Mad7: An IP friendly CRISPR enzyme

Johan Rojek[1], Yogesh Basavaraju[1], Saranya Nallapareddy[1], Roland Baumgartner[2], Sanne Schoffelen[1], Tanya Lipscomb[2], ryan gill[1], and Lasse Pedersen[1]

[1]Danmarks Tekniske Universitet
[2]Artisan Bio

March 07, 2024

Abstract

Mad7 is a CRISPR enzyme, similar to Cas9 and Cas12a, which is of great interest to industry and academia due to its permissive licensing agreement, which states that everyone is free to use it but not distribute it commercially. This paper demonstrates that Mad7 genome editing works in CHO cells and provides the CHO engineering community with the computational tools needed to generate Mad7 compatible gRNAs against the CHO genome.

Introduction

CRISPR enzymes, like Cas9, have revolutionized the world of genome editing, as most recently evidenced by the 2020 Nobel Prize in Chemistry being awarded "for the development of a method for genome editing". Several other CRISPR enzymes have been described in its wake, such as the widely used Cas12a (formerly Cpf1) (Zetsche et al. 2015). In fact, due to the influx of new CRISPR-Cas variants, the classification scheme has recently been expanded to 2 classes, 6 types, and 33 subtypes (Makarova et al. 2020).

Agreeable terms of use, such as reasonably priced licensing, are of particular interest to the world of commercial genome engineering, in particular, but not limited to, biotechnology and pharma companies. A CRISPR enzyme called Mad7 (also known as ErCas12a) discovered by the company Inscripta comes with a particularly agreeable license that allows free use for academic *and* commercial research and development purposes. According to Inscripta: "*Royalties are only attached to the use of the MAD7 nuclease if a commercial manufacturing process uses the MAD7 nuclease on an ongoing basis or a product physically contains the MAD7 nuclease* ".

However, besides a few unreviewed texts by Inscripta themselves, only four peer-reviewed studies have currently tested Mad7 (Wierson et al. 2019; Liu et al. 2020; Price et al. 2020; Jarczynska et al. 2021). In summary, they demonstrate that Mad7 can generate indel efficiencies similar to Cas9 in various organisms, that Mad7 can be used for targeted DNA insertions, and that a deactivated variant can facilitate CRISPR-based transcriptional inhibition.

In this paper, we tested the ability of Mad7 to produce indels in CHO cells. We tested two different Mad7 variants with either 1 or 4 nuclear localization sequences (NLS) and two different delivery protocols: ribonucleoprotein (RNP) and plasmid-based delivery.

We found that Mad7 works in CHO cells using both RNP and plasmid-based delivery protocols but that the 4xNLS variant was clearly superior to the 1xNLS. Additionally, we present two pieces of software: One is an update to our gRNA finding tool CRISPy (Ronda et al. 2014), now called CRISPyR, that works with

Mad7 PAM sequences, and the other is "Hamplicon", a tool for automatization of indel detection analysis of NGS amplicon data.

Methods

gRNA design and preparation

CRISPyR (https://github.com/laeblab/crispy) was used to identify Mad7 gRNAs as demonstrated in the supplemental Jupyter notebook (https://github.com/laeblab/Mad7_CHO_analysis). Briefly described, a Chinese Hamster genome sequence was obtained from NCBI. CRISPyR was used to find all Mad7 compatible target sites from the exons of the Timp1 and the BGN gene. 4 gRNAs per gene were selected, targeting the first exon of either gene. Sequences are provided in supplementary table 1.

gRNAs for Mad7 RNPs were purchased as "Alt-R CRISPR-Cpf1 crRNA" from Integrated DNA Technologies (IDT). gRNAs for Cas9 RNPs were purchased as "TrueGuide Synthetic gRNA" from ThermoFisher. The gRNAs were prepared by reconstituting the gRNA in IDTE buffer to a concentration of 100. gRNA for plasmids were purchased as DNA oligos (IDT) and cloned into the gRNA expression plasmid (PL0088_-CLED) using USER cloning with primers designed using AMUSER (Genee et al. 2015).

