



Research article

In vitro analysis of site specific nuclease selectivity by NGS

Vincent Brondani*

GenAccurate, Colmar, France

* **Correspondence:** Email: genaccurate@gmail.com.

Abstract: Nucleases currently used in genome engineering induce hydrolysis of DNA phosphate backbone in a sequence-specific manner. The RNA guided nucleases describe today are recognizing a sequence with two distinct molecular interactions: first, like a restriction endonuclease, by direct interaction between the protein and the DNA; and second, by hybridization of the guide RNA with the target DNA sequence. Here we report an in vitro assay to assess the cleavage specificity and the selectivity of the nucleases. The assay is designed using a plasmid encompassing the DNA target site degenerated at positions determined on structural feature. The results demonstrate that the Cpf1 RNA guided nuclease is highly specific for the target sequence, nevertheless its substrate selectivity is low compare to a restriction endonuclease.

Keywords: Nucleases; in vitro cleavage; specificity; selectivity; NGS

Abbreviations: BamHI: *Bacillus amyloliquefaciens* H I; Cas: CRISPR-associated protein; ClaI: *Caryophanon latum* I; Cpf1: CRISPR from *Prevotella* and *Francisella* 1; CRISPR: Clustered regularly interspaced short palindromic repeat; crRNA: CRISPR RNA; DSB: Double strand Break; DMEM: Dulbecco's modified Eagle Medium; DMSO: Dimethyl sulfoxide; EcoRI: *E. coli* RY13 I; FokI: *Flavobacterium okeanokoites*; MBP: Maltose binding protein; NGS: Next generation sequencing; PAM: Protospacer adjacent motif; SmaI: *Serratia marcescens* I; TALEN: transcription activator-like effector nucleases; ZFN: Zinc finger nuclease.

1. Introduction

The first bacterial nucleases discovered and commonly used in laboratory for molecular biology applications are the restriction endonucleases [1]. Those enzymes recognize short sequences, most commonly palindromic, and catalyze the cleavage of the DNA at site. Because of the length of the target sequences, they are not adapted to perform the genome engineering in eukaryotic cells. In recent

years, the zinc finger nucleases (ZFNs) [2], the transcription activator-like effector nucleases (TALENs) [3], and more recently, the bacterial adaptive CRISPR-Cas systems [4–6] were established for the genome engineering of various cell types. The ZFNs and TALENs nucleases are DNA binding domains that are fused to the DNA cleavage domain from the *FokI* restriction endonuclease, whereas the CRISPR-Cas system is using a guide RNA that directs the nuclease to the target DNA sequence. The CRISPR-Cas ribonucleoprotein (RNP) complex catalyzes the DNA cleavage at target sites. Two of the most common nucleases for CRISPR-Cas systems are SpCas9 [7] and Cpf1 proteins [8] and are different in their structural organization and DNA cleavage properties. The analysis is focused on the Cpf1 nucleases that are considered as the most specific and selective for the target sequence [9]. Cpf1 requires only a crRNA to bring the specificity for a target DNA (23 base pairs in length) to the RNP complex. In addition to the DNA sequence recognized by the guide RNA, the Cpf1 protein interacts with a short DNA region (5'-TTTv-3') adjacent to the target DNA sequence, termed Protospacer Adjacent Motif (PAM). The PAM sequence is located upstream to the target DNA sequence and the DNA cleavage induces a 5' overhangs ends double strand break (DSB) that is distal to the PAM sequence. Compared to the previously techniques already described [10,11] with complex DNA substrate in a single reaction, the assay is designed to analyze, and compare nuclease specificity and selectivity of nucleases, using a plasmid library. The library is designed with a short DNA sequence that is degenerated at a selected region according to the structural feature of the nuclease complex with the DNA. The assay described here is interesting to analyze the *in vitro* activity of engineered nuclease and to demonstrate their site specificity and selectivity.

2. Materials and methods

2.1. Cell culture and plasmids

HEK 293 (ATCC CRL-1573) cells were cultivated at 37 °C, 5% CO₂ in Dulbecco's modified eagle's medium containing 4.5 mg/L D-glucose (DMEM, Fisher scientific, USA), Glutamine 2.5 mM, supplemented with 10% heat inactivated fetal bovine serum (FBS, Gibco, USA) and antibiotics solution (100 U/ml penicillin and 100 µg/mL).

