ROESING, MIRANDA R., M.S. Screening Highly-Specific Hairpin Guide RNAs for CRISPR-Cas Gene Editing. (2021)

Directed by Eric A. Josephs. 69 pp.

This thesis outlines the development of a screening method to improve the specificity of the CRISPR effector, Cas9, in gene editing applications. We accomplish this by identifying modifications to the RNA-cofactor of Cas9, its guide RNA (gRNA), that determines the DNA sequence recognized by the effector, where changes to the gRNA secondary structure are hypothesized to limit the potential for mutations at unintended sites. The identification of highly specific gRNA sequences can potentially lead to highly specific gene editing techniques that can be used to treat genetic diseases such as Tay-Sachs disease, cystic fibrosis, and sickle cell anemia without off-target mutations. Randomized libraries of gRNAs with extra nucleotides that can form secondary structures such as “hairpins” with the targeting segment of the gRNA (hairpin-gRNAs or hp-gRNAs) are screened with SpyCas9 in the presence of DNA with the intended target, and DNA containing “off-target” sequences that are similar but not an exact match for the target. The results from this screen can provide insights to the biophysical requirements that dictate target recognition and the potential for design rules for high-fidelity hp-gRNAs for any genetic target for use in effective gene therapeutics.

# SCREENING HIGHLY-SPECIFIC HAIRPIN GUIDE RNAS FOR CRISPR-CAS GENE EDITING

by

Miranda R. Roesing

A THESIS

Submitted to

the Faculty of The Graduate School at

The University of North Carolina at Greensboro

in Partial Fulfillment

of the Requirements for the Degree

Master of Science

Greensboro

2021

Approved by

Dr. Eric A. Josephs

Committee Chair

APPROVAL PAGE

This thesis written by Miranda R. Roesing has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair Dr. Eric A. Josephs

Committee Members Dr. Dennis LaJeunesse

Dr. Kristen Dellinger

Dr. Jianjun Wei

4/5/2021

Date of Acceptance by Committee

4/1/2021

Date of Final Oral Examination

ACKNOWLEDGEMENTS

I was very fortunate to receive a great deal of support and assistance during this journey.

I would first like to thank my advisor, Dr. Eric A. Josephs, who has been a consistently supporting and motivating mentor, and whose expertise has been invaluable. His ingenuity, assistance, and encouragement made this master's thesis possible. I am very grateful that I had the opportunity to be a part of Dr. Josephs' lab.

I would like to thank all of the faculty and staff of the Joint School of Nanoscience and Nanoengineering and the University of North Carolina Greensboro for their support in this academic endeavor. A special thanks to my committee members, Dr. Dennis LaJeunesse, Dr. Kristen Dellinger, and Dr. Jianjun Wei for their guidance and support throughout this process.

I would like to thank all of the post-doctoral fellows, undergraduates, and fellow graduate students who have helped me throughout this journey. A special thank you goes to the previous and current members of our lab; Parth Desai, Robert Glass, Tanjina Islam, Hillary Dimig, Tinku Supakar, Dr. Rammyani Bagchi, Dr. Rachel Tinker-Kulberg.

Finally, I would like to thank family and friends. This could not have been done without the support provided by my parents, my sister, my family here in Greensboro, and my friends near and far. Thank you for all of the happy distractions that have boosted morale.

This work was supported by the National Institutes of Health / National Institute of General Medical Sciences [1R35GM133483] and start-up funds from UNC Greensboro and the Joint School of Nanoscience and Nanoengineering (JSNN). The work was performed in part at JSNN, a member of the Southeastern Nanotechnology Infrastructure Corridor (SENIC) and National Nanotechnology Coordinated Infrastructure (NNCI), which is supported by the National Science Foundation [Grant ECCS-1542174].

TABLE OF CONTENTS

LIST OF TABLES ..... vii

LIST OF FIGURES ..... viii

CHAPTER I: INTRODUCTION ..... 1

Gene Editing Systems ..... 1

Basics of Prokaryotic Immunity ..... 2

Class II Type II CRISPR-Cas Systems in Gene Editing ..... 5

Methods in Gene Editing and Epigenetics ..... 8

Off-target Activity, Prevalence, and Prediction ..... 12

The SpyCas9 Conformational Checkpoint ..... 16

Manipulation of the SpyCas9 Effector to Increase Specificity ..... 18

Optimized Design of GRNAs to Enhance Specificity ..... 21

The Effect of GRNA Secondary Structure on SpyCas9 Specificity ..... 22

The Effect of GRNA 5' Hairpin Extensions on SpyCas9 Specificity ..... 23

Identification of Design Rules for Hairpin Extended GRNAs ..... 26

CHAPTER II: A PLATFORM FOR SCREENING OF SPYCAS9 ACTIVITY ..... 28

Aim ..... 28

Methods ..... 28

Results and Discussion ..... 34

CHAPTER III: SCREENING FOR OFF-TARGET ACTIVITY ..... 36

Aim ..... 36

Methods ..... 37

Results and Discussion ..... 44

CHAPTER IV: SCREENING OF HAIRPIN LIBRARIES ..... 50
Aim ..... 50
Methods ..... 50
Results and Discussion ..... 53
CHAPTER V: SUMMARY AND FUTURE WORK ..... 62
REFERENCES ..... 64

LIST OF TABLES

Table 1. Plasmid construct variants and their origin of replication, resistance markers, and key regulatory components ..... 31

Table 2. Strain generation via multiple transformations for on-target activity screening ..... 33

Table 3. Selective plating conditions for assessment of SpyCas9 activity ..... 34

Table 4. Generation of bacterial strains for SpyCas9 plasmid based and library based off-target activity screenings ..... 40

Table 5. Sequences of off-target sites and target sites introduced into pgRNA-EMX1 for screening of off-target activity of SpyCas9 ..... 42

Table 6. Generation of bacterial strains for hairpin library specificity screening ..... 52

Table 7. Isolated hairpin sequences from Sanger sequencing of hairpin library screening in non-selective conditions with tetraloop 5'-UUCG-3' ..... 54

Table 8. Isolated hairpin sequences from Sanger sequencing of hairpin library screening in selective conditions with tetraloop 5'-UUCG-3' ..... 55

Table 9. Hairpin extended guide RNAs from screening of hairpin library with 4 potential tetraloops ..... 58

LIST OF FIGURES

Figure 1. Basic structure of ZFs and TALEs ..... 2

Figure 2. Schematic of the CRISPR adaptive immunity mechanism ..... 4

Figure 3. Mechanism of DNA target recognition and scission by Cas9 ..... 6

Figure 4. Nomenclature of SpyCas9 gRNA modules ..... 7

Figure 5. Mechanisms of non-homologous end joining and homology-directed repair ..... 9

Figure 6. Mechanism of wild type SpyCas9 versus catalytically inactive SpyCas9 ..... 10

Figure 7. Overview of Cas9n base editing mechanism ..... 11

Figure 8. Mechanism of prime editor pegRNA complex ..... 12

Figure 9. Workflow of CIRCLE-seq ..... 14

Figure 10. Workflow of CHANGE-seq ..... 15

Figure 11. Schematic of RNA-induced structural activation of SpyCas9 for target cleavage ..... 16

Figure 12. Schematic of Sniper-Cas9 directed evolution ..... 19

Figure 13. Structure of the dual-dCas9:*Fok*I:DNA complex primed for scission ..... 20

Figure 14. Favored and disfavored nucleotides along highly active sgRNAs ..... 21

Figure 15. Basic 5’ extended hairpin structure attached to a sgRNA ..... 23

Figure 16. Mechanism of 5’ hairpin extension interference of dSpyCas9 binding to mismatched target ..... 24

Figure 17. Differences in on-target and off-target induction of indels between wild type, truncated, and hairpin extended sgRNAs for EMX1 spacer 1 ..... 26

Figure 18. Modification of Sniper-screen protocol to select for only highly specific hp-gRNAs ..... 27

Figure 19. Expression cassette for aTc inducible SpyCas9 ..... 29

Figure 20. Expression cassette for aTc inducible guide RNA containing the EMX1 spacer ..... 29

Figure 21. Expression cassette for the CcdB toxin under a pBAD promoter ..... 30

Figure 22. Screening of SpyCas9 on-target activity by colony count ..... 35

Figure 23. Expression cassettes of pgRNA-EMX1 variants ..... 38

Figure 24. Expression cassette for the library of 15 off-targets and the target without a SpyCas9 recognizable PAM ..... 39

Figure 25. Colony count assay for off-target activity ..... 46

Figure 26. Correlation between experimental duplicates of screening for off-target activity in toxin suppressed conditions ..... 47

Figure 27. Correlation between experimental duplicates of screening for off-target activity in selective conditions ..... 48

Figure 28. Depletion in off-target reads before and after screening for SpyCas9:gRNA-EMX1 off-target activity ..... 49

Figure 29. Cassette for pgRNA-EMX1 hairpin library variants ..... 51

Figure 30. Predicted secondary structures and their free energies of hp-gRNAs isolated from non-selective conditions ..... 56

Figure 31. Predicted secondary structures and their free energies of hp-gRNAs isolated from selective conditions ..... 57

Figure 32. Predicted of secondary structure of isolated hairpins from screening of hairpin library with library of 4 potential tetraloops ..... 59

Figure 33. Gel electrophoresis results from the *in vitro* digestion of gRNA-EMX1, “s8”, “s9”, and “s10” ..... 61

Figure 34. Replicate of gel electrophoresis results from the *in vitro* digestion of gRNA-EMX1, “s8”, “s9”, and “s10” ..... 61

# CHAPTER I: INTRODUCTION

## GENE EDITING SYSTEMS

The ability to alter genetic material in living organisms in a highly direct and specific manner has been greatly sought after for both reverse genetics applications and treating genetic diseases. Historically, gene editing platforms involve the use of protein complexes that damage DNA followed by mutagenic repair via the host cellular machinery. Earlier gene editing technologies such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) were shown to be relatively restrictive due to their complexity and complicated design (Fig. 1).[1] The first of these technologies, zinc finger proteins, were discovered as DNA binding proteins in *Xenopus* and later coupled to the restriction enzyme *FokI* to enable DNA double strand breaking activity.[2] Further optimization of ZFNs resulted in a large spectrum of gene editing capabilities, including repair of mutations, deletions, insertions, inversions, duplications, and translocations.[3] Despite the versatility of ZFN technologies, another class of engineered DNA-binding nuclease effectors, TALENs, emerged and gained rapid interest for their limitless targeting range and elevated rate of DNA scission. TALENs originate from highly conserved repeats in domains within *Xanthomonas* bacterial proteins which are injected into host plant cells, altering plant gene expression and thus enabling a more advantageous environment for bacterial colonization. Further engineering of TALENs resulted in modular repeat domains that could be programmed to recognize virtually any DNA sequence and cleave the site with the assistance of a *FokI* domain.[3]

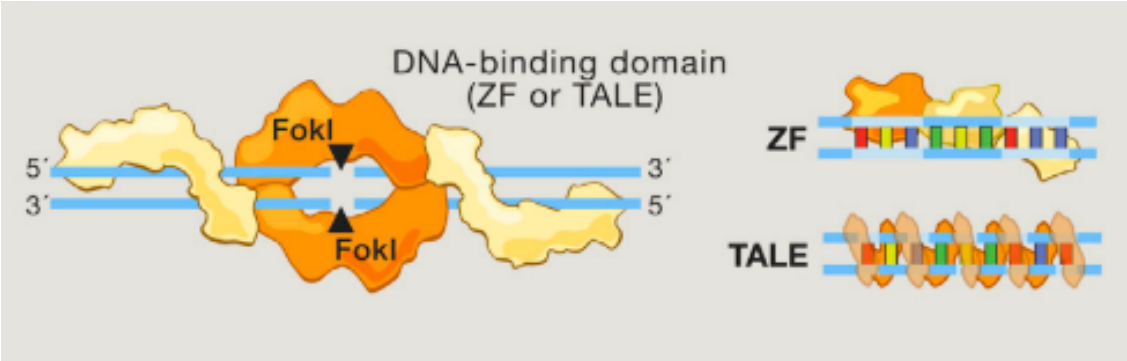



Figure 1. **Basic structure of ZFs and TALEs**.[4] Engineered variants of these naturally occurring DNA-binding proteins complexed with the *Fok*I restriction enzyme enabled site-specific scission of dsDNA.