Design, expression and production of Mad7 variants

Two versions of Mad7 were designed. A 1xNLS (pNic28-EcMad7-1xNLS) and a 4xNLS (pNic28-EcMad7-4xNLS) version. Both are polyhistidine-tagged for protein purification.

E. coli BL21star (DE3) cells were transformed with the His-tagged Mad7 expression vector. 2xYT medium supplemented with kanamycin was inoculated with a single colony and incubated overnight at 37 °C. The culture was diluted in 1-2 L 2xYT medium to an OD_{600} of 0.1 and grown at 37 °C to an OD_{600} of 0.6. At this point, the culture was placed in an ice bath for 15-20 min. Next, IPTG was added to a final concentration of 0.2 mM, and protein expression was allowed to take place at 18 °C overnight (18-20 h).

Cells were harvested by centrifugation and resuspended in lysis buffer (20 mM Tris, 500 mM NaCl and 10 mM imidazole, pH 8.0) supplemented with "cOmplete Protease Inhibitor Cocktail" (Merck, cat# 11697498001). After resuspension, Benzonase nuclease (Merck, cat# E1014-5KU, 10 μL per 40 mL lysate) and lysozyme (Merck, cat# 10837059001, 1 mg/mL lysate) were added, and the cell suspension was left on ice for 30 min. Cells were disrupted on an Avestin EmulsiFlex C-5 homogenizer (15-20 Kpsi), and the insoluble cell debris was removed by centrifugation (15,000 g, 4 °C, 15 min). All subsequent chromatography steps were carried out at 10 °C. The cleared lysate was loaded on a 5 mL HisTrap FF column (Cytiva, cat# 17525501). The resin was washed with 10 column volumes of wash buffer (lysis buffer, but with 20 mM imidazole), and the protein was eluted with 10 column volumes of elution buffer (lysis buffer, but with 250 mM imidazole). Fractions containing the protein (typically 13.5 mL) were pooled and diluted to 25 mL in dialysis buffer (250 mM KCl, 20 mM HEPES and 1 mM DTT, and 1 mM EDTA, pH 8.0). The sample was dialyzed against 1 L of dialysis buffer at 10 °C using a dialysis membrane tubing with a molecular weight cut-off of 6-8 kDa (Spectra/Por standard grade regenerated cellulose, 23 mm wide). The dialysis buffer was replaced after 1-2 hours, and dialysis continued overnight. The next day, the dialyzed sample was diluted two-fold in 10 mM Hepes (pH 8) and immediately loaded on a 5 mL HiTrap Heparin HP column (Cytiva, cat# 17040601), pre-equilibrated with buffer A (20 mM Hepes, 150 mM KCl, pH 8.0). The resin was washed with 2 column volumes of buffer A, and the protein was eluted using a linear gradient from 0 to 50 % of buffer B (20 mM Hepes, 2M KCl, pH 8.0) over 12 column volumes. Fractions containing the protein were pooled (typically 10-15 mL) and up-concentrated to 2 mL using a centrifugal filter unit (Amicon® Ultra-15, 30K MWCO; centrifugation at 4 °C). A final chromatography step was performed by injecting the up-concentrated sample on a 120 mL Superdex200 gel filtration column (Cytiva #28989335) with 50 mM sodium phosphate, 300 mM NaCl, 0.1 mM EDTA, pH 7.5 as separation buffer. Fractions of interest were pooled and up-concentrated

to at least 20 mg/mL (concentration determined by measuring absorbance at 280 nm on a NanoDrop 2000 (ThermoFisher) with a percent solution extinction coefficient (Abs 0.1%) of Mad7 of 1.1). The concentrated protein solution was supplemented with glycerol (20 % (v/v) final concentration) and DTT (1 mM final concentration), snap-frozen in liquid nitrogen, and stored at -80 °C. Approximately 20 mg of nuclease was isolated from 1 L of E. coli culture.

