AIP Conference Proceedings

RESEARCH ARTICLE | FEBRUARY 28 2019

Efforts to improve the efficiency and specificity of CRISPR-Cas9 techniques FREE

Yuanyuan Wang

AIP Conf. Proc. 2079, 020015 (2019)
https://doi.org/10.1063/1.5092393

CrossMark

View Online
Export Citation

Articles You May Be Interested In

Design and *in vitro* test of sgRNA for the CRISPR/Cas9 plasmid construct of the *SQS* gene of *Artemisia annua* L.

AIP Conference Proceedings (January 2022)

Development of CRISPR/Cas9 plasmid for multiple sites genome editing in oil palm (*Elaeis guineensis* Jacq.)

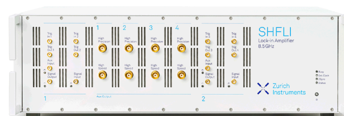
AIP Conference Proceedings (April 2019)

Disruption of electrostatic contacts in the HNH nuclease from a thermophilic Cas9 rewires allosteric motions and enhances high-temperature DNA cleavage

J. Chem. Phys. (December 2022)

04 October 2023 20:48:14

500 kHz or 8.5 GHz?
And all the ranges in between.
Lock-in Amplifiers for your periodic signal measurements



Find out more

AIP Publishing

Efforts to Improve the Efficiency and Specificity of CRISPR-Cas9 Techniques

Yuanyuan Wang[a)]

School of Bioscience, University of Kent, Kent CT2 7BQ, United Kingdom

[a)]Corresponding author: yuanyuan__wang@126.com

Abstract. The CRISPR-Cas9 technique derived from bacterial immune system has been used in the majority of genome engineering applications due to its advantages of low cost and flexibility. In order to achieve efficient modifications to target genome in different cell types by CRISPR-Cas9 genome editing, three basic components are required, including 20nt sequence of sgRNA that could recognize desired DNA sequences, Cas9 protein for introducing double-strand breaks and the NGG PAM sequence for the binding between Cas9 proteins and the target genomic sites. However, the CRISPR-Cas9 technique today are facing some difficulties in adoption of further genome applications, especially those requiring high precision modifications. This is mainly because of the off-target effects, in which the spCas9 nuclease cleaves unwanted genomic sequences. To address this, a number of strategies have been developed. In this article, we mainly reviewed the factors in determining the activity of Cas9 proteins and concluded the approaches to reduce the off-target effects.

04 October 2023 20:48:14

INTRODUCTION

In the past, the genetic recombination based on the Mendelian law was the predominant method for the genetic research and it was difficult to conduct precise genetic modifications. Until recent years, the emerging of targeted genome engineering with artificial DNA endonucleases provided opportunities for scientists to manipulate genes at any site in different cell types or organisms, which has greatly revolutionized biology, biotechnology and medicine. Today three genome editing techniques are frequently used, which are zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR-Cas9 system.

ZFNs and TALENs are widely used in academia and industrial research due to high efficiency of DNA cleavage. They consist of two basic components of DNA recognition proteins and DNA cleavage FOK**I** endonuclease. The DNA recognition protein can direct the cleavage protein of a desired DNA region, and based on the protein-DNA interaction, the double-stranded breaks (DSBs) at targeted sites are generated for gene deletion, replacement or addition. [1-2]. The generated DSBs can activate the cellular repair pathways, which are error-prone non-homologous end joining (NHEJ) and high-fidelity homologous recombination (HR). NHEJ is the predominant DNA damage repairing pathway occurring in the absence of a repair template, which results in the gene insertions or deletions at the targeted sites. The HR-mediated DNA damage repair pathway is only observed when homologous sequences are present, which facilitates highly targeted gene inactivation, replacement or addition [3-4].

The Site-specific genome editing techniques needs to redesign and re-program a new set of nucleases. The difficulties in engineering the proteins involved in the ZFNs and TALENs limit them for wider applications. Therefore, CRISPR-Cas9 technique becomes an alternative and predominant method to manipulate DNA sequences, which relies on a simple base-pair complementarity between the guide RNA (sgRNA) and target genomic regions to introduce DSBs [5-6] (Figure 1). Specifically, the CRISPR-Cas9 was derived from type II microbial adaptive immune system, where bacteria protect themselves by destroying genomic information of invading phages and plasmids [7]. The CRISPR represents a family of DNA repeats, whose loci consists of repetitive sequences separated by nonrepetitive DNA sequences named spacers. Studies revealed that the spacers are short stretches of DNA sequences from the phage genome. By transcription and further processing, they serve as sgRNA to direct Cas9 proteins to the foreign genetic

2018 International Conference on Biotechnology and Bioengineering (8th ICBB)
AIP Conf. Proc. 2079, 020015-1–020015-5; https://doi.org/10.1063/1.5092393
Published by AIP Publishing. 978-0-7354-1807-3/$30.00

sequences following a 5'-NGG protospacer-adjacent motif (PAM) [8-9]. Thereafter, the Cas9 nucleases cleave the DNA strands that are complementary to 20nt sequence within the sgRNA.

