

Review

Function and Regulation of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) / CRISPR Associated (Cas) Systems

Corinna Richter, James T. Chang and Peter C. Fineran *

Department of Microbiology and Immunology, University of Otago, PO Box 56, Dunedin 9054, New Zealand; E-Mails: ricco896@student.otago.ac.nz (C.R.); james.chang@otago.ac.nz (J.T.C.)

* Author to whom correspondence should be addressed; E-Mail: peter.fineran@otago.ac.nz; Tel.: +64-3-479-7735; Fax: +64-3-479-8540.

Received: 4 September 2012; in revised form: 10 October 2012 / Accepted: 11 October 2012 / Published: 19 October 2012

Abstract: Phages are the most abundant biological entities on earth and pose a constant challenge to their bacterial hosts. Thus, bacteria have evolved numerous ‘innate’ mechanisms of defense against phage, such as abortive infection or restriction/modification systems. In contrast, the clustered regularly interspaced short palindromic repeats (CRISPR) systems provide acquired, yet heritable, sequence-specific ‘adaptive’ immunity against phage and other horizontally-acquired elements, such as plasmids. Resistance is acquired following viral infection or plasmid uptake when a short sequence of the foreign genome is added to the CRISPR array. CRISPRs are then transcribed and processed, generally by CRISPR associated (Cas) proteins, into short interfering RNAs (crRNAs), which form part of a ribonucleoprotein complex. This complex guides the crRNA to the complementary invading nucleic acid and targets this for degradation. Recently, there have been rapid advances in our understanding of CRISPR/Cas systems. In this review, we will present the current model(s) of the molecular events involved in both the acquisition of immunity and interference stages and will also address recent progress in our knowledge of the regulation of CRISPR/Cas systems.

Keywords: phages; plasmids; horizontal gene transfer; CRISPR; Cas; cascade; PAM; crRNA; resistance

1. Introduction

Bacteria and Archaea are frequently exposed to stresses such as infection from bacteriophages (phage) and other genetic elements. These events can result in horizontal gene transfer (HGT), which is mediated by transduction, transformation or by conjugation of mobile elements such as plasmids [1,2]. HGT can be beneficial for adaptation and survival by rapidly promoting the transfer of genes encoding antibiotic resistance, virulence factors, or the ability to degrade toxic compounds [1]. However, in the absence of any selective advantage, foreign genetic elements can have a fitness cost. This is most evident when prokaryotes are infected by virulent phages. The sheer scale of viral abundance ($\sim 10^{31}$), diversity and rates of infection ($\sim 10^{25}/s$) [3,4] has resulted in the evolution of prokaryotic defenses to maintain the balance in this arms race [5].

Bacteria possess multiple methods to regulate genetic flux and resist phage infection. These mechanisms include the mutation or masking of cell surface receptors, restriction-modification, abortive infection and the clustered regularly interspaced short palindromic repeats (CRISPR) systems [5,6]. CRISPR systems are a widespread mechanism that equips bacteria with a sequence-specific heritable ‘adaptive immune system’ that has a genetic memory of past genetic incursions (for recent reviews see [7–9]).

CRISPRs use small non-coding RNAs for defense and function in conjunction with CRISPR associated (Cas) proteins. The mechanism of CRISPR/Cas interference involves three phases (Figure 1). Firstly, resistance is acquired via the integration of short sequences from foreign genetic elements (termed spacers) into repetitive genetic elements known as CRISPR arrays. Secondly, CRISPR arrays are then transcribed and processed into small RNAs (crRNAs) by Cas proteins. In the third and final step, targeting of the invading phage or plasmid is mediated by a Cas protein complex that contains crRNAs. During this stage, the crRNA-Cas protein complex then interferes, in a sequence-specific manner, with the foreign nucleic acids.

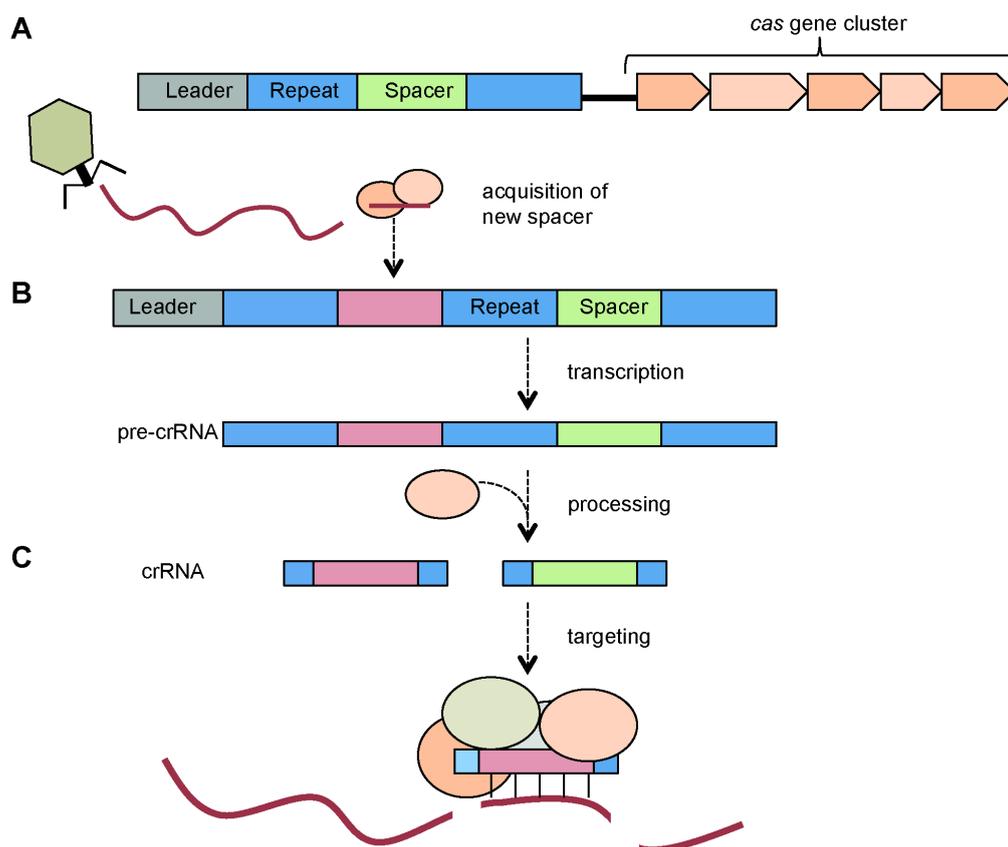
Despite sharing mechanistic similarities, there is significant diversity amongst CRISPR/Cas systems. There are three major CRISPR/Cas system types (I-III), which are characterized by a signature protein [10]. The main types are further divided into subtypes (e.g., I-A to I-F) based mainly on the presence of a subtype-specific set of Cas proteins and for some types the repeat sequence of the associated CRISPR array is also taken into account. The last five years have seen rapid advances in our mechanistic/molecular understanding of CRISPR/Cas systems. In this review, we provide an historical context, highlight the current models for the adaptation, processing and interference stages and finally, discuss the regulation of the CRISPR/Cas systems.

2. Discovery of CRISPR/Cas Systems

The first report of interspaced palindromic repeat sequences was made by Ishino *et al.* who detected five 29 bp repeats with 32 bp spacers near the *iap* gene of *Escherichia coli* [11]. Further reports were made for *Mycobacterium tuberculosis* [12], *Haloferax* spp. [13] and *Archaeoglobus fulgidus* [14,15]; however, they were not detected in eukaryotic or virus sequences [15]. Mojica and co-workers performed a comparative *in silico* study of those repetitive elements to determine structure and sequence similarity, as well as their phylogenetic distribution [16] showing that CRISPRs display a

high degree of homology between phylogenetically distant species and a wide distribution in bacteria and archaea. To date CRISPR/Cas systems have been found in almost 50% of bacterial and 85% of archaeal genome sequences available [17].

Figure 1. Overview of clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated (Cas) adaptive immunity. (a) Adaptation. The CRISPR arrays are composed of short repeats and intervening sequences derived from foreign invaders. Upon infection with a foreign element (e.g., phages or plasmids), part of the genome is typically incorporated into the leader end of the CRISPR array and the repeat is duplicated. The CRISPR arrays are located adjacent to a cluster of *cas* genes; **(b) crRNA generation.** The CRISPRs are transcribed into pre-crRNAs that are then processed into mature crRNAs; **(c) Interference.** The crRNA, in a complex with Cas proteins, binds and degrades the target nucleic acid of the invading element.



