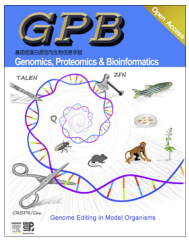
Genomics Proteomics Bioinformatics 13 (2015) 336–344

HOSTED BY

ELSEVIER

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com



REVIEW

Genome Editing and Its Applications in Model Organisms

Dongyuan Ma [a], **Feng Liu** [*,b]

State Key Laboratory of Membrane Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

Received 29 October 2015; revised 3 December 2015; accepted 16 December 2015
Available online 4 January 2016

Handled by Xiangdong Fang

KEYWORDS

Genome editing;
CRISPR/Cas;
Zebrafish;
Disease model;
Gene therapy

Abstract Technological advances are important for innovative biological research. Development of molecular tools for DNA manipulation, such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the clustered regularly-interspaced short palindromic repeat **(CRISPR)/CRISPR**-associated **(Cas)**, has revolutionized **genome editing**. These approaches can be used to develop potential therapeutic strategies to effectively treat heritable diseases. In the last few years, substantial progress has been made in **CRISPR/Cas** technology, including technical improvements and wide application in many model systems. This review describes recent advancements in **genome editing** with a particular focus on **CRISPR/Cas**, covering the underlying principles, technological optimization, and its application in **zebrafish** and other model organisms, **disease modeling**, and **gene therapy** used for personalized medicine.

Introduction

Since the discovery of the DNA double helix in 1953, many basic biological concepts pertaining to the genome, such as gene transcription and translation, genetic code and epigenetic modification, have been established by developing multiple experimental techniques. These include enzymes for *in vitro* DNA manipulations (such as polymerases, restriction endonucleases, and DNA ligases), recombinant DNA technology, *in vitro* DNA synthesis, site-specific mutagenesis, and whole-genome sequencing. Nonetheless, site-specific modification within genomes has remained a major challenge.

Genome editing, namely, refers to editing the nucleotides of the genome with engineered nucleases in cultured cells or living organisms. In the past decade, several types of engineered nucleases have been developed, including zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the recent clustered regularly-interspaced short palindromic repeat (CRISPR) systems. These nucleases, in particular CRISPR systems, immensely facilitate the wide application of genome editing in various biological research fields. More importantly, genome editing holds great promise in potential clinical applications such as gene therapy. In this review, we will briefly describe the features and development of these three editing methods and then mainly focus on the latest CRISPR technology, *Science*'s 2015 Breakthrough of the Year [1], and its application.

* Corresponding author.
E-mail: liuf@ioz.ac.cn (Liu F).
[a] ORCID: 0000-0002-9710-5970.
[b] ORCID: 0000-0003-3228-0943.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

http://dx.doi.org/10.1016/j.gpb.2015.12.001

ZFNs

ZFNs were discovered in 1996 and subsequently employed in genetic engineering experiments with *Drosophila* and mammalian cells in 2002 [2–4]. Unlike the previously reported approaches relying on DNA base-pair recognition, such as oligonucleotides, reverse splicing, or small molecules, the site-directed ZFNs act through DNA/protein recognition [2,4–7]. ZFNs are composed of a zinc finger-mediated DNA binding domain for DNA recognition and a nuclease activity domain of FokI for DNA cleavage [2]. ZFNs can cause double-strand DNA breaks (DSBs). Subsequently, insertion or deletion at the site of the genomic DSB can be induced by imprecise non-homologous end joining (NHEJ)-mediated repair, whereas point mutations or insertions from oligonucleotide or plasmid donor templates can be introduced by precise homology-directed repair (HDR)-mediated repair [8]. Over the past decade, ZFNs were optimized and widely used in, for example, targeted gene knockout in the genomes of mammalian cells to generate genetically distinct $DHFR^{-/-}$ cell lines, establishing OCT4-eGFP human embryonic stem cells (hESCs) or targeting *PITX3* in induced pluripotent stem (iPS) cells, and heritable gene disruption in mouse and zebrafish [9–12].

TALENs

Plant pathogen *Xanthomonas* can secret TALEs upon infection of various host species, which facilitate bacterial infection or trigger defense by binding to promoter regions to activate effector-specific genes or R genes of the host plants [13,14]. TALEs recognize specific DNA sequences via DNA-binding domains composed of nearly identical 34-amino acid repeated units. Two hypervariable amino acid residues at positions 12 and 13, named repeat-variable diresidues (RVDs), are required for target site specificity [15,16]. Therefore, RVDs have been manipulated to generate the programmable DNA-binding proteins and used for site-directed genome editing [15–19]. Similar to ZFNs, the sequence-independent FokI nuclease, found in *Flavobacterium okeanokoites*, functions as the site-specific nuclease for TALEN assays when the target sites are recognized by different TALEs.

