CRISPR/Cas9, a decade of genome editing tools to fix the DNA

Lúcia Santos, PhD
Department of Physiology, University College Cork, Cork T12 K8AF, Ireland; lsantos@ucc.ie

https://doi.org/10.57098/SciRevs.Biology.1.1.3
Received June 30, 2022. Accepted July 14, 2022.

Abstract: This year marks the 10th anniversary of the CRISPR/Cas9 genome editing discovery. Since its discovery in 2012, the CRISPR/Cas9 system has become an indispensable tool in many research fields. This system has been extensively characterized and further optimized to broaden its editing capabilities. Depending on the DNA modification to make, there are now available several editing agents. In this review, we provide an overview of the CRISPR/Cas9 system and how it can be used to fix the DNA using the traditional repair mechanisms non-homologous end joining (NHEJ) and homology-directed repair (HDR), and the most recent gene editing approaches - base editing and prime editing.

Keywords: microbiology, CRISPR, gene editing, kill switch

1. The CRISPR/Cas9 system

Jennifer Doudna and Emmanuelle Charpentier reported for the first time in 2012 the potential of CRISPR/Cas9 as a genome editing tool, a discovery that yielded the researchers the Nobel Prize in Chemistry in 2020 (1). Since then, the ability of Cas9 nuclease to introduce site-specific changes in the DNA has been extensively studied in many different research fields, from medicine to agriculture. However, the clustered regularly interspaced short palindromic repeats/CRISPR-associated proteins (CRISPR/Cas) have their origin in the adaptive immune system of archaea and bacteria (2,3). This defense mechanism uses RNA-guided nucleases to cleave foreign genetic elements and consists of three main stages: acquisition, expression, and interference (4). In the acquisition stage, a complex of Cas proteins binds to the invading genetic elements and cleaves a portion of the target DNA, called protospacer (4). Then, at the expression stage, the spacers are transcribed and processed into mature CRISPR RNAs (crRNAs) and, lastly, at the interference stage, the Cas protein is guided by the crRNA to recognize and cleave foreign nucleic acid molecules (4–6).

The CRISPR systems can be classified into two different classes (class I and class II), which are further divided into six different types (type I - VI) (4). The composition of the effector module distinguishes CRISPR/Cas class I from class II (4). The class I effector module has several Cas proteins that work together to bind and process the target, while class II systems have a unique crRNA-binding protein that is equivalent to the whole class I effector module (4). Class I includes types I, III, and IV, and class II the types II, V, and VI (4). The different types of CRISPR/Cas systems recognize and cleave DNA (type I, II, and V), RNA (type VI), or both (type III) (7). The effect of type IV on DNA or RNA is still unknown (7).

1.1 CRISPR-associated protein 9

CRISPR-associated protein 9 (Cas9) belongs to the class II-type II CRISPR system (1). In type II systems, an additional RNA molecule is needed – the transactivating CRISPR RNA (tracrRNA) (4). This RNA molecule is responsible for: i) the pre-crRNA processing by RNase III, forming the mature guide RNA (gRNA) composed of the crRNA:tracrRNA complex, and ii) activation of the crRNA-guided DNA cleavage by Cas9 (1). The crRNA has 42 nucleotides (nt), the first 20-nt at the 5'-end correspond to the spacer sequence, and the other 22-nt pair with the 5'-end of the tracrRNA (1). The remaining nucleotides of the tracrRNA are free to interact with the Cas9 protein (1). The 10-12 nucleotides at the 3'-end of the 20-nt crRNA form the seed sequence that confers DNA targeting specificity (8). While mis-
matches in this sequence impair target DNA binding and cleavage, a high level of homology with other DNA regions lead to off-target effects (i.e. the ability of the gRNA to recognize other than the target DNA sequences) (8). To avoid off-target effects, researchers have engineered different Cas9 proteins to produce high-fidelity Cas9 variants that have reduced non-specific DNA interactions maintaining on-target activity (9–14).

The Cas9 protein has a bilobed structure composed of the recognition (REC) lobe and the nuclease (NUC) lobe. As the names indicate, the REC domain recognizes the gRNA sequence, and the NUC lobe cleavages the double-stranded DNA. In addition, the NUC lobe recognizes the protospacer-adjacent motif (PAM) sequence (8,15). The NUC lobe is further divided into two domains - the HNH and RuvC domains - each responsible to cut one of the DNA strands. The first cuts the strand complementary to the gRNA sequence (target strand) while the second cleaves the DNA containing the PAM sequence (non-target strand) (Fig. 1) (15).

Mutations in these nuclease domains, either HNH (H840A) or RuvC (D10A), produce nickase variants of the Cas9 (nCas9), which induce nicks in only one of the DNA strands (16). Furthermore, when both domains carry these mutations that result in a nuclease deactivated Cas9 variant (dCas9) lacking its catalytic activity (16).

