Check for updates

Universal genome cutters: from selfish genetic elements to antivirus defence and genome editing tools

Eugene V. Koonin (National Library of Medicine, USA)

Over the last 3 years, the new generation of genome editing and engineering tools based on the bacterial CRISPR-Cas (Clustered Regularly Interspaced Palindromic Repeats and CRISPR-associated proteins) has attracted enormous attention in both the scientific community and the popular media[1,2]. Although exaggerations certainly cannot be avoided along the way, the enthusiasm is well justified. Indeed, CRISPR-Cas provides a means to cut into any DNA molecule at any desirable site with an unprecedented, virtually perfect specificity, thus making realistic nearly any genome modification at a minimal expense of time and resources. The CRISPR-Cas systems present one of those rare, truly lucky cases where a fundamental discovery that changes our ideas of how evolution works is also of immense practical value.

CRISPR-Cas: an RNA-guided machine of Lamarckian evolution

CRISPR-Cas is the bacterial and archaeal system of adaptive immunity that employs the same molecular principle as eukaryotic RNA interference (RNAi), namely, complementarity of short guide RNA molecules to the target, RNA in the case of eukaryotic RNAi, but DNA in the case of CRISPR. Short as they are (typically, around 40 nucleotides in the case of CRISPR), these RNA guides contain enough information to ensure, at least in theory, absolute specificity of target recognition. The final step of defence, namely the guide RNA-directed target cleavage by the effector nuclease, is analogous between CRISPR-Cas and RNAi, although the nucleases themselves are unrelated. However, the upstream steps could not be more different (Figure 1). The RNAi system simply cuts the double-stranded form of a viral RNA genome into pieces and either uses these directly as guide RNAs or first amplifies them with the help of a dedicated RNA-dependent RNA polymerase. The CRISPR–Cas system goes through a far more complex route. The CRISPR defence response starts with the adaptation step whereby foreign (e.g. bacteriophage) DNA is recognized by the Cas1-Cas2 protein complex after which Cas1, a transposase-like enzyme, cuts out a piece of DNA and with the help of the microbial repair machinery inserts it into the CRISPR array. The CRISPR locus including the phage-specific spacers is then transcribed, the long transcript is processed into mature CRISPR (cr) RNAs consisting of the spacer (guide) and parts of the adjacent repeats, the crRNA is loaded onto the effector nuclease, which in at least some CRISPR variants also participates in the crRNA maturation, and the latter cleaves the cognate region of the target (Figure 1).

Once this mechanism of the CRISPR-Cas function in microbial immunity was deciphered, the startling realization was that this system embodies the Lamarckian evolutionary scenario[3]. Indeed, CRISPR-Cas mediates direct adaptation of an organism to an environmental factor (specifically, resistance to a virus) through directed, stably inheritable modification of the host genome. The key problem for such a mechanism, as for many defence strategies, is self- versus non-self-discrimination, and it is still far from being clear how good different variants of CRISPR-Cas are at recognizing invading foreign DNA. Some have been reported to efficiently recognize foreign DNA, in particular, small genomes of viruses and plasmids thanks to their high replication activity, reaffirming the adaptive, Lamarckian character of microbial adaptive immunity. Surprisingly, however, other CRISPR-Cas variants seem to incorporate spacers indiscriminately which often leads to autoimmunity. In this case, the operation of the system involves selection of the fittest variants (those that contain a virus-specific spacer) and appears to be more Darwinian than Lamarckian. Thus, even fundamental aspects of CRISPR-induced genome evolution appear to differ between different types of these adaptive

Downloaded from http://portlandpress.com/biochemist/article-pdf/38/3/4/852465/bio03803004.pdf by guest on 19 May 2025