The plasmid expressing pCMV-CPF1-MBP tag is encoding cDNA at the C terminal part of Cpf1 gene. The crRNA was co-expressed using a pcrRNA expression plasmid. The oligonucleotides corresponding to the target site were cloned in the pTarget vector and the oligonucleotides corresponding to the library were cloned in the pUC19 vector (between EcoRI and BamHI restriction sites). The annealed oligonucleotides were ligated overnight with the T4 DNA ligase, in a molecular ratio of one linearized (10 mg of vector) vector for 10 inserts, in a final volume of 100 mL. *E. Coli* competent cells were transformed with the ligation reaction and plated on agarose plates with ampicillin. The clones were collected and the library was amplified by a three hours culture at 37 °C and the DNA was prepared using an anion exchange resin (Quiagen, Germany). The sequences of the oligonucleotides are the following:

Target site sense: AATTCTTTATTCAGTGCTTTGATGTAATCCG

Target site antisense: GATCCGGATTACATCAAAGCACTGGAATATAAG

Library sense: AATTCTTTATTCAGTGCNNNNATGTAATCCG

Library antisense: GATCCGGATTACATNNNNGCACTGGAATATAAG

2.2. Protein expression

HEK293 cells were cultivated in a 10 cm dish at 80% confluency and were transfected using lipofectamine 2000 reagent (100 L) with 50 g pCMV-CPF1-MBP, with or without 10 g of the guide RNA expression vector. 48 h post transfection, the cells were collected, and Cpf1-MBP fusion protein were purified with amylose resin (New England Biolabs, USA). The cells were washed with PBS and re-suspended in 500 L of Lysis buffer containing Tris-HCl pH 8 100 mM, KCl 50 mM, MgCl₂ 20 mM, Triton x-100 0.1%, RNase Inhibitor (200 Unit/mL: New England Biolabs, USA) and Protease inhibitor minus EDTA (Merk, USA). The cells were broken by thermal shock (freezing in liquid nitrogen and thawing at 42 °C). After centrifugation for 10 min at 15000 rpm, the supernatants containing the recombinant proteins were incubated for 2 h with amylose beads (150 L). The beads were washed three times with 1mL of lysis buffer, the recombinant proteins were eluted with a Lysis buffer adjusted at 20 mM maltose (Merk, USA). The purified samples were characterized using a Nano drop spectrophotometer and compared for their 260/280 ratio to analyze the co-purification of the crRNA. The results demonstrate that about 5% of the purified protein correspond to nucleotides. The molecular weight ratio between the Cpf1-MBP protein and the crRNA is 1/10.

2.3. Analysis of nuclease specific activity

The *in vitro* cleavage assay was performed in lysis buffer, using 500 ng of a linearized pTarget plasmid as a substrate, and 1 g of purified Cpf1-MBP RNP in a final volume of 20 L. After 15 min incubation at 37 °C, the reactions were arrested by addition of EDTA, proteinase K, and incubated at 60 °C for 10 min. The DNA fragments were analyzed on a 1% agarose gel with ethidium bromide.

2.4. Analysis of the nuclease substrate selectivity

The *in vitro* cleavage assay was performed in lysis buffer, using 50 ng of the pUC19 library plasmid substrate and 1 g of the Cpf1-MBP RNP in a final volume of 10 L or 20 units of the *Clal* enzyme (NEB, USA). The reactions were preincubated for 1 h on ice followed by an incubated for 15 min 37 °C and used to transform 100 L *E-Coli DH5 alpha* competent cells (Enzynomics, USA). 10% of the *E-Coli* transformation reactions were plated on ampicillin agarose plates and 90% of the reaction were used for a liquid cell culture in LB broth with Ampicillin at 30 °C. The plasmids were prepared using plasmid miniprep kit (Quiagen, Germany) and the target region was amplified using the following primers:

NGS-m13F:

ACATCTTTCCTACACGACGCTCTTCCGATCTNNNCAGAAACAGTACGAC

NGS-m13R:

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNTGTAACGACGGCCAGT

The PCR amplification was performed using the following cycle (x20): Denaturation 15 s., Annealing 15 s. and elongation 15 s. The PCR reaction were amplified for a second round (x20) with the Next Generation Sequencing (NGS) adapter primers and purified using a PCR purification Kit (Quiagen, Germany). The purified PCR fragment were sequenced with an Illumina miniseq. The extracted NGS data were analyzed at the website <http://www.rgenome.net/cas-analyzer>. The score for each sequenced were normalized and adjusted to 1000 for the highest sequence score.

