Introduction of CRISPR/Cas system, as a genetic interference pathway and essential factor in genome evolution

Tahere Seifi [1], Tina Shaffaf [1], Seyed Reza Kazemi Nezhad [1*]

[1]Department of Genetics, Faculty of Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran

*Corresponding author: Seyed Reza Kazemi Nezhad, Department of Genetics, Faculty of Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran. Email: kazemi_reza@yahoo.de

ABSTRACT

The clustered regularly interspaced short palindromic repeats / (CRISPR) associated protein system is an adaptive immune system in bacteria and archaea. This system mimics eukaryotic RNA interference (RNAi) system and can play an important role in gene regulation and immune system. CRISPR interference (CRISPRi) defends the bacteria against invader agents, such as phages and plasmids. As a result, it can provide a genetic interference pathway depending on the integration of a small stretch of invader DNA of the bacterial or archaeal genome, which acts as a heritable genetic memory to defense against viruses or phages.

INTRODUCTION

Acquisition of a new gene, which confers a selective advantage, is an essential factor in genome evolution. A great proportion of bacterial and archaeal genomes contain genes derived from exchanging the genetic material among related or unrelated species [1]. This phenomenon is known as Horizontal Gene Transfer (HGT). HGT occurs with transformation, conjugation or transduction [2]. A small fraction of acquired genes confers selective advantage, thus microorganisms evolve many genetic mechanisms to prevent HGT, such as DNA restriction or surface exclusion [2].

During the past two decades, clustered regularly interspaced short palindromic repeats (CRISPR) systems have been identified as novel genetic interference pathways limiting at least two major mechanisms of HGT consist of conjugation and transduction. Mechanisms such as eukaryotic RNAi and related pathways are analogous to CRISPRi [3, 4]. Unlike other defense mechanisms, CRISPRi is an adaptive and heritable immune system and enables bacteria to control the sequence length, and to be reprogrammed to reject the invading DNA molecule, which has not been previously encountered [3]. In this review, we discuss about the CRISPR structure and functions, and mention various applications, which have advantages in basic and medical genetics research.

Discovery of the CRISPR system

In 1987, Ishino and his colleagues cloned and sequenced *iap* gene, which is responsible for alkaline phosphatase isozyme conversion in *E. coli* [5]. He found that there was a set of 29 nucleotide (nt) repeat downstream of *iap* gene. These repeats were separated by non-repetitive, and similar short sequences known as spacers. Similar repeats were detected when the whole genome sequencing projects were carried out in bacteria and archaea [6, 11]. Now a days, CRISPR is introduced as a family of repeats present in many bacterial, and archaeal species. This sequence was nominated as CRISPR in 2002 [38]. In fact, this term can reflect the loose structure (Fig 1). A 'Leader' sequence precedes a cluster of repeats which is an AT rich region with several hundreds base pairs length which has intraspecies conservation [38].
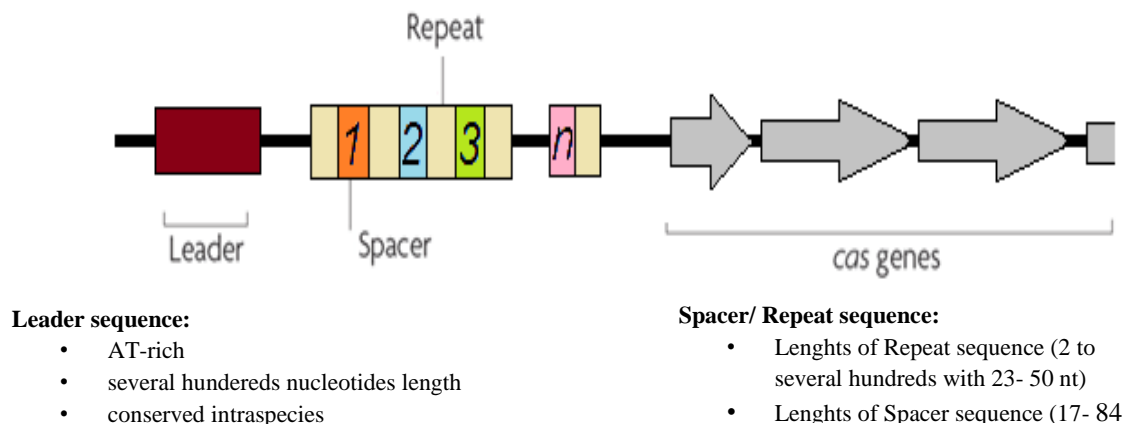

Leader sequence:
- AT-rich
- several hundereds nucleotides length
- conserved intraspecies

Spacer/ Repeat sequence:
- Lenghts of Repeat sequence (2 to several hundreds with 23- 50 nt)
- Lenghts of Spacer sequence (17- 84

Figure 1. CRISPR array structure. CRISPR array involves an AT-rich leader sequence with several hundreds nucleotides length, which is conserved intraspecies. Second CRISPR array segment is Repeat/spacer sequence. The number of repeat sequences is from two to a few hundreds with 23- 50nt length. Spacer sequences are 17-84nt long matched with sequences from mobile genetics elements, such as bacteriophages and plasmids.

As mentioned, CRISPR/Cas system provides adaptive immune system against invading agents [3, 4]. This immunity is based on the integration of the invader DNA stretch into the CRISPR loci as proto-spacer which acts as genetic memory for protection against viruses or phages [3, 4]. The space sequence is matched with the foreign sequences, mobile genetic elements [13, 15]. This provides the possibility of a potential base pairing that can enable a sequence-based interference with phage infection, gene expression or both. These suggest that the function of CRISPR system is analogous to eukaryotic RNAi [17, 18].

Analysis of the CRISPR loci in *Yersinia pestis* strains demonstrated that the form of new spacer acquisition is polarized [15], meaning that new units have been added at one end of the cluster close to the next end [15, 16]. Note that *Cas* genes are essential for CRISPR function. In 2008, Marraffini *et.al*. Showed that limitation of Plasmid conjugation can be induced using CRISPRi in *Staphylococcus epidermidis* [19]. This research demonstrated an association between CRISPR/*Cas* gene loci and prevention of HGT in bacteria; Thus, CRISPR/Cas system is known as the immunity system.