Plasmid construction

The plasmids used for transfection were constructed with uracil-specific excision reagent (USER) cloning. The DNA fragments for USER cloning were generated by PCR amplification with Phusion U Hot Start PCR Master Mix (Thermo Scientific, Cat #F533S) and uracil-containing primers designed with Amuser 1.0 (Genee et al. 2015; Lund et al. 2014). DNA fragments were generated and assembled to construct the final plasmids: Mad7-1X-NLS and Mad7-4X-NLS. The gRNA plasmids were also constructed with USER cloning using overhang primers holding the specific gRNA spacers to generate DNA fragments that would self-ligate in-frame with the U6 promoter. All plasmids were sequence-verified. The plasmid vectors were transformed into *Mach1* cells (Invitrogen, cat #C869601), expanded in 2xYT medium, and isolated using Plasmid Midi Kit (Qiagen, cat #12143) to obtain transfection grade plasmids.

Cell cultivation

CHO-S cells (Gibco, cat# A1155701) were maintained in CD-CHO media (Gibco, cat# 10743029) supplemented with 8 mM L-Glutamine (Gibco, cat# A2916801) in Erlenmeyer shake flasks with vent caps (Corning, cat #431143 and #431144) at 37 °C in a humidified atmosphere with 5% CO_2 on an orbital shaker platform rotating at 120 rpm with a 25 mm shaking amplitude. The cells were passaged every 2-3 days at least three times before transfection, making sure the cells were in exponential growth on the day of transfection.

Plasmid transfection

On the day of transfection, 1.5 x 10e6 cells were seeded per well in a 12-well plate (Falcon, cat#351143). The transfections were with Freestyle MAX Reagent according to the manufacturer's (Gibco, cat#16447100) with a total of 1.8 μg of plasmid (1:1 ratio of gRNA:Mad7). After transfection, the cells were incubated for 48 hours at 37 °C in a humidified atmosphere with 5% CO_2 on an orbital shaker platform rotating at 120 rpm with a 25 mm shaking amplitude before cell harvest for genomic DNA extraction.

Electroporation of CHO cells with RNPs

RNP mixtures were prepared at a 1.2:1 molar ratio between gRNA and nuclease. DPBS, gRNA, and nuclease were added to PCR plates (Sarstedt, cat#72.1978.202) in the specified order for a total volume of 5 μL with either 10 pmol RNP/sample or 50 pmol RNP/sample. The solutions were gently mixed, incubated at room temperature for 15 min, and kept at 4 °C until transfection.

Electroporations were performed using the SG Cell Line 96-well Nucleofector Kit (Lonza cat# V4SC-3096) with the Amaxa Nucleofector 96-well Shuttle (Lonza, cat# AAM-1001S) and the 4D-NucleofectorTM Core Unit (Lonza, cat#AAF-1002B).

The cells were prepared for transfection by spinning down the cultures at 200 g for 5 minutes, removing the supernatant completely by aspiration, and carefully resuspending the cells in 20 μL Nucleofector supplement/Nucleofector Solution solution per 200.000 cells, and distributing 20 μL cell solution to each of the 5 μL gRNA/RNP complexes. The entire volumes (25 μL of cell + RNP complex solution) were transferred to the Nucleocuvette plate, placed in the Lonza 96-well electroporation shuttle, and electroporated using Lonza program 96-FF-137. The Nucleocuvette plates were added 80 μL prewarmed media to each well, and 45 μL was transferred from the Nucleocuvette plates into 96-well F-bottom plates (Corning Cat#351172)

3

containing 155 μL prewarmed medium. The plates were then incubated for 48 hours at 37 celcius in a humidified atmosphere with 5% CO_2 before cell harvest for genomic DNA extraction.

DNA extraction, amplicons generation, and preparation for NGS

48 hours post-transfection, the cells were mixed by pipetting 100 μL up and down. 100 μL cells were then transferred to 96-well PCR plates (Sarstedt, Cat#72.1978.202), and the plates were centrifuged at 1,000 g for 10 minutes at room temperature. The supernatants were removed by rapid inversion onto a KimWipe. 100 μL QuickExtract (Lucigen, cat# QE9050) was added to each well, and the cells were resuspended by pipetting up and down several times. Finally, the plates were incubated at 65 °C for 15 minutes and 95 °C for 5 minutes and stored at -80 °C.