3. Results and discussion

In this report, the design, and the results of an *in vitro* nuclease assay to analyze the specificity and the selectivity of a nuclease for its target sequence, is presented. For this purpose, a cleavage assay with a plasmid library was designed, coupled with a deep sequencing analysis by NGS for the characterization of the un-cleaved substrates. Thus, to analyze the guide RNA substrate selectivity of the *Cpf1* RNP complexes, a pUC19-Target plasmid library was generated that contain an altered target site for a specific target sequence. The target site in the library is a degenerated sequence between the positions 11 to 14 of the corresponding crRNA guide sequence. Hence, the plasmid library encompassing the 256 5'-NNNN-3' possible sequences with only one perfect matches to the guide RNA. In theory, because of large mismatches between the guide RNA and the target DNA, if the nuclease is highly selective for its target sequence, only the correct target sequence must be cleaved within the library. The nucleotide positions degenerated in the target sequence, depicted in the Figure 1A (nucleotides represented in red within the structure), were selected because they are in the core structure of the RNP complex. Nevertheless, this structural region of the *Cpf1* protein is not compact and offer the possibility of the protein architecture to adapt with nucleotides mismatches. In addition, the selected nucleotides are distal to the PAM sequence, and close enough to the catalytic site to disturb its activity (distance of 6 and 9 nucleotides from the cleavage site for the untargeted and targeted strand respectively). The library 5'-NNNN-3' constructed for this assay was analyzed by sanger sequencing capillary electrophoresis (Figure 1B) but this assay is not suitable to discriminate individual sequences and their statistical distribution. To analyze the products before, and after cleavage by the nuclease, the NGS experimentation is required.

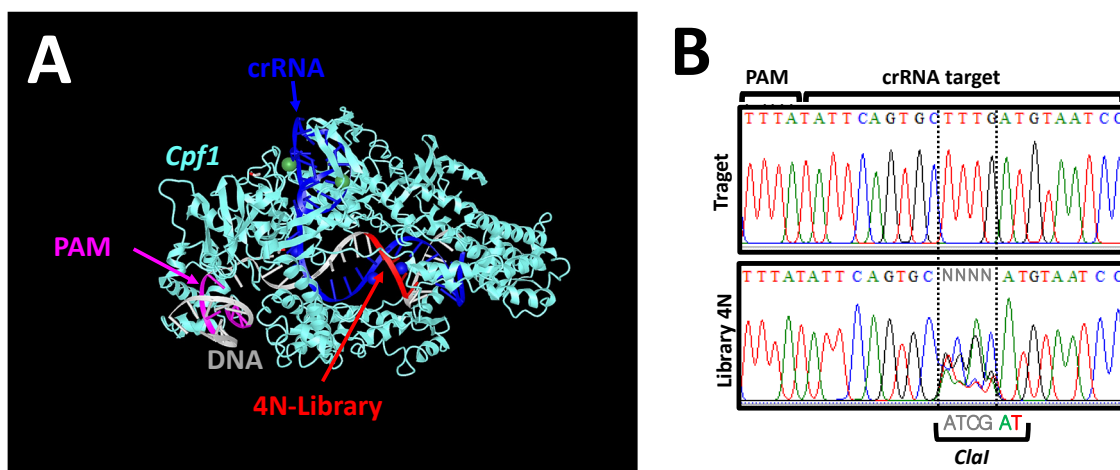


Figure 1. (A) Structural representation of the *Cpf1* RNP associated with its target DNA. The *Cpf1* protein (Light Blue) associated the crRNA (Dark Blue) and the DNA (Grey) is represented to observe the position of the PAM sequence (Magenta) and the 4 nucleotides degenerated in the substrate plasmid 4N library (Red). (B) Sequencing of the target site and the Library 4N: The sequences show the target site (upper panel) cloned in pTarget plasmid and the library 4N (Lower Panel) cloned in pUC19. The PAM (5'-TTTA-3') and the crRNA sequences are underlined, the sequence corresponding to the *Clal* is represented under the library sequence.

