

Characterizing Cas9 Protospacer-Adjacent Motifs with High-Throughput Sequencing of Library Depletion Experiments

Jonathan L. Braff,^{1,4} Stephanie J. Young,^{1,2,3,4} Kevin M. Esvelt,¹ and George M. Church^{1,2,5}

¹Wyss Institute for Biologically Inspired Engineering, Harvard Medical School, Boston, Massachusetts 02115;

²Department of Genetics, Harvard Medical School, New Research Building, Boston, Massachusetts 02115;

³Program in Medical Engineering and Medical Physics, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

This protocol outlines a general approach for characterizing the protospacer-adjacent motifs (PAMs) of Cas9 orthologs. It uses a three-plasmid system: One plasmid carries Cas9 and its tracrRNA, a second targeting vector contains the spacer and repeat, and the third plasmid encodes the targeted sequence (as the protospacer) with varying PAM sequences. It leverages the Cas9 nuclease activity to cleave and destroy plasmids that bear a compatible PAM. The level of depletion of a library of targeted plasmids after Cas9-mediated selection can then be assessed by deep sequencing to reveal candidate PAMs for downstream validation.



MATERIALS

It is essential that you consult the appropriate Material Safety Data Sheets and your institution's Environmental Health and Safety Office for proper handling of equipment and hazardous materials used in this protocol.

Reagents

Agarose gel (1%) and reagents for agarose gel electrophoresis

Antibiotics (three compatible selection agents to be used for cloning and plasmid maintenance)

Choose plasmid backbones (see below) that contain the appropriate genetic resistance markers for each selection agent.

Competent *Escherichia coli* and reagents for transformation

Use chemically competent cells for CRISPR system expression and targeting vector cloning (Step 3). For PAM library construction (Step 5) and subsequent steps, any nuclease-deficient, high-efficiency cloning strain should work. We use NEB Turbo Electrocompetent E. coli (New England BioLabs C2986K).

DNA for plasmid construction

CRISPR system DNA (Cas9 protein of interest, medium-strength proC constitutive promoter, and cognate tracrRNA with native promoter and terminator)

Codon-optimize the Cas9 protein sequences, if needed, using JCAT (Java codon adaptation tool; www.jcat.de) (Grote et al. 2005). The Cas9 gene and trans-activating CRISPR RNA (tracrRNA) DNAs can be obtained via polymerase chain reaction (PCR) amplification of genomic DNA from the organism carrying the Cas9 system of interest, or synthesized by a gene synthesis service, such as gBlocks (IDT). Example DNA sequences are available in Esvelt et al. 2013. See Step 1 and Troubleshooting.

⁴These authors contributed equally to this work.

⁵Correspondence: gchurch@genetics.med.harvard.edu

Targeting vector DNA (strong J23100 promoter, unique 20-nt spacer sequence, and repeat sequence for the CRISPR system of interest)

Select spacer sequences that will be functional in Escherichia coli (for two example sequences, see Esvelt et al. 2013). At least two versions of the targeting construct are required to implement proper controls; see Step 2.

Glycerol (10% v/v in ddH₂O), sterile

Isothermal assembly mix

Use Gibson Assembly Master Mix (New England BioLabs E2611S) or prepare isothermal assembly mix in the laboratory as described by Gibson et al. (2009).

LB liquid medium

In addition, prepare plates containing LB medium solidified with agar.

PCR mix

Use a high-fidelity polymerase with proofreading activity such as KAPA HiFi HotStart ReadyMix PCR Kit (Kapa Biosystems) or Phusion polymerase (New England BioLabs). Use nuclease-free water for PCR.

Plasmid backbones for construction of three-plasmid system

Each vector type (the CRISPR system expression vector, the targeting vectors, and the PAM libraries) will need to coexist in a single cell with two other types of plasmids, so select plasmid backbones with appropriately compatible origins and selection markers. For example, the three plasmids can be p15a origin with chloramphenicol resistance, colE1 origin with kanamycin resistance, and cloD13 origin with spectinomycin resistance. Plasmids can be obtained from Addgene.

Reagents for construction of protospacer-adjacent motif (PAM) libraries (see Step 4)

Sequencing reagents (Illumina MiSeq)

Equipment

Equipment for agarose gel electrophoresis

Incubator at 37°C

Microcentrifuge

PCR tubes

Sequencing platform (Illumina MiSeq)

Shaking incubator at 37°C

Spin columns for PCR purification, plasmid DNA preparation, and plasmid mini-prep (QIAGEN)

Thermal cycler

METHOD

Constructing Plasmids

1. Using isothermal assembly (Gibson et al. 2009), create the CRISPR system expression vector containing the Cas9 protein of interest preceded by a medium-strength proC constitutive promoter, immediately followed by its cognate tracrRNA with its native promoter and terminator.

See Troubleshooting.

2. Using isothermal assembly, create at least two versions of the targeting vector, each containing a strong J23100 promoter followed by a unique 20-nt spacer sequence immediately adjacent to the repeat sequence for the CRISPR system of interest.

The two versions of the targeting vector will not need to coexist in a cell together, so they can and should contain identical plasmid backbones. This design controls for any biases in PAM library construction and selection by allowing the comparison between library depletion in a properly targeted (i.e., spacer matches the protospacer) experiment versus library depletion in a nontargeted (i.e., spacer does not match the protospacer) experiment.

3. Transform each of the three assemblies into a chemically competent *E. coli* cloning strain. Grow cells overnight on LB agar plates with the appropriate antibiotic for selection. Verify the sequences of the selected clones and perform a plasmid DNA preparation of correct clones.
4. Construct two PAM libraries (one for each of the targeting plasmids), each containing a protospacer that matches one of the spacers from Step 2, followed by an 8-bp degenerate region immediately 3' of the spacer where the PAM site would typically be located.