Terminology for these repeated elements was inconsistent and included interspaced short sequence repeats (SSR) [18], **s**pacer **i**nterspaced **d**irect **r**epeats (SPIDR) [15] or short **r**egularly spaced **r**epeats (SRSRs) [16]. For simplicity, the name CRISPR (**c**lustered **r**egularly **i**nterspaced **s**hort **p**alindromic **r**epeats) was accepted since it reflects the most important characteristics [19].

An important step towards the understanding of the role of CRISPRs was the identification of four genes closely located to the CRISPR arrays [19], which were termed CRISPR-associated (*cas*) genes 1-4. Cas1-4 were only present in genomes containing CRISPR arrays and their predicted functions included helicase and nuclease activities, which led to the hypothesis that CRISPR/Cas systems might be involved in DNA repair [20].

A critical discovery came in 2005 with three independent studies showing that spacers matched sequences of extrachromosomal origin, including phages, prophages and plasmids [21–23]. Furthermore, there was a positive correlation between the presence of spacers matching a particular phage and phage resistance [22,23]. In combination with detailed bioinformatics analyses of the Cas proteins, it was hypothesized that CRISPR/Cas systems provide an RNAi-like mechanism of resistance against invading genetic elements [21–25].

The first direct evidence that CRISPR/Cas could protect bacteria from phages or plasmids was provided by two key studies. Firstly, Barrangou *et al.* obtained phage resistant mutants following phage challenge of *Streptococcus thermophilus*. The phage resistant strains had incorporated new spacers matching to the viral genome while removal or addition of spacers matching to the phages resulted in phage sensitivity or resistance, respectively [26]. Subsequently, Marraffini and Sontheimer showed that CRISPR/Cas systems can also prevent both conjugation and transformation of plasmids in *Staphylococcus epidermidis* [27]. Recently, it was shown that the *Streptococcus pneumoniae* CRISPR/Cas can prevent natural transformation. Non-capsulated *S. pneumoniae* with CRISPRs programmed to target capsule genes were unable to be transformed into capsulated strains during infection in mice [28].

3. Components of CRISPR/Cas Systems

CRISPR/Cas systems comprise the following critical elements: a CRISPR array, an upstream leader sequence and the *cas* genes (Figure 1A) that are discussed below [19,29].

3.1. CRISPR Arrays: Repeats, Spacers and Leader Sequence

In the CRISPR array, repeats alternate with spacer sequences [11,30–32]. Within arrays, the repeats are typically identical in terms of length and sequence, but small differences do occur [16,29]. In particular, the repeat at the end of an array is often truncated or deviates more from the consensus sequence [17,33]. Repeats are usually between 23 and 47 bp in length and those of the same subtype share a consensus sequence in type I-C, I-E, I-F, and II [10,34,35]. Many repeat sequences show palindromes or short inverted repeats [16] and a predicted stable secondary structure in the form of a stem-loop [34]. However, there are repeats, which are not palindromic and are predicted to be unstructured. This difference in the ability to form secondary structures also has implications for the mechanism of pre-crRNA processing, which will be discussed later [10,34].

In contrast, the spacer sequences are mostly unique within a genome. Many spacers have been matched to sequences originating from extra-chromosomal sources such as phage or plasmids and other transferable elements [21–23,36]. Indeed, it is the spacers that confer the sequence-specific immunity against those extra-chromosomal agents [26,33,37]. The sequences in the foreign genome from which spacers are derived are termed protospacers. It is likely that the majority of spacers are phage- or plasmid-derived, but the limited depth of sequence data on these abundant and diverse elements leads to an underestimation. Interestingly, some CRISPR/Cas systems also possess spacers that match sequences elsewhere within their own genome. Their function, if any, is not resolved, but it has been proposed that they have been accidentally incorporated [38]. Spacer length can vary slightly throughout an array (typically by 1–2 nt) and spacer lengths up to 72 bp have been reported, but usually the size is similar to that of the repeats in the same array [17].

The third component of the CRISPR array is the leader sequence, which is located upstream of the first repeat [15]. This AT-rich sequence is about 200–500 bp long and includes the promoter necessary for transcription of the array (see below) [39–42]. The leader region is also important for the acquisition of new spacers [43].

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 [35]. The length of CRISPR arrays appears to correlate with the degree of activity in spacer acquisition, with longer arrays being more active than short arrays and those with degenerate repeat sequences [33].

3.2. Cas Proteins

In close proximity to the CRISPR array are genes that encode the CRISPR associated (Cas) proteins (Figure 1A) [19,20,24,25,44]. Cas proteins provide the enzymatic machinery required for the acquisition of new spacers from, and targeting, invading elements.

CRISPR/Cas systems are currently classified into type I, II and III, based on the phylogeny and presence of particular Cas proteins [10]. There is further division within each type into subtypes (e.g., type I is composed of type I-A to I-F). The Cas proteins are important for the differentiation between both the major CRISPR/Cas types and the subtypes. Two main groups of Cas proteins can be distinguished. The first group, which includes the core Cas proteins, is found across multiple types or subtypes. Cas1 and Cas2 (Cas2 is sometimes fused to another Cas protein) are found in every Cas operon across the three main types, and Cas1 is considered the universal marker of CRISPR/Cas systems [10,24,25,44]. Cas3, Cas9, and Cas10 are each specific for one major type, serving as signature proteins for type I, II, and III, respectively [10].

The second major group consists of proteins only found within the gene clusters of one particular subtype [10,25]. For example, the Cse and Csy proteins are subtype-specific for type I-E and type I-F systems, respectively. The subtype specific proteins of some systems form a complex involved in targeting and interference, which in the type I-E system is referred to as Cascade (CRISPR-associated complex for antiviral defense) [45–49]. These complexes will be discussed in detail below. 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 [18]. In summary, the *cas* cluster for each CRISPR/Cas subtype contains genes encoding Cas1 and Cas2, the signature protein defining the major type and a set of subtype specific proteins. Interestingly, the total number of proteins differs between the types of system, even though the functional principles are similar in the cases studied thus far. It is possible that unexpected functions may reside within Cas proteins that are yet to be characterized.

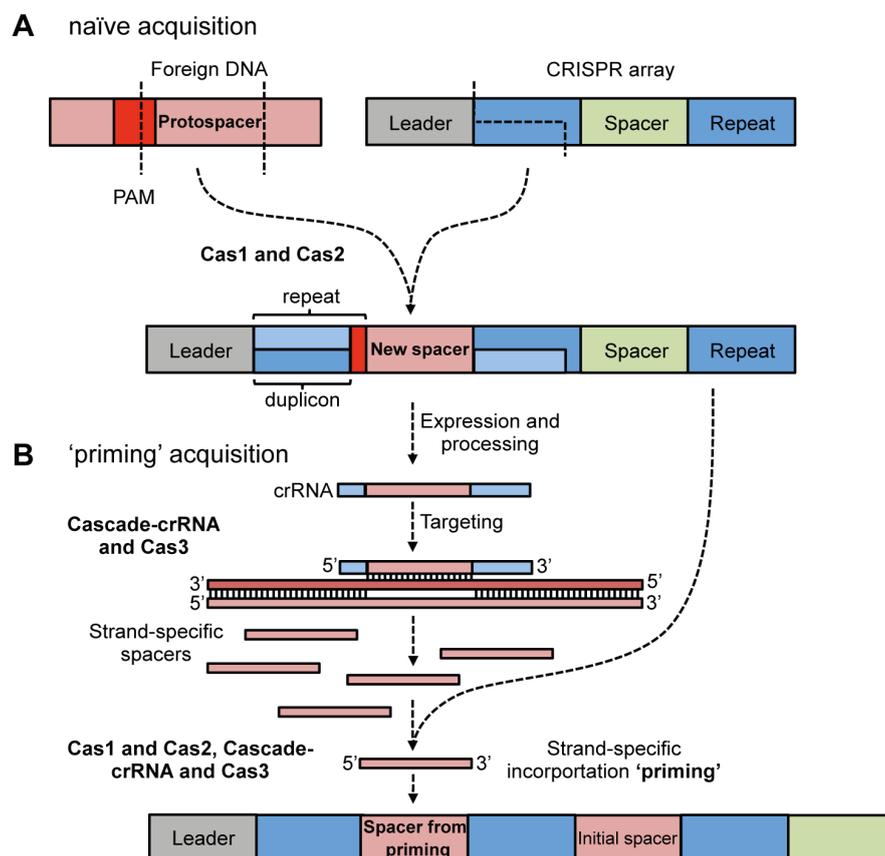
4. CRISPR/Cas Mechanism

As highlighted in Figure 1, the mechanism of CRISPR/Cas defense involves three stages. Resistance must first be acquired by integrating spacers into the CRISPR arrays. These arrays are then transcribed and processed into short crRNAs. Finally, a crRNA-Cas ribonucleoprotein complex targets the invading nucleic acid for degradation. In the following sections we outline the processes of spacer acquisition, crRNA biogenesis and interference.

