Strategies for Using CRISPR-Cas9 Orthologs to Perform Gene Editing Applications

Zhikang Suo

Chongqing Nankai Secondary School, Chongqing 401122, China

Abstract. CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats -Cas9) is primarily the adaptive immune system developed by bacteria and archaea to defend themselves from an invasion of the virus. CRISPR-Cas9 is considered currently the most potential gene editing technology in 21 century due to its highest operationality and modifiability among the gene-editing tech that humans have ever found. Orthologous proteins are the proteins of different species which originally develop from one ancestor and still hold the same function as the original protein. This research mainly concludes and organizes the strategies for two specific types of Cas9 orthologs in the application of gene editing.

Keywords: CRISPR-Cas9 Orthologs; dCas9; Gene Editing Applications; Thermophiles; Transcriptional Regulation; Epigenetic Modification; Genome Imaging.

1. Introduction

1.1 Backgrounds of CRISPR-Cas9

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) was first discovered by Yoshizumi Ishino among noncoding DNA areas in the genome of E.Coli and demonstrates an unusual pattern of one piece of repeat gene following a piece of spacer gene repetitively. But CRISPR wasn't connected with the adaptive immune system of bacteria until 2005 when it was demonstrated by three isolated experimenting groups that some spacer sequences among it comes from DNA of the virus bacteriophage which has attacked the bacteria. In 2007, CRISPR was first proved by experimental evidence to be involved in the adaptive immune system of bacteria and later certified to play a key role in the recognition and breaking of invaded virus DNA sequences through Cas protein complexity. Until nowadays, generally, CRISPR systems could be mainly divided into two Classes, Class I and Class II. Class I contains type I, type III, and type IV CRISPR system, which utilize multiple effector proteins' complexity to recognize and cut the target gene.[1] Class II includes type II, type V, and type VI CRISPR systems, which commonly achieve recognition and cleavage of the target gene through one single large effector protein. [1] Among them, type I, II, and V CRISPR systems especially edit DNA, type VI solely identifies and cleaves RNA, and type III could edit both DNA and RNA, but the detailed effecting mechanism of the type IV system and whether it effect on DNA or RNA haven't been discovered yet. [2] Owing to the high complexity of Class I CRISPR systems' effecting process and limitation of target organism of the type IV system, type V systems and type II systems, which hold the comparatively simpler structures of affecting mechanism and process, become currently the most widely researched and applied CRISPR systems. And CRISPR-Cas9 just belongs to the type II system.

1.2 Effecting Process of Original CRISPR-Cas9 and Limitation of it in Application

Type II systems' original effecting process could be divided into three conjoint steps: acquisition of spacer sequence, formation of integrated Cas9-gRNA complexity, and interference of target gene. Firstly, when a new virus-like bacteriophage invades and injects its DNA into the bacteria, the dimer formed by Cas1 and Cas2 would primarily cleave one piece of sequence, protospacer, on the DNA of the invaded virus through recognition of specific PAM (Protospacer Adjacent Motif) that is next to the protospacer. And in the type II system, PAM is normally NGG, which means one random nucleotide and two Guanine orderly from 5' to 3'. After cleaving, the dimer is responsible for integrating the protospacer into the CRISPR on its own genome, and the protospacer is commonly

inserted between the leader sequence and the first repeat in the CRISPR. Secondly, when the same species of virus invade this bacteria again, the whole CRISPR sequence is firstly transcribed into Pre-crRNA (Pre-CRISPR RNA) and the tracrRNA(trans-activating CRISPR RNA), which could be paired with the repeats, would also be transcribed. After the tracrRNA combined with a transcribed repeat on the Pre-crRNA, Cas9 and RNase III (Ribonuclease III) together would shear the Pre-crRNA and then create pieces of mature tracrRNA-crRNA dsRNA, called gRNA (double-stranded RNA). Eventually, since Cas9 could be divided into two lobes, one recognition lobe that is composed of REC I, REC II, bridge helix, PI (PAM-interacting) domains, and a nuclease lobe composed of RuvC, and HNH domains, when the mature gRNA combined with REC I, Cas9-gRNA complexity would begin to search target sequence on genes of the invasive virus and then unwind the target gene once the PI domain contacting the matched PAM sequence of NGG through recognizing the specific groove. Then once over 10~12 downstream nucleotides next to the PAM could be paired with the transcribed spacer in the gRNA, the bridge helix will be activated to initiate the cleavage of the target strand and non-target strand through HNH and RuvC respectively. Otherwise, if the number of matched nucleotides is under 10, the complexity would quickly separate from this piece of the gene until the PI domain recognize the next paired PAM system. After the double strand of the target gene is broken, the resulting two pieces of breaking DNA would commonly be restored and combined again through NHEJ (Non-homologous end joining), which hold a high possibility of resulting in insertion and deletion on the gene. Therefore, the target gene is likely to be inactivated or be translated into dysfunction protein/proteins, which eventually disrupt the translation and package of the virus inside the bacteria.