To characterize the cleaved products of the library following the nuclease treatment, a subtractive/differential analysis was performed on the un-cleaved substrates. In order to select and clearly isolated the un-cleaved plasmids of the library, a biological filter was used after the DNA cleavage reaction, by transforming *E-Coli*. Indeed, only circular DNA are capable to replicate in bacteria and allow to stringently separate circular DNA substrates from linear products. Two parameters were taken into account for the analysis. First, the density of clones on agarose plates following the *E-Coli* transformation that reflect quantity of the plasmids linearized. Second, the quantitative assessment by NGS of each 256 sequences present within the library, before and after the nuclease treatment. The sequences of the plasmids in the samples were tested by PCR amplification of the nuclease target region, and deep sequencing analysis of the amplicon. First, the technical conditions and the biological filter were validated. Thus, following the *E-Coli* transformation, the NGS experiments for both samples, the library and biological filter must correlate. The deep sequence analysis of the original library, and following its transformation and amplification in *E-Coli* with overnight culture at 30 or 37 °C, were compared. The suboptimal temperature allows a linear cell growth, whereas the optimal temperature is with an exponential phase and might accentuate the sequence distribution differences within the library. The As expected, the analysis of the NGS data (Figure 2) show that the correlation coefficient (R) is high for both conditions, but the distribution of the dots in the graphs are different, reflecting that the representation of the sequences is not identical to the initial “library” with the *E-Coli* culture at 37 °C. On the contrary, the culture condition at the 30 °C shows a linear distribution of the dots on the diagonal. In addition, the subtraction of “filter” to the “library” (left Panel for both graphs, the values and the corresponding sequences are sorted from the smaller to the biggest with the “Filter” data set) show a distribution of the dots on the x-axis. The 30 °C cell culture condition is accurate to characterize the sequences that are not amplified following the nuclease reaction.

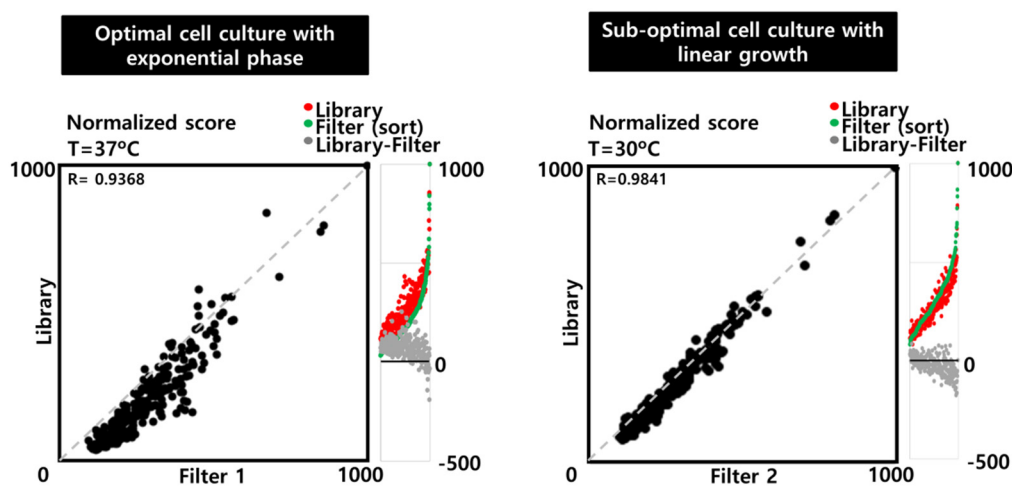


Figure 2. Diagram of the analysis of the biological filter to discriminate circular and linear DNA in the *in vitro* nuclease reactions: The NGS results of the two conditions tested for the plasmid library amplification Filter 1 (37 °C) and Filter 2 (30 °C) are compared to the original Library to show their correlation. The R value is calculated for both conditions and the distribution of the sequences on the left panel (red dots: original library, green dots: Filter, grey dots: value of the Filter minus the original library).

