Editor’s cut: DNA cleavage by CRISPR RNA-guided nucleases Cas9 and Cas12a

Thomas Swartjes, Raymond H.J. Staals and John van der Oost
Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, Netherlands
Correspondence: John van der Oost (john.vanderoost@wur.nl)

Discovered as an adaptive immune system of prokaryotes, CRISPR–Cas provides many promising applications. DNA-cleaving Cas enzymes like Cas9 and Cas12a, are of great interest for genome editing. The specificity of these DNA nucleases is determined by RNA guides, providing great targeting adaptability. Besides this general method of programmable DNA cleavage, these nucleases have different biochemical characteristics, that can be exploited for different applications. Although Cas nucleases are highly promising, some room for improvement remains. New developments and discoveries like base editing, prime editing, and CRISPR-associated transposons might address some of these challenges.

Introduction
Horizontal gene transfer plays a crucial role in the evolution of prokaryotes [1–4]. Invading genetic material may give fitness advantages to recipient cells, but also may pose a substantial risk. Therefore, bacteria and archaea have evolved a range of mechanisms to protect themselves from foreign nucleic acids [5]. One such defense mechanism is the CRISPR–Cas system (clustered regularly interspaced short palindromic repeats and CRISPR-associated). CRISPR–Cas has been shown to provide adaptive immunity against mobile genetic elements (e.g. viruses) in three stages: adaptation, expression, and interference (Figure 1)[6,7].

In the adaptation stage (Figure 1A), a complex of Cas1 and Cas2 (and sometimes Cas4) acquires nucleotide stretches called protospacers. Protospacers are selected based on the presence of a protospacer adjacent motif (PAM) and are preferentially derived from foreign DNA or RNA [8–12]. Acquired protospacers are incorporated as spacers in the CRISPR array on the host chromosome, where they are flanked by repeats [13]. Spacer acquisition has been demonstrated to also involve host proteins as well as CRISPR interference itself [14–19].

During the expression stage (Figure 1B), the CRISPR array is transcribed into a long precursor CRISPR RNA (pre-crRNA) that is subsequently processed into small CRISPR RNAs (crRNAs), each containing a single spacer flanked by parts of the repeat. Next, Cas proteins and the mature crRNA (also referred to as the guide RNA) assemble to form a ribonucleoprotein (RNP) complex.

In the interference stage (Figure 1C), RNPs recognize PAMs and interrogate adjacent sequences for complementary to the crRNA. The RNPs then trigger degradation of the foreign DNA or RNA, either by nuclease activity of the RNP itself or through recruitment of another nuclease, for instance, the helicase/nuclease Cas3 [20]. CRISPR–Cas systems are classified in Class 1 (type I, III, IV) and Class 2 (type II, V, VI) [21]. Class 1 systems use multi-subunit Cas complexes (consisting of multiple different Cas proteins) for interference. In contrast, Class 2 systems are characterized by their single, multi-domain protein nucleases.

Genome editing
Among applications of CRISPR–Cas, genome editing has attracted the most attention by far. Eukaryotic genome editing uses one of two major DNA double-stranded break (DSB) repair pathways: non-homologous end-joining (NHEJ) or homologous recombination (HR) [22]. NHEJ resolves DSBs...
in the DNA by ligating both ends, often resulting in small insertions or deletions (indels) [23]. In the case of targeted protein-coding genes, indels are likely to cause frameshifts, effectively knocking out these genes. HR, on the other hand, requires a homologous repair template (ssDNA or dsDNA), which is used for more precise DSB repair [22]. For genome editing, one can supply an artificial repair template with the desired sequence, flanked by sequences homologous to the targeted locus. HR will then incorporate the repair template to repair the DSB, thereby introducing the designed edit. Due to these differences in resolving DSBs, NHEJ is useful for knocking out existing genes, and HR can be used for precise insertions or replacement of DNA. Whereas eukaryotic cells possess both HR and NHEJ repair systems, the HR system is highly cell cycle dependent, being

