MAD7: a novel RNA-directed nuclease for precision gene editing

Andrew Garst, Eileen Spindler, Clint Davis, Chris Abraham, Charlie Johnson, Miles W. Andrea Edwards, Josh Shorenstein, Craig Struble, Paul Hardenbol, and Ryan T. Gill

Inscripta, Inc. 5500 Central Ave #220 Boulder, CO 80301

Abstract

RNA-guided DNA nuclease systems are widely used in genome editing. Low diversity of currently available nucleases in terms of guide RNA characteristics, nuclease size, and PAM target sequences, as well as commercial restrictions on their use, highlight the need for novel systems. Inscripta has built the capability to identify and generate RNA-guided DNA endonucleases. Here we describe MAD7, a system that targets TTTT PAMs and shows both cutting and editing activity in E. coli and S. cerevisiae.

MAD7 Structure Relative to AsCpf1

Figure 1. Overlay of MAD7 inferred structure (green) on the crystal structure of the related ortholog AsCpf1 (gray) in complex with guide RNA (pseudoknot) and DNA (red). A. Expanded view of the PAM recognition domain structure showing hydrogen bonds with DNA (indicated by dashed lines). B. Comparison of PAM recognition domain amino acid (AA) residues of MAD7 and 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.

Genome Coverage of MAD7 and Other Nucleases

Figure 2. Percent of E. coli and S. cerevisiae genomes within 20 or 50 bases of TTTC, TTTN, NGG, and TTN loci. MAD7 covers 90% of both genomes within 50 bases of TTTC. This close PAM proximity ensures that efficient editing should be possible at most genomic targets.

MAD7 Cuts YTTN

Figure 3. 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 HNHN-Target-NH domain. 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.

MAD7 Cuts Diverse Loci in E. coli

Figure 4. Cleavage across 96 loci in E. coli. 71/96 loci show complete depletion, 5/96 loci were not present in either transformation, 3/96 loci were not cut and therefore overrepresented in the resulting sequencing library.

Editing Designs

Figure 5. Examples of genome target, guide RNA spacer, and homology arm sequences for editing designs for loci in E. coli and S. cerevisiae. PAM sequences are highlighted in yellow, and bases highlighted. Edits in S. cerevisiae were detected by amplifying the CAN1 locus from individual colonies and detecting the presence or absence of the edited sequence by sensitivity to digestion with EcoRI.

MAD7 Induced Editing Efficiency

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 were detected by amplifying the CAN1 locus from individual colonies and detecting the presence or absence of the edited sequence by sensitivity to digestion with EcoRI.

Figure 6. Examples of edits. IGV screenshots showing alignments in the ruvA locus with edited bases highlighted. Edits in E. coli were detected by extracting genomic DNA from individual colonies, generating sequencing libraries with Nextera library prep kits and sequencing by NGS. Edits in S. cerevisiae were detected by amplifying the CAN1 locus from individual colonies and detecting the presence or absence of the edited sequence by sensitivity to digestion with EcoRI.

Conclusions

- We have identified MAD7, a novel gene-editing nuclease
- MAD7 is most similar to, but has low overall homology with, AsCpf1
- Genome coverage can be >90% with MAD7
- MAD7 cuts TTTT PAM sequences in 74% of loci tested
- MAD7 shows robust editing in both E. coli and S. cerevisiae