Due to the high similarity of TALE recognition sequences, a complicated procedure is required to generate programmable proteins that target specific sites on the genomic DNA, which limits the wide use of TALENs in genome engineering. The presence of extensive identical repeat sequences confers a huge technical challenge to clone repeated TALE arrays for different DNA target sites. To this end, several modified methods have been developed to enable rapid TALE assembly, including the 'Golden Gate' platform [20,21], high-throughput solid-phase based sequential ligation systems [22,23], and ligation-independent cloning techniques [24]. For target site recognition, it's extremely important that the sequence of the TALE binding sites should start with a thymine (T). Moreover, the length of the target site and the spacer between the two TALEN arms are also very important for the formation of the FokI dimer and editing efficiency.

CRISPR/Cas

CRISPR/CRISPR-associated (Cas) systems exist in prokaryotes to mediate bacterial adaptive immune defense against viruses or invading nucleic acids, as the first infection experiments showed that CRISPR/Cas confer resistance against lytic phages of *Streptococcus thermophilus* [25]. Brouns et al. found that mature CRISPR RNAs (crRNAs) work with Cas proteins to provide prokaryotes with antiviral defense by interfering with virus proliferation [26]. In 2012, Jinek et al. showed that the dual-RNA structure formed by crRNA and trans-activating crRNA (tracrRNA) is sufficient to direct *Streptococcus pyogenes* type II Cas9 protein (spCas9) to cleave specific target DNA sequences *in vitro* [27]. *In vitro* DNA cleavage by spCas9 and dual-RNAs reveals the potential of this system for genome editing. Subsequently, the RNA-guided editing tool for mammalian genomes was established using an engineered type II bacterial CRISPR system in 2013 [28,29].

In the CRISPR/Cas system, crRNA–tracrRNA, also referred as the guide RNA (gRNA), recognizes the target sites on the genome, and then recruits Cas9 protein for precise cleavage at specific endogenous genomic loci [28,29]. During this process, synthesis of the gRNA, composed of a specific 20-bp crRNA and the universal tracrRNA, can be driven by a U6 polymerase III promoter *in vivo* or by a phage RNA polymerase, such as T7 RNA polymerase, *in vitro* [29–31]. The first nucleotide of the gRNA target site should be a guanine (G) for U6-directed transcription and two guanines (GG) for T7-directed transcription [28–31]. The most important region for target site selection by the CRISPR/Cas system is the protospacer adjacent motif (PAM) sequence, NGG, which mediates stimulation of the Cas9 nuclease activity [29]. Thus, compared to ZFNs and TALENs, the easy programmability of gRNAs is the most advantageous feature of CRISPR/Cas system (**Table 1**). Therefore, CRISPR/Cas has been quickly applied to generate mutations in different organisms, to establish various disease models, and for the use in gene correction and therapy [32].

CRISPR/Cas in model organisms

The modified type II CRISPR/Cas, including the human codon–optimized versions of Cas9 and the specific gRNA, was first shown to work efficiently in HEK 293T cells, human leukemia K562 cell line, murine cell lines, and PGP1 iPS cells, using the *adeno-associated virus integration site 1* (*AAVS1*) or *empty spiracles homeobox 1* (*EMX1*) loci as target genes in February 2013 [28,29]. Soon after, in March 2013, the synthesized Cas9 mRNA and gRNA targeting *fumarate hydratase* (*fh*) were shown to work *in vivo* to induce targeted genetic modifications in zebrafish as efficiently as ZFN and TALENs [33]. A month later, it was reported that Cas9/gRNA efficiently induced biallelic conversion of *etsrp* and *gata5* in zebrafish somatic cells and resulted in the abnormal intersegment vessels and cardia bifida, respectively, recapitulating the phenotype of $etsrp^{y11}$ and fau^{tm236a} mutants described previously [34]. Later on, CRISPR/Cas-mediated gene editing was used to efficiently disrupt five genes simultaneously in mouse ESCs. Meanwhile, mice with biallelic mutations in *Tet1* and *Tet2*