

Review



**Cite this article:** Koonin EV, Makarova KS. 2019 Origins and evolution of CRISPR-Cas systems. *Phil. Trans. R. Soc. B* **374**: 20180087. <http://dx.doi.org/10.1098/rstb.2018.0087>

Accepted: 24 October 2018

One contribution of 17 to a discussion meeting issue ‘The ecology and evolution of prokaryotic CRISPR-Cas adaptive immune systems’.

**Subject Areas:**

evolution, genomics, microbiology

**Keywords:**

adaptive immunity, mobile genetic elements, signalling, gene shuffling

**Author for correspondence:**

Eugene V. Koonin  
e-mail: [koonin@ncbi.nlm.nih.gov](mailto:koonin@ncbi.nlm.nih.gov)

# Origins and evolution of CRISPR-Cas systems

Eugene V. Koonin and Kira S. Makarova

National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD 20894, USA

EVK, 0000-0003-3943-8299

CRISPR-Cas, the bacterial and archaeal adaptive immunity systems, encompass a complex machinery that integrates fragments of foreign nucleic acids, mostly from mobile genetic elements (MGE), into CRISPR arrays embedded in microbial genomes. Transcripts of the inserted segments (spacers) are employed by CRISPR-Cas systems as guide (g)RNAs for recognition and inactivation of the cognate targets. The CRISPR-Cas systems consist of distinct adaptation and effector modules whose evolutionary trajectories appear to be at least partially independent. Comparative genome analysis reveals the origin of the adaptation module from casposons, a distinct type of transposons, which employ a homologue of Cas1 protein, the integrase responsible for the spacer incorporation into CRISPR arrays, as the transposase. The origin of the effector module(s) is far less clear. The CRISPR-Cas systems are partitioned into two classes, class 1 with multisubunit effectors, and class 2 in which the effector consists of a single, large protein. The class 2 effectors originate from nucleases encoded by different MGE, whereas the origin of the class 1 effector complexes remains murky. However, the recent discovery of a signalling pathway built into the type III systems of class 1 might offer a clue, suggesting that type III effector modules could have evolved from a signal transduction system involved in stress-induced programmed cell death. The subsequent evolution of the class 1 effector complexes through serial gene duplication and displacement, primarily of genes for proteins containing RNA recognition motif domains, can be hypothetically reconstructed. In addition to the multiple contributions of MGE to the evolution of CRISPR-Cas, the reverse flow of information is notable, namely, recruitment of minimalist variants of CRISPR-Cas systems by MGE for functions that remain to be elucidated. Here, we attempt a synthesis of the diverse threads that shed light on CRISPR-Cas origins and evolution.

This article is part of a discussion meeting issue ‘The ecology and evolution of prokaryotic CRISPR-Cas adaptive immune systems’.

## 1. Introduction

Thanks to the unprecedented success of the Cas9, Cas12 and Cas13 endonucleases as genome editing tools, during the last decade, biochemical activities, structures, comparative genomic and at least some of the biological functions of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated proteins) systems and individual Cas proteins have been studied in exquisite detail [1–10]. The CRISPR-Cas are adaptive (acquired) immune systems that store the memory of encounters with foreign DNA, primarily that of mobile genetic elements (MGE), in unique spacer sequences derived from MGE and inserted into CRISPR arrays. The transcripts of the CRISPR spacers are used to recognize the cognate sequences and direct Cas nucleases to their unique target sites upon new encounters with familiar MGEs, resulting in the inactivation of the latter.

Like all defence mechanisms, CRISPR-Cas systems evolve in the regime of a perennial arms race with MGE, which results in the rapid evolution of some of

the *cas* gene sequences, primarily effector module components [11], and remarkable diversification of the gene composition and organization of the CRISPR-*cas* loci. This molecular diversity underlies the diversification of the molecular mechanisms of CRISPR-mediated defence [8,12,13].

Along with eukaryotic RNA interference (RNAi) and prokaryotic Argonaute-centred defence mechanisms, the CRISPR-Cas belong to nucleic acid-guided defence systems [14–18]. Arguably, among these mechanisms, CRISPR-Cas systems are the most biologically complex because, in contrast with the innate immunity mechanisms, such as those of the Argonaute-based systems and most of the forms of eukaryotic RNAi, but similarly to the piRNA branch of RNAi, CRISPR-Cas possess an integral capacity of creating immune memory and thus represent bona fide adaptive immunity [19–22].

Complete CRISPR-*cas* loci consist of a CRISPR array, that is, two to several hundred direct, often partially palindromic, normally exact repeats (25–35 bp each), separated by unique spacers (typically 30–40 bp each), and the adjacent cluster of multiple *cas* genes which are organized in one or more operons encoding both the adaptation and the effector modules, often along with accessory genes [13,23]. The CRISPR-Cas immune response includes three distinct but often intertwined stages: (i) adaptation, (ii) pre-crRNA (pre-CRISPR RNA) expression and processing, and (iii) interference.

During the adaptation stage, a complex of Cas proteins binds to a target DNA molecule and, typically, after encountering a distinct, short (2–4 bp) motif known as PAM (Protospacer-Adjacent Motif), introduces two double-strand (ds) breaks into the target DNA. The released segment, the protospacer, is then inserted between two repeats in the CRISPR array (most often, into the proximal repeat unit that immediately follows the leader sequence), so that it becomes a spacer [24,25]. The CRISPR array is then repaired by the cellular repair machinery, resulting in the duplication of the proximal repeat [26–28]. Some CRISPR-Cas systems employ an alternative mechanism of adaptation, namely, spacer acquisition from RNA (transcripts of a DNA genome of an MGE) via reverse transcription by a reverse transcriptase (RT) that is encoded in the CRISPR-*cas* locus and, in most cases, fused to the Cas1 protein [29,30].

At the expression-processing stage, the CRISPR array is typically transcribed into a single, long transcript, the pre-crRNA, that is processed to generate mature crRNAs by a distinct complex of Cas proteins, a dedicated processing nuclease (Cas6), a single large Cas protein or an external RNase [31,32].

At the final, interference stage, the crRNA that remains bound to the processing complex is employed as the guide (gRNA) to recognize the protospacer or a closely similar sequence in the genome of a virus or a plasmid which is then cleaved and inactivated by a Cas nuclease which can be either a component of the same effector complex or a separate Cas protein [33,34].

The brief description above is an over-simplified scheme that, out of necessity, misses many important details of CRISPR-Cas functions. Such details can be found in numerous recent reviews on different facets of CRISPR-Cas biology [1–9,24,25,31–34].

At both the structural and the functional levels, the CRISPR-Cas systems have a distinct modular organization [13,23]. The two principal components of the CRISPR-Cas systems are the adaptation and effector modules. In most of

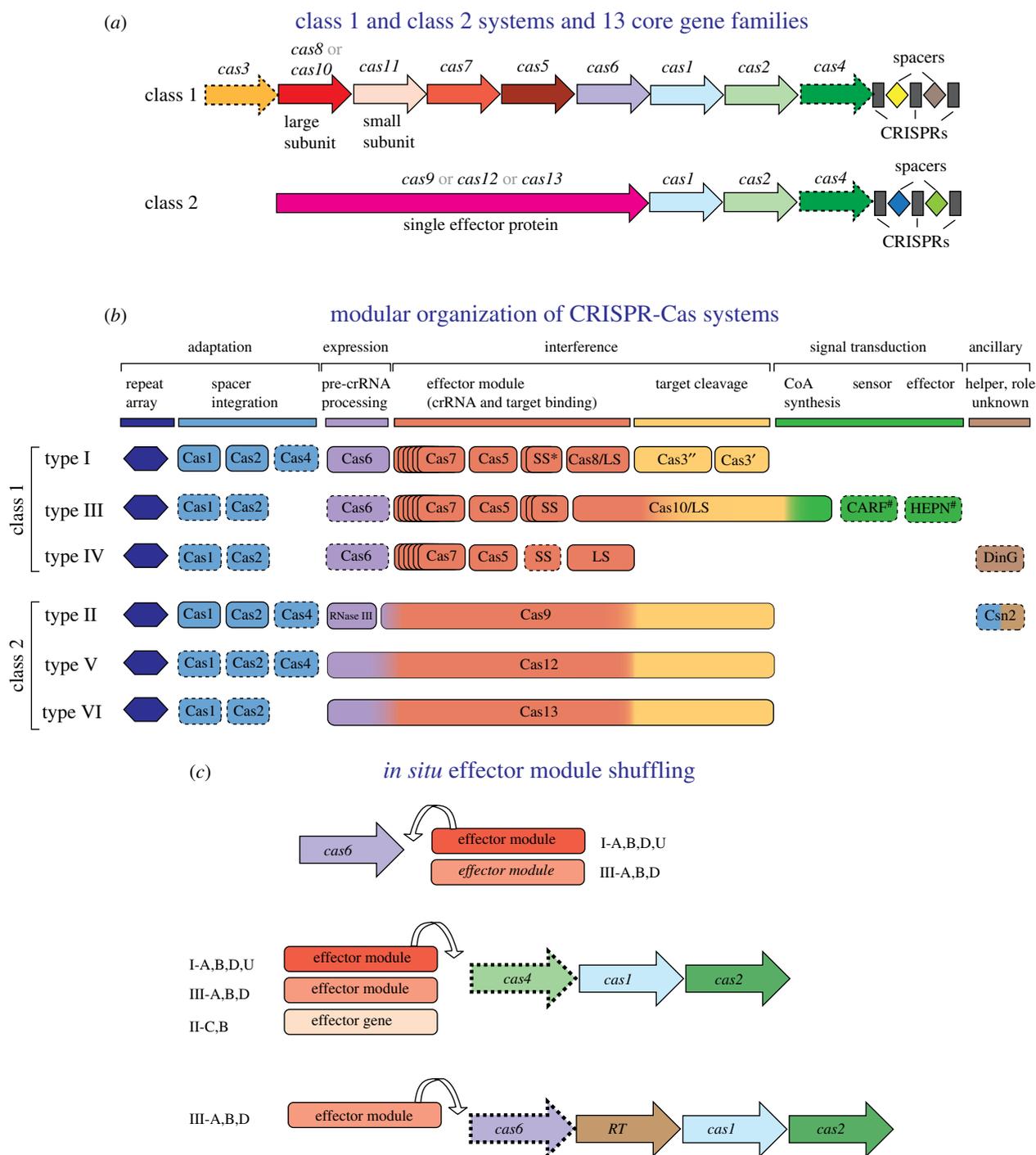
the CRISPR-Cas systems, the adaptation module consists of the Cas1 and Cas2 proteins, which form a complex in which Cas1 is the enzymatically active subunit, namely, the endonuclease (integrase) involved in the cleavage of both the source, protospacer-containing DNA and the CRISPR array, whereas Cas2 forms the structural scaffold of the complex [24,35–37]. In many CRISPR-Cas systems (see below), additional Cas proteins, such as Cas4, Cas3, Cas9 or RT, also contribute to the adaptation stage, in some cases forming fusions with Cas1 or Cas2 [30,38–40]. In contrast with the comparatively simple and uniform architecture of the adaptation module, the effector modules are highly diverse among CRISPR-Cas systems, and their variation forms the basis of the current CRISPR-Cas classification [8,13], which is described in the next section.

The extraordinary, compared with other defence systems in prokaryotes, complexity and diversity of CRISPR-Cas systems implies a complex evolutionary history. Phylogenomic studies have revealed a pervasive trend in CRISPR-Cas evolution, namely, the contributions of several classes of MGE to the ultimate origin and the subsequent diversification of the CRISPR-Cas systems, in particular the adaptation modules [4,13,23,41–43]. However, the origin of the most prevalent forms of the effector modules remains a much harder problem. In this article, we attempt to synthesize the available clues on the origins and evolution of different components of CRISPR-Cas systems along with the evidence of the reverse trend, that is, recruitment of CRISPR-Cas and their components by MGE.