The discovery of both ZFNs and TALENs provide the foundation for targeted gene editing biotechnologies in a variety of model organisms. More recently, gene editing biotechnologies use enzymes derived from the bacterial immune response to foreign genetic material. The discovery of this prokaryotic immune response known as CRISPR (clustered regularly interspaced short palindromic repeats) has revolutionized gene editing therapies. The drastic increase in the use of CRISPR associated (Cas) effectors shortly after their discovery can be attributed to their abundance across bacterial species and their extreme simplicity.[5] Cas effectors already possess a nucleolytic domain, and when complexed with a programmable RNA cofactor, they can target any genomic site of interest. The sole dependence on the RNA cofactor to target DNA or RNA sequences has been monumental in the application of these systems for potential gene editing therapeutics.[6] Complicated re-engineering of new proteins comprised of multiple repeat domains, such as TALENs or ZFNs, is no longer a requirement for targeted gene editing.

## BASICS OF PROKARYOTIC IMMUNITY

To understand how Cas effectors are used in gene editing, it is instructive first to discuss their origin as a prokaryotic defense mechanism against foreign nucleic acids. About 50% of bacteria and 90% of archaea have evolved complex adaptive immune systems for protection against phage infection and extrachromosomal DNA.[7] In these prokaryotes, the source of immunity is a

genomic locus composed of an AT-rich leader sequence, an array of CRISPR repeats separated by short DNA sequences known as "spacers", and a Cas effector.[8] There are two classes and six types of CRISPR-Cas systems, all with unique modular organizations due to the evolutionary divergence that provides prokaryotes with immunity against the introduction of foreign DNA or RNA. Class I is defined by multi-subunit effector complexes of types I, III, and IV. Type I interference involves cleavage of large segments of DNA as the activated helicase-nuclease Cas3 can move unidirectionally along DNA molecules of interest and introduce strand breaks.[9] Class I CRISPR-Cas systems have been employed in genome editing applications but are more difficult to manipulate due to their multi-subunit effectors, and, when activated, they usually degrade large portions of genome instead of making specific changes to individual genes.[10] Type III complexes destroy both RNA and DNA by cleaving target strands in 6 nucleotide increments and are hypothesized to originate from transcriptional-dependent interference.[11]

There are three types of single unit effector complexes of Class II CRISPR-Cas systems with variable targeting abilities. The CRISPR effector of type II CRISPR systems, Cas9, produces a blunt DSB and is most commonly used in gene editing applications. The Type V effector, Cas12, produces two staggered cuts on target DNA. The effectors of Type VI systems, Cas13, recognize and cleave RNA.[4] These diverse interference mechanisms and variations in molecular requirements of these CRISPR-Cas systems provide the potential for various biotechnological applications.

There are three discrete steps in the Type II CRISPR response; spacer acquisition, CRISPR-RNA (crRNA) biogenesis and processing, and interference that results in targeted destruction of phage or plasmid sequences that are complementary to the crRNA spacer (Fig. 2).[12] The detection of foreign genetic material inside a prokaryotic cell containing this immune response triggers transcription of elements of the CRISPR locus. In most CRISPR-Cas systems, the *cas1* and *cas2* gene products are nucleases that form a complex that enables capture of small segments of the invader's genetic material and mediates spacer acquisition.[8] Upon initial detection of an infection Cas1 and Cas2 incorporate segments of the foreign genetic material in between CRISPR repeats, resulting in an array of alternating CRISPR repeats and spacers.[13]

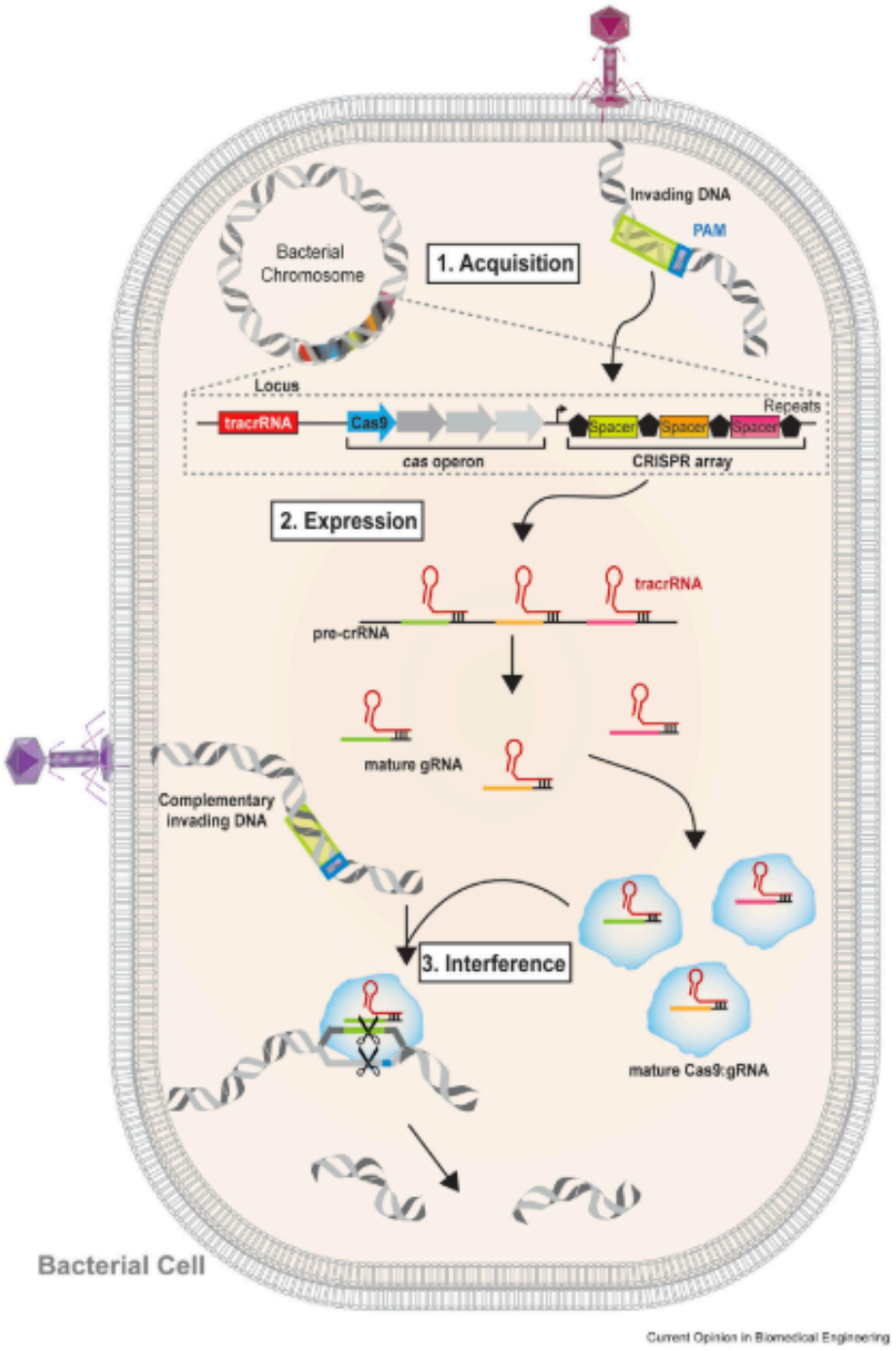



Figure 2. **Schematic of the CRISPR adaptive immunity mechanism**.[12] In nearly all CRISPR-Cas systems, prokaryotic CRISPR loci codes for RNA-guided endonuclease activity to recognize and degrade targeted sequences based on immune memory.

Upon subsequent infection, the expression stage of immunity is initiated. In Type II CRISPR systems, which are most commonly used in gene editing, the CRISPR repeats and spacers are transcribed into a pre-CRISPR-RNA (pre-crRNA) that is later processed into multiple mature crRNAs and trans-activating RNAs (tracrRNAs), each containing a spacer. When combined, the tracrRNA and crRNA form a chimeric RNA that complex that binds with the Cas endonuclease and programs targeted cleavage of phage DNA or RNA.[14] The spacer portion of the gRNA binds complementary to the phage or plasmid DNA downstream of a motif called the protospacer adjacent motif (PAM) and initiates unwinding of DNA by the Cas effector, resulting in R-loop formation.[15] The catalytic domains of the Cas effector then act as molecular scissors and create a double strand break (DSB), ultimately destroying foreign DNA.

## CLASS II TYPE II CRISPR-CAS SYSTEMS IN GENE EDITING

The repurposing of Class II Type II CRISPR-Cas systems has rapidly emerging applications in gene editing therapeutics and diagnostics. The main components of these systems, the Cas effector and engineered single guide (sgRNA) that drives site-specific DNA double strand break (DSB) at a site complementary to the sgRNA spacer, are easily programmable and have relatively low complexity (Fig. 2).[5] Due to simplicity in engineering these sequence-specific, RNA-guided, DNA-targeting endonucleases, they have been repurposed for their gene targeting ability in various model organisms.[1,16]

Class II Type II CRISPR-Cas systems are viable gene editing candidates due to their efficacy, predictability, and customizable nature. There are three core genes that make up Type II CRISPR-Cas9 systems (cas1, cas2, and cas9), and the Cas9 protein is required for gene editing.[17] *Streptococcus pyogenes* Cas9 (SpyCas9) is the most extensively investigated endonuclease in human genome editing. Cas9, the effector responsible for scission in type II systems, has few molecular requirements for targeted DNA cleavage compared to previously studied sequence-specific nuclease technologies. A short DNA motif, known as the protospacer adjacent motif (PAM) is recognized by Cas9 and the DNA sequence complementary to the targeting region of the RNA cofactor binds to the RNA cofactor immediately next to the PAM (Fig. 3).[5]

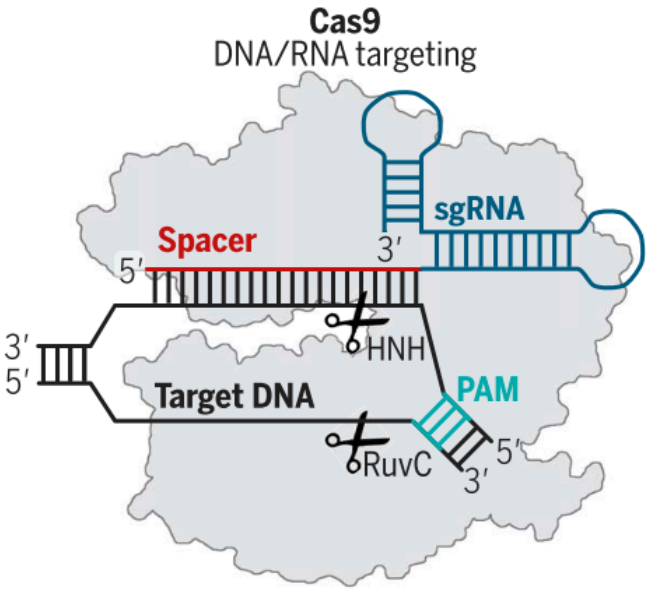



Figure 3. **Mechanism of DNA target recognition and scission by Cas9**.[5] The targeting 20 bp "spacer" region of the sgRNA recognizes and binds complementarily to the 20 bp "protospacer" of the target region of the double stranded DNA immediately 3' of the PAM site, recognized by SpyCas9. The formation of this "R-loop" structure within the active site of SpyCas9 enables scission of both DNA strands, carried out by the HNH and RuvC nuclease domains.

The PAM site recognized by Cas9 allows for freedom in sgRNA design and target site selection.[18] The well characterized target site for SpyCas9 is a 20-bp sequence complementary to the gRNA flanked downstream by a 5'-NGG-3' PAM, or a more weakly recognized 5'-NAG-3' PAM.[19] The SpyCas9 PAM is recognized by two tryptophan residues in loops of the effector.[20] Upon recognition of a target sequence SpyCas9 introduce formation of an R-loop, followed by a DSB and subsequent host cell mutagenic repair mechanisms, useful for making a gene inoperative. Since SpyCas9 has been the most intensively studied Cas9 analog, it is the prime candidate to further our understanding of the implications of Cas9 structure as well as the structural impact of gRNAs on gene targeting activity and specificity.

