

Staggered vs. Blunt Cutting of Cas9 in the Context of Targeted Sequencing

Rachel A. Case¹, Chloe H. Ziemelis¹, Monica J. Holmes¹, Gregory A. Farnum¹, Brian D. Athey^{1,2}
¹Department of Computational Medicine and Bioinformatics, ²Department of Psychiatry, University of Michigan, Ann Arbor.

Abstract

CYP2D6 plays a critical role in drug metabolism, affecting about 25% of medications that are processed in the body. As such, it is taken into important consideration when genetic profiles are being evaluated for medications. However, CYP2D6 has many variations, which are hard to detect with the current technology. This project aims to create verifiable assays that provide the necessary information about CYP2D6 through Oxford Nanopore Sequencing and targeted sequencing using Cas9.

Originally, Cas9 was regarded as a blunting enzyme (an enzyme that cuts DNA with no base pair overhangs) and used to cut DNA at certain sites in the genome, allowing for targeted regions to be modified with sequencing adapters and sequenced. During our initial studies, it was noticed that the number of reads are unequally distributed from the initial cut site, leading to the idea that Cas9 does not actually cut blunt ends and instead leaves the DNA with various overhang structures.

This experiment is meant to verify the presence and ligation efficiency of adaptors attached to DNA fragments using gel electrophoresis, with focus on detecting 1, 2, 3, and 4 base pair overhangs on both the 5' and 3' ends after Cas9 cutting. By using duplexed adaptors with various known overhangs and ligating them to Cas9 cut DNA and blunt DNA, it should be possible to determine the resulting pattern of overhang structures. A combination of Oxford Nanopore Sequencing and Gel Electrophoresis will be used to determine the results. The expected result is showing Cas9's cutting mode pattern and the presence of specific overhang structures. In conclusion, having a set assay will provide a better understanding of Cas9 and more accuracy in clinical lab results, leading to improved drug metabolism predictions.

Introduction

Cas9, the enzyme being studied in this lab, is a part of a newly harnessed technique used to cut DNA called CRISPR. Discovered by Jennifer Doudna and Emmanuelle Charpentier in 2013, it originates from a bacterial defense system to invading viruses. By storing the viral DNA sequence in the CRISPR locus and using the virus's protospacer adjacent motif (PAM) site (essentially a unique, identifying DNA sequence), the CRISPR locus can translate CRISPR RNAs, also known as crRNAs, to target the invading virus by duplexing with tracerRNAs and guiding Cas9, an endonuclease, to cut the dsDNA into two pieces. (Bekaert, 2022)

Originally, Cas9 was believed to cut blunt ends on DNA. However, when using an Oxford Nanopore Sequencer to profile DNA (a tool that passes molecules through nanopores and measures the change in electrical current to result in a combination of reads) a bias was observed in the number of reads. The reads, which are sequences of base pairs, were found to be unequally distributed from the initial cut site. Because of this bias, we believe that Cas9 is forming overhangs on the cut DNA, making one strand longer than the other.

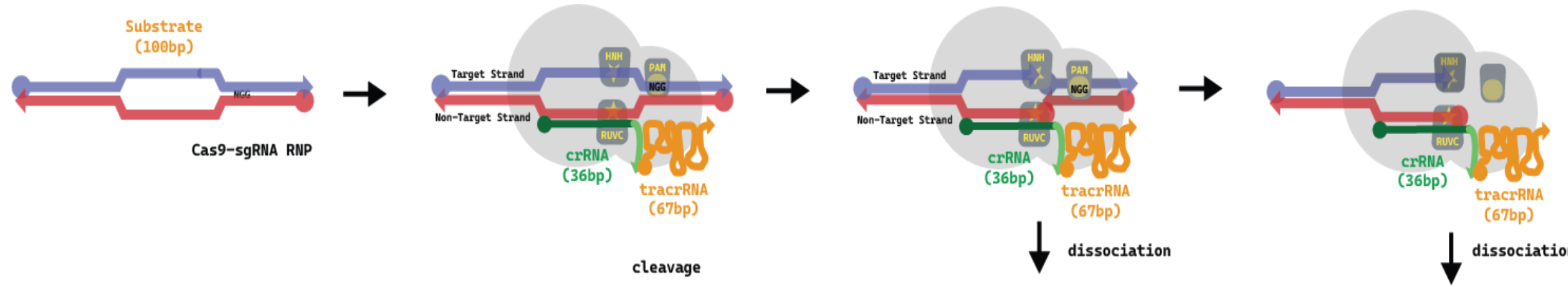


Figure 1: Cas9 Cut Substrate. Cas9 cuts a substrate three base pairs upstream at the PAM site (an NGG sequence), when the Cas9 guide RNA matches the twenty base pair target site directly upstream (5').