The study was performed with a *CpfI*-MBP RNP complex isolated from mammalian cells, in order to determine the activity of the nuclease extracted from the same environment than in genome engineering experiment (i.e.: exposure to the cellular enzymes that can induce post-translational modification of the protein or post-transcriptional of the crRNA). Because of the MBP tag fusion at the C-terminal region of the *CpfI* protein, the full-length protein *CpfI*-MBP can be purified in a single step from cellular extract. The ribonucleoproteic complex was extracted from the Hek293 cells transfected with a *CpfI*-MBP expression vector and the crRNA expression plasmid. The fusion protein *CpfI*-MBP purified with the amylose resin (Figure 3A) was tested for its *in vitro* cleavage activity using a linearized plasmid containing the target site cloned adjacent to an *EcoRI* restriction site (Figure 3B). The linear DNA substrate was incubated with the *EcoRI* restriction enzyme as a control for the size of the *CpfI* RNP cleaved products. The *CpfI*-MBP RNP complex show the same pattern than the *EcoRI* control reaction, demonstrating the high specificity of the purified *CpfI*-MBP RNP complex for the target site. The *CpfI*-MBP RNP complex was then tested for its substrate selectivity using the pUC19-Target plasmid library. A control was performed with the restriction enzyme *Clal* (5'-ATCGAT-3'), routinely used in molecular biology, that cleave a single site present in a unique plasmid within the library. Indeed, there is no *Clal* cleavage site in the original pUC19 plasmid, and only one sequence (corresponding to the 5'-ATCG-3' altered sequence) in the library that is a full-length palindromic site for the *Clal* restriction enzyme. Thus, over the 256 plasmids from the library, a single plasmid contain the *Clal* target, and the others correspond to a 5'-NNNNTA-3' sequence degeneration covering more than the half of the palindromic sequence recognize by this restriction enzyme. The results of the experiment (Figure 3C) shows the bacterial culture dishes obtained after 15 min of incubation with the nucleases. The plates indicate that the number of transformed bacteria (10% of the reaction) is identical between the "filter" and the "*Clal*" samples, whereas the number of clones is dramatically reduced for the sample "*CpfI*-MBP". With a longer incubation time (1 h) the library amplification is completely abolished for the reaction with the *CpfI*-MBP RNP but not with *Clal* restriction enzyme (data not shown). Two additional experiment were performed to control the unspecific activity for the restriction enzyme. First, with the restriction enzyme *EcoRI* that cleave all plasmids of the library, and completely abolish the *E.Coli* library amplification (No clones were obtained). Second, with the restriction enzyme *XmaI* that do not cleave any plasmid of the library. This observation demonstrate that the *CpfI*-MBP RNP is able to cleave all the plasmids from the library but not the *Clal* restriction enzyme. The plasmids isolated from the bacteria were then tested by NGS sequencing to characterize the cleaved products. The deep sequencing analysis of the different samples is depicted in the diagrams of the sequence distribution. The sequences were sorted from the smallest to the biggest number based on the "Filter" sample to demonstrate that bacterial amplification is equivalent to the original plasmids within the library. The data analysis with the *Clal* nuclease sample (middle Panel) show the deletion of a single sequence (5'-ATCG-3'). The experiment with the *XmaI* enzyme that do not cleave any plasmids within the library is similar to the control experiment, reflecting the very high selectivity of the restriction enzyme for the its target site. On the contrary, the *CpfI*-MBP RNP (right panel) shows that several sequences disappear from the library. Those sequences correspond to the target site (5'-TTTG-3') and to the sequences with a single nucleotide mismatch (Figure 3D). This observation demonstrated that the *CpfI*-MBP RNP affinity for the target sequence, and the sequences with a single nucleotide mismatch, is higher and are preferentially selected by the complex.

