Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems

Citation: Mohanraju, Prarthana et al. “Diverse Evolutionary Roots and Mechanistic Variations of the CRISPR-Cas Systems.” Science 353, 6299 (August 2016): aad5147 © 2016 American Association for the Advancement of Science

As Published: http://dx.doi.org/10.1126/science.aad5147

Publisher: American Association for the Advancement of Science (AAAS)

Persistent URL: http://hdl.handle.net/1721.1/113195

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

[Summary page]

Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems

Prarthana Mohanraju[1], Kira S. Makarova[2], Bernd Zetsche[3],
Feng Zhang[3], Eugene V. Koonin[2], John van der Oost[1*]

Background: Prokaryotes have evolved multiple systems to combat invaders such as viruses and plasmids. Examples of such defence systems include receptor masking, restriction-modification (R-M systems), DNA interference (Argonaute), bacteriophage exclusion (BREX or PGL) and abortive infection, all of which act in an innate, non-specific manner. In addition, prokaryotes have evolved adaptive, heritable immune systems, i.e. clustered regularly interspaced palindromic repeats (CRISPR) and the CRISPR-associated proteins (CRISPR-Cas). Adaptive immunity is conferred by the integration of DNA sequences from an invading element into the CRISPR array (adaptation), which is transcribed into long pre-CRISPR (pre-cr) RNAs and processed into short crRNAs (expression), which guide Cas proteins to specifically degrade the cognate DNA on subsequent exposures (interference).

Advances: A plethora of distinct CRISPR-Cas systems are represented in genomes of most archaea and almost half of the bacteria. The latest CRISPR-Cas classification scheme delineates two classes that are each subdivided into three types. Integration of biochemistry and molecular genetics has contributed significantly to revealing many of the unique features of the variant CRISPR-Cas types. Additionally, structural analysis and single molecule studies have further advanced our understanding of the molecular basis of CRISPR-Cas functionality. Recent progress includes relevant steps in the adaptation stage, when fragments of foreign DNA are processed and incorporated as new spacers into the CRISPR array. In addition, three novel CRISPR-Cas types (IV, V, and VI) have been identified, and in particular, the type V interference complexes have been experimentally characterized. Moreover, the ability to easily program sequence-specific DNA targeting and cleavage by CRISPR-Cas components, as demonstrated for Cas9 and Cpf1, allows for the application of CRISPR–Cas components as highly effective tools for genetic engineering and gene regulation in a wide range of eukaryotes and prokaryotes.The pressing issue of off-target cleavage by the Cas9 nuclease is being actively addressed using structure-guided engineering.

Outlook: Although our understanding of the CRISPR-Cas system has increased tremendously over the past few years, much remains to be done. About the evolution of CRISPR-Cas systems, the continuing discovery of novel CRISPR-Cas variants will provide direct tests of the recently proposed modular scenario. The recent discovery and characterization of new CRISPR-Cas types with many unique features implies that our current knowledge has relatively limited power for predicting the functional details of distantly related variants. Hence, newly discovered CRISPR-Cas systems need to be thoroughly dissected with the aforementioned multi-disciplinary approaches to gain insight in their biological role, to unravel their molecular mechanism, and to harness their potential for biotechnology. As to the biology, key outstanding questions include the ecological roles of microbial adaptive immunity, the high rates of its horizontal transfer, and the co-evolution of CRISPR-Cas and phage-encoded anti-CRISPR proteins. Relatively little is known about the regulation of the CRISPR-Cas expression, and about the roles of CRISPR-Cas in processes other than defence. With respect to CRISPR-Cas mechanism, details on the connection between the adaptation stage and the interference stage in primed spacer acquisition remain elusive. A key aspect of CRISPR-Cas that is poorly understood at present concerns the mechanism(s) of self/non-self discrimination. Preliminary data show that these mechanisms differ substantially among CRISPR variants, and more research is required to obtain the full picture. As previously found for distinct types of Class 1 effector complexes (Cascade / Cmr), recent comparison of Class 2 types (Cas9 / Cpf1) has revealed, along with the overall architectural similarity, significant structural and mechanistic differences. It may well be that these

variations eventually could translate into complementary applications. Apart from innovative tools for basic research, they CRISPR-associated effector complexes will be instrumental for developing next generation of antiviral prophylactics and therapeutics. For applications in human gene therapy, improved methods for efficient and safe delivery of Cas9/Cpf1 and their guide RNAs to cells and tissues are still desired. In conclusion, further unravelling of the fundamentals of CRISPR-Cas structure, functions and biology, and characterization of new Cas effector proteins in particular, is crucial for optimizing and further expanding the diverse applications of CRISPR-Cas systems.

[[1] Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, Netherlands

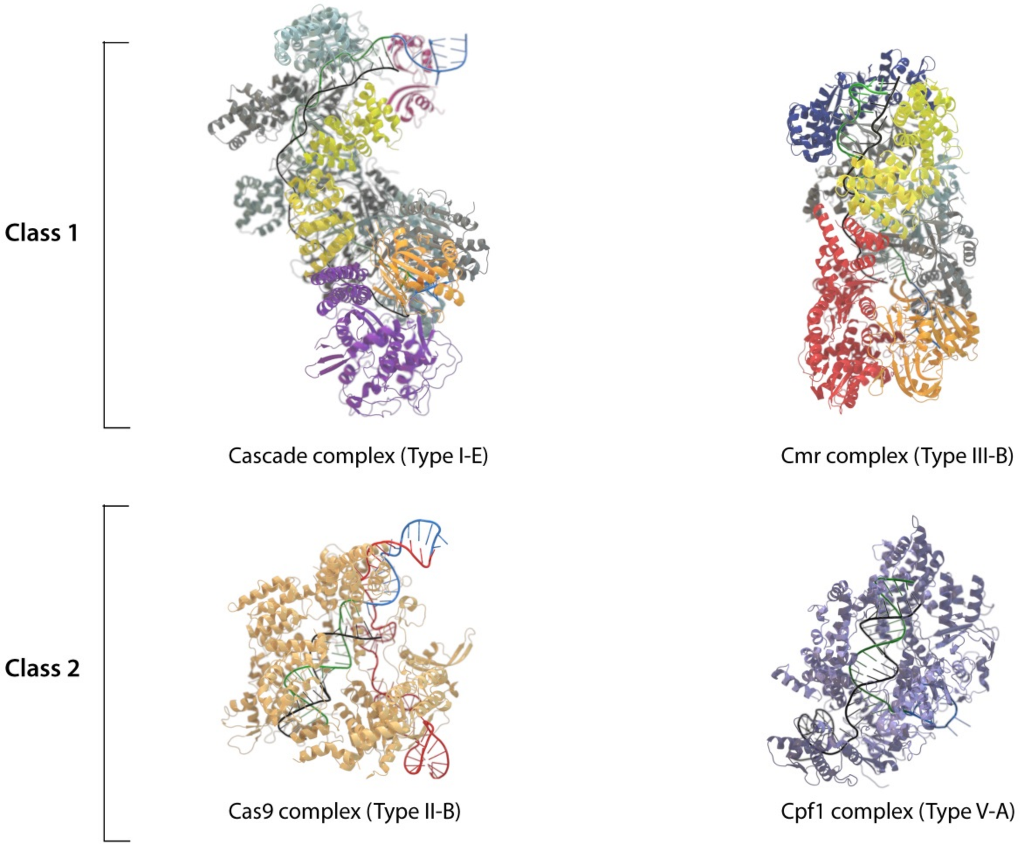
[2] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

[3] Broad Institute of MIT and Harvard, Cambridge, MA 02142

*To whom correspondence should be addressed: john.vanderoost@wur.nl]

[Figure title]: **Evolution of CRISPR-Cas systems resulted in incredible structural and functional diversity**

[Figure caption]: Class 1 CRISPR-Cas systems are considered as the ancestral systems. The Class 2 systems, evolved from Class 1 via insertion of transposable elements encoding various nucleases, and are nowadays being used as novel tools for genome editing



Title:

Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems

Abstract:

Adaptive immunity had been long thought of as an exclusive feature of animals. However, the discovery of the CRISPR-Cas defense system, present in almost half of prokaryotic genomes, proves otherwise. Due to the everlasting parasite-host arms race, CRISPR-Cas has rapidly evolved through horizontal transfer of complete loci or individual modules, resulting in extreme structural and functional diversity. CRISPR-Cas systems are divided into two distinct classes that each consist of three types and multiple subtypes. In this Review, we discuss recent advances in CRISPR-Cas research that reveal elaborate molecular mechanisms and provide a plausible scenario of CRISPR-Cas evolution. We also describe the latest developments of a wide range of CRISPR-based applications.

Main Text:

Bacteria and archaea suffer constant predation by viruses which are extremely abundant in almost all environments [1]. Accordingly, bacteria and archaea have evolved a wide range of antivirus defense mechanisms [2]. As viruses generally have high rates of mutation and recombination, they have the potential to rapidly escape these prokaryotic defense systems. Thus, the hosts' defenses must also adjust and evolve rapidly, leading to an on-going virus-host arms race. Protective systems provide innate immunity at all stages of the parasite's infection cycle, via receptor masking, restriction–modification (R–M systems), DNA interference (Argonaute), bacteriophage exclusion (BREX or PGL) and abortive infection [2-8].

The innate immunity strategies are complemented by an adaptive immune function of the prokaryotic clustered regularly interspaced short palindromic repeats (CRISPR) and the associated Cas proteins [9, 10] system. Diverse variants of the CRISPR-Cas systems are present in the examined genomes of most archaea and almost half of the bacteria [2]. Here we discuss insights into the evolution and functionality of Class 1 and Class 2 CRISPR-Cas. This progress has enabled the development of sophisticated tools for genetic engineering in molecular biology, biotechnology and molecular medicine.

CRISPR-Cas defence

The CRISPR-Cas systems provide protection against mobile genetic elements (MGEs), in particular viruses and plasmids, by sequence-specific targeting of foreign DNA or RNA [9, 11-15]. A CRISPR-*cas* locus generally consist of an operon of CRISPR-associated (*cas*) genes and a CRISPR array composed of a series of direct repeats interspaced by variable DNA sequences (known as 'spacers') (Figure 1a). The repeat sequences and lengths as well as the number of repeats in CRISPR arrays vary broadly, but all arrays possess the characteristic arrangement of alternating repeat and spacer sequences. The spacers are key elements of adaptive immunity as they store the "memory" of an organism's encounters with specific MGEs acquired as a result of a previous unsuccessful infection [16-19]. This memory enables the recognition and neutralization of the invaders upon subsequent infections [9].

CRISPR-mediated adaptive immunity involves three steps: adaptation, expression, and interference [14, 20-23] (Figure 1b). During the adaptation step, fragments of foreign DNA (known as protospacers) from invading elements are processed and incorporated as new spacers into the CRISPR array. The expression step involves the transcription of the CRISPR array, which is followed by processing of the precursor transcript into mature CRISPR RNAs (crRNAs). The crRNAs are assembled with one or more Cas proteins into CRISPR ribonucleoprotein (crRNP) complexes. The interference step involves crRNA-directed cleavage of invading cognate virus or plasmid nucleic acids by Cas nucleases within the crRNP complex [14, 20, 24]. The multifaceted and modular architecture of the CRISPR-Cas systems also allows it to play non-defense roles, ranging from biofilm formation and cell differentiation to virulence [25-27].

CRISPR–Cas diversity, classification and evolution

The rapid evolution of highly diverse CRISPR-Cas systems is thought to be driven by the continuous arms race with the invading MGEs [28, 29]. The latest classification scheme for CRISPR-Cas systems takes into account the repertoire of *cas* genes, the sequence similarity between Cas proteins and the locus architecture, and includes two classes that, at present, are subdivided into 6 types and 19 subtypes [30, 31]. The key feature of the organization and evolution of the CRISPR-Cas loci is their pronounced modularity. The module responsible for the adaptation step is largely uniform among the diverse CRISPR-Cas systems and consists of the *cas*1 and *cas*2 genes, both of which are essential for the acquisition of spacers. In many CRISPR-Cas variants, the adaptation module also includes the *cas*4 gene. By contrast, the CRISPR-Cas effector module which is involved in the maturation of the crRNAs as well as in target recognition and cleavage, shows a far greater versatility (Figure 2a) [30].

