The CRISPR-Cas9 System: A New Dawn in Gene Editing

Nurul Husna Shafie*, Mohamed Saleem, Emmanuel Jairaj Moses, Siti Razila Abdul Razak, and Narazah Mohd Yusoff*

Advanced Medical and Dental Institute (AMDI), Universiti Sains Malaysia, Bandar Putra Bertam, 13200 Kepala Batas, Penang, Malaysia

Abstract

The CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats with CRISPR-associated protein 9) system is a genome editing system that is easy to design, highly specific, efficient, robust, and well suited for high-throughput and multiplexed gene editing for a variety of cell types and organisms. In this review we describe current applications of this new system, which is growing in popularity and is increasingly being employed to selectively control gene expression on a genome-wide scale.

Keywords: Genome editing; CRISPR; Gene expression

Abbreviations: ZFN: Zinc Finger Nuclease; TALEN: Transcription Activator-Like Effector Nucleases; CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats; Cas: CRISPR Associated Endonuclease; ZFP: Zinc Finger Peptide; DSB: Double Stranded Break; sgRNA: Single Guide RNA; crRNA: Crispr RNA; trRNA-Trans Activating RNA; PAM: Protospacer Adjacent Motif; HR: Homologous Recombination; NHEJ: Non-Homologous End Joining; DMD: Duchenne Muscular Dystrophy

History of Gene Editing

Gene editing technology is a new tool that can be used to introduce targeted modifications into the genome. Currently, there are three well-defined technologies for gene editing: Zinc Finger Nucleases (ZFNs), Transcription Activator-Like Effector Nucleases (TALENs), and Clustered Regularly Interspaced Palindromic Repeats (CRISPRs) with CRISPR-associated (Cas) nucleases. Each of these systems is characterized by an adaptable sequence-specific DNA binding domain and a nuclease domain that creates a double-strand cleavage. The site-specific DNA binding domains of the ZFN and TALEN systems are based on chimeric protein, whereas the CRISPR-Cas system utilizes an RNA molecule.

ZFN technology represents the first generation of engineered nucleases for genome editing. ZFNs consist of an engineered Zinc Finger Peptide (ZFP) fused to the cleavage domain of the FokI restriction enzyme. The engineered DNA binding protein domain specifically localizes the ZFN at a predetermined location in the genomic sequence of interest and facilitates targeted genome editing by creating double-strand cleavage by the catalytic domain of the FokI endonuclease. ZFNs have relatively low resolution in unique sequence recognition and hence low specificity. This is partly because the ZFP region contains a tandem array of Cys2-His2 zinc fingers, and each motif (finger) recognizes 3 base pairs (bp) of nucleotides [1]. Thus, unique target sequence recognition in the eukaryotic genome is difficult. Over time the method has been improved considerably. For example, to improve the DNA binding specificity of ZFN, the array of fingers in the ZFP has been optimally increased from three to six to bind an 18 bp target, which enables the ZFN dimers to specify 36 bp of DNA per cleavage locus [2]. Even with this improvement, however, the tool remains inefficient and may confound biological interpretations.

TALENs, the second generation of genome engineering nucleases, were discovered in 2010 [3] and gained rapid momentum among researchers because these nuclease molecules permit more predictable and specific binding to target DNA [4]. A TALEN has two functional domains: a transcription activator-like effector domain, which is an oligopeptide array of modules (each module constitutes 33–35 amino acids) from the bacterium Xanthomonas sp., and a cleavage domain of FokI nuclease. Because there are four different modules, one for each nucleotide base, constructing customizable sequence-specific TALENs is a convenient method to target nearly any sequence of interest [5,6]. TALENs function as obligate heterodimers in which each monomer binds 15-20 bp of DNA that flanks a 12-24 bp spacer region. Target DNA sequence recognition occurs in the central domain of the tandem repeats. Compared with ZFNs [7], TALENs are easy to manufacture, several times cheaper, and functionally better than ZFNs. However, because TALENs are much larger molecules than ZFNs, efficient delivery into the cell may be difficult.

CRISPR together with Cas proteins form the CRISPR-Cas system [8], which is the newest gene editing method. CRISPRs constitute a family of short DNA repeats that are important components of the adaptive immune system in bacteria and archaea. These elements protect the microbes against various viral invasions. Cas proteins have functional domains that are similar to nucleases, helicases, polymerases, and polynucleotide-binding proteins [9]. Originally, the CRISPR-Cas system was divided into eight subtypes [9]. However, a new system is introduced because the previous classification did not take into consideration the distant relationships between various Cas proteins [10]. In this new classification, the CRISPR-Cas system is divided into three different types. The type I and III systems involve the specialized Cas endonucleases which process the pre-crRNAs and once mature, the crRNA will assemble into a large Cas protein complex. The complex is capable to recognize and cleave nucleic acids complementary to the crRNA [11]. The type II, which discuss here, CRISPR-Cas9 system is characterized as a small RNA-based immune system.
system of bacteria and archaea [8], and recently it was developed for efficient genome engineering [12,13]. Cas9 is an RNA-guided DNA nuclease enzyme that provides an effective means of introducing targeted loss-of-function mutations at specific sites in the genome [12,14] by generating DNA Double Stranded Breaks (DSBs) at specific genomic loci. This system is easy to design, highly specific, efficient, and well-suited for high-throughput and multiplexed gene editing for a variety of cell types and organisms [15].