4.1. Adaptation via Spacer Acquisition

The least characterised stage of CRISPR/Cas defense is adaptation, whereby new spacers are acquired from phages and plasmids and, with duplication of the repeat, are added to the leader proximal end of the CRISPR array (Figure 2). A number of metagenomic studies of phage-bacterial dynamics in environmental [50,51] and human [52] niches have provided clear evidence that CRISPR/Cas systems actively acquire new spacers from mobile genetic elements over short time-scales. The first demonstration of spacer incorporation from phages and plasmids in the laboratory was in the type II-A system of *S. thermophilus* [26,53]. Very recently, studies have observed spacer acquisition in the *E. coli* type I-E [43,54,55], *P. aeruginosa* type I-F [56], *Streptococcus agalactiae* type II-A [57] and *Sulfolobus solfataricus* type I and III-B [58] systems. The most well characterized system is the *E. coli* type I-E, so this will be the focus of the following discussion.

Figure 2. Model of CRISPR/Cas adaptation (based on type I-E systems). **(a)** Naïve acquisition. Cas1 and Cas2 are required for acquisition of new spacers. The first repeat at the leader end of the CRISPR array is duplicated and incorporates a new spacer sequence from a protospacer. The final nucleotide of the repeat is not duplicated but is provided by the PAM nucleotide immediately adjacent to the protospacer; **(b)** Priming acquisition. Expression of the CRISPR array and generation of the crRNAs against the foreign DNA results in binding / targeting, which is hypothesized to aid in the generation of precursors for integration (possibly single- or double-stranded). Productive interference is not required for priming. Priming acquisition requires Cascade-crRNA, Cas3 and Cas1 and Cas2 and results in new spacers derived from the same strand as the initial spacer.



4.1.1. Cas Proteins Required for Adaptation

In the *E. coli* type I-E system, Cas1 and Cas2 are required for spacer incorporation from plasmids and phages [43,55]. These recent results are consistent with the dispensable role of Cas1 and Cas2 in crRNA biogenesis / targeting in type I-E [45], type II-A [59] and type III-A [60] systems. The type I-E data is of broad relevance because Cas1 and Cas2 are conserved across all CRISPR/Cas types [10,24]. It is interesting that an insertion mutation of *csn2* inhibited spacer acquisition in the *S. thermophilus* type II-A system, suggesting that, in addition to Cas1 and Cas2, other Cas proteins might play a role [26].

Exactly how Cas1 and Cas2 function is unclear, but biochemical and structural studies offer some clues [61–63]. Cas1 is a metal-dependent endonuclease and the *Pseudomonas* Cas1 cleaves dsDNA, generating ~80 bp fragments [61]. The 80 bp dsDNA fragments generated from Cas1 are larger than expected for integration (~32–33 nt), indicating that other factors are involved. Cas1 dimers contain a stirrup-like structure and a positively charged surface that together might be involved in binding dsDNA [9]. Cas2 from *S. solfataricus* cleaves ssRNA at U-rich regions *in vitro* [64], but ssRNA or ssDNA binding or cleavage was not detected for *Desulfovibrio vulgaris* Cas2 [65]. Recently, the *Bacillus halodurans* Cas2 dimer was shown to contain Mg²⁺-dependent endonuclease activity against dsDNA and generated 120 bp products [66]. Despite these studies the exact mechanistic roles of Cas1 and Cas2 in acquisition are unknown.

There is increasing evidence that some Cas1 and Cas2 proteins may function together directly (e.g., via protein-protein interactions). For example, the type I-A system from *Thermoproteus tenax* contains a Cas1-Cas2 fusion protein, which interacts with Csa4 and Csa1, in what the authors name Cascis (CRISPR associated complex for the integration of spacers) [67]. It is noteworthy that the type I-F Cas3 is a Cas2-Cas3 hybrid, containing an N-terminal Cas2-like domain fused to Cas3 [10,24].

4.1.2. Protospacer Selection and Incorporation into CRISPR Arrays

The minimal CRISPR requirements for spacer acquisition in the type I-E system are a single ‘repeat’ and 60 bp of upstream sequence [43]. As the CRISPR promoter is not contained within this 60 bp, this suggests that CRISPR transcription is not required for incorporation, but indicates this region might be recognized by Cas1 and/or Cas2 to enable the directionality of integration at the leader end of the CRISPR array.

In type I-E arrays with multiple repeats, the leader proximal repeat is the one that is duplicated (Figure 2A) [43]. In addition, the protospacer added to the CRISPR array contains the last nt of the PAM motif, which becomes the final nt of the 5' repeat (Figure 2A) [54,55,68]. Therefore, only the first 28 nt of the repeat are duplicated and thus, this unit has been called the ‘duplicon’ [68] and might provide fidelity to the direction of incorporation [69]. This duplicon mechanism cannot apply to all CRISPR/Cas types since the last nucleotide of some PAMs (e.g., CCN in *S. solfataricus*) is not conserved, so could not contribute to the conserved repeats [58]. Interestingly, internal spacer acquisition was recently observed for one CRISPR array in *S. solfataricus* [58], which challenges the dogma that spacer acquisition is always leader proximal [69].

One critical question is how do the CRISPR/Cas systems acquire foreign DNA and not DNA from their own chromosome? In *S. thermophilus*, examination of the protospacers in the phage genomes

enabled the identification of a short 3' protospacer adjacent motif (PAM) [37]. PAM sequences have since been identified via bioinformatics for other CRISPR/Cas systems [70]. In *E. coli*, spacer selection was shown to require the presence of a PAM [43], but the short length of PAMs cannot provide the level of discrimination to avoid sampling chromosomal DNA. A self / non-self mechanism does occur, since the incorporation of spacers from the chromosome is rare [43]. It is possible that DNA modification, such as by a restriction-modification-like mechanism, might play a role.

4.1.3. Priming Acquisition

In the type I-E system there is now evidence that acquisition is divided into two stages; (1) naïve- and (2) priming-based acquisition [54,55]. As discussed above naïve acquisition is the initial spacer incorporation from a new foreign element. During priming, the initial spacer 'primes' the array for the addition of multiple invader-derived spacers from the same DNA strand as the initial spacer (Figure 2B) [54,55]. Cas1, Cas2, Cascade, Cas3 and a crRNA that matches the invading element are required for priming [55]. The nuclease activity of Cas3 was proposed to be required for the feedback loop, generating precursors for incorporation following the initial targeting event [54]. However, spacers that match the foreign DNA but do not support targeting, due to single nucleotide mutations, still promote priming [55]. Priming is proposed to allow adaptation to phage and plasmids that mutate and escape the initial spacers [55] and multiple spacers would increase resistance and decrease the frequency of escape.

In summary, details of CRISPR adaptation are beginning to come to light, but there are still many questions regarding the exact mechanisms of protospacer recognition, spacer generation, repeat duplication, integration of the new spacer and the roles of the Cas proteins.