Classification and Structure of the CRISPR loci

CRISPR sequences are presented in approximately 40% of *Bacteria* and 90% of *Archaea*, which have at least one CRISPR locus [20]. This marked difference may be caused by a bias in sequences of bacteria. This bias is presented in long established laboratory strains that have had no bacteriophage exposure for many generations, so their genome lacks the CRISPR loci.

CRISPR sequence consists of the repeat-spacer units ranging from just a few to several hundred numbers, the average is 66. The highest number is observed in the genome of the thermophilic bacterium, *Chloroflexus sp. Y400fl* with 374 units in one of its four CRISPR loci. The Repeat units have various sequences, even among CRISPR loci in the same genome. They exhibit limited dyad symmetry. There are 12 categories of CRISPR loci based on similarity of the sequences and their potential to form stem-loop structure [20]. A 23-50 nt range is reported as repeat length with an average of 31 nt, and 17-84 nt range is reported as spacer sequence length with an average length of 36 nt. The presence of various sets of *Cas* genes increases the complexity of CRISPR loci. More than 40 *Cas* gene families have been identified [17, 24]. There are eight CRISPR subtypes based on the presence of 45 gene families associated with CRISPR loci (established by Haft *et.al*.) [24] (Table 1). *Cas1-cas6* was identified as gene families, which are considered 'core' *Cas* genes. *Cas1* and *Cas2* are introduced in all CRISPR loci (in a few cases

Cas2 is encoded as a combined domain of Cas3 protein) [24, 25].

Acquisition of new spacer sequences

At the molecular level, CRISPR function can be divided into three phases during which the new spacers are incorporated in the CRISPR array: Acquisition, Expression and Interference (Fig 2a), the CRISPR RNAs (crRNAs) are expressed, processed (Fig 2b) and CRISPRi occurs (Fig 2c) [26]. During the first step, a new spacer, who is known as proto-spacer, integrates in the invader genome. This procedure allows the microorganism to increase the potential of rapid adaptation with the environmental invaders. Thus, this step can be referred to as 'adaptation' phase [3, 26]. The spacer incorporation has an unknown molecular mechanism in detail. In *Ecoli*, Cas1 and Cas2 are necessary for the pre-existing function of spacers [27]. Because of the widespread presence of these proteins in CRISPR/Cas systems, it is thought that they participate in adaptation. The short conserved flanking sequences of proto-spacer (2-5bp) in invader genome are named 'CRISPR motifs' or 'proto-spacer adjacent motif' (PAM), which are classified based on their CRISPR sequences [20, 25]. The motif presence indicates that selection of proto-spacer is not random and a recognition signal is needed for the selected target sequence that is considered as a new spacer [25]. Mutation in CRISPR motifs or PAM sequences can cause CRISPR immunity evading which indicates the flanking sequences role during the defense (or interference) phase [21, 25].

The integration stores the genetic information to use it as inhibitor of the invader agents, specifically during 'defense' phase of CRISPRi [4]. CrRNA transcription correlates the adaptation phase with the defense phase [28, 29]. Transcription of CRISPR initiates from the leader sequence. Studies on thermophiles revealed that the camp regulator protein can upregulate *Cas* genes and camp signal transduction can guide CRISPR response in this bacterium [28-30].

CRISPR transcription and processing

CRISPR transcript (CRISPR precursor RNA) is cleaved and processed into a small crRNA by Cas proteins. Cas proteins are multi-complex proteins known as CRISPR associated complexes for antiviral defense (cascade) [31]. The components of cascade may be different in each subtype. Experimental data revealed that pre-crRNA processing occurs endo-nucleolytically at a specific site of each repeat
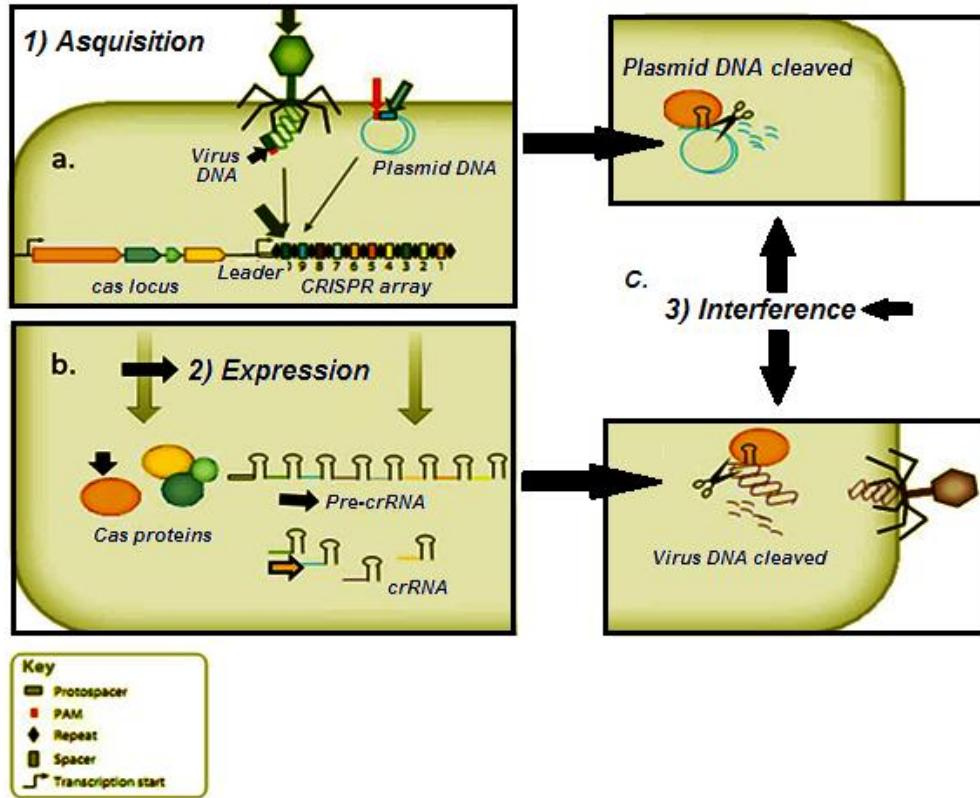
Figure 2. Three phases of CRISPR/Cas function. (a) Acquisition, new spacers are incorporated in CRISPR array, (b) Expression, crRNAs are expressed and processed, (c) Interference. The Expressed and processed RNA matches with the invader genome, Cas proteins can recognize this structure. After that, invader genome will be cleaved with Cas helicase and nuclease activity.