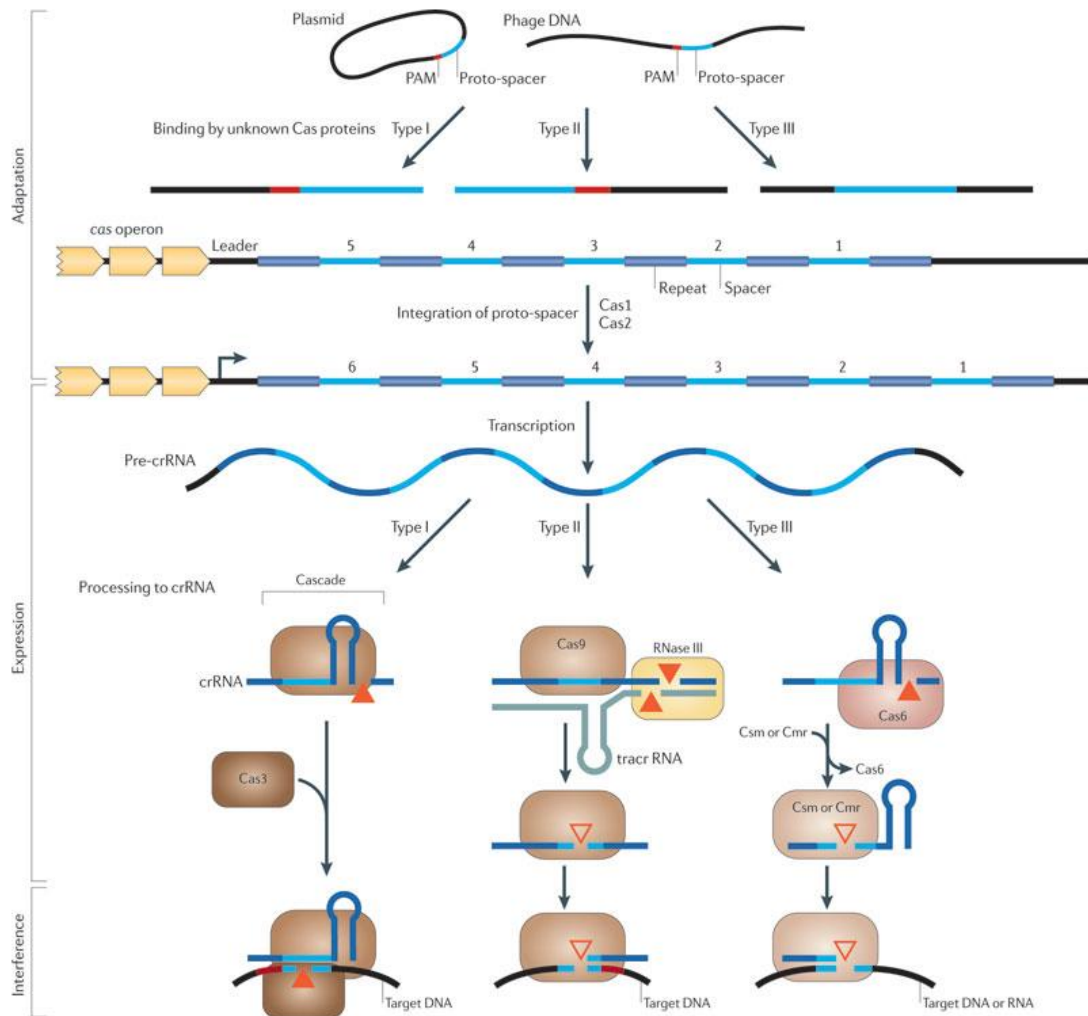


Figure 1. The figure shows the 3 steps of the immune response by the 3 major types of CRISPR-Cas systems. *Reproduced with permission from Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P,Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol. 2011; 9(6):467-77*

immunity systems. Regardless of these specifics, though, all CRISPR-Cas systems are elaborate devices for evolutionary engineering of microbial genomes.

Origin and evolution of CRISPR-Cas: contributions from diverse mobile elements

Given the sky-rocketing interest in CRISPR-Cas as genome editing tools, comparative genomics of these systems over the last few years has been pursued quite diligently. The most obvious result of this research is that CRISPR-Cas loci in archaeal and bacterial genomes are highly complex and extremely diverse, in terms of both gene arrangement (Figure 2) and amino acid sequences of the Cas proteins[4]. Yet, there is definitely method in this madness. The organization of the CRISPR-Cas systems is distinctly modular whereby a uniform adaptation module that consists of the Cas1 and Cas2 genes (and sometimes one or two additional genes) combines with diverse effector modules (Figure 2). With respect to the architecture of the effector modules, the CRISPR-Cas systems fall into two major classes[4]. In Class 1, the effector modules consist of multiple genes coding for proteins that form subunits of elaborate effector complexes involved in crRNA processing and target recognition and cleavage. In contrast, in Class 2, the entire effector module consists of a single, albeit very large protein, such as the now famous Cas9 used in most of the genome engineering applications and the more recently discovered Cpf1[5]. It is the relative simplicity of the Class 2 effector modules that makes them so practicable for genome engineering.