In the CRISPR-Cas9 technique, endonuclease Cas9 is able to cleave any DNA sequences by redesigning the sgRNA sequences. It becomes even more promising and powerful when aiming to introduce an exogenous sequences of choice (donor DNA) to replace the target bases by activating the HR pathway of the DNA DSBs. To achieve this, four steps are generally carried out. First, design of sgRNA to target the gene(s) of interest and generation of expression vectors for Cas9, sgRNA and donor DNA. Several online tools now are available to design the sgRNA, such as CRISPR-ERA which currently supports for the design of 9 organisms [10]. DNA oligos should be synthesized to construct the sgRNA after the selection of sgRNA sequences (step2). Thereafter, the constructed CRISPR-Cas9 complex are introduced into the cells to insert, delete or replace genes. Finally, the introduced mutations are examined by polymerase chain reaction (PCR) genotyping and confirmed by sequencing [11].

Owing to its advantages of high efficiency, simplicity of sgRNA design to target desired sequences and low-cost, the CRISPR-Cas9 technique brings new era for genome regulation and organization in living cells across different organisms. However, some issues prevent the CRISPR-Cas9 from more applications, particularly the off-target effects and low specificity. Therefore, it is necessary to find some strategies to enhance the specificity of the CRISPR-Cas9 genome editing. In this review article, we will focus on the efforts of scientific community to improve the activity of Cas9 proteins, and briefly discuss the applications and future directions of the CRISPR-Cas9 technique.

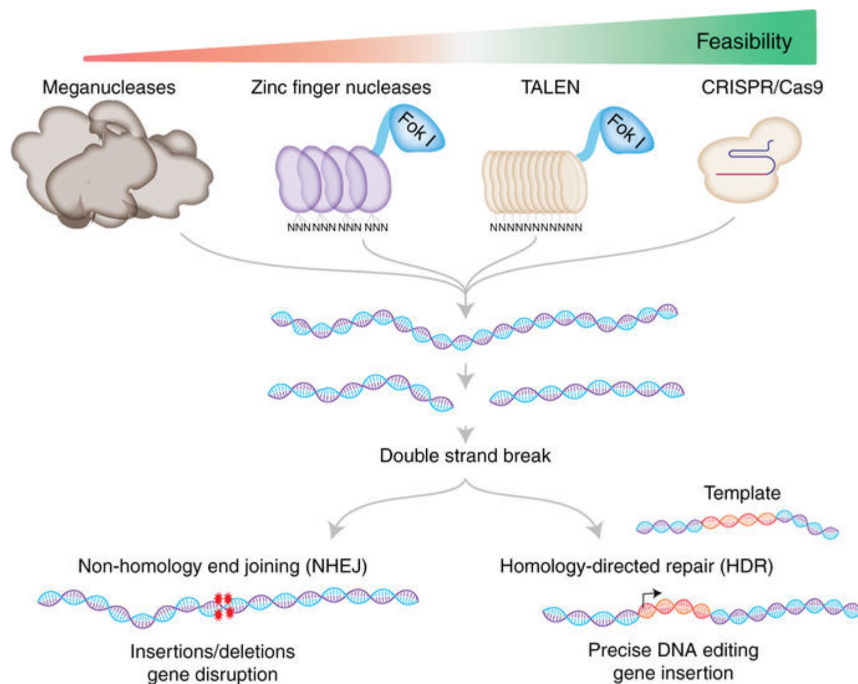


FIGURE 1. The working principle of the predominant genomic engineering techniques [12]. The working principle of major techniques for genetic engineering. The Meganucleases are engineered restriction enzymes that could recognize long stretches of DNA sequences. The zinc finger nuclease and TALENs consist of two basic components of DNA recognition proteins and DNA cleavage FOK**I** endonuclease. CRISPR-Cas9 technique relies on a simple base-pair complementarity between the sgRNA and target genomic regions. All these tools are able to introduce double-strand breaks that can be repaired either by error- prone non-homology end joining or homology-directed repair. Non-homology end joining results in random gene insertions or deletions at the targeted sites, homology-directed repair facilitates highly targeted gene inactivation, replacement or addition.