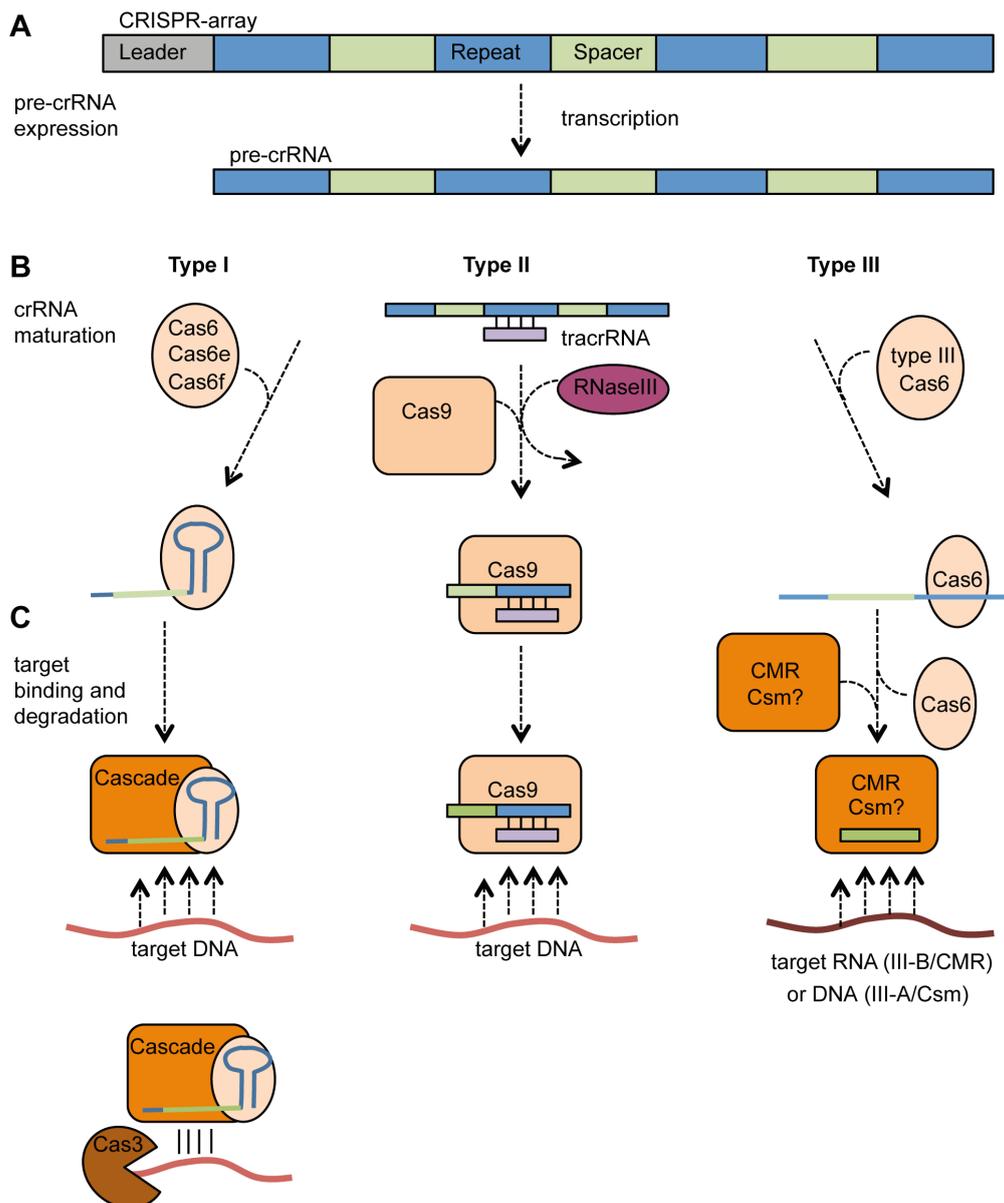
4.2. Expression, crRNA Generation and Interference

Following the acquisition of spacers, the CRISPR array is expressed as a long pre-crRNA that is processed into crRNAs. The mature crRNAs form part of a ribonucleoprotein complex, which targets and degrades the foreign genetic material. As will be outlined below, these processes are similar, yet differ between type I, II, and III systems (Figure 3).

4.2.1. Type I Systems

In type I CRISPR/Cas systems, regions within each repeat form a stem-loop, which is bound by the processing endoribonuclease. The endoribonucleases in type I systems belong to the group of Cas6 proteins. Characterized members of this group include Cas6b in type I-B [71], Cas5d in type I-C [49], Cas6e (CasE, Cse3) in type I-E [45,72,73] and Cas6f (Csy4) in type I-F [42,74]. The pre-crRNA is cleaved in a sequence-specific manner at the downstream base of the stem-loop, producing the mature crRNA consisting of a short 5' repeat handle followed by the spacer sequence and the stem-loop of the next repeat (Figure 3B) [49,71,74,75]. The crRNA and the endoribonuclease stay associated, and might serve as nucleation point [48,49,76] for the formation of Cascade (Figure 3C). Cascade is a ribonucleoprotein complex formed by the subtype-specific Cas proteins and the crRNA, initially identified in type I-E, but later also found in type I-C and I-F [9,45,46,48,49].

Figure 3. crRNA generation and target interference in type I, II and III CRISPR/Cas systems. (a) Transcription of the CRISPR array into a pre-crRNA; (b) Processing of the pre-crRNA into mature short crRNAs. In type I, RNA cleavage is performed by Cas6-homologues, which bind the repeat stem-loop and stay associated for Cascade formation. In type II, tracrRNA is required for binding and processing of the pre-crRNA by Cas9 and RNaseIII. In type III, Cas6 binds to non-structured repeats and processes the pre-crRNA into crRNA and then dissociates; (c) Target binding and cleavage. Type I Cascade binds the DNA target before recruiting Cas3 for degradation. In type II, Cas9 stays associated with the tracrRNA:crRNA complex after processing and subsequently binds and cleaves target DNA. The type III-B CMR-complex binds spacer sequence and targets RNA. It is hypothesized that a type III-A Csm complex forms and this system targets DNA.



The type I-E Cascade has a shape referred to as seahorse-like and has a hexameric backbone formed by Cas7 (CasC, Cse4). The stem-loop of the crRNA is bound to the endoribonuclease Cas6e, which forms the “head” of the seahorse. Cse1 (CasA) forms the “tail” and binds the 3' end of the crRNA at

the other end. Cse1 is bridged to Cas6 by a Cse2 (CasB) dimer, and linked by Cas5 (CasD) to the sixth Cas7 protein of the backbone [46,47]. The I-F and I-C complexes share common features with the I-E Cascade in terms of their shape, but also show variation, which is mainly due to different proteins involved. Different numbers of subtype-specific proteins constitute Cascade, three in the type I-C, four in I-F systems, while five distinct proteins are present in the type I-E Cascade [10]. Additional variation occurs at the level of the stoichiometry of individual protein subunits in different subtype Cascade complexes. When three or four proteins are used, the endoribonuclease, with the crRNA bound via the stem-loop, is located at one end of the complex. A homo-multimer of six proteins (Csy3 and Csd2 for I-F, and I-C, respectively) forms the neck or backbone of the complex, which holds the spacer and ensures it is extended for optimal presentation to invading nucleic acids. The end of the complex is formed by a hetero-dimer [48,49]. A type I-A complex from *S. solfataricus* (termed aCASCADE by the authors) has also been partially characterized and contained the backbone protein, Cas7 (Csa2), Cas5a, Csa5, Cas6 and crRNAs [77]. Although not directly shown, based on the data from other type I systems, the type I-A Cascade is likely to consist of the Cas6 endoribonuclease bound to the crRNA, a backbone of six Cas7 proteins and a Cas5a/Csa5 heterodimer.

During the final step of interference, the type I Cascade containing a specific crRNA recognizes and binds the appropriate target DNA (Figure 3C). In type I-E systems, a short loop within CasA recognizes potential targets via the PAM sequence and it is proposed to bind them before the target DNA is incorporated into Cascade [78]. In addition to the PAM requirement in the target, a non-contiguous 7 nt sequence (nucleotides in position 1–5 and 7–8) within the first 8 nt of the spacer, relative to the 5' end of the spacer within the crRNA, are essential for initial target recognition and hence are termed the 'seed sequence' [48,79]. Cascade with bound crRNA specifically binds to negatively supercoiled target DNA [80]. The negative supercoiled topology serves as energy source for formation of an R-loop. Furthermore, Cascade induces DNA bending before Cas3 is recruited for subsequent DNA degradation (Figure 3C) [80]. Cas3 is the signature protein of type I systems and possesses ATP-dependent helicase activity as well as the ability to cleave ssDNA via an HD nuclease domain [81–83]. Even though the exact mechanism is not yet fully understood, it is hypothesized that once recruited, Cas3 further unwinds and cleaves the target DNA in a 5' to 3' direction, while Cascade dissociates from the target and is recycled [80,83].

4.2.2. Type II Systems

A very different mechanism of CRISPR expression and crRNA maturation was discovered in the type II-A system of *Streptococcus pyogenes* [84]. An abundant short *trans*-encoded transcript is essential for processing of pre-crRNA into mature crRNA. Therefore, these RNAs are termed *trans*-activating crRNAs (tracrRNA). These tracrRNAs contain a 25 bp stretch with almost perfect complementarity to the type II CRISPR repeats within the pre-crRNA (Figure 3B). The type II repeats do not form stem-loops and the authors suggest this deficiency is overcome by pairing with the tracrRNA [84].