Amplicons of the target regions were generated using primers with specific overhang adaptors (primers in supplemental table 1). Forward overhang: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[locus-specific sequence]-3'. Reverse overhang: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[locus-specific sequence]-3'. 2 μL gDNA mixture was used as template for touchdown PCR using Phusion High-Fidelity PCR Master Mix with HF Buffer (Thermo Scientific, cat#F-531L) and the following thermocycler conditions: 95 °C for 7 min; 20 cycles of 95 °C for 45 s, 69.5 °C to 59.5 °C [-0.5 °C/cycle] for 30 s, 72 °C for 30 s; 40 cycles of 95 °C for 45 s, 59 °C for 30 s, 72 °C for 30 s; 72 °C for 7 min. The amplicons were run on a 2% agarose gel for size verification.

Agencourt AMPure XP beads (Beckman Coulter, cat#A63881) were used to purify the amplicons and wash away free primers and primer-dimers according to the manufacturer's protocol. Index PCR was carried out using the Nextera XT Index Kit V2 Sets (Illumina) and the KAPA HiFi HotStart ReadyMix (Roche, cat#7958935001). Each reaction consisted of 5 μL amplicon DNA, 2.5 μL Nextera XT Index 1 primer (N7XX), 2.5 μL Nextera XT Index 2 primer (S5XX), 12.5 μL KAPA HiFi HotStart ReadyMix, and 2.5 μL nuclease-free water. The following Thermocycler program was used for the index PCR: 95 °C for 3 min; 8 cycles of 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s; 72 °C for 5 min. The amplicons were then purified with Agencourt AMPure XP beads, and the amplicon library product sizes were validated using a Bioanalyzer DNA 1000 Kit (Agilent, cat#5067-1504) with the Agilent 2100 Bioanalyzer. The library was quantified with a Qubit dsDNA HS assay kit (Invitrogen, cat#Q32851) and the Qubit Fluorometer, diluted to 10 nM, and submitted for NGS.

NGS data analysis and visualization

NGS data were analyzed and visualized using a Jupyter Lab notebook (https://github.com/laeblab/Mad7_-CHO_analysis). All figures and results can be replicated by running the code in the notebook against the raw data uploaded to Zenodo (https://doi.org/10.5281/zenodo.5020293).

Statistics

A one-way ANOVA was used to compare the percentage of indels in pools of cells treated with RNP + gRNA versus the mean of each amplicon's controls in each setup. If significant differences were found, an all versus all comparison was made using Tukey's range test.

Results

Software for gRNA design and amplicon analysis

Since no software was available for designing Mad7 compatible gRNA's, we extended our Cas9 gRNA software, CRISPy (Ronda et al. 2014), to also work with Mad7. For performance reasons, the program was

rewritten in RUST. The program is freely available online (https://github.com/laeblab/crispy).

The software works by first indexing all potential gRNA sites in a user-supplied genome sequence. Afterward, this index can be used to search for gRNAs in specified regions, and the discovered gRNAs will then be compared to all other potential gRNA sites in the genome to evaluate potential off-targets. The program is highly performant and can be run on a standard laptop.

Additionally, to minimize the time spent analyzing NGS amplicon data from genome editing, we developed a tool called "Hamplicons", as it uses hamming distance to analyze amplicons. This tool is also freely available online (https://github.com/laeblab/hamplicons).

The user supplies hamplicons with a fasta file containing the expected wild-type amplicons as well as the raw sequencing (fastq) files, and then hamplicons analyzes all the fastq files for the presence of any of the amplicons and generates a report of indels observed in the data. Out of the box, Hamplicons is set up to work with our preferred method for generating amplicon sequencing data, consisting of Illumina paired-end sequencing with a short overlap between read-pairs. The folder and file structure generated by Illumina sequencers is recognized and used to correctly pair and group files.

The indel analysis recognizes whether the merged read-ends are identical/similar to the wild-type amplicon sequences using hamming distance. When a read is recognized as a particular amplicon, then the length of the merged read is compared to the expected amplicon length, and any discrepancies are reported as indels.

Mad7 with 4xNLS is efficient for generating indels in CHO cells

To evaluate whether Mad7 works in CHO cells, we generated eight gRNAs (supplementary table 1) targeting two genes; BGN and Timp1. These genes were chosen because we had previously successfully edited them using Cas9 and thus knew that these targets were accessible for the CRISPR enzyme (Kol et al. 2020). Additionally, we repeated Bgn and Timp1 KO experiments from that study using an RNP protocol which, as can be seen in supplementary figure 1, was highly efficient.