Figure 1. CRISPR–Cas adaptive immunity.
Schematic overview of the steps involved in adaptive immunity as provided by CRISPR–Cas. (A) Adaptation stage, featuring the introduction of viral DNA/RNA (curved gray line) with protospacer 1 (dark blue) and an adjacent PAM (red). Protospacer 1 is incorporated in the bacterial genome as spacer 1 (dark blue) flanked by two repeats (orange) in the CRISPR array alongside spacer 2 (green). (B) Expression stage, wherein a Cas protein (light blue) is produced and pre-crRNA is transcribed and processed into mature crRNAs, which subsequently form functional RNP complexes with the Cas protein. (C) Interference stage, where a second infection is targeted for degradation by crRNA-guided Cas proteins, binding to matching protospacers adjacent to PAMs.
mostly active during S-phase [24]. DNA-cleaving Cas enzymes can be used to create DSBs in desired locations, making them useful in directing genome editing for eukaryotes. In prokaryotes, Cas nucleases are not used to direct DNA repair, but rather to cleave the original DNA sequence. These DSBs kill cells with the original DNA sequence, effectively counter selecting unedited cells and enriching for edited cells.

Where most other DNA nucleases rely on protein–DNA binding for target recognition, Cas nucleases use crRNA–DNA complementarity to determine specificity. New crRNAs can be designed and produced rapidly, providing superior versatility for target specificity of Cas nucleases. In this review, we focus on Class 2 Cas nucleases like Cas9 and Cas12a because these combine targeting versatility with the simplicity of single protein nucleases, making them the preferred tools for genome editing.

**Cas9**

Because of its early discovery, the most frequently used Cas protein is the type II nuclease Cas9. The characterization of Cas9 from *Streptococcus pyogenes* (SpCas9) highlighted the potential of CRISPR–Cas for genome editing applications [25,26]. Since then, Cas9 has been used for applications in organisms from all domains of life, ranging from improving microbial cell factories and crop plants to clinical trials aimed at repairing genetic disorders in human patients [27–36].

Cas9 is a single, multi-domain DNA nuclease that is guided by RNA [25]. The Cas9 protein can be divided into the recognition (REC) lobe and the nuclease (NUC) lobe [37] (Figure 2A). The NUC lobe consists of a PAM-interacting (PI) domain and two nuclease domains: HNH and RuvC. The archetype SpCas9 protein is a polypeptide of 1368 amino acids, but shorter variants from other organisms have also been described [38–40].

In type II systems, CRISPRs are transcribed as a single pre-crRNA. The maturation of the guides requires a transactivating crRNA (tracrRNA), which partially base pairs to the repeat regions of the pre-crRNA [41]. In the presence of Cas9, these partial dsRNA fragments are processed by the non-Cas ribonuclease RNase III [41]. To circumvent this processing step, the mature tracrRNA and crRNA can be fused by a short synthetic linker, and expressed as a single-guide RNA (sgRNA) that binds Cas9 to form a functional RNP complex [25].

The resulting RNP complex surveys the cells for a complementary target sequence. Single-molecule analysis revealed that the Cas9 RNP complex encounters DNA stretches through three-dimensional diffusion [42]. In the absence of a PAM, these RNP–DNA interactions are very short-lived [42]. If the RNP encounters a PAM, however, binding lasts substantially longer [42]. The PAM can, therefore, be regarded as an initial quality control that Cas9 uses in its surveillance for DNA targets and might allow the nuclease to find target DNA sequences more efficiently. SpCas9 recognizes its PAM (5’-NGG), directly downstream of the target sequence (Figure 2D) [25]. As the 5’-NGG motif occurs quite often, the PAM is generally not limiting Cas9 applications, unless a very precise cleavage position is desired. It should be noted that PAM recognition by Cas9 can be either strict or very relaxed in different Cas9 orthologs [39,43–45].

PAM binding by Cas9 enables local separation of the DNA strands directly upstream of the PAM, allowing interrogation of the target sequence for base pairing with the crRNA [42,46]. Successful base-pairing interactions between the interrogated target DNA strand and the crRNA extend the melting of the DNA strands [42]. The DNA strand displacement proceeds towards the PAM-distal region, eventually resulting in a complete R-loop configuration [42] (Figure 2B–D). Because of the direction (PAM proximal to distal), PAM-proximal matching between crRNA and the target DNA are disproportionately important for RNP binding [42,47]. This important PAM-proximal stretch is referred to as the seed region; in case of mismatches between this seed and the target strand, R-loop formation, and subsequent cleavage are generally aborted [25,42,47]. A certain level of mismatches in the PAM-distal target sequence are tolerated, which is advantageous for its natural role in defense against viruses with high mutation rates. However, in the case of genome editing, this tolerance can result in undesired binding/cleavage of imperfectly matching DNA sequences, referred to as off-targets [48,49]. More specifically, efficient DNA binding by Cas9 requires at least nine continuous matches between crRNA and target sequence [47]. Sufficient base pairing between crRNA and the DNA target strand stabilizes the R-loop, which is further stabilized by the interaction between the displaced DNA strand and charged residues of the Cas9 protein [37,50]. If the R-loop extends to the most PAM-distal bases, a conformational change is induced that poises the HNH domain in the NUC lobe for cleavage [51].

