, producing blunt-ended cleavage products (Jinek et al., 2012; Garneau et al., 2010).

It is tempting to speculate that, in the Type II systems, Cas9-bound tracrRNA provides a scaffold for the crRNA binding and stabilization similarly to Cascade proteins in Type I and Cmr proteins in Type III systems (Hale et al., 2009; Jore et al., 2011; Zhang et al., 2012; Wiedenheft et al., 2011a; Wiedenheft et al., 2011b; Lintner et al., 2011).

Similar to the seed sequence in type I systems, complementarity between crRNA and target over a 13 bp stretch proximal to the PAM is required for interference (Jinek et al., 2012) and hence, phages with mutations in this region of the protospacer region can evade interference (Deveau et al., 2010) (Fig. 1).

There are two major differences between mature crRNAs in II -A and Type I systems. First, crRNA in Type II lacks a 5'-handle and contains an extended 22-nt 3'-handle generated by the RNase III cleavage within the repeat region in the pre-crRNA:tracrRNA duplex. Second, the spacer fragment in the Type II

crRNA is shorter, because the 5'-end of the spacer sequence is trimmed to 20nt by unknown nuclease(s). Consequently, the spacer in the mature crRNA matches only 20 of the 30-nt protospacer sequence in the invading nucleic acid. The non-matching fragment in the protospacer is not important for the CRISPR-mediated immunity; however, shortening of the protospacer sequence to 19nt or more abrogates CRISPR-mediated plasmid interference (Gasiunas et al., 2012; Jiang et al., 2013b). Three model systems have been used to study mechanisms of invading nucleic acid destruction by Type II systems (Gasiunas et al., 2014).

### 1.6. Type III CRISPR-Cas System

Type III systems are further classified into III-A and III-B subtypes predominantly found in archaeal genomes (Makarova et al., 2011a; Makarova et al., 2011b) and interestingly, type III-B systems are only found in combination with one or more other CRISPR subtypes (Bhaya et al., 2011). Most of our knowledge on the III-A subtype comes from the *Staphylococcus epidermidis* model system, while *Sulfolobus solfataricus* and *Pyrococcus furiosus*, have been used as model systems for III-B.

Type III systems encode the CRISPR-specific endoribonuclease, Cas6, and the subtype-specific signature Cas10 protein that bears palm-domain polymerase-specific sequence motifs (Gasiunas et al., 2014) and is very likely involved in target interference (Richter et al., 2013). Similar to Cas3 proteins of type I systems, Cas10 encodes a HD nuclease domain that is proposed to have similar function in target degradation (Makarova et al., 2011a; Makarova et al., 2011b). Multiple RAMP-family proteins are present in the Type III systems (Makarova et al., 2006; Makarova et al., 2011a; Makarova et al., 2011b; Makarova et al., 2013; Koonin and Makarova, 2013). Intriguingly, two different Type III systems seem to target different nucleic acids. The Type III-A system of *S. epidermidis* contains five Csm proteins and targets DNA in vivo (Marraffini and Sontheimer, 2008; Marraffini and Sontheimer, 2010). DNA targeting by this system does not

require a specific PAM sequence, but sequences complementary to the 8nt 5'-tag of the crRNA are not targeted by this system (Marraffini and Sontheimer, 2010).

The similarity of the III-A and III-B operons suggests that interference in the III-A subtype is indeed mediated by an effector complex rather than a single protein. As a result the putative complex has been termed the CSM complex (Makarova et al., 2011b). Every CRISPR/Cas system apart from the III-B subtype is thought to target dsDNA by forming an R-loop structure, consisting of a heteroduplex between crRNA and the complementary protospacer strand and a ssDNA (single-stranded DNA) non-complementary strand, followed by degradation by the interference nuclease (Brouns et al., 2008; Garneau et al., 2010; Ivancic-Bace et al., 2012). On the other hand, the III-B system from *Sulfolobus islandicus* interferes with plasmid DNA transformation via transcription-dependent DNA targeting and relies on the direct protospacer transcription into RNA (Richter et al., 2013; Richter et al., 2012).