A deeper foray into comparative genomics has revealed the evolutionary roots of the key components of the CRISPR-Cas system[6,7]. In a nutshell, these approaches involve searching for genomic contexts in

Downloaded from http://portlandpress.com/biochemist/article-pdf/38/3/4/852465/bio03803004.pdf by guest on 19 May 2025

New classification of CRISPR-Cas systems

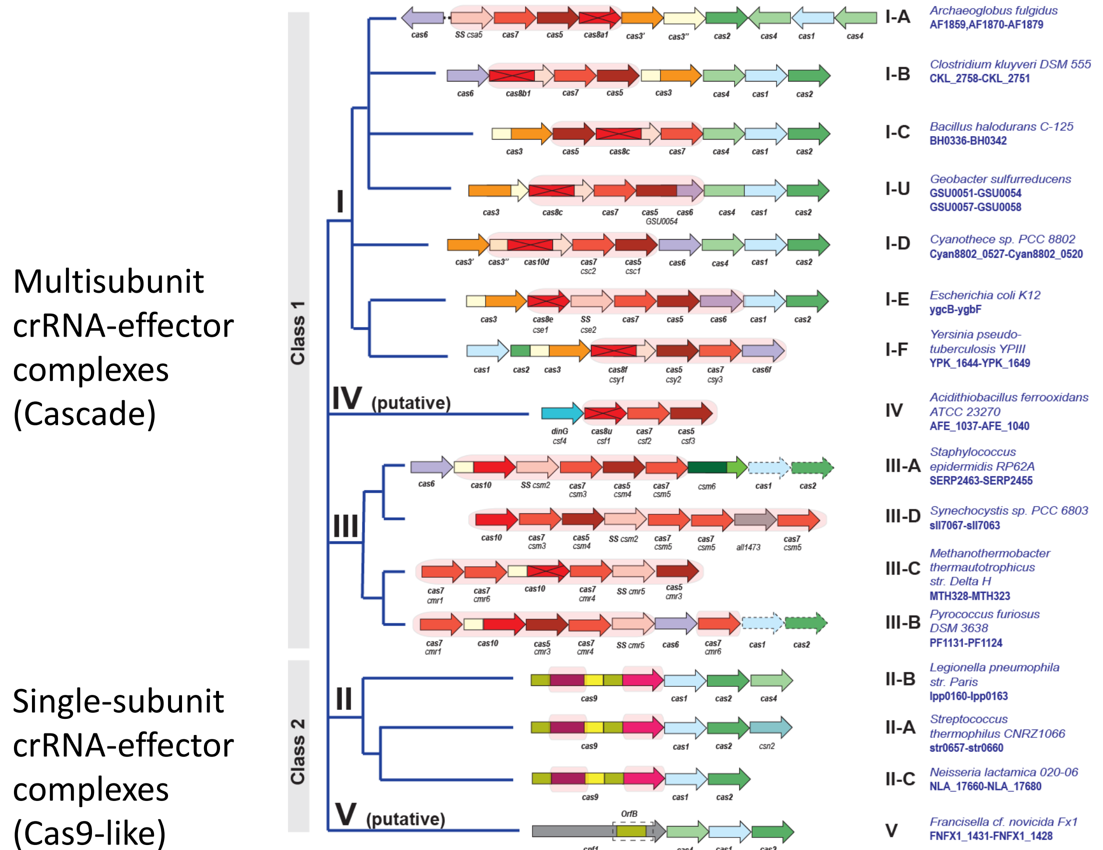


Figure 2. The block arrows denote cas gene, with the transcription direction. The effector complexes (Class 1) and multidomain effector proteins (Class 2) are shown by pink background. *Reproduced with permission from* [4].