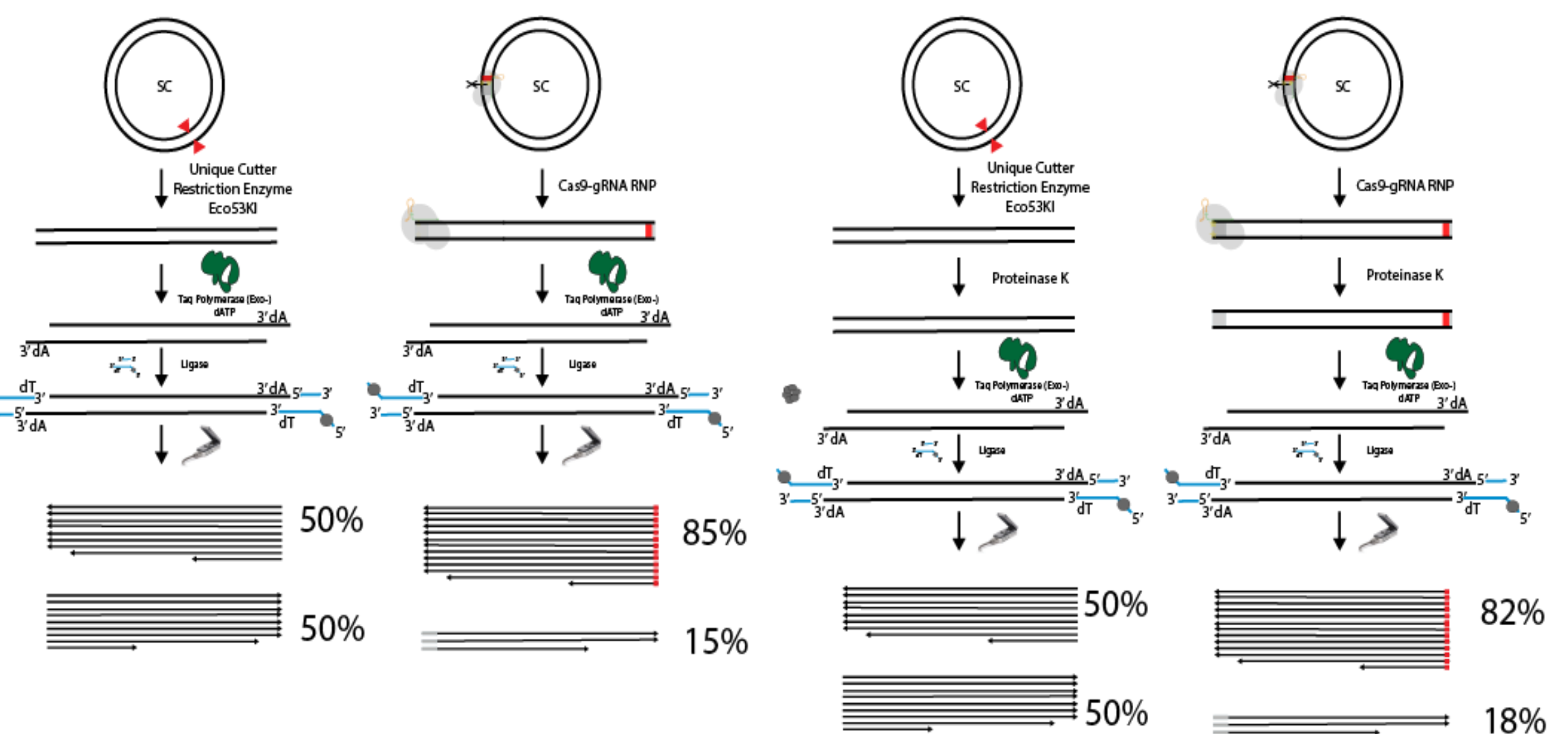


Figure 2: Bias Found in Cas9 Cutting. Previous experiments demonstrating a bias in reads were thought to have Cas9 blocking the ends of DNA. However, several experiments where Proteinase K was used to remove Cas9 prior to attaching the ligation adaptors and the bias was still observed.