In addition to the tracrRNA, Cas9 (Csn1) and host RNase III are needed for the processing of pre-crRNA into mature crRNA. The requirement of RNase III is the first discovery of a non-Cas protein being an essential part of CRISPR/Cas defense machinery (Figure 3B) [84]. Interestingly, the

S. thermophilus type II-A system provides phage resistance in *E. coli* when recombinantly expressed [59], which suggests diverse RNase III enzymes are compatible for crRNA generation. Cas9 is not only involved in crRNA processing but, along with the tracrRNA:crRNA hybrid, recognizes and degrades the target (Figure 3C) [85]. In complex with the tracrRNA:crRNA structure, Cas9 binds the target dsDNA and creates a double strand break by cleaving the complementary and non-complementary strand with its HNH- and RuvC-like domains, respectively [85]. *In vivo*, blunt cleavage of phage and plasmid dsDNA was observed at a position 3 bp from the PAM within the protospacer [53]. The PAM is important for the affinity towards the target and may have implications on the unwinding of the duplex, as cleavage of targets with mutated PAMs is only impaired in dsDNA but not ssDNA. Indeed, phages with PAM mutations also have the ability to avoid targeting by type II systems [37]. Furthermore, 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 [85] and hence, phages with mutations in this region of the protospacer region can evade interference [37].

4.2.3. Type III Systems

In type III CRISPR/Cas systems, expression and interference are predominantly similar to those described for type I systems. However, the significant difference is that the endoribonuclease Cas6 does not bind to a repeat stem-loop but to the first bases of the repeat and that further crRNA maturation steps occur (Figure 3B) [86,87]. Indeed, not all repeats in those systems contain a strong palindrome and those associated with type III-A are predicted to be mostly non-structured [34]. In *Pyrococcus furiosus*, the binding site of Cas6 is located within nt 2–8 of the repeat and cleaves at a distal site between bases 22 and 23 resulting in crRNAs with an 8 nt repeat handle on the 5' end, the spacer sequence and 22 nt of the following repeat at the 3' end [86–88]. Interestingly, this RNA is not the final mature crRNA product. Cas6, unlike the type I endoribonucleases, only transfers the crRNA to the targeting complex termed CMR for type III-B where they are further processed (Figure 3C) [10,89,90]. In the type III-A system of *S. epidermidis*, a similar mechanism for processing by Cas6 was observed and the *csm2*, *csm3* and *csm5* genes were required for further 3' processing and crRNA maturation [60].

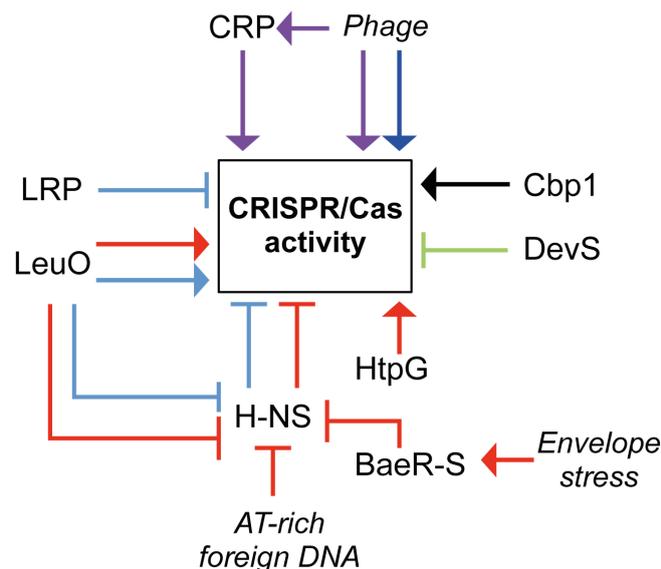
The *P. furiosus* CMR complex includes the full subset of Cmr proteins (Cmr1–6) and can contain two different species of crRNA [90], both with the 8 nt 5' end derived from the repeat and either 31 nt or 37 nt of the spacer [91]. The type III-B CMR complex from *S. solfataricus* contains an additional protein, Cmr7 and a stoichiometry of one protein of each of Cmr1–6 and 6 Cmr7 proteins [92]. Interestingly, the *P. furiosus* CMR complex has been shown to target RNA [89,90,92]. Basepairing between the last 14 nt of the crRNA and the RNA target seem to be essential for cleavage [90]. Target RNA cleavage occurs 14 nt upstream from the 3' end of the two mature crRNA species, generating two cleavage sites [90]. In contrast to type III-B, DNA is the target of type III-A systems [27], but there is not yet data about complex formation for the Csm proteins.

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 [93]. This ensures that only non-self targets are licensed for degradation, but whether the same principle applies to other types remains unknown.

5. Regulation of CRISPR/Cas Systems

As CRISPR/Cas systems function as a defense mechanism their expression might be expected to respond to invasion by extrachromosomal elements. This was recently proven in a shotgun proteomics approach in *S. thermophilus*, which showed an increase in Cas protein expression following phage infection [94]. Likewise, CRISPR and Cas expression were increased by phage infection in *Thermus thermophilus* [39]. Furthermore, CRISPR/Cas systems can be modulated in response to UV and play a role in sensitivity to DNA damage, hinting towards other possible roles besides the neutralization of invading elements [63,67]. However, to date little is known about how CRISPR/Cas systems are regulated in response to those external stimuli or during periods in which the system is not required [63,67]. Under certain conditions it could be viewed as favorable to down-regulate CRISPR/Cas activity when beneficial elements are available. However, most evidence demonstrates up-regulation of CRISPR/Cas upon exposure to phage. Since CRISPR/Cas systems are not 100% effective, some beneficial elements will be acquired and when they provide a selective advantage they are likely to be maintained. An overview of the regulatory inputs is shown in Figure 4.

Figure 4. Regulation of CRISPR/Cas activity. Network model summarizing the regulation of CRISPR/Cas activity. Triangular and flat arrowheads indicate positive and negative effects on CRISPR/Cas activity, respectively for *E. coli* (red), *Salmonella* (cyan), *M. xanthus* (green), *Sulfolobus* (black), *T. thermophilus* (purple) and *S. thermophilus* (blue). For details see text.



Most information about CRISPR/Cas regulation is available for the type I-E systems in *E. coli* and *Salmonella enterica* serovar Typhi backgrounds. The Histone-like nucleoid structuring protein (H-NS) is a global regulator involved in compacting the bacterial chromosome, via high affinity binding to AT-rich, curved DNA (reviewed by Dorman *et al.* [95]). Many promoters are located in close proximity to curved DNA. Binding of H-NS to these promoters prevents RNA polymerase (RNAP) binding, which results in gene-silencing [95]. In *E. coli*, H-NS negatively regulates *cas* gene expression, due to H-NS binding sites near the *cas* operon promoters [41,96]. It is hypothesized that

H-NS will bind to invading nucleic acids when they enter the cell, due to the higher AT-content [97,98]. Sequestration of H-NS is predicted to free the *cas* and/or *LeuO* promoter (see below) for RNAP recognition and thus activate expression, allowing effective CRISPR/Cas-mediated defense [41,96,99].

LRP (lysine-responsive regulatory protein) is a negative regulator for *cas* expression in *S. typhi* [100]. LRP functions in a similar fashion to H-NS by binding to the *cas* promoter and competitively excluding binding of RNAP. This appears to be independent of H-NS. However, LRP is able to bind simultaneously alongside H-NS, indicating that the two proteins interact to generate a nucleosome structure for *cas* gene repression [100]. Interestingly, LRP does not regulate the *cas* operon in *E. coli*, suggesting that the type I-E CRISPR/Cas systems in these two closely-related organisms function differently at the regulatory level [41,100].

LeuO, a LysR-type transcriptional regulator that also responds to amino acid starvation, affects *cas* gene expression via competition with H-NS in *S. Typhi* and *E. coli* [96,100,101]. *LeuO* binds to a region flanking the *cas* promoter and the H-NS binding site. The binding of *LeuO* to the DNA competes with H-NS binding, and relieves the inhibition of *cas* expression by enabling promoter recognition by RNAP [96,100,101]. During amino acid starvation expression of *LeuO* is increased in response to accumulation of the small molecules guanosine 3'-diphosphate 5'-triphosphate and guanosine 3',5'-bis(diphosphate), collectively called (p)ppGpp [102]. Interestingly, accumulation of (p)ppGpp does not occur during phage lambda infection in *E. coli* [103]. In theory, it is possible that infection with other phages triggers amino acid starvation, leading to *LeuO*-dependent activation of the CRISPR/Cas system.

The BaeR-S two-component regulatory system is activated in response to envelope stress (reviewed by MacRitchie *et al.* [104]). Phage infection potentially causes envelope stress by accumulation of viral proteins in the membrane. Upon detection of membrane stress, the histidine sensor kinase BaeS is activated by phosphorylation and in turn activates the BaeR protein. Activated BaeR gains the ability to bind DNA, modulating gene expression [104]. In *E. coli*, the BaeR binding site is located at the H-NS binding site near the *cseI* (*casA*) promoter. Therefore, bound BaeR serves as an antagonist for H-NS binding, predicted to result in the release of the *cas* promoters for RNA polymerase recognition and *cas* expression [105,106].