which a particular cas gene is not associated with other elements of a typical CRISPR-Cas system. Genetic elements discovered by this approach might shed light on the origin of CRISPR-Cas, and a key clue indeed has been gleaned from the non-CRISPR genomic contexts of cas1 homologues. One family of such Cas1 genes has been found embedded in segments of DNA that resemble transposable elements (transposons for short) although they could not be classified into any of the many known families of transposons. A transposon is a selfish genetic element that has the capacity of autonomous replication and integrates into new sites in the host genome. All autonomous, non-defective transposons encode an enzyme, known as transposase (also known as recombinase or integrase), which is a nuclease with distinct properties that cuts a copy of the transposon out of a replicating DNA molecule and facilitates its integration into the new site. The series of reactions involved in transposition is strikingly similar to that occurring during the adaptation stage of the CRISPR immune response, strongly suggesting that the non-CRISPR Cas1 homologues function as transposases in a novel family of transposons that were accordingly dubbed casposons[6] (Figure 3). In addition to Cas1, the casposons encode a DNA polymerase and therefore can be described as "self-synthesizing" transposons, a type of selfish element that so far have been discovered in eukaryotes, but not in archaea or bacteria[8]. The additional few genes differ among the casposons and include various enzymes and predicted DNA-binding proteins. Typical of transposons, the casposons are flanked by inverted terminal repeats. The biology of the casposons remains to be studied in detail but comparative genomic analysis reveals their recent mobility whereas biochemical experiments demonstrate the recombinase activity of the Cas1 homologue[9].

The discovery of the casposons prompted a scenario of evolution for CRISPR-Cas whereby the adaptive immunity system evolved as a result of integration of a casposon next to an innate immunity locus6 (Figure

Downloaded from http://portlandpress.com/biochemist/article-pdf/38/3/4/852465/bio03803004.pdf by guest on 19 May 2025