(8 nt upstream of the spacer). Mature crRNA is Associated with the cascade. CrRNAs form a stem-loop structure because they have palindromic repeats. CrRNA involves a spacer-flanking unit with repeat sequences (Fig 2b). CRISPRs are separated by short spacer sequences matching the DNA of bacteriophage or plasmid. Therefore, crRNAs consist of a spacer sequence can be matched with invader genome and formed base pairs. Finally, specific Cas proteins identify stem-loop structure and cut the invader genome and inactive it.

Recognition of the invading sequences

The base pairing potential between crRNA and the invader genome is mediated by the recognition of crRNA-Cas ribonucleoprotein (crRNP) complex. Base pairing is initiated from a seed sequence, after that pairing is promoted and finally DNAse activity is initiated which causes the cleavage of the invader genome [32-34]. It is noted that the CRISPR machinery recognizes DNA targets rather than RNA ones in numerous species; an exception is type III-B of

CRISPR/Cas system, which recognizes RNA targets [50]. *In vitro* analysis suggests that during RNA cleavage, 3'phosphate and 5'hydroxyl ends are generated at a 14 nt distance from the target sequence which is paired to the 3'terminal of the crRNA [3].

CRISPR-Cas diversity

The dynamic and selective pressures cause rapid evolution of the immune system because of the high potential of the CRISPR/Cas system to adapt with the new environmental agents rapidly. Therefore, it is acceptable that CRISPR/Cas systems are highly diverse. The comparative studies of CRISPR loci demonstrated that the sequences of the CRISPR repeats, the *Cas* genes and *Cas* operon architecture are major differences between various systems. Based on these differences, three main types of CRISPR were determined as type I, type II and type III; these three types are divided into several subtypes themselves [41]. Each type has a specific signature Cas-proteins: Cas3-helicase, Cas9-nuclease, and Cas10 (a large protein with unknown function) are defined for type I, II and III systems, respectively. Type II system is distincted from two other types phylogenetically and structurally. Nomenclature is based on the target, cleaving the invading agents, crRNA and Cas-protein composition of crRNP complexes or cascade [41]. Cas protein composition of type I-

A to type I-F crRNP complexes are determined as cascade, but all type II crRNPs, which are divided to type II-A, II-B and II-C, are known as Cas 9 complexes. The Cas proteins involved in Type III-A and Type III-B crRNPs systems are known as Csm and Cmr complexes.

Type II system is only found in bacteria in composition with other CRISPR-Cas types, whereas Types I and III are found in different combinations among bacteria and archaea which are phylogenetically diverse [39, 42]. It is interesting to note that viral genomes and plasmids have CRISPR-Cas systems, which suggests that HGT provides the possibility of exchanging CRISPR systems [43-47]. Despite the diversity, there are four functional categories for all Cas proteins: Nucleases or recombinase, involved in acquisition of the spacer; ribonucleases, catalyzes the crRNA guides processing; proteins present in crRNA complexes and finally the proteins which have nuclease activity to degrade DNA and RNA targets.

Type I and Type III systems

In these two types, Cas6-like nuclease plays a role in pre-crRNA primary processing. Cas6 proteins have structural variations. Their enzymatic activity for cleaving pre-crRNA is involved in hydrolyzing a single phosphodiester bond in repeated sequences of the transcript. The

crRNA arrangement is consisted of a 5' unusual seahorse shape (binding to DNA changes its conformation) of 8 repeated nucleotides (complete spacer sequence) derived from 3'seahorse shape with a variable size which forms a hairpin structure in some systems. The Cas6 protein delivers primary crRNAs to the respective crRNP complexes [49, 50]. In some subtypes, after transferring the crRNAs to the protein complexes, additional processing occurs to produce different mature crRNAs and the resulted crRNAs can have different nucleotide numbers [51, 52]. Cas5 protein is substituted with Cas6 in type I-C system. It shows the variation in applied proteins as Cas proteins in different types and subtypes of CRISPR systems. In type II systems, pre-crRNA processing occurs completely different; CRISPR locus includes a *Cas* operon, a CRISPR array and a gene encoding a *trans*-activating crRNA (tracrRNA) [53]. The tracrRNA is complementary to the repeat region of the pre-crRNA transcript, which is consisted of 25 nucleotides. The housekeeping ribonuclease RNase III recognizes and cleaves a double stranded region resulted from base pairing between these two RNAs. Primary crRNA processing requires Cas9 protein [53, 54]. This protein probably plays an important role in binding to the RNA molecule and positioning it for being cleaved by RNase III. The association of Cas9 with crRNA-tracrRNA hybrid follows

initiation of the processing by RNase III. The 5'end of the crRNA spacer is 24-27 nt trimmed in length [53]. The mature crRNA-tracrRNA hybrid requires interfering the target. This structure is probably necessary for proper anchoring and positioning of the crRNA in Cas9. Following binding to the target nucleic acid, major rearrangements occur in Cas9 domains. These rearrangements are revealed using crystallography studies.

Assembly of crRNP complexes

Stable crRNP complexes are consisted of mature crRNAs and Cas proteins. The multi-subunit surveillance complexes of type I systems are called Cascade, which are the first identified crRNP complexes in *E. Coli* type I-E system [34, 84]. The structural studies have shown that there are architectural similarities between type I and type III crRNPs [54-57], but Cas9 is the only protein component that presents in type II crRNP complexes [54-57].