An engineered single guide RNA (sgRNA) containing a CRISPR RNA (crRNA) and a partially complementary trans-activating RNA (tracrRNA) are necessary and sufficient for genome editing [12]. The first 20 nucleotides located at the 5′-end of sgRNA, called the spacer, direct the Cas9 nuclease to the complementary 20 nucleotides of the target sequence, where they hybridize. A highly conserved secondary structure downstream of the spacer, called the protospacer, in the presence of a short nucleotide motif (known as the Protospacer Adjacent Motif (PAM)) (Figure 1) must be recognized by the CRISPR-Cas9 nuclease complex before cleavages occurs [13-16]. The CRISPR-Cas9 system’s potential to target genomes appears to require only a PAM sequence (NGG). This simplicity makes it a cutting edge genome editing tool.

Processes in Genome Editing

Developing a genome editing tool requires engineering endonucleases that can create highly efficient and accurate DSBs at a user defined location in the genome and subsequently activate the cellular pathways involved in DSB repair processes via Homologous Recombination (HR)-mediated gene repair or Non-Homologous End Joining (NHEJ). HR uses homologous DNA sequences as templates for precise repair. It involves strand invasion and requires a homologous DNA template to precisely edit a genomic sequence or insert exogenous DNA that results in gene knock out or gene knock in. NHEJ is an error-prone ligation process that results in small insertions or deletions (indel mutations). This process involves the re-ligation of the two broken ends at the cleavage sites and is catalyzed by DNA ligases [16,17]. Indel generation is exploited as a convenient method for gene silencing (knock-out mutation). CRISPR is capable of modifying the chromosomal target by indel mutations at high frequency [18]. In addition, CRISPR-Cas9 allows the simultaneous targeting of several sequences for multiplexed gene editing [14,19] and has the potential for gene replacement by concurrently targeting the sequences upstream and downstream from a given locus [20].

The CRISPR-Cas system is a prokaryotic immune system. The type II CRISPR-Cas9 system is widely used for efficient genome editing [14,19,21-23] and the establishment of gene silencing [24]. When in complex with sgRNA (Figure 1), Cas9 introduces DSBs in a target sequence that is homologous to the spacer moiety of crRNA [11]. The generation of DSBs in the target DNA initiates the genome editing process. Chromosomal DSBs trigger DNA repair either by HR or NHEJ in the absence of a homologous repair template (Figure 2). These repair systems can be harnessed for genome editing [25].

Genome-scale CRISPR knock-out screening

Recent progress in modifying CRISPR has led to extremely efficient gene disruption in many model organisms [12,19,26-28]. For example, the generation of gene encoding serine threonine kinase ROP18 knock outs in the type 1 GT1 strain of Toxoplasma gondii using CRISPR-Cas9 extends reverse genetic techniques to diverse isolates of T. gondii [29]. Chen et al. initially demonstrated that DSBs can be engineered at precise locations in the Caenorhabditis elegans genome by injecting the CRISPR-Cas9 complex, resulting in gene knock out [30]. DSBs also can be used for transgene- instructed gene conversion and allow for the systematic study of gene function in this widely used model organism [30]. High rates of mutagenesis efficiency (75-99%) have been reported when testing one homozygous egfp reporter gene and four endogenous loci in zebra fish [31]. These five genomic loci can be targeted simultaneously, resulting in multiple loss-of-function phenotypes in the same injected fish, which in turn support multiple biallelic gene inactivations [31]. Similarly, the simultaneous disruption of five
genes (Tet1,2,3, Sry, Uty-8 alleles) in mouse embryonic stem cells was observed when CRISPR-Cas-mediated gene editing was employed. Further co-injection of Cas9 mRNA and sgRNAs targeting Tet1 and Tet2 into zygotes generated mice with biallelic mutations in both genes with an efficiency of 80%. This shows that CRISPR-Cas9 is highly efficient at simultaneous targeting of multiple genes in stem cells and mice [19].