In the type III-B system of *P. furiosus*, Cas6 is not an integral part of the interference complex after the crRNA processing, but the 8-nt 5'- repeat tag serves as an anchor for the assembly of a six protein (Cmr1–Cmr6) ribonucleoprotein interference complex. A similar Cmr complex with seven proteins (Cmr1–Cmr7) was identified for *S. solfataricus* and shown to endonucleolytically cleave invading RNA at UA dinucleotides (Zhang et al., 2012; Hale et al., 2009). Targeting of RNA was shown to be PAM-independent for both investigated Cmr complexes. Notably, these two interference complexes differ from all other investigated subtypes, as they specifically target RNA and not DNA (Hale et al., 2009; Hale et al., 2012). However, recently, it could be demonstrated in vivo that Cmr proteins can target also plasmid DNA in a PAM-independent manner (Deng et al., 2013).

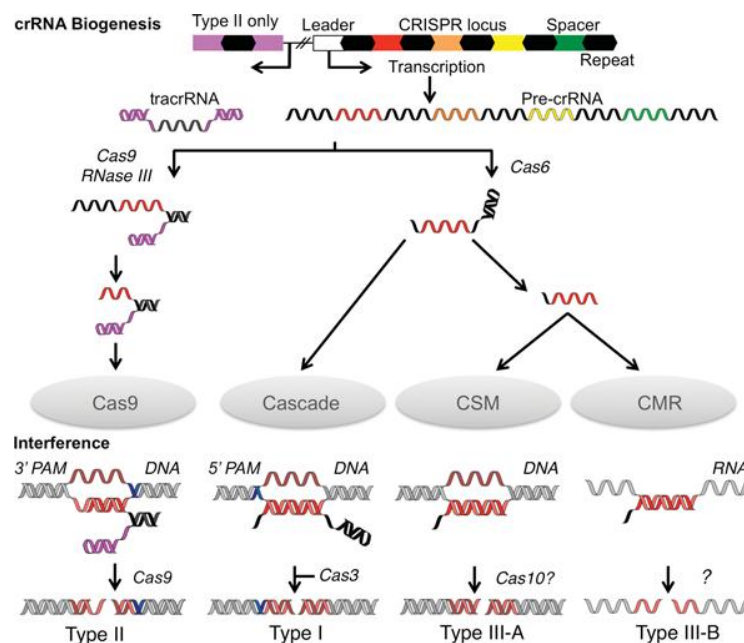
CRISPR-encoded immunity in the Type III CRISPR–Cas systems is consistent with the following general mechanism of interference. Firstly, the CRISPR repeat region is transcribed into a long primary pre-crRNA which undergoes a two-step processing to yield mature crRNA of

two different lengths which contain an 8-nt 5'-handle originating from the repeat sequence and trimmed spacer 3'-end. Despite the differences in repeat sequences (partially palindromic vs. non-palindromic in the III-A and III-B systems, respectively), Cas6 ribonuclease contributes to the primary processing stage. Secondly, the mature crRNA in the Type III-B system is incorporated into an effector complex which targets RNA in vitro using crRNA as a guide. Thirdly, unlike the effector complexes of Type I and Type II systems, the effector complexes of III-A and III-B systems achieve interference in a PAM-independent manner (Gasiunas et al., 2014).

In interference, the Cmr complex scans RNA and crRNA basepairs with a matching protospacer sequence. Two different RNA cleavage mechanisms are proposed. The Cmr complex (consists of six Cas proteins (Cas10, Cmr1, and Cmr3–6) and crRNA) from *P. furiosus* exploits the ruler mechanism to introduce cuts in the target RNA 14nt from the

3'-end of crRNA to yield two product fragments with different lengths, 39 and 45nt (Hale et al., 2009). The Cmr complex of *S. sulfolobus* (consists of seven Cas proteins (Cmr1, Cas10, and Cmr3–7)) guided by crRNA of various length cuts the target RNA in a sequence-specific manner at UA dinucleotides at multiple positions. The Ss-Cmr complex cleaves both target and guide RNA (crRNA) sequences in vitro (Beloglazova et al., 2011) (Fig. 1).