## 2. Diversity, classification and evolutionary modularity of CRISPR-Cas systems

The CRISPR-Cas systems are a universal immune mechanism that, at least in principle, can adapt to defend the host from any MGE. Because of this universal adaptability, CRISPR-Cas systems do diversify as extensively as innate immune systems, such as restriction-modification modules, the ubiquitous and most abundant defence component in archaea and bacteria. Nevertheless, the Cas protein sequences and the genomic organization of CRISPR-*cas* loci display substantial diversity. All CRISPR-Cas systems are divided into two distinct classes, on the basis of the design principles of the effector modules. Class 1 systems have multisubunit effector complexes comprising several Cas proteins, whereas in class 2 systems, the effector is a single, large, multidomain protein [13] (figure 1). Classification of CRISPR-Cas systems is a complicated matter. There are no universal Cas proteins that could be used as phylogenetic markers, and even the phylogeny of the most evolutionarily conserved protein, Cas1, fails to adequately represent the relationships between CRISPR-Cas systems owing to the semi-independent evolution of different modules (see below). Therefore, the existing classification of CRISPR-Cas systems employs multiple criteria including signature *cas* genes, organization of the *cas* operons and phylogenies of conserved Cas proteins.

The two CRISPR-Cas classes are divided into three types each, types I, III and IV in class 1, and types II, V and VI in class 2; each type is characterized by distinct architectures of the effector modules that include unique signature proteins (figure 1). Each type is further classified into multiple subtypes that are distinguished by subtler differences in locus



**Figure 1.** Class 1 and class 2 CRISPR-Cas systems: key features, modular organization and module shuffling. (a) The general architectures of class 1 (multiprotein effector complexes) and class 2 (single-protein effector complexes) CRISPR-Cas systems. Genes are shown as arrows; homologous genes are shown by the same colour. Gene names follow the current nomenclature and classification [8,13]. (b) The principal building blocks of CRISPR-Cas system types. An asterisk indicates the putative small subunit (SS) that might be fused to the large subunit in several type I subtypes [13]. The # next to the CARF and HEPN domain labels indicates that other unknown sensor and effector domains can be involved in the signalling pathway. Dispensable genes are indicated by a dashed outline. The bottom panel schematically shows module shuffling in CRISPR-*cas* loci. (Online version in colour.)

organization and, also, often encode subtype-specific Cas proteins [8,12,13,43]. The mechanisms of pre-crRNA processing in class 1 and class 2 CRISPR-Cas systems notably differ. In class 1 systems, the maturation of crRNAs is catalysed by a dedicated complex of multiple Cas proteins that was first identified in subtype I-E and designated Cascade (CRISPR-associated complex for antiviral defence) [44–48]. The Cascade complex binds the pre-crRNA and recruits an additional Cas protein, Cas6 (or, on rarer occasions, Cas5), which is the nuclease directly responsible for processing. In type II systems, the prototype of class 2, processing is

catalysed by an external bacterial enzyme, RNase III, with the help of an additional RNA species, the *trans*-acting CRISPR (tracr)RNA, encoded within the CRISPR-*cas* locus [49–54]; tracrRNAs have been identified also in subtype V-B systems, although in this case, the cleavage enzyme remains uncharacterized [53–55]. In types V and VI, pre-crRNA processing is catalysed by a distinct, still incompletely characterized nuclease activity of the same large effector protein that is involved in target cleavage [56–59].

Major differences between class 1 and class 2 CRISPR-Cas systems are apparent also at the interference stage. In type I

systems, the processing complex containing the mature crRNA recognizes the protospacer sequence in the target and recruits an additional Cas protein, Cas3, which consists of a helicase domain that unwinds the target dsDNA and the nuclease domain directly responsible for the cleavage [60–62]. In type III, the nuclease involved in the target cleavage is a subunit of the processing complex itself; in this case, no helicase is involved but DNA cleavage requires primary cleavage of RNA transcripts of the target genome by a distinct CRISPR-associated RNase. In class 2 systems, cleavage is performed by the nuclease domain(s) of the large effector protein [49,50,52,54,59,63–69] (see more below).

The adaptation and effector modules of the CRISPR-Cas systems show pronounced autonomy not only with respect to functions and structure, but also evolutionarily (figure 1). The topology of the phylogenetic tree of Cas1, the key subunit of all adaptation complexes and repeat structures, are poorly compatible with the overall classification of the effector modules and phylogenies of individual effector proteins, apparently because of frequent module exchange among CRISPR-Cas systems of different types and subtypes [13]. Numerous examples of such exchange have been reported [30,70–72], including those that occur *in situ*, in the vicinity of several ‘attractor’ genes, namely, *cas6* and various combinations of the adaptation genes *cas1*, *cas2*, *cas4* and *RT* (figure 1). In most cases, the adaptation module, together with the cognate CRISPR array and/or *cas6* gene, remains fixed, as judged by the conservation in related genomes, whereas the effector module genes are shuffled [72]. This *in situ* exchange often involves not only effector modules from different variants of the same CRISPR-Cas subtype but also other subtypes or even other types (figure 1). Type III loci are especially prone to such module recombination: in particular, RT-containing type III adaptation modules have been shown to almost freely combine with effector modules from diverse subtypes of type III systems [30]. The evolutionary pressure to exchange effector modules is likely a consequence of an arms race against rapidly evolving viruses. Viruses encode many anti-CRISPR proteins (Acrs) which so far have been shown to target only components of effector complexes and, typically, show high specificity towards particular variants of CRISPR-Cas systems [73–80]. Exchange of effector modules could provide an escape route from Acrs. The partial evolutionary independence of the adaptation and effector modules is further corroborated by the presence, in numerous bacterial and archaeal genomes, of stand-alone adaptation and, even more often, effector modules [13,43,72]. Notwithstanding all these manifestations of the modularity of the CRISPR-Cas systems, it should be noted that the functional separation of the modules is but an approximation because some Cas proteins, in particular class 2 effectors, appear to be involved in all three stages of the CRISPR response [81,82].

### 3. Origin of the adaptation module and adaptive immunity in prokaryotes from a distinct group of transposons

Genomic surveys of *cas* genes show that Cas1, the endonuclease responsible for spacer integration into CRISPR arrays, is not always encoded within CRISPR-*cas* loci [23].

Examination of the genomic neighbourhoods of ‘stand-alone’ *cas1* homologues has led to the unexpected finding that these *cas1* genes actually are embedded within 12–18 kb regions of genomic DNA flanked by terminal inverted repeats (TIRs) and thus clearly resembling transposable elements [83]. These predicted transposons share two universal genes encoding, respectively, Cas1 and a family B DNA polymerase, and also encompass variable sets of additional genes, mostly encoding diverse nucleases and DNA-binding proteins containing helix–turn–helix (HTH) domains. The biochemical mechanisms of the reactions catalysed by Cas1 during spacer integration into CRISPR arrays and by integrases during transposon integration are closely similar, which naturally leads to the prediction that Cas1 is the integrase of the newly identified group of transposons; accordingly, these predicted transposons were named casposons and their predicted integrases were dubbed casposases [83,84]. Given the invariable presence of a DNA polymerase, casposons appear to be self-synthesizing transposons that direct their own replication during transposition via a copy-and-paste mechanism. This type of transposon so far has not been found in prokaryotes but is common among eukaryotes, many of which harbour polintons which also encode a B family DNA polymerase along with a retrovirus-type integrase that is unrelated to Cas1 [85–87]. The integrase activity of the casposase was promptly confirmed experimentally [88], and moreover, it has been shown that casposons and CRISPR spacers insert into similar target sites [89].

Although casposons are not among the most abundant classes of MGE in prokaryotes, comparative analysis of their gene organization revealed considerable diversity and resulted in the identification of four distinct casposon families which are integrated mostly into archaeal genomes as well as those of some bacteria [90]. Notably, a small group of casposons encode a predicted virus capsid protein, indicating that some of these elements are actually ‘casposviruses’, in a close analogy to the polintons which also possess capsid proteins and are predicted to form virions [91]. Transposition of casposons has not yet been demonstrated directly but comparative genomic analysis of many strains of the archaeon *Methanosarcina mazei* has led to the identification of clear signs of recent mobility, indicating that at least some of the casposons are active transposons [92].

The phylogenetic tree of the Cas1 family splits into two major branches, one of which includes the casposases and the other one consists of the CRISPR-associated Cas1 proteins [83]. Although, technically, the root position is unknown, this tree topology is compatible with a founding role of the casposase in the CRISPR-Cas evolution. The entire CRISPR adaptation module likely originated from a casposon which could have also contributed additional *cas* genes [41]. Although currently known casposons do not encode Cas2 (the key structural subunit of CRISPR adaptation complexes), some encode nucleases homologous to Cas4, a component of the adaptation module in several CRISPR-Cas subtypes, as well as additional nucleases [83,84]. The ancestral casposon configuration including a gene for a Cas2 homologue can be expected to appear in the growing microbial genome datasets. Additionally, the ancestral CRISPR repeats and the leader sequence could evolve from either the TIRs or a duplicated target site of the ancestral casposon [90].

The event that gave rise to the adaptation module and, concomitantly, prokaryotic adaptive immunity could have involved a chance insertion of a casposon into the vicinity of an ancestral innate immunity locus, followed by immobilization of the casposon and elimination of some of its genes, including the DNA polymerase [41]. The ancestral innate immunity system that gave rise to the CRISPR effector module might have functioned by directly engaging guide RNA derived from transcripts of foreign genomes, in an analogy to prokaryotic Argonaute-centred defence systems [16,18]. However, innate immunity systems homologous to CRISPR-Cas effector modules have not been so far identified in bacterial or archaeal genomes, and thus the ancestry of the effector module remains a hard puzzle [23,53]. In the next section, we consider some clues that might allow us to glean a solution.

#### 4. Ancestral class 1 effector modules: origin from a stress-response system?

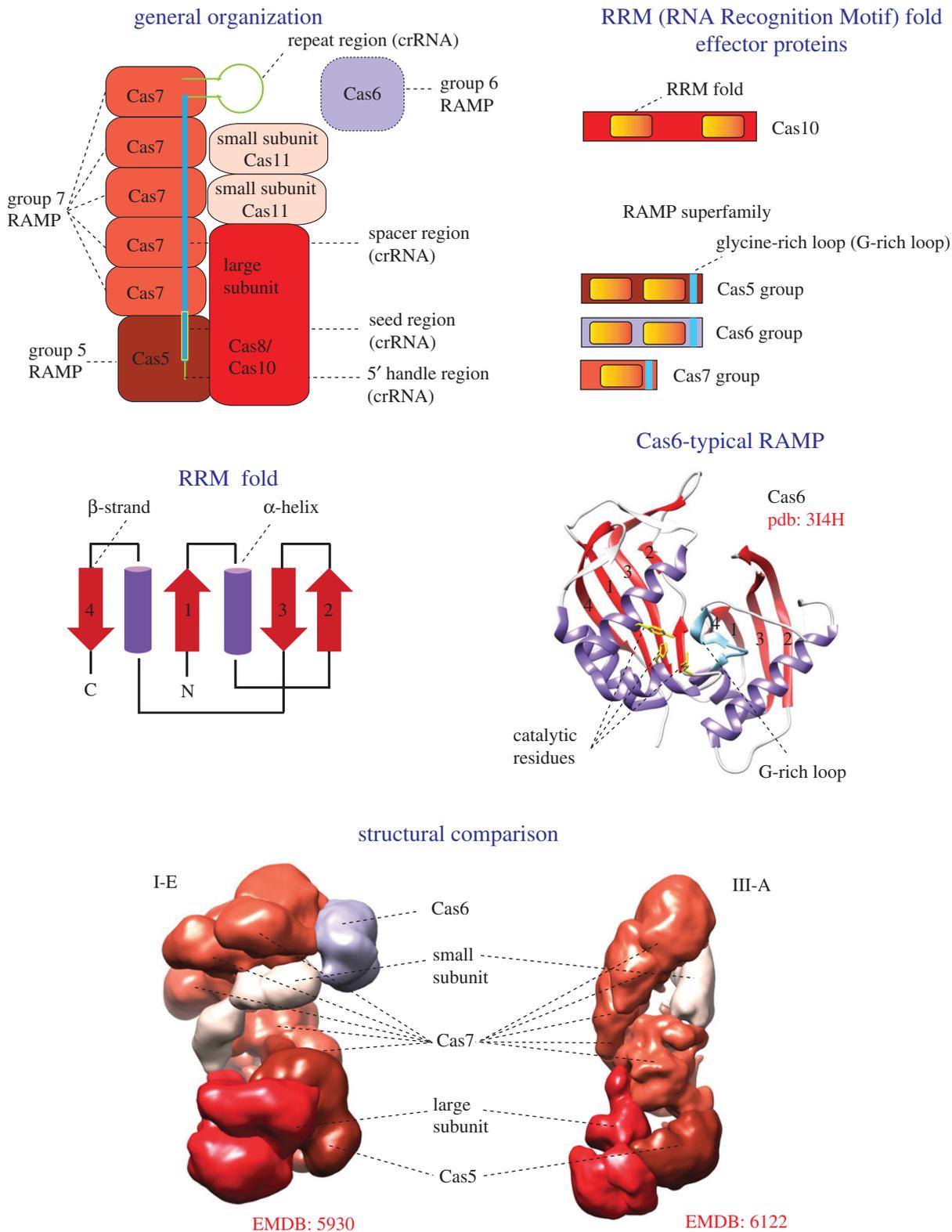
Despite limited direct sequence similarity, there is little if any doubt that the effector complexes of type I and type III that jointly compose the vast majority of class 1 systems share a common ancestry (figure 2). These complexes have strikingly similar overall architectures, and in both cases, the skeleton of the complex is formed by multiple copies of Cas7 protein, a member of the so-called RAMP (Repeat-Associated Mysterious Protein) superfamily [4,46,93,94]. Additionally, the effector complexes of both types contain a single copy of Cas5 protein, another RAMP superfamily member distantly related to Cas7, and the so-called large and small subunits (figures 1 and 2). Because of the generally fast evolution of *cas* genes and the resulting low sequence conservation [11], homology of the large and small subunits between types I and III could not be ascertained at the sequence level; however, the small subunits show significant structural similarity, which implies homology [95]. The large subunit of type III systems, Cas10, is a protein containing two RRM (RNA recognition motif) domains [96,97]. One RRM domain shows highly significant similarity to the palm domain, the catalytic domain of a broad variety of RNA and DNA polymerases and nucleotide cyclases [98,99]. This palm domain is predicted to be an active enzyme, whereas the second RRM domain is inactivated. The type I large subunit is the Cas8 protein, which is highly diverged in sequence even among different type I subtypes and shows no detectable sequence similarity, and only tenuous structural similarity, at best, to Cas10 [23]. This uncertainty notwithstanding, the conservation of the overall structural organization and the Cas7–Cas5 scaffold provides sufficient evidence for the common ancestry of the effector complexes between type I and type III (figure 2). Furthermore, considering that type I and type III systems together represent about 90% of all detected CRISPR-Cas loci, and moreover, amount to 100% among archaea [13], it is not much of a stretch to propose that the common ancestor of the type I and type III effector modules is also the ancestral form of this module for CRISPR-Cas systems in general [23,53]. By extrapolation, the ancestral effector module can be inferred to have contained, at least, the RAMP-based scaffold shared by the modern ones.