The most advantageous attribute of Class II Type II CRISPR-Cas9 systems is the RNA cofactor, as mentioned previously, that can be programmed to match any 20-bp DNA site of interest. For

gene editing platforms, this RNA cofactor can be selected or designed based on various criteria, including sequence uniqueness, secondary structure, and interactions within the Cas9 active site.[14] Functional components of gRNA architecture have been identified downstream of the DNA-targeting spacer sequence and their impact on SpyCas9 function has been characterized (Fig. 4).[14]

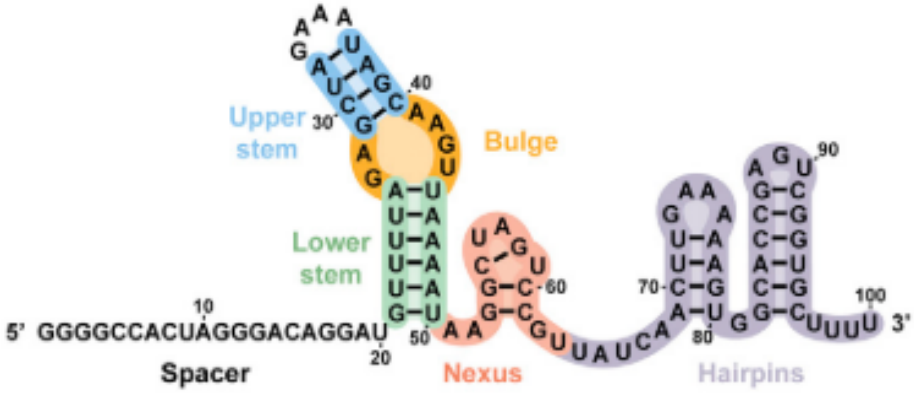



Figure 4. **Nomenclature of SpyCas9 gRNA modules**. The lower stem, bulge, and upper stem are components form the junction of the CRISPR repeat and tracrRNA duplex. The nexus and hairpins are located at the 3' end of the tracrRNA.[14]

It was found that the introduction of mutations to the upper and lower stems did not interrupt SpyCas9 function unless there were consecutive mutations in the lower stem. Furthermore, the first guanine to the 3' end of the spacer at the beginning of the lower stem is a conserved feature in crRNAs of Type II CRISPR systems, but this guanine is not critical for SpyCas9 activity. SpyCas9 function was found to be extremely disrupted by a mutation within the bulge motif of the gRNA. However, removal of the upper stem and replacement of the bulge with a tetraloop motif resulted in an active SpyCas9 in human cells.[14] Multiple biochemical alterations of the nexus motif resulted in a decrease in SpyCas9 activity, but a slight change in nexus conformation can be tolerated, according to cell studies. Disruptions in activity associated with alteration of the nexus and bulge motifs indicate their mediation of essential gRNA:SpyCas9 interactions with target DNA. This study also found that for SpyCas9 systems, the nexus and just one of the hairpin motifs are critical in target DNA interactions and functional versions of these motifs can

be swapped between various orthogonal Cas9 analogs. The identification of vital components in gRNA architecture provides a basis for effective, stable, and potentially more compact gRNAs that provide higher efficacy of SpyCas9:DNA interactions.

The SpyCas9 CRISPR-Cas system provides the ability to accomplish highly efficient gene editing with just two components that are modular in nature. There has been significant progress in characterizing these components and how to appropriately modify and optimize them for therapeutic use. We now understand the basic molecular requirements of SpyCas9, many SpyCas9 analogs, and the most widely used guide RNA cofactor, the basic sgRNA, as well as some of the underlying structural and sequential factors that enhance or diminish SpyCas9 activity. The impact of further modulation of structure the sgRNA component to probe the biophysical interactions that may enhance gene editing capabilities has been investigated, but not extensively. Here, we aim to further optimize the extensively studied SpyCas9 system for enhanced methods in gene editing and epigenetics.

## METHODS IN GENE EDITING AND EPIGENETICS

In mammalian cells, when a DSB is initiated by Cas9 there are two methods of DNA repair that may be triggered. DNA is repaired by either homology directed repair (HDR) or non-homologous end joining (NHEJ) (Fig. 5).[4] NHEJ mechanisms can result in the introduction of insertions or deletions, known as “indels” at the cleavage site that can result in a frameshift or knockout of a gene of interest.[17] When a synthetic “repair template” containing a desired mutation is introduced along with Cas9, HDR can be used to introduce precise sequential changes into target sites with relatively low efficiency through recombination.[21] It is important when designing gene editing methods for therapeutic use that the introduction of indels or undesired proximal DNA sequences are minimal. Avoiding unwanted DSBs in the therapeutic use of Cas9 variants is an effective method of preventing benign or harmful insertions or deletions of genetic information.

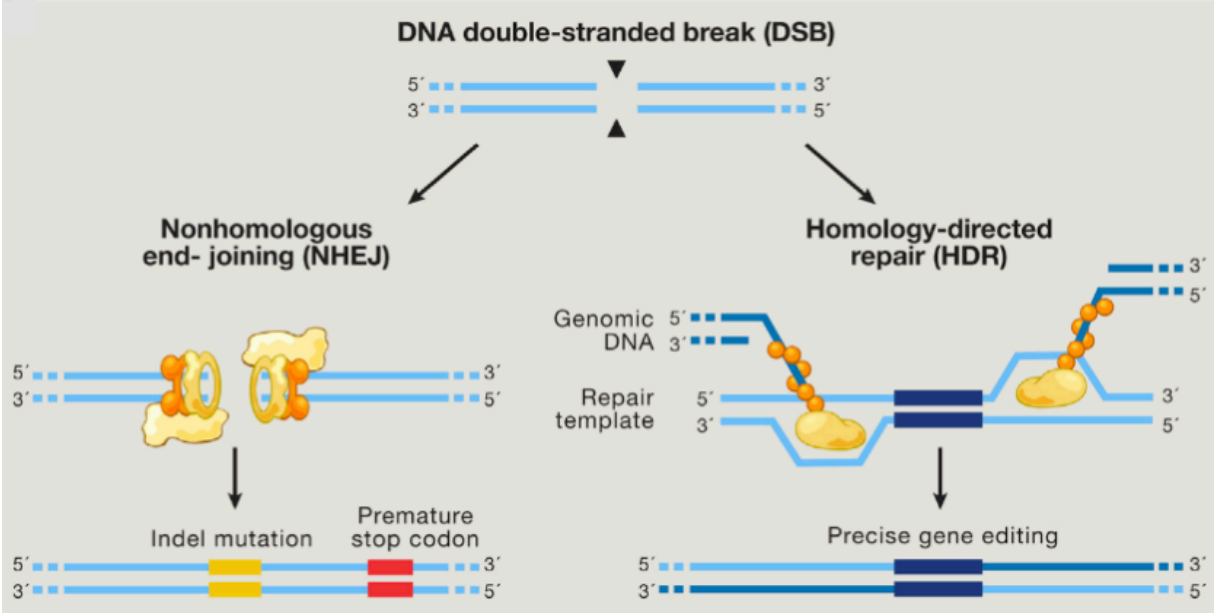



Figure 5. **Mechanisms of non-homologous end joining and homology-directed repair**.[4] Host cellular machinery either repairs a DSB by NHEJ or HDR, resulting in either an indel induction followed by a frameshift or incorporation of a desired repair template.

Due to the simplicity in altering the gRNA cofactor, gRNA libraries have been generated for high throughput screens for genome-wide knockout experiments for efficient forward genetic screens. Identification of gRNA cofactors that complex with SpyCas9 and induce highly targeted DNA cleavage can be accomplished by exposing a large number of ~20 bp oligonucleotides into an environment containing SpyCas9 and target DNA and then sequencing post cleavage. One screening method to determine targeting ability of specific gRNA sequences introduces a pool of approximately 123,000 potentially targeting gRNAs with SpyCas9 and monitors gene knockout of engineered green fluorescent protein (EGFP) in human cells.[22] This method, known as genome-scale CRISPR Knock-Out version 2 (GeCKO v2), was able to isolate 6 targeting gRNAs for approximately 23,000 different genes from the oligonucleotide pool.

Although CRISPR-Cas systems were initially reprogrammed for their ability to induce mutation at a targeted DNA site, other applications quickly emerged. To date, Cas9 systems have been utilized in a wide range of targeted genome editing and epigenetic methods, including

transcriptional silencing, single nucleotide base editing without DSB events, epigenetic modification, and prime editing with a reverse transcriptase.[23, 24, 25, 26] These methods are described below.

Gene regulation with a deactivated Cas9 variant and a targeting gRNA was found to interfere with transcriptional machinery by steric inhibition of polymerase activity.[27] The potential for gene downregulation in the form of transcriptional silencing with the CRISPR-Cas9 toolbox was investigated and a method referred to as CRISPRi was established. CRISPRi takes advantage of a nuclease-dead (dCas9) complexed with a gRNA that binds to a gene of interest, forms an R-loop, and inhibits RNA polymerase from initiation of mRNA transcript synthesis (Fig. 6).[23]

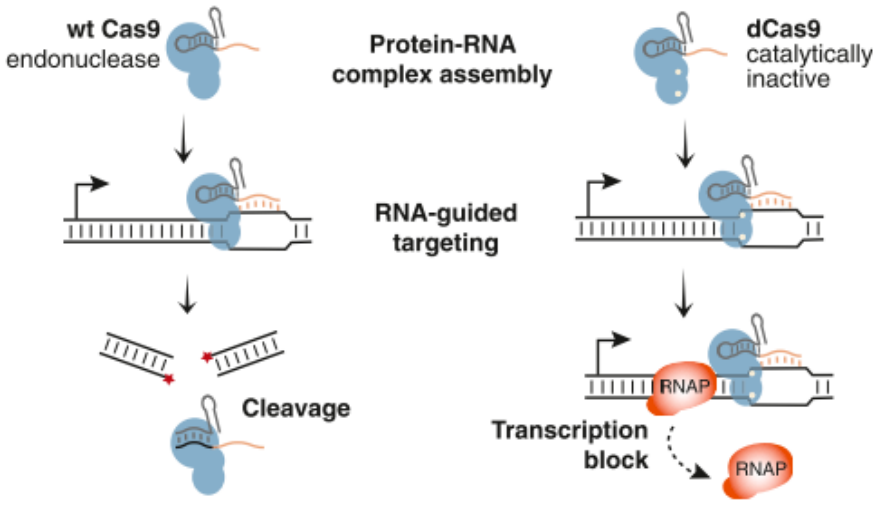



Figure 6. **Mechanism of wild type SpyCas9 versus catalytically inactive SpyCas9**.[27] RNA polymerase binds to DNA to initiate transcription but is sterically inhibited by the R-loop formation catalyzed by bound dCas9.

Multiple target genes were found to be significantly repressed in both *Mycobacteria smegmatis* and *Mycobacteria tuberculosis* without alteration of the genetic sequence. CRISPRi has also been used for gene silencing in human cells.[27] EGFP silencing was demonstrated in HEK293 cells with human codon-optimized dCas9 and an gRNA designed to target EGFP. It was also found that regulatory effects of gene silencing in human cells are more sensitive, and the

proximity of gRNA targeting to the transcriptional start site is important for efficiency of transcription repression. These findings show that gene downregulation can be effectively achieved in *E. coli* and a potential for therapeutic transcriptional silencing in mammalian cells.

Base editors are Cas9 derived gene editing tools that can correct pathogenic point mutations that cause diseases such as Tay-Sachs, color blindness, and sickle-cell anemia without DSBs, but by direct chemical modification of nucleotides.[14] Base editors use a chimeric dCas9 or Cas9 nickase (Cas9n) fused to a cytidine deaminase that converts dC nucleotides to dT. Additionally, David Liu and coworkers have fused together an RNA adenosine deaminase engineered to recognize DNA nucleotides with a Cas9n and carried out several rounds of directed evolution to create an adenine base editor that converts dA to dG and has 50% efficiency in human cells (Fig. 7).[24] Unwanted edits due to indels or HDR are avoided without induction of a DSB. The formulation of this programmable base editor is monumental in therapeutics for diseases that are caused by single point mutations.

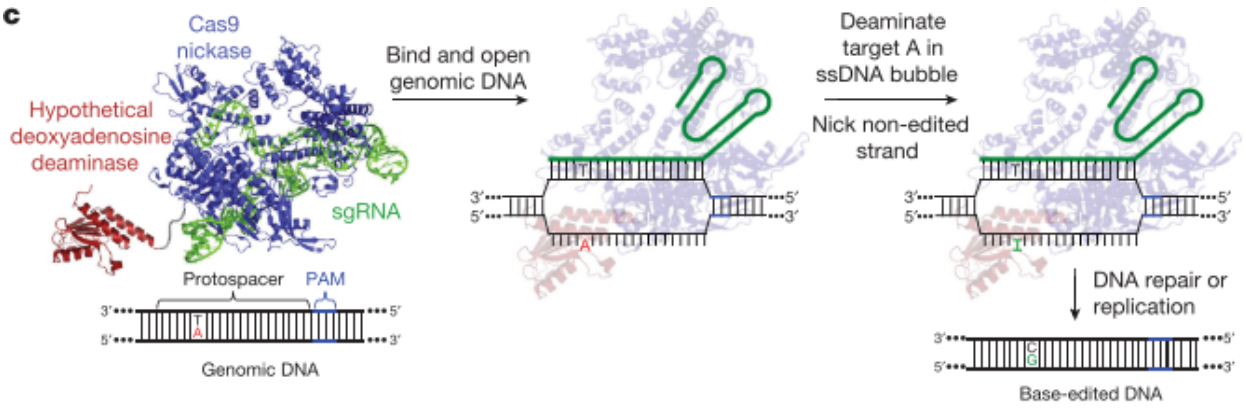



Figure 7. **Overview of Cas9n base editing mechanism**.[24] Deamination of an individual adenine nucleotide to an inosine, read by polymerases as guanosine, is accomplished via the fusion of Cas9n to a complex to a deoxyadenosine deaminase.

