Published Online on 24 January 2018

Proc Indian Natn Sci Acad **84** No. 2 June 2018 pp. 455-477

DOI: 10.16943/ptinsa/2017/49241

Review Article

Cutting it Right: Plasticity and Strategy of CRISPR RNA Specific Nucleases

ANKITA PUNETHA, K N R YOGANAND, SIDDHARTH NIMKAR and B ANAND*
Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India

(Received on 22 September 2017; Revised on 26 December 2017; Accepted on 01 January 2018)

The existence of adaptive immunity in prokaryotes came to light with the discovery of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) in association with CRISPR-associated (Cas) proteins. This RNA mediated defence system confers resistance against the invading mobile genetic elements such as phages and plasmids. The CRISPR-Cas system operates by forming a ribonucleoprotein complex that comprises of an invader derived small RNA and Cas protein(s). Herein the small RNA acts as a guide to recognize the nucleic acid target whereas the Cas proteins facilitate target annihilation. Given the cardinal role adopted by this small RNA, its maturation from the pre-CRISPR transcript forms a pivot for successful adaptive immunity. The mandate to generate the guide CRISPR RNA (crRNA) is fulfilled by specific endoribonuclease, which processes the pre-crRNA transcript in between the repeats to liberate the individual interfering units. Intriguingly, while some endoRNases of the CRISPR system are able to process the pre-crRNA independently, others require participation of additional Cas proteins, which form a multi-protein complex for processing the pre-crRNA. Additionally, some CRISPR variants require non-Cas auxiliary factors to process the pre-crRNA. The mode of crRNA maturation further diversifies as the endoRNases in CRISPR variants coevolve with repeat clusters that exhibit high diversity in sequence and folding. Therefore, the maturation of a specific crRNA requires a distinct mechanistic solution for substrate discrimination by these endoRNases, the understanding of which is essential for appreciating the CRISPR biology. This review highlights the vivid modes adopted by the diverse CRISPR-Cas systems to generate the mature crRNA.

Keywords: CRISPR RNA; CRISPR-Cas System; Guide RNA; Cascade; Cas5; Cas6; Csy4; Cas9; RNase; DNase; Surveillance Complex

Introduction

In order to survive, all organisms must overcome their predators. The prokaryotes and their viral predators coexist in natural and man-made environment and therefore the prokaryotes face a constant threat of getting infected by phages. This results in acute pressure on the microbial community to coevolve with their predators causing an evolutionary arms race between prey and predator. Pitted against a hostile environment, prokaryotes have developed multi-layered antiviral defense systems, which act at various stages of the infection cycle of the invader. These include various innate defense systems like surface exclusion (receptor downregulation or masking), super infection exclusion (Sie systems), restriction-modification systems (R-M and R-M like systems), and abortive infection systems (Abi) (Hyman and Abedon, 2010; Labrie *et al.*, 2010; Westra *et al.*, 2012a). These innate defense mechanisms are diffusive in nature and do not rely on the identity of the predator to elicit a response (Fig 1). Added to this repertoire of arsenals, the recently discovered Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) in association with CRISPR-associated (Cas) proteins endows the bacteria and archaea with an adaptive immunity (Jansen *et al.*, 2002a; Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Pourcel *et al.*, 2005; Makarova *et al.*, 2006; Barrangou *et al.*, 2007; Brouns *et al.*, 2008; Marraffini

*Author for Correspondence: E-mail: banand@iitg.ernet.in

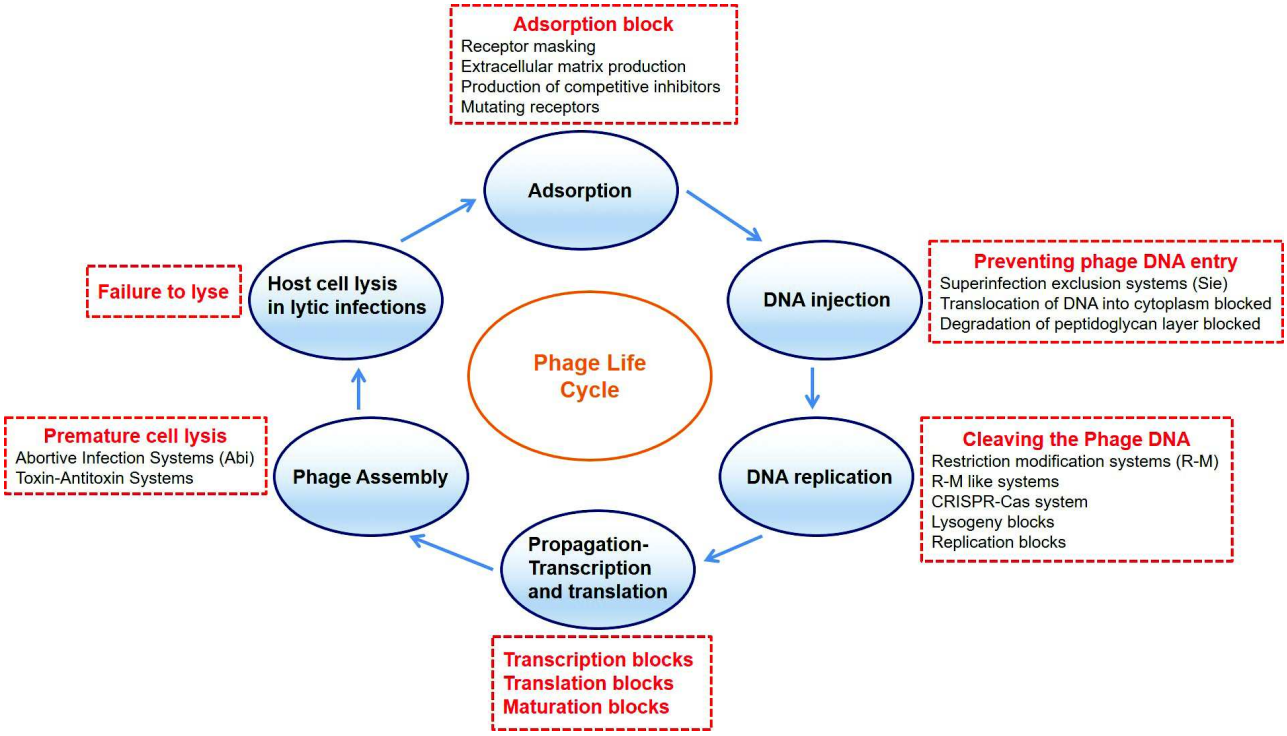

Fig. 1: The various antiviral defense systems of prokaryotes to counter the different stages of phage life cycle. The phage life cycle is shown in the ovals. The various defense mechanisms of the host operating at a particular stage of phage life cycle are shown in rectangle corresponding to that stage

and Sontheimer, 2008; Hale *et al.*, 2009; Garneau *et al.*, 2010; Hale *et al.*, 2012; Zhang *et al.*, 2012; Mojica and Rodriguez-Valera, 2016) (Fig. 2). This system acquires a fragment of foreign nucleic acid during invasion and employs this as an immunological memory in order to specifically recognize and neutralize the recurrent infections. In contrast to other known defense mechanisms in prokaryotes, CRISPR-Cas system is an RNA mediated adaptive immune system that holds functional analogy to the RNA interference (RNAi) in higher organisms. Both these systems utilize a guide RNA to direct the effector protein complex to silence the target (Fire *et al.*, 1998; Barrangou *et al.*, 2007; Brouns *et al.*, 2008; Marraffini and Sontheimer, 2008; Hale *et al.*, 2009; Garneau *et al.*, 2010; Hale *et al.*, 2012; Zhang *et al.*, 2012; Swarts *et al.*, 2014; Swarts *et al.*, 2015; Miyoshi *et al.*, 2016; Doxzen and Doudna, 2017).

Approximately half of all sequenced bacteria and nearly all sequenced archaea harbour one or more CRISPR loci (Grissa *et al.*, 2007; Rousseau *et al.*, 2009; Makarova *et al.*, 2015), which is composed of a series of direct repeats of about 20-50 bp separated by similarly sized unique invader derived spacer sequences (Ishino *et al.*, 1987; Nakata *et al.*, 1989; Jansen *et al.*, 2002b; Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Pourcel *et al.*, 2005) along with *cas* genes in their immediate vicinity (Jansen *et al.*, 2002a; Haft *et al.*, 2005). A low complexity, A/T rich, noncoding sequence can be found upstream of the CRISPR array. This is termed as “leader” and typically this region harbours promoter for CRISPR transcription (Jansen *et al.*, 2002b; Lillestol *et al.*, 2006; Lillestol *et al.*, 2009; Pougach *et al.*, 2010). In those CRISPR arrays that are usually preceded by a leader sequence, individual experimental studies reveal that the leader contains signals for CRISPR-Cas adaptation, which facilitates the spacer incorporation at leader-repeat junction in a polarized fashion. (Jansen *et al.*, 2002b; Lillestol *et al.*, 2009; Bernick *et al.*, 2012; Erdmann and Garrett, 2012; Yosef *et al.*, 2012; Diez-Villasenor *et al.*, 2013; Erdmann *et al.*, 2014; Wei *et al.*, 2015a; Wei *et al.*, 2015b; Nunez *et al.*, 2016; Wang *et al.*, 2016; Yoganand *et al.*, 2017). However, in case of leaderless CRISPR array as in *Neisseria* strains (type II-C), the new spacer integration seems to occur

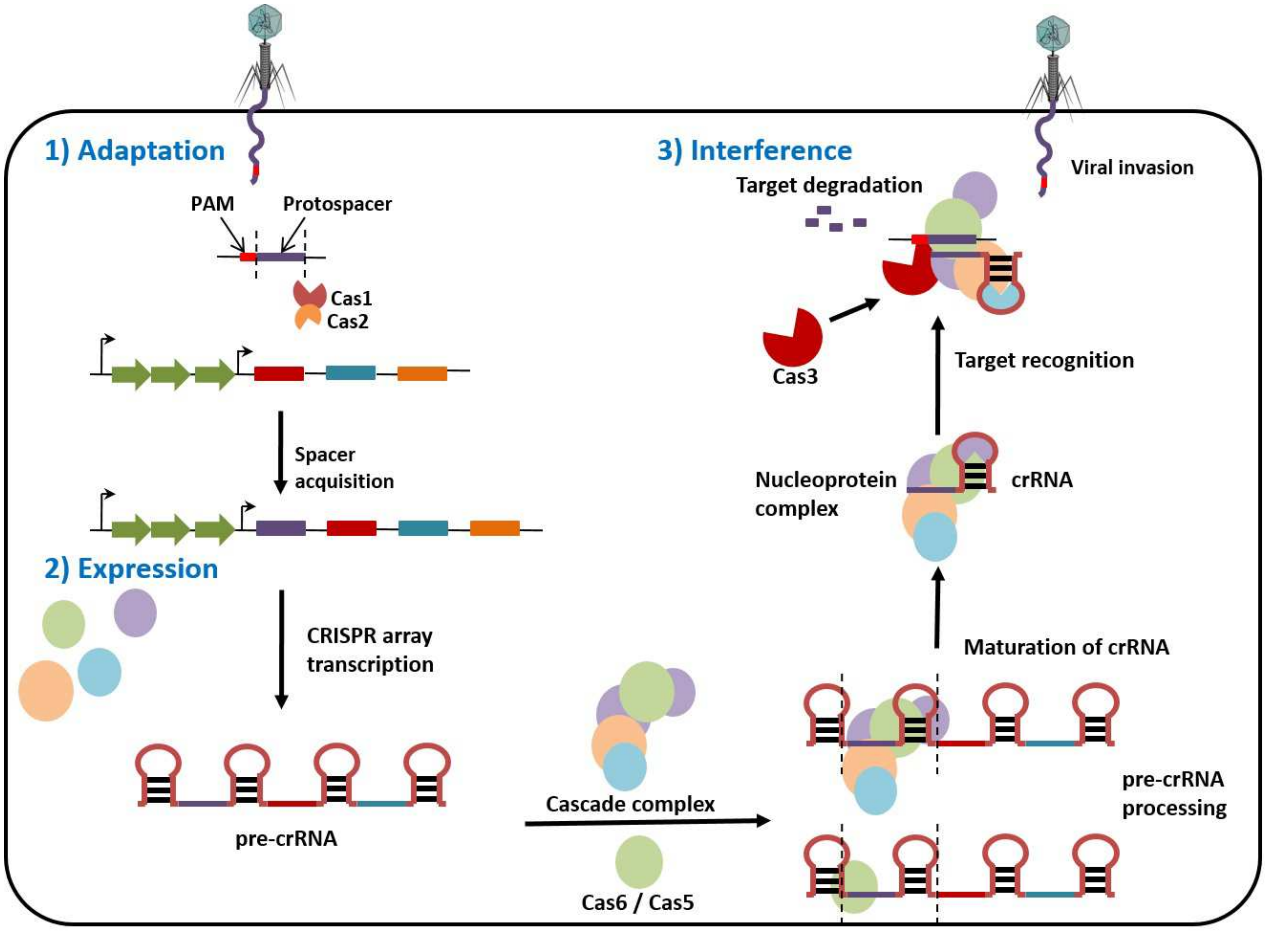

Fig. 2: Mechanistic overview of CRISPR-mediated immunity. The different stages of CRISPR-Cas immunity as exemplified in type I system is shown. Stage 1 – An infected prokaryote incorporates a piece of invading nucleic acid (the protospacer) into a genomic CRISPR locus as a new spacer at the leader proximal end of the CRISPR. Stage 2 – The CRISPR locus is transcribed as a single pre-crRNA transcript that is processed by Cas endoRNase either alone or in association with other Cas proteins into mature crRNAs. Stage 3 – The crRNA aided large multi-protein complex targets the invading nucleic acid. The sequence complementarity between the crRNA spacer and the invading nucleic acid triggers its degradation

downstream of the array (Zhang *et al.*, 2013). In such CRISPR arrays, transcription may occur from the promoters proximal to CRISPR array (Lillestol *et al.*, 2006; Lillestol *et al.*, 2009; Zhang *et al.*, 2013).