Can we reconstruct the evolution of the effector module beyond the common ancestor of type I and type III? A key

clue seems to be that, in type III, the entire effector complex, with the sole exception of the small subunit, is composed of domains with the same structural fold, the RRM fold, which is topologically identical to the widespread ferredoxin-like fold [19]. It seems likely, therefore, that the complex originally evolved by serial duplications and fusions of the RRM domains followed by subsequent extensive divergence (figures 2 and 3) [23]. The direction of evolution appears clear: from the RRM endowed with the polymerase–cyclase activity, as in Cas10, to the common ancestor of the RAMP superfamily which acquired a shared structural feature identifiable in all three groups of RAMPs, the glycine-rich loop (figures 2 and 3). Furthermore, structural comparisons suggest that small subunits from both type I and type III systems are homologous and similar to the C-terminal four-helix bundle domain of Cas10 [95,100]. Thus, it seems plausible that, following a duplication, fission of an ancestral Cas10-like protein could have given rise to the ancestors of both RAMPs and the small subunits (figure 3). Considering that catalytically active RAMPs of all three major families, Cas5, Cas7 and Cas6, are known, an ancestral RAMP also could have been an RNase (figure 3). However, independent origins of RNase activity cannot be ruled out, especially considering that different Cas6 protein have different residues involved in catalysis (figures 2 and 3) [23,101–103].

Recent discoveries suggest clues also to the hardest puzzle, the organization and possible functions of the putative stand-alone ancestor of the CRISPR-Cas effector module. Most of the type III CRISPR-Cas systems encode proteins containing one or both characteristic domains: CARF (CRISPR-Associated Rossmann Fold), a (predicted) nucleotide-binding domain [104], and HEPN (Higher Eukaryote and Prokaryote Nucleotide-binding domain), an RNase that is primarily involved in various defence functions in both prokaryotes and eukaryotes [105]. The widespread examples of CARF–HEPN fusion in type III systems include Csm6 and Csx1; some type III systems instead include proteins that consist of a CARF domain fused to an unrelated nuclease or to a DNA-binding, HTH domain [104]. Years after the enzymatic activities of Cas10 and the HEPN domains as well as the nucleotide-binding capacity of the CARF domains, which implies allosteric regulation of the activity of HEPN (or other nucleases), have been predicted by computational methods [98,104,105], their functions in CRISPR-Cas have remain obscure, although it has been speculated that these proteins are linked through oligonucleotides synthesized by Cas10 [106].

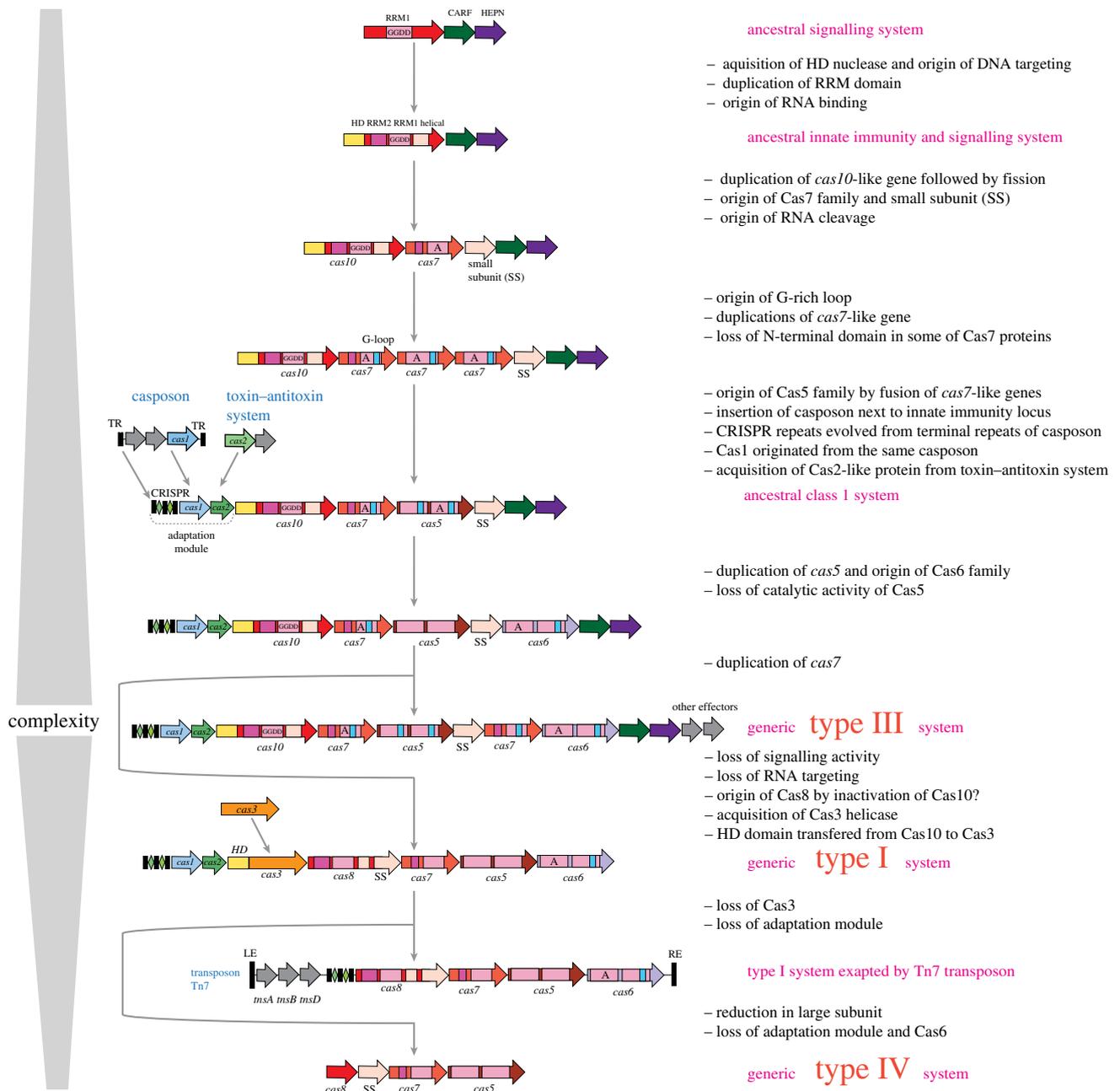
One of the most notable findings in the CRISPR field in the last few years has been the experimental validation of this hypothesis. Two laboratories have independently demonstrated that a dedicated signalling pathway activated by target recognition is central to the immune function of type III CRISPR-Cas systems. Specifically, target binding by the crRNA–effector complex allosterically activates the polymerase activity of Cas10 which catalyses the synthesis of cyclic oligoA (cOA). The produced cOA molecules are bound by the CARF domain of the Csm6 protein (and by inference, other CARF domain-containing proteins), resulting in allosteric activation of the promiscuous RNase activity of the HEPN RNase domain of Csm6 and degradation of both the target RNA and other RNA molecules [107,108]. The outcome of this indiscriminate RNA degradation is thought to be



**Figure 2.** Key features and general organization of class 1 CRISPR-Cas systems. The figure schematically shows the general organization of a class 1 effector complex. The colouring of the shapes corresponds to the colour code for *cas* genes in figure 1. All proteins of class 1 effector complex that contain an RRM domain are schematically shown in a separate panel. The topology diagram of the RRM fold is schematically shown, with numbers corresponding to the typical order of  $\beta$ -strands in the fold. The Cas6 structure with two RRM fold domains, which are numbered according to the topology diagram, is shown as the ribbon diagram. The structural comparison of cryo-electron microscopy models demonstrates a striking similarity of effector complex organization of types I and III, despite the absence of significant sequence similarity between the corresponding subunits. The Electron Microscopy Data Bank (EMDB) codes are indicated for each structure. crRNA, CRISPR RNA; RAMP, Repeat-Associated Mysterious Protein. (Online version in colour.)

the induction of dormancy or programmed cell death (PCD) in response to infection, which could work as a ‘contingency plan’ when immune response fails. Recently, it has been shown that this cOA–Cas10 signalling pathway is tightly

regulated by cOA hydrolysis which is catalysed by CARF domains of a distinct family [109,110]. This regulatory circuit might be part of the microbial cell’s sensor that ‘makes the choice’ between immunity and PCD [111,112].



**Figure 3.** Origin and evolution of class 1 CRISPR-Cas systems. The figure depicts a hypothetical scenario of the origin of class 1 CRISPR-Cas from an ancestral signalling system and its subsequent evolution yielding the extant type III and type I systems, as well as reductive evolution that produced type IV systems and minimalist variants of type I system recruited by Tn7 transposons. The key evolutionary events are described to the right of the images. ‘GGDD’, a key catalytic motif of the cyclase/polymerase domain of Cas10; ‘A’, catalytically active RRM domain of a RAMP protein; RE, LE, right and left end, respectively; TR, terminal repeats. (Online version in colour.)