Methods

There is evidence to suggest that Cas9 may have significant 3' - 5' exonuclease activity post-cleavage. The purpose of this experiment was to show 1, 2, 3, and 4 base pair overhangs on Cas9 cut DNA. To do this experiment, a total of thirty-two single stranded oligonucleotides (short strands of DNA) with known base pair sequences and various overhang structures were ordered and duplexed, meaning their top strands were ligated (lined up and attached) to their corresponding bottom strands. The sixteen double stranded oligonucleotides, or adaptors, are referred to as the target.

Prediction:

By using adaptors with various 1, 2, 3, and 4 base pair overhangs and ligating them to Cas9 cut DNA, a difference in peak size of the distal end of DNA will be observed on the Tape Station, confirming that Cas9 does not cut bluntly.

Before determining that Cas9 is forming overhangs on cut DNA, several control experiments were conducted to ensure the lower limit of detection and ligation efficiency using gel electrophoresis and the TapeStation (a tool that uses fluorescence to detect DNA migration). This included running all duplexed adaptors on a gel, ligating the adaptors and their complementary adaptors to each other, and ligating adaptor mixes to their complementary adaptors with and without ligase.

After the initial control experiments, positive control and negative control experiments were designed to verify the actual experiment. For both experiments, a 100 base pair duplexed strand of DNA (referred to as the substrate) was also ordered, with a cut site between 40 and 60 base pairs. Eight of the adaptors are complementary to the distal PAM cut site (at the 40 base pair mark) of the Cas9 cut exon, and the other eight adaptors are complementary to the proximal PAM cut site (at the 60 base pair mark). The overhangs on the adaptors represent possible overhang structures of the Cas9 cut DNA.