Standard PCR and isothermal assembly methods can be used; for example, we amplify a backbone vector such as pZE21 (ExpressSys) using primers encoding one of the two protospacer sequences followed by eight random bases. We typically split the vector into two pieces, and re-assemble the pieces (now with the added protospacer and PAM sequences) in isothermal assembly.

5. Transform the libraries into a high-efficiency cloning strain. Grow the cells overnight in at least 500 mL of LB liquid medium with the appropriate antibiotic for selection, and perform a plasmid DNA preparation from each library.

When cloning libraries, it is important to plate a few different dilutions of the cells to accurately quantitate the number of transformants and ensure adequate coverage of the library. Aim to have $\sim 1 \times 10^8$ transformants. Simple cloning optimizations and/or more than one transformation of each library may be needed to achieve this library size.

It may be desirable to prepare the plasmid at a slightly larger scale, as libraries can be reused to characterize multiple orthologs.

6. Prepare a total of four recipient strains (each containing a targeting plasmid in the presence of the CRISPR expression plasmid) for pairing with the PAM library (Step 7) as follows.
 - i. Cotransform the CRISPR system expression plasmid and one targeting plasmid into a nuclease-deficient, high-efficiency cloning strain.
 - ii. Repeat Step 6.i using the second targeting plasmid.
 - iii. Grow the cells from each transformation overnight on LB agar plates containing both antibiotics for selection (i.e., one for each plasmid resistance marker).
 - iv. Perform two cell preparations from each strain by pelleting 50 mL of a monoclonal liquid culture in late exponential growth, washing the cells three times with 50 mL of 10% glycerol, and then resuspending the cells in 50 μ L of ddH₂O (double-distilled water).
7. Transform each sample from Step 6 with 200 ng of PAM library DNA, such that each of the targeting strains is paired with each library, for a total of four transformations. Recover the cells for 2 h at 37°C.
8. Dilute the cells 100 \times in LB medium containing all three appropriate antibiotics for selection (i.e., one for each plasmid resistance marker) and grow the cells overnight (12–18 h). In addition, plate several dilutions of the recovery culture from Step 7 to estimate posttransformation library size.

More than 1×10^7 clones corresponds to complete coverage of 65,536 (4^8) PAM sequences.

Performing High-Throughput Sequencing and Analysis

9. Select library clones after 12–18 h of antibiotic selection. Harvest DNA from each of the selections by plasmid mini-prep.
10. Prepare an Illumina sequencing library via PCR amplification of the degenerate region of the PAM library, and sequence using MiSeq.

It is important to consult the sequencing core or Illumina directly regarding current best practice for amplicon sequencing to ensure the amplicon library has sufficient sequence diversity, appropriate adapters, and proper indexing to demultiplex the various selections after sequencing. Paired-end 30-bp reads with custom sequencing primers work well. If custom primers will be used, make sure the melting temperature is high enough for the sequencing platform.
11. Preprocess the sequencing results by merging the paired-end reads and filtering out any pairs without perfect alignment, the protospacer sequences, and the plasmid backbones. Trim the filtered reads, leaving only the degenerate PAM regions.

12. Calculate fold depletion for each candidate PAM by comparing the position weight matrix (PWM) of the spacer–protospacer matching pool with that of the mismatching pool while filtering for the candidate PAM sequence.

A simple Python script called `patternProp3.py` is helpful for this step (see Esvelt et al. 2013; supplemental). A list of candidate PAMs can be generated exhaustively or by inspection. To help generate the list by inspection, `patternProp3.py` outputs not only the total fold depletion for a candidate PAM but also the specific depletion for each base at each position for all reads matching that candidate. One can first look at the fold depletions of the null filter (i.e., NNNNNNNN), and then identify the most depleted bases and use those as the new filter. It is important to be aware that as the filter becomes more complex, there are fewer reads, so the differences may become statistically insignificant.

See Troubleshooting.

(Optional) Validating PAM Sequences

13. To validate a putative PAM identified in the experiment, redesign the library plasmid to encode the PAM and measure depletion.

Alternatively, validation can be completed with a GFP (green fluorescent protein) reporter plasmid, in which the protospacer and PAM are in the 5' UTR (untranslated region) of GFP (which has a strong pR promoter and T7 g10 ribosome binding site), as described by Esvelt et al. (2013).

TROUBLESHOOTING

Problem (Step 1): The experiment requires bacterial expression of novel orthologs, yet performing a PAM-finding experiment requires a priori knowledge of the Cas9 ortholog of interest, specifically the corresponding sequence and directionality of the tracrRNA and crRNA.

Solution: Use tools such as CRISPRfinder (Grissa et al. 2007). Include the native promoter and terminator of the tracrRNA, and place the cassette downstream from Cas9. Minimize the secondary structure of the spacer and repeat.

Problem (Step 12): Depletion biases are present in the library experiment.

Solution: There can be apparent minimum and maximum fold depletion values (Esvelt et al. 2013). Some are low-frequency mutations leading to inactivated Cas9 or target escape. Analysis should focus on moderately depleted sequences. Changing the recovery time may vary the stringency of the selection. In addition, the library selection can be performed in the context of the final activity of interest (i.e., the protocol described here would be for cleavage activity).

ACKNOWLEDGMENTS

This work was supported by U.S. Department of Energy grant DE-FG02-02ER63445 (to G.M.C.) and the Wyss Institute for Biologically Inspired Engineering.

REFERENCES

- Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* 10: 1116–1121.
- Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6: 343–345.
- Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRfinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35: W52–W57.
- Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D. 2005. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* 33: W526–W531.