Determining the requirements for acquisition of new CRISPR immunity elements in *Escherichia coli*

Sarah [redacted][1] | Joseph Wade[2], Lauren Cooper[2], Anne Stringer[2], Carol Smith[2]

[1]Russell Sage College, Troy, NY [2]Wadsworth Center, NYS DOH, Albany, NY

NEW YORK STATE OF OPPORTUNITY. Department of Health | Wadsworth Center

What is CRISPR?

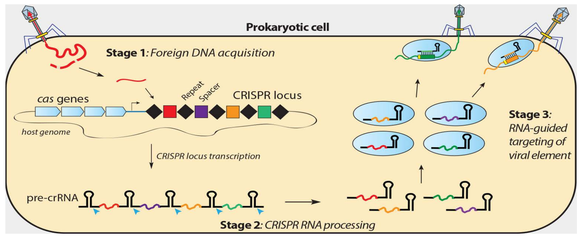

Figure 1. General scheme of CRISPR immunity. Source: The Doudna Lab

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is an adaptive immune system found in roughly 50% of bacteria and 90% of archaea. This system works through three stages: adaptation, biogenesis, and interference (Figure 1). Adaptation, the focus of this project, is when bacteria incorporate new immunity elements ("spacers") from invading DNA molecules. Biogenesis allows spacers to be transcribed into CRISPR RNAs (crRNAs). Interference occurs when a crRNA is bound by Cascade, a complex of CRISPR-associated (Cas) proteins. This complex searches for DNA that has sequence complementary to the crRNA. If an invading DNA molecule, such as a plasmid or bacteriophage genome, is targeted by Cascade, the foreign genetic material will be destroyed, rendering the invader incapable of causing harm.

Adaptation

Adaptation, sometimes called "spacer acquisition", occurs when the cell identifies foreign genetic material and obtains a new segment of DNA ("spacer") from the invader, to incorporate into its genome. There are two types of adaptation: naïve and primed. Naïve adaptation incorporates a spacer from a newly encountered invader whereas priming, the focus of this project, acquires another unique spacer from a previously encountered invader. Each type of adaptation requires different proteins to occur. Primed acquisition additionally requires an existing spacer against an already encountered invader. Adaptation ultimately results in the addition of one new spacer and one repeat to the CRISPR array. Many *cis*- and *trans*-acting factors likely affect spacer acquisition, but not all have been identified.

The Leader Sequence

The leader sequence is a region of DNA directly upstream from the CRISPR array that often separates the array and the *cas* genes. The leader sequence has been shown to play a critical role in adaptation.

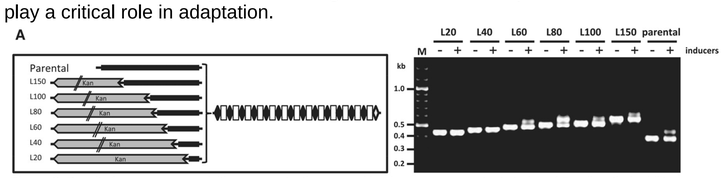

Figure 2. Results from spacer acquisition assays testing several leader sequence mutant strains of *E.coli* on their ability to acquire new spacers. Source: Yosef et. al (2012)

Researchers found that at least 60 bp upstream from the first repeat in the CRISPR array are necessary for spacer acquisition (Figure 2). However, the researchers deleted 20 bp chunks of the leader sequence, meaning that the precise number of bp necessary for spacer acquisition could lie anywhere in the 40-60 bp region.

Figure 3. Sequence alignment of several CRISPR leader sequences in different organisms.

There is also an area of the leader sequence that is conserved between the CRISPR systems of *E. coli* and several other species, which implies some importance to the proper function of the system (Figure 3). It is possible that this area of the genome could act as a binding site for a DNA binding protein to help in spacer acquisition.

Project Overview

The goal of this project was to determine what areas of the leader sequence are necessary for primed spacer acquisition in *E.coli.* Mutant strains lacking the 40-50 bp (Δ40-50) and 50-60 bp (Δ50-60) regions upstream from the first repeat in the CRISPR I array of *E. coli*, as well as a strain with a dysfunctional conserved region (cons*) were created via FRUIT. The 40-50 and 50-60 bp region mutations were created through deletion mutations, whereas each base pair in the conserved region was mutated (GTTGGT → CAACCA) to disable it.