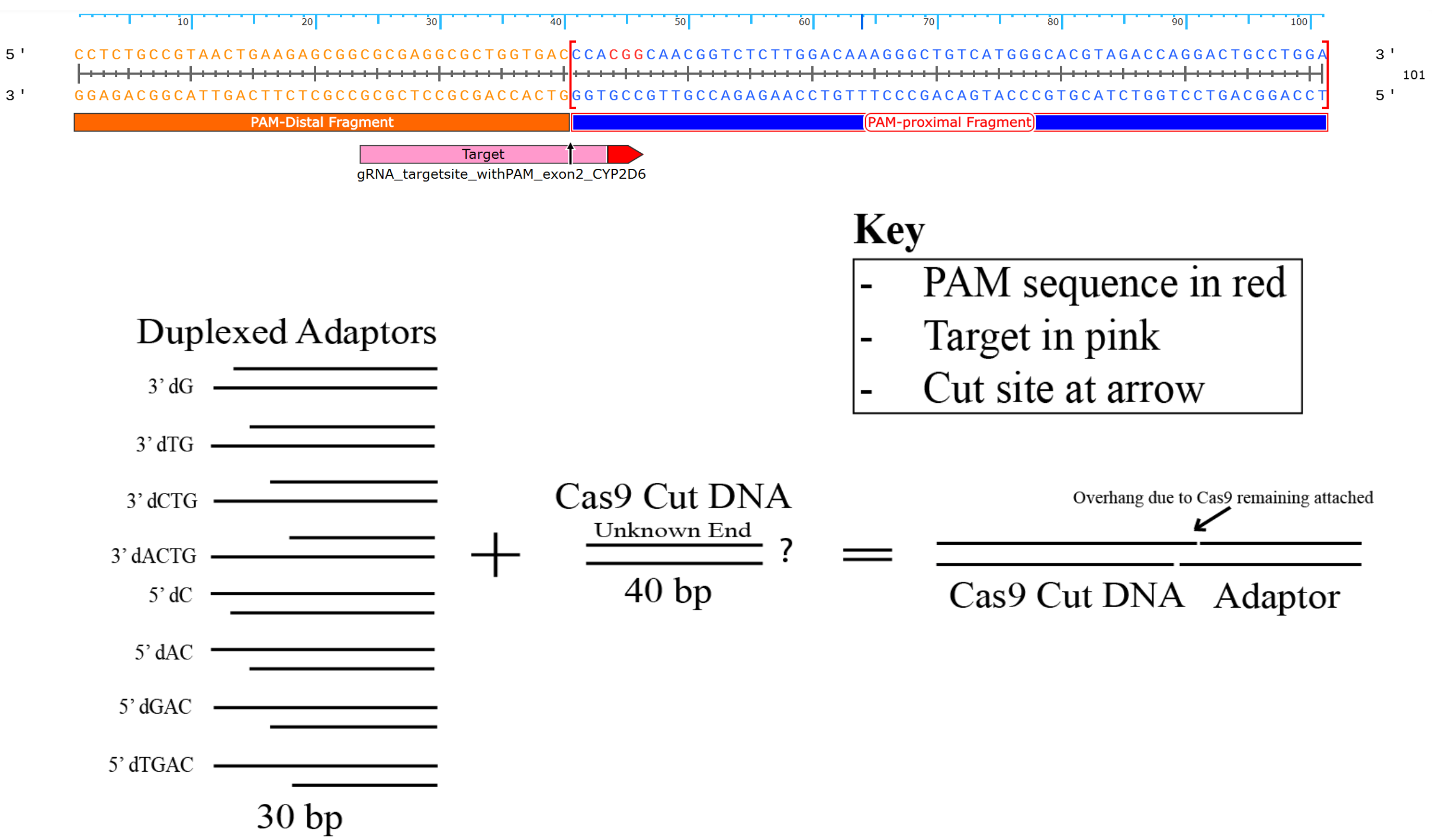


Figure 4: The Sequence of the Substrate and the Adaptors Ligated to the Substrate. The various overhang structures of the adaptors are possible overhang structures of Cas9 cut substrate.

For the positive control, the substrate consisted of a 1:1 ratio of blunt DNA and DNA with overhangs. The substrate was ligated to the complements and adaptors. Because there is a greater ligation efficiency on non-blunt ends, the expected outcome is a higher product peak at both the 70 and 90 base pair marks.

For the negative control, the substrate consisted entirely of blunt DNA. The expected result would show a much lower ligation efficiency at the 70 and 90 base pair marks than the positive control.

For the experimental design, the substrate was Cas9 cut DNA. If there are overhangs present on the PAM distal cut end, then there will be a greater ligation efficiency and higher product peak. This peak would theoretically be in between the peaks shown in the positive and negative controls.

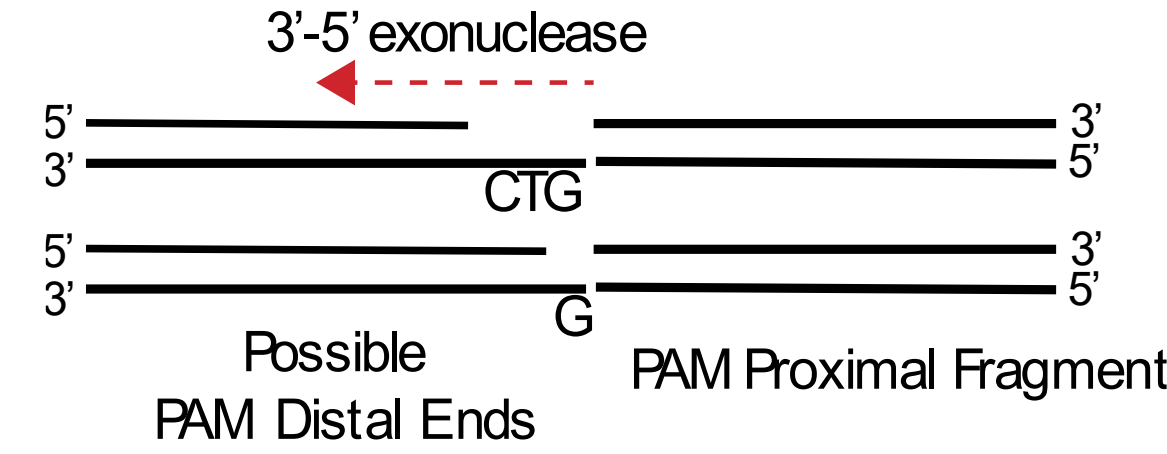


Figure 3: Possible Overhang Structures due to Cas9 Exonuclease Activity. Cas9 may have exonuclease activity because of the bias seen in the reads, despite having the Cas9 removed before sequencing.