The two classes of CRISPR-Cas systems differ fundamentally with respect to the organization of the effector module [30]. Class 1 systems (including Types I, III and IV) are present in bacteria and archaea, and encompass effector complexes composed of four to seven Cas protein subunits in an uneven stoichiometry (e.g. the CRISPR-associated complex for antiviral defense [Cascade] of Type I systems, and the Csm/Cmr complexes of Type III systems). Most of the subunits of the Class 1 effector complexes, in particular Cas5, Cas6 and Cas7, contain variants of the RNA-binding RRM (RNA Recognition Motif) domain [32, 33]. Although the sequence similarity between the individual subunits of Type I and Type III effector complexes is generally low, the complexes share strikingly similar overall architectures, suggestive of a common origin [31, 32, 34, 35]. The ancestral CRISPR-Cas effector complex most likely resembled the extant Type III complexes as indicated particularly by the presence of the archetypal Type III protein, the large Cas10 subunit, which appears to be an active enzyme of the DNA polymerase-nucleotide cyclase superfamily, unlike its inactive Type I counterpart (Cas8) [31-33]

In the less common Class 2 CRISPR-Cas systems (Types II, V and VI) which are almost completely restricted to bacteria, the effector complex is represented by a single, multi-domain protein [30]. The best characterized Class 2 effector is Cas9 (Type II), the RNA-dependent endonuclease that contains two unrelated nuclease domains, HNH and RuvC, that are responsible for the cleavage of the target and the displaced strand, respectively, in the crRNA-target DNA complex [36]. The Type II loci also encode a trans-acting CRISPR (tracr)RNA that evolved from the corresponding CRISPR repeat and is essential for pre-crRNA processing and target recognition in Type II systems [37, 38]. The prototype Type V effector, Cpf1 (subtype V-A), contains only one nuclease domain (RuvC-like) that is identifiable by sequence analysis [39]. However, analysis of the recently solved structure of Cpf1 complexed with the crRNA and target DNA has revealed a second nuclease domain, the fold of which is unrelated to HNH or any other known nucleases. In analogy to the HNH domain in Cas9, the novel nuclease domain in Cpf1 is inserted into the RuvC domain, and it is responsible for cleavage of the target strand [Yamano et al 2016, in press].

Screening of microbial genomes and metagenomes for potential new Class 2 systems [31] has resulted in the identification of three novel CRISPR-Cas variants. These include subtypes V-B and V-C, which resemble Cpf1 in that their predicted effector proteins contain a single, RuvC-like nuclease domain. Cleavage of target DNA by the Type V-B effector, denoted C2c1, has been experimentally demonstrated [31]. Type VI is unique in that its putative effector protein contains two HEPN domains with predicted RNase activity (Figure 2a).

Recent comparative genomic analyses of variant CRISPR-Cas systems (Figure 2b) [31] has revealed a strong modular evolution with multiple combinations of adaptation modules and effector modules, as well as a pivotal contribution of mobile genetic elements to the origin and diversification of the CRISPR-Cas systems. The ancestral prokaryotic adaptive immune system could have emerged via the insertion of a casposon (a recently discovered distinct class of self-synthesizing transposons that appear to encode a Cas1 homolog) next to an innate immunity locus (probably consisting of genes encoding a Cas10 nuclease and possibly one or more RNA binding proteins). Apart from providing the Cas1 nuclease/integrase that is required for recombination during spacer acquisition [40-42], the casposon may also have contributed the prototype CRISPR repeat unit that could have evolved from one of the inverted terminal repeats of the casposon [43]. An additional toxin-antitoxin module that inserted either in the ancestral casposon or in the evolving adaptive immunity locus probably provided the *cas*2 gene, thus completing the adaptation module. The Cas10 nuclease and one or more additional proteins with an RRM fold (the ultimate origin of which could be a polymerase/cyclase that gave rise to Cas10) of the hybrid locus could have subsequently evolved to become the ancestral CRISPR-Cas effector module [31-33, 43].

The widespread occurrence of Class 1 systems in archaea and bacteria, together with the proliferation of the ancient RRM domain in Class 1 effector proteins, strongly suggests that the ancestral CRISPR-Cas belonged to Class 1. Most likely, the multiple Class 2 variants then evolved via several independent replacements of the Class 1 effector locus with nuclease genes that were derived from distinct MGEs (Figure 2b). In particular, Type V effector variants (Cpf1) seem to have evolved from different families of the TnpB transposase genes that are widespread in transposons [31], whereas the Type II effector (Cas9) may have evolved from IscB, a protein with two nuclease domains that belongs to a recently identified distinct transposon family [44]. Notably, Class 2 CRISPR-Cas systems, in their entirety, appear to have been derived from different MGEs: Cas1 from a casposon, Cas2 from a toxin-antitoxin module, and the different effector proteins, such as Cas9 or Cpf1, from respective transposable elements [31].

CRISPR adaptation

The spacers of a CRISPR array represent a chronological archive of previous invader encounters. The captured spacer sequences are integrated into the CRISPR loci after exposure to MGEs, at the leader end of the array that contains the start site of CRISPR transcription [9, 14, 45]. Analysis of invader target sequences (also called protospacers) has revealed a short motif directly adjacent to the target sequence called the protospacer adjacent motif (PAM) [46]. This PAM motif allows self/non-self discrimination by the host, on the one hand, because its presence in alien targets is required for non-self interference, and on the other hand, because its absence in the host's CRISPR array avoids self-targeting [47]. In Class 1-Type I and Class 2-Type II systems, the PAM is not only involved in interference, but also plays a role in spacer selection during the adaptation stage, implying the acquisition of functional spacers only [48, 49]. The PAM is a short (2 to 7 nucleotides), partially redundant sequence that in itself cannot preclude incorporation of spacers from the host DNA due to the low information content of the motif. The short PAM appears to be the result of an evolutionary trade-off between efficient incorporation of spacers from non-self DNA and preventing an autoimmune reaction.

Although host chromosomal fragments can be incorporated as new CRISPR spacers; detection of such events obviously implies that this did not result in a lethal phenotype, either due to a modified PAM and/or to an inactivated CRISPR-Cas effector module [50]. Indeed, in the absence of the effector module, elevated frequencies of self-spacer acquisition occurred in *E. coli* [51]. Similarly, *Streptococcus thermophilus* with a catalytically inactive Cas9 resulted in a major increase of spacers derived from the host genome [52]. In addition, there is a strong preference for the integration of plasmid over chromosomal spacer sequences [51, 53, 54] with plasmid sequences incorporated 100-1000 times more frequently than host DNA [55]. Spacer acquisition in *E. coli* requires active replication of the protospacer-containing DNA [55]. Thus, small, fast replicating plasmid genomes are a much better source of spacers than the large host DNA, findings consistent with acquisition of spacers from an infecting virus genome in the archaeon *Sulfolobus islandicus* requiring its active replication [56]. In *E. coli*, the CRISPR-Cas system derives the spacers primarily from products of RecBCD-catalyzed DNA degradation that are formed during the repair of double-stranded breaks associated with stalled replication forks [57]. Most likely, this. Other possible sources of substrates for CRISPR adaptation include DNA fragments generated either by other defense systems, such as the restriction-modification system [58], or by the CRISPR-Cas system itself [48].

Cas1 and Cas2 play crucial roles in spacer acquisition in all CRISPR-Cas systems [49, 51]. In addition, these proteins can function *in trans* provided the repeats involved are sufficiently similar in size and structure. Accordingly, *cas1* and *cas2* genes are missing in many active CRISPR-Cas loci, in particular of Type III as well as Types IV and VI [30]. Overexpression of Cas1 and Cas2 from the *E. coli* Type I-E system has been shown to be sufficient for the extension of the CRISPR array [51]. Mutations in the active site of Cas1 abolish spacer integration in *E. coli* [51], whereas the nuclease activity of Cas2 is dispensable [54]. In *E. coli*, a complex of a central Cas2 dimer and two flanking Cas1 dimers binds and processes PAM-containing DNA fragments (Figure 3a) [54, 59], after which the newly generated spacers can be integrated into a CRISPR array via a recombination mechanism akin to that of retroviral integrases and transposases [60] (Figure 3b).

In several Type III CRISPR-Cas systems, Cas1 is fused to reverse transcriptase (RT) [20], and recently, it has been shown that these systems are capable of acquisition of RNA spacers by direct incorporation of an RNA segment into the CRISPR array followed by reverse transcription and replacement of the RNA strand by DNA [61]. The biological function of this process remains to be elucidated but regardless, these findings demonstrate remarkable versatility of adaptation pathways.

Spacer acquisition (adaptation) in Type I systems proceeds along two distinct paths: *naïve* acquisition which occurs during an initial infection, and *primed* acquisition when the CRISPR contains a previously integrated spacer that is complementary to the invading DNA [62]. According to the proposed model, naïve spacer adaptation involves five steps (numbered 1-5 in Figure 3b). (1) Fragmentation of (mainly) invasive nucleic acids by non-Cas systems (e.g. by RecBCD after stalling replication fork, or by restriction enzymes (R-M)) [55, 58]), or by CRISPR-associated nucleases [48]. Although this step may be non-essential, it probably enhances the efficiency of the overall process and its specificity toward invading DNA. (2) Selection of DNA fragments for (proto)spacers by scanning for potential PAMs (after partial target unwinding) by one of the four Cas1 subunits of the Cas1-Cas2 complex [63]. (3) Measuring of the selected protospacer generating fragments of the correct size with 3' hydroxyl groups by Cas1 nuclease. (4) Nicking of both strands of the leader-proximal repeat of the CRISPR array at the 5' ends through a direct nucleophilic attack by the generated 3' OH-groups, resulting in covalent links of each of the strands of the newly selected spacer to the single-stranded repeat ends. (5) Second strand synthesis and ligation of the repeat flanks, by a non-Cas repair system [45, 60].

Primed spacer adaptation so far has been demonstrated only in Type I systems [49, 64, 65]. This priming mechanism constitutes a positive-feedback loop that facilitates the acquisition of new spacers from formerly encountered genetic elements [66]. Priming can occur even with spacers that contain several mismatches, making them incompetent as guides for targeting the cognate foreign DNA [66]. Based on PAM selection, functional spacers are preferentially acquired during naïve adaptation. This initial acquisition event triggers a rapid priming response after subsequent infections. Priming appears to be a major pathway of CRISPR adaptation, at least for some Type I systems [64]. Primed adaptation strongly depends on the spacer sequence [67], and the acquisition efficiency is highest in close proximity to the priming site. In addition, the orientation of newly inserted spacers indicates a strand bias, which is consistent with the involvement of single-stranded adaption intermediates [68]. According to one proposed model [69], replication forks in the invader's DNA are blocked by the Cascade complex bound to the priming crRNA, enabling the RecG helicase and the Cas3 helicase/nuclease proteins to attack the DNA. The ends at the collapsed forks then could be targeted by RecBCD which provides DNA fragments for new spacer generation [69]. Given that the use of crRNA for priming has much less strict sequence requirements than direct targeting of the invading DNA, priming is a powerful strategy that might have evolved in the course of the host-parasite arms race to reduce the escape by viral mutants, to provide robust resistance against invading DNA, as well as to enhance self-non-self-discrimination. Naive as well as primed adaptation in subtype I-F system of *P. aeruginosa* CRISPR-Cas require both the adaptation and the effector module [68]

In the Type II-A system, the Cas9-tracrRNA complex and Csn2 are involved in spacer acquisition along with the Cas1-Cas2 complex [52, 70]; the involvement of Cas9 in adaptation is likely to be a general feature of Type II systems. Although the key residues of Cas9 involved in PAM recognition are dispensable for spacer acquisition, they are essential for the incorporation of new spacers with the correct PAM sequence [70]. The involvement of Cas9 in PAM recognition and protospacer selection [70] suggests that in Type II systems Cas1 may have lost this role. Similarly, Cas4 that is present in subtypes IA-D and II-B has been proposed to be involved in the CRISPR adaptation process, and this prediction has been validated experimentally for Type I-B [64]. Cas4 is absent in subtype II-C system of *Camplylobacter jejuni*. Nonetheless, a conserved Cas4-like protein found *in Camplyobacter* bacteriophages can activate spacer acquisition to use host DNA as an effective decoy to bacteriophage DNA. Bacteria that acquire self-spacers and escape phage infection have to overcome CRISPR-mediated autoimmunity either by loss of the interference functions, leaving them susceptible to foreign DNA invasions, or tolerate changes in gene regulation [71]. Furthermore, in subtypes I-U and V-B, Cas4 is fused to Cas1 which implies cooperation between these proteins during adaptation. In Type I-F systems, Cas2 is fused to Cas3 [19] suggestive of a dual role for Cas3 [20], i.e. involvement in adaptation, in addition to its role in interference. These findings support the coupling between the adaptation and interference stages of CRISPR-Cas defense during priming.