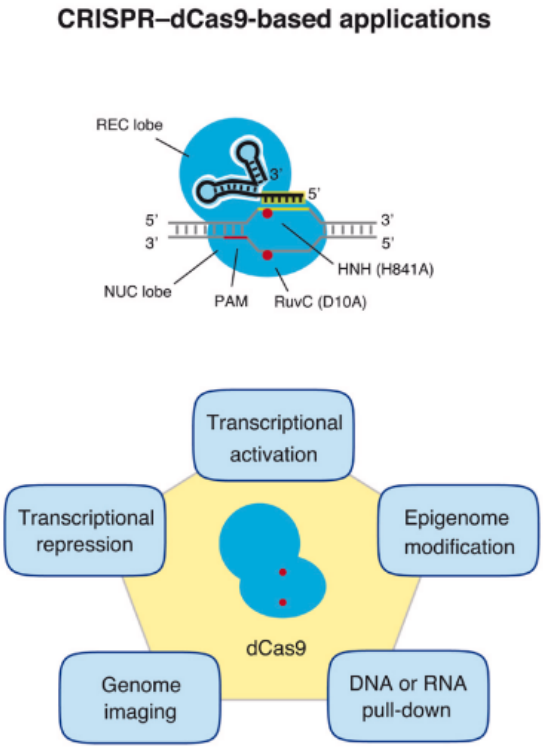


Fig 1. Summary of CRISPR–dCas9-based genome engineering applications, citated from [9]

Since the CRISPR-Cas9 would initiate the cutting of the combined gene after 10~12 nucleotides paired with the spacer gene and 20 nucleotides are supposed to be the common length of spacer sequence in CRISPR-Cas9, the off-target effect of the original CRISPR-Cas9 system is happened in a high frequency, especially when it's used for recognizing target gene on mammals or plants, which are carried with a genome made up of billions of nucleotides. Although the defending mechanism of

this system are supposed to work perfectly in the bacteria and archaea since it could achieve the goals of interrupting virus reproduction inside the bacteria or archaea even if the CRISPR-Cas9 complexity gets off-target from the target gene, the off-target effect would probably cause huge distortion on the result or process of the experiment if the system is directly used in gene editing. Owing to that reason, the necessity of improving the specificity of CRIPR-Cas9 when it's utilized as a gene editing item is evident. But due to the low stability of designing more specific CRISPR-Cas9 through directly operating random mutation on Cas9 genes and the restricting property of present most commonly applied Spy Cas9 whose only combined gRNA could be designed, the high diversity of natural CRISPR-Cas9 orthologs in species of bacteria and archaea, selecting specific orthologs of CRISPR-Cas9 among the diverse natural CRISPR-Cas9 orthologs and matching each of them with a specific type of gene editing application tends to be a more feasible way for improving the specificity of CRISPR-Cas9. Therefore, this paper would discuss and propose some strategies for utilizing two particular types of Cas9 orthologs, particularly the dCas9 of them, in different areas of gene editing.

2. Contents

2.1 Strategies of Thermophilic Cas9 orthologs Applications in Thermophiles

According to the research achievement of Giedrius Gasiunas and Joshua K. Young et al, five orthologs of Cas9, involving Cme2, Cme4, Ghy1, Esp1, and Nsa Cas9, were demonstrated to be thermostable by holding the melting temperatures higher than 50 °C. Their further research has shown that the Cme2, Ain, Cme3and Sth1A Cas9 can remain their cleavage activity when they are tested combined with the Nsa Cas9, detached from *Nitratifractor salsuginis,* one type of deep-sea thermophilic bacteria. Intriguingly, Nsa Cas9 solely could even keep being active at temperatures higher than 60 °C, which means that the CRISPR-Cas 9 system involved Nsa Cas9 could be applied in gene editing of the thermophilic bacterium. [3] Nsa Cas9 and other Cas9 proteins with the properties of tolerating high temperatures are supposed to have at least 2 specific gene-editing applications in thermophiles.