The ability of CRISPR/Cas systems to target DNA raised the question of how they avoid targeting their own CRISPR arrays, which have perfect complementarity to the crRNAs they produce. For the type III-A system, this is avoided by requiring spacer: protospacer complementarity and an absence of base-pairing to the 5' handle in the crRNA (Marraffini et al., 2010). This ensures that only non-self targets are licensed for degradation, but whether the same principle applies to other types remains unknown (Richter et al., 2012).



**Figure 1.** Schematic representation of crRNA biogenesis and CRISPR interference. Processing events involving nucleic acids are coloured; repeats (black), spacers (red–green) and tracrRNA (magenta). For clarity, a single spacer (red) was used to illustrate the processes, although in actual systems all spacers are processed. Targets are shown in other red shades (lighter for the complementary strand and darker for the non-complementary). The PAMs are shown in blue. The pre-crRNA and interference nucleases are indicated along with the interference complexes (Reeks et al., 2013).

### 1.7. Characterization of Cas proteins

The Cas proteins are important for the differentiation between both the major CRISPR/Cas types and the subtypes. The cas gene products were further classified into ~45 distinct families (Haft et al., 2005); that number was later reduced to ~25 families (Makarova et al., 2006).

Two partially independent subsystems of Cas proteins can be distinguished (Makarova et al., 2011b; Richter et al., 2012). The first group is found across multiple types or subtypes, consists of an information processing module and requires the universally present core proteins, Cas1 and Cas2, which are involved in new spacer acquisition (Makarova et al., 2011b; Pougach et al., 2010; Makarova et al., 2006; Haft et al., 2005). The second, or executive, subsystem is required for processing of primary CRISPR transcripts (crRNA) and recognition and degradation of invading foreign nucleic acid, and is quite diverse. For instance, in certain CRISPR sub-types, the multisubunit Cascade is involved in the processing of the crRNA in type I systems (Lintner et al., 2011; Nam et al., 2012a; Wiedenheft et al., 2011a), whereas in other types a single multifunctional protein (Cas3, Cas9 and Cas10) may play this role (Makarova et al., 2011b). In addition, there are several repeat-associated mysterious proteins (RAMPs) that constitute a large superfamily of Cas proteins (Jansen et al., 2002; Barrangou et al., 2007; Przybilski et al., 2011).

All three CRISPR-Cas systems contain two universal genes: cas1, a metal-dependent nuclease that cleaves ssDNA and dsDNA, generating ~80 bp DNA fragments from dsDNA, with no sequence specificity that could be involved in the integration of the alien DNA (spacer) into CRISPR cassettes (Wiedenheft et al., 2009; Marraffini and Sontheimer, 2009; Makarova et al., 2011b). The Cas1 structure reveals a novel fold with a two-domain architecture (Wiedenheft et al., 2009). This ubiquitous, highly conserved protein can be used as a scaffold to investigate the evolution of the CRISPR-Cas system (Makarova et al., 2011b) and cas2, a metal dependent

endoribonuclease, that also appears to be involved in the spacer acquisition stage. The small Cas2 protein cleaves ssRNAs in U-rich regions. Crystal structures of Cas2 from several species have been solved, revealing a ferredoxin fold, which is not common for endoribonucleases (Beloglazova et al., 2008).

Otherwise, the three types of CRISPR-Cas systems substantially differ in their sets of constituent genes, and each is characterized, respectively, by a unique signature gene. The signature genes for the three types are, respectively, Cas3, a superfamily 2 helicase, composed of two domains: an HD domain that has metal-dependent nuclease activity on double-stranded oligonucleotides (Han and Krauss, 2009) and a DEAD/H box helicase domain (Makarova et al., 2006; Sinkunas et al., 2011), cas9 (a large protein containing a predicted RuvC-like and HNH nuclease domains) (Makarova et al., 2011a) and cas10 (a protein containing a domain homologous the palm domain of nucleic acid polymerases and nucleotide cyclases) (Makarova et al., 2011b).