Biogenesis of crRNAs

The short mature crRNAs contain spacer sequences that are the guides that are responsible for the specificity of CRISPR-Cas immunity [12]. They associate with one or more Cas proteins to form effector complexes that target invading MGEs through crRNA:target sequence-specific recognition. The CRISPR arrays are transcribed as long precursors, known as pre-crRNA, that may contain secondary structured elements (hairpins) in those cases where the CRISPR contains palindromic repeats. The processing of the pre-crRNA typically yields 30-65 nt mature crRNAs that consist of a single spacer flanked by a partial repeat at either one or both ends [12, 72].

The pathways of crRNA biogenesis differ between the different CRISPR-Cas types. In Class 1 systems, the Cas6 protein is critical for the primary processing of pre-crRNA. Cas6 is a metal-independent endoribonuclease that recognizes and cleaves a single phosphodiester bond in the repeat sequences of a pre-crRNA transcript [12, 73, 74]. Members of the Cas6 family contain two RRM-type RNA-binding domains. The primary cleavage by Cas6 results in crRNAs containing a repeat-derived 5' "handle" of 8 nucleotides with a 5' hydroxyl group, followed by the complete spacer sequence and a repeat-derived 3' handle of variable size that in some subtypes forms a hairpin structure with a 3' phosphate or a cyclic 2'–' phosphate [12, 73, 75]. The Cas6 family proteins show considerable structural variation that might reflect the cleavage specificity [72, 76, 77].

In Type I-E and I-F systems, the Cas6 ribonuclease is a single-turnover enzyme that remains attached to the crRNA cleavage product. In these cases, Cas6 is a subunit of a multi-subunit Cascade complex [12, 78] (Figure 4a). In the Type I-F systems, the crRNP complex consists of the crRNA, Cas6f and Csy1, Csy2 and Csy3 proteins [79-81]. In other systems (subtypes I-A, I-B, I-D and III-A-D), Cas6 is not associated with the crRNA-processing complex. The absence of a Cas6 subunit in the complex correlates with the lack of a hairpin structure of the 3' handle and a variable 3' end. The absence of a *cas*6 gene in Type I-C is complemented by another double RRM-fold subunit, Cas5d, which has adopted the role of the endoribonuclease that in other subtypes is carried out by Cas6 [82]. Some systems co-existing in the same species have been demonstrated to share the same set of guides, e.g. Type III-A (Csm) and Type III-B (Cmr) of *Thermus thermophilus* [83], Type III-B (Cmr), Type I-A (Csa) and Type I-G (Cst) of *Pyrococcus furiosus* [84]. Given that the Type III loci usually lack *cas*6 genes, a single stand-alone Cas6 nuclease is likely to be responsible for the supply of crRNAs to the Type III complexes in *T. thermophilus* [83]. In *P. furiosus,* Cas6 nuclease of Type I generates the crRNAs from all CRISPR loci for the different co-existing complexes [84]. Cas6–based processing of pre-crRNA in Type III systems is typically followed by a sequence-unspecific trimming at the 3' end (by yet to be identified RNases) to yield mature crRNAs with a defined 8-nt 5' end and a variable 3' end [34, 85, 86].

Type II systems employ a unique mechanism for crRNA biogenesis whereby processing depends on Cas9, a host RNase-III and the tracrRNA that base pairs with the repeats of the pre-crRNA [36, 37, 72] (Figure 4b). The cleaved crRNA–tracrRNA hybrid is bound and stabilized by Cas9, triggering a conformational change towards a state compatible with target scanning, recognition and interference [36, 37, 87]. Trimming of the 5' end of the crRNA probably occurs by a non-Cas RNase. The absence of Type II systems in archaea is consistent with the absence of RNase-III genes in most archaeal genomes [88]. In the Type II-C system of *Neiseria meningitidis*, short intermediate crRNA guides are transcribed from multiple promoters embedded within the repeats of the CRISPR array, implying that the system does not require RNase-III [89]. (Figure 4c). Expression of tracrRNA has also been demonstrated for the subtype V-B system, suggestive of a crRNA processing pathway analogous to that in Type II. By contrast, in subtype V-A and Type VI systems, no tracrRNA is co-expressed with the pre-crRNA [31, 39]. Class 2 CRISPR-Cas systems lacking tracrRNA can be expected to function using novel mechanisms of crRNA biogenesis, including processing by other host RNases or by the effector proteins (such as Cpf1) themselves.

A third variant of guide maturation has recently been described for the Cpf1 effector complex, a Class 2 system that (unlike Cas9) does not associate with a tracrRNA. It has been demonstrated that Cpf1 has an intrinsic RNase activity that allows for the primary processing of the pre-crRNA to crRNA guides with a 5' hairpin [90][Zetsche & Mohanraju 2016]. The biosynthesis of crRNAs by Cpf1 system is metal-, sequence- and structure- dependant [90]. Secondary processing of CRISPR guides probably occurs via a non-Cas ribonuclease; maturation of Cas9-associated guides occurs by trimming at the 5' end (Figure 4B) whereas in Cpf1 the 3' flanks of the crRNA are removed (Figure 6).

Target interference

Selection of CRISPR-Cas targets is a step-wise process that relies on recognition of a non-self sequence, a complementary spacer of which is stored in the CRISPR locus. In most cases, with the exception of the RNA targeting Type III systems, cognate protospacer sequences flanked by a PAM sequence are recognized by a CRISPR ribonucleoprotein (crRNP) complex (Type I Cascade, Type II Cas9, Type V Cpf1 (Figure 5)) and specifically degraded [12, 14, 39]. In addition, selection of an appropriate target sequence depends on a so-called seed sequence on the guide [78, 91]. The seed is a 7-8 bp sequence in close proximity to the PAM. Matching PAM and seed sequences are crucial for target interference [78, 91, 92] and act as a quality control step that is required for the complete displacement of the non-complementary strand of the target DNA by the crRNA guide, the so-called R-loop conformation. Downstream of the seed region, mismatches between spacer and protospacer are tolerated to some extent (see below) [91].

In Type I systems, the Cascade RNP complex scans DNA for complementary target sites, initially by identifying an appropriate PAM motif, followed by partial melting and base pairing by the guide's seed sequence, and eventually by formation of a complete R-loop structure [75, 93]. Upon reaching a PAM-proximal mismatch, the R-loop propagation stalls and the interference is aborted [94]. When base pairing between guide and protospacer is complete, the R-loop structure appears to be locked in a state to license DNA degradation by the Cas3 nuclease/helicase [12, 19, 94].

Single-molecule experiments with *E. coli* Cascade demonstrate that crRNA-guided Cascade exhibits two distinct binding modes for matching and mismatched targets, which trigger either interference (matching target) or primed spacer acquisition (mismatched target). Unlike the interference of matching targets, mismatched targets are recognized with low fidelity, as indicated by a short-lived binding. The latter association is PAM- and seed-independent and can involve base pairing by any part of the crRNA spacer. In this case, the Cascade complex does not adopt a conformation that allows docking of Cas3 [95], precluding DNA interference. Instead, this Cascade-target complex primes the formation of a spacer acquisition complex that consists of Cas3 and Cas1-Cas2, and generates DNA fragments that are integrated as new spacers in the CRISPR array [93]. These dual roles of Cascade allow for efficient degradation of bona fide targets and priming the acquisition of new spacers from mismatched targets (e.g. from viral escape mutants) as an update of the CRISPR memory [95].

Although Type III systems are structurally related to the Type I system (Figure 5) [34, 35, 59, 96-100], they show some substantial mechanistic variations. Initial analyses indicated that Csm (III-A) complexes target DNA [13] whereas Cmr (III-B) complexes target RNA [11, 101, 102]. However, it has recently been demonstrated that both Type III complexes are transcription-dependent DNA nucleases [83, 103], i.e. they initially recognize their target through specific interaction of the crRNA guide with a complementary nascent mRNA, after which cleavage occurs of the flanking DNA sequences [104-109]. Robust interference by these systems relies on the concerted cleavage of the transcript RNA and the transcribed DNA. The Cas7-like backbone subunits (Csm3, Cmr4) are responsible for the ribonuclease activity, typically resulting in cleavage of the target RNA at 6 nucleotide intervals [83, 98, 102, 103, 110-112]. Binding of the Cmr complex to its complementary RNA target induces a conformational change [35, 98] that results in activation of the Cas10 DNA-cleaving subunit (Csm1/Cmr2) [105, 106, 108]. Disruption of the ribonucleases active sites (in Csm3/Cmr4), at least in some cases, does not hamper the activation of the DNA nuclease activity of the complexes [103, 105]. Exonucleolytic cleavage of single-stranded DNA and RNA by recombinant *S. epidermidis* Csm1 (Cas10), *T. maritima* and *P. furiosus* Cmr2 has been demonstrated *in vitro* [105, 106, 113]. In the *S. epidermidis* system, a Csx1 ortholog (Csm6) provides an auxiliary RNA-targeting activity that operates in conjunction with the RNA- and DNA-targeting endonuclease activities of the Csm effector complex [114-116]; in the *P. furiosus* Cmr system Csx1 appears not to be an essential component [103]. The relative contribution of the different nuclease subunits appears to vary in the different Type III systems and under different conditions, and awaits further characterization.

Another unique feature of Type III systems concerns the mechanism of self/non-self discrimination. Genetic analyses have revealed that Type III systems do not use the PAM-based “non-self-activation” mechanism of Type I (Cascade), Type II (Cas9) and Type V (Cpf1). The mechanism employed by the *S. epidermidis* Csm system apparently involves crRNA- or protein-based recognition of the repeats in the CRISPR locus, resulting in “self-inactivation” [117, 118]. However, the DNA cleavage activity of the *P. furiosus* Cmr complex was recently reported to require the presence of a short sequence adjacent to the target sequence within the activating target RNA, i.e. an RNA protospacer-adjacent motif (rPAM) [106]. Additional analysis is required to reveal whether the reported motifs are typical features that distinguish the two sub-types.