The discovery of the cOA–Cas10 signalling pathway suggests a plausible possibility which, in fact, has been discussed previously solely on the basis of protein domain and gene neighbourhood analysis, that the ancestor of the CRISPR-Cas effectors was a stress-response system that triggered PCD upon activation by an alarmone, such as cOA [106,111] (figure 3). A potential candidate for the role of such a signal transduction system with a defence function has been identified in the genomes of several bacteria, namely, a protein that contains a single palm domain homologous to the active polymerase–cyclase domain of Cas10 [106]. The potential effector coupled to this predicted signalling enzyme remains unknown, although, in a few cases, both CARF and HEPN domains are fused to the minimal *cas10*-like gene [106]. Thus, the tight link between the three

domains apparently emerged very early during evolution. Under the proposed model, the ancestral signalling system evolved to become an innate immunity mechanism as a result of the acquisition of an HD DNase domain and serial duplications, fusions and diversification of the RRM domains which yielded a complex endowed with the RNA-binding and -processing capacity (figure 3). The complexity of the organization of many type III loci encoding an active cOA synthetase and the growing number of ancillary proteins that are likely to be involved in the cOA signalling pathway as effectors imply a key role of the cOA pathway in defence mechanisms. Notably, in this model, the function of innate, and eventually, adaptive immunity evolved from a primordial PCD/dormancy induction form of defence (figure 3). These type III systems with dual functions are the most

complex among all CRISPR-Cas systems. The loss of the signalling capacity through inactivation of the palm domain of Cas10 results in the loss of ancillary proteins, specialization on DNA targeting and the overall reduced complexity of the effector complex organization. This is the proposed course of evolution for type I systems as well as some type III variants (figure 3). Further reduction and loss of the DNA cleavage capacity, most probably, gave rise to the type IV CRISPR-Cas systems and other derivative CRISPR-Cas variants that were recruited by mobile elements to assist their replication and/or transposition (figure 3 and see below).

## 5. Class 2 effectors: multiple cases of nuclease recruitment from mobile genetic elements

Class 2 effectors are radically different from those of class 1 in that all the effector functions are concentrated in a single protein [13]. In-depth analysis of the protein sequences of class 2 effectors revealed a striking feature: they are all homologous to nucleases encoded by different classes of MGE [42,43]. All type II and type V effectors (Cas9 and Cas12 proteins, respectively) share a domain that belongs to the RuvC-like endonuclease family which belongs to the RNase H fold which is common to a great variety of nucleases and some other proteins [113] (figure 4). However, the sequence similarity between the RuvC-like domains of Cas9 and Cas12, and even between different subtypes within each type is low such that these proteins can be recognized as homologues only by highly sensitive sequence profile searches or structural comparisons [43,53]. Outside of the RuvC-like domain, the sequences of Cas9 and Cas12 show no similarity to each other and appear not to be homologous [43,53]. The structures of several Cas9 proteins [51,114–116], Cas12a (Cpf1) [66,67] and Cas12b (C2c1) [54,66] complexed with the guide RNA, target DNA and, in the cases of Cas9 and Cas12b, tracrRNA have been reported. All these effector proteins share similar size and overall shape which is a bilobed, ‘jaw-like’ structure accommodating the target DNA and the gRNA between the lobes. However, beyond the RuvC-like domains, the structures cannot be superimposed [117]. The RuvC-like domains of Cas9, Cas12a and Cas12b contain inserts, in similar but not identical positions, that represent non-homologous domains, respectively, the HNH family nuclease domain in Cas9 and unique, non-enzymatic domains that facilitate target cleavage in the Cas12 proteins [54,66,67,118].

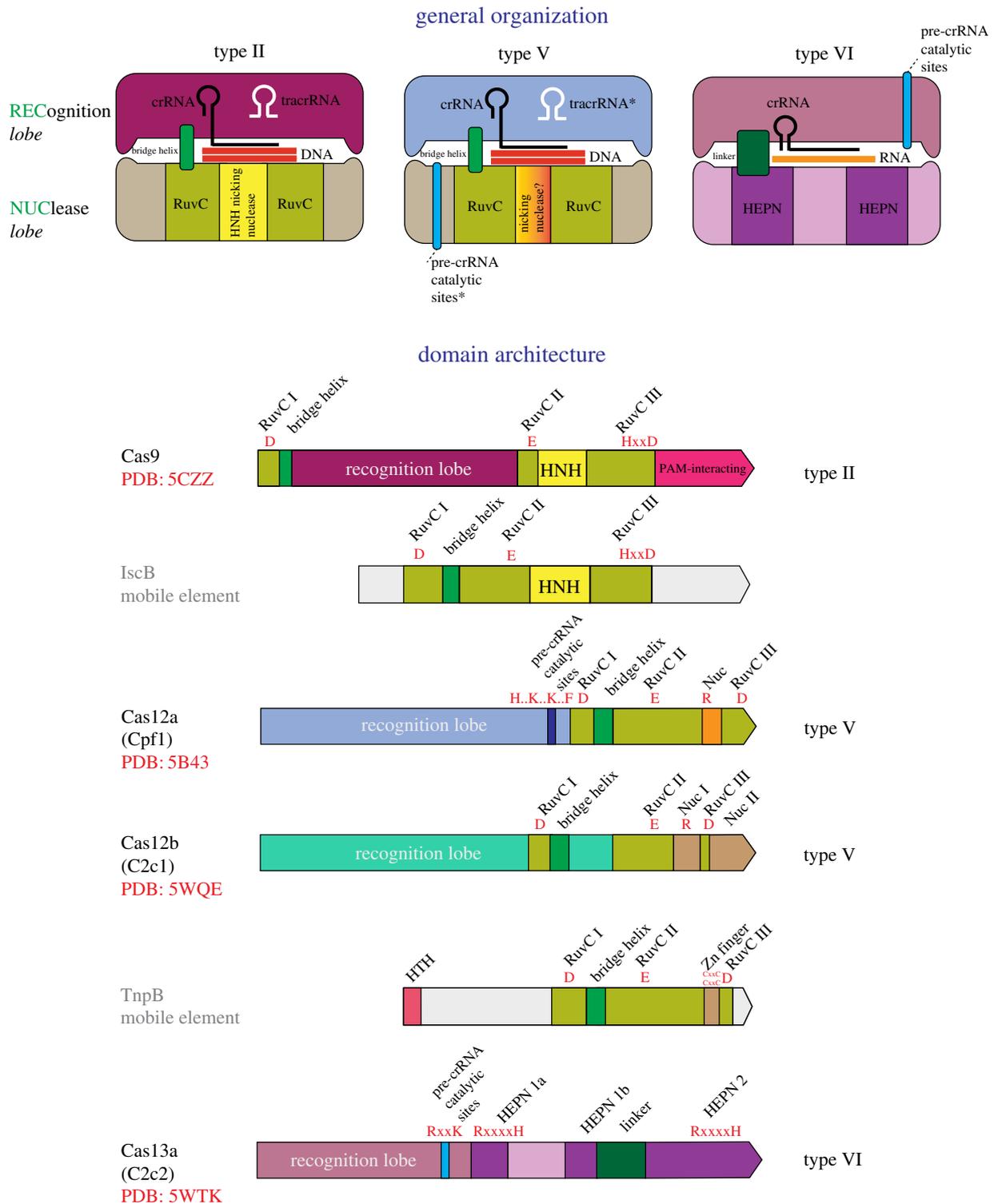
An essential clue to the evolutionary origin of Cas9 and Cas12 was the observation that, apart from the members of the same family of Cas protein, both show the highest sequence similarity to the RuvC-domain-containing TnpB proteins of the IS605 and other related families of transposons [42,53]. The *tnpB* genes are among the most abundant genes in bacterial and archaeal genomes, and are encoded either by autonomous transposons, which additionally encode a transposase (TnpA), or more frequently, by non-autonomous transposons, including the eukaryotic Fanzor elements, in which TnpB is the only protein product [119]. The role of TnpB in transposons remains unclear, given that this protein is not required for transposition and, actually, appears to downregulate it [120], but the perfect conservation of the

RuvC-like endonuclease catalytic sites in most TnpB sequences indicates that these proteins are active nucleases.

Unexpectedly, the effectors of type II and different subtypes of type V (Cas9 and Cas12a, 12b, 12c, respectively) showed the highest similarity to different groups of TnpB proteins, suggesting independent origins from the same ancestral protein family [53]. Because of the low sequence conservation, no reliable phylogenetic trees could be constructed for the type II and type V effectors together with the TnpB proteins. However, the ancestry of Cas9 could be readily traced to a distinct family of transposons (denoted ISC, after Insertion Sequences Cas9-related) that are found primarily in Cyanobacteria and encode IscB proteins which share a signature domain architecture with Cas9, namely, the insertion of an HNH endonuclease domain into the TnpB (RuvC-like) domain [42,121] (figure 5).

The likely path of evolution from TnpB to the type V effectors became more tractable with the identification of a distinct variety of putative CRISPR-Cas systems that were designated subtype V-U (after uncharacterized). Subtype V-U loci lack adaptation modules and typically consist solely of TnpB homologues encoded next to CRISPR arrays [43]. The putative V-U effectors are much smaller than Cas9 or Cas12 but are similar in size or only slightly larger (400–600 amino acid residues) than typical, transposon-encoded TnpB proteins. In contrast with Cas9 and Cas12, the TnpB homologues encoded in the V-U loci are strongly similar to the transposon-encoded TnpB. Accordingly, the V-U proteins could be included in robust phylogenetic trees which convincingly support independent origins of (at least) five distinct groups of putative V-U2 effectors from different TnpB subfamilies [43]. The experimental demonstration of the functionality of the type V-U systems is pending. However, the evolutionary conservation of five distinct variants of subtype V-U in diverse bacteria, the finding that the spacer sequences are completely different even in closely related V-U loci and the presence of multiple phage-specific spacers jointly show that at least some V-U variants are functional CRISPR-Cas systems [43].

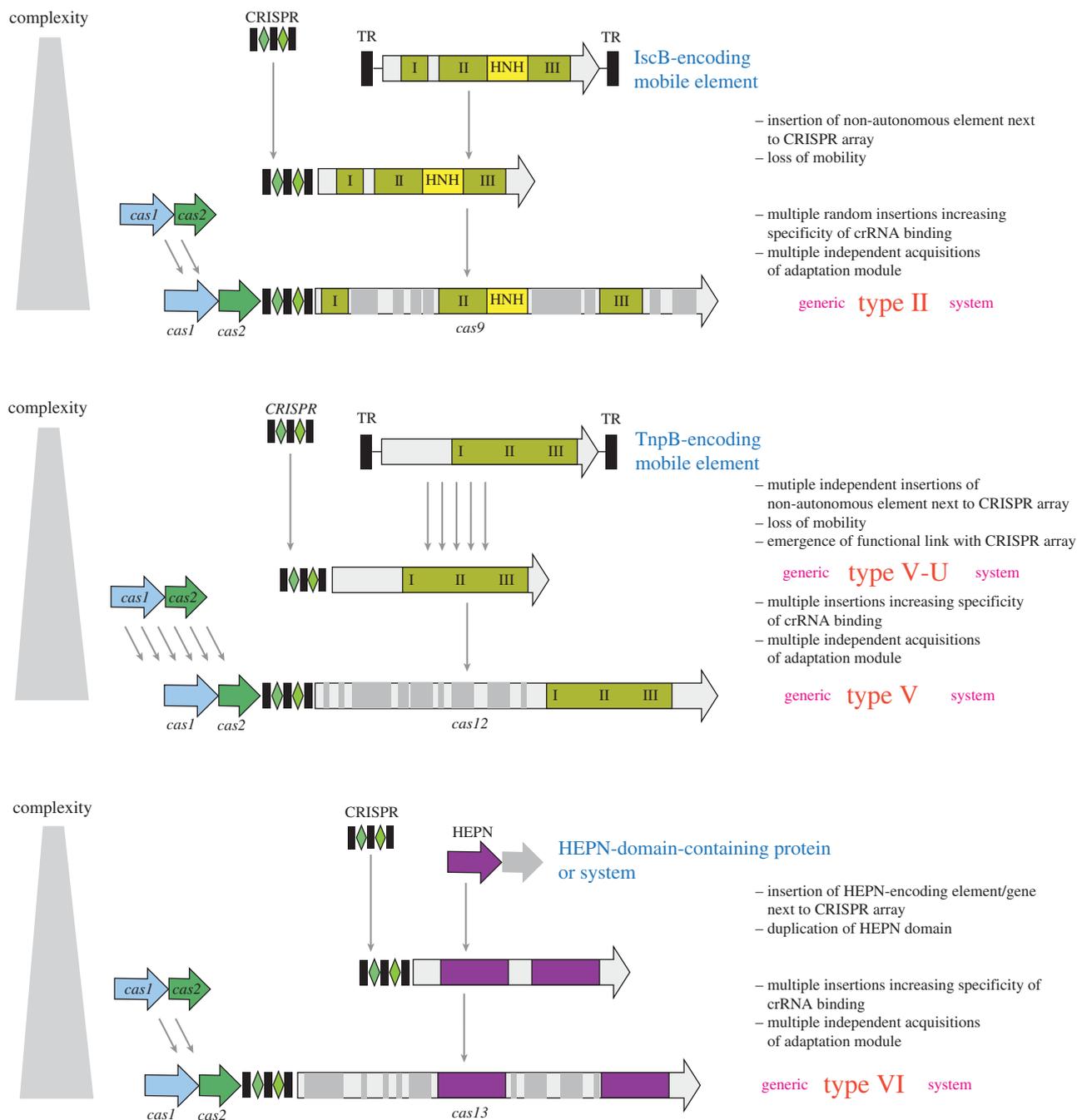
The close similarity between the predicted V-U effectors and transposon-encoded TnpB proteins implies that subtype V-U represents recently evolved, ‘baby’ CRISPR-Cas systems. The evolutionary scenario for type II and type V systems (figure 5) starts with random insertion of the non-autonomous TnpB-encoding transposons next to CRISPR arrays. Given the enormous abundance of these transposons in bacterial and archaeal genomes, such random insertions can be expected to occur frequently, and indeed, apart from the evolutionarily conserved V-U variants, a number of apparently spurious juxtapositions of CRISPR arrays and *tnpB* genes have been detected [43]. The subsequent evolution would involve parallel, independent ‘maturing’ of the effectors via acquisition of additional domains the sources of which remain obscure and might have involved both internal duplications and recombination [43] (figure 5). The gained portions of the proteins are unrelated between different subtypes, but the convergent outcome is the emergence of an effector protein that is large and flexible enough to accommodate the complex of the crRNA with the target DNA (figure 5). It should be emphasized that not only the effector modules but also the adaptation modules of class 2 systems were acquired independently from different class 1 variants (figure 5) [43].