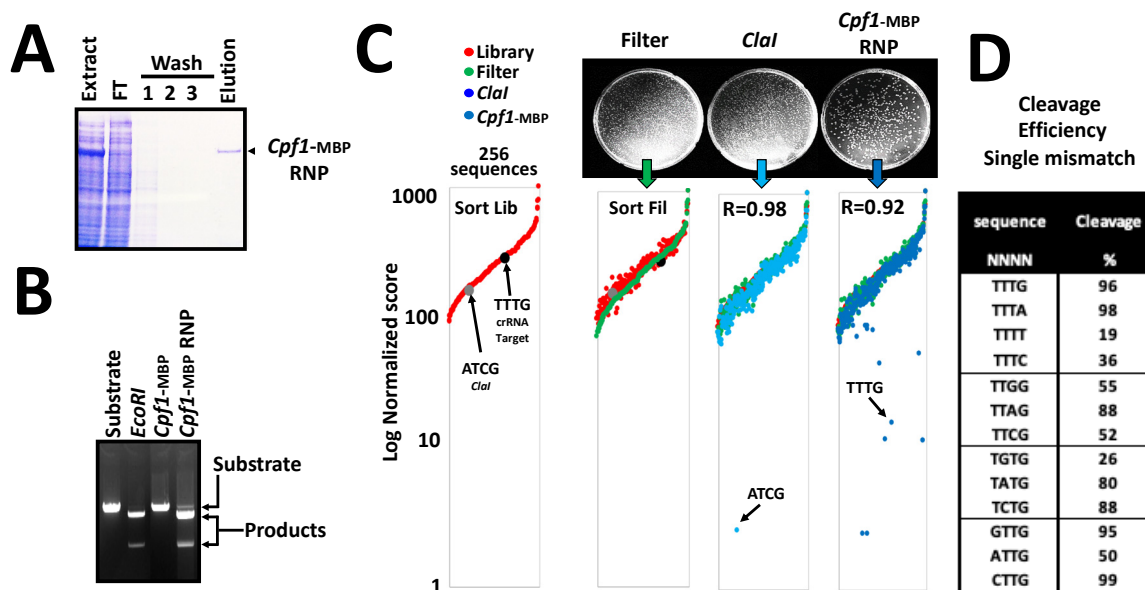


Figure 3. (A) *Cpf1*-MBP RNP purification: The protein extract, the flow through (FT), the washing steps (1, 2 and 3) and the elution of the *Cpf1*-MBP RNP were analyzed on a Denaturing SDS page 4–20% acrylamide gel. (B) In vitro Cleavage activity of the *Cpf1*-MBP RNP on a specific target site: the *Cpf1*-MBP purified with (*Cpf1*-MBP RNP) and without the crRNA were tested for their cleavage activity on a linearized plasmid substrate in parallel with an *EcoRI* reaction to control the products size. (C) Selectivity of the *Cpf1*-MBP RNP on the plasmid substrate library (4N): The top panel shows the plates of bacterial cultures for the different tested samples (Filter, *ClaI* and *Cpf1*-MBP). The diagram (Logarithmic scale) of the distribution for the 256 sequences within the library (left panel, Sort Lib., Red dots) are depicted with the sequence corresponding to the *ClaI* (light grey dot) site and the crRNA (Black dot). with the filter (Sort Fil., green dots) and analyzed together with the *ClaI* sample (Light Blue dots) or the *Cpf1*-MBP RNP sample (Dark blue dots). (D) Table of the sequences with one mismatch cleaved by the *Cpf1*-MBP RNP: Each sequence is depicted with the percentage of reduction compared to the value from the Filter.

4. Conclusions

The results of our study concluded that the assay developed is precise to analyze both, the specificity and the substrate selectivity of nucleases. We demonstrated that the *Cpf1*-MBP RNP is highly specific to its target sequence but is not highly selective for its DNA substrate as it is for the restriction enzymes. Indeed, we observed a clear reduction of all the clones within the library compared to the *ClaI* restriction enzyme. This in turn suggests that the *Cpf1*-MBP RNP might induce a large number of unspecific cleavages in the genome of a eukaryotic cells, even with a highly specific crRNA for its target site. Hence, the molecular engineering of the *Cpf1* enzyme is required to obtain a highly selective enzyme that stringently recognizes its target site in the genome.

Acknowledgements

The author is grateful to his previous colleagues from the Pasteur Institute to provide their help for the screening assay.

Conflict of interest

The author has declared no conflict of interest.

References

1. Kühnlein U, Linn S, Arber W (1969) Host specificity of DNA produced by *Escherichia coli*, XI. In vitro modification of phage fd replicative form. *Proc Natl Acad Sci USA* 63: 556–562.
2. Bibikova M, Carroll D, Segal DJ, et al. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* 21: 289–297.
3. Christian M, Cermak T, Doyle EL, et al. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186: 757–761
4. Gasiunas G, Barrangou R, Horvath P, et al. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 109: E2579–E2586.
5. Jinek M, Chylinski K, Fonfara I, et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816–821.
6. Zetsche B, Gootenberg JS, Abudayyeh OO, et al. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163: 759–771.
7. Nishimasu H, Ran FA, Hsu PD, et al. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156: 935–949
8. Yamano T, Nishimasu H, Zetsche B, et al. (2016) Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* 165: 949–962.
9. Kim D, Kim J, Hur JK, et al. (2016) Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol* 34: 863–868.
10. Kamps-Hughes N, Quimby A, Zhu Z, et al. (2013) Massively parallel characterization of restriction endonucleases. *Nucleic Acids Res* 41: e119.
11. Murugan K, Seetharam AS, Severin AJ, et al. (2020) CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects: EDITORS'PICK: Cas12a nickase activities. *J Biol Chem* 295: 5538–5553.



AIMS Press

© 2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)