Class 2 systems require only a single protein for interference. In Type II, the crRNP complex involved in target recognition and degradation consists of Cas9 bound to the crRNA guide base-paired with the tracrRNA [37]. The crystal structures of Cas9 reveal two distinct lobes that are involved in target recognition and nuclease activity (Figure 5). The positively charged groove at the interface of the two lobes accommodates the crRNA-DNA heteroduplex [119, 120]. A major step in Cas9 activation is the re-orientation of the structural lobes upon crRNA/tracrRNA loading, which results in the formation of a central channel that accommodates the target DNA [119]. Binding and cleavage of the target DNA by the Cas9-crRNA effector complex depend on the recognition of an appropriate PAM located at the 3’ end of the protospacer [92], which serves as a licensing element in subsequent for DNA strand displacement and R-loop formation. The PAM motif resides in a base-paired DNA duplex. Sequence-specific PAM readout by Arg 1333 and Arg 1335 in Cas9 positions the DNA duplex such that the +1 phosphate group of the target strand interacts with the phosphate lock loop [121]. This promotes local duplex melting, allowing the Cas9–RNA complex to probe the identity of the nucleotides immediately upstream of the PAM. Base pairing between a 12 nucleotide seed sequence of the guide RNA and the target DNA strand [92] drives further stepwise destabilization of the target DNA duplex and directional formation of the guide-RNA–target-DNA heteroduplex [121]. This R-loop triggers a conformational change of the two nuclease domains (HNH and RuvC) of Cas9, which adopt an active state that allows for the completion of interference by cleavage of both target strands [120, 122]. Cas9 generates a blunt end double-strand break, typically located 3 nucleotides from the 3´ end of the protospacer [14, 123]. Recently, however, PAM-independent single-stranded targeting by *Neisseria meningitidis* Cas9 has been described [124].

Similar to Type II, the effector modules of Type V systems consist of a large multi-domain protein complex (Cpf1 and C2c1 in subtypes V-A and V-B, respectively). Like Cas9, these proteins encompass a RuvC-like nuclease domain and an arginine-rich, bridging helix. However, in contrast to

Cas9, the RuvC-like domain of Cpf1 and C2c1 is more compact and the HNH domain is missing (Figure 6). Subtype V-B systems resemble Type II with respect to the requirement for a tracrRNA, both for processing and for interference. In contrast, Cpf1-crRNA (Type V-A) complexes are single RNA-guided endonucleases that cleave target DNA molecules in the absence of a tracrRNA [39]. A model is presented for a step-wise cleavage of the target DNA, i.e. initial RuvC-dependent cleavage of the displaced strand, followed by cleavage of the target strand by the novel nuclease domain. The observation that inactivation of the RuvC domain prevents cleavage of both strands of the target DNA [39, 90] suggests that the novel nuclease is allosterically activated by the RuvC cleavage event. Although allosteric control has also been demonstrated in interference by Cas9 [122], details appear to differ. Both Cpf1 and C2c1 from different bacteria efficiently cleave target DNA containing a well-defined T-rich PAM at the 5' end of the protospacer (5'-PAM) [31, 39] in contrast to the more variable, G-rich 3'-PAM sequence of Cas9 [125]. Structural analysis has shown that Cpf1 recognizes its PAM through a combination of base and shape readout, in which several PAM-interacting amino acid residues that are conserved in the Cpf1 family are involved [Yamano et al 2016]. Another unique feature of the Cpf1 endonuclease is the generation of staggered double-stranded DNA break with 4 or 5-nt 5' overhangs [39]; in the Cpf1 structure, the unique nuclease domain is positioned such as to cleave the target strand outside the heteroduplex as opposed to the HNH domain of Cas9 in which the active site contacts the target within the heteroduplex [Yamano et al 2016] (Figure 6).

The Type VI systems contain a unique effector protein (C2c2) with two predicted HEPN domains. Because all experimentally characterized members of the HEPN superfamily possess RNase activity [126], it appears likely that Type VI systems target RNA. The Type VI systems remain to be functionally characterized, but crRNA processing has been demonstrated in the absence of tracrRNA, suggestive of mechanistic analogies with subtype V-A (Cpf1) [31].

Phages are constantly evolving multiple tactics to avoid, circumvent or subvert prokaryotic defense mechanisms [8]. Phages can evade CRISPR interference through single-nucleotide substitution in the protospacer region or in the conserved protospacer-adjacent motif [46]. Additionally, *P. aeruginosa* phages encode several proteins affecting the activity of Type I-E and I-F systems [127]. Diverse sequences of these proteins and mechanisms of action, coupled with the strong selection imposed by different antiviral systems, suggesting an abundance of anti-CRISPR proteins yet to be discovered. Strikingly, some bacteriophages themselves encode a CRISPR-Cas system that in this case functions as an anti-defense device targeting an antiphage island of the bacterial host and thus enabling productive infection [128]. Together, these findings emphasize the complexity of the virus-host arms race in which CRISPR-Cas systems are involved and suggest that many important aspects of this race remain to be characterized.

Very recently, an unexpected claim has been published on the existence of a CRISPR-like defense system in a giant mimivirus infecting unicellular eukaryotes (amoeba) [129]. This system named 'mimivirus virophage resistance element' (MIMIVIRE) has been proposed to protect certain mimivirus strains from the Zamilon virophage, a small virus that parasitizes on mimiviruses. However, the MIMIVIRE locus lacks CRISPR-like repeats or a Cas1 homolog and encodes only very distant, generic homologs of two Cas proteins (a helicase and a nuclease that belong to the same protein superfamilies as Cas3 and Cas4, respectively, but lack any specific relationship with these Cas proteins). Thus, any analogy between this putative eukaryotic giant virus defense systems and CRISPR-Cas should be perceived with extreme caution.

Genome editing applications

The molecular features of CRISPR-Cas systems, particularly Class 2 systems with single protein effectors, have made them attractive starting points for researchers interested in developing programmable genome editing tools (Table 1). In 2013, the first reports of harnessing Cas9 for multiplex gene editing in human cells appeared [130-133]. These studies have demonstrated that Cas9 could efficiently create indels at precise locations and that by supplying exogenous repair templates, insertion of a new sequence at target sites could be achieved via homologous recombination. A "dead" Cas9 (dCas9) variant with inactivating mutations in the HNH and RuvC domains binds DNA without cutting, providing a programmable platform for recruiting different functional moieties to target sites. The dCas9 has been used for transcriptional activation and repression [134-137], for localizing fluorescent protein labels [138], and for recruiting histone modifying enzymes [139, 140]. Other applications of Cas9 include building gene circuits [141-143], creating new anti-microbials [144], antivirals [145-147], and large-scale gain- and loss-of-function screening [148-151] (Table 1).

The genome editing toolbox has been expanding through the discovery of novel Class 2 effector proteins, such as Cpf1 [39]. Cpf1 does not require tracrRNA, further simplifying the system for genome editing applications. In addition, it generates sticky ends, which could potentially increase the efficiency of insertion of new DNA sequences relative to the blunt ends created by Cas9 [39].

Central to the success of any Cas-based genome editing tool is the specificity of the enzyme, and many approaches to increase specificity have been reported. For example, "double-nicking", which utilizes dimers of two Cas9 variants, each mutated to create a nick in one strand of the DNA, improves specificity by requiring two target matches to create the double-strand break [152, 153]. Another tactic is to control the amount of Cas9 in the cell via an inducible system that expresses a low level of Cas9 [154, 155]. Shortening the region of complementarity in the guide RNA also reduces off-target cleavage [156]. Finally, structure-guided engineering has been used to mutate specific residues in Cas9, to weaken the interaction with the non-target strand or to decrease non-specific interactions with the target DNA site, favoring cleavage at sites that are perfectly complementary to the guide RNA and reducing off-target effects to undetectable levels at many sites [157, 158].

A major outstanding challenge for realizing the full potential of Cas-based genome editing, including its use as a therapeutic, is efficient and tissue-specific delivery. Some progress has been made in this area, including the use of a smaller Cas9 ortholog [159], which is more amenable to packaging into viral vectors. Other approaches are also being pursued, including non-viral methods for delivery of DNA or mRNA by nanoparticles [160] and electroporation [161], or direct delivery of Cas9 protein [162]. Another issue is the long-term effects of Cas9 expression in heterologous eukaryotic cells, which remain unexplored. Last but not least, there are ethical concerns about the potential for editing the human genome as well as those raised by the possibility of using Cas-based gene drives for ecosystem engineering [163] that must be fully considered.

Outlook

The intensive research over the past few years on structural and functional features of variant CRISPR-Cas systems has revealed that they encompass many homologous components and share common mechanistic principles but also show enormous variability. A key aspect of this variability is module shuffling which involves frequent recombination of adaptation and effector modules coming from different types of CRISPR-Cas within the same locus. Apart from major differences in the architectures of the effector complexes, functional diversity of CRISPR-Cas includes versatile mechanisms of crRNA guide processing, self/non-self discrimination, and target cleavage. The versatility of Class 2 systems in particular, where distinct subtypes apparently evolved via independent recombination of adaptation modules with widely different effectors, is notable given the potential of these systems as genome editing tools. The in-depth analysis of a few well characterized CRISPR systems has revealed key structural and mechanistic features. However, the continuing discovery of novel CRISPR-Cas variants and new molecular mechanisms implies that our current insights have limited power for predicting functional details of distantly related variants. Hence, such new CRISPR-Cas systems need to be meticulously analysed to understand the biology of prokaryotic adaptive immunity and harness its potential for biotechnology. In this review, we could not cover in any detail several fascinating aspects of CRISPR-Cas biology such as co-evolution of immune systems with viruses, the interplay between CRISPR-Cas activity and horizontal gene transfer, or non-immune functions of CRISPR-Cas. The complexity and extreme variability of the CRISPR-Cas systems ensure that researchers in this field will have much to do for many years to come.

Figure legends

Figure 0 – CRISPR-Cas functions – natural & synthetic (Graphical abstract)

Figure 1 -Overview of the CRISPR–Cas (Clustered Regularly Interspaced Short palindromic repeats–CRISPR-associated proteins) system.

(a) Architecture of Class 1 (multi-protein effector complexes) and Class 2 (single-protein effector complexes) CRISPR-Cas systems.

(b) CRISPR–Cas adaptive immunity is mediated by CRISPR RNAs (crRNAs) and Cas proteins, which form multi-component CRISPR ribonucleoprotein (crRNP) complexes. The first stage is adaptation, which occurs upon entry of an invading mobile genetic element (in this case, a viral genome). Cas1 (blue) and Cas2 (yellow) proteins select and process the invading DNA, and thereafter, a protospacer (orange) is integrated as a new spacer at the leader end of the CRISPR array (repeat sequences (grey) that separate similarly-sized, invader-derived spacers (multiple colors)). During the second stage, expression, the CRISPR locus is transcribed and the pre-crRNA is processed into mature crRNA guides by Cas (e.g. Cas6) or non-Cas proteins (e.g. RNase-III). During the final interference stage, the Cas-crRNA complex scans invading DNA for a complementary nucleic acid target, after which the target is degraded by a Cas nuclease.

Figure 2 – CRISPR diversity and evolution.

(a) Modular organization of the CRISPR-Cas systems. LS, Large Subunit, SS, Small Subunit. A putative small subunit that might be fused to the large subunit in several Type I subtypes is indicated by an asterisk. Cas3 is shown as fusion of two distinct genes encoding the helicase Cas3' and the nuclease HD Cas3''; in some Type I systems, these domains are encoded by separate genes. Functionally dispensable components are indicated by dashed outlines. Cas6 is shown with a thin solid outline for Type I because it is dispensable in some systems, and by a dashed line for Type III because most systems lack this gene and use the Cas6 provided in trans by other CRISPR–cas loci. The two colours for Cas4, Cpf1 and C2c1, and three colours for Cas9 reflect the contributions of these proteins to different stages of the CRISPR-Cas response (see text). Modified with permission from [30]. The question mark indicates currently unknown components.

(b) Evolutionary scenario for the CRISPR-Cas systems. Abbreviations: TR, terminal repeats; TS, terminal sequences; HD, HD-family endonuclease; HNH, HNH-family endonuclease; RuvC, RuvC-family endonuclease; HEPN, putative endoribonuclease of HEPN-superfamily. Genes and portions of genes shown in grey denote sequences that are thought to have been encoded in the respective mobile elements but were eliminated in the course of evolution of CRISPR-Cas systems. Modified with permission from [31].

Figure 3 – Spacer acquisition.