Type I crRNP complexes

Cas5, Cas6, Cas7 and a single 61-nucleotide crRNA are known as the core complex of type I-E crRNP of *E. Coli*. These complex proteins are less tightly associated with the Cse1 and Cse2 subunits. One Cse1, two Cse2s, one Cas5, six Cas7s and Cas6 (in I-E and I-F subtypes) are the complete complexes that have uneven subunit

stoichiometries. Features of type I and III crRNP complexes are provided at the time of alternative translation of encoding polycistronic mRNAs [58]. The helical backbone is composed of Cas7 (Csy3 in type I-F system), Cas8 (known as Csy1 in type I-F system), Cas5 (known as Csy2 in type I-F system) and the crRNA is consisted of Cascade complexes of type I-C [59] and type I-F systems [60]. Cas3' and Cas3'' as truncated domains of Cas3, act as a helix, and a nuclear domain respectively, and form different parts of the complex.

Type II crRNP complexes

Analysis of Cas9 proteins from the type II-A system of *Streptococcus pyogenes* shows that Cas9 has a conserved architecture consisting of two distinct lobes, the recognition lobe (the α-helix structure) and the nuclease lobe. The first lobe is involved in coordinating with the guild RNA (gRNA), and the second lobe is responsible for PAM recognition and the target DNA cleavage. Reorientation of two structural lobes following binding to the single guide RNA (sgRNA) causes a central channel formation and then DNA substrates will become able to attach [56]. The groove located at the interference surface of two lobes has a positive charge and

filled with an RNA-DNA hetro-duplex. As a result, Cas9-sgRNA-ssDNA target structure is performed. HNH and Ruv C, as two nuclear domains, are not sufficient and cleavage does not happen. R-Loop configuration is required to reach the ultimate cleavage-competent state of Cas9. Because of CRISPR/Cas9 importance in genetic regulating, editing and engineering, performance of type II CRISPR/Cas9 is illustrated as a schematic figure (Fig 3).

Type III crRNP complexes and similarities with type I systems

Csm and Cmr complexes were determined from type III-A crRNA of *S. solfataricus* and two type III-B crRNP complexes from *Pyrococcus furiosus* and *Thermus thermophilus*, respectively (61). Csm3 in type IIIA and Cmr4 in type III-B systems are consisted of a multi-copy backbone that belongs to Type III complexes. This complex is a structural homolog of Cas7. The Cmr complexes from *P. furious* are the same as crRNA in cascade because of their positioning along the backbone. The Subunits of Type I and III systems can be introduced as small and large subunits.
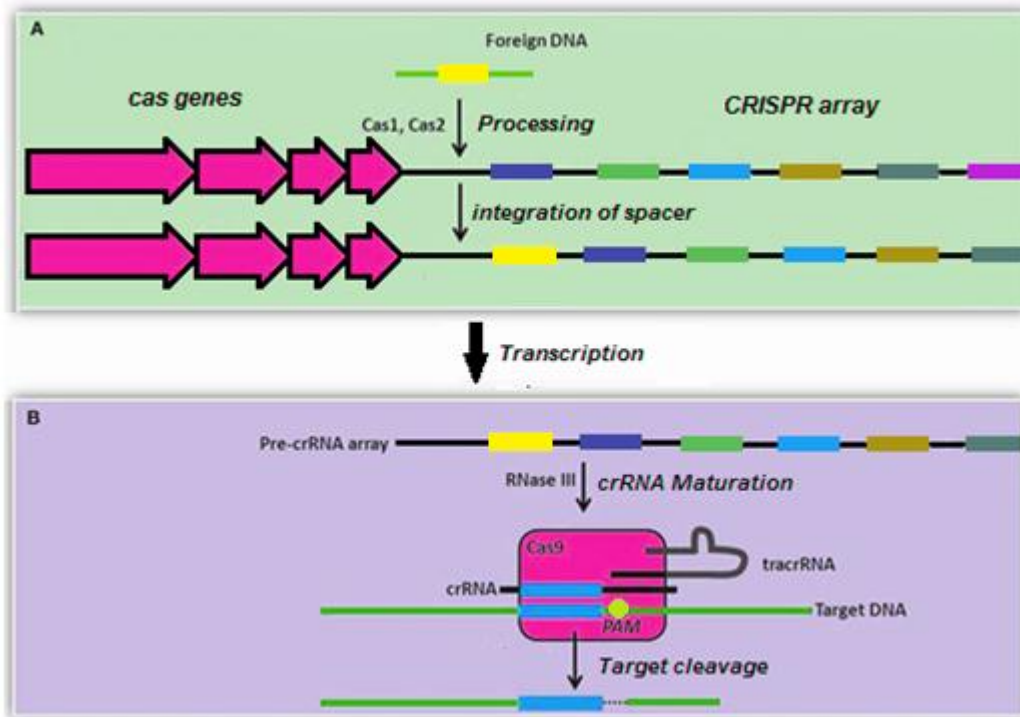
Figure 3. Three phases of Type II CRISPR/Cas9 function. (A) Invader DNA is recognized by Cas1 and Cas2, and processed (red) and integrated at the leader end of the CRISPR array (red) (adaptation or acquisition phase, blue), (B) For protection against invader DNA in the next attack, CRISPR is transcribed as a single transcript (pre-crRNA), then it is matured into a small targeting crRNA using RNase III and tracrRNA. CrRNA and tracrRNA complex associate with Cas9, and then hybridization of the complementary DNA sequence is mediated by spacer sequence. Cas9 cleaves the targeted DNA at downstream of PAM (red circle) (Effector or expression and interference phases, pink).

The small subunit of Type III-A complexes consists of three copies of Csm2 and Cmr5 in Type III-A and Type III-B complexes, respectively [61, 62]. Type I and Type III crRNA complexes have conserved position of the large substitution, which reveals that their roles are analogue. The small subunits of type I and Type III systems are conserved in some parts of the structure [61-63].