The HNH domain then cleaves the target strand, which consistently creates a break between the 3rd and 4th nucleotides upstream of the PAM [25,52]. The conformational change in the HNH domain is also required for
nuclease activity of the RuvC domain which cleaves the displaced strand, generally between either the 3rd and 4th or between the 4th and 5th bases upstream of the PAM [25,53,52]. Collectively, this results in a DSB with blunt ends or a short overhang. Notably, cleavage by Cas9 falls within the seed region of the target sequence [25]. If the dsDNA break is repaired through NHEJ, the resulting insertions or deletions will lead to disruption of the seed region, and hence the altered target sequence is no longer cleaved.

After cleavage, Cas9 remains stably bound to both ends of the protospacer. While the targeted strand is in heteroduplex with the crRNA, the cleaved ends of the displaced strand are less structured and might engage the RuvC active site repeatedly. This would explain the observed time-dependent exonuclease activity on the displaced strand after initial cleavage [52].
Cas12a
Although Cas9 has proven to be a powerful tool for genome editing, other Cas nucleases may be interesting alternatives. The type V nuclease Cas12a (previously called Cpf1) is a good example of this. Despite having been characterized several years later than Cas9, Cas12a offers some advantages that may favor the latter enzyme for specific applications. One of the most used variants of Cas12a is the enzyme from *Acidaminococcus sp. BV3L6* (AsCas12a). Like Cas9, Cas12a has by now been applied in a wide range of species including bacteria, yeast, plants, and in human cells [54–58].

As a Class 2 nuclease, Cas12a is a single multi-domain protein that induces RNA-guided dsDNA cleavage [59]. Like Cas9, the Cas12a structure can be subdivided in a REC and a NUC lobe [60,61] (Figure 3A). The NUC lobe of Cas12a lacks an HNH domain, but does contain a RuvC domain, for DNA cleavage [60]. In

![Figure 3. AsCas12a.](image)

**Figure 3. AsCas12a.**
(A) The layout of the AsCas12a protein from C- (left) to N-terminus (right). Featuring the REC lobe (gray) and the domains of the NUC lobe: WED (pink), PI (red), RuvC (light blue), BH (green), Nuc (yellow). The last residue position for each domain is given below the domains. (B) Full structure (PDB: 5B43 [60]) of AsCas12a in complex with crRNA (orange), target DNA strand (dark blue), and displaced DNA strand (black). Protein domains are colored in accordance with (A). Visualized using PyMOL [109]. (C) Same as B with some residues from the REC lobe and WED domain removed to allow a view of the sgRNA and DNA. (D) Schematic overview of the R-loop formed by Cas12a RNP. DNA and RNA strands are colored in accordance with (B). The direction of R-loop formation is indicated with a dashed arrow. Approximate RuvC cleavage positions in the DNA are represented by (light blue) triangles, but both strands are cleaved by a single catalytic site [62].
Indeed, Cas12a has been found to be less toxic than Cas9 in some contexts [54,74]. The displaced strand to be trimmed likely due to subsequent interactions with the RuvC catalytic site [69]. In contrast with the relatively stable interaction of Cas9 with its cleavage products, the Cas12a PAM-distal cleavage product is readily released after cleavage [42,47,68]. This release of the PAM-distal DNA might allow ssDNA activity on ssDNA by Cas12a after cleavage of an initial specific ssDNA activity is unlikely to be problematic as DNA hardly occurs in a single-stranded state [42]. Because of this, there might be a strict trade-off between PAM range and on-target activity.