Grand evolutionary scenario for CRISPR-Cas

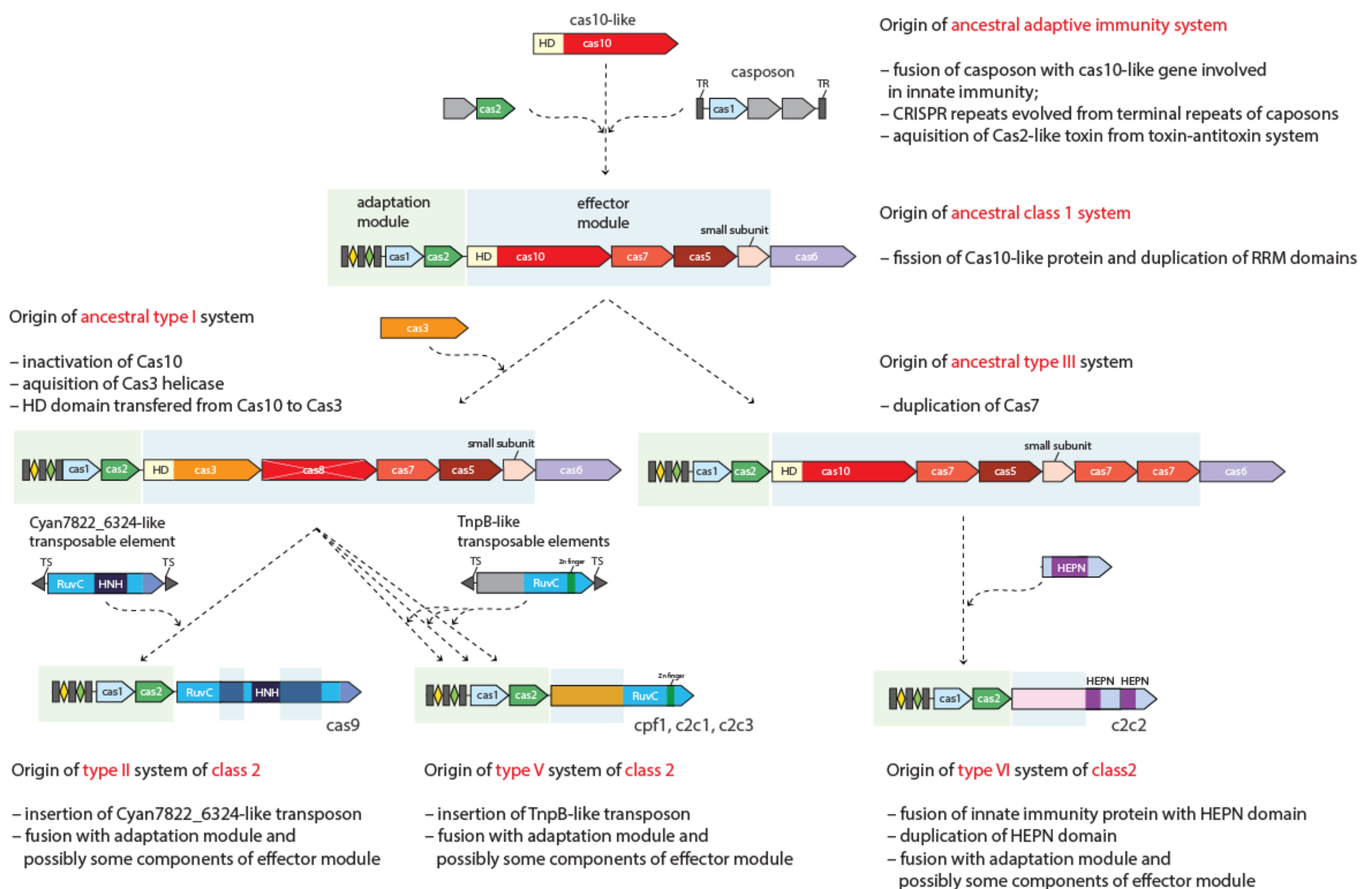


Figure 3. Evolutionary scenario for the CRISPR-Cas systems: multiple contributions of transposable elements. *Reproduced with permission from[7].*

3). Under this scenario, the casposon gave rise to the adaptation modules of CRISPR-Cas as well as the repeat themselves whereas the primordial innate immunity system became the Class 1 effector module. This reconstruction of CRISPR evolution presents a striking parallel to the history of vertebrate adaptive immunity where a completely unrelated transposon (of the Transib family of eukaryotic transposons) apparently gave rise to V(D)J recombination, the key mechanism of antibody diversification. Furthermore, an analogous series of events involving yet another distinct transposon apparently gave rise to the molecular machinery that is responsible forthe elimination of numerous intergenic regions during the maturation of the macronucleus in ciliates[10].

Strikingly, the contribution of transposons to the evolution of CRISPR-Cas is not limited to the casposons. Recently, a concerted effort, certainly driven in part by the keen interest in potential new tools for genome engineering, has been undertaken to explore the diversity of Class 2 CRISPR-Cas systems[7]. Several new (sub)types of Class 2 systems were discovered (Figure 3), and these findings shed new, unexpected light on the evolution of microbial adaptive immunity. The target-cleaving nuclease domains of type II and type V effector proteins turned out to be related to different groups of TnpB family nucleases that are encoded by diverse bacterial and archaeal transposons. The type VI effector proteins contain unrelated RNase domains of the higher eukaryotes and prokaryotes nucleotide-binding (HEPN) family that are involved in a variety of defence functions in both prokaryotes and eukaryotes. Thus, Class 2 CRISPR effectors seem to have evolved on multiple, independent occasions, primarily or exclusively from genes of transposable elements that replaced Class 1 effector genes. Notably, under this scenario, Class 2 systems derive entirely from different types of mobile elements including the casposons, different groups of

Downloaded from http://portlandpress.com/biochemist/article-pdf/38/3/4/852465/bio03803004.pdf by guest on 19 May 2025

IS (insertion sequence) transposons encoding TnpB proteins, and in the case of type II, a self-splicing intron that gave rise to the HNH nuclease domain of Cas9.