Target surveillance and interference

The invader genome targeting proceeds in a stepwise manner [26]. This process contains several steps: (1) Scanning the invader genome for Proto-spacer sequence to find out which one complements crRNA. (2) Discriminating itself from outsider. (3) Base pairing between the seed region and the complementary proto-spacer (7-8nt length) following extension of pairing causes strand displacement [64-66]. Hybridization of crRNA, and target strand forms R-Loop structure which cause changes in crRNP complex

conformation [61, 67-69]. This transition in structure provides a signal to recruit a trans-acting nuclease (for type I and type III-A systems) or leads to activate of intrinsic nuclease activity (for type II and type III-B systems). Nonspecific interaction between crRNP complex and the invader genome results in scanning the invader genome. PAM is recognized by one of the subunits of cascade, this association between crRNP, and PAM motifs may destabilize the duplex of DNA. Thus, crRNA is enabled to hybridize with target DNA. Pairing occurs between the seed region of crRNA spacer (close to the PAM motif), and the proto-spacer complementary to the target. Hybridization between crRNA, and target DNA generates R-Loop, which causes a major change in the conformation of cascade and local bending of target DNA [70, 71]. Recruitment of Cas3 nuclease-helicase is caused by these structural changes. Helicase activity of Cas3, which depends on ATP which is combined with metal-dependent nuclease activity, causes completely degradation of target DNA [71, 72] The displaced strand of R-Loop is cleaved by endo-nucleotide cleavage followed by a 3' to 5' exo-nucleotide degradation [71-73], the other DNA strands undergo endonucleolytic, and exonucleolytic degradations [71, 72-75].

Type II systems

Type II system interference is completely different from type I and type III systems. Cas9-RNP complex is consisted of Cas9, crRNA and tracrRNA which perform the interference in this type [53-55, 76]. Cas9 targets a DNA fragment in a step-wise manner [77]. SgRNA works as a leader and causes structural rearrangements that lead to a central channel formation binding to the target DNA [56]. The DNA is scanned for a PAM motif by Cas9-RNA complex. Type I, type II systems scan, and find DNA using similar pathways. PAM motif is located on displaced strand of target DNA and positioned to the 3´seahorse shape of crRNA guide. Cas9 nuclease lobe determines PAM recognition using the flexible loops contained two tryptophans in its C-terminal domain [77]. DNA binding, DNA strands displacement and initiating of R-Loop formation in PAM require interaction with the PAM motif. Base pairing initiates in seed sequence for over 12 nt and proceeds to the distal end of the target sequence [77]. At the final stage, Cas9 cleaves the DNA using the nucleolytic activity of two active sites in nuclease lobe. The DNA strand paired with the crRNA is cleaved by a HNH-like nuclease domain and also the displaced DNA strand is cleaved by a Ruv-like nuclease domain [54, 56]. So, a blunt double-stranded end at a specific site with 3 nt distance

of 3'end of the proto-spacer is resulted [76]. Cas9 is known as a single-turnover enzyme [77].

Type III systems

Type III-A system complex is named Csm and contains Csm1-Csm5 and crRNA [78, 79]. Details of how the type III crRNP complexes target the proto-spacer is unknown. Type III-A systems discriminate self-DNA from non-self DNA in a PAM-independent manner [80]. This process is performed with Cas10-like protein Csm1 (large subunit) mediation. It is proposed that crRNA contains a seed sequence at the 5'end of the spacer region [38]. The Csm6 (a helicase and nuclease associated with Csm complex) is involved in DNA interference. Csm6 is known as Csx1, and its analogue with Cas3 in type I systems.

The complex of type III-B systems is Cmr which contains Cmr1-Cmr6 [78, 79]. The special feature of these complexes is that they target RNA rather than DNA [81]. RNA cleavage proceeds from 3'end to 5'end [62]. The spacer contains a seed-like sequence located at the 5'end. Results have shown that these complexes have multiple catalytic sites. Cmr-α, and Cmr-β are two Cmr complexes in *Sulfolobus islandicus*. Cmr-α is composed of Cmr1-Cmr6, that target plasmid DNA using a Csx1-dependent mechanism that requires targets transcription [82]. This seems to be a functional analogue of the aforesaid Csm complex of III-A systems. But Cmr-β complex is composed of Cmr1-Cmr7 in *Sulfolobus islandicus* and targets RNA *in vitro* (51), but its catalytic mechanism is different from the mechanism of III-B system in *P. furious* and *T. Thermophiles* [61, 62]. These results reveal that the mechanisms and functions of heteroduplex formation in type III systems are diverse [82]. Molecular details of crRNA binding by crRNP complexes are diverse between different CRISPR/Cas systems. Interestingly, rice protein mediates base-pairing pattern. Strings of RecA molecules, in complex with an ssDNA template, invade a dsDNA helix. After the displacement of the non-complementary strand, the new hybrid formed of complementary DNA strands provide a space that is locally allowed to adopt a class B conformation. This process resembles the crRNA–target hybrid configuration of the Cascade R loop.

The different structure of crRNA-target hetroduplex is provided by a single stretch resulted from crRNAs binding to the targets and increases the flexibility of crRNAs. In type II systems, the sgRNA base pairing with the 20 nt of target DNA is mediated by Cas9 protein [57]. The RNA targets in type III-B systems are similar to eukaryotic RNAi because of using a 21 nt guide RNA [83].

CRISPR system as a genetic editing and genetic engineering tool

In 2012, for the first time, Jennifer Doudna and her colleagues published their findings about genome editing capabilities of bacterial immunity system introduced by Danisco group [54, 76]. They showed that Cas9 protein involved in type II system can cleave the target dsDNA via detection of a base-paired structure formed between the activating tracrRNA, and the targeting crRNA. In addition, they demonstrated that Cas9 endonuclease family can be programmed by a single RNA molecule to cleave specific DNA sites; these observations raised the exciting possibility of developing a simple and versatile RNA-directed targeting, and editing. Cas9, the hallmark of type II systems, has been hypothesized to be involved in both crRNA maturation and crRNA-guided DNA interference [84]. Type II of CRISPR/Cas system has a powerful potential to be applied in genetic editing, and engineering (Fig. 4) [84]. This system has a good function in basic and applied molecular and medical genetic parallel to TALEN and ZFN (Zinc-Finger Nuclease) techniques [87-89]. Although, working with CRISPR is simple and more sufficient than TALEN and ZFN [89]. Type II of CRISPR/Cas system has been applied to discover a drug for muscle disease in zebrafish models [90].