The CRISPR-Cas system shows high diversity owing to the dynamic evolution of CRISPR locus, involving numerous rearrangements and horizontal transfer of complete or individual modules and therefore this precludes a simple phylogenetic classification. Recent classification schemes of CRISPR-Cas systems that evolved over the years combine the analysis of signature protein families, features of the architecture of *cas* loci and the information of the effector modules involved in interference and categorize them into two distinct classes, six types and nineteen subtypes (Haft *et al.*, 2005; Makarova *et al.*, 2006; Makarova *et al.*, 2015; Makarova *et al.*, 2017a; Makarova *et al.*, 2017b; Shmakov *et al.*, 2017). Class 1 employs multi-protein effector module for target interference whereas Class 2 utilizes multi-domain single effector protein. These highly diverse CRISPR-Cas variants display major structural and functional differences in their mode of generating resistance against invading nucleic acids. Overall, the CRISPR immunity can be operationally distinguished into three stages – adaptation, maturation and interference (Fig. 2) and the functional mechanism seems to vary among the different types of CRISPR-Cas system.

Stage 1: Adaptation

CRISPR adaptation involves acquiring the immunological memory of infection in the form of nucleic acids from invading mobile genetic elements (MGEs) (Barrangou *et al.*, 2007; Brouns *et al.*, 2008; Deveau *et al.*, 2008; Datsenko *et al.*, 2012; Yosef *et al.*, 2012; Wei *et al.*, 2015b; Jackson *et al.*, 2017). The invader-derived nucleic acid prior to integration into the CRISPR array as "spacers" is referred as protospacer. The integration of new protospacers occurs at one end of the CRISPR array in a polarized fashion and this biased integration seems to maintain the chronology of the infection (Yosef *et al.*, 2012; McGinn and Marraffini, 2016; Nunez *et al.*, 2016; Wright and Doudna, 2016; Wright *et al.*, 2017; Yoganand *et al.*, 2017). At the end of adaptation process, a new spacer gets integrated into the CRISPR array followed by the duplication of the repeat. The selection of protospacer fragment for integration is based on the presence of a small sequence motif (2-7 nt) referred to as protospacer adjacent motif (PAM), which can be located either at the 5'- or the 3'-end of the protospacer (Deveau *et al.*, 2008; Horvath *et al.*, 2008; Mojica *et al.*, 2009; Wei *et al.*, 2015a). In CRISPR-Cas systems, two modes of spacer adaptation are known (i) Naïve adaptation which occurs during a maiden infection and it involves the participation of adaptation machinery alone (Yosef *et al.*, 2012; Arslan *et al.*, 2014; Hooton and Connerton, 2014) and (ii) Primed adaptation that occurs when the CRISPR contains a previously integrated spacer that is complementary to the invading DNA and requires both adaptation and interference machinery (Datsenko *et al.*, 2012; Fineran *et al.*, 2014; Hynes *et al.*, 2014; Li *et al.*, 2014; Künne *et al.*, 2016; Rao *et al.*, 2016; Staals *et al.*, 2016). Adaptation leads to highly efficient and selective acquisition of spacers with consensus PAM and most spacers are capable of protecting the host (Datsenko *et al.*, 2012; Swarts *et al.*, 2012; Li *et al.*, 2014; Richter *et al.*, 2014; Wei *et al.*, 2015b). The adaptation module is largely invariant across CRISPR-Cas systems. Cas1 and Cas2 that are ubiquitously conserved across CRISPR-Cas systems are essential for protospacer integration. In addition, this process is orchestrated by several host derived accessory factors (Ivanèiæ-Baæe *et al.*, 2015; Nunez *et al.*, 2016; Fagerlund *et al.*, 2017; Wright *et al.*, 2017; Yoganand *et al.*, 2017).

Stage 2: Maturation

In the second stage termed as maturation, the CRISPR locus is transcribed to form a single primary transcript called the pre-CRISPR RNA (pre-crRNA) by RNA polymerase. Subsequently, the pre-crRNA is cleaved endonucleolytically by specific endoribonuclease to yield mature crRNA, which then binds to Cas effector proteins and serves as guide in the third stage of CRISPR-mediated defense (Carte *et al.*, 2008; Deltcheva *et al.*, 2011; Sashital *et al.*, 2011; Garside *et al.*, 2012; Nam *et al.*, 2012; Fonfara *et al.*, 2016). Thus, based on its function, the crRNA is also referred as prokaryotic silencing (psiRNA) (Hale *et al.*, 2009) or guide RNA (Brouns *et al.*, 2008; Carte *et al.*, 2008). The mechanism of crRNA maturation is elaborated in later part of the section (see below).

Stage 3: Interference

The last stage of CRISPR-Cas system is Interference where the invading plasmid or phage DNA is targeted by nucleolytic cleavage thereby protecting the host from invasive attack. Degradation of the target DNA is initiated with the identification of protospacer by an interference complex (Brouns *et al.*, 2008; Garneau *et al.*, 2010). In Class 1, interference complex comprises of a multi-subunit ribonucleoprotein (RNP) complex called as CRISPR- associated complex for antiviral defence (Cascade) whereas, in Class 2, a single protein degrades the target. With the exception of type III subtype in Class 1, recognition of target by the CRISPR machinery requires the presence of a PAM (Mojica *et al.*, 2009; Marraffini and Sontheimer, 2010; Semenova *et al.*, 2011; Gasiunas *et al.*, 2012; Jinek *et al.*, 2012; Pyenson *et al.*, 2017). The primary function of PAM sequence is to protect the host genome against self-targeting and to selectively target the invading DNA (Westra *et al.*, 2013). PAM is recognised by the interference complex and mutations in PAM sequence leads to target evasion (Sapranauskas *et al.*, 2011; Westra *et al.*, 2012b; Hayes *et al.*, 2016). In addition, cleavage of protospacer depends on the sequence located proximal to the PAM called as seed sequence. Seed sequences are usually 8-9 nt long and are crucial for interference (Semenova *et al.*, 2011; Wiedenheft *et al.*, 2011b). Target binding relies on the base pairing between the protospacer and the spacer region of crRNA, which

is a part of the interference complex. crRNA binding leads to the displacement of the non-complementary strand, forming a DNA/RNA hybrid called as R-loop. In addition, the target cleavage mechanism varies strikingly between Class 1 and Class 2 (Hille and Charpentier, 2016; Mohanraju *et al.,* 2016; Nishimasu and Nureki, 2016; Wright *et al.,* 2016; Makarova *et al.*, 2017a; Makarova *et al.*, 2017b).

In this review, we focus on the recent advancements on the distinct modes of maturation of the crRNA among CRISPR variants, which are presented in the following section.

Mechanism of CRISPR RNA Maturation

The maturation stage can be further divided into two sequential steps (i) the transcription of CRISPR locus and (ii) the crRNA processing, both of which are required for successful interference.

The Transcription of CRISPR Locus

Transcription of a CRISPR locus into a primary transcript or pre-crRNA was first observed in high throughput analyses of non-coding RNAs in the archaea *Archaeoglobus fulgidus* and *Sulfolobus solfataricus* P2 (Tang *et al.*, 2002; Tang *et al.*, 2005). The transcripts ranged from a minimum length corresponding to the distance between two successive repeats in the CRISPR cluster to higher order multiples of this single repeat-spacer unit. The detected sequences corresponded to various positions of the CRISPR array suggesting that the whole locus is transcribed as a long transcript, which is subsequently processed into smaller repeat- spacer units. Later, the transcription of CRISPR loci was shown in a number of species, such as *Escherichia coli* (Brouns *et al.*, 2008; Pougach *et al.*, 2010; Pul *et al.*, 2010), *Thermus thermophilus* (Agari *et al.*, 2010), *Xanthomonas oryza*e (Semenova *et al.*, 2009), *Pyrococcus furiosus* (Hale *et al.*, 2009), *Staphylococcus epidermidis* (Marraffini and Sontheimer, 2008), *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius* (Tang *et al.*, 2005; Lillestol *et al.*, 2009). All these studies suggested the unidirectional transcription from the leader proximal end of the locus. The analysis of the transcription start-sites and leader regions of the *Sulfolobales* revealed putative BRE and TATA box motifs within 25 nt of the transcription start site in the leader sequence. This suggested the existence of promoter in the leader region (Lillestol *et al.*, 2009). Also, the reverse transcripts of the repeat clusters were detected in *S. solfataricus* and *S. acidocaldarius* (Lillestol *et al.*, 2009), suggesting the existence of putative BRE and TATA box elements downstream of the CRISPR arrays, but their processing seems to be less efficient and therefore it remains unknown whether they produce functional repeat-spacer units. Generally, Cas proteins and pre-crRNA are expressed constitutively but under certain conditions these levels can be regulated, suggesting a feedback mechanism to monitor the presence of invasive nucleic acid. The process has striking differences among various CRISPR-Cas systems, which highlight the remarkable ability of CRISPR systems to adapt and evolve according to environmental pressures. The following types of regulation have been observed in the transcription of CRISPR arrays and *cas* genes

1. crRNAs are often quantitatively identified as dominant form of small RNAs in bacteria and archaea. This suggests the constitutive expression of CRISPR loci, which can be further induced by viral challenge. This is consistent with a surveillance mode of action and has been observed in archaea (Tang *et al.*, 2002; Hale *et al.*, 2009; Lillestol *et al.*, 2009; Semenova *et al.*, 2009). Further, expression can be upregulated by cAMP receptor protein in response to phage infection (Agari *et al.*, 2010). This pathway also gets activated during carbon limitation stress. Another study suggests the upregulation of *cas* gene expression in response to envelope stress (Perez-Rodriguez *et al.*, 2011).

2. The negative regulation of *cas* operon by DevS along with the *dev* operon, which controls developmental stages has been observed in *Myxococcus xanthus* (Viswanathan *et al.*, 2007). In *E. coli*, transcription is suppressed by the Heat-stable Nucleoid Structuring protein (H-NS) - a typical transcriptional repressor in gram-negative bacteria - which binds to the promoter region in the leader sequence of the CRISPR locus (Pul *et al.*, 2010). This repression is relieved by the transcriptional regulator LeuO, by binding to the same genomic region and reversing the cooperative binding of H-NS

dimers along the DNA and also by directly or indirectly promoting the CRISPR-associated transcription (Westra *et al.*, 2010).

Biogenesis of CRISPR RNA

The biogenesis of crRNA involves processing of the CRISPR array (consisting of repeat-spacer units) by endoRNases, which cleave in-between the repeats of pre-CRISPR transcript and liberate each spacer unit flanked by partial repeat sequence on both sides. This in some cases may get further trimmed on edges in order to become a mature crRNA. Distinct set of enzymes is employed to process pre-crRNAs in various types of CRISPR-Cas systems (Makarova *et al.*, 2015; Mohanraju *et al.*, 2016; Makarova *et al.*, 2017a; Makarova *et al.*, 2017b; Shmakov *et al.*, 2017). These endoRNases perform two specific functions First, they recognize and process the precursor transcript to generate the mature form of crRNAs and second they retain the mature crRNA for subsequent loading onto the respective effector proteins or complexes that mediate interference. In some CRISPR variants such as type I-A, I-B, I-D, II A-C and III A-B, the endoRNase that processes the pre-CRISPR transcript doesn't seem to be a part of the effector complex (Makarova *et al.*, 2017a; Makarova *et al.*, 2017b).