Several epigenetic modification technologies have been developed to regulate gene expression in the form of phenotypic or transcriptional control. Epigenetic control for downstream gene regulation has been demonstrated by harnessing the transcriptional regulation potential of dCas9

and complexing the nuclease dead protein with a histone acetyltransferase (HAT).[25] The dCas9-HAT complex combined with targeting gRNAs resulted in robust activation of four different human genes in HEK293T cells. This novel method of CRISPR activation and enhancement provides a template for targeted delivery of other enzymes used in epigenetic research.

Prime editing is a more recent breakthrough in the quest for targeted Cas9 genome editing for therapeutic applications with minimal introduction of indels. This "search and replace" method navigates around the need for a DNA DSB with a prime editing guide RNA (pegRNA) guided Cas9n "prime editor" (PE) complexed to an engineered reverse-transcriptase (Fig. 8).[26] Here, the reverse transcriptase facilitates editing via polymerization using the pegRNAs 3' extension that contains a short editing sequence as a template. "Search and replace" genome editing carries the potential for the facile replacement of small portions of genes that encode genetic diseases with low off-target effects or indels.

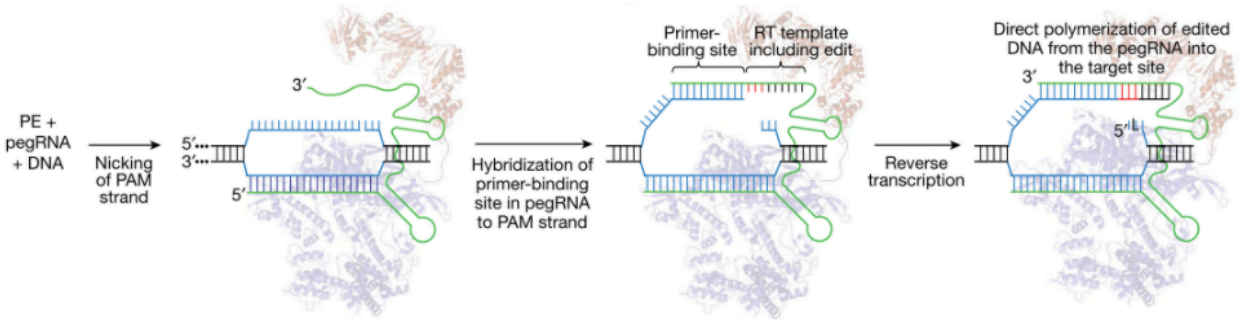



Figure 8. **Mechanism of prime editor pegRNA complex**.[26] In the prime editing process, the Cas9n nicks the target site and exposes a 3' hydroxyl group that initiates reverse transcription of an extension of the pegRNA that codes the desired edit.

## OFF-TARGET ACTIVITY, PREVALENCE, AND PREDICTION

Despite the numerous advantages and uses of Cas9 in genome editing, there is a significant prevalence of genome modification at sites in the same genome where sequences are similar but not identical to the target, especially in mammalian cells. This is a problem in gene editing, where unwanted edits to genomic loci with similar sequences to the target may occur and

produce phenotypically disadvantageous effects. These off-target sites typically contain small differences within the "protospacer" sequence (the sequence targeted by the gRNA spacer) that are not strongly discriminated by the CRISPR effector.[28] Off-target activity needs to be understood and eliminated in order for SpyCas9 to be utilized for human gene therapies. The specificity of SpyCas9 systems is hypothesized to be largely controlled by a series of conformational changes of the effector, R-loop formation, and gRNA architecture.[29]

Prediction of Cas9 off-target sites is a challenging process for which various high throughput methods have been developed. One method to determine off-target binding potential uses dCas9 and a library of targets sites with incorporated mutations.[19] The dCas9 binding to on and off-target sites was identified by tracking the position of fluorescently labeled dCas9-gRNA complexes and subsequently using high throughput Next Generation Sequencing (NGS) to detect mutations in off-target sites that allowed dCas9 binding. Data from this technique indicates that mismatches in the PAM distal region, the region of nucleotides located farthest from the PAM, are more tolerable for gRNA recognition and therefore more prevalent in off-target activity, while mismatches in the ~7 bp seed region closest to the PAM are better discriminated. Furthermore, mismatches outside the seed region between positions -8 and -20 bp showed very limited effects on dCas9 binding.[19] These findings show that the mechanism of strand invasion affects the dCas9-gRNA association to off-target sites and the importance of mutation position in off-target effects.

CIRCLE-seq is an *in vitro* method that was developed to effectively detect all possible Cas9 cleavage sites in a genome in a rapid, comprehensive manner.[28] The circularization of genomic DNA containing all potential Cas9 cleavage sites, followed by Cas9 cleavage and paired-end high throughput sequencing yields pairs of reads that give all off-target sequences per gRNA (Fig. 9). CIRCLE-seq introduced tagged insertions at known off-target sites for EMX1 and VEGFA site 1, and showed enhanced detection of off-target sequences where integration was possible 3 bp away from the PAM.[28] This reference genome independent method can be used to identify several patient specific off-target sites that may have one or more single nucleotide polymorphisms (SNPs) that differ from the majority of the population. The CIRCLE-seq

technique has provided data that is essential for further improvement of CRISPR-Cas9 gene editing and minimization of off-target effects. This method enabled the identification of off-targets that had not previously been discovered.

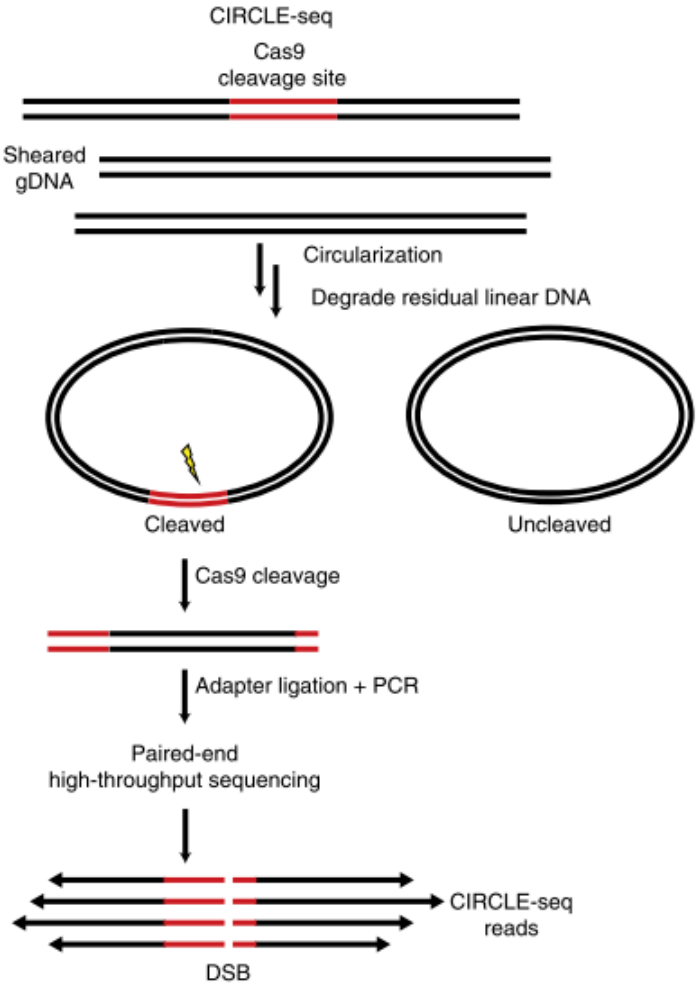



Figure 9. **Workflow of CIRCLE-seq**.[28] Genomic DNA was cut into workable segments, some of which contained a Cas9 recognition site, and circularized. Cas9 cleavage of circular genomic DNA resulted in linear DNA segments with portions of the cut off-target site on either end. PCR amplification followed by next generation sequencing provided a dataset of Cas9 off-target sites in the genome.

More recently, a higher throughput and more comprehensible method of *in-vitro* detection of Cas9 off-target activity was developed by Tsai and coworkers. In an attempt to fully understand principles of Cas9 specificity, 110 sgRNAs were screened against 13 therapeutically important human genes via a method known as, "circularization for high-throughput analysis of nuclease genome-wide effects by sequencing", or CHANGE-seq (Fig. 10).[30]

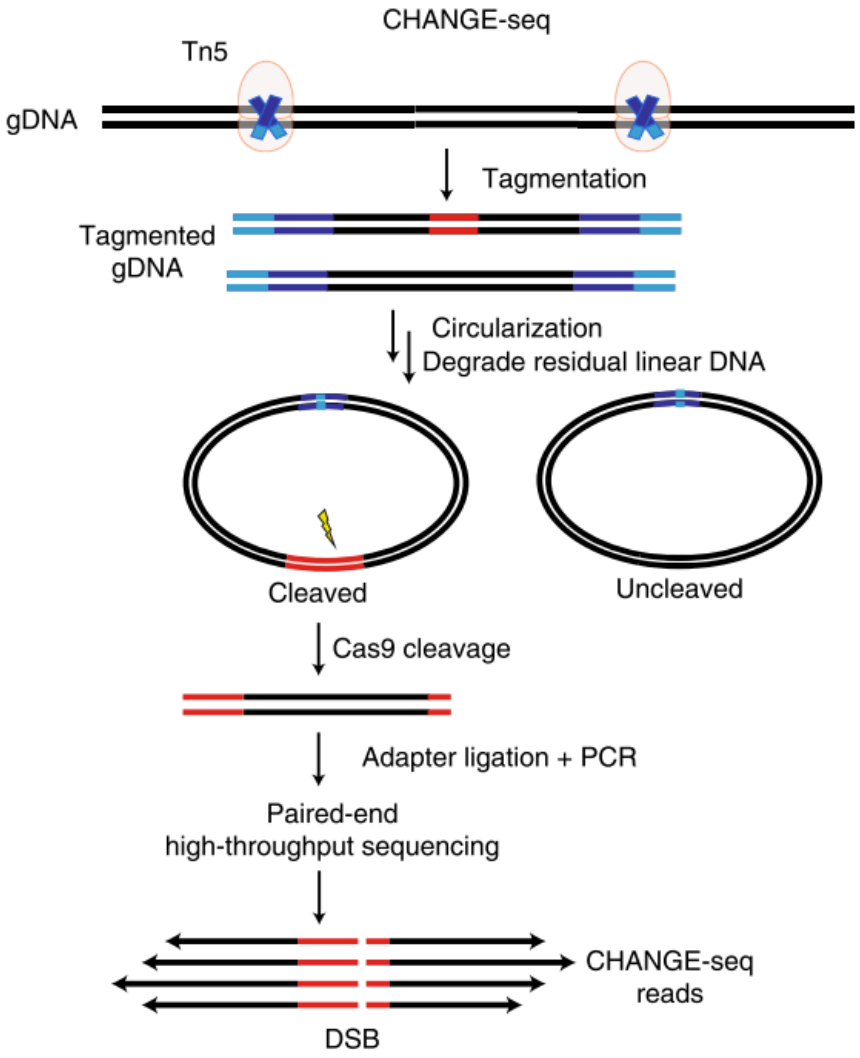



Figure 10. **Workflow of CHANGE-seq**.[30] This method uses a Tn5 transposome to tagment genomic DNA for facile circularization of DNA libraries, followed by Cas9 cleavage, PCR amplification of newly cut sites, and deep sequencing to piece together the sites recognized and cut by Cas9.

When compared to CIRCLE-seq, CHANGE-seq requires far less genomic DNA input and requires far less processing steps, while yielding a high enrichment efficiency.[30] Their highly scalable screen generated the largest dataset of Cas9 targets and off-targets to date, enabling a thorough characterization of genome wide Cas9 activity. They found that guanine frequency in target sequences are associated with increased specificity and that characteristic mismatch combinations and positions yield decreased Cas9 specificity.

