INSCRIPTA™

# Initial characterization of CRISPR-MAD7 in microbe systems

## Introduction

CRISPR-enabled genome editing is revolutionizing various aspects of biotechnology. While there is a broad range of known CRISPR systems (1-2), the most widely used is the CRISPR-Cas9 system that utilizes a nucleic acid-directed endonuclease from *Streptococcus pyogenes* (3-5). It has become increasingly clear, however, that a system using only one nucleic acid-guided endonuclease may not be sufficient to address the breadth of possible applications with the desired levels of performance.

To address the need for an expanded range of nucleic acid-directed endonucleases, Inscripta is developing new classes of RNA-guided endonucleases called MADzymes, the name inspired by the biological diversity found on the island of Madagascar. These nucleases are being engineered to have improved features such as different PAM recognition sequences, different cut efficiencies, reduced sizes, or different enzyme kinetics.

By developing and broadly releasing these enzymes, Inscripta strives to promote widespread exploration and adoption of CRISPR tools in both academic and commercial settings. Inscripta has now released the MAD7 enzyme from its nuclease discovery and engineering platform. Preliminary studies show that MAD7 has the ability to target, cut, and edit specific PAM sites in microbes using CRISPR systems.

## MAD7 Overview

MAD7 is 1263 amino acids in length and exhibits homology to the Type V (Cpf1-like) family of CAS enzymes. Structural modelling of MAD7 shows similarity to the *Acidaminococcus sp.* Cpf1 (6-7), and although overall amino acid sequence identity is only 31%, the sequence of the PAM recognition domain is highly conserved between these two enzymes (Figure 1).

Driven by this similarity, we cloned a codon-optimized version of MAD7 and tested it using design rules related to those previously described for Cpf1 orthologs (6-7). We evaluated gRNA spanning a range of design criteria, including different AT-rich PAM sites, spacer lengths, and repeat sequences.

We determined that MAD7 can efficiently utilize YTTN PAMs (Figure 2), which are frequent throughout the *E. coli* and *S. cerevisiae* genomes; approximately 95% of the genome lies within 50 nucleotides (NT) of a YTTN site in both species. MAD7 functions with gRNA designs that include a 21-NT truncation of the 5′ repeat sequence, a TTTN or CTTN PAM targeting region, and a 21-NT spacer that is directly adjacent to the 3′ end of the targeted PAM.

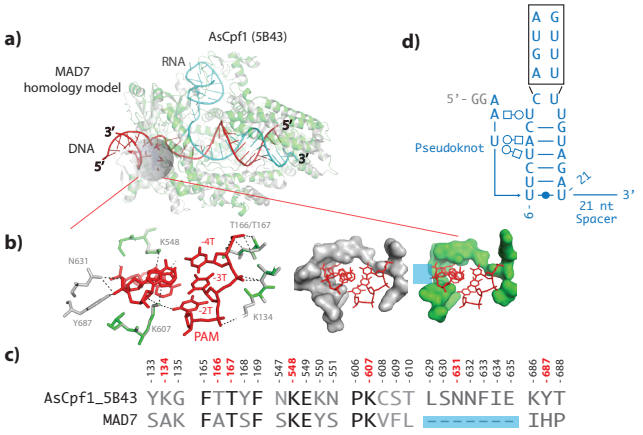


**Figure 1. A.** Overlay of MAD7 inferred structure (*green*) on the crystal structure of the related ortholog AsCpf1 (*grey*) in complex with guide RNA (gDNA; *aqua*) and DNA (*red*). **B.** Expanded view of the PAM recognition domain structure showing hydrogen bonds with DNA (*indicated by dashed lines*). **C.** Comparison of PAM recognition domain amino acid (AA) residues of MAD7 and AsCpf1 showing hydrogen bond contacts to DNA (*red*) and position relative to AsCpf1. The MAD7 nuclease used in these studies is 1263 AA in size and shows 31% overall sequence identity with AsCpf1. **D.** Secondary structure of crRNA, pseudoknot (*blue*) with base pairing indicated using Leontis-Westhoff nomenclature and a dinucleotide GG (*grey*). The box indicates variations to the loop region of the pseudoknot that have been validated in *E. coli*. Numbering indicates the nucleotide position relative to the transcription start site.

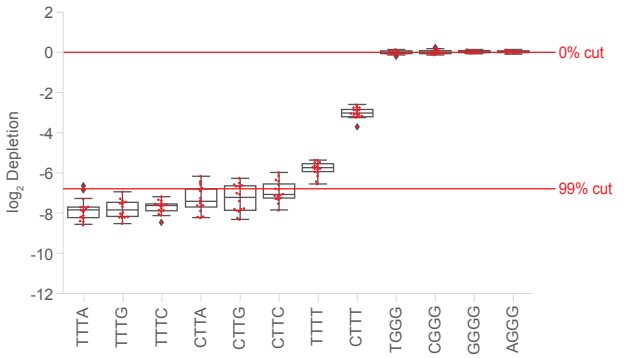


**Figure 2.** Cleavage efficiency of MAD7 across YTTN and NGGG PAMs. Data represents the average depletion observed for 6 different guides for each PAM, with each dot representing an independent measurement. MAD7-expressing *E. coli* (MAD7$^{+}$) was transformed with a library of self-targeting plasmids with a NNNN-Target-NN sequence. Completeness of cut was determined by amplifying the target sequence and comparing the depletion of normalized read counts in MAD7+ cells to read counts observed in MAD7- cells. Notably, some loss in activity was observed with a T at the 3′ terminus of the PAM.