(a) Crystal structure of the complex of Cas1-Cas2 bound to the dual-forked DNA (PDB accession 5DQZ). The target DNA is shown in dark blue and the Cas1 and Cas2 dimers of the complex are indicated in blue, and yellow, respectively.

(b) Model explaining the capture of new DNA sequences from invading nucleic acid and the subsequent DNA integration into the host CRISPR array. The numbers on the left correspond to the order of events as described in the text. The dashed lines indicate nucleotides, the nucleotides C and N on the two sides of the protospacer are shown in red and green to clarify the orientation.

Figure 4 – Guide expression and processing.

(a) Generation of CRISPR RNA (crRNA) guides in Type I and Type III CRISPR–Cas systems. Primary processing of the pre-crRNA is catalysed by Cas6, which typically results in a crRNA with a 5' handle of 8 nucleotides, a central spacer sequence and (in some subtypes) a longer 3' handle. Shown here is the guide processing (red triangles) for subtype I-E by Cas6e. The occasional secondary processing of the 3' end of crRNA is catalysed by unknown ribonuclease(s).

(b) In Type II CRISPR–Cas systems, the repeat sequences of the pre-crRNA hybridize with complementary sequences of transactivating CRISPR RNA (tracrRNA). The double-stranded RNA is cleaved by RNase-III (red triangles); further trimming of the 5’ end of the spacer is carried out by by unknown ribonuclease(s) (pink).

(c) CRISPR with transcriptional start site (TSS) in repeats.

Figure 5 – CRISPR RNP complexes.

Crystal structures of the CRISPR-ribonucleoproteins (crRNP) complexes responsible for target interference. Type I-E Cascade complex (PDB accession 4QYZ) and Type III-A Cmr complex (PDB accession 3X1L) from Class 1 and Type II-B Cas9 complex (PDB accession 4OO8) and Type V-A Cpf1 complex (PDB accession [Yamano et al 2016]) from Class 2. Colors of nucleic acid fragments are same as in Figure 6.

Figure 6 – Target interference.

Genomic loci architecture of the components of Class 1 and Class 2 CRISPR-Cas systems and schematic representation of target interferences for the different subtypes. The dsDNA (target) is shown in black, the CRISPR RNA (crRNA) repeat in blue, the spacer region of the CRISPR RNA repeat in green, and the transactivating CRISPR RNA (tracrRNA) is in red.

Supplementary Data

Movie - CRISPR RNP complexes (as in Fig.5) – draft is ready, will be optimized soon.

References

1. Koonin, E.V. and V.V. Dolja, *A virocentric perspective on the evolution of life.* Curr Opin Virol, 2013. **3**(5): p. 546-57.
2. Makarova, K.S., Y.I. Wolf, and E.V. Koonin, *Comparative genomics of defense systems in archaea and bacteria.* Nucleic Acids Res, 2013. **41**(8): p. 4360-77.
3. Tock, M.R. and D.T. Dryden, *The biology of restriction and anti-restriction.* Curr Opin Microbiol, 2005. **8**(4): p. 466-72.
4. Swarts, D.C., et al., *DNA-guided DNA interference by a prokaryotic Argonaute.* Nature, 2014. **507**(7491): p. 258-61.
5. Goldfarb, T., et al., *BREX is a novel phage resistance system widespread in microbial genomes.* EMBO J, 2015. **34**(2): p. 169-83.
6. Makarova, K.S., et al., *Defense islands in bacterial and archaeal genomes and prediction of novel defense systems.* J Bacteriol, 2011. **193**(21): p. 6039-56.
7. Chopin, M.C., A. Chopin, and E. Bidnenko, *Phage abortive infection in lactococci: variations on a theme.* Curr Opin Microbiol, 2005. **8**(4): p. 473-9.
8. Samson, J.E., et al., *Revenge of the phages: defeating bacterial defences.* Nat Rev Microbiol, 2013. **11**(10): p. 675-87.
9. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.
10. Jansen, R., et al., *Identification of genes that are associated with DNA repeats in prokaryotes.* Mol Microbiol, 2002. **43**(6): p. 1565-75.
11. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex.* Cell, 2009. **139**(5): p. 945-56.
12. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-4.
13. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA.* Science, 2008. **322**(5909): p. 1843-5.
14. Garneau, J.E., et al., *The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA.* Nature, 2010. **468**(7320): p. 67-71.
15. Terns, M.P. and R.M. Terns, *CRISPR-based adaptive immune systems.* Curr Opin Microbiol, 2011. **14**(3): p. 321-7.
16. Mojica, F.J., et al., *Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements.* J Mol Evol, 2005. **60**(2): p. 174-82.
17. Bolotin, A., et al., *Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin.* Microbiology, 2005. **151**(Pt 8): p. 2551-61.
18. Pourcel, C., G. Salvignol, and G. Vergnaud, *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): p. 653-63.
19. Makarova, K.S., et al., *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**: p. 7.
20. van der Oost, J., et al., *CRISPR-based adaptive and heritable immunity in prokaryotes.* Trends Biochem Sci, 2009. **34**(8): p. 401-7.
21. Wiedenheft, B., S.H. Sternberg, and J.A. Doudna, *RNA-guided genetic silencing systems in bacteria and archaea.* Nature, 2012. **482**(7385): p. 331-8.
22. Barrangou, R. and L.A. Marraffini, *CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity.* Mol Cell, 2014. **54**(2): p. 234-44.
23. Marraffini, L.A., *CRISPR-Cas immunity in prokaryotes.* Nature, 2015. **526**(7571): p. 55-61.
24. Gasiunas, G., T. Sinkunas, and V. Siksnys, *Molecular mechanisms of CRISPR-mediated microbial immunity.* Cell Mol Life Sci, 2014. **71**(3): p. 449-65.
25. Westra, E.R., A. Buckling, and P.C. Fineran, *CRISPR-Cas systems: beyond adaptive immunity.* Nat Rev Microbiol, 2014. **12**(5): p. 317-26.
26. Sampson, T.R. and D.S. Weiss, *CRISPR-Cas systems: new players in gene regulation and bacterial physiology.* Front Cell Infect Microbiol, 2014. **4**: p. 37.
27. Louwen, R., et al., *The role of CRISPR-Cas systems in virulence of pathogenic bacteria.* Microbiol Mol Biol Rev, 2014. **78**(1): p. 74-88.
28. Takeuchi, N., et al., *Nature and intensity of selection pressure on CRISPR-associated genes.* J Bacteriol, 2012. **194**(5): p. 1216-25.
29. Koonin, E.V. and Y.I. Wolf, *Evolution of the CRISPR-Cas adaptive immunity systems in prokaryotes: models and observations on virus-host coevolution.* Mol Biosyst, 2015. **11**(1): p. 20-7.
30. Makarova, K.S., et al., *An updated evolutionary classification of CRISPR-Cas systems.* Nat Rev Microbiol, 2015. **13**(11): p. 722-36.
31. Shmakov, S., et al., *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems.* Mol Cell, 2015. **60**(3): p. 385-97.
32. Makarova, K.S., et al., *Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems.* Biol Direct, 2011. **6**: p. 38.
33. Makarova, K.S., Y.I. Wolf, and E.V. Koonin, *The basic building blocks and evolution of CRISPR-CAS systems.* Biochem Soc Trans, 2013. **41**(6): p. 1392-400.
34. Rouillon, C., et al., *Structure of the CRISPR interference complex CSM reveals key similarities with cascade.* Mol Cell, 2013. **52**(1): p. 124-34.

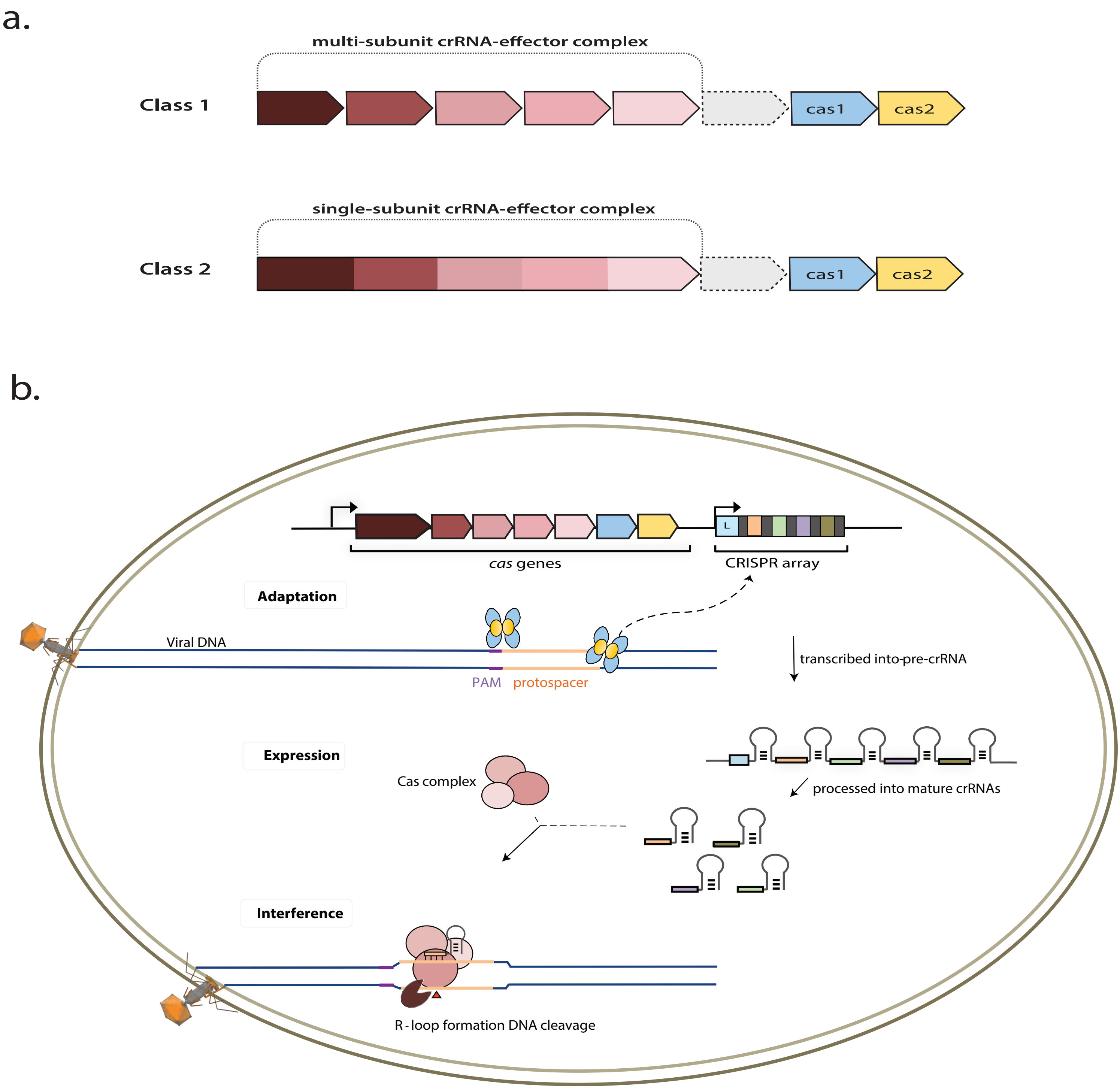
35. Spilman, M., et al., *Structure of an RNA silencing complex of the CRISPR-Cas immune system.* Mol Cell, 2013. **52**(1): p. 146-52.
36. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.
37. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-7.
38. Chylinski, K., et al., *Classification and evolution of type II CRISPR-Cas systems.* Nucleic Acids Res, 2014. **42**(10): p. 6091-105.
39. Zetsche, B., et al., *Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System.* Cell, 2015. **163**(3): p. 759-71.
40. Krupovic, M., et al., *Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity.* BMC Biol, 2014. **12**: p. 36.
41. Hickman, A.B. and F. Dyda, *The casposon-encoded Cas1 protein from Aciduliprofundum boonei is a DNA integrase that generates target site duplications.* Nucleic Acids Res, 2015. **43**(22): p. 10576-87.
42. Krupovic, M., et al., *Recent mobility of casposons, self-synthesizing transposons at the origin of the CRISPR-Cas immunity.* Genome Biol Evol, 2016.
43. Koonin, E.V. and M. Krupovic, *Evolution of adaptive immunity from transposable elements combined with innate immune systems.* Nat Rev Genet, 2015. **16**(3): p. 184-92.
44. Kapitonov, V.V., K.S. Makarova, and E.V. Koonin, *ISC, a novel group of bacterial and archaeal DNA transposons that encode Cas9 homologs.* J Bacteriol, 2015.
45. Amitai, G. and R. Sorek, *CRISPR-Cas adaptation: insights into the mechanism of action.* Nat Rev Microbiol, 2016. **14**(2): p. 67-76.
46. Deveau, H., et al., *Phage response to CRISPR-encoded resistance in Streptococcus thermophilus.* J Bacteriol, 2008. **190**(4): p. 1390-400.
47. Mojica, F.J., et al., *Short motif sequences determine the targets of the prokaryotic CRISPR defence system.* Microbiology, 2009. **155**(Pt 3): p. 733-40.
48. Swarts, D.C., et al., *CRISPR interference directs strand specific spacer acquisition.* PLoS One, 2012. **7**(4): p. e35888.
49. Datsenko, K.A., et al., *Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system.* Nat Commun, 2012. **3**: p. 945.
50. Stern, A., et al., *Self-targeting by CRISPR: gene regulation or autoimmunity?* Trends Genet, 2010. **26**(8): p. 335-40.
51. Yosef, I., M.G. Goren, and U. Qimron, *Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli.* Nucleic Acids Res, 2012. **40**(12): p. 5569-76.
52. Wei, Y., R.M. Terns, and M.P. Terns, *Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation.* Genes Dev, 2015. **29**(4): p. 356-61.
53. Diez-Villasenor, C., et al., *CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli.* RNA Biol, 2013. **10**(5): p. 792-802.
54. Nunez, J.K., et al., *Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity.* Nat Struct Mol Biol, 2014. **21**(6): p. 528-34.
55. Levy, A., et al., *CRISPR adaptation biases explain preference for acquisition of foreign DNA.* Nature, 2015. **520**(7548): p. 505-10.
56. Erdmann, S., S. Le Moine Bauer, and R.A. Garrett, *Inter-viral conflicts that exploit host CRISPR immune systems of Sulfolobus.* Mol Microbiol, 2014. **91**(5): p. 900-17.
57. Wigley, D.B., *RecBCD: the supercar of DNA repair.* Cell, 2007. **131**(4): p. 651-3.
58. Dupuis, M.E., et al., *CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance.* Nat Commun, 2013. **4**: p. 2087.
59. Wang, J., et al., *Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems.* Cell, 2015. **163**(4): p. 840-53.
60. Nunez, J.K., et al., *Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity.* Nature, 2015. **519**(7542): p. 193-8.
61. Silas, S., et al., *Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein.* Science, 2016. **351**(6276): p. aad4234.
62. Fineran, P.C. and E. Charpentier, *Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information.* Virology, 2012. **434**(2): p. 202-9.
63. Arslan, Z., et al., *Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system.* Nucleic Acids Res, 2014. **42**(12): p. 7884-93.
64. Li, M., et al., *Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process.* Nucleic Acids Res, 2014. **42**(4): p. 2483-92.
65. Richter, C., et al., *Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer.* Nucleic Acids Res, 2014. **42**(13): p. 8516-26.
66. Fineran, P.C., et al., *Degenerate target sites mediate rapid primed CRISPR adaptation.* Proc Natl Acad Sci U S A, 2014. **111**(16): p. E1629-38.
67. Xue, C., et al., *CRISPR interference and priming varies with individual spacer sequences.* Nucleic Acids Res, 2015. **43**(22): p. 10831-47.
68. Vorontsova, D., et al., *Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery.* Nucleic Acids Res, 2015. **43**(22): p. 10848-60.
69. Ivancic-Bace, I., et al., *Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity.* Nucleic Acids Res, 2015. **43**(22): p. 10821-30.
70. Heler, R., et al., *Cas9 specifies functional viral targets during CRISPR-Cas adaptation.* Nature, 2015. **519**(7542): p. 199-202.