Diversity in crRNA Maturation

The six types and the nineteen subtypes of CRISPR-Cas employ distinct endoRNases to process the pre-crRNA transcript (Makarova *et al.*, 2015; Mohanraju *et al.*, 2016; Makarova *et al.*, 2017a; Makarova *et al.*, 2017b; Shmakov *et al.*, 2017). The current repertoire of the known endoribonucleases that are involved in maturation of the crRNA in currently known CRISPR-Cas system is shown in Table 1. Interestingly, the above classification of the CRISPR-Cas system is based exclusively on the diversity of the *cas* component. However, another dimension to this classification can be added based on the diversity of the CRISPR component (Kunin *et al.*, 2007). This will reveal specific relationships between CRISPR-Cas subtypes and is also advantageous in instances where CRISPR arrays occur in the absence of *cas* genes. The repeats present in various CRISPR- Cas types and subtypes differ in their sequence and structure, and can attain a structured architecture in the form of stem-loop or remain unstructured. The propensity of each repeat sequence to form stable secondary structures, typically a stem-loop structure, depends on the palindromic nature of the repeat sequence. Based on the sequence similarity, the CRISPR repeats can be organized into multiple clusters, resulting in some clusters with stable, highly conserved RNA secondary structures, while others lack the detectable structures. The alignment of the repeats, followed by clustering of the sequence similarity, generated 33 repeat clusters, 12 of which contained 10 or more members, with the largest cluster (cluster 1) containing 94 repeat sequences (Kunin *et*

Table 1: Distinct crRNA processor of various CRISPR-Cas system

Type	EndoRNase
I-A	Cas6
I-B	Cas6
I-C	Cas5
I-U	Cas5-Cas6
I-D	Cas6
I-E	Cas6
I-F	Cas6
III-A	Cas6
III-B	Cas6
III-C	
III-D	
IV	
II-A	RNase III #
II-B	RNase III #
II-C	RNase III #
V-A	Cpf1/Cas12a
V-B	C2c1/Cas12b*
V-C	C2c3/Cas12c*
V-U	
VI-A	C2c2/Cas13a1
VI-B	Cas13b*
VI-C	Cas13a2*

The crRNA processing endoRNases that belong to different subtypes of CRISPR-Cas system are shown. The Type I-U system shows the fusion of the Cas5-Cas6 proteins, the functionality of which needs to be explored. In type II system, the requirement of Cas9 to direct RNase III to specifically cleave the repeat is shown by hash (#). The possibility of the effector nucleases involvement in crRNA maturation is indicated by star (*). The boxes coloured with pink indicate that the crRNA processor is yet to be identified in those CRISPR- Cas systems

al., 2007). Some clusters contained repeats from organisms as distantly related as archaea and bacteria, supporting the horizontal transfer of the CRISPR-Cas systems between microorganisms. Based on this repeat clustering, the various CRISPR-Cas systems can be associated with diverse clusters – differing in the sequence and structure of repeats and also on the phyletic distribution as shown in Table 2. Among the structured clusters, variations were observed in sequence, the length of the stem and its position relative to the 5'- or 3'-end of the repeat. For example, in cluster 4 the stem is typically 5 bp long and is found in the middle of the repeat, whereas in cluster 3 it is typically 7 bp long, and is found towards the 5'-end of the repeat. Thus, the specific recognition of a repeat RNA requires a distinct mechanistic solution for substrate discrimination by endoRNases, suggesting variations in the mechanism of substrate recognition and processing among CRISPR variants.

Type I and Type III CRISPR Systems

A CRISPR-specific endoribonuclease Cas6 associated with most type I (except I-C) and type III (except III-C and III-D) binds and cleaves the repeat elements in a sequence and structure specific manner (Brouns *et al.*, 2008; Carte *et al.*, 2010; Gesner *et al.*, 2011; Lintner *et al.*, 2011; Sashital *et al.*, 2011; Wang *et al.*, 2011; Haurwitz *et al.*, 2012; Sternberg *et al.*, 2012; Shao and Li, 2013; Shao *et al.*, 2016). These proteins are part of the RAMP (Repeat Associated Mysterious Proteins) superfamily which encompasses a plethora of protein families having tandem or single ferredoxin-like folds, also called as RNA recognition motifs (RRM) for RNA binding (Haurwitz *et al.*, 2010; Wang *et al.*, 2011). In most type I systems (except I-A and I-B), Cas6 associates with other Cas proteins to form a large RNP surveillance complex referred to as Cascade that takes part in both the crRNA maturation and invader

Table 2: Clustering of repeats belonging to different types of CRISPR system

	Repeat Clusters											
Repeat architecture	1	2	3	4	5	6	7	8	9	10	11	12
Phyletic distribution	bacterial	bacterial	bacterial	bacterial	bacterial/archaeal	archaeal	archaeal	bacterial	archaeal	bacterial	archaeal	bacterial
Repeat folding	unfolded	Folded	folded	folded	folded	unfolded	unfolded	folded	unfolded	unfolded	unfolded	folded
type I-A / Apern												
type I-B / Hmari												
type I-B / Tneap												
type I-C / Dvulg												
type I-E / Ecoli												
type I-F / Ypest												
type II-A / Nmeni												
type III-A / Mtube												

The information is derived from Kunin et al (2007) and the classification scheme that is in vogue is presented. Association of the specific repeat cluster with particular CRISPR variant is indicated by cell filled with green colour. The repeat folding and phyletic distribution of the various clusters among the CRISPR variants are also shown. The clusters (1-12) differ in the sequence, size and position of stem in the repeat. More than one kind of repeat cluster can be associated with a particular Cas subtypes. Exceptionally, cluster 5 is not associated exclusively with any of the recognized Cas subtypes and occurs in genomes that contain other CRISPR repeat clusters as well. It was found to be associated with most of the core *cas* genes (*cas*1-4 and *cas*6) and lacks any of the additional type-defining genes (Kunin *et al.*, 2007)

silencing (Brouns *et al.*, 2008; Lintner *et al.*, 2011; Hochstrasser and Doudna, 2015; Hille and Charpentier, 2016; Makarova *et al.*, 2017a). In such cases, Cas6 acts as a single turnover catalyst and remains bound to the matured crRNA after cleavage to become part of the interference complex (Brouns *et al.*, 2008; Jore *et al.*, 2011; Sashital *et al.*, 2011; Haurwitz *et al.*, 2012; Sternberg *et al.*, 2012; Jackson *et al.*, 2014; Niewoehner *et al.*, 2014; Zhao *et al.*, 2014; Hayes *et al.*, 2016). However, in case of type III systems, pre- CRISPR transcript is processed by Cas6 in solitude and not as part of the effector complex (Carte *et al.*, 2008; Carte *et al.*, 2010; Wang *et al.*, 2011; Shao and Li, 2013; Shao *et al.*, 2016; Makarova *et al.*, 2017a). Further, Cas6/III does not become a constituent of interference complex (Hale *et al.*, 2009; Zhang *et al.*, 2012; Rouillon *et al.*, 2013; Spilman *et al.*, 2013; Benda *et al.*, 2014; Staals *et al.*, 2014; Osawa *et al.*, 2015; Taylor *et al.*, 2015; Makarova *et al.*, 2017a), with the exception being Csm/III-A complex from *Streptococcus thermophilus* that shows weak transient interactions with Cas6 (Tamulaitis *et al.*, 2014). This possibly grants Cas6/III the flexibility required to associate with multiple subtypes that potentially differ at the interference stage. This allows sharing of crRNA processing pathways, with the mature crRNAs ultimately getting loaded onto specific effector complex. Type III system displays two kinds of effector complexes – Csm (Type III-A/D) or Cmr (Type III-B/C) (Hale *et al.*, 2009; Zhang *et al.*, 2012; Hatoum-Aslan *et al.*, 2013; Rouillon *et al.*, 2013; Tamulaitis *et al.*, 2014; Zhang *et al.*, 2016), which show distant relationship with Type I effector complex (Rouillon *et al.*, 2013; Osawa *et al.*, 2015; Taylor *et al.*, 2015).

A primary processed crRNA typically contains 8 nt repeat sequence at the 5'-end (often called 5' handle), a spacer sequence (guide) (Brouns *et al.*, 2008; Haurwitz *et al.*, 2010; Lintner *et al.*, 2011) and a variable 3'-end in some instances, which stems from further trimming at 3'-end by as yet unidentified nuclease(s) (Hale *et al.*, 2008; Hatoum-Aslan *et al.*, 2011; Richter *et al.*, 2012; Zhang *et al.*, 2012; Hatoum-Aslan *et al.*, 2013; Rouillon *et al.*, 2013; Shao and Li, 2013; Tamulaitis *et al.*, 2014). Intriguingly, the extent of processing appears to be effector complex-specific (Hale *et al.*, 2009; Lintner *et al.*, 2011; Zhang *et al.*, 2012; Hatoum-Aslan *et al.*, 2013; Rouillon *et al.*, 2013; Tamulaitis *et al.*, 2014). For example, *S. solfataricus* Cmr RNA component shows two different populations – some with a short 3' handle, while others with very little repeat-derived sequence at the 3'-end (Zhang *et al.*, 2012). Similarly, *S. epidermidis* Csm and *P. furiosus* Cmr RNA component show two mature crRNAs, differing by 6 nt (Lintner *et al.*, 2011; Hatoum-Aslan *et al.*, 2013). However, the *S. solfataricus* Csm appears to have a single form of bound crRNA, comprising of spacer bounded by 8 nt 5' handle and 3 nt 3' handle (Rouillon *et al.*, 2013). By contrast, the RNA component of *E. coli* Cascade and *Pseudomonas aeruginosa* Csy complex show no secondary processing and retains the unprocessed 3'-end in mature crRNA (Brouns *et al.*, 2008; Jore *et al.*, 2011; Wiedenheft *et al.*, 2011a; Haurwitz *et al.*, 2012). After binding to the effector complex, the crRNA may undergo secondary processing by a ruler-like mechanism operating from the 5' handle of the primary crRNA (Hatoum-Aslan *et al.*, 2011). The crRNA-binding subunits that form the backbone of the effector complexes may protect definite length of crRNA, and excess may get trimmed of from the 3'-end (Hatoum-Aslan *et al.*, 2011; Hatoum-Aslan *et al.*, 2013; Rouillon *et al.*, 2013). Thus, the differential processing of crRNA seems to depend on their ultimate destination in effector complex, with the complex defining the final length of the crRNA.

Though these crRNA processors share no detectable sequence similarity, they all adopt ferredoxin-like folds (Ebihara *et al.*, 2006; Gesner *et al.*, 2011; Sashital *et al.*, 2011; Wang *et al.*, 2011). However, despite their shared fold and structural topology, each subtype specific endoRNase exhibits remarkably different mechanism for target RNA recognition and cleavage (Fig. 3). This functional versatility is related to the specific repeat family of each subtype, which can form structured, unstructured or weakly structured repeats, that influences the mode of recognition of repeats and binding by the respective Cas proteins (Kunin *et al.*, 2007).