Mobile genetic elements, defence against genetic parasites and DNA repair: common themes in the evolution of genome manipulation

Genomes of all organisms are subject to constant assault from various kinds of DNA damage causing mutations as well as genetic parasites. Hence the evolution of diverse and elaborate systems of repair and defence in all cellular life forms. The underlying common theme that unifies mobile genetic elements, defence systems that protect cells against them and repair systems is that they all are essentially devices for genome manipulation. Indeed, the main "goal" of mobile elements is to integrate into the host genome and in the case of more aggressive elements, such as self-synthesizing transposons and viruses, to facilitate their own replication. Hence the characteristic repertoire of enzymatic activities encoded by these elements, such as various nucleases, helicases and in many cases polymerases. Defence systems face tasks that are, in principle, similar: they have to specifically inactivate the genomes of parasites, while sparing the host genome. The adaptive immunity systems additionally have to modify the host genome in a specific fashion so as to retain memory of past infection (CRISPR) or to generate combinatorial diversity that is required to recognize all possible antigens (vertebrate immune system).

Due to the common functional requirements, evolution of selfish genetic elements is tightly intertwined with the evolution of repair and defence systems such that genes of selfish elements are often recruited for functions in defence or repair and vice versa. The nucleases Cas1 and RuvC/TnpB discussed above are prime examples. Indeed, Cas1 functions as the recombinase in the casposons (most likely, its primary activity) and as the enzyme responsible for adaptation (spacer integration) in CRISPR-Cas. Furthermore, Cas1 mutants of *Escherichia coli* display a repair-deficient phenotype indicating that the nuclease activity of Cas1 is also involved in repair, most likely in the resolution of stalled replication forks11. Quite similarly, the RuvC nuclease (in this case, called TnpB) is encoded by numerous transposons, where its specific role remains to be elucidated, several subtypes of Class 2 CRISPR-Cas, where it functions as the target-cleaving nuclease, and in bacterial repair as a Holliday junction resolvase[12]. Thus, each of these nuclease families has the triple functionality in selfish elements, defence and repair. Additional examples of such multitasking can be presented but it seems likely that many more remain to be discovered, in particular, by systematics search for non-transposon-associated functions of recombinases.

Selfish elements often recruit complete microbial defence systems that they deploy as means of counterdefence. A notable example is the recently discovered incorporation of a CRISPR-Cas system by several bacteriophages; many other bacteriophages encode dedicated anti-CRISPR proteins[13]. In a striking case of arms race in action, the phage CRISPR-Cas system has been shown to specifically target a host restriction-modification (RM) module that otherwise prevents phage reproduction.

The link between selfish elements and defence systems has another dimension that I cannot discuss in much detail in this short article but have to mention at least in brief. To different extents, microbial defence systems themselves display features of selfish elements. Such properties are particularly prominent in RM and toxin-antitoxin modules that experience rampant horizontal gene transfer, typically, via plasmid vectors, and have evolved strategies to make microbial cells "addicted" to those plasmids[14].

Genome engineering: from nature to laboratory

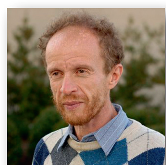
As discussed above, both transposable elements and microbial defence systems are naturally evolved devices for genome engineering in which sequence-specific nucleases are the active moieties. Therefore, in retrospect, there is nothing surprising in the fact that these enzymes have been so successfully adopted to genome engineering in molecular biology laboratories. The first generation of such tools are restriction endonucleases that transformed molecular biology in the 1970s. All their utility notwithstanding, restriction endonucleases have major limitations because they cleave DNA at strictly defined, short recognition sites. Therefore, a genomic DNA, even that of a virus, let alone a complex cellular life form such as an animal, typically is cleaved into multiple fragments, hampering efficient genome engineering. To some extent, these limitations have been overcome by isolation of numerous restriction endonucleases with diverse specificities. Nevertheless, this first generation of DNA editing tools remained far behind the ultimate goal of genome engineers, namely facile and highly selective modification of any desirable site in the genome. Largely, this limitation of restriction endonucleases stems from the fact that they are components of innate immunity that evolved to target any foreign DNA that does not carry the protective modification: hence short recognition sites. In contrast, adaptive immunity evolved to target sequences that are long enough to

Downloaded from http://portlandpress.com/biochemist/article-pdf/38/3/4/852465/bio03803004.pdf by guest on 19 May 2025

typically ensure highly specific recognition of a unique site in the target genome. Equally important, CRISPR-Cas incorporates fragments of the target genomes nearly randomly, so that there are no pre-defined requirements for the recognized sites. Thus, CRISPR-Cas is a virtually perfect tool for genome editing that can be programmed to cleave any desirable part of the genome. Furthermore, although CRISPR-Cas shows some low but non-negligible level of erroneous, off-target cleavage, relatively simple approaches have been developed to enhance its specificity.