## Endonuclease activity of MAD7

To test the endonuclease activity of MAD7, we designed gRNA for 96 different loci across the *E. coli* genome and tested for functional cutting using a multiplex depletion assay. We observed that expression of MAD7 in standard microbial expression vectors in combination with gRNA resulted in depletion in 74% of the transformed strains (i.e., 71 of 96 strains targeted at different loci were depleted in the mixed pool, as would be expected in the case of active MAD7 cutting).

## Precision editing using MAD7

To assess the utility of MAD7 in precision editing via homology-directed repair (HDR), we co-transfected MAD7 in microbial expression plasmids with both gRNA and donor-DNA plasmids as previously described (8). Editing precision and efficiency in *E. coli* and *S. cerevisiae* were assessed by targeted sequencing. Robust editing was observed, with the majority of designs resulting in very high editing efficiencies in *E. coli* (Table 1).

## Conclusion

We have identified a novel RNA-guided nuclease, MAD7, that shows low overall homology to AsCpf1. MAD7 targets YTTN PAM sites, which occur at regular frequency within *E. coli* and *S. cerevisiae* genomes. We have demonstrated the targeted nuclease activity of MAD7 in gRNA depletion assays and have shown precision editing via HDR in both *E. coli* and *S. cerevisiae*. This work demonstrates the utility of MAD7 for engineering bacterial and eukaryotic genomes and its potential to impact academic and commercial research in the energy, healthcare, and agricultural sectors.

## References


1. Makarova *et al.*, 2015. *Nature Reviews Microbiology.* **13**:722-736.
2. Lander, E. 2016. *Cell.* **164**:18-28.
3. Jinek, M. Chylinksi, K., Fonfara, I., Hauer, M., Doudna, J., and Charpentier, E. 2012. *Science.* **337**:816-821.
4. Cong, L., Ran, F., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P., Wu, X., Jiang, W., Marraggini, L., Zhang, F. 2013. *Science.* **339**:819-823.
5. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. 2013. *Nature Biotechnology.* **31**: 233–239.
6. Zetsche, B. *et al.* 2015. *Cell.* **163**:759-771.
7. Zetsche, B. *et al.* 2017. *BioRxiv.* https://doi.org/10.1101/134015
8. Garst,A., Bassalo, M., Pines, G., Lynch, S., Halweg-Edward, A., Liu, R., Liang, L. Wang, Z., Zeitoun, R., Alexander, W. and Gill, R.T. 2017. *Nature Biotechnology.* **35**:48-55.


| Locus | PAM | Cell | Homology arm length, bases | Edited bases | Colonies tested | Edited colonies | Edit efficiency |
|---|---|---|---|---|---|---|---|
| alaC | TTTC | *E. coli* MG1655 | 132 | 6 | 13 | 7 | 54% |
| cyaA | TTTC | *E. coli* MG1655 | 132 | 5 | 2 | 2 | 100% |
| fucK | TTTA | *E. coli* MG1655 | 132 | 7 | 8 | 8 | 100% |
| helD | TTTA | *E. coli* MG1655 | 132 | 5 | 7 | 7 | 100% |
| lolD | TTTC | *E. coli* MG1655 | 132 | 8 | 8 | 8 | 100% |
| lpxM | TTTC | *E. coli* MG1655 | 132 | 4 | 8 | 8 | 100% |
| narQ | TTTC | *E. coli* MG1655 | 132 | 7 | 8 | 8 | 100% |
| phoB | TTTG | *E. coli* MG1655 | 132 | 5 | 8 | 8 | 100% |
| rhsB | TTTC | *E. coli* MG1655 | 132 | 5 | 5 | 5 | 100% |
| ruvA | TTTA | *E. coli* MG1655 | 132 | 5 | 8 | 8 | 100% |
| waaC | TTTA | *E. coli* MG1655 | 132 | 6 | 8 | 0 | 0% |
| xylF | TTTC | *E. coli* MG1655 | 132 | 6 | 8 | 7 | 88% |
| | | | | | | | |
| CAN1 | TTTC | *S. cerevisiae* S288c | 134 | 6 | 82 | 54 | 66% |

**Table 1.** Edits in *E. coli* loci were selected to introduce non-synonymous changes to the genome. Edit status was determined by whole-genome shotgun sequencing. Edits in *S. cerevisiae* S288C were the introduction of an EcoRI recognition sequence into the genome. Edit status was determined by amplifying the locus by PCR and cutting with EcoRI followed by gel electrophoresis.

INSCRIPTA™