In *E. coli*, the crRNA is processed by Cas6/I-E as part of Cascade (Brouns *et al.*, 2008; Wiedenheft *et al.*, 2011a) (Fig. 3A). The structure of Cas6/I-E from *T. thermophilus* (Gesner *et al.*, 2011; Sashital *et al.*, 2011) reveals the presence of tandem ferredoxin-like ($\beta\alpha\beta\beta\alpha\beta$) fold. The repeat sequences of this system form a stable hexanucleotide stem with

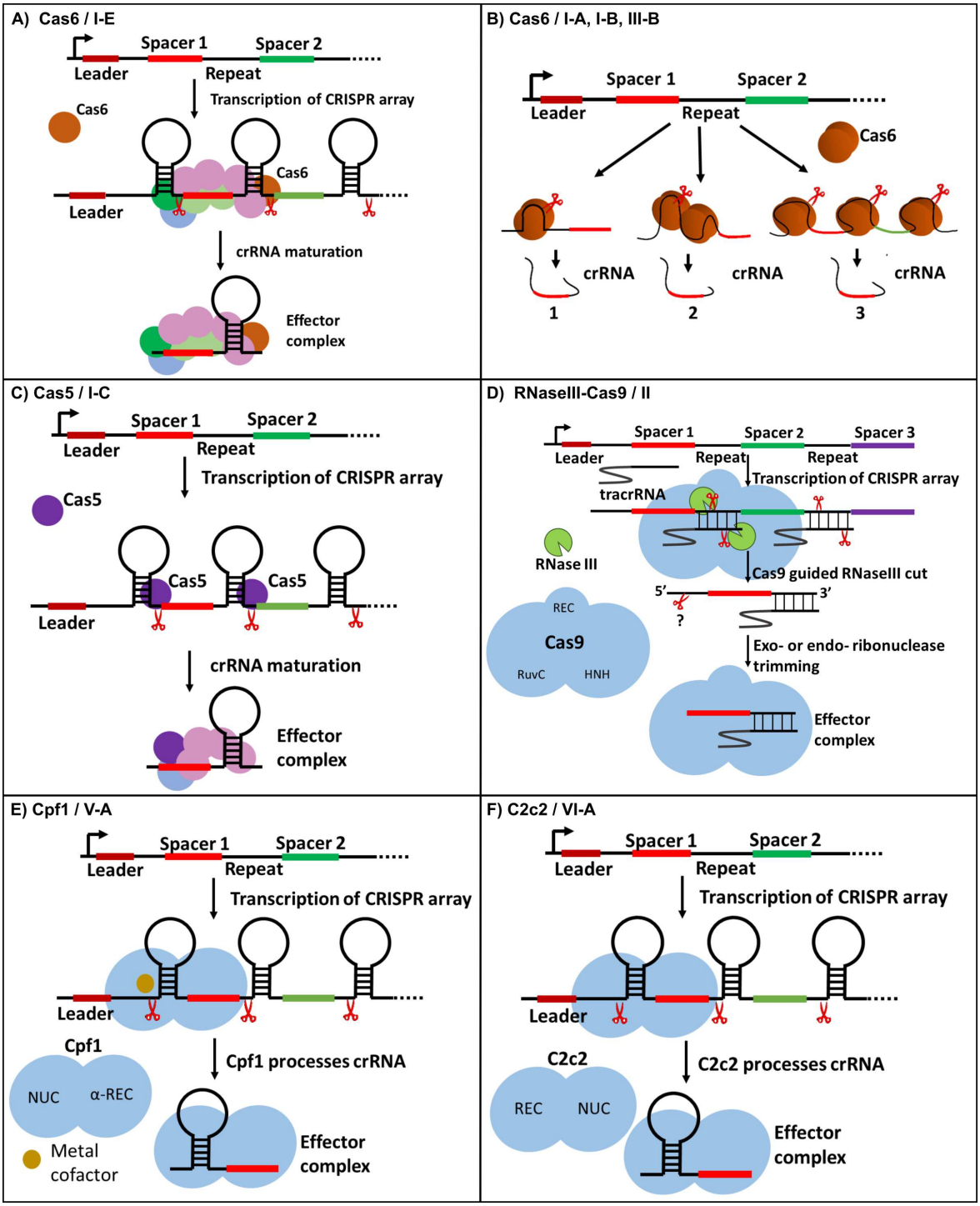

Fig. 3: Schematic representation of the CRISPR RNA (crRNA) maturation in CRISPR variants. (A) The crRNA maturation in type I-E system is shown, in which Cas6 (shown in brown) processes the structured repeat as a part of Cascade. (B) The distinct modes utilized by Cas6 to recognize the unstructured repeat in type I-A, I-B and III-B. (C) The case of type I-C where Cas5 (shown in purple) takes over the role of Cas6 to process the structured repeat. (D) The maturation in type II systems where RNase III processes the unstructured repeat tracr RNA duplex in presence of Cas9. (E) In type V-A system, Cpf1 processes the structured repeats upstream of the repeat stem, with an additional requirement of metal cofactor. (F) In type VI-A system, C2c2, processes the structured repeat upstream of the stem. In addition, it is an RNA-guided RNase, which is in contrast to most CRISPR systems, which show an RNA-guided DNA targeting. The distinct lobes and domain of the mutidomain effector proteins are shown in blue. The point of cleavage is indicated by a scissor. RNase III is shown in green colour

a tetranucleotide loop. The phosphate backbone of the 3' region of the RNA makes electrostatic contacts with positively charged residues in both of the ferredoxin-like domains, while the 5' region remains exposed to solvent. Here, the N-terminal ferredoxin domain (N- ferredoxin) of Cas6/I-E is involved in RNA hydrolysis and interacts accordingly with the lower portion of the RNA stem containing the scissile phosphate, as well as with the two unpaired nucleotides at the 3'-end of the RNA. On the other hand, the C-terminal ferredoxin domain (C-ferredoxin), which is involved in the recognition of RNA substrate, includes the major groove interacting β-hairpin (β6-β7), that contacts the middle and upper regions of the stem-loop through its several positively charged residues. In addition to the major groove binding hairpin of the protein, a second β-hairpin (â11 and β12) contacts the base of the RNA stem. Intriguingly, the free and RNA-bound structures of Cas6/I-E show notable differences. The regions including the β-hairpin (β6-β7) and a loop connecting strands β11 and β12 are disordered in the free state, indicating that interactions with the RNA stabilize their conformation (Fig. 4A). Further, the unwinding of the terminal base pair of the repeat stem-loop takes place, which is necessary for cleavage at a G-A bond at the 3'-end of the base of the repeat stem-loop (Gesner *et al.*, 2011; Sashital *et al.*, 2011). The catalytic residues Tyr23, His26, Arg27 and Arg158 are involved in RNA cleavage (Fig. 5A).

In *Pseudomonas aeruginosa*, Cas6/I-F adopts an N-terminal ferredoxin-like fold (N- ferredoxin) but its C-terminal region adopts an extended conformation (C-ext), although the basic secondary structure connectivity resembles a ferredoxin-like fold (Haurwitz *et al.*, 2010) (Fig. 4B). The repeat sequences of this system form a stable five base pair stem with a pentaloop, with the cleavage occurring at the 3'-end of the stem. An arginine rich helix in the C-terminal domain interacts extensively with the major groove of the RNA stem-loop and uses two amino acid side chains to read out the identity of the bottom two base pairs of the hairpin. The base of the stem of the repeat RNA is positioned in the positively charged cleft between the two domains (Fig. 4B). Cas6/I-F makes a sequence specific interaction with the first single stranded nucleotide upstream of the stem-loop, but does not interact with any of the nucleotides downstream of the stem-loop (Haurwitz *et al.*, 2010;

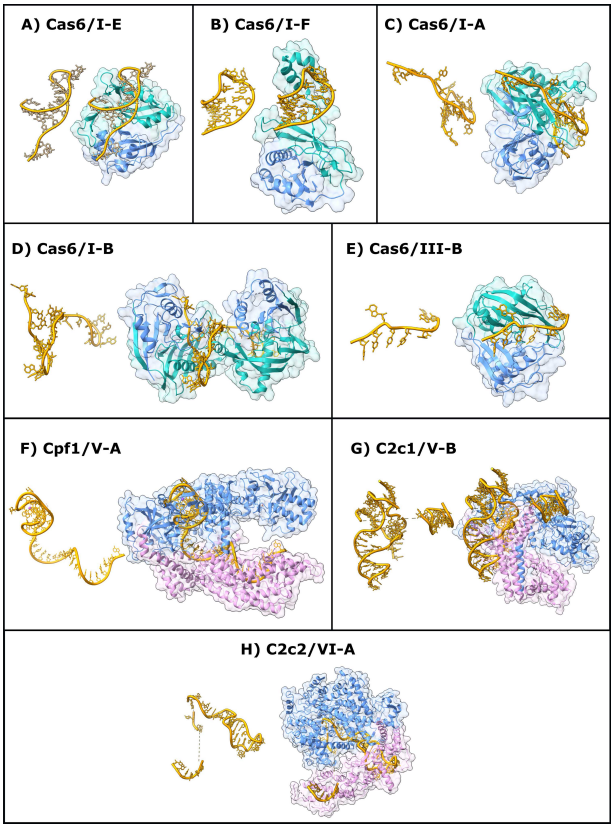

Fig. 4: The distinct modes of repeat RNA recognition by Cas endoRNases. The crystal structures of (A) Cas6e from *T. thermophilus* with 20 nt repeat RNA (PDB ID 2Y8W), (B) Cas6f from *P. aeruginosa* with 16 nt repeat RNA (PDB ID 2XLK), (C) Cas6 from *S. solfataricus* with 24 nt repeat RNA (PDB ID 4ILL), (D) Cas6 from *M. maripaludis* with 31 nt repeat RNA (PDB ID 4Z7K) (E) Cas6 from *P. furiosus* with 10 nt repeat RNA (PDB ID 3PKM), (F) Cpf1 from *Acidaminococcus* sp. *bv3l6* with with 43 nt guide RNA (PDB ID 5B43), (G) C2c1 from *A. acidoterrestris* with 112 nt sgRNA (PDB ID 5U34), and (H) C2c2 from *L. shahii* with 58 nt crRNA (PDB ID 5WTK) are shown. The bound RNA in some cases is only the product mimic or the minimum cleavable repeat instead of the complete repeat RNA. The N-terminal is shown in blue and C-terminal in cyan and RNA in orange (A-E). The recognition (REC) lobe and nuclease (NUC) lobe were displayed in purple and blue respectively (F-H). This figure was rendered using Chimera (Pettersen *et al.*, 2004)

Haurwitz *et al.*, 2012). Sequence specific hydrogen bonds tether the substrate in the active site so that the cleavage takes place immediately downstream of the hairpin, *i.e.*, 8 nt upstream of the spacer sequence. Residues His29 and Ser148 are involved in processing of crRNA and mutation in these residues abolishes cleavage reflecting their importance (Fig. 5B). Cas6/

I-F remains bound to the cleavage product via the base-specific interactions with the RNA, enabling the subsequent use of the mature crRNA by Cascade/I-F complex (Sternberg *et al.*, 2012; Rollins *et al.*, 2015).

A representative of the Cas6 family protein associated with subtypes I-A, I-B, I-D, III-A and III-B has been characterized in *P. furiosus*, *S. solfataricus* and *Methanococcus maripaludis* (Carte *et al.*, 2010; Wang *et al.*, 2011; Shao and Li, 2013; Shao *et al.*, 2016). Although the architecture of Cas6 from these subtypes also consists of two ferredoxin-like domains, the molecular mechanism for recognition and cleavage of the pre-crRNA seems to have evolved to accommodate the unstructured repeat RNA associated with these subtypes (Kunin *et al.*, 2007) (Fig. 3B). In Cas6/III-B the conserved positively charged central cleft between the two ferredoxin-like domains is responsible for interaction with single stranded repeat RNA, where conserved residues form contacts with nucleotides near the 5'-end of the CRISPR repeat, anchoring it in position for the cleavage at 3'-end of the repeat on the opposite surface of the protein (*vide* 3 in Fig. 3B). The RNA likely wraps around the protein to the opposite face where Cas6/III-B cleaves the RNA in an A-A dinucleotide motif (Carte *et al.*, 2008; Carte *et al.*, 2010; Wang *et al.*, 2011) (Fig. 4E). The catalytic and binding sites are distantly located, and are linked by substrate, which is interacting weakly or transiently with the signature Gly-rich loop (*vide* 3 in Fig. 3B). The cleavage of the pre- crRNA transcript occurs 8 nt upstream of each spacer generating the conserved 5' handle and variable 22 nt repeat derived sequence at the 3'-end. This processing is mediated by a catalytic triad comprising of Tyr31, His52 and Lys46 (Fig. 5E). The product remains bound to Cas6 until transferred to the respective effector complex (Cmr/III-B complex). The mature crRNAs may undergo further trimming at the 3'-end (Hale *et al.*, 2009; Carte *et al.*, 2010; Wang *et al.*, 2011).

Another interesting variation is observed in Cas6/I-A, which can induce the formation of 3 bp stem-loop upon binding to RNA for efficient processing by active site residues (Shao and Li, 2013) (Fig. 3B (*vide* 1), 4C and 5C). In yet another instance, two Cas6/I-B molecules bind to two distinct sites (a cleavage site and a distal site) on the long CRISPR repeat, *i.e.*, dual binding (*vide* 2 in Fig. 3B). One molecule of the Cas6/I-B that is bound to the distal site recognizes a 2 base pair stem and an AAYAA loop and supplies a tyrosine residue as a nucleobase mimic, that interacts with an adenine, which helps in stabilizing the stem and facilitates efficient cleavage of the pre-crRNA by another molecule of Cas6/I-B bound to the cleavage site (Shao *et al.*, 2016) (Fig. 3B (*vide* 2), 4D and 5D). Despite the diverse modes of RNA recognition, Cas6 cleavage produces the crRNA with an 8 nt repeat derived sequence at the 5'-end (Brouns *et al.*, 2008; Carte *et al.*, 2008; Carte *et al.*, 2010; Hatoum-Aslan *et al.*, 2011; Jore *et al.*, 2011; Wang *et al.*, 2011; Hatoum-Aslan *et al.*, 2013).