**Figure 4.** Key features, general organization and domain architectures of class 2 systems. Schematic of the complexes of effector proteins, with the target DNA or RNA, guide RNA and (for type II) tracrRNA shown on the top of the figure, and the domain architectures of the effector proteins depicted underneath. The catalytic residues of the effector nuclease domain and, for Cas12a and Cas13a, the residues shown to be required for pre-crRNA processing are indicated in red. The Protein Data Bank (PDB) codes are included for proteins with solved structures. HTH, helix–turn–helix DNA-binding domain. The tracrRNA, the pre-crRNA processing catalytic sites and the nicking, target strand-cleaving nuclease of the type V effectors are denoted by asterisks to indicate that they are each present only in subsets of the type II and type V effectors. The catalytic amino acid residues of the target-cleaving and pre-crRNA processing nucleases are shown in red. The small blue boxes show the approximate location of pre-crRNA processing nuclease domain. I, II, III are the distinct amino acid motifs that jointly compose the catalytic site of the RuvC-like nuclease. In the motif signature, ‘x’ stands for any amino acid, and ‘.’ indicates that the catalytic residues are separated by a small, variable number of non-conserved residues. Adapted from [8], with permission. (Online version in colour.)

In addition to two distinct classes of MGE, casposons and TnpB-encoding transposons, microbial toxin–antitoxin (TA) modules seem to have been important contributors to the evolution of CRISPR-Cas systems. The most common type

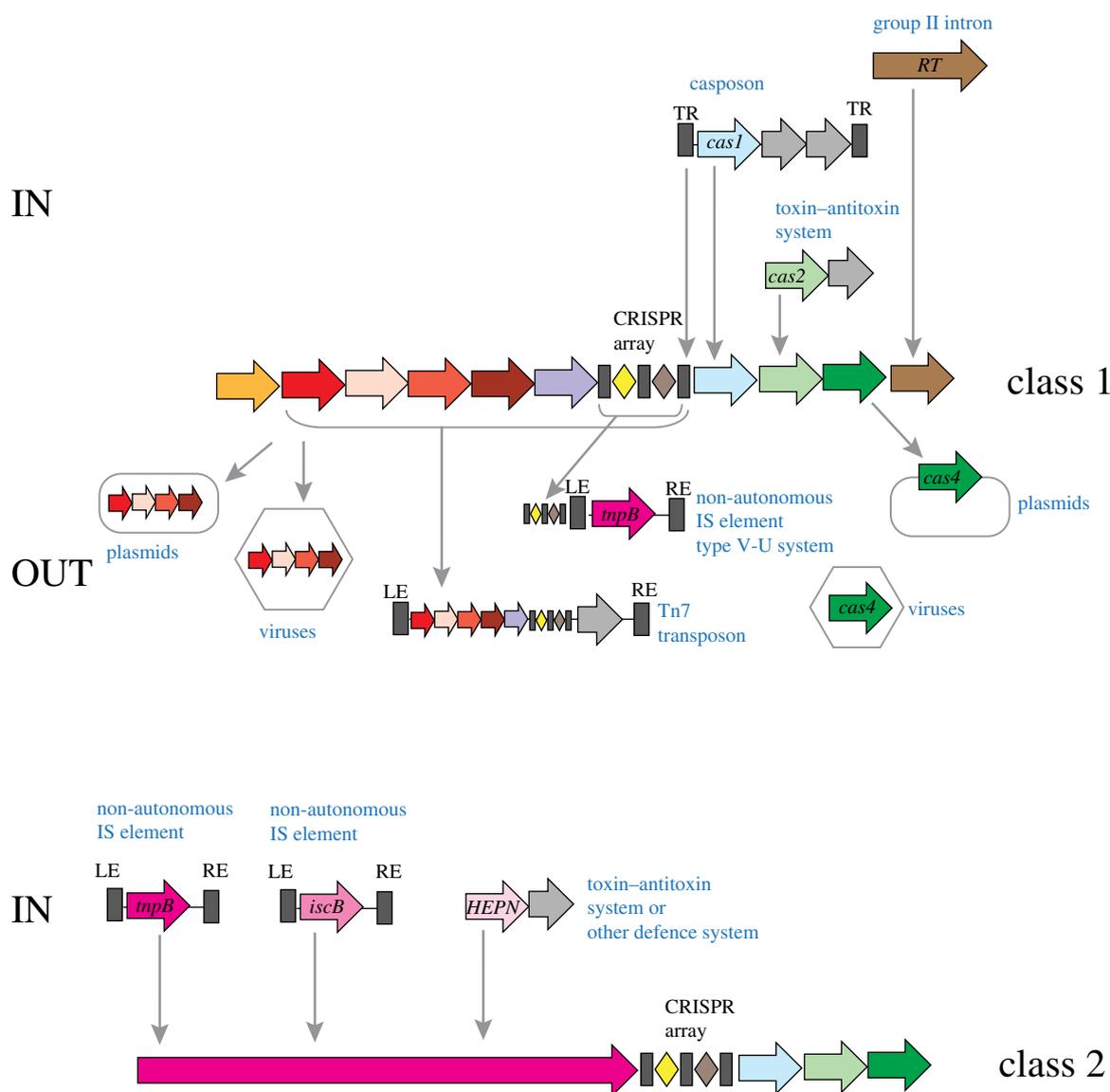
II TA modules consist of two proteins, a toxin and an anti-toxin, of which the antitoxin is substantially less stable than the toxin and is eliminated by proteolysis under stress, resulting in toxin activation [122–126]. The most common toxin



**Figure 5.** Origin of the class 2 CRISPR-Cas effectors from MGE. The figure depicts a hypothetical scenario of the origin of class 2 CRISPR-Cas from non-autonomous transposons, for type II and type V systems, and from a defence system (toxin–antitoxin module) for type VI systems. IscB and TnpB are the inferred ancestors of the type II (Cas9) and type V (Cas12) effectors, respectively. Inserts that could have contributed to increased specificity and efficiency of the effectors are shown by grey rectangles of variable size. I, II and III are the distinct amino acid motifs that jointly compose the catalytic site of the RuvC-like nuclease. TR, terminal repeats. (Online version in colour.)

domains are the interferases, RNases that indiscriminately cleave mRNAs inside the ribosome, resulting in microbial dormancy or cell death [127,128]. The interferases belong to several unrelated protein families including HEPN, RelE and VapD, a distinct variant of the RRM fold [105,124]. Although the TA modules lack mechanisms of active mobility, they nevertheless qualify as MGE because they are typically transferred on plasmids and are ‘addictive’ to the host cells, which die if they do not receive the TA-carrying plasmid upon segregation owing to the difference in stability between the toxin and antitoxin proteins [123,125,126]. In addition to plasmids, many TA loci are found in bacterial and archaeal chromosomes, and are thought to induce dormancy or PCD as an ‘altruistic’ defence strategy [129,130].

At least two unrelated classes of TA appear to have contributed to the evolution of CRISPR-Cas. The structural subunit of the adaptation complex, Cas2, belongs to the VapD family of interferases [19]. The interferase catalytic site is intact in the majority of the Cas2 proteins but is disrupted in some and is not required for adaptation [36]. Thus, the role of the demonstrated nuclease activity of Cas2 [131–133] in CRISPR-Cas function remains uncertain, and it cannot be ruled out that Cas2 functions as a toxin inducing dormancy or PCD when the immune function of CRISPR-Cas fails [112,134]. As discussed above, the *cas2* gene might have become a component of CRISPR-Cas systems via a casposon that gave rise to the adaptation module (figures 3 and 6). Given that some casposons encode also nucleases



**Figure 6.** 'IN and OUT': exchange of components between CRISPR-Cas systems and MGE. The figure shows a hypothetical scenario of *cas* gene acquisition by evolving CRISPR-Cas systems from MGE (IN) and by MGE from CRISPR-Cas systems (OUT). Genes are shown by arrows. The colouring corresponds to distinct *cas* genes and is the same as in figure 1. Grey arrows denote any genes that are considered to be unrelated to CRISPR-Cas. Specific acquisition events are shown for class 1 and class 2 systems separately. Arrows indicate the inferred direction of the gene flow. HEPN, RNase of the HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding domain) superfamily; RE, LE, right and left end, respectively; RT, reverse transcriptase; TR, terminal repeats. (Online version in colour.)

homologous to Cas4 [83,90], this Cas protein that is part of the adaptation module in several CRISPR-Cas subtypes also could originate from that ancestral casposon.

As also discussed above, the HEPN domain, the other abundant toxin RNase, is present in many Cas proteins in type III and in all type VI systems. In type III systems, the HEPN domains present in accessory proteins, such as Csm6 and Csx1, are the endpoint RNA-cleaving effectors of the cOA-Cas10 signalling pathway and likely originate from the ancestral signal transduction module (see above; figure 3). The type VI effectors (Cas13) are large proteins containing two diverged HEPN domains that are both required for target RNA cleavage and the promiscuous RNase activity that is activated by target binding [43,53,59,68,69]. The presence of two HEPN domains is a unique signature of the type VI effectors that testifies to their common origin despite the extreme divergence of the HEPN domain sequences and the lack of detectable sequence similarity outside of these domains. The extremely low sequence conservation typical of the HEPN domains precludes reliable phylogenetic

analysis and thus confident identification of the specific ancestors. Thus, it is unclear whether the HEPN domains in type VI effectors originate from type III HEPN-containing proteins or, independently, from toxins. Regardless, a 'maturing' path of evolution resembling that proposed for the type II and type V effectors appears likely for Cas13 (figure 5).

## 6. Origin of reverse transcriptase-containing adaptation modules for group II introns

A large subset of type III CRISPR-Cas systems, in addition to the regular process of spacer capture from foreign DNA, are capable of acquiring spacers from RNA (typically, transcripts of an invading DNA genome) that is reverse-transcribed by a CRISPR-associated RT [29]. As pointed out above, the RT is a component of a distinct variety of adaptation modules that have been shown to combine promiscuously with diverse variants of type III effector modules or found in a stand-alone form, adjacent to a CRISPR array [30]. In these

adaptation modules, the RT is often fused to Cas1 or, alternatively, is encoded by a gene adjacent to *cas1*. Phylogenetic analysis of the RT superfamily shows that most of the CRISPR-associated RTs form a strongly supported clade that is affiliated with the RTs of group II introns [30,135,136]. Thus, the RT–Cas1 fusion represented in diverse type III loci appears to have emerged at a single point in evolution, conceivably, as a result of a random insertion of a group II intron into a type III CRISPR–*cas* locus (figure 6). There is a clear, striking parallel between this evolutionary scenario and that for the origin of type II and V effectors from TnpB-encoding transposons.