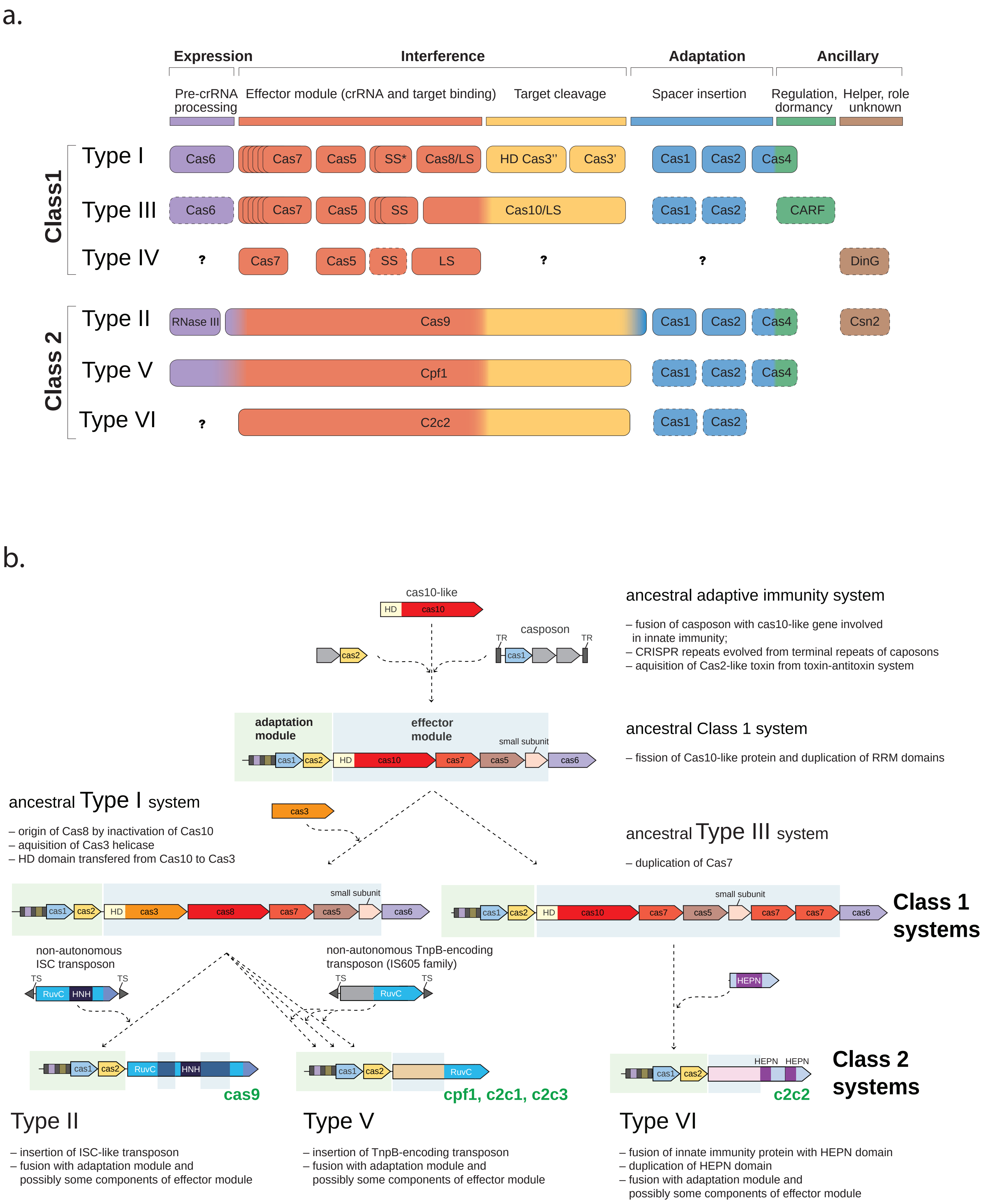
71. Hooton, S.P. and I.F. Connerton, *Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein.* Front Microbiol, 2014. **5**: p. 744.
72. Charpentier, E., et al., *Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity.* FEMS Microbiol Rev, 2015. **39**(3): p. 428-41.
73. Carte, J., et al., *Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes.* Genes Dev, 2008. **22**(24): p. 3489-96.
74. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea.* Nat Rev Genet, 2010. **11**(3): p. 181-90.
75. Jore, M.M., et al., *Structural basis for CRISPR RNA-guided DNA recognition by Cascade.* Nat Struct Mol Biol, 2011. **18**(5): p. 529-36.
76. Reeks, J., J.H. Naismith, and M.F. White, *CRISPR interference: a structural perspective.* Biochem J, 2013. **453**(2): p. 155-66.
77. Niewoehner, O., M. Jinek, and J.A. Doudna, *Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases.* Nucleic Acids Res, 2014. **42**(2): p. 1341-53.
78. Wiedenheft, B., et al., *RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10092-7.
79. Rollins, M.F., et al., *Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa.* Nucleic Acids Res, 2015. **43**(4): p. 2216-22.
80. Wiedenheft, B., et al., *Structures of the RNA-guided surveillance complex from a bacterial immune system.* Nature, 2011. **477**(7365): p. 486-9.
81. Haurwitz, R.E., et al., *Sequence- and structure-specific RNA processing by a CRISPR endonuclease.* Science, 2010. **329**(5997): p. 1355-8.
82. Nam, K.H., et al., *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): p. 1574-84.
83. Staals, R.H., et al., *RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus.* Mol Cell, 2014. **56**(4): p. 518-30.
84. Majumdar, S., et al., *Three CRISPR-Cas immune effector complexes coexist in Pyrococcus furiosus.* RNA, 2015. **21**(6): p. 1147-58.
85. Zhang, J., et al., *Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity.* Mol Cell, 2012. **45**(3): p. 303-13.
86. Hatoum-Aslan, A., et al., *A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs.* J Biol Chem, 2013. **288**(39): p. 27888-97.
87. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.* Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-86.
88. Garrett, R.A., et al., *CRISPR-Cas Adaptive Immune Systems of the Sulfolobales: Unravelling Their Complexity and Diversity.* Life (Basel), 2015. **5**(1): p. 783-817.
89. Zhang, Y., et al., *Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis.* Mol Cell, 2013. **50**(4): p. 488-503.
90. Fonfara, I., et al., *The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA.* Nature, 2016.
91. Semenova, E., et al., *Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10098-103.
92. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.* Nature, 2014. **507**(7490): p. 62-7.
93. Redding, S., et al., *Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System.* Cell, 2015. **163**(4): p. 854-65.
94. Rutkauskas, M., et al., *Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection.* Cell Rep, 2015.
95. Blosser, T.R., et al., *Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex.* Mol Cell, 2015. **58**(1): p. 60-70.
96. Jackson, R.N., et al., *Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli.* Science, 2014. **345**(6203): p. 1473-9.
97. Mulepati, S., A. Heroux, and S. Bailey, *Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target.* Science, 2014. **345**(6203): p. 1479-84.
98. Taylor, D.W., et al., *Structural biology. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning.* Science, 2015. **348**(6234): p. 581-5.
99. Osawa, T., et al., *Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog.* Mol Cell, 2015. **58**(3): p. 418-30.
100. Jackson, R.N. and B. Wiedenheft, *A Conserved Structural Chassis for Mounting Versatile CRISPR RNA-Guided Immune Responses.* Mol Cell, 2015. **58**(5): p. 722-8.
101. Zebec, Z., et al., *CRISPR-mediated targeted mRNA degradation in the archaeon Sulfolobus solfataricus.* Nucleic Acids Res, 2014. **42**(8): p. 5280-8.
102. Staals, R.H., et al., *Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus.* Mol Cell, 2013. **52**(1): p. 135-45.
103. Tamulaitis, G., et al., *Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus.* Mol Cell, 2014. **56**(4): p. 506-17.
104. Peng, W., et al., *An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference.* Nucleic Acids Res, 2015. **43**(1): p. 406-17.
105. Estrella, M.A., F.T. Kuo, and S. Bailey, *RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex.* Genes Dev, 2016. **30**(4): p. 460-70.