Notwithstanding the mode of RNA recognition that seems to vary between type I and III, the endoRNases seem to follow a metal independent acid-base hydrolysis mechanism producing a cyclic 2'-3' phosphate intermediate and the final product having 5' hydroxyl group (5'-OH) and 3' phosphate (3'-P) ends (Carte *et al.*, 2008; Haurwitz *et al.*, 2010; Gesner *et al.*, 2011; Jore *et al.*, 2011; Sashital *et al.*, 2011; Wiedenheft *et al.*, 2011b). The deprotonated hydroxyl at the 2' position of the ribose functions as a nucleophile. The catalytic sites of all characterized Cas6-like enzymes are composed of an invariant histidine residue and a tyrosine residue in the active site along with a variable lysine or serine (Carte *et al.*, 2008; Haurwitz *et al.*, 2010; Gesner *et al.*, 2011; Sashital *et al.*, 2011). However, the relative positions of these residues are poorly conserved, which might explain the observed functional variations in Cas6 activity. Moreover, to orient the substrate correctly, these RAMP proteins employ a glycine rich loop, which is typically located towards the C- terminus.

While Cas6 is pervasive among most type I systems, the only exception to this is observed in type I-C, wherein Cas6 is absent and its role is adopted by Cas5, which is shown to be involved in pre-crRNA processing (Garside *et al.*, 2012; Nam *et al.*, 2012; Koo *et al.*, 2013; Punetha *et al.*, 2014) (Fig. 3C). Intriguingly, Cas5 in other type I systems appears to be inert and the gain of function with respect to RNA processing is uniquely seen in type I-C. Cas5/I-C processes the pre-CRISPR transcript individually as well as part of the Cascade/I-C complex in a metal-independent fashion (Punetha *et al.*, 2014). Unlike type I-E, Cascade/I-C comprises of only three Cas

proteins *viz.*, Cas5, Csd1 and Csd2. Remarkably, Csd1/I-C also exhibits a metal-independent endoRNase activity similar to Cas5/I-C. This parallel processing of the crRNA seems to be an evolutionary adaptation for eliciting a rapid immune response and confers a selective advantage against genome predators. Further, the stoichiometry of the constituents of the RNP complex in type I-C may differ from type I-E as Csd1/I-C seems to be a fusion of its functional homolog Cse1/I-E and Cse2/I-E (Punetha *et al.*, 2014). Apart from the RNase activity, Cas5/I-C and Csd1/I-C also exhibit promiscuous DNase activity that is selectively promoted in the presence of divalent metals (Punetha *et al.*, 2014). Remarkably, the active site residues in Cas5/I-C show considerable functional overlap in both RNA and DNA hydrolysis (Punetha *et al.*, 2014). The mechanism of nucleic acid hydrolysis in type I-C also seems to follow a general acid base catalytic mechanism. Tyr46, Lys116 and His117 of Cas5/I-C seem to be attractive candidates to assume the analogous role as proposed for the equivalent residues in Cas6/III-B (Tyr31 Lys52 His46) (Carte *et al.*, 2008) and the archetypal enzyme RNase A catalytic triad (His12 Lys41 His119), wherein the Tyr is replaced by His12 (Raines, 1998).

A salient feature of the CRISPR-Cas system is that an organism can harbour more than one CRISPR variants. This can be attributed to horizontal gene transfer either via plasmids that harbour CRISPR-*cas* loci or by other gene transfer mechanisms such as transposon activity, which result in the movement of CRISPR-*cas* loci across widely diverged lineages (Godde and Bickerton, 2006; Horvath *et al.*, 2009; Portillo and Gonzalez, 2009). For example, *T. thermophilus* harbours three types of surveillance complexes (Cascade/I-E, Csm/III-A and Cmr/III-B) and three Cas6 endoRNases (TTHB192, TTHA0078 and TTHB231), which are associated with different repeat clusters and have distinct mode of substrate recognition (Sashital *et al.*, 2011; Niewoehner *et al.*, 2014; Staals *et al.*, 2014). Another mosaic CRISPR-Cas system is found in *S. solfataricus*, which harbours three types of surveillance complexes (Cascade/type I-A, Csm/type III-A and Cmr/type IIIB), five Cas6 paralogues and two different CRISPR repeat families (Sokolowski *et al.*, 2014).

Type II CRISPR System

Unlike type I and type III systems, the crRNA maturation in type II involves a distinct Cas protein referred as Cas9, which operates together with RNase III – a host factor – to facilitate CRISPR RNA processing (Deltcheva *et al.*, 2011; Jinek *et al.*, 2012) (Fig. 3D). Additionally, a small RNA referred as tracrRNA (trans-activating CRISPR RNA), which is located upstream of the CRISPR array and complementary to the repeat region of pre- CRISPR RNA plays an important role in crRNA maturation. In addition, it is essential for crRNA mediated DNA recognition and Cas9 mediated targeting *in vitro*, even for crRNAs that bypass processing (Jinek *et al.*, 2012; Zhang *et al.*, 2013). Cas9 promotes the base pairing between the tracrRNA and the repeat sequence of the pre-crRNA, which then becomes a substrate for double strand specific RNase III (Fig. 3D). The cleavage ensues at specific positions within the duplex region producing crRNA units that consist of a complete spacer sequence flanked by the partial repeats. Subsequently, the extended processing of the partial repeat sequence is initiated at the 5'-end of the spacer by an unidentified nuclease. The processing continues until few nucleotides inside the spacer sequence and this concludes the dual step mechanism of crRNA maturation (Deltcheva *et al.*, 2011). The mature crRNA comprises of a 20 nt spacer derived sequence at the 5'-end and a 19-22 nt repeat derived sequence at the 3'-end. This feature is strikingly different from the mature crRNAs found in type I and III, which harbour a repeat derived sequence at the 5'-end.

Even within type II, a variant mechanism of crRNA maturation is noticed in the type II-C represented by *Neisseria meningitides*, which shows an RNase III-independent mechanism (Zhang *et al.*, 2013). In this, the terminal 9 nt of each repeat harbours a promoter sequence (an extended -10 box promoter element), resulting in transcription of 9-10 nt downstream sequence, *i.e.*, within the spacer. This generates 5'-end of crRNA directly by transcription bypassing the processing event, yielding the mature crRNA (Zhang *et al.*, 2013). Although trimming at the 3'-end may occur, which involves RNase III and tracrRNA but this is dispensable for interference (Zhang *et al.*, 2013). The dispensability of pre-crRNA processing represents a remarkable mechanistic variation in crRNA generation among type II variants.

Yet another interesting feature of type II system is that it seems to be absent in archaea and restricted to bacteria, and this seems plausible with the absence of gene encoding RNase III in most archaeal genomes (Garrett *et al.*, 2015).

The multidomain Cas9 (the effector of type II system) has a bi-lobed architecture, consisting of recognition (REC) and nuclease (NUC) lobes (Fig. 3D) (Sapranauskas *et al.*, 2011; Gasiunas *et al.*, 2012; Jinek *et al.*, 2012). The NUC lobe contains two nuclease domains A RuvC-like domain and a HNH domain and therefore might mediate the second cleavage in crRNA that occurs at a fixed distance within the spacers during crRNA maturation (Deltcheva *et al.*, 2011). The association of Cas9 with the dual RNA structure composed of mature crRNA tracrRNA duplex triggers a conformational change, resulting in activation of ternary silencing complex that is suitable for target scanning, recognition, and interference (Deltcheva *et al.*, 2011; Gasiunas *et al.*, 2012; Jinek *et al.*, 2012). The reorientation of Cas9 structural lobes generates a positively charged central cleft between the two lobes to bind both the guide RNA and target DNA duplex (Jinek *et al.*, 2014; Nishimasu *et al.*, 2014). The RuvC and HNH domains of the NUC lobe of Cas9 cleave the displaced non-complementary and the complementary target strand, respectively, in the crRNA target DNA complex (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012; Chen *et al.*, 2014). Interestingly, an allosteric communication ensures concerted action of both RuvC-HNH nuclease domains during targeting (Sternberg *et al.*, 2015; Jiang *et al.*, 2016). The binding and targeting by Cas9 depends on the recognition of PAM at the 3'-end of protospacer in addition to crRNA target DNA complementarity, that authorizes subsequent DNA strand displacement and R-loop formation (a three-stranded structure, composed of crRNA target DNA hybrid and the displaced DNA strand) (Anders *et al.*, 2014; Sternberg *et al.*, 2014). Upon target DNA binding, Cas9 HNH domain undergoes several conformational transitions before adopting its active state displaying a remarkable conformational activation (Sternberg *et al.*, 2015; Dagdas *et al.*, 2017; Huai *et al.*, 2017; Shibata *et al.*, 2017).

The crRNA and tracrRNA can be combined into a chimeric single guide RNA (sgRNA) for simplifying the system for sequence-specific DNA targeting (Jinek *et al.*, 2012; Butt *et al.*, 2017). Thus, any DNA sequence of interest can be targeted by changing the guide RNA sequence (spacer) within the crRNA. Further, a nuclease deficient Cas9 (dCas9, a variant with inactivating mutations in RuvC and HNH domains) can be fused to an effector domain to drive required function. The vivid applications of CRISPR-Cas9 technology include – multiplex gene editing (creating indels at precise locations) (Cho *et al.*, 2013; Cong *et al.*, 2013; Jinek *et al.*, 2013; Mali *et al.*, 2013; Butt *et al.*, 2017; Kumagai *et al.*, 2017), epigenome editing (Hilton *et al.*, 2015; Kearns *et al.*, 2015; Liu *et al.*, 2016; Chen *et al.*, 2017; O'Geen *et al.*, 2017), transcriptional activation and repression (Bikard *et al.*, 2013; Gilbert *et al.*, 2013; Konermann *et al.*, 2013; Qi *et al.*, 2013; Kim *et al.*, 2017; Machens *et al.*, 2017), imaging system for fluorescent labelled macromolecules in live cell (Chen *et al.*, 2013; Fu *et al.*, 2016; Nelles *et al.*, 2016; Takei *et al.*, 2017; Zhou *et al.*, 2017), building gene circuits (Kiani *et al.*, 2014; Liu *et al.*, 2014; Nissim *et al.*, 2014; Gander *et al.*, 2017), genome-wide screening (Shalem *et al.*, 2014; Wang *et al.*, 2014; Schmierer *et al.*, 2017), developing antimicrobials (Bikard *et al.*, 2014; Park *et al.*, 2017), and antivirals (Ebina *et al.*, 2013; Hu *et al.*, 2014; Ramanan *et al.*, 2015; Dampier *et al.*, 2017; Huang and Nair, 2017; Scott *et al.*, 2017) to name a few. Thus, this system exhibits immense promise for precise gene editing to cure genetic diseases.

Type IV and Type V CRISPR Systems

In the recently identified type IV system, the mechanism underlying the crRNA maturation is yet to be determined (Makarova *et al.*, 2015; Makarova *et al.*, 2017a). In type V-A system a single multifunctional protein, Cpf1 (Cas12a), which is similar to that of Cas9/II, is involved in both maturation and interference stages (Zetsche *et al.*, 2015). However, in contrast to Cas9/II, Cpf1/V-A directly processes pre-crRNA without the requirement of auxiliary factors (Fonfara *et al.*, 2016) (Fig. 3E). Cpf1/V-A cleaves pre- crRNA upstream of a stem-loop structure formed within the CRISPR repeats and generates intermediate crRNAs, which are further processed to form mature crRNAs. Since Cpf1/V-A exhibits both RNase and DNase activity, the dual nuclease activity requires sequence and structure specific binding to the stem-loop of crRNA repeats, for which it uses distinct active domains and cleaves nucleic acids in the presence of magnesium or calcium (Yamano *et*

al., 2016). Cpf1/V-A has a bi-lobed architecture consisting of aá-REC lobe and a NUC lobe (Dong *et al.*, 2016; Yamano *et al.*, 2016) (Fig. 4F). The α-REC lobe contains two REC domains at the N-terminal region, and the NUC lobe consists of a RuvC domain, wedge (WED), Nuc domain, PAM-interacting (PI) domain and a Bridge helix (BH) domain. The catalytic residues (His843, Lys852, Lys869 and Phe873) involved in crRNA processing reside in the WED domain of NUC lobe (Fonfara *et al.*, 2016) (Fig. 5F). In contrast to type I and III (Cas6 and Cas5/I-C) endoRNases, which are metal independent processors of crRNA, Cpf1/V-A requires a metal cofactor for the RNA processing (Fig. 3E). Further, in contrast to Cas9/II, Cpf1/V-A contains only RuvC domain and lacks the HNH domain, suggesting a distinct molecular mechanism being operational at interference level. The crRNA-target DNA duplex binds at the central positively charged cleft formed between the REC and NUC lobes, as revealed by the structure of Cpf1 from *Acidaminococcus* sp. (Gao *et al.*, 2016).