Emergency of new era for genome sequencing reopens annotation of new genomic variations which are valuable to investigate their effects on phenotypes [85]. CsRISPR/Cas system can be applied to induce Indel variations in desired sequences using its matching and nuclease ability (Fig. 4a). In addition, CRISPR/Cas9 can be attached to activator or silencer motifs and be used as a tool to up-regulate or down-regulate the expression of interested gene (Fig. 4b). GFP-CRISPR/Cas construct is used for imaging studies to locate a genomic locus (Fig. 4c). These applications are used to study the mechanism of upstream or downstream sequences to control gene expression [85, 86].

CONCLUSION

During the past two decades, impressive progress has been made in understanding many aspects of the molecular features of CRISPR/Cas systems. CRISPR/Cas system is defined as a heritable adaptive immunity system in bacteria, and archaea. The discovery of this system reveals the genome evolution mediated by horizontal gene transfer. The molecular mechanism of this system has shown the coincide regulation to provide defense against environmental agents. This protection is prepared by a varied set of proteins [71-75]. The variations in environmental agents are driving forces of genome evolution. Understanding the details of the molecular

mechanism of CRISPR/Cas systems proceeds to point that the CRISPR/Cas applications are emerged and applied in genome editing, and engineering [85]. The CRISPR/Cas system has significant applications in basic researches, which is a good advantage for studying the molecular mechanisms. The emergence of the whole genome sequencing causes detection of the novel variations in the genome. The ability of CRISPR/Cas system in Indel variation introduction has made it as a powerful tool to perform a functional assay for new annotated variations.
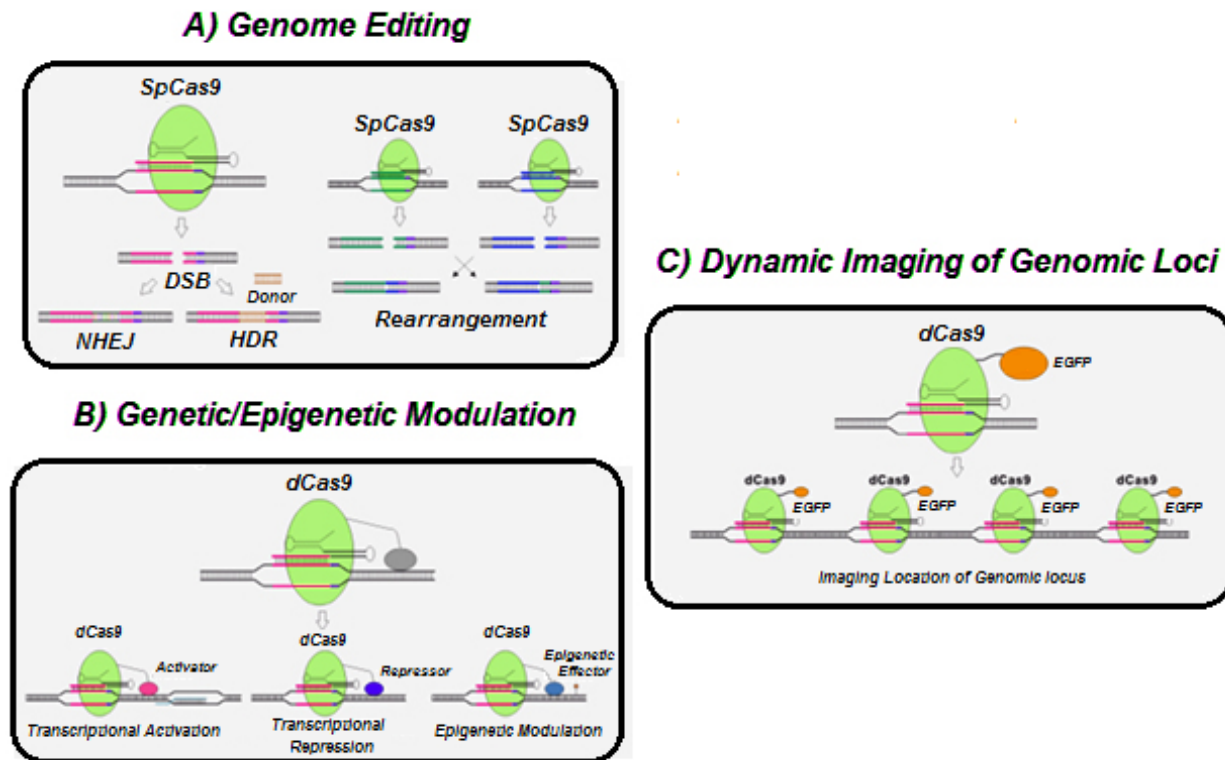

Figure 4. CRISPR/Cas application in (a) Genome Editing, (b) genetic/Epigenetic Modulation and (c) Dynamic imagining of Genomic loci. Cas9 (dCas9) is a form of Cas9 mutated in nuclease domain. dCas9 is attached to the VP64 and P65 subunits of NF-KB, and KRAB domain as activator and suppressor proteins, respectively. It can be used to control the target gene expression. (C) dCas9 attached to EGFP, useful for molecular genetic studies like telomere movements studies.

REFERENCES

[1]. Nakamura Y, Itoh T, Matsuda H, Gojobori T. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet*, 2004; 36(7): 760–6.

[2]. Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol*, 2005; 3(9): 711–21.

[3]. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet,* 2010; 11(3): 181- 90.

[4]. Bhaya D, Davison M, and Barrangou R. CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annu Rev Genet*, 2011; 45: 273– 97.

[5]. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol*, 1987; 169(12): 5429–33.

[6]. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG. Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science,* 1996; 273(5278): 1058–73.

[7]. Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, Quackenbush J, Lee NH. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus. *Nature*, 1997; 390(6658): 364–70.

[8]. Nelson KE1, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima. *Nature,* 1999; 399(6734): 323–29.

[9]. Charlebois RL, Chow C, Clausen IG, Curtis B, Doolittle WF, Duguet M. Completing the sequence of the Sulfolobus solfataricus P2 genome. *Extremophiles,* 1998, 2: 305–12.

[10]. Kawarabayasi Y1, Hino Y, Horikawa H, Yamazaki S, Haikawa Y, Jin-no K. Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, Aeropyrum pernix K1. *DNA Res*, 1999; 6: 83–101.