The Cas proteins known as RAMPs (Repeat-Associated Mysterious Proteins) are present in several copies in both type I and III systems. Some of the RAMPs have been shown to possess sequence- or structure-specific RNase activity that is involved in the processing of pre-crRNA transcripts (Brouns et al., 2008; Hale et al., 2009). The crystal structures of several RAMPs have been solved and indicate that they contain one or two domains which display distinct versions of the RNA recognition motif (RRM) or ferredoxin fold (Lintner et al., 2011; Wang et al., 2011; Haurwitz et al., 2010). The RNA-binding RAMP domain is present in the Cas5, Cas6, Cas7 and Cmr3 protein families and RAMP-like domains are found in Cas2 and Cas10 (Reeks et al., 2013; Makarova et al., 2011b). Cas5 and Cas6, previously annotated as core Cas proteins (cas 1-6) as well, represent a group of distantly related Cas proteins referred to as RAMPs; they appear to have similar 3D structures, and share at least a C-terminal glycine-rich loop (Makarova et al., 2002).



### 1.8. Type U CRISPR-Cas systems

The subtypes I-U, II-U and III-U are introduced for systems that lack currently defined subtype specific signature genes but either might fit one of the established subtypes on the basis of further structure and sequence analysis, or potentially could become founders of new subtypes (Makarova et al., 2011b). An unusual CRISPR-Cas system has been recently identified in several bacterial genomes, e.g., *Acidithiobacillus ferrooxidans* ATCC 23270, denoted type U as it did not contain signature genes of any of the three CRISPR-Cas types (Garneau et al., 2010). This system is associated neither with the two ubiquitous core cas genes, cas1 or cas2, nor with any other signature genes of the three CRISPR-Cas types or the 10 subtypes. The *A. ferrooxidans* system consists of four genes denoted csf1, csf2, csf3 and csf4. The Csf2 protein is a Cas7 group RAMP closely related to the Csm3 subfamily. Csf3 is yet another diverged RAMP protein that might be functionally analogous to the Cas5 group (Makarova et al., 2011a).

### References

- Abedon, S.T., 2012. Bacterial 'immunity' against bacteriophages. *Bacteriophage*. 2(1), 50–54.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 315, 1709–1712.
- Barrangou, R., 2013. CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip Rev RNA*. 4, 267–78.
- Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., Koonin, E.V., Edwards, A.M., Savchenko, A., Yakunin, A.F., 2008. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J Biol Chem*. 283(29), 20361–20371.
- Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., Yakunin, A.F., 2011. Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J*. 30, 4616–4627.
- Bhaya, D., Davison, M., Barrangou, R., 2011. CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annu. Rev. Genet*. 45, 273–97.
- 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., van der Oost, J., 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*. 321, 960–964.
- Carte, J., Wang, R., Li, H., Terns, R.M., Terns, M.P., 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes. Dev*. 22(24), 3489–3496.
- Chylinski, K., Le Rhun, A., Charpentier, E., 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol*. 10 (5), 726–37.
- Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. 471, 602–607.
- Deng, L., Garrett, R.A., Shah, S.A., Peng, X., She, Q., 2013. A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Mol. Microbiol*. 87, 1088–1099.
- Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., Moineau, S., 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol*. 190, 1390–1400.
- Deveau, H., Garneau, J.E., Moineau, S., 2010. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol*. 64, 475–493.
- Dicarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., Church, G.M., 2013. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucl. Acids. Res*. 41(7), 4336–43.
- Diez-Villasenor, C., Almendros, C., Garcia-Martinez, J., Mojica, F.J., 2010. Diversity of CRISPR loci in *Escherichia coli*. *Microbiology*. 156, 1351–61.
- Fineran, P.C., Charpentier, E., 2012. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology*. 434, 202–9.
- Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiácovo, M.P., Church, G.M., Calarco, J.A., 2013. Heritable genome editing in *C. Elegans* via a CRISPR-Cas9 system. *Nat. Methods*. 10 (8), 741–3.
- Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A., Moineau, S., 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. 468, 67–71.
- Gasiunas, G., Barrangou, R., Horvath, P., Siksnys, V., 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA*. 109: E2579–E2586.
- Gasiunas, G., Sinkunas, T., Siksnys, V., 2014. Molecular mechanisms of CRISPR mediated microbial immunity. *Cell. Mol. Life Sci*. 71, 449–465.
- Gyles, C., Boerlin, P., 2014. Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. *Vet. Pathol*. 51(2), 328–340.
- Haft, D.H., Selengut, J., Mongodin, E.F., Nelson, K.E., 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*. 1(6), e60.
- Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V., III; Graveley, B.R., Terns, R.M., Terns, M.P., 2012. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell*. 45, 292–302.

- Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., Terns, M.P., 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*. 139, 945–956.
- Han, D., Krauss, G., 2009. Characterization of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2. *FEBS Lett.* 583(4), 771–776.
- Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., Doudna, J.A., 2010. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*. 329 (5997), 1355–1358.
- Horvath, P., Barrangou, R., 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science*. 327(5962), 167–70.
- Horvath, P., Coute-Monvoisin, A.C., Romero, D.A., Boyaval, P., Fremaux, C., Barrangou, R., 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int. J. Food Microbiol.* 131, 62–70.
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., Nakata, A., 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* 169 (12), 5429–5433.
- Ivancic-Bace, I., Al Howard, J., Bolt, E.L., 2012. Tuning in to interference: R-loops and Cascade complexes in CRISPR immunity. *J. Mol. Biol.* 422, 607–616.
- Jansen, R., Embden, J.D., Gastra, W., Schouls, L.M., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43, 1565–1575.
- Jiang, W., Bikard, D., Cox, D., Feng Zhang, Z., Luciano, A., Marraffini, L.M., 2013b. RNA-guided editing of bacterial genomes using CRISPR–Cas systems. *Nat. Biotechnol.* 31, 233–239.
- Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., Weeks, D.P., 2013a. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucl. Acids Res.* 41 (20), e188.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. 337(6096), 816–821.
- Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Edze, R., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, Ü., Wurm, R., Wagner, R., Beijer, M.R., Barendregt, A., Zhou, K., Snijders, A.P.L., Dickman, M.J., Doudna, J.A., Boekema, E.J., Heck, A.J.R., van der Oost, J., Brouns, S.J.J., 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* 18, 529–536.
- Koonin, E.V., Makarova, K.S., 2013. CRISPR–Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biol.* 10, 1–8.
- Kunin, V., Sorek, R., Hugenholtz, P., 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* 8 (4), R61.
- Lange, S.J., Alkhnbashi, O.S., Rose, D., Will, S., Backofen, R., 2013. CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucl. Acids Res.* 41 (17), 8034–8044.
- Levin, B., 2010. Nasty viruses, costly plasmids, population dynamics, and the conditions for establishing and maintaining CRISPR-mediated adaptive immunity in bacteria. *PLoS Genet.* 6(10), 1–12.
- Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q., Copie, V., Young, M.J., White, M.F., Lawrence, C.M., 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.
- Makarova, K.S., Aravind, L., Grishin, N.V., Rogozin, I.B., Koonin, E.V., 2002. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucl. Acids Res.* 30(2), 482–496.
- Makarova, K.S., Aravind, L., Wolf, Y.I., Koonin, E.V., 2011a. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR–Cas systems. *Biol. Direct.* 6, 38.
- Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., Koonin, E.V., 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.* 1, 7.
- Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., van der Oost, J., Koonin, E.V., 2011b. Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* 9 (6), 467–477.
- Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2009. Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct.* 4 (19), 1–38.
- Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2013. Comparative genomics of defence systems in archaea and bacteria. *Nucl. Acids Res.* 1–18.
- Marraffini, L.A., Sontheimer, E.J., 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*. 322, 1843–1845.
- Marraffini, L.A., Sontheimer, E.J., 2009. Invasive DNA, chopped and in the CRISPR. *Structure*. 17(6), 786–788.
- Marraffini, L.A., Sontheimer, E.J., 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genetics.* 11, 181–190.
- Mojica, F.J., Díez-Villasenor, C., Garcia-Martinez, J., Almendros, C., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology.* 155, 733–740.
- Mojica, F.J., Díez-Villaseñor, C., Soria, E., Juez, G., 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* 36 (1), 244–246.
- Mulepati, S., Bailey, S., 2011. Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). *J. Biol. Chem.* 286, 31896–31903.
- Mulepati, S., Orr, A., Bailey, S., 2012. Crystal structure of the largest subunit of a bacterial RNA-guided immune complex and its role in DNA target binding. *J. Biol. Chem.* 287, 22445–22449.
- Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., Delisa, M.P., Ke, A., 2012a. Cas5d protein processes

- pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure*. 20, 1574–1584.
- Nam, K.H., Huang, Q., Ke, A., 2012b. Nucleic acid binding surface and dimer interface revealed by CRISPR-associated CasB protein structures. *FEBS Lett.* 586, 3956–3961.
- Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L., Severinov, K., 2010. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.* 77, 1367–1379.
- Pougach, K.S., Lopatina, A.V., Severinov, K., 2012. CRISPR adaptive immunity systems of prokaryotes. *Mol. Biol.* 46(2), 175–182.
- Przybilski, R., Richter, C., Gristwood, T., Clulow, J.S., Vercoe, R.B., Fineran, P.C., 2011. Csy4 is responsible for CRISPR RNA processing in *Pectobacterium atrosepticum*. *RNA Biol.* 8, 517–528.
- Reeks, J., Naismith, J.H., White, M.F., 2013. CRISPR interference: a structural perspective. *Biochem. J.* 453, 155–166.
- Richter, C., Chang, J.T., Fineran, P.C., 2012. Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems. *Viruses*. 4, 2291–311.
- Richter, H., Randau, L., Plagens, A., 2013. Exploiting CRISPR/Cas: Interference Mechanisms and Applications. *Int. J. Mol. Sci.* 14, 14518–14531.
- Sampson, T.R., Weiss, D.S., 2013. Alternative Roles for CRISPR/Cas Systems in Bacterial Pathogenesis. *PLOS Pathogens*. 9 (10), 1–3.
- Sapranaukas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., Siksnys, V., 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides plasmid immunity in *Escherichia coli*. *Nucl. Acids Res.* 39(21), 9275–82.
- Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., Siksnys, V., 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO. J.* 30, 1335–1342.
- Sorek, R., Lawrence, C.M., Wiedenheft, B., 2013. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* 82, 237–266.
- Stern, A., Keren, L., Wurtzel, O., Amitai, G., Sorek, R., 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* 26(8), 335–340.
- van Belkum, A., Scherer, S., van Alphen, L., Verbrugh, H., 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol. Mol. Biol. Rev.* 62, 275–293.
- Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., Jaenisch, R., 2013. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell*. 153 (4), 910–918.
- Wang, R., Preamplume, G., Terns, M.P., Terns, R.M., Li, H., 2011. Interaction of the Cas6 ribonuclease with CRISPR RNAs: recognition and cleavage. *Structure*. 19(2), 257–264.
- Wiedenheft, B., Sternberg, S.H., Doudna, J.A., 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*. 482, 331–338.
- Weinberger, A.D., Gilmore, M.S., 2012. CRISPR-Cas: To Take Up DNA or Not—That Is the Question. *Cell Host & Microbe*. 12(2), 125–126.
- Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., Nogales, E., 2011b. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*. 477, 486–489.
- Wiedenheft, B., Sternberg, S.H., Doudna, J.A., 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*. 482, 331–338.
- Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J., Dickman, M.J., Doudna, J.A., 2011a. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci.* 108, 10092–10097.
- Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W., Doudna, J.A., 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR mediated genome defense. *Structure*. 17(6), 904–912.
- Wilkinson, R., Wiedenheft, B., 2014. A CRISPR method for genome engineering. *F1000Prime Reports*. 6, 3.
- Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.V., Naismith, J.H., Spagnolo, L., White, M.F., 2012. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol. Cell*. 45, 303–313.