After binding the gRNA, the Cas9 protein becomes catalytically active and searches for a suitable PAM sequence (8). This is a 3-nt sequence located downstream of the spacer in the nontarget sequence. The PAM sequence varies according to the organism it derives from (8). Streptococcus pyogenes (SpCas9) is the most used Cas9 nuclease, and it recognizes any 5′-NGG-3′ sequence (“N” stands for any nucleotide). Once a suitable PAM sequence is found, the gRNA binds the target DNA sequence, and if there is perfect complementarity between the two, the Cas9 cleavages the double-stranded DNA 3-nt upstream of the PAM sequence (Fig. 1) (8). Even though the occurrence of an “NGG” PAM is relatively common in the human genome, the need for a specific motif for targeting limits the DNA target sites to a subset of sequences. To overcome this limitation, researchers have engineered SpCas9 variants that recognize a wider array of PAM sequences (17–25).

![Figure 1 - CRISPR/Cas9 genome editing](image)

The gRNA (in green) binds the target region and the Cas9 nuclease cuts the double-stranded DNA 3 bp upstream of the PAM sequence (in red). The DNA can then be repaired by two different mechanisms: non-homologous end joining (NHEJ) or homology-directed repair (HDR). In NHEJ, a mixture of nucleotides can be inserted or deleted at the cut site, forming indels. If a donor DNA template is provided the HDR repair mechanism is triggered, and the DNA is precisely repaired. Key: Inserted nucleotides in yellow; Precise edit in purple.
2. Genome editing tools

2.1 CRISPR/Cas9 genome editing

As mentioned above, following gRNA binding to the complementary DNA sequence the Cas9 nuclease cleaves the DNA creating a double-stranded break (DSB). The default mechanism by which the DNA can be repaired is called non-homologous end joining (NHEJ) (2). This repair mechanism introduces a mixture of nucleotide insertions and deletions (indels) at the cut site, which can cause gene knockout when occurring at coding regions (Fig. 1) (2). Besides indel formation, DSB is also associated with chromosomal translocations and p53 gene activation (26–28).

Another DNA repair mechanism is homology-directed repair (HDR). This is a high-fidelity repair mechanism; however, it is less efficient than NHEJ as it mainly occurs in the S phase of the cell cycle (29). In the presence of a repair template containing homology arms flanking the desired edit, HDR is triggered and leads to precise repair of the genome (Fig. 1) (2). Nevertheless, as the two repair mechanisms can occur in the same cell in different alleles, even when an exogenous donor DNA template is used there is indel formation.

2.2 Base Editing

To introduce specific single nucleotide changes in the DNA and avoid the unwanted indels created by DSB, researchers have developed base editing. These genome editing tools allow the irreversible conversion of one base into another in a direct and programmable manner. Contrary to CRISPR/Cas9-mediated HDR, base editing does not require a donor DNA template and as a Cas9 nickase variant is used it does not generate DSB and, consequently, very low levels or no indels are produced. To date, base editors enable the introduction of all four transition mutations (C-to-T; T-to-C; A-to-G; G-to-A) and two transversions (C-to-G and G-to-C).

2.2.1 Cytosine base editing

The first base editors to be developed were the cytosine base editors (CBEs), in which a Cas9 (D10A) nickase variant is fused with a cytidine deaminase (30). Cytidine deaminase converts the C:G pair into a U:G mismatch (30) (Fig. 2). Then, Cas9 nickase cuts the non-edited DNA strand (target strand), favoring U:G to U:A editing (Fig. 2) (30). Cytidine deaminases, such as APOBEC1 and cytidine deaminase 1 (CDA1), can deaminate any cytidine in positions 4-8 of the protospacer (editing window), counting PAM sequence as positions 21 to 23 (Fig. 2) (30,31).

Since the development of the first cytosine base editor version (BE1), many enhancements have been done to improve editing efficiency. Besides the deaminase and the Cas9 nickase, the last base editor version (BE4max) has modified nuclear localization signals (NLS) and codon usage, and two uracyl N-glycosylase inhibitor (UGI) domains were also added (32). These two UGI domains inhibit uracil N-glycosylase (UNG), part of the base excision repair (BER) pathway, avoiding U:G mismatch recognition and reversion back to C:G pair.

One of the main disadvantages of CBEs and other base editing systems is that deaminases not only change the target nucleotide but also all others present in the editing window. To overcome this problem, base editors with narrower editing windows have been developed (33). Another disadvantage is that not always there is a suitable PAM sequence available that puts the target base in the correct editing window. The use of different Cas9 variants, such as SpRYCas9, can increase the range of target sequences (24,25).