In a newly identified type V-B system, a guide tracrRNA mediated C2c1 (Cas12b) site-specifically cleaves both the strands of target DNA (Yang *et al.*, 2016; Lewis and Ke, 2017). C2c1/V-B shows the cleavage properties similar to Cpf1/V-A and distinct from Cas9/II. But the requirement of both crRNA and tracrRNA, is in sharp contrast to Cpf1/V-A, which only requires crRNA for targeting (Yang *et al.*, 2016). The structure of C2c1 from *Alicyclobacillus acidoterrestris* in complex with a chimeric single-molecule guide RNA (sgRNA comprises of a crRNA covalently linked to a tracrRNA) exhibits a bi-lobed architecture consisting of a REC and NUC lobe (Liu *et al.*, 2017a) (Fig. 4G). In C2c1, the sgRNA scaffold forms a tetra-helical structure with a distinct mechanism of assembly that differs from Cas9/II or Cpf1/V-A endonuclease (Liu *et al.*, 2017a) (Fig. 4G). The crRNA binds in the central cleft of C2c1, while the tracrRNA binds in an external surface groove. Although C2c1/V-B lacks a PI domain, the PAM duplex has a similar binding position found in Cpf1/V-A. Hitherto, the involvement of C2c1 in crRNA biogenesis either as a solo or in association with auxiliary factor remains to be determined.

Type VI CRISPR System

The type VI-A effector, C2c2 (Cas13a1), is responsible for both pre-crRNA processing and target RNA cleavage (East-Seletsky *et al.*, 2016). The effectors of type VI are unrelated to the type II and type V effectors and contain two ribonuclease

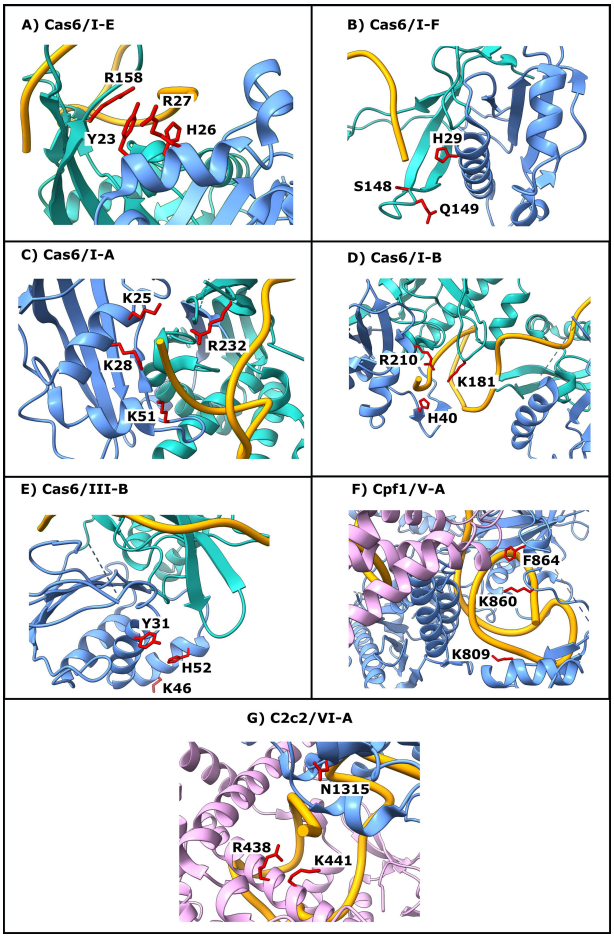

Fig. 5: The active site residues of Cas endoRNases in CRISPR variants. The catalytic residues of the distinct Cas endoRNase are highlighted. The crystal structures of (A) Cas6e from *T. thermophilus* with 20 nt repeat RNA (PDB ID 2Y8W), (B) Cas6f from *P. aeruginosa* with 16 nt repeat RNA (PDB ID 2XLK), (C) Cas6 from *S. solfataricus* with 24 nt repeat RNA (PDB ID 4ILL), (D) Cas6 from *M. maripaludis* with 31 nt repeat RNA (PDB ID 4Z7K) (E) Cas6 from *P. furiosus* with 10 nt repeat RNA (PDB ID 3PKM), (F) Cpf1 from *Acidaminococcus* sp. *bv3l6* with with 43 nt guide RNA (PDB ID 5B43), (G) C2c1 from *A. acidoterrestris* with 112 nt sgRNA (PDB ID 5U34), and (H) C2c2 from *L. shahii* with 58 nt crRNA (PDB ID 5WTK) are shown. The bound RNA in some cases is only the product mimic or the minimum cleavable repeat instead of the complete repeat RNA and is shown in orange. The N-terminal is shown in blue and C-terminal in cyan (A-E). The recognition (REC) lobe and nuclease (NUC) lobe were displayed in purple and blue, respectively (F-H). This figure was rendered using Chimera (Pettersen *et al.*, 2004)

domains of the HEPN superfamily (Liu *et al.*, 2017b; Makarova *et al.*, 2017b; Shmakov *et al.*, 2017). In *Leptotrichia buccalis* (LbuC2c2), the two unique HEPN domains of LbuC2c2 (Cas13a) were shown to be important for interference but not for the crRNA processing. The residue (R1079) involved in crRNA maturation was found to be located in the C-terminal domain of LbuC2c2 (East-seletsky *et al.*, 2016). The recently available structure of *Leptotrichia shahii* (LshC2c2) exhibits a bi-lobed structure, which is reminiscent of all other Class 2 effectors. It contains a REC lobe (1-498) with an N-terminal domain (NTD) and a Helical-1 domain and a NUC lobe (499-1389) with two HEPN domains and a Helical-2 domain (Liu *et al*., 2017b) (Fig. 3F and 4H). The two RNase catalytic sites responsible for cleaving pre-crRNA and target RNA are independently located on REC lobe (Helical-1 domain) and NUC lobe (HEPN domains), respectively. The surface of the Helical-1 domain facing the NTD domain is positively charged which forms a cleft for pre-crRNA-binding. The Helical-1 domain of LshC2c2 reveals the involvement of Lys471 in substrate orientation and Arg438 and Lys441 in pre-crRNA processing (Fig. 4H and 5G). Mutation of Asn1315 in HEPN2 domain reduced the LshC2c2 activity suggesting that it is important but not directly involved in catalysis. Among these critical residues for pre-crRNA processing, Arg438 is conserved and Lys441 is variable (Fig. 5G). The lack of conservation of Lys441 suggests that the cleavage position varies between pre-crRNAs associated with C2c2 of different species.

Further, the crRNA maturation doesn't seem to require tracrRNA, which is in contrast to Cas9/II and C2c1/V-B (Deltcheva *et al.*, 2011; Chylinski *et al.*, 2013; Chylinski *et al.*, 2014; Shmakov *et al.*, 2015; Wu *et al.*, 2017). C2c2 is an RNA-guided RNase, which upon target RNA recognition, shows promiscuity towards other host RNAs and apparently causes cell toxicity or death (Abudayyeh *et al.*, 2016). The effectors of type VI-A could have evolved from HEPN domain-containing RNases that serve as accessory proteins in certain Class 1 CRISPR-Cas system (Liu *et al.*, 2017b; Makarova *et al.*, 2017b; Shmakov *et al.*, 2017). The crRNA maturation in type VI-B harbouring Cas13b and type VI-C harbouring Cas13a2 needs to be further investigated (Makarova *et al.*, 2017b; Smargon *et al.*, 2017).

Conclusion

The discovery of the disparate Cas nucleases has significantly enriched the knowledge on nucleic acid recognition and catalysis that lies at the heart of CRISPR immunity. The different CRISPR-Cas systems showcase the impressive range of molecular strategies utilized by Cas endoribonucleases to impart RNA specificity and cleavage. Thus, the high diversity in the modes of crRNA recognition highlights the exceptional plasticity of CRISPR-Cas systems and reflects the co-evolution of the endoRNases with their associated repeat sequences. The fascinating variations displayed by the ever-expanding CRISPR-Cas systems allude to yet more surprises on the mechanism of CRISPR RNA maturation in the years to come.

Acknowledgement

We gratefully acknowledge the past and present grants from Department of Biotechnology (DBT) [BT/PR14402/BRB/10/830/2010, BT/08/IYBA/2014-15, BT/406/NE/UEXCEL/2013, BT/PR5511/MED/29/631/2012, BT/341/NE/TBP/2012, BT/PR15925/NER/95/141/2015], Board of Research in Nuclear Sciences (BRNS) [2011/20/37B/15/BRNS/2266] and Science and Engineering Research Board (SERB) [YSS/2014/000286], Govt. of India. We sincerely thank all members of the MAB lab for their critical comments and suggestions. BA is a recipient of INSA Young Scientist Medal and DBT-Innovative Young Biotechnologist Award.

References

Abudayyeh O O, Gootenberg J S, Konermann S, Joung J, Slaymaker I M, *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector *Science* **353** aaf5573

Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, *et al.* (2010) Transcription profile of Thermus thermophilus CRISPR systems after phage infection *J Mol Biol* **395** 270- 281

Anders C, Niewoehner O, Duerst A and Jinek M (2014) Structural

basis of PAM-dependent target DNA recognition by the Cas9 endonuclease *Nature* **513** 569-573

Arslan Z, Hermanns V, Wurm R, Wagner R and Pul U (2014) Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system *Nucleic Acids Res* **42** 7884-7893

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes *Science* **315** 1709-1712

Benda C, Ebert J, Scheltema R A, Schiller H B, Baumgartner M, *et al.* (2014) Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in

Cmr4 *Mol Cell* **56** 43-54

Bernick D L, Cox C L, Dennis P P and Lowe T M (2012) Comparative genomic and transcriptional analyses of CRISPR systems across the genus Pyrobaculum *Front Microbiol* **3** 251

Bikard D, Euler C W, Jiang W, Nussenzweig P M, Goldberg G W, *et al.* (2014) Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials *Nat Biotechnol* **32** 1146-1150

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, *et al.* (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system *Nucleic Acids Res* **41** 7429-7437

Bolotin A, Quinquis B, Sorokin A and Ehrlich S D (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin *Microbiology* **151** 2551-2561

Brouns S J, Jore M M, Lundgren M, Westra E R, Slijkhuis R J, *et al.* (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes *Science* **321** 960-964

Butt H, Eid A, Ali Z, Atia M A M, Mokhtar M M, *et al.* (2017) Efficient CRISPR/Cas9- Mediated Genome Editing Using a Chimeric Single-Guide RNA Molecule *Front Plant Sci* **8** 1441

Carte J, Pfister N T, Compton M M, Terns R M and Terns M P (2010) Binding and cleavage of CRISPR RNA by Cas6 *RNA* **16** 2181-2188

Carte J, Wang R, Li H, Terns R M and Terns M P (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes *Genes Dev* **22** 3489-3496

Chen B, Gilbert L A, Cimini B A, Schnitzbauer J, Zhang W, *et al.* (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system *Cell* **155** 1479-1491

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

Chen T, Gao D, Zhang R, Zeng G, Yan H, *et al.* (2017) Chemically Controlled Epigenome Editing through an Inducible dCas9 System *J Am Chem Soc* **139** 11337-11340

Cho S W, Kim S, Kim J M and Kim J S (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease *Nat Biotechnol* **31** 230-232

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

Chylinski K, Makarova K S, Charpentier E and Koonin E V (2014) Classification and evolution of type II CRISPR-Cas systems *Nucleic Acids Res* **42** 6091-6105

Cong L, Ran F A, Cox D, Lin S, Barretto R, *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems *Science* **339** 819-823