Results

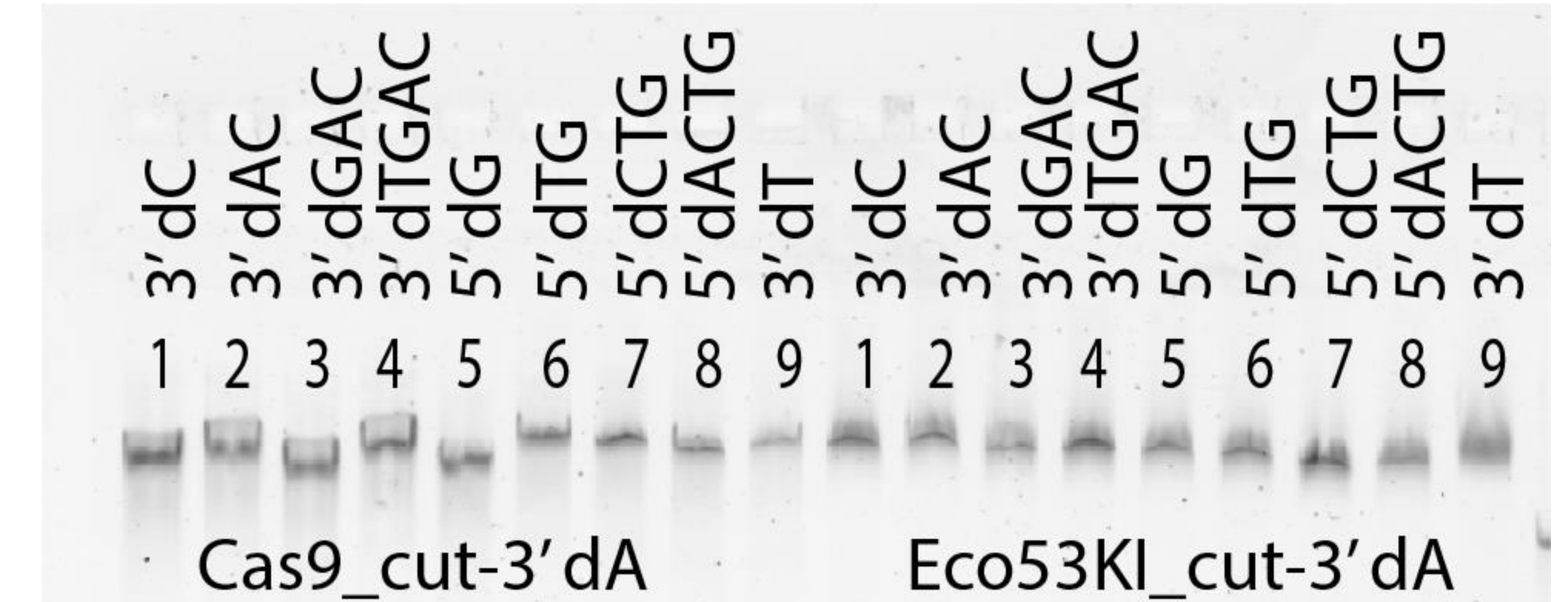


Figure 5: Gel of Cas9 Cut Substrate vs. Eco53KI Cut Substrate. The Cas9 cut substrate shows gel bands at various base pair lengths, unlike the Eco53 cut substrate, which is consistent in base pair length. This is evidence that Cas9 does not cut blunt, and may be proved using the TapeStation.

Conclusion & Discussion

The conclusion is that Cas9 does not produce blunt cuts, but instead creates overhangs which will be seen in the base pair shifts on the TapeStation. The TapeStation results will be conclusive, and the through controls show that the results are either negative or positive. If there are overhangs, they will be determined. In addition, the ligation efficiency may be calculated from the control experiments and compared, further showing the presence of overhangs. Because Cas9 is currently used as a blunt cutting enzyme, the future Cas9 reliant experiments will have to be altered to account for the overhang structure formation, thereby increasing their verifiability. With the new knowledge that Cas9 forms overhangs, the unexplainable presence of biases in sequencing reads may be understood, impacting clinical sequencing techniques and drug metabolism predictions.

To take this experiment further, Capillary Electrophoresis could be used to view the exact base pair overhang structures. The TapeStation, while very sensitive, does not have the ability to show 1, 2, 3, and 4 base pair changes. For this experiment's purpose, it worked well to show an overall general shift in the peak due to more ligation to overhang structures. However, it might be helpful to know which overhang structures are formed, and why Cas9 cuts the way it does.