The high-temperature protein G (HtpG) chaperone affects the CRISPR/Cas system in *E. coli* by stabilizing Cas3. If HtpG is absent, the CRISPR/Cas system is no longer able to efficiently prevent phage infection [107]. However, HtpG is not present in all bacteria containing CRISPR/Cas systems. Indeed, most archaeal strains that contain Cas3 homologues do not carry an HtpG homolog, suggesting that it is not universally required. Therefore, other bacteria may require other chaperones in order for CRISPR/Cas interference to occur [108].

Other systems that affect CRISPR/Cas regulation were identified in *Myxococcus xanthus*, *T. thermophilus* and Sulfolobales. In *Myxococcus*, the *dev* operon is co-transcribed with *cas* genes, and this operon is negatively auto-regulated by DevS (a Cas5 protein) [109]. *T. thermophilus* contains 12 CRISPRs, and 3 different *cas* operons, type I-E, type III-A and type III-B [10,39]. During phage infection, the *cas* genes for the type I-E and type III-A systems were up-regulated via CRP (cAMP receptor protein) [39,110]. When CRP is bound to cAMP, it recognizes promoters and activates gene expression. Interestingly, inactivation of *crp* did not abolish the *cas* activation, suggesting both *crp*-dependent and -independent pathways. In addition, some *cas* genes and CRISPR arrays were

regulated by phage infection in a CRP-independent manner [39]. In *Sulfolobus islandicus* and *S. solfataricus*, the expression of pre-crRNA is regulated by Cbp1 (CRISPR DNA repeat binding protein), which directly binds the DNA repeats and might influence the activity of promoters and terminators present within some spacers or repeats [111,112]. In summary, CRISPR/Cas systems can be regulated at the level of transcription of the *cas* genes and CRISPR arrays and post-translational level on the Cas proteins (Figure 4). The picture of how these systems are regulated is far from complete and diverse and complex regulatory strategies exist.

6. Conclusions

CRISPR/Cas systems are widespread and versatile prokaryotic defense mechanisms providing adaptive and heritable immunity to extrachromosomal elements. In recent years research performed on these systems has led to a deeper understanding of the underlying mechanistic principles. Despite this rapid increase in knowledge, many questions remain to be answered. For example, challenges in the future include elucidating why some prokaryotes carry multiple CRISPR/Cas types and how their activity is coordinated. Furthermore, the adaptation step is still poorly understood, especially with respect to the specific roles of the Cas proteins involved and how foreign DNA is selected for integration. Major progress has been made regarding structure, function and interactions of some Cas proteins and their complexes. However, for many subtype-specific proteins, functional and structural information is still lacking and their precise role in the CRISPR/Cas mechanism remains unknown. Finally, the question, how CRISPR/Cas activity is regulated and to what extent it plays a role in areas not related to defense, e.g., DNA repair, needs to be addressed. Future research on CRISPR/Cas systems will also open up vast opportunities to utilize them, such as genome engineering for the creation of more robust industrial starter strains or as a gene targeting or gene silencing mechanism similar to that used in eukaryotic cells.

Acknowledgments

We thank Rita Przybilski for critically reading the manuscript and the referees for their constructive input. This work was supported by a University of Otago Research Grant and a Rutherford Discovery Fellowship to PCF from the Royal Society of New Zealand. CR was supported by a University of Otago Postgraduate Scholarship and a DAAD Doktorandenstipendium.

Conflict of Interest

The authors declare no conflict of interest.

References and Notes

1. Frost, L.S.; Leplae, R.; Summers, A.O.; Toussaint, A. Mobile genetic elements: The agents of open source evolution. *Nat. Rev. Microbiol.* **2005**, *3*, 722–732.
2. Koonin, E.V.; Wolf, Y.I. Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **2008**, *36*, 6688–6719.

3. Fuhrman, J.A. Marine viruses and their biogeochemical and ecological effects. *Nature* **1999**, *399*, 541–548.
4. Hendrix, R.W. Bacteriophage genomics. *Curr. Opin. Microbiol.* **2003**, *6*, 506–511.
5. Labrie, S.J.; Samson, J.E.; Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **2010**, *8*, 317–327.
6. Petty, N.K.; Evans, T.J.; Fineran, P.C.; Salmond, G.P. Biotechnological exploitation of bacteriophage research. *Trends Biotechnol.* **2007**, *25*, 7–15.
7. Marraffini, L.A.; Sontheimer, E.J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* **2010**, *11*, 181–190.
8. Terns, M.P.; Terns, R.M. CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* **2011**, *14*, 321–327.
9. Wiedenheft, B.; Sternberg, S.H.; Doudna, J.A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **2012**, *482*, 331–338.
10. Makarova, K.S.; Haft, D.H.; Barrangou, R.; Brouns, S.J.; Charpentier, E.; Horvath, P.; Moineau, S.; Mojica, F.J.; Wolf, Y.I.; Yakunin, A.F.; *et al.* Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **2011**, *9*, 467–477.
11. Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **1987**, *169*, 5429–5433.
12. Groenen, P.M.; Bunschoten, A.E.; van Soolingen, D.; van Embden, J.D. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol. Microbiol.* **1993**, *10*, 1057–1065.
13. Mojica, F.J.; Ferrer, C.; Juez, G.; Rodríguez-Valera, F. Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* **1995**, *17*, 85–93.
14. Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; *et al.* The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **1997**, *390*, 364–370.
15. Jansen, R.; van Embden, J.D.; Gaastra, W.; Schouls, L.M. Identification of a novel family of sequence repeats among prokaryotes. *OMICS* **2002**, *6*, 23–33.
16. Mojica, F.J.; Díez-Villaseñor, C.; Soria, E.; Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* **2000**, *36*, 244–246.
17. Grissa, I.; Vergnaud, G.; Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **2007**, *8*, 172.
18. van Belkum, A.; Scherer, S.; van Alphen, L.; Verbrugh, H. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 275–293.
19. Jansen, R.; Embden, J.D.; Gaastra, W.; Schouls, L.M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **2002**, *43*, 1565–1575.
20. Makarova, K.S.; Aravind, L.; Grishin, N.V.; Rogozin, I.B.; Koonin, E.V. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.* **2002**, *30*, 482–496.

21. Mojica, F.J.; Díez-Villaseñor, C.; García-Martínez, J.; Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **2005**, *60*, 174–182.
22. Bolotin, A.; Quinquis, B.; Sorokin, A.; Ehrlich, S.D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **2005**, *151*, 2551–2561.
23. Pourcel, C.; Salvignol, G.; Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **2005**, *151*, 653–663.
24. Makarova, K.S.; Grishin, N.V.; Shabalina, S.A.; Wolf, Y.I.; Koonin, E.V. 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* **2006**, *1*, doi:10.1186/1745-6150-1-7.
25. Haft, D.H.; Selengut, J.; Mongodin, E.F.; Nelson, K.E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **2005**, *1*, e60.
26. Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D.A.; Horvath, P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **2007**, *315*, 1709–1712.
27. Marraffini, L.A.; Sontheimer, E.J. CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. *Science* **2008**, *322*, 1843–1845.
28. Bikard, D.; Hatoum-Aslan, A.; Mucida, D.; Marraffini, L.A. CRISPR Interference Can Prevent Natural Transformation and Virulence Acquisition during *In Vivo* Bacterial Infection. *Cell Host Microbe* **2012**, *12*, 177–186.
29. Lillestøl, R.K.; Redder, P.; Garrett, R.A.; Brugger, K. A putative viral defence mechanism in archaeal cells. *Archaea* **2006**, *2*, 59–72.
30. Nakata, A.; Amemura, M.; Makino, K. Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J. Bacteriol.* **1989**, *171*, 3553–3556.
31. She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-Weiher, C.C.; Clausen, I.G.; Curtis, B.A.; De Moors, A.; *et al.* The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 7835–7840.
32. van Embden, J.D.; van Gorkom, T.; Kremer, K.; Jansen, R.; van der Zeijst, B.A.; Schouls, L.M. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. *J. Bacteriol.* **2000**, *182*, 2393–2401.
33. Horvath, P.; Romero, D.A.; Côté-Monvoisin, A.C.; Richards, M.; Deveau, H.; Moineau, S.; Boyaval, P.; Fremaux, C.; Barrangou, R. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* **2008**, *190*, 1401–1412.
34. Kunin, V.; Sorek, R.; Hugenholtz, P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **2007**, *8*, doi:10.1186/gb-2007-8-4-r61.
35. Grissa, I.; Vergnaud, G.; Pourcel, C. CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **2007**, *35*, W52–W57.
36. Brodt, A.; Lurie-Weinberger, M.N.; Gophna, U. CRISPR loci reveal networks of gene exchange in archaea. *Biol. Direct* **2011**, *6*, doi:10.1186/1745-6150-6-65.