## THE SPYCAS9 CONFORMATIONAL CHECKPOINT

Understanding the conformational change checkpoint in Cas9 activity is also a crucial component in determining ways to minimize off-target activity. Structural characterization of SpyCas9 has elucidated the organization of functional and supporting units in the effector.[31] RNA mediated conformational activation was revealed with 2.6 Å-resolution by negative-strain single particle electron microscopy of three-dimensional reconstructions of SpyCas9:gRNA and SpyCas9:gRNA:DNA.[20] The channel that binds target DNA is formed via gRNA binding and subsequent activation of a conformational change of the central unit of the Cas9 effector. It was deduced that the arginine-rich region and PAM-binding loops functions as a hinge to enable conformation change and favor gRNA target DNA binding (Fig. 11).[20] These findings indicate that there are biophysical requirements between gRNA and Cas9 domains and that gRNA binding dictates the activity of SpyCas9.

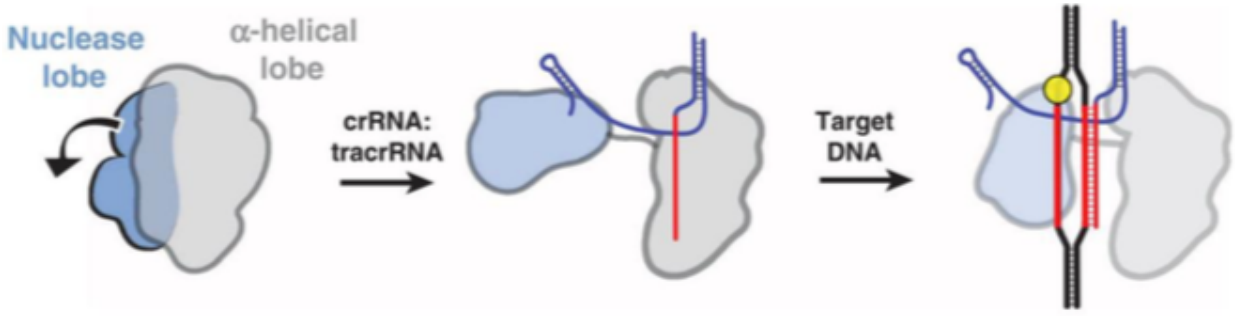



Figure 11. **Schematic of RNA-induced structural activation of SpyCas9 for target cleavage**.[20] The SpyCas9 conformational checkpoint is initiated by binding of the

crRNA:tracrRNA duplex, creating a channel which primes the effector for DNA binding followed by a DSB.

Previous work with CRISPR-Cas9 systems prominent in gene editing have verified the protein structures. Stability remains a crucial component in the function of the SpyCas9:RNA:DNA complex.[32] Single molecule studies of SpyCas9 have been conducted to probe the effects of gRNA binding on SpyCas9 conformational activation and activity. A conformational shift in SpyCas9 due to interactions between the 14th-17th nucleotides from the PAM in the spacer and the gRNA has been found to stabilize the complex, according to single molecule studies.[29]

High resolution characterization of SpyCas9 accompanied by gRNA with their DNA target have been identified to reveal functional interactions.[33] Recently, a group investigated the dynamic stretching of target DNA and binding efficiencies of gRNA:SpyCas9 complex.[12] Although gRNA:Cas9:target sequence interactions have been investigated, the interactions between amino acid residues of SpyCas9 have not been extensively probed. The functional interactions between catalytic HNH and Ruv-C domains and within SpyCas9 noncatalytic domains are also unknown. Protein folding may also affect the ability of Cas enzymes to stay bound to target DNA post-cleavage or release and bind to another target.

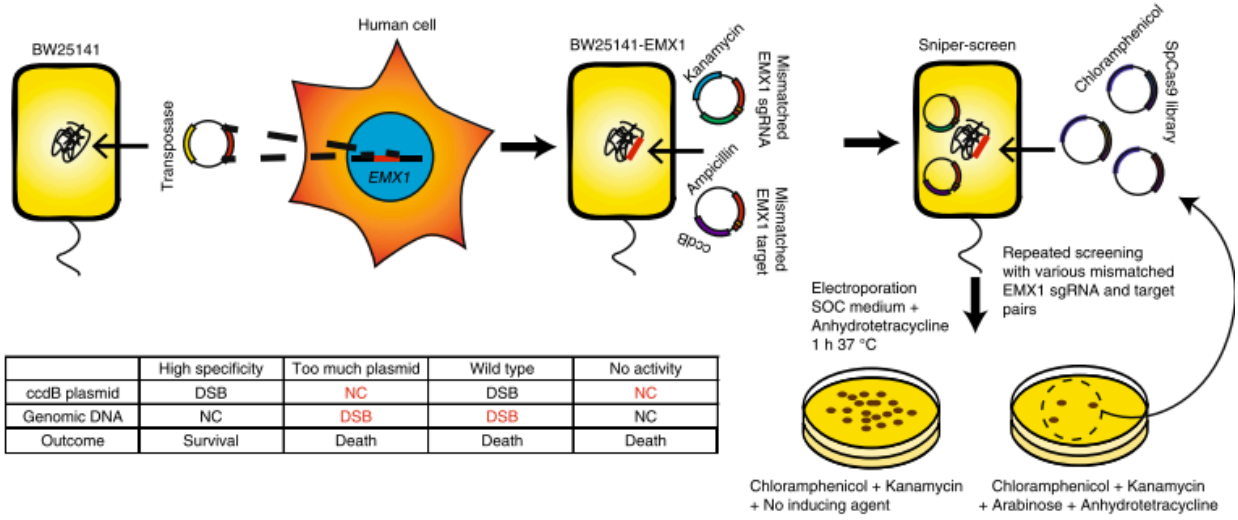
Many structural analyses have not explained the off-target cleavage of SpyCas9 at sites without 100% complementarily to the target DNA. Doudna and coworkers developed a Forester resonance energy transfer (FRET)-based investigation to determine conformational control in response to various gRNAs and target sites.[34] They found that specific motifs at both ends of the gRNA are required to trigger a conformational rearrangement of Cas9 to form a closed conformation, and that a specific ratio of gRNA to SpyCas9 in the hinge state was necessary for Cas9 closure. Conformational analysis of the HNH domain in the presence of gRNAs and dsDNA showed that the HNH domain equilibrates differently for different off-target sites. Interestingly, off-targets with more than 4 bp mismatches were cleaved more slowly, and off-target sites with 1-3 bp mismatches near the PAM distal end were cleaved at an elevated rate in comparison to other off-targets.[34] These findings show that gRNA architecture and conformational activation of Cas9 mediates cleavage of on and off-target sites.

## MANIPULATION OF THE SPYCAS9 EFFECTOR TO INCREASE SPECIFICITY

Various approaches have been developed in order to mitigate unwanted edits arising from off-target Cas9 activity. Limiting Cas9 exposure is another option, where the native or mutated endonuclease is complexed with enzymes that restrict off-target activity.[35,36] Directed evolution of various Cas9 analogs has resulted in mutated effectors with enhanced targeting abilities.[31] Adjustments can be made to the gRNA in the form of truncation, extension, chemical modification, or mutation. Modulation of either component of the SpyCas9 system has been investigated for fine tuning of targeting efficiency. Also, off-target effects may vary depending on relative ratios of gRNA:Cas9:target DNA, where fine tuning the exposure of components to each other can enhance cleavage specificity.

The specificity of Cas9 has been altered by mutation of essential amino acid residues to enhance targeted endonuclease activity. The first attempt to create a high-fidelity Cas9 variant resulted in the design of eSpCas9(1.1) that improved off-target activity. However, it was later determined that the mechanism of off-target reduction in eSpCas9(1.1) relies on an intermediate inactive state that also partially inhibits on-target activity. To further optimize on-target and off-target activity, HypaCas9, was developed by mutation of residues within the REC3 lobe that enhances effector binding to the RNA-DNA duplex.[31] If there are mutations in the spacer, these REC3 mutations prevent REC2 from securing the HNH domain at the conformational checkpoint for cleavage. Mutation of residues within lobes of SpyCas9 can impact the overall R-loop formation and other conformational changes required for Cas9 cleavage.

Directed evolution to generate ultra-specific SpyCas9 that does not diminish on-target activity resulted in the development of Sniper-Cas9.[37] This positive screening method allows for simultaneous detection of on-target and off-target activity of mutated SpyCas9s against the human *EMX1* gene by a SpyCas9 library transfected into *E. coli* containing the EMX1 site in its genome (Fig. 12).




| | High specificity | Too much plasmid | Wild type | No activity |
|---|---|---|---|---|
| ccdB plasmid | DSB | NC | DSB | NC |
| Genomic DNA | NC | DSB | DSB | NC |
| Outcome | Survival | Death | Death | Death |

Figure 12. **Schematic of Sniper-Cas9 directed evolution**.[37] A target sequence from the human *EMX1* gene was first integrated into the genome of *E. coli* strain BW25141 via Tn7 transposon integration. Then, plasmids containing mismatched sgRNAs and toxin plasmids containing a mismatched target sites were transformed into this strain. Upon addition of a library of $10^7$ SpyCas9 mutants a screening for highly specific SpyCas9 specificity was achieved via multiple rounds of screening for survival of only Cas9 mutants that cut the CcdB plasmid and not the genomic EMX1 site.

After multiple rounds of evolution, the Sniper-Cas9 variant was isolated due to the display of higher frequency of on-target activity in comparison to SpCas9(1.1) and HypaCas9 with the same gRNAs. It was also found that high Sniper-Cas9 specificity was not cell-type specific. This intensive positive and negative screening for a highly specific SpyCas9 resulted in an evolved superior Cas9 variant for gene editing applications.

Limiting exposure by a sterically hindered complex composed of mutated Cas9 was attempted to enhance on-target DNA cleavage. Mutation of one residue in each of the HNH and Ruv-C catalytic lobes of Cas9 results in an effector that induces single strand DNA (ssDNA) nicks rather than a DSB. A system of two Cas9n-gRNA complexes were designed to elongate the base pairing length between gRNAs and target site to increase DSB specificity.[35] gRNA with 23 bp offset pairs were introduced to on and off-target sites for the EMX1 gene in HEK293T cells and

off-target activity was reduced. This method was found to be 100-fold more specific than wild type Cas9 in targeting the EMX1 and VEGFA genomic loci.[35] Despite the efficacy of double nicking, this enzyme construct requires complex design and would likely be difficult to deliver therapeutically.

Another method of limiting Cas9 exposure to potential off-target sites was adapted from earlier gene editing technologies, complexing a *Fok*I restriction endonuclease in between two dCas9 enzymes.[36] gRNAs of each dCas9 were designed to flank the target sequence either 15 bp apart or 24 bp apart to anchor *Fok*I for cleavage at recognition site (Fig. 13).

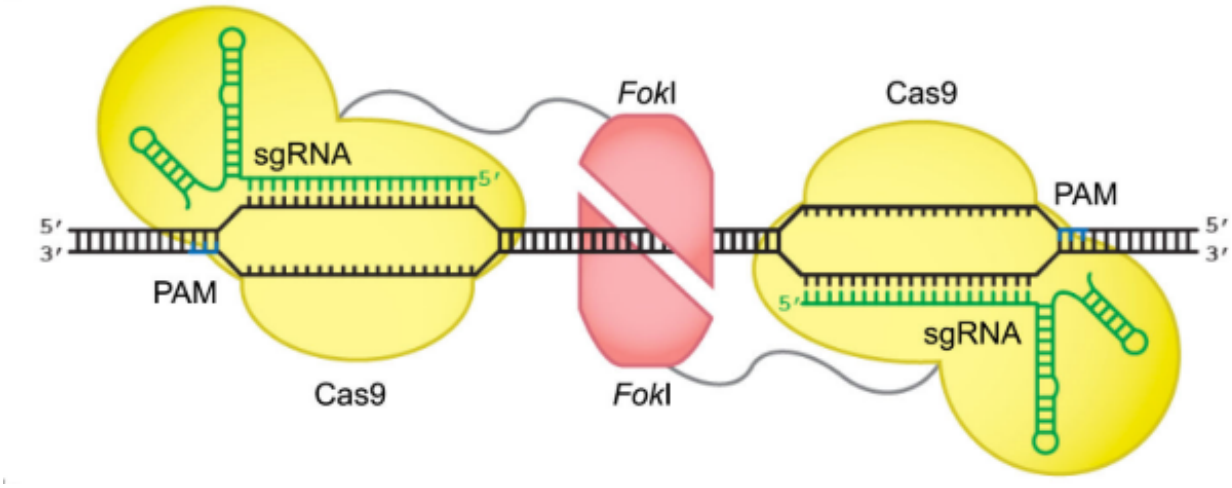



Figure 13. **Structure of the dual-dCas9:*Fok*I:DNA complex primed for scission**.[36] The *Fok*I restriction enzyme was connected to two dCas9s via a linker composed of glycine and serine residues. This multi-unit complex requires a dual Cas9 recognition site for scission, resulting in decreased off-target effects.