OPTIMIZATION OF THE SGRNA TO ENHANCE ACTIVITY OF THE CAS9 PROTEINS

The off-target effects, in which the Cas9 proteins will cleave genomic sequences that are not complementary with sgRNA, are one of the major issues that impede potential applications of the CRISPR. Cutting efficiency and DNA targeting specificity of the Cas9 proteins can be achieved by the artificial sgRNA formed by a fusion of 20nt

04 October 2023 20:48:14

nucleotides at 5’-terminal for DNA recognition and a set of nucleotides at 3’-terminal for Cas9 binding. On this basis, it seems that optimization of the activity and specificity of the sgRNA can subsequently eliminate the off-target effects of the Cas9 proteins.

In the past years, impacts of the sgRNA sequence features on Cas9 activity have been widely characterized by using various methods, such as Genome-wide Unbiased Identification of DSBs Enabled by Sequencing (GUIDE-seq), next-generation sequencing (BLESS) and Digernome-seq. It was revealed that the 3'- extension of sgRNA could facilitate highly efficient Cas9 cleavage in vitro [13], and perfect base-pairing within 10-12 bp of the sgRNA at 5’-end could directly determine associated Cas9 activity [14]. These findings illustrated that programming the 3’-end of sgRNA sequence could enhance the Cas9 activity, and perfect matches between the 5’-terminal of sgRNA sequence and its target regions are required when designing the sgRNA constructs. In addition to this, scientists identified the impacts of mismatches between the sgRNA with target sequences on Cas9 specificity by examining Cas9-mediated DNA cleavage with single mismatches and more than one mismatches within different sgRNA sites. One of the major findings was that single-base mismatch at the last 10-12 nucleotides of the 20nt sgRNA targeting region in 3’-ternimal could completely abolish the Cas9-mediated DNA cleavage. However, the mismatches at the 5’-terminal of the sgRNA could be better tolerated than those at 3’-end. [15]. This further proved that the sgRNA is an important factor for Cas9 activity and provided some parameters to consider for the generation of sgRNA design constructs.

There has been several approaches to try and reduce the off-target effects of the Cas9 nucleases. Increasing the GC-content was one of them, which could stabilize the hybrids of RNA and DNA and thereby facilitating higher efficiencies of the Cas9-midiated genome editing [16].

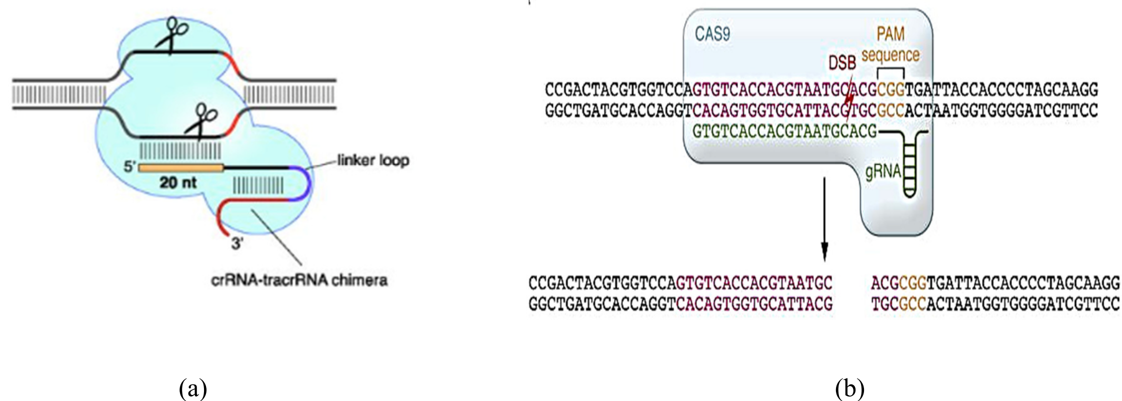


FIGURE2. The schematics of CRISPR-Cas9 techniques. (a) The Cas9 protein guided by a short chimeric sgRNA. [8, 13]. The chimeric sgRNA is formed by a fusion of 20nt nucleotides at 5’-terminal for DNA recognition and a set of nucleotides at 3’-terminal for Cas9 binding. (b) The DSBs introduced by Cas9 protein [17]. The sgRNA can direct Cas9 proteins to the foreign genetic sequences following a 5’-NGG protospacer-adjacent motif (PAM). The Cutting efficiency and DNA targeting specificity of the Cas9 proteins can be achieved by sgRNA.