37. Deveau, H.; Barrangou, R.; Garneau, J.E.; Labonté, J.; Fremaux, C.; Boyaval, P.; Romero, D.A.; Horvath, P.; Moineau, S. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **2008**, *190*, 1390–1400.
38. Stern, A.; Keren, L.; Wurtzel, O.; Amitai, G.; Sorek, R. Self-targeting by CRISPR: Gene regulation or autoimmunity? *Trends Genet.* **2010**, *26*, 335–340.
39. Agari, Y.; Sakamoto, K.; Tamakoshi, M.; Oshima, T.; Kuramitsu, S.; Shinkai, A. Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *J. Mol. Biol.* **2009**, *395*, 270–281.
40. Pougach, K.; Semenova, E.; Bogdanova, E.; Datsenko, K.A.; Djordjevic, M.; Wanner, B.L.; Severinov, K. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.* **2010**, *77*, 1367–1379.
41. Pul, Ü.; Wurm, R.; Arslan, Z.; Geissen, R.; Hofmann, N.; Wagner, R. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* **2010**, *75*, 1495–1512.
42. Przybilski, R.; Richter, C.; Gristwood, T.; Clulow, J.S.; Vercoe, R.B.; Fineran, P.C. Csy4 is responsible for CRISPR RNA processing in *Pectobacterium atrosepticum*. *RNA Biol.* **2011**, *8*, 517–528.
43. Yosef, I.; Goren, M.G.; Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **2012**, *40*, 5569–5576.
44. Godde, J.S.; Bickerton, A. The repetitive DNA elements called CRISPRs and their associated genes: Evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* **2006**, *62*, 718–729.
45. Brouns, S.J.; Jore, M.M.; Lundgren, M.; Westra, E.R.; Slijkhuis, R.J.; Snijders, A.P.; Dickman, M.J.; Makarova, K.S.; Koonin, E.V.; van der Oost, J. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **2008**, *321*, 960–964.
46. Jore, M.M.; Lundgren, M.; van Duijn, E.; Bultema, J.B.; Westra, E.R.; Waghmare, S.P.; Wiedenheft, B.; Pul, Ü.; Wurm, R.; Wagner, R.; *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **2011**, *18*, 529–536.
47. Wiedenheft, B.; Lander, G.C.; Zhou, K.; Jore, M.M.; Brouns, S.J.; van der Oost, J.; Doudna, J.A.; Nogales, E. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **2011**, *477*, 486–489.
48. Wiedenheft, B.; van Duijn, E.; Bultema, J.B.; Waghmare, S.P.; Zhou, K.; Barendregt, A.; Westphal, W.; Heck, A.J.; Boekema, E.J.; Dickman, M.J.; *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10092–10097.
49. Nam, K.H.; Haitjema, C.; Liu, X.; Ding, F.; Wang, H.; DeLisa, M.P.; Ke, A. Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System. *Structure* **2012**, *20*, 1574–1584.
50. Andersson, A.F.; Banfield, J.F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **2008**, *320*, 1047–1050.
51. Tyson, G.W.; Banfield, J.F. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* **2008**, *10*, 200–207.
52. Pride, D.T.; Salzman, J.; Haynes, M.; Rohwer, F.; Davis-Long, C.; White, R.A., 3rd.; Loomer, P.; Armitage, G.C.; Relman, D.A. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *ISME J.* **2012**, *6*, 915–926.

53. Garneau, J.E.; Dupuis, M.E.; Villion, M.; Romero, D.A.; Barrangou, R.; Boyaval, P.; Fremaux, C.; Horvath, P.; Magadan, A.H.; Moineau, S. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **2010**, *468*, 67–71.
54. Swarts, D.C.; Mosterd, C.; van Passel, M.W.; Brouns, S.J. CRISPR interference directs strand specific spacer acquisition. *PLoS One* **2012**, *7*, e35888.
55. Datsenko, K.A.; Pougach, K.; Tikhonov, A.; Wanner, B.L.; Severinov, K.; Semenova, E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **2012**, *3*, doi:10.1038/ncomms1937.
56. Cady, K.C.; Bondy-Denomy, J.; Heussler, G.E.; Davidson, A.R.; O'Toole, G.A. The CRISPR/Cas Adaptive Immune System of *Pseudomonas aeruginosa* Mediates Resistance to Naturally Occurring and Engineered Phages. *J. Bacteriol.* **2012**, *194*, 5728–5738.
57. Lopez-Sanchez, M.J.; Sauvage, E.; Da Cunha, V.; Clermont, D.; Ratsima Hariniaina, E.; Gonzalez-Zorn, B.; Poyart, C.; Rosinski-Chupin, I.; Glaser, P. The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Mol. Microbiol.* **2012**, *85*, 1057–1071.
58. Erdmann, S.; Garrett, R.A. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol. Microbiol.* **2012**, *85* 1044–1056.
59. Sapranauskas, R.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P.; Siksnys, V. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **2011**, *39*, 9275–9282.
60. Hatoum-Aslan, A.; Maniv, I.; Marraffini, L.A. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 21218–21222.
61. Wiedenheft, B.; Zhou, K.; Jinek, M.; Coyle, S.M.; Ma, W.; Doudna, J.A. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **2009**, *17*, 904–912.
62. Han, D.; Lehmann, K.; Krauss, G. SSO1450—A CAS1 protein from *Sulfolobus solfataricus* P2 with high affinity for RNA and DNA. *FEBS Lett.* **2009**, *583*, 1928–1932.
63. Babu, M.; Beloglazova, N.; Flick, R.; Graham, C.; Skarina, T.; Nocek, B.; Gagarinova, A.; Pogoutse, O.; Brown, G.; Binkowski, A.; *et al.* A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol. Microbiol.* **2011**, *79*, 484–502.
64. Beloglazova, N.; Brown, G.; Zimmerman, M.D.; Proudfoot, M.; Makarova, K.S.; Kudritska, M.; Kochinyan, S.; Wang, S.; Chruszcz, M.; Minor, W.; *et al.* A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.* **2008**, *283*, 20361–20371.
65. Samai, P.; Smith, P.; Shuman, S. Structure of a CRISPR-associated protein Cas2 from *Desulfovibrio vulgaris*. *Acta Crystallogr. F Struct. Biol. Cryst. Commun.* **2010**, *66*, 1552–1556.
66. Nam, K.H.; Ding, F.; Haitjema, C.; Huang, Q.; Delisa, M.P.; Ke, A. Double-stranded Endonuclease Activity in *B. halodurans* Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated Cas2 Protein. *J. Biol. Chem.* **2012**, Epub ahead of print.

67. Plagens, A.; Tjaden, B.; Hagemann, A.; Randau, L.; Hensel, R. Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *J. Bacteriol.* **2012**, *194*, 2491–2500.
68. Goren, M.G.; Yosef, I.; Auster, O.; Qimron, U. Experimental Definition of a Clustered Regularly Interspaced Short Palindromic Duplicon in *Escherichia coli*. *J. Mol. Biol.* **2012**, *423*, 14–16.
69. Westra, E.R.; Brouns, S.J. The rise and fall of CRISPRs—Dynamics of spacer acquisition and loss. *Mol. Microbiol.* **2012**, *85*, 1021–1025.
70. Mojica, F.J.; Diez-Villasenor, C.; Garcia-Martinez, J.; Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **2009**, *155*, 733–740.
71. Richter, H.; Zoepfel, J.; Schermuly, J.; Maticzka, D.; Backofen, R.; Randau, L. Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. *Nucleic Acids Res.* **2012**, Epub ahead of print.
72. Sashital, D.G.; Jinek, M.; Doudna, J.A. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat. Struct. Mol. Biol.* **2011**, *18*, 680–687.
73. Gesner, E.M.; Schellenberg, M.J.; Garside, E.L.; George, M.M.; MacMillan, A.M. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat. Struct. Mol. Biol.* **2011**, *18*, 688–692.
74. Haurwitz, R.E.; Jinek, M.; Wiedenheft, B.; Zhou, K.; Doudna, J.A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **2010**, *329*, 1355–1358.
75. Sternberg, S.H.; Haurwitz, R.E.; Doudna, J.A. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* **2012**, *18*, 661–672.
76. Haurwitz, R.E.; Sternberg, S.H.; Doudna, J.A. Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *EMBO J.* **2012**, *31*, 2824–2832.
77. Lintner, N.G.; Kerou, M.; Brumfield, S.K.; Graham, S.; Liu, H.; Naismith, J.H.; Sdano, M.; Peng, N.; She, Q.; Copie, V.; *et al.* Structural and Functional Characterization of an Archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Complex for Antiviral Defense (CASCADE). *J. Biol. Chem.* **2011**, *286*, 21643–21656.
78. Sashital, D.G.; Wiedenheft, B.; Doudna, J.A. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell.* **2012**, *46*, 606–615.
79. Semenova, E.; Jore, M.M.; Datsenko, K.A.; Semenova, A.; Westra, E.R.; Wanner, B.; van der Oost, J.; Brouns, S.J.; Severinov, K. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10098–10103.
80. Westra, E.R.; van Erp, P.B.; Kunne, T.; Wong, S.P.; Staals, R.H.; Seegers, C.L.; Bollen, S.; Jore, M.M.; Semenova, E.; Severinov, K.; *et al.* CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Mol. Cell* **2012**, *46*, 595–605.
81. Howard, J.A.; Delmas, S.; Ivančić-Baće, I.; Bolt, E.L. Helicase dissociation and annealing of RNA-DNA hybrids by *Escherichia coli* Cas3 protein. *Biochem. J.* **2011**, *439*, 85–95.
82. Sinkunas, T.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P.; Siksnys, V. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **2011**, *30*, 1335–1342.