106. Elmore, J.R., et al., *Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system.* Genes Dev, 2016. **30**(4): p. 447-59.
107. Deng, L., et al., *A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus.* Mol Microbiol, 2013. **87**(5): p. 1088-99.
108. Samai, P., et al., *Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity.* Cell, 2015. **161**(5): p. 1164-74.
109. Goldberg, G.W., et al., *Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting.* Nature, 2014. **514**(7524): p. 633-7.
110. Hale, C.R., et al., *Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex.* Genes Dev, 2014. **28**(21): p. 2432-43.
111. Benda, C., et al., *Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4.* Mol Cell, 2014. **56**(1): p. 43-54.
112. Ramia, N.F., et al., *Essential structural and functional roles of the Cmr4 subunit in RNA cleavage by the Cmr CRISPR-Cas complex.* Cell Rep, 2014. **9**(5): p. 1610-7.
113. Ramia, N.F., et al., *Staphylococcus epidermidis Csm1 is a 3'-5' exonuclease.* Nucleic Acids Res, 2014. **42**(2): p. 1129-38.
114. Sheppard, N.F., et al., *The CRISPR-associated Csx1 protein of Pyrococcus furiosus is an adenosine-specific endoribonuclease.* RNA, 2016. **22**(2): p. 216-24.
115. Jiang, W., P. Samai, and L.A. Marraffini, *Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity.* Cell, 2016. **164**(4): p. 710-21.
116. Niewoehner, O. and M. Jinek, *Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6.* RNA, 2016. **22**(3): p. 318-29.
117. Marraffini, L.A. and E.J. Sontheimer, *Self versus non-self discrimination during CRISPR RNA-directed immunity.* Nature, 2010. **463**(7280): p. 568-71.
118. van der Oost, J., et al., *Unravelling the structural and mechanistic basis of CRISPR-Cas systems.* Nat Rev Microbiol, 2014. **12**(7): p. 479-92.
119. Jinek, M., et al., *Structures of Cas9 endonucleases reveal RNA-mediated conformational activation.* Science, 2014. **343**(6176): p. 1247997.
120. Nishimasu, H., et al., *Crystal structure of Cas9 in complex with guide RNA and target DNA.* Cell, 2014. **156**(5): p. 935-49.
121. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease.* Nature, 2014. **513**(7519): p. 569-73.
122. Sternberg, S.H., et al., *Conformational control of DNA target cleavage by CRISPR-Cas9.* Nature, 2015. **527**(7576): p. 110-3.
123. Magadan, A.H., et al., *Cleavage of phage DNA by the Streptococcus thermophilus CRISPR3-Cas system.* PLoS One, 2012. **7**(7): p. e40913.
124. Zhang, Y., et al., *DNase H Activity of Neisseria meningitidis Cas9.* Mol Cell, 2015. **60**(2): p. 242-55.
125. Hsu, P.D., E.S. Lander, and F. Zhang, *Development and applications of CRISPR-Cas9 for genome engineering.* Cell, 2014. **157**(6): p. 1262-78.
126. Anantharaman, V., et al., *Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing.* Biol Direct, 2013. **8**: p. 15.
127. Bondy-Denomy, J., et al., *Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins.* Nature, 2015. **526**(7571): p. 136-9.
128. Seed, K.D., et al., *A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity.* Nature, 2013. **494**(7438): p. 489-91.
129. Levasseur, A., et al., *MIMIVIRE is a defence system in mimivirus that confers resistance to virophage.* Nature, 2016. **531**(7593): p. 249-52.
130. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.
131. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.
132. Jinek, M., et al., *RNA-programmed genome editing in human cells.* Elife, 2013. **2**: p. e00471.
133. Cho, S.W., et al., *Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease.* Nat Biotechnol, 2013. **31**(3): p. 230-2.
134. Qi, L.S., et al., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-83.
135. Bikard, D., et al., *Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system.* Nucleic Acids Res, 2013. **41**(15): p. 7429-37.
136. Konermann, S., et al., *Optical control of mammalian endogenous transcription and epigenetic states.* Nature, 2013. **500**(7463): p. 472-6.
137. Gilbert, L.A., et al., *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-51.
138. Chen, B., et al., *Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system.* Cell, 2013. **155**(7): p. 1479-91.
139. Hilton, I.B., et al., *Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers.* Nat Biotechnol, 2015. **33**(5): p. 510-7.
140. Kearns, N.A., et al., *Functional annotation of native enhancers with a Cas9-histone demethylase fusion.* Nat Methods, 2015. **12**(5): p. 401-3.
141. Kiani, S., et al., *CRISPR transcriptional repression devices and layered circuits in mammalian cells.* Nat Methods, 2014. **11**(7): p. 723-6.
142. Liu, Y., et al., *Synthesizing AND gate genetic circuits based on CRISPR-Cas9 for identification of bladder cancer cells.* Nat Commun, 2014. **5**: p. 5393.

143. Nissim, L., et al., *Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells.* Mol Cell, 2014. **54**(4): p. 698-710.
144. Bikard, D., et al., *Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials.* Nat Biotechnol, 2014. **32**(11): p. 1146-50.
145. Ebina, H., et al., *Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus.* Sci Rep, 2013. **3**: p. 2510.
146. Ramanan, V., et al., *CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus.* Sci Rep, 2015. **5**: p. 10833.
147. Hu, W., et al., *RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection.* Proc Natl Acad Sci U S A, 2014. **111**(31): p. 11461-6.
148. Shalem, O., et al., *Genome-scale CRISPR-Cas9 knockout screening in human cells.* Science, 2014. **343**(6166): p. 84-7.
149. Wang, T., et al., *Genetic screens in human cells using the CRISPR-Cas9 system.* Science, 2014. **343**(6166): p. 80-4.
150. Konermann, S., et al., *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex.* Nature, 2015. **517**(7536): p. 583-8.
151. Gilbert, L.A., et al., *Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.* Cell, 2014. **159**(3): p. 647-61.
152. Mali, P., K.M. Esvelt, and G.M. Church, *Cas9 as a versatile tool for engineering biology.* Nat Methods, 2013. **10**(10): p. 957-63.
153. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.* Cell, 2013. **154**(6): p. 1380-9.
154. Zetsche, B., S.E. Volz, and F. Zhang, *A split-Cas9 architecture for inducible genome editing and transcription modulation.* Nat Biotechnol, 2015. **33**(2): p. 139-42.
155. Davis, K.M., et al., *Small molecule-triggered Cas9 protein with improved genome-editing specificity.* Nat Chem Biol, 2015. **11**(5): p. 316-8.
156. Fu, Y., et al., *Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.* Nat Biotechnol, 2014. **32**(3): p. 279-84.
157. Slaymaker, I.M., et al., *Rationally engineered Cas9 nucleases with improved specificity.* Science, 2016. **351**(6268): p. 84-8.
158. Kleinstiver, B.P., et al., *High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.* Nature, 2016.
159. Ran, F.A., et al., *In vivo genome editing using Staphylococcus aureus Cas9.* Nature, 2015. **520**(7546): p. 186-91.
160. Platt, R.J., et al., *CRISPR-Cas9 knockin mice for genome editing and cancer modeling.* Cell, 2014. **159**(2): p. 440-55.
161. Qin, W., et al., *Efficient CRISPR/Cas9-Mediated Genome Editing in Mice by Zygote Electroporation of Nuclease.* Genetics, 2015. **200**(2): p. 423-30.
162. D'Astolfo, D.S., et al., *Efficient intracellular delivery of native proteins.* Cell, 2015. **161**(3): p. 674-90.
163. Esvelt, K.M., et al., *Concerning RNA-guided gene drives for the alteration of wild populations.* Elife, 2014: p. e03401.

END


a.
multi-subunit crRNA-effector complex
Class 1
cas1
cas2
single-subunit crRNA-effector complex
Class 2
cas1
cas2
b.
L
cas genes
CRISPR array
Adaptation
Viral DNA
PAM
protospacer
transcribed into-pre-crRNA
Expression
Cas complex
processed into mature crRNAs
Interference
R - loop formation DNA cleavage



ancestral adaptive immunity system

– fusion of casposon with cas10-like gene involved in innate immunity;
– CRISPR repeats evolved from terminal repeats of caposons
– aquisition of Cas2-like toxin from toxin-antitoxin system

ancestral Class 1 system

– fission of Cas10-like protein and duplication of RRM domains

ancestral Type I system

– origin of Cas8 by inactivation of Cas10
– aquisition of Cas3 helicase
– HD domain transfered from Cas10 to Cas3

ancestral Type III system

– duplication of Cas7

Type II

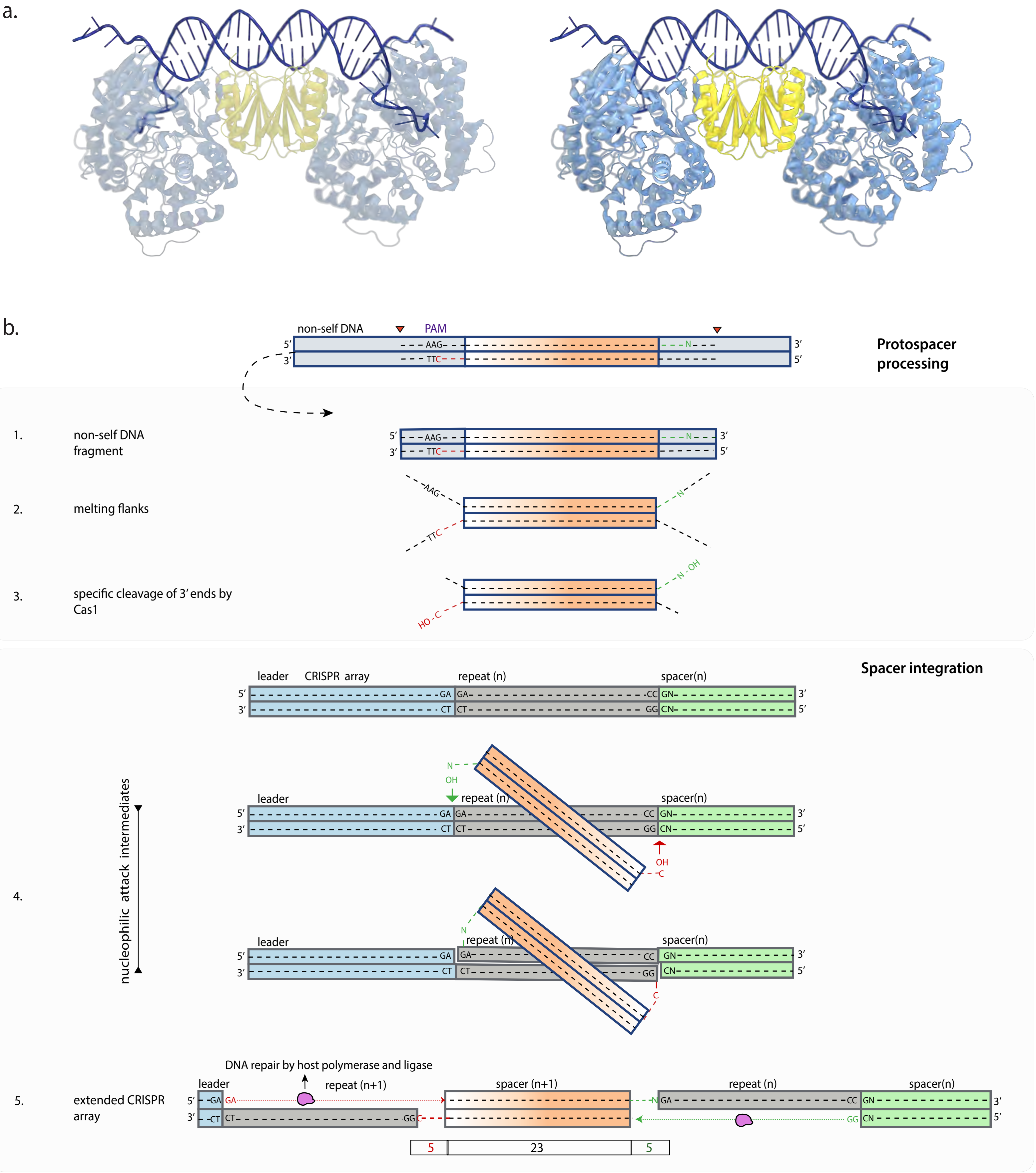
– insertion of ISC-like transposon
– fusion with adaptation module and possibly some components of effector module