**Challenges for genome editing**

Although both Cas9 and Cas12a are highly promising tools, some room for improvement remains for their optimal performance in genome editing. One main constraint of Cas nucleases is that the targeting space is limited by the requirement for the presence of a properly positioned PAM. To overcome this limitation, rational engineering and/or random mutagenesis have been used to produce Cas9 and Cas12a enzymes with altered, or less stringent PAM recognition [46,75–78]. It is tempting to propose efforts towards a hypothetical PAM-less Cas nuclae. However, the PAM recognition is generally coupled to the initial opening of the dsDNA target. In addition, more permissive Cas nucleases would spend more time on non-target sequences [42].
In contrast with Cas nucleases, prokaryotic Argonaute proteins are a class of guided nucleases that do not require a PAM for their activity. Because currently characterized prokaryotic Argonaute proteins do not trigger DNA unwinding, their dsDNA cleavage activity is far lower than for Cas9 and Cas12a [79,80]. Hence, it will be a major challenge to develop an efficient PAM-independent Cas nuclease.

An alternative to PAM-less nucleases is to work towards a comprehensive ‘toolbox’ of Cas nucleases that each recognize a different PAM. A Cas nuclease with an appropriate PAM could then be chosen based on the desired target site. The current targeting range is most limiting for applications where the exact location of cleavage is critical, including single nucleotide engineering by base editing (see below) [75].

The presence of a PAM does not always guarantee efficient protospacer cleavage. In addition to the PAM, the nature of the crRNA/protospacer sequence has been found to strongly affect the cleavage efficiency of Cas nucleases [48,49]. Apart from target accessibility related to chromatin structure, the variable parts of the (pre-)crRNA sequences may result in the formation of inappropriate secondary structures. Mis-folding of these guides will affect the formation of functional RNP complexes, resulting in decreased editing efficiency.

However, high cleavage efficiency does not always result in efficient eukaryotic genome editing. HR efficiency in eukaryotic cells varies substantially depending on the cell type and on the cell cycle phase. In addition, the NHEJ system may outcompete the HR system in repairing DSBs, which might lead to low levels of the desired HR. In conclusion, for precise genome editing, HR efficiency is a major bottleneck, rather than DSB formation by Cas nucleases. Hence, apart from strategies to improve HR, HR-independent strategies (see below) may be promising alternatives [81].

Possibly the most important challenge for applications of Cas nucleases is to solve the problem of off-target cleavage. Indeed sequences partially complementary to the crRNA have been shown to be cleaved at relevant rates [48,82]. For human therapeutics, it is critical that these off-targets are reduced to the bare minimum to enable safe genome editing. Compared with Cas9, Cas12a is substantially more specific due to its decreased tolerance for mismatches during DNA/RNA heteroduplex formation, at least in vivo [70]. Characterization of more natural Cas nucleases might reveal even more specific enzymes. In addition, rational engineering combined with laboratory evolution has been used to successfully generate several Cas9 variants with substantially increased specificity [50,83–88].

**Beyond DNA cleavage by Cas9 and Cas12a**

Several exciting developments (Table 1) have recently been established in the field of Class II CRISPR nucleases. These developments further highlight the potential of the programmable DNA-specificity of Cas nucleases. Besides Cas9 and Cas12a, other Cas nucleases may prove useful genome editing tools, including Cas12b and Cas12e nucleases for which genome editing has already been demonstrated [63,89]. In addition, a designed variant of a Class 1 complex (Cascade fused to a FokI nuclease domain) has recently been described as an alternative genome editing system [90].

Active nucleases like Cas9 and Cas12a can be converted into nickases or be rendered catalytically inactive by mutating conserved residues in the catalytic sites [25,91,59,60]. Binding of such catalytically inactive Cas nucleases can be used to inhibit gene transcription (CRISPR inhibition, CRISPI), e.g. by blocking a promoter to prevent binding of RNA polymerase [91–93]. On the other hand, fusions of a Cas nuclease to the RNA polymerase omega subunit can be used to enhance the recruitment of RNA polymerase, promoting expression (CRISPR activation, CRISPARa) of downstream genes [92–94]. This type of targeted gene regulation can be used to study gene functions and is a versatile tool in the field of synthetic biology.

Nickase and catalytically inactive versions of Cas nucleases also offer a tantalizing alternative to the DSB-induced genome editing. The programmable specificity of Cas nucleases can be used to direct fused cytidine or adenine deaminase enzymes to specific DNA stretches. Such base editing has so far been reported for conversions of C*G pairs to T*A (using cytidine deaminase fused Cas9 or Cas12a), or A*T to G*C (using an evolved and engineered DNA adenine deaminase fused Cas9) [95–98].