83. Beloglazova, N.; Petit, P.; Flick, R.; Brown, G.; Savchenko, A.; Yakunin, A.F. Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J.* **2011**, *30*, 4616–4627.
84. Deltcheva, E.; Chylinski, K.; Sharma, C.M.; Gonzales, K.; Chao, Y.; Pirzada, Z.A.; Eckert, M.R.; Vogel, J.; Charpentier, E. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **2011**, *471*, 602–607.
85. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337*, 816–821.
86. Carte, J.; Wang, R.; Li, H.; Terns, R.M.; Terns, M.P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **2008**, *22*, 3489–3496.
87. Carte, J.; Pfister, N.T.; Compton, M.M.; Terns, R.M.; Terns, M.P. Binding and cleavage of CRISPR RNA by Cas6. *RNA* **2010**, *16*, 2181–2188.
88. Wang, R.; Preamplume, G.; Terns, M.P.; Terns, R.M.; Li, H. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: Recognition and cleavage. *Structure* **2011**, *19*, 257–264.
89. Hale, C.R.; Majumdar, S.; Elmore, J.; Pfister, N.; Compton, M.; Olson, S.; Resch, A.M.; Glover, C.V., 3rd.; Graveley, B.R.; Terns, R.M.; *et al.* Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell* **2012**, *45*, 292–302.
90. Hale, C.R.; Zhao, P.; Olson, S.; Duff, M.O.; Graveley, B.R.; Wells, L.; Terns, R.M.; Terns, M.P. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **2009**, *139*, 945–956.
91. Hale, C.; Kleppe, K.; Terns, R.M.; Terns, M.P. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* **2008**, *14*, 2572–2579.
92. Zhang, J.; Rouillon, C.; Kerou, M.; Reeks, J.; Brugger, K.; Graham, S.; Reimann, J.; Cannone, G.; Liu, H.; Albers, S.V.; *et al.* Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol. Cell* **2012**, *45*, 303–313.
93. Marraffini, L.A.; Sontheimer, E.J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **2010**, *463*, 568–571.
94. Young, J.C.; Dill, B.D.; Pan, C.; Hettich, R.L.; Banfield, J.F.; Shah, M.; Fremaux, C.; Horvath, P.; Barrangou, R.; VerBerkmoes, N.C. Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus*. *PLoS One* **2012**, *7*, e38077.
95. Dorman, C.J. H-NS, the genome sentinel. *Nat. Rev. Microbiol.* **2007**, *5*, 157–161.
96. Westra, E.R.; Pul, Ü.; Heidrich, N.; Jore, M.M.; Lundgren, M.; Stratmann, T.; Wurm, R.; Raine, A.; Mescher, M.; van Heereveld, L.; *et al.* H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* **2010**, *77*, 1380–1393.
97. Lucchini, S.; Rowley, G.; Goldberg, M.D.; Hurd, D.; Harrison, M.; Hinton, J.C. H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog.* **2006**, *2*, e81.
98. Navarre, W.W.; Porwollik, S.; Wang, Y.; McClelland, M.; Rosen, H.; Libby, S.J.; Fang, F.C. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* **2006**, *313*, 236–238.
99. Mojica, F.J.; Díez-Villaseñor, C. The on-off switch of CRISPR immunity against phages in *Escherichia coli*. *Mol. Microbiol.* **2010**, *77*, 1341–1345.

100. Medina-Aparicio, L.; Rebollar-Flores, J.E.; Gallego-Hernández, A.L.; Vázquez, A.; Olvera, L.; Gutiérrez-Ríos, R.M.; Calva, E.; Hernández-Lucas, I. The CRISPR/Cas Immune System Is an Operon Regulated by LeuO, H-NS, and Leucine-Responsive Regulatory Protein in *Salmonella enterica* Serovar Typhi. *J. Bacteriol.* **2011**, *193*, 2396–2407.
101. Hernández-Lucas, I.; Gallego-Hernández, A.L.; Encarnación, S.; Fernández-Mora, M.; Martínez-Batallar, A.G.; Salgado, H.; Oropeza, R.; Calva, E. The LysR-type transcriptional regulator LeuO controls expression of several genes in *Salmonella enterica* serovar Typhi. *J. Bacteriol.* **2008**, *190*, 1658–1670.
102. Majumder, A.; Fang, M.; Tsai, K.J.; Ueguchi, C.; Mizuno, T.; Wu, H.Y. LeuO expression in response to starvation for branched-chain amino acids. *J. Biol. Chem.* **2001**, *276*, 19046–19051.
103. Drahos, D.J.; Hendrix, R.W. Effect of bacteriophage lambda infection on synthesis of *groE* protein and other *Escherichia coli* proteins. *J. Bacteriol.* **1982**, *149*, 1050–1063.
104. MacRitchie, D.M.; Buelow, D.R.; Price, N.L.; Raivio, T.L. Two-component signaling and gram negative envelope stress response systems. *Adv. Exp. Med. Biol.* **2008**, *631*, 80–110.
105. Perez-Rodriguez, R.; Haitjema, C.; Huang, Q.; Nam, K.H.; Bernardis, S.; Ke, A.; DeLisa, M.P. Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. *Mol. Microbiol.* **2011**, *79*, 584–599.
106. Baranova, N.; Nikaido, H. The *baeSR* two-component regulatory system activates transcription of the *yegMNOB* (*mdtABCD*) transporter gene cluster in *Escherichia coli* and increases its resistance to novobiocin and deoxycholate. *J. Bacteriol.* **2002**, *184*, 4168–4176.
107. Yosef, I.; Goren, M.G.; Kiro, R.; Edgar, R.; Qimron, U. High-temperature protein G is essential for activity of the *Escherichia coli* clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20136–20141.
108. Chen, C.C.; Ghole, M.; Majumder, A.; Wang, Z.; Chandana, S.; Wu, H.Y. LeuO-mediated transcriptional derepression. *J. Biol. Chem.* **2003**, *278*, 38094–38103.
109. Viswanathan, P.; Murphy, K.; Julien, B.; Garza, A.G.; Kroos, L. Regulation of *dev*, an operon that includes genes essential for *Myxococcus xanthus* development and CRISPR-associated genes and repeats. *J. Bacteriol.* **2007**, *189*, 3738–3750.
110. Shinkai, A.; Kira, S.; Nakagawa, N.; Kashiwara, A.; Kuramitsu, S.; Yokoyama, S. Transcription activation mediated by a cyclic AMP receptor protein from *Thermus thermophilus* HB8. *J. Bacteriol.* **2007**, *189*, 3891–3901.
111. Deng, L.; Kenchappa, C.S.; Peng, X.; She, Q.; Garrett, R.A. Modulation of CRISPR locus transcription by the repeat-binding protein Cbp1 in *Sulfolobus*. *Nucleic Acids Res.* **2012**, *40*, 2470–2480.
112. Peng, X.; Brugger, K.; Shen, B.; Chen, L.; She, Q.; Garrett, R.A. Genus-specific protein binding to the large clusters of DNA repeats (short regularly spaced repeats) present in *Sulfolobus* genomes. *J. Bacteriol.* **2003**, *185*, 2410–2417.