First and foremost, those high-temperature tolerant Cas9 orthologs could be used to elevate the efficiency of industrial enzyme production in thermophiles. Since, to survive in extreme high-temperature environments like thermal springs near volcanoes or abyssal hydrothermal vent chimneys, the enzyme produced by thermophiles, particularly Hyperthermophiles, normally have high thermal-stability and catalytic efficiency, those thermophile-produced enzymes like cellulase, protease, and insulinase are extensively used in industrial production and purification. Since inactive dead Cas9 (dCas9), which has inactivated nuclease lobes but is able to recognize and incorporate with target gene, has been utilized to achieve regulation of gene expression through protein fusion, Nsa Cas9 could also be artificially devised to Nsa dCas9 through introducing silence mutations in genes which encode HNH and RuvC domains (such as the combination of H840A and D10A mutations) and then realize the enhancement of thermophile enzymes production in a couple of ways. [4] To be more specific, this type of high temperature tolerating dCas9 could be fused with specific demethylase, like ten-eleven translocation protein (TET) (involved TET1, TET2, TET3), to enhance the expression of the target gene which could be translated to the enzyme for industrial application through mediating the collection of demethylase's structure domain in enhancer sequence or promoter sequence. Similarly, this dCas9 could also be fused with an activator or promoting transcription factor directly to improve the yielding efficiency of the enzymes encoded by the target gene in the thermophiles. This dCas9 can achieve the acetylation of histone around those target gene to promote expression of the target genes through the fusion with the particular acetyltransferase in thermophiles. What's more important, owing to the fact that the histone of some extreme thermophile would automatically form tetramer even octamer, which could protect against exposing DNA but interrupt the expression of genes in it at the same time, the enhancing mechanism of fusion acetyltransferase and dCas9 could be extremely useful. Admittedly, despite enhancement, this mechanism may also hurt the DNA inside the histone. But since most of those applied thermophiles hold the DNA with a high proportion of G-

C base pairs (some even over 60%), they are not supposed to be undermined or unwound under the temperature which dCas9 formed by Nsa Cas9 be active in. Additionally, according to a new study, thermophilic Cas9 orthologs can also achieve the goal of improving specific thermophilic enzyme production efficiency by chromatin topology. Specifically, a group of researchers realize the formation of a chromatin loop between the enhancer and promoter sequence of the gene in K562 hematopoietic cells and HEK293T cells by connecting two dimerize protein domains (ABI1 and PYL1) with two independent dCas9 orthologs and demonstrates that this chromatin loop significantly increases the expression of one specific locus in K562 but not another.[4,5] For another independent research, researchers demonstrate that DNA looping could also be formed in prokaryotic cells through strong heterodimerizing leucine zippers, which means that no far this method could be applied to prokaryotic cells like thermophilic bacterium and then elevate the enzymes production efficiency through it.[4,6]

2.2 Strategies for Applications of Long PAM Sequence Cas9 Orthologs

In the light of the previous research (done by Giedrius Gasiunas and Joshua K. Young et al) of determining each par PAM sequence recognized by Cas9 orthologs in 10 selected representative clades (from Clade I to X), amounts of orthologs that hold a PAM sequence over the 7 positions were identified and most of them are concentrated in Clade VI.[3] Because the PI domain interacts and recognizes with the corresponding PAM by forming specific hydrogen bonds with the base pairs in the PAM and interacting with a major or minor groove near the PAM. For instance, in Spy Cas9, the PI domain just recognizes the NGG PAM sequence on the nontarget strand by forming 2 specific hydrogen bonds between Arg1335 and one guanine (the G at 3rd position) and 2 similar bonds between Arg1333 and another guanine (the G at 2nd position). Therefore, as the number of base pairs PAM sequences increases, the stability and specificity of the combination between Cas9 and DNA are supposed to be improved. For that reason, the dCas9 orthologs with the large number of base pairs (over or equal to 4) in the corresponding PAM are more fitted with a application of CRISPRi(CRISPR interference) technology, accomplishing interfering expression of the target gene through the property of the dCas9 that it tends to stay on the strand without leaving after combination, and live cell chromatin imaging, which works through a fusion of dCas9 and fluorescent signal protein.[4,7,8] For example, since CRISPRi has been successfully applied in selecting and testing the effecting of a new medicine on specific repetitive gene group in cancer cells, the dCas9 designed from those with a high base number in PAM could further improve the specificity and precision of this application through decreasing the off-target effect led by inadequate recognition of PAM sequence. What's more, this type of dCas9 could also be utilized for lengthening the present time of fluorescent signal in the target chromatin locus in living cells through the higher strength between the dCas9 and target gene, since the higher number of bonds between this type of dCas9 orthologs and its corresponding PAM sequence compared with other types could bring a stronger force of connection between these two.