## 7. Derived CRISPR-Cas variants: reductive evolution and exaptation for non-defence functions

In addition to the diversification of the immune functions, an emerging trend in the evolution of CRISPR-Cas systems is the emergence of derived, defective variants that lose the adaptation and interference capacities, and are exapted for roles other than adaptive immunity. In particular, such minimalist CRISPR-Cas variants are carried by transposons and plasmids [137] (figures 3 and 6). The most common of these is a minimal version of subtype I-F that is encoded by a large family of Tn7-like transposons; smaller groups of Tn7-like transposons encode similarly degraded subtype I-B systems [138]. Phylogenetic analyses of both Tn7 genes and Cas7, the most highly conserved protein in these minimalist CRISPR-Cas systems, identified a single event of subtype I-F capture by a transposon and two independent events of subtype I-B acquisition [138]. All these Tn7-encoded CRISPR-Cas variants lack both the adaptation module and the Cas3 protein that is required for target cleavage by type I systems and, accordingly, are incapable of either adaptation or interference. However, they encompass all the subunits of the pre-crRNA processing complex and therefore can be inferred to generate mature crRNAs and recognize the target DNA (figure 3). The Tn7-encoded CRISPR-Cas loci encompass short CRISPR arrays, some of which contain spacers that target plasmids, bacteriophages sharing hosts with the respective transposons or chromosomal sequences adjacent to integration sites. The transposon-encoded CRISPR-Cas systems remain to be studied experimentally. Nevertheless, the predicted ability of these systems to recognize but not to cleave cognate targets suggests the intriguing possibility that they facilitate insertion of transposons into MGE by generating R-loops at the target sites [138]. In evolutionary terms, the transposon-encoded CRISPR-Cas systems clearly are derived forms that evolved from the respective complete systems.

Type IV CRISPR-Cas systems represent another minimalist variant unrelated to those carried by Tn7-like transposons. Type IV loci are typically carried by plasmids and, in some cases, by prophages of diverse bacteria. Analogous to the transposon-encoded type I variant, type IV systems, with a few exceptions, lack an adaptation module and consist of *cas5*, *cas7* and *cas8* genes, and in certain cases, also *cas6*, along with an additional gene, which in different type IV variants encodes either a DinG family DNA helicase or an uncharacterized small protein [8,13]. Type IV loci rarely

include CRISPR arrays and accordingly can be predicted to use *in trans* arrays located in another region of the same plasmid away from or on the host chromosome. As with the transposon-encoded variants, the functions of type IV CRISPR-Cas systems remain obscure but, given their almost exclusive localization on plasmids, it should be expected that they facilitate maintenance and/or enhance the mobility of plasmids via as yet unknown mechanisms.

A recent systematic screening of microbial genomes for CRISPR-linked genes [139] has led to the discovery of an apparently defective variant of I-E systems that lacks the *cas3* gene and, by implication, cannot cleave targets. Instead, these loci encode an NTPase of the STAND (Signal Transduction ATPases with Numerous Domains) superfamily [140] which is implicated in signalling processes, possibly stress-induced PCD.

The subtype V-U loci discussed above also constitute a variety of ‘minimal’ CRISPR-Cas systems. One of the five distinct groups within subtype V-U, V-U5, encompasses a TnpB homologue that is predicted to be inactivated as a result of the replacement of the catalytic amino acid residues in the RuvC-like nuclease domain [43]. Thus, this system is predicted to perform functions that do not involve target cleavage.

Taken together, these examples show that the reductive evolution of CRISPR-Cas systems leading to their exaptation (recruitment) for non-defence (or at least, not involving target cleavage) functions occurred on multiple independent occasions. So far, none of these functions has been explored experimentally, so the study of these systems appears to be a wide-open research direction that is bound to yield new insights into microbial physiology. More instances of defective CRISPR-Cas systems can be expected to emerge with the advances of genomic and metagenomics, particularly, in MGE genomes.

## 8. The emerging synthesis on CRISPR-Cas evolution

The CRISPR-Cas systems are highly complex molecular ensembles, and as such, undoubtedly are products of a complicated succession of evolutionary events. Moreover, as with all complex systems, the spectre of irreducible complexity looms: a satisfactory evolutionary scenario is expected to account for the functionality of intermediate stages [141–144]. Perhaps, surprisingly, we believe that the findings discussed above provide enough clues for a plausible overall scenario (figures 3, 5 and 6).

The single over-arching theme of CRISPR-Cas evolution is the evolutionary entanglement between these systems of microbial adaptive immunity and various types of MGE. Strikingly, at least four unrelated MGE varieties have contributed to CRISPR-Cas evolution: (i) casposons that gave rise to the adaptation module, (ii) group II introns that donated the RT to a distinct variety of type III adaptation modules, (iii) non-autonomous IS605-like transposons, the ancestors of type II and type V effectors, and (iv) a TA module that apparently contributed Cas2.

The evolution of class 2 CRISPR-Cas systems clearly involved multiple acquisitions of ancestral MGE genes encoding nucleases that subsequently evolved into CRISPR-Cas effectors. In all likelihood, these genes were captured as a

result of chance insertion of the respective MGE into pre-existing CRISPR-*cas* loci or next to orphan CRISPR arrays [43,137]. A puzzling aspect of this part of CRISPR-Cas evolution is the switch of the pre-crRNA processing from the Cas6-mediated mechanism characteristic of class 1 to the effector-catalysed and tracrRNA-dependent mechanisms in class 2. The tracrRNA which is required to recruit RNase III for processing apparently evolved on multiple occasions in different type II and type V systems [145], suggesting that the autonomous, effector-dependent processing is ancestral in class 2. The provenance of this mechanism remains an enigma on which the detailed study of the ancestral MGE-encoded nucleases might shed light.

Importantly, the evolutionary link between CRISPR-Cas and MGE is a two-way street: complete CRISPR-Cas loci, their reduced versions or individual components were repeatedly recruited by various MGE and adopted for antidefence as well as, apparently, for other, still uncharacterized functions.

When the evolution of class 1 effector modules is considered, the second key theme comes up, namely, serial duplication of ancestral genes followed by extreme diversification (figure 3). The ancestral unit is the RRM domain. The expansive RAMP superfamily undoubtedly evolved via a series of RRM duplications. We postulated that the founder of this superfamily also emerged through a duplication of the ancestral, enzymatic RRM domain in the evolutionary progenitor of Cas10, but this connection remains tenuous. The core of the ancestral effector complex could have been a signals transduction system consisting of the ancestor of Cas10 (a cOA polymerase) and a CARF-HEPN effector that triggered dormancy or PCD in response to infection or other forms of stress. This part of the CRISPR-Cas evolution

scenario is, admittedly, the weakest and requires the most effort in genome mining and structural comparison to complete the reconstruction convincingly.

The third major trend is the reductive evolution of CRISPR-Cas systems which led to defective variants that apparently were recruited for functions other than adaptive immunity. Notably, many if not most of such variants reside in MGE [137] and, presumably, contribute to the reproduction of those elements, although the mechanisms involved remain enigmatic.

On a more general plane, the contributions of MGE to the evolution of an adaptive immunity system and, conversely, the recruitment of defence systems or their components by MGE for antidefence (as in some bacterial viruses that encode complete CRISPR-Cas systems) or other functions (as in the case of defective systems discussed here) fits the 'guns for hire' concept [146]. Within this framework, enzymes are shuttled between defence systems and MGE, going to the 'highest bidder' (the elements providing the highest reproductive benefit to the respective genes), given that the activities involved, such as those of transposase, site-specific or promiscuous (in the case of TA) nuclease, helicase or RT, are closely analogous if not identical. This concept is applicable far beyond CRISPR-Cas: for example, different, unrelated transposons were the ancestors of key components of the adaptive immunity system in vertebrates [41] and DNA diminution system in ciliates [147].

**Data accessibility.** This article has no additional data.

**Competing interests.** We declare we have no competing interests.

**Funding.** E.V.K. and K.S.M. are supported by intramural funds of the US Department of Health and Human Services (to the National Library of Medicine).

## References

- Sorek R, Lawrence CM, Wiedenheft B. 2013 CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* **82**, 237–266. (doi:10.1146/annurev-biochem-072911-172315)
- Wright AV, Nunez JK, Doudna JA. 2016 Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* **164**, 29–44. (doi:10.1016/j.cell.2015.12.035)
- Komor AC, Badran AH, Liu DR. 2017 CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **169**, 559. (doi:10.1016/j.cell.2017.04.005)
- Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J. 2016 Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147. (doi:10.1126/science.aad5147)
- Hsu PD, Lander ES, Zhang F. 2014 Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278. (doi:10.1016/j.cell.2014.05.010)
- Barrangou R, Marraffini LA. 2014 CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell.* **54**, 234–244. (doi:10.1016/j.molcel.2014.03.011)
- Doudna JA, Charpentier E. 2014 Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096. (doi:10.1126/science.1258096)
- Koonin EV, Makarova KS, Zhang F. 2017 Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78. (doi:10.1016/j.mib.2017.05.008)
- Barrangou R, Horvath P. 2017 A decade of discovery: CRISPR functions and applications. *Nat. Microbiol.* **2**, 17092. (doi:10.1038/nmicrobiol.2017.92)
- Wu WY, Lebbink JHG, Kanaar R, Geijsen N, van der Oost J. 2018 Genome editing by natural and engineered CRISPR-associated nucleases. *Nat. Chem. Biol.* **14**, 642–651. (doi:10.1038/s41589-018-0080-x)
- Takeuchi N, Wolf YI, Makarova KS, Koonin EV. 2012 Nature and intensity of selection pressure on CRISPR-associated genes. *J. Bacteriol.* **194**, 1216–1225. (doi:10.1128/JB.06521-11)
- Makarova KS *et al.* 2011 Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477. (doi:10.1038/nrmicro2577)
- Makarova KS *et al.* 2015 An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736. (doi:10.1038/nrmicro3569)
- Shabalina SA, Koonin EV. 2008 Origins and evolution of eukaryotic RNA interference. *Trends Ecol. Evol.* **23**, 578–587. (doi:10.1016/j.tree.2008.06.005)
- Koonin EV. 2017 Evolution of RNA- and DNA-guided antiviral defense systems in prokaryotes and eukaryotes: common ancestry vs convergence. *Biol. Direct* **12**, 5. (doi:10.1186/s13062-017-0177-2)
- Hur JK, Olovnikov I, Aravin AA. 2014 Prokaryotic Argonautes defend genomes against invasive DNA. *Trends Biochem. Sci.* **39**, 257–259. (doi:10.1016/j.tibs.2014.04.006)
- Hutvagner G, Simard MJ. 2008 Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.* **9**, 22–32. (doi:10.1038/nrm2321)
- Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, van der Oost J. 2014 The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21**, 743–753. (doi:10.1038/nsmb.2879)
- Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. 2006 A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and