The CRISPR effector nucleases only cleave DNA in the desired site but do not perform the other reactions required for genome editing such as elimination of additional sequences, filling in gaps and ligation of inserted DNA fragments. However, all these steps can be successfully relegated to the cellular repair machinery, again, thanks to the functional connections between defence and repair functions.

The current toolkit of genome editing, impressive as it is in its ability to introduce changes into the genome almost at will, is only the beginning. Additional possibilities for both engineering and regulation are already emerging from the newly discovered CRISPR-Cas variants, and more is likely to be offered by novel defence systems that remain to be discovered. ■



Eugene V. Koonin is the leader of the Evolutionary Genomics Group at the National Center for Biotechnology Information. He received his Ph.D. in Molecular Biology in 1983 from the Department of Biology, Moscow State University, joined the NCBI in 1991 and became a Senior Investigator in 1996. His group is pursuing several research directions in evolutionary genomics of prokaryotes, eukaryotes and viruses, along with general theoretical problems of the evolution of life, such as principles and models of host-parasite coevolution. He is the author of "The Logic of Chance: The nature and origin of biological evolution" (2011) and the founder and Editor-in-Chief of Biology Direct, an Open Access, open peer-review journal. Dr. Koonin is a Fellow of the American Academy of Arts and Sciences, and American Academy of Microbiology, a member of the National Academy of Sciences of the USA, a Foreign Associate of the European Molecular Biology Organization, and Doctor Honoris Causa of Universite Aix-Marseille (France).

References

1. Doudna, J.A. and Charpentier, E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science **346**, 1258096.
2. Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature **526**, 55–61
3. Koonin, E.V. and Wolf, Y.I. (2016) Just how Lamarckian is CRISPR-Cas immunity: the continuum of evolvability mechanisms. Biol. Direct **11**, 9
4. Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., et al. (2015) An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. **13**, 722–736
5. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., et al. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell **163**, 759–771
6. Koonin, E.V. and Krupovic, M. (2015) Evolution of adaptive immunity from transposable elements combined with innate immune systems. Nat. Rev. Genet. **16**, 184–192
7. Shmakov, S., Abudayyeh, O.O., Makarova, K.S., et al. (2015) Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Mol. Cell. **60**, 385–397
8. Krupovic, M. and Koonin, E.V. (2016) Self-synthesizing transposons: unexpected key players in the evolution of viruses and defense systems. Curr. Opin. Microbiol. **31**, 25–33
9. Hickman, A.B. and Dyda, F. (2015) The casposon-encoded Cas1 protein from *Aciduliprofundum boonei* is a DNA integrase that generates target site duplications. Nucleic Acids Res. **43**, 10576–10587
10. Chalker, D.L. and Yao, M.C. (2011) DNA elimination in ciliates: transposon domestication and genome surveillance. Annu. Rev. Genet. **45**, 227–246
11. Babu, M., Beloglazova, N., Flick, R., et al. (2011) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol. Microbiol. **79**, 484–502
12. Gorecka, K.M., Komorowska, W. and Nowotny, M. (2013) Crystal structure of RuvC resolvase in complex with Holliday junction substrate. Nucleic Acids Res. **41**, 9945–9955
13. Bondy-Denomy, J. and Davidson, A.R. (2014) To acquire or resist: the complex biological effects of CRISPR-Cas systems. Trends Microbiol. **22**, 218–225
14. Ramisetty, B.C. and Santhosh, R.S. (2016) Horizontal gene transfer of chromosomal Type II toxin-antitoxin systems of *Escherichia coli*. FEMS Microbiol. Lett. **363**

Downloaded from http://portlandpress.com/biochemist/article-pdf/38/3/4/852465/bio03803004.pdf by guest on 19 May 2025