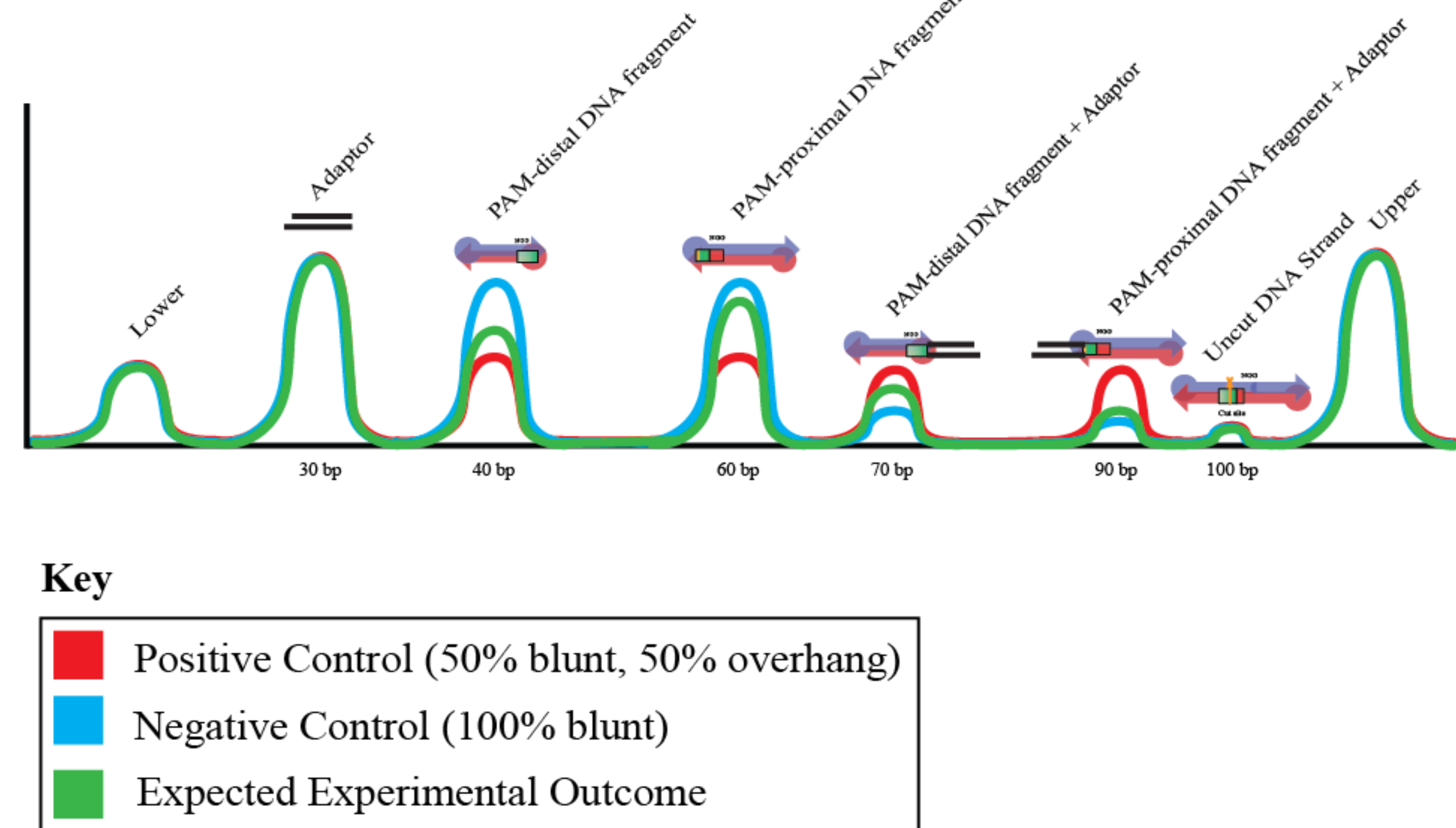


Figure 6: Expected Negative, Positive, and Experimental Results. The expected experimental result is in between the negative and positive controls. There is more ligation in the positive control due to the presence of overhangs, and there is very little ligation in the negative control due to the blunt ends. Because Cas9 cut DNA is theorized to have overhangs on the distal end, there is more ligation at the 70 base pair mark.

Acknowledgements

Supported by Phenomics Health Inc. We thank the Advanced Genomics Core at the University of Michigan for running TapeStation samples. Bekaert, B., Boel, A., Cosemans, G., De Witte, L., Menten, B., & Heindryckx, B. (2022). CRISPR/Cas gene editing in the human germline. *Seminars in Cell & Developmental Biology*, 131. <https://doi.org/10.1016/j.semedb.2022.03.012>