2.2.2 Adenine base editing

The adenine base editors (ABEs) were developed following the same rationale behind CBEs development (34). These base editors deaminate any adenosine in the non-target DNA strand and convert it into inosine (I) (Fig. 2) (34). Inosine pairs with cytosine enabling the conversion of an A:T base pair
into a G:C pair (Fig. 2) (34). As there are no enzymes in Nature known to deaminate adenosine in DNA, to develop ABEs, Escherichia coli (E. coli) tRNA adenosine deaminase enzyme (TadA) was evolved to function on DNA (34). The ABEs result from the fusion of wild-type non-catalytic TadA monomer and the evolved TadA monomer with a nickase Cas9 (34). Typically, the ABEs can convert any A:T to G:C in positions 4–7 of the protospacer (34–36).

2.2.3 C-to-G base editing

The first base editors developed (CBEs and ABEs) were intended to introduce the four base transitions, however, base transversions have occasionally been observed as byproducts of these base editors (37). To develop the C-to-G base editor (CGBE1), researchers took advantage of these unexpected editing outcomes and engineered BE4max (see section 3.2.1) by removing the two UGI domains and adding an E. coli UNG (eUNG) enzyme to its carboxy-terminal (38). A shorter version, miniCGBE1, lacking the eUNG domain was also developed. These C-to-G editors can edit any cytidine at positions 5–7 in the protospacer, being position 6 the most efficient (Fig. 2) (38).

2.3 Prime editing

As mentioned above CRISPR/Cas9 can cut the double-stranded DNA that can be repaired by the NHEJ mechanism, introducing a mixture of indels at the target site. However, there are occasions, for example, genetic diseases, where a specific gene editing approach is required. The homology-directed repair can be used to install precise DNA changes, but it relies on an exogenous donor DNA template, indel formation is not completely avoided, and it is inefficient in most relevant cell types. As an alternative, researchers developed base editing however, this editing approach is not able to perform most nucleotide transversions nor targeted insertions or deletions. To overcome this problem, researchers came up with a new versatile and precise genome editing method called prime editing. (39). Prime editing uses Cas9 (H840A) nickase fused to an engineered M-MLV reverse transcriptase (RT) and a prime editing gRNA (pegRNA) that not only specifies the target site as well as it encodes the desired edit (39). The Cas9 nickase cuts the nontarget strand of the DNA exposing a 3’ flap that binds to the primer binding site (PBS) of the RNA template serving as a primer for the RT (Fig. 3) (39,40). Then, the RT extends the 3’ flap and copies the edit sequence of the pegRNA (39,40). The endogenous endonuclease FEN1 excises 5’ flaps and the edited 3’ flap hybridizes with the unedited complementary strand (Fig. 3) (39,40).

However, as only one strand of DNA is edited there are mismatches formation that can be resolved naturally in favor of the desired editing (40). To further improve prime editing efficiency, researchers co-transfected a standard gRNA targeting the complementary strand allowing the Cas9 nickase to nick the unedited strand. Nicking the unedited strand bias mismatch repair (MMR) in favor of the edited sequence by using the edited DNA strand as a template (39,40). Another strategy for prime editing improvement is the co-expression of a dominant negative MMR protein (MLH1dn) to transiently inhibit MMR and, consequently, enhance editing efficacy (41).

Figure 3 – Prime editing. The pegRNA complex (in green) binds to the target region and the Cas9 (H840A) nickase cuts the non-target DNA strand 3bp upstream of the PAM site (in red). DNA nicking creates a 3’ flap that interacts with the primer binding site (PBS) located at the 3’ end of the pegRNA. The DNA/RNA hybrid serves as a primer site for the new DNA synthesis and RT polymerase uses the RT template (in light purple) to extend the 3’ flap and copy the edit (in dark purple) also present in the pegRNA. The unedited 5’ flap is removed by FEN1 (in black) and the edited 3’ flap hybridises with the unedited complementary strand resulting in precise DNA editing.
3. Conclusion

Since its discovery, CRISPR/Cas9 technology has been evolving at a fast pace. This technology has revolutionized genetic engineering, enabling many advances in medicine (e.g. treatment of human genetic diseases) or in agriculture (e.g. improvement of food crops) (42). It evolved into a precise genome editing tool that allows making nearly any DNA change with almost no undesired editing byproducts (43). However, more efforts are needed to further improve CRISPR/Cas9 editing capabilities and to understand the consequences of editing the genome.

References


19. Lee CM, Cradick TJ, Bao G. The neisseria meningitidis CRISPR-Cas9 system enables specific genome editing in mammalian cells. *Molecular Therapy.* 2016;24(3). https://doi.org/10.1038/mt.2016.8


**Conflicts of Interest**
The authors state no conflict of interest.