ORTHOGONAL CAS9 PROTEINS FOR GENOME EDITING

The most commonly used Cas9 protein is derived from *Streptococcus pyogenes* (spCas9), which is able to mediate single activity at a variety of genomic locus within any given cells. However, the spCas9 proteins have a number of limitations. First, the mismatches between the sgRNA sequences and target sequences can lead to off-target effects of spCas9 proteins. To address this, a double-nicking (DN) approach recently has been developed by combining two sgRNAs and a pair of SpCas9 nickase D10A mutants (Cas9n) to introduce DNA cleavage at desired sites. Compared with the wild-type Cas9, the specificity of CRISPR-Cas9 has been significantly improved by 50x to 1,000× [5-6].

Furthermore, the range of sequences that can be identified by the spCas9 nucleases is limited by a need for a PAM with sequence NGG. Even if the N could be any nucleotide, and the PAM sequence can broadly be found in the genomes of different organisms, the NGG PAM still cannot cover the full genomes. And importantly, the spCas9 proteins cannot mediate different activities at different targets simultaneously. Therefore, new Cas9 variants with alternative PAM specificity are required to replace the spCas9 proteins for the DNA cleavage in more organisms. Kleinstiver, B. P. *et al.* identified and characterized two Cas9 variants from Streptococcus thermophilus (St1Cas9) and Staphylococcus aureus (SaCas9) [18]. They reported that these two Cas9 variants facilitated efficient cas9-

04 October 2023 20:48:14

mediated modifications to the genome in bacterial systems and human cells. However, whether these two Cas9 variants can introduce DSBs in more organisms and keep high specificity is still unknown.

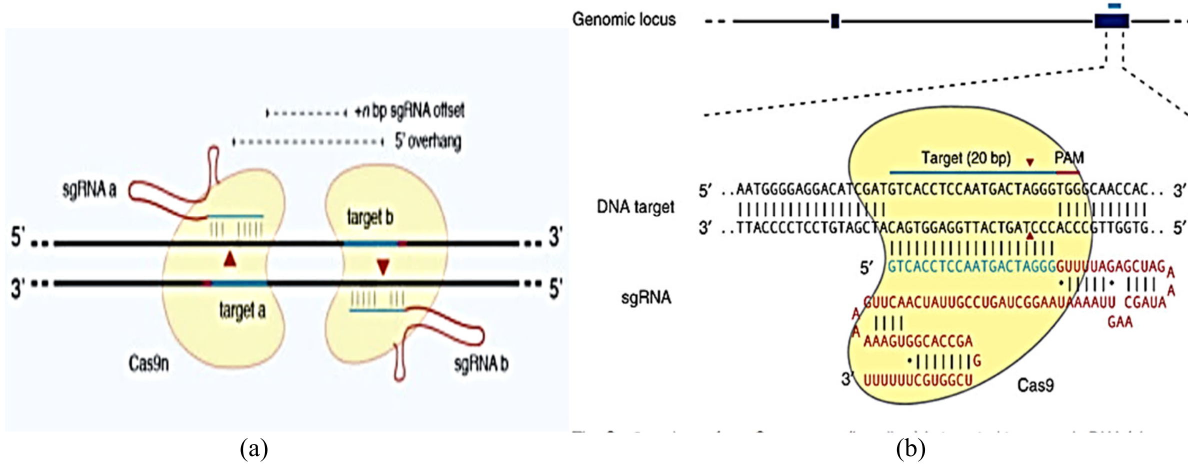


FIGURE 3. The Orthogonal Cas9 proteins for DNA cleavage. (a) Schematic of spCas9 protein binding to target genomic locus next to NGG PAM [5]. Sequences that can be identified by the spCas9 nucleases is limited by a need for a PAM with sequence NGG, the N could be any nucleotide. (b) Schematic of double-nicking approach [6]. The double-nicking approach introduces DNA cleavage at desired sites by combining two sgRNAs and a pair of SpCas9 nickase D10A mutants (Cas9n).

PAM SEQUENCE AND THE OPTIMAL INJECTION CONCENTRATIONS OF SGRNA AND CAS9 PROTEINS

The PAM sequence and sgRNA concentration have been reported to potentially influence the efficiency and specificity of Cas9 proteins, which brings more concerns on the design of CRISPR-Cas9 constructs.