Here, the targeting mechanics of the gRNA:dCas9 complex are used while the DNA cleavage is executed by *Fok*I. This method introduces spatial requirements and doubles the molecular requirements for a DSB. Due to this they were able to significantly eliminate off-target activity at 11 off-target sites and maintain 5-10% on-target activity for the VEGFA genomic locus.[36] While this method minimized off-target activity significantly, it also reduced wild type SpyCas9 on-target activity. Due to the complexity in design of this engineered multi-unit effector and the

relative reduction in on target efficacy, this method is not particularly advantageous for gene editing therapeutics.

## OPTIMIZED DESIGN OF GRNAS TO ENHANCE SPECIFICITY

A screening of the activity of 1,841 sgRNAs to identify key sequence features for enhanced sgRNA design led to the development of an online tool, recently renamed CRISPick, for generating highly specific and effective SpyCas9 sgRNAs. Multiple sgRNAs were screened for specificity in both human and mouse cell lines, creating a large enough dataset to develop a predictive model for key sequence features of sgRNA that enhance Cas9 activity, universally.[38] The most active and specific sgRNAs were assessed for nucleotide preferences along the entire target sequence (Fig. 14).

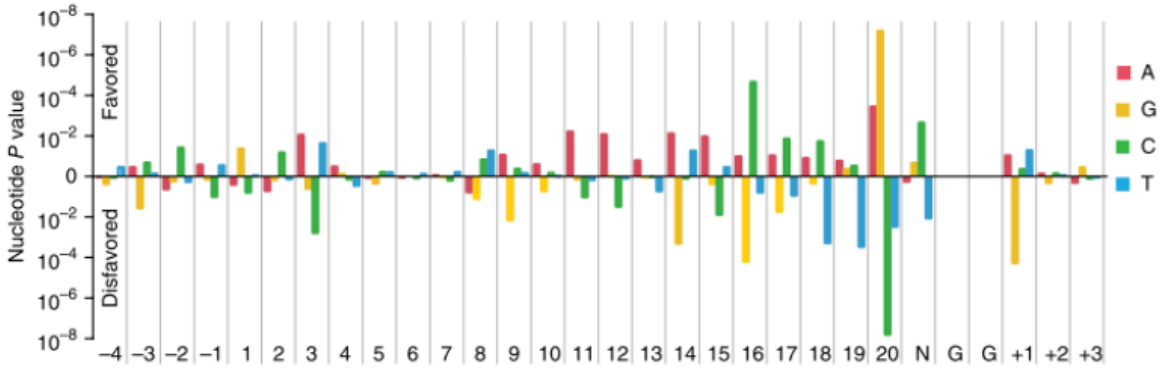



Figure 14. **Favored and disfavored nucleotides along highly active sgRNAs.** A pool of 1,841 sgRNAs were screened for their activity while targeting coding sequences. Interestingly, strong preference for guanine at position 20, cytosine at position 16, and adenines from positions 11-15 were found.[38]

This was the first genome-wide sgRNA library screening that held a large emphasis on-target activity criteria, which enabled the identification of more highly active guides per gene of interest and stricter design rules.

An attempt to further optimize sgRNA design rules provided a dataset of positive and negative screens for detection of sgRNA effects in mouse and human genomes. Doench and coworkers

characterized the previously discovered sgRNA libraries genome-wide based on rules determined from the above-mentioned screen and assessed these libraries for off-target effects. They were able to significantly contribute to a knowledge base about guides by introducing a machine learning algorithm developed to predict sgRNA design rules by quantifying off-target propensities per given spacer and genome.[39] The quantification of a means to predict "off-target potential" of sgRNAs was measured by the cutting frequency determination (CFD) score, which was found to reliably predict off-targetability for 89 sgRNAs and surpass the accuracy of former off-target prediction models.

## THE EFFECT OF GRNA SECONDARY STRUCTURE ON SPYCAS9 SPECIFICITY

Methods of increasing SpyCas9 specificity mentioned above involve the manipulation of just the SpyCas9 effector. Introducing various secondary structures via extensions or truncations of gRNAs to limit off-target activity without diminishing on-target recognition involves the manipulation of the smaller and simpler component of the system.

The use of truncated or extended non-canonical gRNAs and their effect of DNA unwinding of the 5' end of the protospacer was observed with a single molecule FRET assay.[40] In this investigation, the crRNA portion of gRNAs were truncated or extended while the tracrRNA portion was held constant. gRNAs truncated to 18 nucleotides were found to have a much lower fraction of unwound DNA by effector recognition than complete gRNAs. Furthermore, 17 nucleotide long gRNAs were nearly undetectable. Due to a lack of unwinding data at the protospacer, the specificity of truncated gRNAs could not be defined. Another study found that truncation of gRNAs by two nucleotides at their 5' ends were able to accommodate more mismatches in the protospacer.[29]

DNA unwinding by several engineered Cas9 variants with extended gRNAs having either one or two non-hybridizing guanines to the 5' end was investigated for DNA targets with 0-3 mismatches. The DNA unwinding data shows that specificity of wild type SpyCas9 and SniperCas9 was significantly enhanced with the addition of one guanine to the 5' end of the gRNA, while on-target DNA unwinding remained the same.[40] It was also found that the addition

of the second 5’ guanine resulted in a decrease of on-target recognition for many of the engineered Cas9 variants. Mainly, an increase in extension length overall increased target DNA unwinding specificity.[40] To enhance the highly specific effect of single guanine elongation to the 5’ end of gRNAs, engineered Cas9 variants have been developed that cap this extension with glutamine and tyrosine residues so that the gRNA does not interfere with the nucleolytic lobe of the effector and compromise activity.[41] These findings provide insight for specific gRNA design.

## THE EFFECT OF GRNA 5’ HAIRPIN EXTENSIONS ON SPYCAS9 SPECIFICITY

Extension of the 5’ end of sgRNAs results in a variety of secondary structures of extended sgRNAs, known as, x-sgRNAs. These secondary structures are due to the ability of RNA to bind to itself, sometimes using non-canonical base pairs, creating a wide range of folding motifs. One subset of these secondary structures that result from 5’ elongation of sgRNAs is the RNA hairpin. Hairpin sgRNAs (hp-sgRNAs) themselves have potential to form various folding patterns, and even the same hairpin sequence can have multiple secondary structures, each of which can be observed depending on energetics of their environment (Fig. 15).[42]

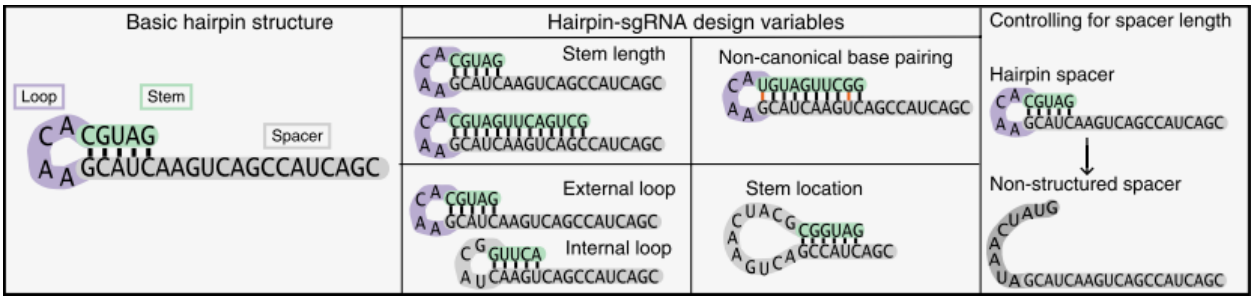



Figure 15**. Basic 5’ extended hairpin structure attached to a sgRNA**.[42] The basic hairpin structure is composed of a stem and loop, most commonly with the loop containing four nucleotides. Variations of these hp-sgRNA structures arise from alteration of stem length, the potential for the PAM distal end of the spacer to bind to itself, and non-canonical base pairing. Without base pairing between a component of the hairpin extension and the spacer, we see a non-structured spacer.

Investigation of dCas9 and wild type SpyCas9 DNA binding interactions in complex with different guides aides in the understanding of conformational gating dictated by the 14$^{th}$-17$^{th}$ nucleotides of a target region. Due to the likelihood of Cas9:gRNA:DNA complexes forming, even at sites with up to 10 mismatches, truncated or hairpin extended gRNAs that interfere with this target-binding region of the spacer were mapped. Interestingly, gRNAs that interfered with the 14$^{th}$-17$^{th}$ nucleotide "conformational gate" either via truncation or extension resulted in heightened Cas9 specificity. Monte Carlo simulations showed that truncated gRNAs were found to improve specificity via a kinetic effect rather than an energetic effect while extended gRNAs enabled the full process of strand invasion and were more stably bound to DNA.[29] While truncated gRNAs showed potential for increased specificity, there is an importance in the presence of the 19$^{th}$ and 20$^{th}$ nucleotides on the 5' end of the protospacer, perhaps aiding in re-annealing the nucleotides positions critical for conformational gating (Fig. 16).

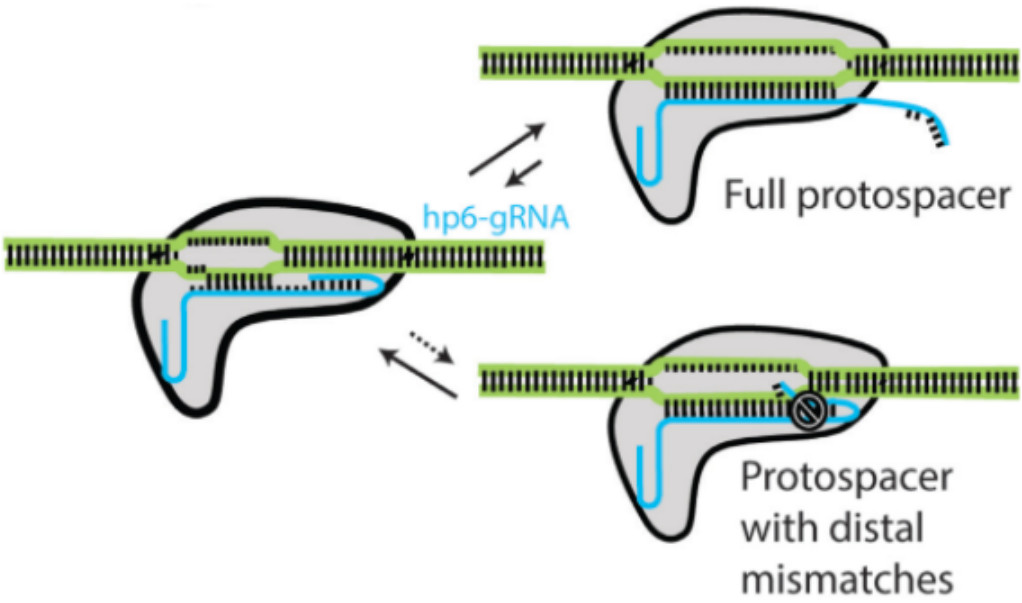



Figure 16. **Mechanism of 5' hairpin extension interference of dSpyCas9 binding to mismatched target**. In the presence of the true target, or "full protospacer", the hairpin structure opens and allows full melting of the spacer to the protospacer. In the presence of an imperfect target site, the hairpin structure preferentially binds to the gRNA spacer, preventing full melting to the target and inhibiting dSpyCas9 binding.[29]

Extended gRNAs with hairpin structures were found to diminish off-target binding while retaining full stability of strand invasion. It is hypothesized that given the narrow Cas9 DNA-binding channel, the geometry of a closed hairpin structure is ideal for carrying out the entirety of strand invasion while prohibiting the conformational change necessary for Cas9 activity at off-target sites.[29]

Further exploration of hp-sgRNAs showed an overall increase in Cas9 specificity with various pre-designed hairpin extensions. Since the RNA-RNA duplex of hp-sgRNAs can be accommodated within the DNA binding loop of the Cas effector, this secondary structure was ideal for investigation of energetic differences between the hp-sgRNA binding with either a perfect DNA target site or a site containing off-target mismatches.[42] Hp-sgRNAs and non-structured sgRNAs were screened for their ability to alter gene activation via dCas9-P300 trans-activator binding to a DNA target site in human cells, and it was confirmed that only the hairpin secondary structures, not simply extensions, were essential for alteration in gene activation.[42] This assessment of binding efficiency of the dCas9-P300 complex provided the baseline requirement for SpyCas9 on-target activity of these hp-sgRNAs, an essential prerequisite for determination of specificity.