We tested two versions of Mad7, one with 1xNLS and one with 4xNLS, at two different concentrations, high (50 pmol/sample) (figure 1) and low (10 pmol/sample) (supplementary figure 2), using the RNP protocol.

As shown in figure 1, the Mad7 with 4xNLS gave high indel levels and worked significantly better than the 1xNLS. The latter was rarely statistically significant compared to the controls. The low concentration did not result in significant indel rates.

Figure 1 : Genomic indel rates generated by Mad7 RNP together with 1 of 8 different gRNAs targeting either BGN or Timp1. "Mad7 1xNLS" (blue) is Mad7 with 1 NLS and the indicated gRNA, "Mad7 4xNLS" (red) is Mad7 with 4 NLS and the indicated gRNA, "Mad7 RNP, no EP" (yellow) is Mad7 without any gRNA, "+EP, No RNP" (green) is electroporation without RNP, but with indicated gRNA, "wt" (gray) is wild type cells with no RNP nor electroporation. Error bars are standard deviations of 3 replicates.

Plasmid based expression of Mad7 also works

We tested co-transfection of a Mad7 expression plasmid and a gRNA expression plasmid as an alternative to the RNP protocol. Plasmid transfection has the advantage of not requiring the prior production and purification of the Mad7 protein. We targeted the same genes using the identical gRNA sequences as in the RNP experiment. As shown in figure 2, the plasmid-based protocol can also generate indels but shows lower levels of indels than the RNP protocol. Similar to the RNP protocol, though less distinct, the 4xNLS version is superior to the 1xNLS version.

Figure 2 : Genomic indel rates generated by Mad7 expression plasmid together with 1 of 4 different gRNA expression plasmids targeting either BGN or Timp1. "Mad7 1xNLS" (blue) is Mad7 with 1 NLS and the indicated gRNA, "Mad7 4xNLS" (red) is Mad7 with 4 NLS and the indicated gRNA, "wt" (gray) is wild type cells with no RNP nor electroporation, "Mad7 1xNLS No gRNA" (green) is the Mad7 1xNLS expression

plasmid, but with no gRNA, "Mad7 4xNLS No gRNA" (green) is the Mad7 4xNLS expression plasmid, but with no gRNA. Error bars are standard deviations of 3 replicates.

Conclusion

This paper, demonstrates that Mad7 can be used for genome engineering in CHO cells and provide gRNA design and indel analysis software. The results show that nuclear localization sequences are essential for the efficiency of the Mad7, with the 4xNLS version outperforming the 1xNLS version in the majority of circumstances. Futhermore, there are notable differences in indel percentages between the individual gRNAs, which is reminiscent of early Cas9 gRNA design. For this reason, we believe future work should focus on improving the gRNA prediction software, possibly by repeating the excellent work conducted to advance the Cas9 platform.

Acknowledgements

Bibliography

Genee, Hans Jasper, Mads Tvillinggaard Bonde, Frederik Otzen Bagger, Jakob Berg Jespersen, Morten O. A. Sommer, Rasmus Wernersson, and Lars Rønn Olsen. 2015. "Software-Supported USER Cloning Strategies for Site-Directed Mutagenesis and DNA Assembly." ACS Synthetic Biology 4 (3): 342–49. https://doi.org/10.1021/sb500194z.

Jarczynska, Zofia D., Jakob K. H. Rendsvig, Nichlas Pagels, Veronica R. Viana, Christina S. Nødvig, Ferdinand H. Kirchner, Tomas Strucko, Michael L. Nielsen, and Uffe H. Mortensen. 2021. "DIVERSIFY: A Fungal Multispecies Gene Expression Platform." ACS Synthetic Biology 10 (3): 579–88. https://doi.org/10/gjmcp2.

Kol, Stefan, Daniel Ley, Tune Wulff, Marianne Decker, Johnny Arnsdorf, Sanne Schoffelen, Anders Holmgaard Hansen, et al. 2020. "Multiplex Secretome Engineering Enhances Recombinant Protein Production and Purity." Nature Communications 11 (1, 1): 1908. https://doi.org/10/gjmhn8.