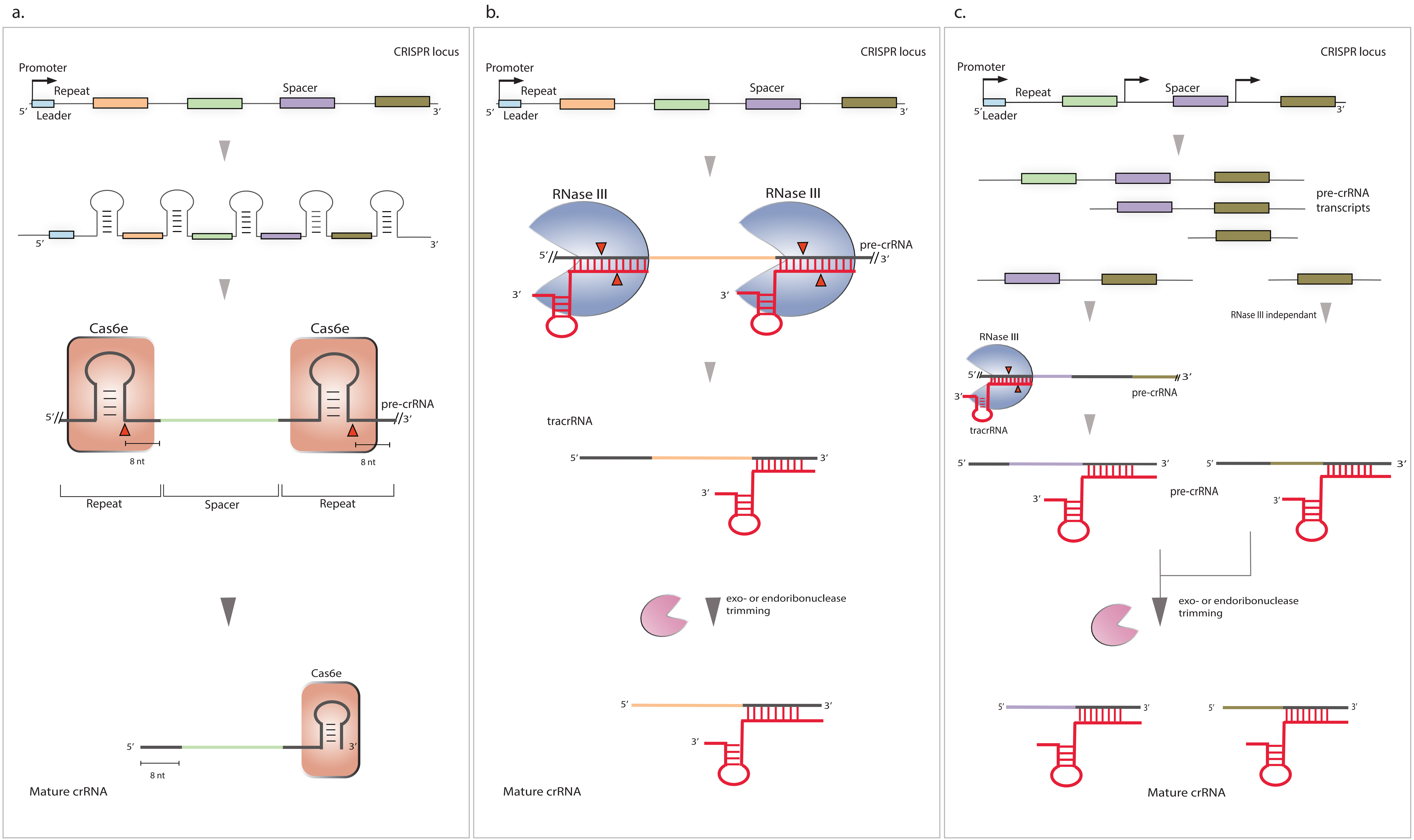
Type V

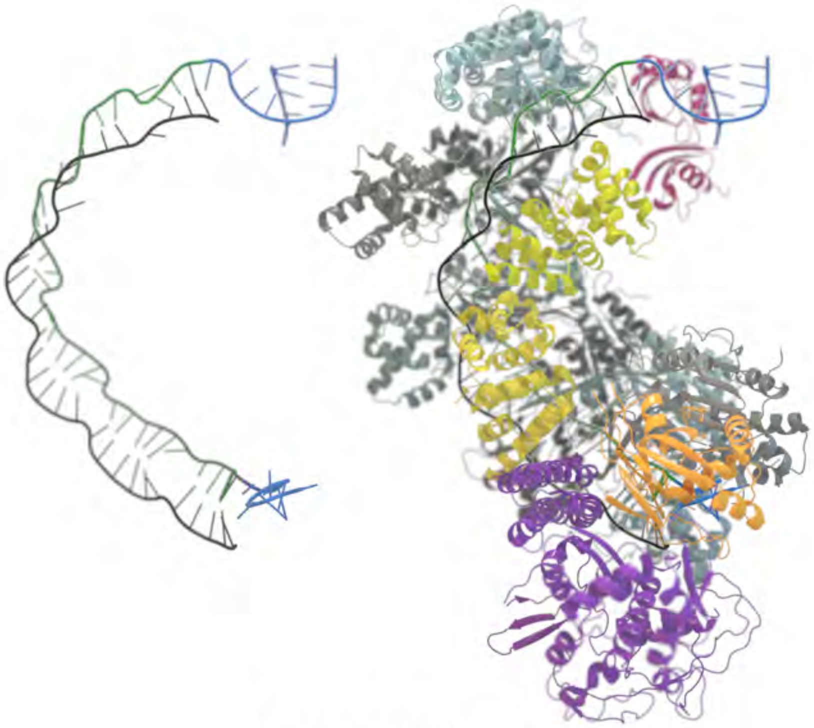
– insertion of TnpB-encoding transposon
– fusion with adaptation module and possibly some components of effector module

Type VI

– fusion of innate immunity protein with HEPN domain
– duplication of HEPN domain
– fusion with adaptation module and possibly some components of effector module


a.
b.
non-self DNA
PAM
5′
3′
AAG
TTC
N
3′
5′
Protospacer processing
1.
non-self DNA fragment
2.
melting flanks
3.
specific cleavage of 3′ ends by Cas1
HO - C
N - OH
Spacer integration
leader
CRISPR array
repeat (n)
spacer(n)
GA
CT
CC
GG
GN
CN
OH
N
C
4.
nucleophilic attack intermediates
5.
extended CRISPR array
DNA repair by host polymerase and ligase
repeat (n+1)
spacer (n+1)
5
23
5

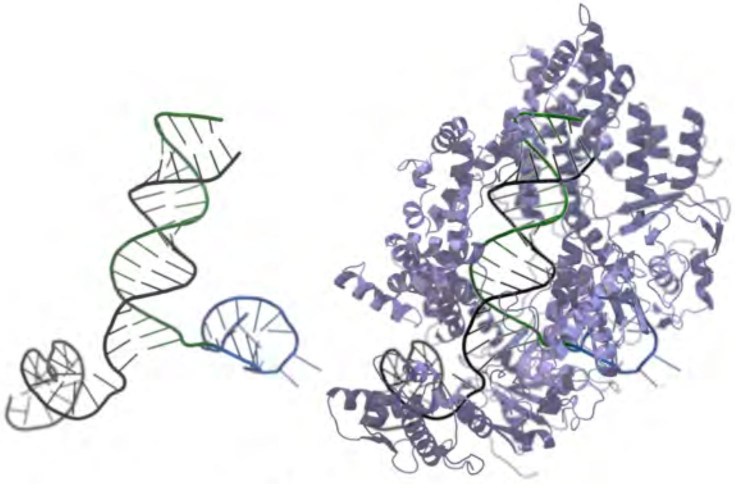

a.
CRISPR locus
Promoter
Repeat
Spacer
5′
Leader
3′
5′
3′
Cas6e
Cas6e
pre-crRNA
5′//
//3′
8 nt
8 nt
Repeat
Spacer
Repeat
Cas6e
5′
3′
8 nt
Mature crRNA
b.
CRISPR locus
Promoter
Repeat
Spacer
5′
Leader
3′
RNase III
RNase III
pre-crRNA
5′//
//3′
3′
3′
tracrRNA
5′
3′
3′
exo- or endoribonuclease trimming
5′
3′
3′
Mature crRNA
c.
CRISPR locus
Promoter
Repeat
Spacer
5′
Leader
3′
pre-crRNA transcripts
RNase III independant
RNase III
5′//
//3′
3′
pre-crRNA
tracrRNA
5′
3′
5′
3′
pre-crRNA
3′
3′
exo- or endoribonuclease trimming
5′
3′
5′
3′
Mature crRNA



Cascade complex (Type I-E)

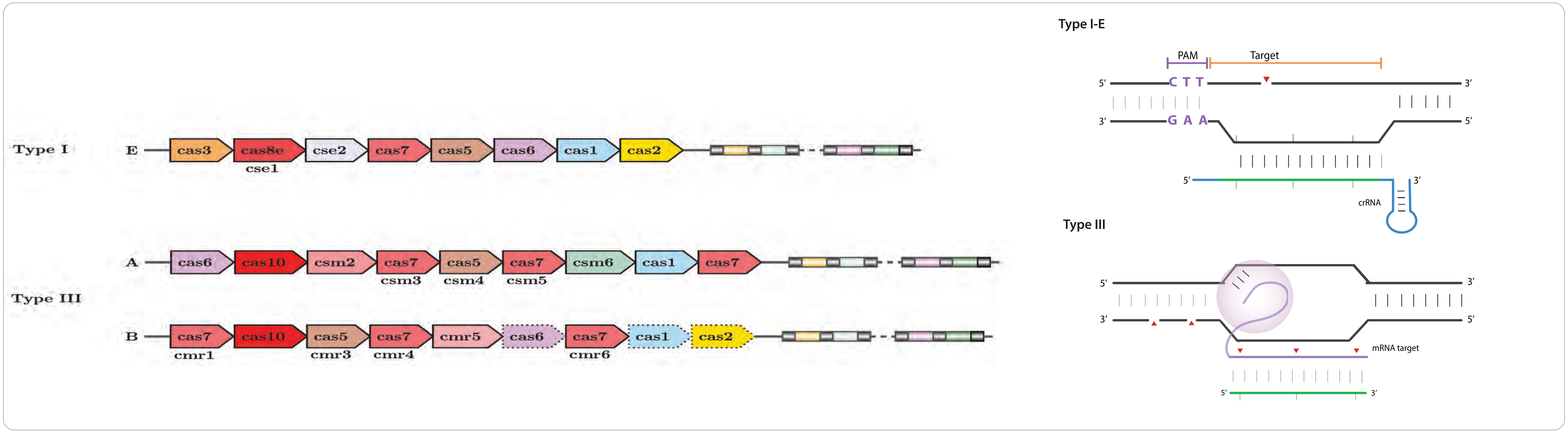
Cmr complex (Type III-B)

Cas9 complex (Type II-B)



Cpf1 complex (Type V-A)

Class 1



Class 2