Dagdas Y S, Chen J S, Sternberg S H, Doudna J A and Yildiz A (2017) A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9 *Sci Adv* **3** eaao0027

Dampier W, Sullivan N T, Chung C H, Mell J C, Nonnemacher M R, *et al.* (2017) Designing broad-spectrum anti-HIV-1 gRNAs to target patient-derived variants *Sci Rep* **7** 14413

Datsenko K A, Pougach K, Tikhonov A, Wanner B L, Severinov K, *et al.* (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system *Nat Commun* **3** 945

Deltcheva E, Chylinski K, Sharma C M, Gonzales K, Chao Y, *et al.* (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III *Nature* **471** 602-607

Deveau H, Barrangou R, Garneau J E, Labonte J, Fremaux C, *et al.* (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus *J Bacteriol* **190** 1390-1400

Diez-Villasenor C, Guzman N M, Almendros C, Garcia-Martinez J and Mojica F J (2013) CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli *RNA Biol* **10** 792-802

Dong D, Ren K, Qiu X, Zheng J, Guo M, *et al.* (2016) The crystal structure of Cpf1 in complex with CRISPR RNA *Nature* **532** 522-526

Doxzen K W and Doudna J A (2017) DNA recognition by an RNA-guided bacterial Argonaute *PLoS ONE* **12** e0177097

East-Seletsky A, O'Connell M R, Knight S C, Burstein D, Cate J H, *et al.* (2016) Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection *Nature* **538** 270-273

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

Ebina H, Misawa N, Kanemura Y and Koyanagi Y (2013) Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus *Sci Rep* **3** 2510

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

Erdmann S, Le Moine Bauer S, Garrett R A (2014) Inter-viral conflicts that exploit host CRISPR immune systems of Sulfolobus *Mol Microbiol* **91** 900-917

Fagerlund R D, Wilkinson M E, Klykov O, Barendregt A, Pearce F G, *et al.* (2017) Spacer capture and integration by a type I-F Cas1–Cas2-3 CRISPR adaptation complex *Proc Natl Acad Sci U S A* **114** E5122-E5128

Fineran P C, Gerritzen M J H, Suárez-Diez M, Künne T, Boekhorst J, *et al.* (2014) Degenerate target sites mediate rapid primed CRISPR adaptation *Proc Natl Acad Sci U S A* **111** E1629-E1638

Fire A, Xu S, Montgomery M K, Kostas S A, Driver S E, *et al.* (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans *Nature* **391** 806-811 Fonfara I, Richter H, Bratovic M, Le Rhun A, Charpentier E (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA *Nature* **532** 517-521

Fu Y, Rocha P P, Luo V M, Raviram R, Deng Y, *et al.* (2016) CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci *Nat Commun* **7** 11707

Gander M W, Vrana J D, Voje W E, Carothers J M and Klavins E (2017) Digital logic circuits in yeast with CRISPR-dCas9 NOR gates *Nat Commun* **8** 15459

Gao P, Yang H, Rajashankar K R, Huang Z and Patel D J (2016) Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition *Cell Res* **26** 901-913

Garneau J E, Dupuis M E, Villion M, Romero D A, Barrangou R, *et al.* (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA *Nature* **468** 67-71

Garrett R A, Shah S A, Erdmann S, Liu G, Mousaei M, *et al.* (2015) CRISPR-Cas Adaptive Immune Systems of the Sulfolobales Unravelling Their Complexity and Diversity *Life (Basel)* **5** 783-817

Garside E L, Schellenberg M J, Gesner E M, Bonanno J B, Sauder J M, *et al.* (2012) Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases *RNA* **18** 2020-2028

Gasiunas G, Barrangou R, Horvath P and Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria *Proc Natl Acad Sci U S A* **109** E2579-2586

Gesner E M, Schellenberg M J, Garside E L, George M M and Macmillan A M (2011) Recognition and maturation of effector RNAs in a CRISPR interference pathway *Nat Struct Mol Biol* **18** 688-692

Gilbert L A, Larson M H, Morsut L, Liu Z, Brar G A, *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes *Cell* **154** 442-451

Godde J S and Bickerton A (2006) The repetitive DNA elements called CRISPRs and their associated genes evidence of horizontal transfer among prokaryotes *J Mol Evol* **62** 718-729 Grissa I, Vergnaud G and Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats *BMC Bioinformatics* **8** 172

Haft D H, Selengut J, Mongodin E F and 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** e60

Hale C, Kleppe K, Terns R M and Terns M P (2008) Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus *RNA* **14** 2572-2579

Hale C R, Majumdar S, Elmore J, Pfister N, Compton M, *et al.* (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, *et al.* (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex *Cell* **139** 945-956

Hatoum-Aslan A, Maniv I and Marraffini L A (2011) Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site *Proc Natl Acad Sci U S A* **108** 21218-21222

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

Haurwitz R E, Jinek M, Wiedenheft B, Zhou K and Doudna J A (2010) Sequence- and structure-specific RNA processing

by a CRISPR endonuclease *Science* **329** 1355-1358

Haurwitz R E, Sternberg S H and Doudna J A (2012) Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA *EMBO J* **31** 2824-2832

Hayes R P, Xiao Y, Ding F, van Erp P B, Rajashankar K, *et al.* (2016) Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli Nature* **530** 499-503

Hille F, Charpentier E (2016) CRISPR-Cas biology, mechanisms and relevance *Philos Trans R Soc Lond B Biol Sci* **371** 20150496

Hilton I B, D'Ippolito A M, Vockley C M, Thakore P I, Crawford G E, *et al.* (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers *Nat Biotechnol* **33** 510-517

Hochstrasser M L and Doudna J A (2015) Cutting it close CRISPR-associated endoribonuclease structure and function *Trends Biochem Sci* **40** 58-66

Hooton S P and Connerton I F (2014) Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein *Front Microbiol* **5** 744

Horvath P, Coute-Monvoisin A C, Romero D A, Boyaval P, Fremaux C, *et al.* (2009) Comparative analysis of CRISPR loci in lactic acid bacteria genomes *Int J Food Microbiol* **131** 62-70

Horvath P, Romero D A, Coute-Monvoisin A C, Richards M, Deveau H, *et al.* (2008) Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus *J Bacteriol* **190** 1401-1412

Hu W, Kaminski R, Yang F, Zhang Y, Cosentino L, *et al.* (2014) RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection *Proc Natl Acad Sci U S A* **111** 11461-11466

Huai C, Li G, Yao R, Zhang Y, Cao M, *et al.* (2017) Structural insights into DNA cleavage activation of CRISPR-Cas9 system *Nat Commun* **8** 1375

Huang Z and Nair M (2017) A CRISPR/Cas9 guidance RNA screen platform for HIV provirus disruption and HIV/AIDS gene therapy in astrocytes *Sci Rep* **7** 5955

Hyman P and Abedon S T (2010) Bacteriophage host range and bacterial resistance *Adv Appl Microbiol* **70** 217-248

Hynes A P, Villion M and Moineau S (2014) Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages *Nat Commun* **5** 4399

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

Ivanèiæ-Baæe I, Cass S D, Wearne S J and Bolt E L (2015) Different genome stability proteins underpin primed and naïve adaptation in E. coli CRISPR-Cas immunity *Nucleic Acids Res* **43** 10821-10830

Jackson R N, Golden S M, van Erp P B, Carter J, Westra E R, *et al.* (2014) Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli *Science* **345** 1473-1479

Jackson S A, McKenzie R E, Fagerlund R D, Kieper S N, Fineran P C, *et al.* (2017) CRISPR-Cas Adapting to change *Science* **356** eaal5056

Jansen R, Embden J D, Gaastra W and Schouls L M (2002a) Identification of genes that are associated with DNA repeats in prokaryotes *Mol Microbiol* **43** 1565-1575

Jansen R, van Embden J D, Gaastra W and Schouls L M (2002b) Identification of a novel family of sequence repeats among prokaryotes *Omics* **6** 23-33

Jiang F, Taylor D W, Chen J S, Kornfeld J E, Zhou K, *et al.* (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage *Science* **351** 867-871

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna J A, *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity *Science* **337** 816-821

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

Jore M M, Lundgren M, van Duijn E, Bultema J B, Westra E R, *et al.* (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade *Nat Struct Mol Biol* **18** 529-536 Kearns N A, Pham H, Tabak B, Genga R M, Silverstein N J, *et al.* (2015) Functional annotation of native enhancers with a Cas9-histone demethylase fusion *Nat Methods* **12** 401-403

Kiani S, Beal J, Ebrahimkhani M R, Huh J, Hall R N, *et al.* (2014) CRISPR transcriptional repression devices and layered circuits in mammalian cells *Nat Methods* **11** 723-726

Kim S K, Seong W, Han G H, Lee D H and Lee S G (2017) CRISPR interference-guided multiplex repression of endogenous competing pathway genes for redirecting metabolic flux in Escherichia coli *Microb Cell Fact* **16** 188

Konermann S, Brigham M D, Trevino A, Hsu P D, Heidenreich M, *et al.* (2013) Optical control of mammalian endogenous transcription and epigenetic states *Nature* **500** 472-476

Koo Y, Ka D, Kim E J, Suh N and Bae E (2013) Conservation and variability in the structure and function of the Cas5d

endoribonuclease in the CRISPR-mediated microbial immune system *J Mol Biol* **425** 3799-3810

Kumagai H, Nakanishi T, Matsuura T, Kato Y and Watanabe H (2017) CRISPR/Cas-mediated knock-in via non-homologous end-joining in the crustacean Daphnia magna *PLoS ONE* **12** e0186112

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

Künne T, Kieper S N, Bannenberg J W, Vogel A I M, Miellet W R, *et al*. (2016) Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation *Mol Cell* **63** 852-864

Labrie S J, Samson J E and Moineau S (2010) Bacteriophage resistance mechanisms *Nat Rev Microbiol* **8** 317-327

Lewis K M and Ke A (2017) Building the Class 2 CRISPR-Cas Arsenal *Mol Cell* **65** 377-379

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

Lillestol R K, Redder P, Garrett R A and Brugger K (2006) A putative viral defence mechanism in archaeal cells *Archaea* **2** 59-72

Lillestol R K, Shah S A, Brugger K, Redder P, Phan H, *et al.* (2009) CRISPR families of the crenarchaeal genus Sulfolobus bidirectional transcription and dynamic properties *Mol Microbiol* **72** 259-272

Lintner N G, Kerou M, Brumfield S K, Graham S, Liu H, *et al.* (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE) *J Biol Chem* **286** 21643-21656

Liu L, Chen P, Wang M, Li X, Wang J, *et al.* (2017a) C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism *Mol Cell* **65** 310-322

Liu L, Li X, Wang J, Wang M, Chen P, *et al.* (2017b) Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities *Cell* **168** 121-134

Liu X S, Wu H, Ji X, Stelzer Y, Wu X, *et al.* (2016) Editing DNA Methylation in the Mammalian Genome *Cell* **167** 233-247.e217

Liu Y, Zeng Y, Liu L, Zhuang C, Fu X, *et al.* (2014) Synthesizing AND gate genetic circuits based on CRISPR-Cas9 for identification of bladder cancer cells *Nat Commun* **5** 5393 Machens F, Balazadeh S, Mueller-Roeber B and Messerschmidt K (2017) Synthetic Promoters and Transcription Factors for Heterologous Protein Expression in Saccharomyces cerevisiae *Front Bioeng Biotechnol* **5** 63

Makarova K S, Grishin N V, Shabalina S A, Wolf Y I and Koonin E V (2006) A putative RNA- interference-based immune system in prokaryotes computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action *Biology Direct* **1** 7

Makarova K S, Wolf Y I, Alkhnbashi O S, Costa F, Shah S A, *et al.* (2015) An updated evolutionary classification of CRISPR-Cas systems *Nat Rev Microbiol* **13** 722-736

Makarova K S, Zhang F and Koonin E V (2017a) SnapShot Class 1 CRISPR-Cas Systems *Cell* **168** 946-946

Makarova K S, Zhang F and Koonin E V (2017b) SnapShot Class 2 CRISPR-Cas Systems *Cell* **168** 328-328

Mali P, Esvelt K M and Church G M (2013) Cas9 as a versatile tool for engineering biology *Nat Methods* **10** 957-963

Marraffini L A and Sontheimer E J (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA *Science* **322** 1843-1845

Marraffini L A and Sontheimer E J (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity *Nature* **463** 568-571

McGinn J and Marraffini Luciano A (2016) CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration *Mol Cell* **64** 616-623