In another study, the eyeless gene in Daphnia magna was knocked out using CRISPR-Cas, suggesting that it is a useful marker gene in this system [32]. The CRISPR-Cas9 system also has been used to generate stable knock-out cell line models in human endometrial cell lines [33], human myeloid leukemia cells [34], and human melanoma cells [35] for genetic screening. These findings suggest that the knock-out cell line models generated by the CRISPR-Cas9 system could be used to complement mouse knock-out models, which will offer a new strategy for investigating the function of genes in differentiated cells and tissues. For example, studies of the disruption of four genes (ApoE, B2m, Prf1, and Prkdc) in rats by co-injection of Cas9 mRNA and sgRNA into one-cell fertilized eggs demonstrated the potential of the CRISPR system to efficiently and reliably generate gene knock-out rats [36]. Furthermore, RNA guided endonucleases (RGENs) containing Cas9 protein and sgRNA complexes efficiently induced mutations in the mouse Prkdc gene in up to 93% of newborn mice with minimal toxicity [37], which suggests CRISPR-Cas9-mediated mutagenesis in animals model has been achieved.

**Genome-scale CRISPR knock-in screening**

Knock-in methods rely on HR between engineered DNA and a targeted locus. This approach enables proteins to be modified at specific loci and to generate fluorescent protein fusions [30,38]. Homologous repair of Cas9-induced DSBs has been demonstrated in multiple organisms [38-40]. For example, CRISPR-Cas9-mediated knock-in of DNA cassettes into the zebrafish genome at predetermined target sites occurred at a very high rate via homology-independent DSB repair pathways [20].

DiCarlo et al. demonstrated high frequencies of oligonucleotide recombination with a transient gRNA CRISPR system in yeast [39]. They showed that co-transformation of a gRNA plasmid and a donor DNA in cells resulted in almost 100% donor DNA recombination frequency. In addition, CRISPR-Cas9 achieved knock in to destabilization domain-tag the essential gene Treacher Collins-Franceschetti syndrome 1 in human 293T cells, leading to rapid modulation of protein levels in mammalian cells [40].

Gratz et al. reported that a 50 nucleotide modification was successfully introduced into the yellow locus by Cas9-induced HR in Drosophila [41] by co-injection of Cas9 mRNA, gRNAs against yellow and single-stranded oligos as donor DNA for recombination. In another study, Xue et al. used the Cas9 transgenic system to generate knock-in mutations in Drosophila by insertion of a large piece of 2 kb heterologous DNA aided by the use of a visible marker [42].

**CRISPR Advanced Gene Therapy**

Xie et al. used CRISPR-Cas genome editing technology to correct disease-causing mutations in cells from beta-thalassemia patients; the correction in human induced pluripotent stem cells restored normal function and provided a rich source of cells for transplantation [43]. Ebina et al. reported that CRISPR aimed at disrupting HIV-1 provirus may be capable of eradicating viral genomes from infected individuals by editing the HIV-1 genome and blocking its expression. They found that the CRISPR-Cas9 system is able to remove internal viral genes from the host cell chromosome [44]. Zhen et al. recently demonstrated that the CRISPR-Cas9 targeting promoter of human papillomavirus oncoproteins (E6 and E7) resulted in accumulation of p53 and p21 proteins and markedly suppressed the proliferation of cervical cancer cells in vitro and in vivo [45].

In another study, a dominant cataract-causing mutation in the Crygc gene in mice was corrected using CRISPR-Cas9, thereby demonstrating the potential of this system for efficient correction of a genetic disease [46]. The CRISPR-Cas9 genome editing system also has the potential to repair the cystic fibrosis transmembrane conductance receptor locus by HR in cultured intestinal stem cells of cystic fibrosis patients [47]. Long et al. [48] recently used the CRISPR-Cas9 gene editing system to correct the dystrophin mutation in developing mdx mice; a model for Duchenne Muscular Dystrophy (DMD) suggests that CRISPR-Cas9 gene therapy might work in animals. The corrected cells may ultimately generate many healthy muscle fibers where this strategy may one day allow correction of disease-causing mutations in the muscle tissue of patients with DMD [48].

**Concluding Remarks and Future Directions**

The simplicity of programming CRISPR-Cas9 has contributed to its rapid implementation in genome engineering. Its use has resulted in rapid generation of genome-scale knock-out libraries for complex model systems, including human cells and animal disease models, and it has potential for use in ex vivo gene therapy in humans. Nevertheless, the CRISPR system may have limitations that have yet to be identified. The delivery and specificity of the gRNA is still not fully understood. Hence, the effectiveness of the CRISPR delivery system must be measured and further validated in order to utilize this system as a safe and reliable tool, especially in the treatment of human diseases.

**Acknowledgement**

N.H.S and M.S thank the Universiti Sains Malaysia (USM) for a Post-Doctoral Fellowship. E.J.M thanks USM for an Academic Staff Training Scheme. N.M.Y acknowledges USM for RUI grant 1001/CIPPT/812095.

**References**