There are three known critical energetic processes undergoing a dynamic equilibrium during strand invasion. These are the hybridization of the target DNA, the hybridization of the spacer to the target DNA, and the breaking or forming of the spacer secondary structure.[42] This model for strand invasion was tested for hp-sgRNAs and resulted in a strong correlation between time bound to DNA and increase in gene expression, confirming that secondary structure of the sgRNA component can alter invasion kinetics and R-loop stability which therefore activates SpyCas9. Since hp-sgRNAs were determined to activate SpyCas9 in the presence of the target site, the next step was to determine whether hp-sgRNAs could enhance gene editing specificity by limiting off-target effects in human cells. For the EMX1 spacer 1 (Fig. 17), VEGFA spacer 1, and VEGFA spacer 2 human genomic loci, all optimized hp-sgRNAs that were designed significantly and reliably increased SpyCas9 specificity, much more than that of truncated gRNAs.[42]

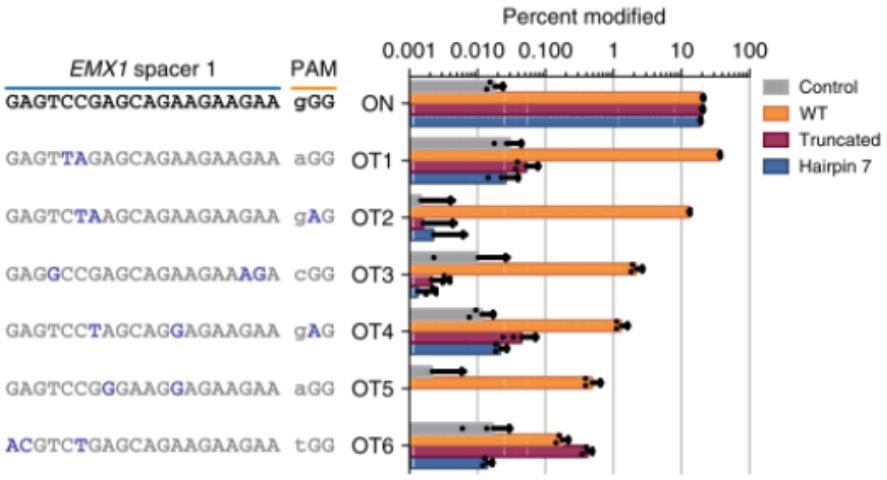



Figure 17**. Differences in on-target and off-target induction of indels between wild type, truncated, and hairpin extended sgRNAs for EMX1 spacer 1**.[42] Percentages of modified genomic sites are shown for each sgRNA variant.

Similar observations of enhanced specificity were made for other CRISPR systems like Cas12 despite their differences from Cas9, hinting that a similar mechanism occurs for all effectors where the secondary structure of the hp-sgRNAs impedes the energetics of R-loop formation and conformational change at off-target sites.

## IDENTIFICATION OF DESIGN RULES FOR HAIRPIN EXTENDED GRNAS

Up until this point, key design rules for guides have been identified for the canonical gRNA for various genes of interest, mostly by nucleotide sequence.[38,39] The tendency of gRNA PAM distal nucleotide additions to increase specificity has been confirmed, and hairpin extensions have been shown to reliably limit off-target effects.[42] Currently, since we do not have concrete design rules for highly specific hp-gRNAs, each hairpin guide needs to be identified, generated, and tested by hand. Individually examining hp-gRNAs for specificity in this way is a laborious and time-consuming process. We aim to characterize the properties of hairpin extensions and identify the most important features of a hairpin, the features that have a direct effect on SpyCas9 specificity.

Here, we develop a screening method to select for highly specific hp-gRNAs and provide a dataset for the EMX1 gene. The screening method for specific hp-gRNAs was adapted from the

screen used for directed evolution of Sniper-Cas9, an *E. coli* based selection method for isolation of highly specific active SpyCas9 mutants.[37] As shown in Figure 18, our method introduces library of randomized hp-gRNAs into a recombinogenic deficient *E. coli* strain in the presence of a plasmid expressing SpyCas9 and a toxin plasmid with an inserted target site.

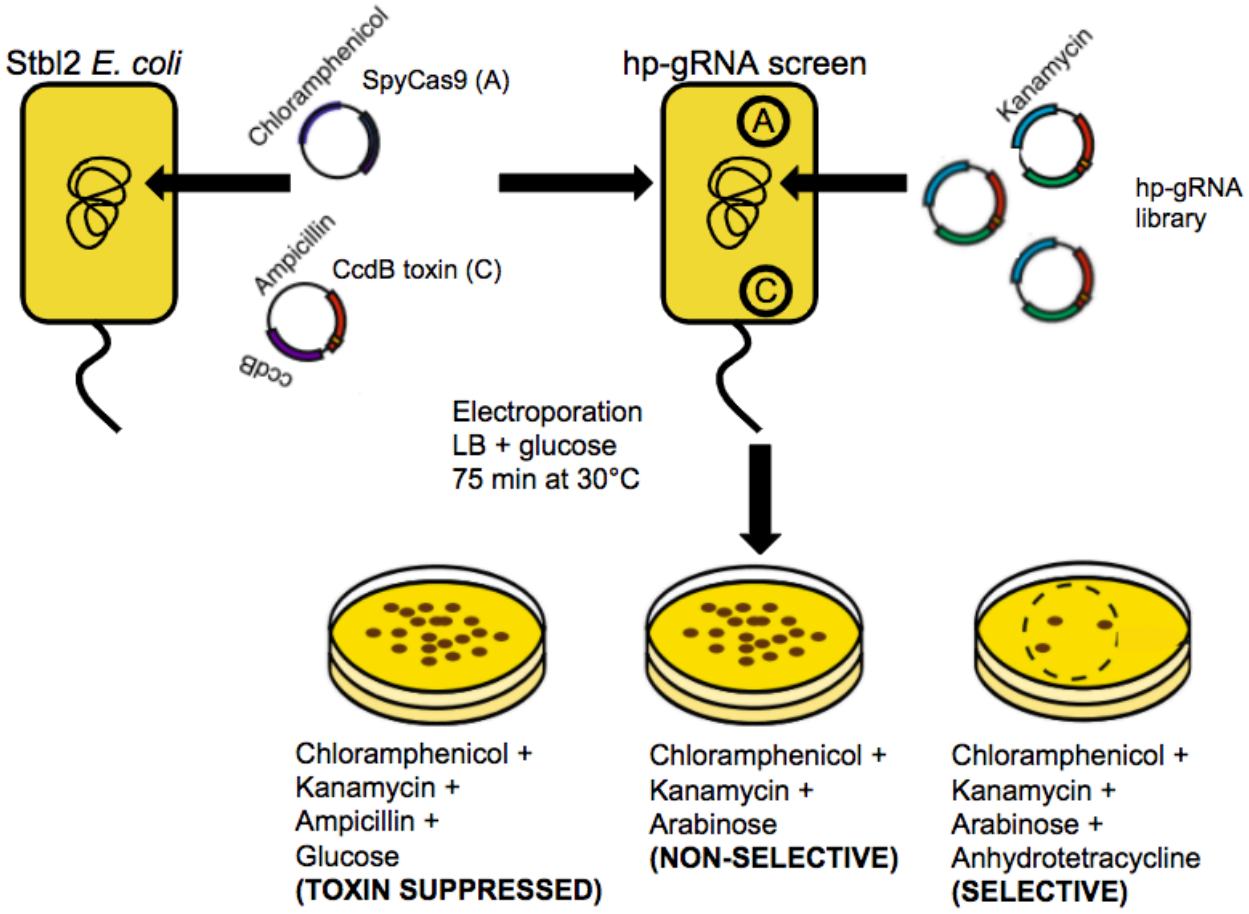



Figure 18. **Modification of Sniper-screen protocol to select for only highly specific hp-gRNAs**. The hairpin library transformants are then plated in selective, non-selective, and toxin suppressed conditions to screen for the highly specific hp-gRNA variants, those that do not disrupt on-target activity but suppress activity at off-targets.

CHAPTER II: A PLATFORM FOR SCREENING OF SPYCAS9 ACTIVITY

## AIM

To develop a method for high-throughput screening of gRNAs with enhanced specificity, we first developed a platform to screen for SpyCas9:gRNA cleavage activity. To do so, we engineered a strain of *E. coli* that allowed for inducible, simultaneous expression of SpyCas9 and a gRNA targeted to a plasmid containing a gene for a toxin (CcdB) in *E. coli*. In *E. coli*, plasmids cleaved by Cas9 are rapidly degraded, and so this "survival by on-target cleavage" assay provides a positive selection method for directly detecting SpyCas9 activity while confirming the ability of the CcdB protein to kill the *E. coli*. We used a gRNA with a well-characterized spacer targeting the human *EMX1* gene and cloned the EMX1 protospacer and flanking sequence into the plasmid with *ccdB*. This initial assessment of on-target activity is essential in providing a baseline for the targetability of the guide containing the EMX1 spacer, as all future experiments also require on-target activity. We found that this approach would allow for strong selection of gRNAs based on their ability to cleave the target, the first essential step in a screen for highly specific gRNAs. Here, we successfully developed a screen for baseline SpyCas9 on-target activity.

## METHODS

### PLASMID CONSTRUCTION

#### Construction of aTc-inducible SpyCas9 (pSpyCas9)

SpyCas9 was PCR amplified from pwtCas9-bacteria (Addgene #44250)[27] and inserted via HIFI assembly (NEB #E5520S) into a vector containing a chloramphenicol resistance marker, pBbA2c-RFP (Addgene #35326), along with a gBlock purchased from Integrated DNA Technologies (Coralville, IA) containing an anhydrotetracycline inducible promoter PLtetO-1.

The resulting construct was transformed into competent NEBDH5α competent *E. coli* cells. We named this plasmid "pSpyCas9" (Fig. 19)).

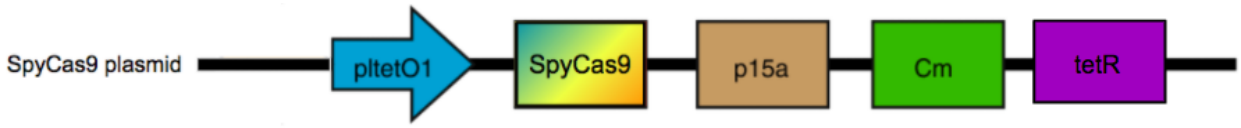



Figure 19. **Expression cassette for aTc inducible SpyCas9**. SpyCas9 is under the pLTetO-1 promoter with a chloramphenicol resistance marker and a p15a origin of replication, yielding a relatively low copy number.[37]

**Construction of aTc-inducible gRNA cassette (pgRNA)**

The gRNA cassette for transcription under anhydrotetracycline control was constructed by PCR amplification of the vector backbone containing a kanamycin resistance marker, pBbS2k-RFP (Addgene #35330), and inserting a gBlock containing a golden gate assembly site and the PLtetO-1 promoter via HIFI assembly. This product was then PCR amplified and the EMX1 spacer site was phosphorylated, annealed, and ligated with the PCR amplified vector containing kanamycin resistance and the pL-tetO-1 promoter. This plasmid for expressing gRNA was named "pgRNA", and when the gRNA spacer sequence targeted the EMX1 site was named "pgRNA-EMX1" (Fig. 20).

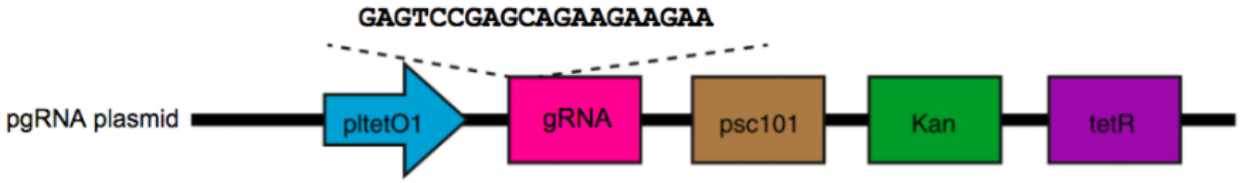



Figure 20. **Expression cassette for aTc inducible guide RNA containing the EMX1 spacer**. The gRNA-EMX1 is under the pLTetO-1 promoter with a kanamycin resistance marker and a psc101 origin of replication, yielding a low copy number.

**Construction of CcdB toxin plasmid with on-target site (pToxin-EMX1)**

A gBlock containing the EMX1 target site with its 15 flanking bp in its human genomic context (Integrated DNA Technologies) was inserted via Gibson assembly into a XbaI and SphI double digested vector containing ampicillin resistance, p11.LacY.wtx1 (Addgene #69056). The vector provided an arabinose inducible CcdB toxin protein under a pBAD promoter. The original p11.LacY.wtx1 plasmid is referred to here as "pToxin" with no EMX1 target site and "pToxin-EMX1" when it contains the EMX1 target site (Fig. 21).