Liu, Zhenyi, John A. Schiel, Elena Maksimova, Žaklina Strezoska, Guojun Zhao, Emily M. Anderson, Yumei Wu, et al. 2020. "ErCas12a CRISPR-MAD7 for Model Generation in Human Cells, Mice, and Rats." The CRISPR Journal 3 (2): 97–108. https://doi.org/10.1089/crispr.2019.0068.

Lund, Anne Mathilde, Helene Faustrup Kildegaard, Maja Borup Kjær Petersen, Julie Rank, Bjarne Gram Hansen, Mikael Rørdam Andersen, and Uffe Hasbro Mortensen. 2014. "A Versatile System for USER Cloning-Based Assembly of Expression Vectors for Mammalian Cell Engineering." Edited by Mark Isalan. PLoS ONE 9 (5): e96693. https://doi.org/10.1371/journal.pone.0096693.

Makarova, Kira S., Yuri I. Wolf, Jaime Iranzo, Sergey A. Shmakov, Omer S. Alkhnbashi, Stan J. J. Brouns, Emmanuelle Charpentier, et al. 2020. "Evolutionary Classification of CRISPR–Cas Systems: A Burst of Class 2 and Derived Variants." Nature Reviews Microbiology 18 (2, 2): 67–83. https://doi.org/10.1038/s41579-019-0299-x.

Price, Marcus A., Rita Cruz, James Bryson, Franck Escalettes, and Susan J. Rosser. 2020. "Expanding and Understanding the CRISPR Toolbox for Bacillus Subtilis with MAD7 and dMAD7." Biotechnology and Bioengineering 117 (6): 1805–16. https://doi.org/10.1002/bit.27312.

Ronda, Carlotta, Lasse Ebdrup Pedersen, Henning Gram Hansen, Thomas Beuchert Kallehauge, Michael J. Betenbaugh, Alex Toftgaard Nielsen, and Helene Faustrup Kildegaard. 2014. "Accelerating Genome Editing in CHO Cells Using CRISPR Cas9 and CRISPy, a Web-Based Target Finding Tool." Biotechnology and Bioengineering 111 (8): 1604–16. https://doi.org/10.1002/bit.25233.

Wierson, Wesley A., Brandon W. Simone, Zachary WareJoncas, Carla Mann, Jordan M. Welker, Bibekananda Kar, Michael J. Emch, et al. 2019. "Expanding the CRISPR Toolbox with ErCas12a in Zebrafish and Human Cells." The CRISPR Journal 2 (6): 417–33. https://doi.org/10.1089/crispr.2019.0026.

Zetsche, Bernd, Jonathan S. Gootenberg, Omar O. Abudayyeh, Ian M. Slaymaker, Kira S. Makarova, Patrick Essletzbichler, Sara E. Volz, et al. 2015. "Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System." Cell 163 (3): 759–71. https://doi.org/10.1016/j.cell.2015.09.038.

Supplementary material

Figure S1 : Genomic indel rates generated by Cas9 RNP together with 1 of 8 different gRNAs targeting either BGN or Timp1. "wt" (gray) is wild type cells with no RNP nor electroporation. "+EP, No RNP" (green) is electroporation without RNP, "Cas9" (red) is Cas9 RNPs with electroporation. Error bars are standard deviations of 3 replicates.

Figure S2 : Genomic indel rates generated by Mad7 RNP, in low concentration, together with 1 of 8 different gRNAs targeting either BGN or Timp1. "Mad7 1xNLS" (blue) is Mad7 with 1 NLS and the indicated gRNA, "Mad7 4xNLS" (red) is Mad7 with 4 NLS and the indicated gRNA, "Mad7 RNP, no EP" (yellow) is Mad7 without any gRNA, "+EP, No RNP" (green) is electroporation without RNP, but with indicated gRNA, "wt" (gray) is wild type cells with no RNP nor electroporation. Error bars are standard deviations of 3 replicates.

Hosted file

SuppTable1_Oligos.docx available at https://authorea.com/users/430079/articles/710789-mad7-an-ip-friendly-crispr-enzyme