[11]. Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, Yamamoto S, Sekine M, Baba S, Kosugi H, Hosoyama A, Nagai Y, Sakai M, Ogura K, Otsuka R, Nakazawa H, Takamiya M, Ohfuku Y, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Kikuchi H. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, Pyrococcus horikoshii OT3. *DNA Res*, 1998; 5(2): 55–76.

[12]. Weiss D, Sampson T. CRISPR-Cas systems: new players in gene regulation and bacterial physiology. *Front Cell Infect Microbiol*, 2014; 4: 4-37.

[13]. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 2005; 151(Pt 8): 2551–61.

[14]. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol*, 2005; 60(2): 174–82.

[15]. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 2005; 151(Pt 3): 653–63.

[16]. Barrangou R, and Horvath P. CRISPR: New Horizons in Phage Resistance and Strain Identification. *Annu Rev Food Sci Technol*, 2012; 3: 143-62.

[17]. Lillestøl RK, Redder P, Garrett RA, Brugger K. A Putative viral defence mechanism in archaeal cells. *Archaea*, 2006; 2(1): 59–72.

[18]. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A Putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*, 2006; 1: 7.

[19]. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, 2008; 322(5909): 1843–45.

[20]. Shah SA, Erdmann S, Mojica FJ, Garrett RA. Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol*, 2013; 10(5): 891–99.

[21]. Swarts DC, Mosterd C, Van Passel MW, Brouns SJ. CRISPR interference directs strand specific spacer acquisition. *PLoS ONE*, 2012; 7(4): e35888.

[22]. Diez-Villasenor C, Guzman NM, AlmendrosC, Garcia-Martinez J, Mojica FJ. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR–Cas I-E variants of Escherichia coli. *RNA Biol*, 2013; 10(5): 792–802.

[23]. Erdmann S. Garrett RA. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol. Microbiol*, 2012; 85: 1044–56.

[24]. Yosef I, Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res,* 2012; 40(12): 5569–76.

[25]. Vercoe RB, Chang JT, Dy RL, Taylor C, Gristwood T, Clulow JS, Richter C, Przybilski R, Pitman AR, Fineran PC. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet*, 2013; 9: e1003454.

[26]. Van Der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci*, 2009; 34(8): 401–07.

[27]. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 2008; 321(5891): 960–64.

[28]. Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, Shinkai A. Transcription profile of Thermus thermophilus CRISPR systems after phage infection. *J. Mol. Biol*, 2010; 395(2): 270–81.

[29]. Shinkai A, Kira S, Nakagawa N, Kashihara A, Kuramitsu S, and Yokoyama S. Transcription activation mediated by a cyclic AMP receptor protein from Thermus thermophilus HB8. *J. Bacteriol*, 2007; 189(10): 3891–3901.

[30]. Chattoraj P, Banerjee A, Biswas S, Biswas I. ClpP of Streptococcus mutans differentially regulates expression of genomic islands, mutacin production and antibiotics tolerance. *J. Bacteriol,* 2010; 192(5): 1312-13.

[31]. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 2008; 321(5891): 960–64.

[32]. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 2007; 315(5819): 1709–12.

[33]. MarraffinI LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, 2008; 322(5909): 1843–45.

[34]. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol*, 2008; 190(4): 1390–1400.

[35]. Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA. Structural basis for DNase activity of a conserved protein implicated in CRISPR mediated genome defense. *Structure*, 2009; 17(6): 904–12.

[36]. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 2009; 155(Pt 3): 733–40.

[37]. Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol*, 2007; 8(4): R61.

[38]. Jansen R, Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol*, 2002; 43(6): 1565–75.

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

[40]. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*, 2006; 1: 7.

[41]. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. Evolution and classification of the CRISPR–Cas systems. *Nat Rev Microbiol*, 2011; 9(6): 467–77.

[42]. Heidrich N, Vogel J. Same same but different: new structural insight into CRISPR–Cas complexes. *Mol Cell*, 2013; 52(1): 4–7.

[43]. Fonfara I, Le Rhun A, Chylinski K, Makarova KS, Lécrivain AL, Bzdrenga J, et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR–Cas systems. *Nucleic Acids Res*, 2013; 42(4): 2577–90.

[44]. Godde JS, Bickerton A. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol*, 2006; 62(6): 718–29.

[45]. Lillestøl RK, Shah SA, Brügger K, Redder P, Phan H, Christiansen J, Garrett RA. CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. Mol. *Microbiol*, 2009; 72(1): 259–72.

[46]. Seed KD, Lazinski DW, Calderwood SB, Camilli A. A bacteriophage encodes its own CRISPR-Cas adaptive response to evade host innate immunity. *Nature,* 2013; 494(7438): 489–91.

[47]. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res*, 2011; 21(10): 1616–25.

[48]. Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. Rapid evolution of the human gut virome. *PNAS*, 2013; 110(30): 12450–55.

[49]. Carte J, Wang R, Li H, Terns RM, Terns MP. Cas6 is an endoribonuclease that generates

guide RNAs for invader defense in prokaryotes. *Genes Dev,* 2008; 22(24): 3489–96.

[50]. Hatoum-Aslan A, Maniv I, Marraffini LA. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor-processing site. *PNAS*, 2011; 108(52): 21218–22.

[51]. Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, Graham S, Reimann J, Cannone G, Liu H, Albers SV, Naismith JH, Spagnolo L, White MF. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell*, 2012; 45(3): 303–13.

[52]. Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover CV 3rd, Graveley BR, Terns RM, Terns MP. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol Cell*, 2012; 45(3): 292–302.

[53]. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzade ZA, Eckert MR, Vogel J, Charpentier E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 2011; 471(7340): 602–07.

[54]. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012; 337(6096): 816–21.

[55]. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS*, 2012; 109(39): E2579–86.

[56]. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 2014; 343(6176): 1247997.

[57]. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of cas9 in complex with guide RNA and target DNA. *Cell*, 2014; 156(5): 935–49.

[58]. Quax TE, Wolf YI, Koehorst JJ, Wurtzel O, van der Oost R, Ran W, Blombach F, Makarova KS, Brouns SJ, Forster AC, Wagner EG, Sorek R, Koonin EV, van der Oost J. Differential translation tunes uneven production of operon-encoded proteins. *Cell Rep*, 2013; 4(5): 938–44.