The NGG PAM adjacent to target genomic locus is required for spCas9-mediated DNA cleavage. However, studies showed that the NAG PAM could also serve as an alternative non-canonical PAM for the DNA cleavage at *EMX* locus in human cells. To test whether the NAG PAM is universal for other sequences in human cells and investigate the DNA cleavage efficiency of reported PAMs including NAG, NGG and NGA, scientists designed three sgRNAs to target GFP genes with NGA, NGG, and NAG PAMs. They found the NAG PAM was not a universal non-canonical PAM for Cas9 activity in human cells, and the NGA PAM resulted in efficient Cas9-mideated DNA cleavage [19].

Studies have revealed that the sgRNA concentration also serves as an important factor in determining the specificity of Cas9 proteins. It has been reported that reducing the concentrations of sgRNA could reduce off-target effects [15]. More specifically, Ren, X. *et al*. have introduced sgRNA vectors into the fly embryos to target *white1* and *white2* genes in order to identify the relationships between sgRNA concertation and the mutagenesis. They found the ideal injection concentration of the sgRNA is around 75 ng/μl to facilitate efficient mutagenesis to fly (2014).

Whether the enzymic concentration can affect the activity of Cas9 proteins is still arguable. Some studies revealed strong impacts whereas some did not observe any relationships. This is because that the Cas9 activity is complex and it differs from the different genomic sites or cell types.

CONCLUSIONS

Efforts to improve the specificity of Cas9 proteins and reduce the off-target effects are still in progress. In this review article, we focused on the analyzation of major factors that can affect the specificity of the CRISPR-Cas9 techniques, including the 20nt sequence of sgRNA, Cas9 proteins, the PAM sequences and the injection concentration of sgRNA. Although there are some obstacles, especially the limitations of the spCas9 proteins which might lead to immunogenesis when being adopted into human-beings, the CRISPR-Cas9 technique will undoubtedly continue to serve as a major technique for a variety of research, ranging from cell studies to medicine, particularly the investigation of gene functions. For the future directions, the CRISPR-Cas9 technique serves as a tool for the treatment of genetic

04 October 2023 20:48:14

diseases; it is likely that the solutions to improve the specificity and reduce the off-target effects need to be thoroughly evaluated.

REFERENCES

1. Y. G. Kim, J. Cha, S. Chandrasegaran, Proc Natl Acad Sci USA. **93**, 1156-1160 (1996).
2. C. Mussolino, T. Cathomen, Curr Opin Chem. Biol, **23**, 644-650 (2012).
3. M. R. Capecchi, Sci, **244**, 1288-1292 (1989).
4. J. S. Kim, T. B. Krasieva, H. Kurumizaka, D. J. Chen, A. M. R. Taylor and K. Yokomori, J Cell Physiol, **170**, 341-347 (2005).
5. F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott and F. Zhang, Trends in Genetics Tig, **8**, 2281 (2013).
6. F. A. Ran, P. D. Hsu, C. Y. Lin, J. S. Gootenberg, S. Konermann, A. E. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang, F. Zhang, Cell, **154**, 1380-1389 (2013).
7. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, Nature, **482**, 331-338 (2012).
8. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, Sci, **337**, 816–821 (2012).
9. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, Nature, **482**, 331-338 (2012).
10. H. Liu, X. Wang, L. Qi, Bio-protocol, **7**, (2017)
11. K. Xie, B. Minkenberg, Y. Yang Y. Bio-protocol, **4**, (2014)
12. M. Adli. Nat Commun, **9**, (2018).
13. M. Jinek, A. East, A. Cheng, S. Lin, E. Ma, and J. Doudna, Elife, **2**, e00471 (2013)
14. P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem , T. J. Cradick, L. A. Marraffini, G. Bao, F. Zhang, Nat Biotechnol, **31**, 827–832 (2013)
15. Y. F. Fu, J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung and J. D. Sander. Nat Biotechnol, **31**, 822-826 (2013)
16. N. Sugimoto, S. Nakano, M. Katoh, A. Matsumura, H. Nakamuta, T. Ohmichi, M. Yoneyama, M. Sasaki, Biochemistry **34**, 11211-11216 (1995).
17. R. M. Gupta and K. Musunuru, Clin Invest, **124**, 4154-61 (2014).
18. B. P. Kleinstiver, M. S. Prew, S. Q. Tsai, V. V. Topkar, N. T. Nguyen, Z. L. Zheng, A. P. W. Gonzales, Z. Y. Li, R. T. Peterson, J. R. J. Yeh, M. J. Aryee and J. K. Joung, Nature, **523**, 481 (2015).
19. Y. L. Zhang, X. L. Ge, F. Y. Yang, L. P. Zhang, J. Y. Zheng, X. F. Tan, Z. B. Jin, J. Qu and F. Gu, Sci Rep, **4**, 5405 (2014).

04 October 2023 20:48:14