Miyoshi T, Ito K, Murakami R and Uchiumi T (2016) Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute *Nat Commun* **7** 11846

Mohanraju P, Makarova K S, Zetsche B, Zhang F, Koonin E V, *et al.* (2016) Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems *Science* **353** aad5147

Mojica F J, Diez-Villasenor C, Garcia-Martinez J and Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system *Microbiology* **155** 733-740

Mojica F J, Diez-Villasenor C, Garcia-Martinez J and Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements *J Mol Evol* **60** 174- 182

Mojica F J M and Rodriguez-Valera F (2016) The discovery of CRISPR in archaea and bacteria *FEBS J* **283** 3162-3169

Nakata A, Amemura M and Makino K (1989) Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome *J Bacteriol* **171** 3553-3556

Nam K H, Haitjema C, Liu X, Ding F, Wang H, *et al.* (2012)

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

Nelles D A, Fang M Y, O'Connell M R, Xu J L, Markmiller S J, *et al.* (2016) Programmable RNA Tracking in Live Cells with CRISPR/Cas9 *Cell* **165** 488-496

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

Nishimasu H and Nureki O (2016) Structures and mechanisms of CRISPR RNA-guided effector nucleases *Curr Opin Struct Biol* **43** 68-78

Nissim L, Perli S D, Fridkin A, Perez-Pinera P and Lu T K (2014) Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells *Mol Cell* **54** 698-710

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

O'Geen H, Ren C, Nicolet C M, Perez A A, Halmai J, *et al.* (2017) dCas9-based epigenome editing suggests acquisition of histone methylation is not sufficient for target gene repression *Nucleic Acids Res* **45** 9901-9916

Osawa T, Inanaga H, Sato C and Numata T (2015) Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog *Mol Cell* **58** 418-430

Park J Y, Moon B Y, Park J W, Thornton J A, Park Y H, *et al.* (2017) Genetic engineering of a temperate phage-based delivery system for CRISPR/Cas9 antimicrobials against Staphylococcus aureus *Sci Rep* **7** 44929

Perez-Rodriguez R, Haitjema C, Huang Q, Nam K H, Bernardis S, *et al.* (2011) Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in Escherichia coli *Mol Microbiol* **79** 584-599

Pettersen E F, Goddard T D, Huang C C, Couch G S, Greenblatt D M, *et al.* (2004) UCSF Chimera—a visualization system for exploratory research and analysis *J Comput Chem* **25** 1605-1612

Portillo M C, Gonzalez J M (2009) CRISPR elements in the Thermococcales evidence for associated horizontal gene transfer in Pyrococcus furiosus *J Appl Genet* **50** 421-430

Pougach K, Semenova E, Bogdanova E, Datsenko K A, Djordjevic M, *et al.* (2010) Transcription, processing and function of CRISPR cassettes in Escherichia coli *Mol Microbiol* **77** 1367-1379

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

Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, *et al.* (2010) Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS *Mol Microbiol* **75** 1495-1512

Punetha A, Sivathanu R and Anand B (2014) Active site plasticity enables metal-dependent tuning of Cas5d nuclease activity in CRISPR-Cas type I-C system *Nucleic Acids Res* **42** 3846-3856

Pyenson N C, Gayvert K, Varble A, Elemento O and Marraffini L A (2017) Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape *Cell Host Microbe* **22** 343-353.e343

Qi L S, Larson M H, Gilbert L A, Doudna J A, Weissman J S, *et al*. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression *Cell* **152** 1173-1183

Raines R T (1998) Ribonuclease A *Chem Rev* **98** 1045-1066

Ramanan V, Shlomai A, Cox D B, Schwartz R E, Michailidis E, *et al.* (2015) CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus *Sci Rep* **5** 10833

Rao C, Guyard C, Pelaz C, Wasserscheid J, BondyiDenomy J, *et al.* (2016) Active and adaptive Legionella CRISPRiCas reveals a recurrent challenge to the pathogen *Cell Microbiol* **18** 1319-1338

Richter C, Chang J T and Fineran P C (2012) Function and Regulation of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR Associated (Cas) Systems *Viruses* **4** 2291

Richter C, Dy R L, McKenzie R E, Watson B N, Taylor C, *et al.* (2014) Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer *Nucleic Acids Res* **42** 8516-8526

Rollins M F, Schuman J T, Paulus K, Bukhari H S, Wiedenheft B (2015) Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa *Nucleic Acids Res* **43** 2216-2222

Rouillon C, Zhou M, Zhang J, Politis A, Beilsten-Edmands V, *et al.* (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade *Mol Cell* **52** 124-134

Rousseau C, Gonnet M, Le Romancer M and Nicolas J (2009) CRISPI a CRISPR interactive database *Bioinformatics* **25** 3317-3318

Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R,

Horvath P, *et al.* (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli *Nucleic Acids Res* **39** 9275-9282

Sashital D G, Jinek M and Doudna J A (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3 *Nat Struct Mol Biol* **18** 680-687 Schmierer B, Botla S K, Zhang J, Turunen M, Kivioja T, *et al.* (2017) CRISPR/Cas9 screening using unique molecular identifiers *Mol Syst Biol* **13** 945

Scott T, Moyo B, Nicholson S, Maepa M B, Watashi K, *et al.* (2017) ssAAVs containing cassettes encoding SaCas9 and guides targeting hepatitis B virus inactivate replication of the virus in cultured cells *Sci Rep* **7** 7401

Semenova E, Jore M M, Datsenko K A, Semenova A, Westra E R, *et al*. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence *Proc Natl Acad Sci U S A* **108** 10098-10103

Semenova E, Nagornykh M, Pyatnitskiy M, Artamonova II and Severinov K (2009) Analysis of CRISPR system function in plant pathogen Xanthomonas oryzae *FEMS Microbiol Lett* **296** 110-116

Shalem O, Sanjana N E, Hartenian E, Shi X, Scott D A, *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells *Science* **343** 84-87

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

Shao Y, Richter H, Sun S, Sharma K, Urlaub H, *et al.* (2016) A Non-Stem-Loop CRISPR RNA Is Processed by Dual Binding Cas6 *Structure* **24** 547-554

Shibata M, Nishimasu H, Kodera N, Hirano S, Ando T, *et al.* (2017) Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy *Nat Commun* **8** 1430

Shmakov S, Abudayyeh O O, Makarova K S, Wolf Y I, Gootenberg J S, *et al.* (2015) Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems *Mol Cell* **60** 385-397

Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, *et al.* (2017) Diversity and evolution of class 2 CRISPR-Cas systems *Nat Rev Microbiol* **15** 169-182

Smargon A A, Cox D B, Pyzocha N K, Zheng K, Slaymaker I M, *et al.* (2017) Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28 *Mol Cell* **65** 618-630.e617

Sokolowski R D, Graham S and White M F (2014) Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system *Nucleic Acids Res* **42** 6532-6541

Spilman M, Cocozaki A, Hale C, Shao Y, Ramia N, *et al.* (2013) Structure of an RNA silencing complex of the CRISPR-Cas immune system *Mol Cell* **52** 146-152

Staals R H, Zhu Y, Taylor D W, Kornfeld J E, Sharma K, *et al.* (2014) RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus *Mol Cell* **56** 518-530

Staals R H J, Jackson S A, Biswas A, Brouns S J J, Brown C M, *et al.* (2016) Interference- driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR–Cas system *Nat Commun* **7** 12853

Sternberg S H, Haurwitz R E and Doudna J A (2012) Mechanism of substrate selection by a highly specific CRISPR endoribonuclease *RNA* **18** 661-672

Sternberg S H, LaFrance B, Kaplan M and Doudna J A (2015) Conformational control of DNA target cleavage by CRISPR-Cas9 *Nature* **527** 110-113

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

Swarts D C, Hegge J W, Hinojo I, Shiimori M, Ellis M A, *et al.* (2015) Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA *Nucleic Acids Res* **43** 5120-5129

Swarts D C, Jore M M, Westra E R, Zhu Y, Janssen J H, *et al.* (2014) DNA-guided DNA interference by a prokaryotic Argonaute *Nature* **507** 258-261

Swarts D C, Mosterd C, van Passel M W and Brouns S J (2012) CRISPR interference directs strand specific spacer acquisition *PLoS ONE* **7** e35888

Takei Y, Shah S, Harvey S, Qi L S and Cai L (2017) Multiplexed Dynamic Imaging of Genomic Loci by Combined CRISPR Imaging and DNA Sequential FISH *Biophys J* **112** 1773-1776

Tamulaitis G, Kazlauskiene M, Manakova E, Venclovas C, Nwokeoji A O, *et al.* (2014) Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus *Mol Cell* **56** 506-517

Tang T H, Bachellerie J P, Rozhdestvensky T, Bortolin M L, Huber H, *et al.* (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus *Proc Natl Acad Sci U S A* **99** 7536-7541

Tang T H, Polacek N, Zywicki M, Huber H, Brugger K, *et al.* (2005) Identification of novel non-coding RNAs as

potential antisense regulators in the archaeon Sulfolobus solfataricus *Mol Microbiol* **55** 469-481

Taylor D W, Zhu Y, Staals R H, Kornfeld J E, Shinkai A, *et al.* (2015) Structural biology. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning *Science* **348** 581-585

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

Wang R, Li M, Gong L, Hu S and Xiang H (2016) DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in Haloarcula hispanica *Nucleic Acids Res***44** 4266-4277

Wang R, Preamplume G, Terns M P, Terns R M and Li H (2011) Interaction of the Cas6 Riboendonuclease with CRISPR RNAs Recognition and Cleavage *Structure* **19** 257-264 Wang T, Wei J J, Sabatini D M and Lander E S (2014) Genetic screens in human cells using the CRISPR-Cas9 system *Science* **343** 80-84

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

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

Westra E R, Pul U, Heidrich N, Jore M M, Lundgren M, *et al.* (2010) H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO *Mol Microbiol* **77** 1380-1393

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

Westra E R, Swarts D C, Staals R H, Jore M M, Brouns S J, *et al.* (2012a) The CRISPRs, they are a-changin' how prokaryotes generate adaptive immunity *Annu Rev Genet* **46** 311- 339

Westra E R, van Erp P B, Kunne T, Wong S P, Staals R H, *et al.* (2012b) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3 *Mol Cell* **46** 595-605

Wiedenheft B, Lander G C, Zhou K, Jore M M, Brouns S J, *et al.* (2011a) Structures of the RNA-guided surveillance complex from a bacterial immune system *Nature* **477** 486-489

Wiedenheft B, van Duijn E, Bultema J B, Waghmare S P, Zhou K, *et al.* (2011b) RNA- guided complex from a bacterial immune system enhances target recognition through seed sequence interactions *Proc Natl Acad Sci U S A* **108** 10092-10097

Wright A V and Doudna J A (2016) Protecting genome integrity during CRISPR immune adaptation *Nat Struct Mol Biol* **23** 876-883.

Wright A V, Liu J-J, Knott G J, Doxzen K W, Nogales E, *et al.* (2017) Structures of the CRISPR genome integration complex *Science* **357** 1113

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

Wu D, Guan X, Zhu Y, Ren K and Huang Z (2017) Structural basis of stringent PAM recognition by CRISPR-C2c1 in complex with sgRNA *Cell Res* **27** 705-708

Yamano T, Nishimasu H, Zetsche B, Hirano H, Slaymaker I M, *et al.* (2016) Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA *Cell* **165** 949-962

Yang H, Gao P, Rajashankar K R and Patel D J (2016) PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease *Cell* **167** 1814-1828

Yoganand K N R, Sivathanu R, Nimkar S and Anand B (2017) Asymmetric positioning of Cas1–2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system *Nucleic Acids Res* **45** 367-381

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

Zetsche B, Gootenberg J S, Abudayyeh O O, Slaymaker I M, Makarova K S, *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system *Cell* **163** 759- 771

Zhang J, Graham S, Tello A, Liu H and White M F (2016) Multiple nucleic acid cleavage modes in divergent type III CRISPR systems *Nucleic Acids Res* **44** 1789-1799

Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, *et al.* (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity *Mol Cell* **45** 303-313

Zhang Y, Heidrich N, Ampattu B J, Gunderson C W, Seifert H S, *et al.* (2013) Processing- independent CRISPR RNAs limit natural transformation in Neisseria meningitidis *Mol Cell* **50** 488-503

Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, *et al.* (2014) Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli *Nature* **515** 147-150

Zhou Y, Wang P, Tian F, Gao G, Huang L, *et al.* (2017) Painting a specific chromosome with CRISPR/Cas9 for live-cell imaging *Cell Res* **27** 298-301.