- hypothetical mechanisms of action. *Biol. Direct* **1**, 7. (doi:10.1186/1745-6150-1-7)
20. Marraffini LA, Sontheimer EJ. 2010 CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* **11**, 181–190. (doi:10.1038/nrg2749)
  21. van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ. 2009 CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* **34**, 401–407. (doi:10.1016/j.tibs.2009.05.002)
  22. Koonin EV, Makarova KS. 2009 CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 Biol. Rep.* **1**, 95. (doi:10.3410/B1-95)
  23. Makarova KS, Wolf YI, Koonin EV. 2013 The basic building blocks and evolution of CRISPR–Cas systems. *Biochem. Soc. Trans.* **41**, 1392–1400. (doi:10.1042/BST20130038)
  24. Amitai G, Sorek R. 2016 CRISPR-Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* **14**, 67–76. (doi:10.1038/nrmicro.2015.14)
  25. Jackson SA, McKenzie RE, Fagerlund RD, Kieper SN, Fineran PC, Brouns SJ. 2017 CRISPR-Cas: adapting to change. *Science* **356**, eaal5056. (doi:10.1126/science.aal5056)
  26. Ivancic-Bace I, Cass SD, Wearne SJ, Bolt EL. 2015 Different genome stability proteins underpin primed and naive adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res.* **43**, 10 821–10 830. (doi:10.1093/nar/gkv1213)
  27. Killelea T, Bolt EL. 2017 CRISPR-Cas adaptive immunity and the three Rs. *Biosci. Rep.* **37**, BSR20160297. (doi:10.1042/BSR20160297)
  28. Liu T *et al.* 2017 Coupling transcriptional activation of CRISPR–Cas system and DNA repair genes by Csa3a in *Sulfolobus islandicus*. *Nucleic Acids Res.* **45**, 8978–8992. (doi:10.1093/nar/gkx612)
  29. Silas S, Mohr G, Sidote DJ, Markham LM, Sanchez-Amat A, Bhaya D, Lambowitz AM, Fire AZ. 2016 Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase–Cas1 fusion protein. *Science* **351**, aad4234. (doi:10.1126/science.aad4234)
  30. Silas S *et al.* 2017 On the origin of reverse transcriptase-using CRISPR–Cas systems and their hyperdiverse, enigmatic spacer repertoires. *MBio* **8**, e00897-17. (doi:10.1128/mBio.00897-17)
  31. Charpentier E, Richter H, van der Oost J, White MF. 2015 Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR–Cas adaptive immunity. *FEMS Microbiol. Rev.* **39**, 428–441. (doi:10.1093/femsre/fuv023)
  32. Hochstrasser ML, Doudna JA. 2015 Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends Biochem. Sci.* **40**, 58–66. (doi:10.1016/j.tibs.2014.10.007)
  33. Nishimasu H, Nureki O. 2017 Structures and mechanisms of CRISPR RNA-guided effector nucleases. *Curr. Opin. Struct. Biol.* **43**, 68–78. (doi:10.1016/j.sbi.2016.11.013)
  34. Plagens A, Richter H, Charpentier E, Randau L. 2015 DNA and RNA interference mechanisms by CRISPR–Cas surveillance complexes. *FEMS Microbiol. Rev.* **39**, 442–463. (doi:10.1093/femsre/fuv019)
  35. Wang J, Li J, Zhao H, Sheng G, Wang M, Yin M, Wang Y. 2015 Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR–Cas systems. *Cell* **163**, 840–853. (doi:10.1016/j.cell.2015.10.008)
  36. Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. 2014 Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528–534. (doi:10.1038/nsmb.2820)
  37. Nunez JK, Lee AS, Engelman A, Doudna JA. 2015 Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature* **527**, 535–538. (doi:10.1038/nature15760)
  38. Li M, Wang R, Zhao D, Xiang H. 2014 Adaptation of the *Haloarcula hispanica* CRISPR–Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res.* **42**, 2483–2492. (doi:10.1093/nar/gkt1154)
  39. Kunne T, Kieper SN, Bannenberg JW, Vogel AI, Miellel WR, Klein M, Depken M, Suarez-Diez M, Brouns SJ. 2016 Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. *Mol. Cell* **63**, 852–864. (doi:10.1016/j.molcel.2016.07.011)
  40. Fagerlund RD *et al.* 2017 Spacer capture and integration by a type I-F Cas1–Cas2-3 CRISPR adaptation complex. *Proc. Natl Acad. Sci. USA* **114**, E5122–E5128. (doi:10.1073/pnas.1618421114)
  41. Koonin EV, Krupovic M. 2015 Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat. Rev. Genet.* **16**, 184–192. (doi:10.1038/nrg3859)
  42. Chylinski K, Makarova KS, Charpentier E, Koonin EV. 2014 Classification and evolution of type II CRISPR–Cas systems. *Nucleic Acids Res.* **42**, 6091–6105. (doi:10.1093/nar/gku241)
  43. Shmakov S *et al.* 2017 Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182. (doi:10.1038/nrmicro.2016.184)
  44. Brouns SJ *et al.* 2008 Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964. (doi:10.1126/science.1159689)
  45. Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, Doudna JA, Nogales E. 2011 Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489. (doi:10.1038/nature10402)
  46. Rouillon C *et al.* 2013 Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol. Cell* **52**, 124–134. (doi:10.1016/j.molcel.2013.08.020)
  47. Spilman M, Cocozaki A, Hale C, Shao Y, Ramia N, Terns R, Terns M, Li H, Staggs S. 2013 Structure of an RNA silencing complex of the CRISPR–Cas immune system. *Mol. Cell* **52**, 146–152. (doi:10.1016/j.molcel.2013.09.008)
  48. Staals RH *et al.* 2013 Structure and activity of the RNA-targeting Type III-B CRISPR–Cas complex of *Thermus thermophilus*. *Mol. Cell* **52**, 135–145. (doi:10.1016/j.molcel.2013.09.013)
  49. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. 2011 CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607. (doi:10.1038/nature09886)
  50. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012 A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821. (doi:10.1126/science.1225829)
  51. Jinek M *et al.* 2014 Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997. (doi:10.1126/science.1247997)
  52. Jiang F, Doudna JA. 2017 CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **46**, 505–529. (doi:10.1146/annurev-biophys-062215-010822)
  53. Shmakov S *et al.* 2015 Discovery and functional characterization of diverse class 2 CRISPR–Cas systems. *Mol. Cell* **60**, 385–397. (doi:10.1016/j.molcel.2015.10.008)
  54. Liu L, Chen P, Wang M, Li X, Wang J, Yin M, Wang Y. 2017 C2c1-sgRNA complex structure reveals RNA-guided DNA cleavage mechanism. *Mol. Cell* **65**, 310–322. (doi:10.1016/j.molcel.2016.11.040)
  55. Yang H, Gao P, Rajashankar KR, Patel DJ. 2016 PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR–Cas endonuclease. *Cell* **167**, 1814–1828. (doi:10.1016/j.cell.2016.11.053)
  56. East-Seletsky A, O’Connell MR, Knight SC, Burstein D, Cate JH, Tjian R, Doudna JA. 2016 Two distinct RNase activities of CRISPR–C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270–273. (doi:10.1038/nature19802)
  57. Fonfara I, Richter H, Bratovic M, Le Rhun A, Charpentier E. 2016 The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521. (doi:10.1038/nature17945)
  58. Zetsche B *et al.* 2017 Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34. (doi:10.1038/nbt.3737)
  59. Liu L, Li X, Wang J, Wang M, Chen P, Yin M, Li J, Sheng G, Wang Y. 2017 Two distant catalytic sites are responsible for C2c2 RNase activities. *Cell* **168**, 121–134. (doi:10.1016/j.cell.2016.12.031)
  60. Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011 Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30**, 1335–1342. (doi:10.1038/emboj.2011.41)
  61. Gong B, Shin M, Sun J, Jung CH, Bolt EL, van der Oost J, Kim JS. 2014 Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc. Natl Acad. Sci. USA* **111**, 16 359–16 364. (doi:10.1073/pnas.1410806111)
  62. Redding S, Sternberg SH, Marshall M, Gibb B, Bhat P, Guegler CK, Wiedenheft B, Doudna JA, Greene EC. 2015 Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR–Cas system. *Cell* **163**, 854–865. (doi:10.1016/j.cell.2015.10.003)
  63. Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012 Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in

- bacteria. *Proc. Natl Acad. Sci. USA* **109**, E2579–E2586. (doi:10.1073/pnas.1208507109)
64. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA. 2016 Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871. (doi:10.1126/science.aad8282)
  65. Zetsche B *et al.* 2015 Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771. (doi:10.1016/j.cell.2015.09.038)
  66. Dong D *et al.* 2016 The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526. (doi:10.1038/nature17944)
  67. Yamano T *et al.* 2016 Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962. (doi:10.1016/j.cell.2016.04.003)
  68. Abudayyeh OO *et al.* 2016 C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573. (doi:10.1126/science.aaf5573)
  69. Smargon AA *et al.* 2017 Cas13b is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28. *Mol. Cell* **65**, 618–630. (doi:10.1016/j.molcel.2016.12.023)
  70. Garrett RA, Vestergaard G, Shah SA. 2011 Archaeal CRISPR-based immune systems: exchangeable functional modules. *Trends Microbiol.* **19**, 549–556. (doi:10.1016/j.tim.2011.08.002)
  71. Vestergaard G, Garrett RA, Shah SA. 2014 CRISPR adaptive immune systems of Archaea. *RNA Biol.* **11**, 156–167. (doi:10.4161/rna.27990)
  72. Hudaiberdiev S, Shmakov S, Wolf YI, Terns MP, Makarova KS, Koonin EV. 2017 Phylogenomics of Cas4 family nucleases. *BMC Evol. Biol.* **17**, 232. (doi:10.1186/s12862-017-1081-1)
  73. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. 2013 Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432. (doi:10.1038/nature11723)
  74. Bondy-Denomy J. 2018 Protein inhibitors of CRISPR-Cas9. *ACS Chem. Biol.* **13**, 417–423. (doi:10.1021/acscchembio.7b00831)
  75. Pawluk A *et al.* 2016 Naturally occurring off-switches for CRISPR-Cas9. *Cell* **167**, 1829–1838. (doi:10.1016/j.cell.2016.11.017)
  76. Pawluk A, Davidson AR, Maxwell KL. 2018 Anti-CRISPR: discovery, mechanism and function. *Nat. Rev. Microbiol.* **16**, 12–17. (doi:10.1038/nrmicro.2017.120)
  77. Pawluk A, Staals RH, Taylor C, Watson BN, Saha S, Fineran PC, Maxwell KL, Davidson AR. 2016 Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat. Microbiol.* **1**, 16085. (doi:10.1038/nrmicrobiol.2016.85)
  78. Maxwell KL. 2017 The anti-CRISPR story: a battle for survival. *Mol. Cell* **68**, 8–14. (doi:10.1016/j.molcel.2017.09.002)
  79. He F, Bhoobalan-Chitty Y, Van LB, Kjeldsen AL, Dedola M, Makarova KS, Koonin EV, Brodersen DE, Peng X. 2018 Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nat. Microbiol.* **3**, 461–469. (doi:10.1038/s41564-018-0120-z)
  80. Hynes AP *et al.* 2018 Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. *Nat. Commun.* **9**, 2919. (doi:10.1038/s41467-018-05092-w)
  81. Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D, Marraffini LA. 2015 Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199–202. (doi:10.1038/nature14245)
  82. Heler R, Wright AV, Vucelja M, Bikard D, Doudna JA, Marraffini LA. 2017 Mutations in Cas9 enhance the rate of acquisition of viral spacer sequences during the CRISPR-Cas immune response. *Mol. Cell* **65**, 168–175. (doi:10.1016/j.molcel.2016.11.031)
  83. Krupovic M, Makarova KS, Forterre P, Prangshvili D, Koonin EV. 2014 Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol.* **12**, 36. (doi:10.1186/1741-7007-12-36)
  84. Krupovic M, Koonin EV. 2016 Self-synthesizing transposons: unexpected key players in the evolution of viruses and defense systems. *Curr. Opin. Microbiol.* **31**, 25–33. (doi:10.1016/j.mib.2016.01.006)
  85. Kapitonov VV, Jurka J. 2006 Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl Acad. Sci. USA* **103**, 4540–4545. (doi:10.1073/pnas.0600833103)
  86. Pritham EJ, Putliwala T, Feschotte C. 2007 Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* **390**, 3–17. (doi:10.1016/j.gene.2006.08.008)
  87. Krupovic M, Koonin EV. 2015 Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat. Rev. Microbiol.* **13**, 105–115. (doi:10.1038/nrmicro3389)
  88. Hickman AB, Dyda F. 2015 The casposon-encoded Cas1 protein from *Aciduliprofundum boonei* is a DNA integrase that generates target site duplications. *Nucleic Acids Res.* **43**, 10 576–10 587. (doi:10.1093/nar/gkv1180)
  89. Beguin P, Charpin N, Koonin EV, Forterre P, Krupovic M. 2016 Casposon integration shows strong target site preference and recapitulates protospacer integration by CRISPR-Cas systems. *Nucleic Acids Res.* **44**, 10 367–10 376. (doi:10.1093/nar/gkw821)
  90. Krupovic M, Beguin P, Koonin EV. 2017 Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol.* **38**, 36–43. (doi:10.1016/j.mib.2017.04.004)
  91. Yutin N, Backstrom D, Ettema TJG, Krupovic M, Koonin EV. 2018 Vast diversity of prokaryotic virus genomes encoding double jelly-roll major capsid proteins uncovered by genomic and metagenomic sequence analysis. *Viol. J.* **15**, 67. (doi:10.1186/s12985-018-0974-y)
  92. Krupovic M, Shmakov S, Makarova KS, Forterre P, Koonin EV. 2016 Recent mobility of casposons, self-synthesizing transposons at the origin of the CRISPR-Cas immunity. *Genome Biol. Evol.* **8**, 375–386. (doi:10.1093/gbe/evw006)
  93. Jackson RN *et al.* 2014 Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* **345**, 1473–1479. (doi:10.1126/science.1256328)
  94. Jackson RN, Wiedenheft B. 2015 A conserved structural chassis for mounting versatile CRISPR RNA-guided immune responses. *Mol. Cell* **58**, 722–728. (doi:10.1016/j.molcel.2015.05.023)
  95. Reeks J, Graham S, Anderson L, Liu H, White MF, Naismith JH. 2013 Structure of the archaeal Cascade subunit Csa5: relating the small subunits of CRISPR effector complexes. *RNA Biol.* **10**, 762–769. (doi:10.4161/rna.23854)
  96. Zhu X, Ye K. 2012 Crystal structure of Cmr2 suggests a nucleotide cyclase-related enzyme in type III CRISPR-Cas systems. *FEBS Lett.* **586**, 939–945. (doi:10.1016/j.febslet.2012.02.036)
  97. Shao Y, Cocozaki AI, Ramia NF, Terns RM, Terns MP, Li H. 2013 Structure of the Cmr2-Cmr3 subcomplex of the Cmr RNA silencing complex. *Structure* **21**, 376–384. (doi:10.1016/j.str.2013.01.002)
  98. Makarova KS, Aravind L, Grishin NV, Rogozin IB, Koonin EV. 2002 A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.* **30**, 482–496. (doi:10.1093/nar/30.2.482)
  99. Makarova KS, Aravind L, Wolf YI, Koonin EV. 2011 Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct* **6**, 38. (doi:10.1186/1745-6150-6-38)
  100. Reeks J, Naismith JH, White MF. 2013 CRISPR interference: a structural perspective. *Biochem. J.* **453**, 155–166. (doi:10.1042/BJ20130316)
  101. Carte J, Pfister NT, Compton MM, Terns RM, Terns MP. 2010 Binding and cleavage of CRISPR RNA by Cas6. *RNA* **16**, 2181–2188. (doi:10.1261/rna.2230110)
  102. Richter H, Zoepfel J, Schermuly J, Maticzka D, Backofen R, Randau L. 2012 Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. *Nucleic Acids Res.* **40**, 9887–9896. (doi:10.1093/nar/gks737)
  103. Reeks J, Sokolowski RD, Graham S, Liu H, Naismith JH, White MF. 2013 Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing. *Biochem. J.* **452**, 223–230. (doi:10.1042/BJ20130269)
  104. Makarova KS, Anantharaman V, Grishin NV, Koonin EV, Aravind L. 2014 CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Front. Genet.* **5**, 102. (doi:10.3389/fgene.2014.00102)
  105. Anantharaman V, Makarova KS, Burroughs AM, Koonin EV, Aravind L. 2013 Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biol. Direct* **8**, 15. (doi:10.1186/1745-6150-8-15)
  106. Burroughs AM, Zhang D, Schaffer DE, Iyer LM, Aravind L. 2015 Comparative genomic analyses

- reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res.* **43**, 10 633–10 654. (doi:10.1093/nar/gkv1267)
107. Kazlauskienė M, Kostiuk G, Venclovas C, Tamulaitis G, Siksnys V. 2017 A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357**, 605–609. (doi:10.1126/science.aao0100)
  108. Niewoehner O, Garcia-Doval C, Rostol JT, Berk C, Schwede F, Bigler L, Hall J, Marraffini LA, Jinek M. 2017 Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548. (doi:10.1038/nature23467)
  109. Athukoralage JS, Rouillon C, Graham S, Gruschow S, White MF. 2018 Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. *Nature* **562**, 277–280. (doi: 10.1038/s41586-018-0557-5)
  110. Rouillon C, Athukoralage JS, Graham S, Gruschow S, White MF. 2018 Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *eLife* **7**, e36734. (doi:10.7554/eLife.36734)
  111. Koonin EV, Makarova KS. 2018 Discovery of oligonucleotide signaling mediated by CRISPR-associated polymerases solves two puzzles but leaves an enigma. *ACS Chem. Biol.* **13**, 309–312. (doi:10.1021/acscmbio.7b00713)
  112. Koonin EV, Zhang F. 2017 Coupling immunity and programmed cell suicide in prokaryotes: life-or-death choices. *Bioessays* **39**, 1–9. (doi:10.1002/bies.201600186)
  113. Majorek KA, Dunin-Horkawicz S, Steczkiewicz K, Muszewska A, Nowotny M, Ginalski K, Bujnicki JM. 2014 The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res.* **42**, 4160–4179. (doi:10.1093/nar/gkt1414)
  114. Nishimasu H *et al.* 2015 Crystal structure of *Staphylococcus aureus* Cas9. *Cell* **162**, 1113–1126. (doi:10.1016/j.cell.2015.08.007)
  115. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. 2014 Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949. (doi:10.1016/j.cell.2014.02.001)
  116. Hirano H *et al.* 2016 Structure and engineering of *Francisella novicida* Cas9. *Cell* **164**, 950–961. (doi:10.1016/j.cell.2016.01.039)
  117. Lewis KM, Ke A. 2017 Building the class 2 CRISPR-Cas arsenal. *Mol. Cell* **65**, 377–379. (doi:10.1016/j.molcel.2017.01.024)
  118. Swarts DC, van der Oost J, Jinek M. 2017 Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a. *Mol. Cell* **66**, 221–233. (doi:10.1016/j.molcel.2017.03.016)
  119. Bao W, Jurka J. 2013 Homologues of bacterial TnpB<sub>IS605</sub> are widespread in diverse eukaryotic transposable elements. *Mob. DNA* **4**, 12. (doi:10.1186/1759-8753-4-12)
  120. Pasternak C, Dulerio R, Ton-Hoang B, Debuchy R, Siguière P, Coste G, Chandler M, Sommer S. 2013 ISDra2 transposition in *Deinococcus radiodurans* is downregulated by TnpB. *Mol. Microbiol.* **88**, 443–455. (doi:10.1111/mmi.12194)
  121. Kapitonov VV, Makarova KS, Koonin EV. 2015 ISC, a novel group of bacterial and archaeal DNA transposons that encode Cas9 homologs. *J. Bacteriol.* **198**, 797–807. (doi:10.1128/JB.00783-15)
  122. Gerdes K. 2012 *Prokaryotic toxin-antitoxins*. Berlin, Germany: Springer.
  123. Gerdes K, Christensen SK, Lobner-Olesen A. 2005 Prokaryotic toxin-antitoxin stress response loci. *Nat. Rev. Microbiol.* **3**, 371–382. (doi:10.1038/nrmicro1147)
  124. Makarova KS, Wolf YI, Koonin EV. 2009 Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct* **4**, 19. (doi:10.1186/1745-6150-4-19)
  125. Van Melderen L. 2010 Toxin-antitoxin systems: why so many, what for? *Curr. Opin. Microbiol.* **13**, 781–785. (doi:10.1016/j.mib.2010.10.006)
  126. Van Melderen L, Saavedra De Bast M. 2009 Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genet.* **5**, e1000437. (doi:10.1371/journal.pgen.1000437)
  127. Yamaguchi Y, Inouye M. 2009 mRNA interferases, sequence-specific endoribonucleases from the toxin-antitoxin systems. *Prog. Mol. Biol. Transl. Sci.* **85**, 467–500. (doi:10.1016/S0079-6603(08)00812-X)
  128. Cook GM, Robson JR, Frampton RA, McKenzie J, Przybilski R, Fineran PC, Arcus VL. 2013 Ribonucleases in bacterial toxin-antitoxin systems. *Biochim. Biophys. Acta* **1829**, 523–531. (doi:10.1016/j.bbagr.2013.02.007)
  129. Gelens L, Hill L, Vandervelde A, Danckaert J, Loris R. 2013 A general model for toxin-antitoxin module dynamics can explain persister cell formation in *E. coli*. *PLoS Comput. Biol.* **9**, e1003190. (doi:10.1371/journal.pcbi.1003190)
  130. Harms A, Brodersen DE, Mitarai N, Gerdes K. 2018 Toxins, targets, and triggers: an overview of toxin-antitoxin biology. *Mol. Cell* **70**, 768–784. (doi:10.1016/j.molcel.2018.01.003)
  131. Beloglazova N *et al.* 2008 A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.* **283**, 20 361–20 371. (doi:10.1074/jbc.M803225200)
  132. Ka D, Kim D, Baek G, Bae E. 2014 Structural and functional characterization of *Streptococcus pyogenes* Cas2 protein under different pH conditions. *Biochem. Biophys. Res. Commun.* **451**, 152–157. (doi:10.1016/j.bbrc.2014.07.087)
  133. Dixit B, Ghosh KK, Fernandes G, Kumar P, Gogoi P, Kumar M. 2016 Dual nuclease activity of a Cas2 protein in CRISPR-Cas subtype I-B of *Leptospira interrogans*. *FEBS Lett.* **590**, 1002–1016. (doi:10.1002/1873-3468.12124)
  134. Makarova KS, Anantharaman V, Aravind L, Koonin EV. 2012 Live virus-free or die: coupling of antiviral immunity and programmed suicide or dormancy in prokaryotes. *Biol. Direct* **7**, 40. (doi:10.1186/1745-6150-7-40)
  135. Toro N, Nisa-Martinez R. 2014 Comprehensive phylogenetic analysis of bacterial reverse transcriptases. *PLoS ONE* **9**, e114083. (doi:10.1371/journal.pone.0114083)
  136. Toro N, Martinez-Abarca F, Gonzalez-Delgado A. 2017 The reverse transcriptases associated with CRISPR-Cas systems. *Sci. Rep.* **7**, 7089. (doi:10.1038/s41598-017-07828-y)
  137. Koonin EV, Makarova KS. 2017 Mobile genetic elements and evolution of CRISPR-Cas systems: all the way there and back. *Genome Biol. Evol.* **9**, 2812–2825. (doi:10.1093/gbe/evx192)
  138. Peters JE, Makarova KS, Shmakov S, Koonin EV. 2017 Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl Acad. Sci. USA* **114**, E7358–E7366. (doi:10.1073/pnas.1709035114)
  139. Shmakov SA, Makarova KS, Wolf YI, Severinov KV, Koonin EV. 2018 Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl Acad. Sci. USA* **115**, E5307–E5316. (doi:10.1073/pnas.1803440115)
  140. Leipe DD, Koonin EV, Aravind L. 2004 STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J. Mol. Biol.* **343**, 1–28. (doi:10.1016/j.jmb.2004.08.023)
  141. Darwin C. 1859 *On the origin of species*. London, UK: John Murray.
  142. Stoltzfus A. 1999 On the possibility of constructive neutral evolution. *J. Mol. Evol.* **49**, 169–181. (doi:10.1007/PL00006540)
  143. Gray MW, Lukes J, Archibald JM, Keeling PJ, Doolittle WF. 2010 Irremediable complexity? *Science* **330**, 920–921. (doi:10.1126/science.1198594)
  144. Stoltzfus A. 2012 Constructive neutral evolution: exploring evolutionary theory's curious disconnect. *Biol. Direct* **7**, 35. (doi:10.1186/1745-6150-7-35)
  145. Faure G, Shmakov SA, Makarova KS, Wolf YI, Crawley AB, Barrangou R, Koonin EV. 2018 Comparative genomics and evolution of transactivating RNAs in Class 2 CRISPR-Cas systems. *RNA Biol.* **15**, 1–14. (doi:10.1080/15476286.2018.1493331)
  146. Koonin EV, Krupovic M. 2015 A movable defense. *Scientist*, 1 January 2015. See <https://www.the-scientist.com/features/a-movable-defense-36135>.
  147. Nowacki M, Shetty K, Landweber LF. 2011 RNA-mediated epigenetic programming of genome rearrangements. *Annu. Rev. Genomics Hum. Genet.* **12**, 367–389. (doi:10.1146/annurev-genom-082410-101420)