3. Conclusion

This research mainly concludes and proposes the strategies for two specific types of Cas9 orthologs in the application of gene editing. The Cas9 which can keep being active in high temperatures (for the organism) could be designed and then applied in improving thermophilic enzyme production efficiency through fusion protein operating epigenetic editing. Another type of Cas9 which have a comparatively longer recognizing PAM sequence than presently designed Spy Cas9 could be designed and later applied in testing the effect of a new medicine on target genes and tracing of specific chromatin locus in living cell through respectively CRISPRi and living cells chromatin imaging.

Despite the two types of Cas9 discussed in the research, there are still amounts of potential Cas9 proteins, which could probably be designed into great gene editing items, like those with the long sequence of spacer sequences in gRNA. Therefore, through this research, it's expected that people

could apply those theoretical organized results of Cas9 orthologs to the real application and create some more powerful and specific gene editing items to bring larger benefits for whole human life in the future.

Acknowledgments

Thanks for the great contribution of Prof. Nikolaos G. Sgourakis in this research paper!

References

[1] Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science. 2016 Aug 5;353(6299):aad5147. doi: 10.1126/science.aad5147. PMID: 27493190.

[2] Liu Z, Dong H, Cui Y, Cong L, Zhang D. Application of different types of CRISPR/Cas-based systems in bacteria. Microb Cell Fact. 2020 Sep 3;19(1):172. doi: 10.1186/s12934-020-01431-z. PMID: 32883277; PMCID: PMC7470686.

[3] Gasiunas G, Young JK, Karvelis T, Kazlauskas D, Urbaitis T, Jasnauskaite M, Grusyte MM, Paulraj S, Wang PH, Hou Z, Dooley SK, Cigan M, Alarcon C, Chilcoat ND, Bigelyte G, Curcuru JL, Mabuchi M, Sun Z, Fuchs RT, Schildkraut E, Weigele PR, Jack WE, Robb GB, Venclovas Č, Siksnys V. A catalogue of biochemically diverse CRISPR-Cas9 orthologs. Nat Commun. 2020 Nov 2;11(1):5512. doi: 10.1038/s41467-020-19344-1. PMID: 33139742; PMCID: PMC7606464.

[4] Adli M. The CRISPR tool kit for genome editing and beyond. Nat Commun. 2018 May 15;9(1):1911. doi: 10.1038/s41467-018-04252-2. PMID: 29765029; PMCID: PMC5953931.

[5] Morgan SL, Mariano NC, Bermudez A, Arruda NL, Wu F, Luo Y, Shankar G, Jia L, Chen H, Hu JF, Hoffman AR, Huang CC, Pitteri SJ, Wang KC. Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. Nat Commun. 2017 Jul 13;8:15993. doi: 10.1038/ncomms15993. PMID: 28703221; PMCID: PMC5511349.

[6] Hao N, Shearwin KE, Dodd IB. Programmable DNA looping using engineered bivalent dCas9 complexes. Nat Commun. 2017 Nov 20;8(1):1628. doi: 10.1038/s41467-017-01873-x. PMID: 29158476; PMCID: PMC5696343.

[7] Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013 Feb 28;152(5):1173-83. doi: 10.1016/j.cell.2013.02.022. Erratum in: Cell. 2021 Feb 4;184(3):844. PMID: 23452860; PMCID: PMC3664290.

[8] Qin P, Parlak M, Kuscu C, Bandaria J, Mir M, Szlachta K, Singh R, Darzacq X, Yildiz A, Adli M. Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. Nat Commun. 2017 Mar 14;8:14725. doi: 10.1038/ncomms14725. PMID: 28290446; PMCID: PMC5424063.

[9] Lo A, Qi L. Genetic and epigenetic control of gene expression by CRISPR-Cas systems. F1000Res. 2017 May 25;6:F1000 Faculty Rev-747. doi: 10.12688/f1000research.11113.1. PMID: 28649363; PMCID: PMC5464239.