[59]. Nam KH, Haitjema C, Liu X, Ding F, Wang H, DeLisa MP, Ke A. Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR–Cas system. *Structure*, 2012; 20(9): 1574–84.

[60]. Wiedenhefta B, Duijnc EV, Bultemad JB, Waghmaree SP, Zhoua K, Barendregtc A, Westphalb W, Heckc AJR. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *PNAS*, 2011; 108: 10092–7.

[61]. Spilman M, Cocozaki A, Hale C, Shao Y, Ramia N, Terns R, Terns M, Li H, Stagg S. Structure of an RNA silencing complex of the CRISPR–Cas immune system. *Mol. Cell*, 2013; 52(1): 146–52.

[62]. Staals RHJ, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, Duijn EV, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K, Masuda A, Dohmae N, Schaap PJ, Doudna JA, Heck AJ, Yonekura K, van der Oost J, Shinkai A. Structure and activity of the RNA-targeting Type III-B CRISPR–Cas complex of Thermus thermophilus. *Mol Cell*, 2013; 52(1): 135–45.

[63]. Makarova KS, Wolf YI, Koonin EV. The basic building blocks and evolution of CRISPR–Cas systems. *Biochem. Soc. Trans*, 2013; 41(6): 1392–1400.

[64]. Westra ER1, Swarts DC, Staals RH, Jore MM, Brouns SJ, van der Oost J. The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu Rev Genet*, 2012; 46: 311–39.

[65]. Sashital DG, Wiedenheft B, Doudna JA. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol Cell,* 2012; 46(5): 606–15.

[66]. Sorek R, Lawrence CM. Wiedenheft B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem*, 2013; 82: 237–66.

[67]. Jore MM, Lundgren M, Van Duijn E, Bultema JB, Westra ER, Waghmare SP, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat struct mol biol*, 2011; 1; 18(5): 529-36.

[68]. Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, Doudna JA, Nogales E. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*, 2011; 477(7365): 486–489.

[69]. Hochstrasser ML, Taylor DW, Bhat P, Guegler CK, Sternberg SH, Nogales E, Doudna JA. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *PNAS*, 2014; 111(18): 6618–23.

[70]. Edze R, Westra ER, Nilges B, Erp PBGV, van der Oost J, Dame RT, Brouns SJJ. Cascade-mediated binding and bending of negatively supercoiled DNA. *RNA Bio,* 2012; 9: 1134–8.

[71]. Westra ER, van Erp PB, Künne T, Wong SP, Staals RH, Seegers CL, Bollen S, Jore MM, Semenova E, Severinov K, de Vos WM, Dame RT, de Vries R, Brouns SJ, van der Oost J. CRISPR immunity relies on the consecutive binding and degradation of negatively

supercoiled invader DNA by Cascade and Cas3. *Mol cell*, 2012; 8; 46(5): 595-605.

[72]. Sinkunas T, Gasiunas G, Waghmare SP, Dickman MJ, Rodolphe Barrangou R, Horvath P,and Siksnys V. *In vitro* reconstitution of Cascade-mediated   CRISPR   immunity   in Streptococcus thermophilus. *EMBO J*, 2013; 32(3): 385–94.

[73]. Beloglazova N, Petit P, Flick R, Brown G, Savchenko A, Yakunin AF. Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J,* 2011; 30(22): 4616–27.

[74]. Van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat Rev Microbial*, 2014; 12(7): 479–92.

[75]. Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. Cas3 is a single-stranded   DNA   nuclease   and   ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J*, 2011; 30(7): 1335–42.

[76]. Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 2010; 468(7320): 67–71.

[77]. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR   RNA-guided   endonuclease   Cas9. *Nature,* 2014; 507(7490): 62–7.

[78]. Hatoum-Aslan A, Samai P, Maniv I, Jiang W, Marraffini LA. A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *J Biol Chem,* 2013; 288(39): 27888–97.

[79]. Rouillon C, Zhou M, Zhang J, Politis A, Beilsten-Edmands V, Cannone G, Graham S, Robinson CV, Spagnolo L, White MF. Structure of the CRISPR interference complex CSM reveals key similarities with Cascade. *Mol Cell*, 2013; 52(1): 124–34.

[80]. Marraffini LA, Sontheimer EJ. Self-versus non-self-discrimination during CRISPR RNA-directed immunity. *Nature*, 2010; 463(7280): 568–71.

[81]. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP. RNA-guided RNA cleavage by a CRISPR RNA–Cas protein complex. *Cell*, 2009; 139(5): 945–56.

[82]. Deng L, Garrett RA, Shah SA, Peng X, She Q. A novel interference mechanism by a type IIIB CRISPR–Cmr module in Sulfolobus. *Mol Microbiol*, 2013; 87(5): 1088–99.

[83]. Sheng G, Zhao H, Wang J, Rao Y, Tian W, Swarts DC, van der Oost J, Patel DJ, Wang Y. Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-

mediated DNA target cleavage. *PNAS*, 2014; 111(2): 652–57.

[84]. Sander JD, Joung JK. CRISPR-Cas systems for editting, regulating and targeting genomes. *Nat Biotechnol*, 2014; 32(4): 347- 55.

[85]. Gaj T, Gersbach CA, Barbas III CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol,* 2013; 31(7): 397- 405.

[86]. Zhao Y, Ying Y, Wang Y. Developing CRISPR/Cas9 Technologies for Research and Medicine. *MOJ Cell Sci Report*, 2014; 1(1): 00006.

[87]. Hsu PD, Landerand E.S, Zhang F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell,* 2014, 157(6): 1262- 78.

[88]. Wang H, La Russa M, Qi LS. CRISPR/Cas9 in genome editing and beyond. *Annu rev biochem*, 2016; 85: 227-64.

[89]. Barrangou R, Doudna J.A. Applications of CRISPRtechnologies in research and beyond. *Nat biotechnol*, 2016; 34 (9): 933-41.

[90]. Maves L. Recent advances using zebrafish animal models for muscle disease drug discovery. *Export Opin Drug Discov*, 2014; 9(9): 1033-45.