Determining Protospacer Adjacent Motif Preferences of Industrially Relevant Clostridial Type I-B CRISPR-Cas Systems

Grant A. Rybnicky[1]* (grantrybnicky2023@u.northwestern.edu), Michael Köpke[2], and **Michael C. Jewett**[3]

[1]Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL; [2]LanzaTech, Skokie, IL; [3]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL

http://jewettlab.northwestern.edu/; https://www.lanzatech.com/

Project Goals: We are addressing the challenge of designing, building, and optimizing biosynthetic pathways in cells in an interdisciplinary venture that establishes the clostridia Foundry for Biosystems Design (cBioFAB). Working both *in vitro* and *in vivo*, the goal is to interweave and advance state-of-the-art computational modeling, genome editing, omics measurements, systems-biology analyses, and cell-free technologies to expand the set of platform organisms that meet DOE bioenergy goals. cBioFAB will (i) reconceive how we engineer complex biological systems by linking pathway design, prospecting, validation, and production in an integrated framework, (ii) enable systems-level analysis of the David T. Jones collection, one of the largest collections of clostridia strains in the world, to uncover novel metabolic pathways, regulatory networks, and genome editing machinery, and (iii) open new paths for synthesis of next-generation biofuels and bioproducts from lignocellulosic biomass.

The recent discovery and in-depth characterization of CRISPR-Cas9 has led to a variety of technologies, including genome editing, genome modification, nucleic acid sensing, and next generation antimicrobials. Although CRISPR-Cas9 is a powerful tool, it is representative of only a small fraction all of CRISPR-Cas systems and is toxic when heterologously expressed in many bacteria. Fortunately, about half of all bacteria that have been sequenced encode at least one CRISPR-Cas system in their genome which provides an alternative to CRISPR-Cas9 for genome manipulation. The most prevalent subtype of CRISPR-Cas system is type I-B, which comprises 20% of all naturally occurring CRISPR-Cas systems. This subtype is predominant in Firmicutes, including solventogenic Clostridia which have immediate relevance to industrial bioprocessing. Endogenous type I-B CRISPR-Cas systems have been used to successfully edit the genomes of *Clostridium pasteurianum*, *Clostridium tyrobutyricum*, and *Clostridium thermocellum*, but expansion to other organisms is hindered by the unique protospacer adjacent motif (PAM) sequence of each CRISPR-Cas system required to target a DNA sequence. Past efforts address this by generating nucleotide alignments between CRISPR array spacers and sequences within a variable database, manually curating alignments to hypothesize a few potential PAMs, and then testing these hypotheses individually by plasmid interference assays. Here we present a standardized *in silico* pipeline to predict PAM sequences for a given CRISPR-Cas system from annotated CRISPR spacers and use crude cell-free extracts to characterize all possible 5-

nucleotide PAM sequences for the *Clostridium autoethanogenum* type I-B CRISPR system. The *in silico* PAM prediction pipeline is able to recapitulate experimentally determined PAMs from a variety of CRISPR-Cas systems. The prediction strength and accuracy is heavily dependent on the number of nucleotide alignments generated in the first step which makes it difficult to predict PAMs for CRISPR-Cas systems with few spacer sequences or within organisms that have few sequenced relatives and mobile genetic elements. To test these predictions, we used our recently developed *C. autoethanogenum* cell-free system to experimentally screen PAM sequences by depletion of variants from a 5-nucleotide randomized PAM library. The depletion assay and subsequent validation indicate a conserved CCW PAM 5' to the protospacer, different from what previous sequence alignment-based methods have predicted. We anticipate that this new pipeline and experimental strategy will aid in standardizing PAM prediction and in characterizing PAM preferences in new CRISPR-Cas systems.

We acknowledge the Department of Energy grant DE-SC0018249 for funding of this project.