Experimental Organism

The strain of E.coli used for experimentation is known as CNC Δ*thyA* + pKD46. This strain has a functional CRISPR system with two CRISPR arrays, which are unable to perform primed adaptation due to the deletion of *cas3*. The strain also lacked *thyA* which allowed the strain to be used in FRUIT.

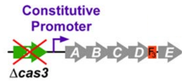

Figure 4. Scheme of CNC strain of *E.coli.*

Methods

- FRUIT (Flexible recombineering using the integration of *thyA*):
  - Scar-free method of introducing deletion and/or point mutations
  - Used to create leader sequence mutations (Δ40-50, Δ50-60, cons*) in the CRISPR I array of CNC strain
  - *thyA* was introduced into cells, that were unable to make thymine, at the desired site of mutation
  - SOEing PCR products were created containing the desired mutations
  - PCR products were transformed into cells, causing *thyA* to be lost by recombination
  - Product was a *thyA*-deficient cell containing the desired mutation
  - *thyA* was inserted back into its native spot on the bacterial genome
  - Strains confirmed via sequencing of PCR products

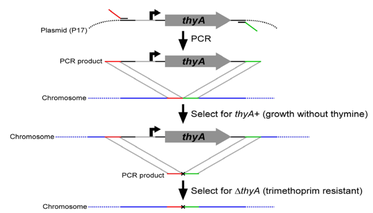

Figure 5. General mechanism of FRUIT. Source: Stringer A.M., et. al (2012)

- Priming Assay
  - Used to view primed adaptation in wild-type leader sequence mutant strains
  - A self-targeting plasmid was transduced into cells containing mutations
    - Provided a template for a new spacer
  - A plasmid containing *cas3* was transduced into cells
    - Allowed for priming acquisition to occur
  - Cells were grown for 5 hours and sampled at set times (T0, T1, T2, & T5)
  - Amplified CRISPR I & II arrays and viewed products on a 12% acrylamide gel
  - Negative result = unexpanded CRISPR array → 1 band
  - Positive result = expanded CRISPR array → 2 or more bands

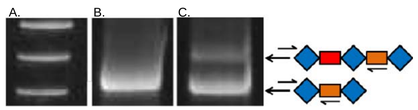

Figure 6. Priming assay results. Panel A shows a DNA ladder. Panel B shows a negative assay result, represented as a single band on an acrylamide gel. Panel C shows a positive result, represented by multiple bands on gel.

Results

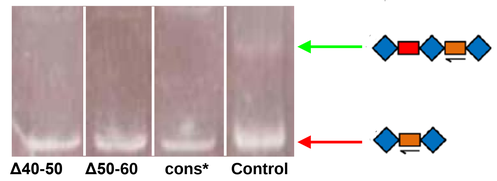

Figure 7. Negative priming assay results for all three CRISPR I leader sequence mutants. Wild type CNC was used as a control. Red Arrow = unexpanded array Green Arrow = expanded array

Conclusion

All three CNC CRISPR I leader sequence mutants were unable to acquire new spacers via primed adaptation after 5 hours of growth. However, all mutants were able to acquire new spacers in the CRISPR II array, which had a wild-type leader sequence. Although this study does not identify the precise base pairs in the leader sequence necessary for adaptation, base pairs in both the 40-50 bp and 50-60 bp regions are important. Many of these positions are conserved between *E. coli* and related species. It is likely that the conserved base pairs are the most important elements. The conserved positions could act as binding sites for one or more DNA binding proteins. However, these results only showed what elements were necessary for priming spacer acquisition, not why. These results should be further explored to try and determine the actual purpose of each element and why it is important to acquisition of new CRISPR immunity elements.

References

Stringer A.M., Singh N., Petrone B.L., Amarasinghe J.J., Mantis N.J., Wade J.T. (2012). FRUIT, a Scar-Free System for Targeted Chromosomal Mutagenesis, Epitope Tagging, and Promoter Replacement in Escherichia coli and Salmonella enterica. PLoS ONE 7(9). doi:10.1371/journal.pone.0044841

The Doudna Lab. Retrieved from: http://rna.berkeley.edu/crispr.html

Yosef I., Goren M.G., Qimron U. (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Research, 5569–5576. doi:10.1093/nar/gks216

Acknowledgements

I would like to thank the Wadsworth Center REU program and REU director Matthew Kohn for the opportunity to pursue this research. This research was supported by NSF REU site award #DBI1359413. I thank all members of the Wade lab for their continued support throughout my REU experience. I would like to further acknowledge Dr. Joseph Wade for devising this project and helping each step of the way.