Compared with DSB-repair-based genome editing, base editing offers major advantages. Most importantly, base conversions do not rely on HR, NHEJ or other mechanisms of DSB repair. Therefore, base editing could theoretically reach higher efficiencies as it is less dependent on cell type and cell cycle phase. In addition, because no DSB is introduced with base editing, it is unlikely to result in major chromosome rearrangements [99]. Recently, cytosine base editing (but not adenine base editing) was found to induce substantial off-target base substitutions on DNA, possibly due to protospacer independent base conversions [100,101]. In addition, both types of base editors were found to convert off-target bases in cellular RNA, although such RNA activity...
could be reduced through protein engineering [102,103]. These studies highlight the importance of genome/transcriptome wide off-target analysis on single cell-derived DNA. Another challenge for base editing is the requirement for precise targeting, which is often limited by PAM availability (see above). Lastly, it is currently difficult to target a single base position for conversion, instead, current base editors may convert any appropriate base within a certain window (although these windows can be very narrow). Whereas this is no problem for gene disruption, it may be a major hurdle for precision gene therapy. Future optimization of base editors is required to address these challenges.

The recent development of prime editing potentially offers a much more versatile method for DSB-independent genome editing [104]. Prime editing uses an extended version of the guide RNA in conjunction with a nickase version of Cas9 fused to reverse transcriptase. The RNA guides the Cas9 fusion to the complementary target on the genome, where Cas9 nicks the displaced strand. The reverse transcriptase then uses the extra part of the guide RNA as a primer, incorporating the guide-encoded edit in the DNA of the displaced strand. Prime editing enables small insertions, deletions, and replacements to be made independent from NHEJ or HR.

Another exciting development is the application of recently discovered CRISPR-associated transposons for both Class I and Class II CRISPR systems [105–108]. These transposons rely on RNA-guided Cas enzymes to direct transposition to positions downstream from protospacers. The involved Cas proteins do not introduce DSBs, but rather guide the transposon proteins [107]. This targeted transposase activity has already been used to integrate large stretches of DNA in the genome of *Escherichia coli* at high efficiency [107,108]. Although it requires transposon proteins in addition to the Cas nuclease protein(s), this method provides a promising alternative to HR-based genomic insertions. The next challenge will be to successfully use these systems for programmable integrations in mammalian cells.

### Perspectives

- The CRISPR–Cas defense system of prokaryotes holds great promise for genome editing in a broad spectrum of organisms, across the entire tree of life. In fact, Cas nuclease-based cleavage and binding of DNA have already proven extremely useful in fundamental research and in microbial and plant biotechnology.
Apart from the initial biochemical and genetic analyses, a range of high-resolution structures and single-molecule studies have shed light on the molecular mechanism of DNA cleavage by Class 2 Cas nucleases. It has also become apparent that there is room for improving targeting space, specificity, and efficiency for genome editing.

Engineering by rational design and/or laboratory evolution may result in improved variants of the current set of Cas nucleases, potentially addressing recent challenges. Moving forward, we expect more developments in engineering Cas nucleases for applications like base editing and prime editing. Finally, it is anticipated that future research will reveal CRISPR systems with new Cas proteins, potentially offering new functions as exemplified by the newly discovered role in RNA-guided transposition.

Abbreviations

AsCas12a, Acidaminococcus sp. BV3L6 Cas12a; BH domain, bridge-helix domain; Cas, CRISPR-associated; CRISPR, clustered regularly interspaced short palindromic repeats; CRISPRa, CRISPR activation; CRISPRi, CRISPR inhibition; crRNA, CRISPR RNA; CTD, C-terminal domain; DSB, double-stranded break; HR, homologous recombination; Indels, insertions and deletions; NHEJ, non-homologous end-joining; Nuc domain, ‘Nuclease’ domain; NUC lobe, nuclease lobe; PAM, protospacer adjacent motif; PI domain, PAM-interacting domain; pre-crRNA, precursor CRISPR RNA; REC lobe, recognition lobe; RNP, ribonucleoprotein; sgRNA, single-guide RNA; SpCas9, Streptococcus pyogenes Cas9; Topo domain, Topo-homology domain; tracrRNA, transactivating crRNA; WED domain, Wedge domain.

Author contributions

T.S. wrote the first draft of the manuscript. All authors have reviewed and revised the manuscript and approved the final version.

Funding


Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

References