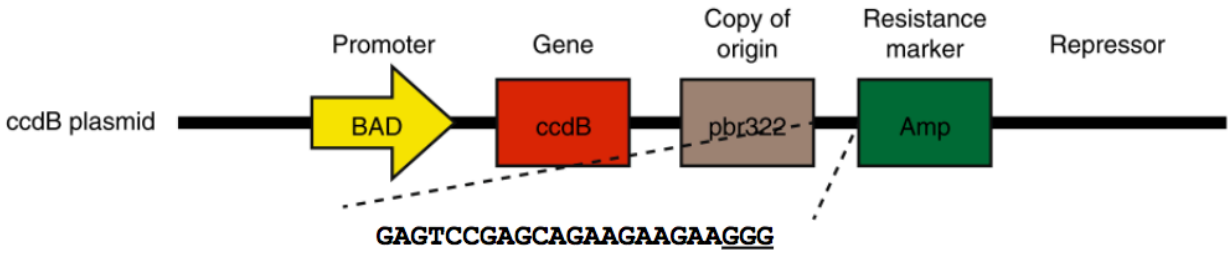



Figure 21. **Expression cassette for the CcdB toxin under a pBAD promoter**. The plasmid containing an arabinose-inducible, glucose-repressible CcdB toxin also contains an ampicillin resistance marker and a pBR322 origin of replication, yielding a much higher copy number than other plasmids used in this experiment (Table 1).

| Plasmid | Origin of replication | Resistance marker | Key components |
|---|---|---|---|
| pSpyCas9 | p15A, copy number ~10 | chloramphenicol | pL-tetO-1 promoter, TetR |
| pgRNA-EMX1<br>pgRNA-EMX1$^{Target}$<br>pgRNA-EMX1$^{OT1}$<br>pgRNA-EMX1$^{noPAM}$<br>pgRNA-EMX1$^{OTLibrary}$<br>pgRNA-EMX1$^{hp1OT1}$<br>pgRNA-EMX1$^{hp4OT1}$ | pSC101, copy number ~5 | kanamycin | pL-tetO-1 promoter, TetR |
| pToxin<br>pToxin-EMX1 | pBR322,copy number ~15-20 | ampicillin | arabinose inducible, glucose repressible, CcdB toxin |
| pTransposon(OT1-6) | pSC101, copy number ~5 | ampicillin | Tn7 insertion capability |

Table 1. **Plasmid construct variants and their origin of replication, resistance markers, and key regulatory components**.

## GENERATION OF BACTERIAL STRAINS

To prevent undesired recombination of similar DNA sequences, all bacterial strains were derived from MAX Efficiency Stbl2 Competent Cells (Invitrogen #10268-019). For this screen, two

bacterial strains were generated to assess SpyCas9:gRNA activity in *E. coli*. First, pSpyCas9 was transformed into Stbl2, resulting in Strain MR001. This provided the base strain for all subsequent plasmid insertions into Stbl2 containing an aTc inducible SpyCas9.

Once pSpyCas9 was transformed into Stbl2, strain MR001 was cultured in LB medium and chloramphenicol overnight. Then 35 µL of overnight growth was grown in 3 mL of fresh LB until reaching mid-exponential phase, an $OD_{600}$ of ~0.6 for electrocompetent cell production, using a "small batch" protocol adapted from the Barrick lab. All subsequent transformation products were made into electrocompetent cells for plasmid uptake using this method, using appropriate resistance markers and suppressing the CcdB toxin with 0.1% glucose when the toxin plasmid was present.

Electrocompetent cells were combined with ~100-150 ng of plasmid on ice for 5 minutes, transferred to a 1 mm electroporation cuvette and electroporated by an 1800V shock. Then 950 µL LB containing 0.1% glucose, warmed to 30ºC, was added within ten seconds of the pulse for recovery. Cells were recovered at 30ºC for approximately 75 minutes, and 70 µL of recovery growth was plated on the appropriate antibiotic with 0.1% glucose if the toxin plasmid was present. All

To screen for SpyCas9:pgRNA-EMX1 scission of the toxin plasmid containing no EMX1 site, pToxin was transformed into an electrocompetent strain MR001. Upon successful transformation, plasmid pgRNA-EMX1 was then transformed into the strain, resulting in Strain MR003 (Table 2). Screening for SpyCas9:pgRNA-EMX1 scission of the toxin plasmid with the EMX1 site, pToxin-EMX1 was transformed into an electrocompetent strain MR001. pgRNA-EMX1 was then transformed into the strain, resulting in Strain MR005 (Table 2).

| Base strain | 1st transformation | 2nd transformation | 3rd transformation | Strain Name |
|---|---|---|---|---|
| Stbl2 | pSpyCas9 | | | MR001 |
| Stbl2 | pSpyCas9 | pToxin | | MR002 |
| Stbl2 | pSpyCas9 | pToxin | pgRNA-EMX1 | MR003 |
| Stbl2 | pSpyCas9 | pToxin-EMX1 | | MR004 |
| Stbl2 | pSpyCas9 | pToxin-EMX1 | pgRNA-EMX1 | MR005 |

Table 2. **Strain generation via multiple transformations for on-target activity screening**.

## SCREENING BY SELECTIVE PLATING

Both Strain MR003 and Strain MR005 were plated in conditions where the CcdB toxin is expressed, and then either in the presence (selective conditions) or absence (non-selective conditions) of SpyCas9 and gRNA expression. The presence of aTc (100 ng/mL) induced the SpyCas9 and gRNA while they were selected with their plasmid resistance markers and arabinose (1.5 mg/mL) promoted expression of the CcdB lethal protein. In conditions where aTc was not present, SpyCas9 and gRNA were not expressed. These plating conditions allowed for direct assessment of SpyCas9:gRNA complex scission of pToxin or pToxin-EMX1. To assess transformation efficiency, the strains were plated in the presence of the resistance markers for all 3 plasmids of interest while suppressing the CcdB toxin with 0.1% glucose (toxin suppressed conditions).

| Selective conditions | Non-selective conditions | Toxin suppressed conditions |
|---|---|---|
| Kanamycin<br>Chloramphenicol<br>Arabinose<br>anhydrotetracycline | Kanamycin<br>Chloramphenicol<br>Arabinose | Kanamycin<br>Chloramphenicol<br>Ampicillin<br>Glucose |

Table 3. **Selective plating conditions for assessment of SpyCas9 activity**.

## RESULTS AND DISCUSSION

This assay screened for SpyCas9 on-target activity by exposing the SpyCas9:gRNA complex to either toxin plasmid with the EMX1 site or without the EMX1 site. Here, survival of the *E. coli* is indicative of SpyCas9 activity. In the presence of pToxin-EMX1, selective conditions where Cas9 and the gRNA were induced along with the toxin allowed for SpyCas9:gRNA-EMX1 scission of the EMX1 site, disabling production of CcdB and allowing for the bacteria to survive. A great deal of colony growth is observed here (Fig 22.), as most the *E. coli* survived. It was also observed that in the presence of pToxin-EMX1, the non-selective conditions resulted in very little colony growth due to repressed SpyCas9:gRNA-EMX1 expression and expressed toxin: here the CcdB protein was expressed and the *E. coli* did not survive. If the CcdB plasmid did not contain the EMX1 protospacer, the SpyCas9:gRNA-EMX1 complex was unable to cleave the plasmid due to the absence of a recognition site, thus allowing the lethal protein to kill the *E. coli*. This occurred in both selective conditions and non-selective conditions where CcdB expression was induced by arabinose, despite the absence or presence of the SpyCas9:gRNA-EMX1 complex. The colony count assay from this screening method suggests that this is a viable platform to selectively screen for SpyCas9 and gRNA on-target activity in *E. coli*.

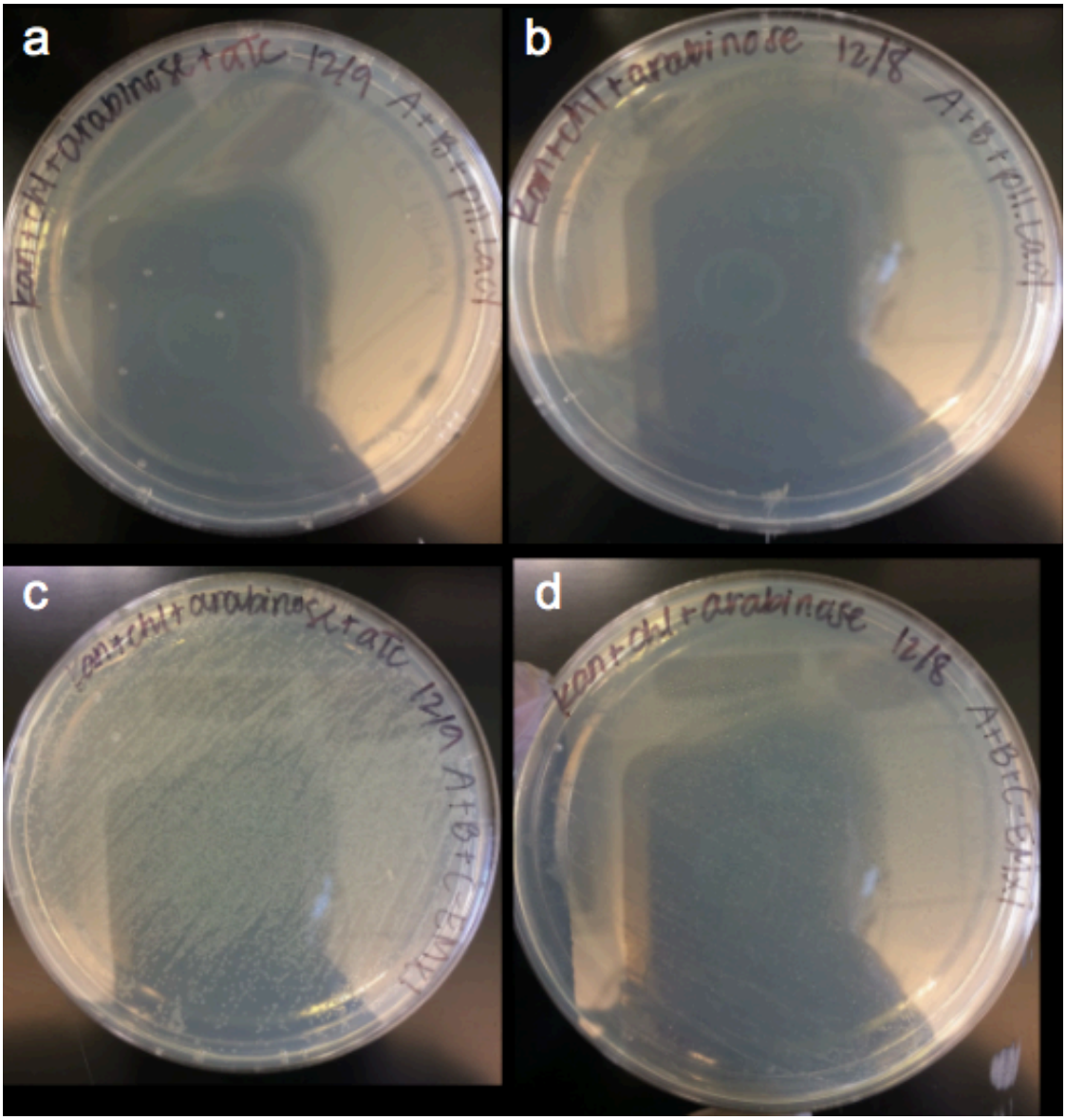



Figure 22. **Screening of SpyCas9 on-target activity by colony count**. (a) Strain MR003 in selective conditions (b) Strain MR003 in non-selective conditions (c) Strain MR005 in selective conditions (d) Strain MR005 in non-selective conditions. Growth was only observed under selective conditions when the Cas9, gRNA, and toxin are expressed if the targeted site is present on the toxin plasmid.

# CHAPTER III: SCREENING FOR OFF-TARGET ACTIVITY

## AIM

The development of an unbiased screening of pgRNA-EMX1 for off-target activity was essential in determining the specificity of gRNAs screened using this assay. For the first part of the screening, the targetability of the guide resulting from the pgRNA-EMX1 cassette with off-target 1 and the target without a PAM (and so not recognizable by the Cas9) was characterized. These off-targets were originally placed into the *E. coli* genome via transposon 7 (Tn7) mediation, where off-target activity would result in cell death, but later placed on the plasmid expressing the gRNA (self-targeting assay), where the *E. coli* could no longer survive in the presence of the antibiotic that the plasmid provided resistance towards. This assay tests the ability of pgRNA-EMX1 to successfully cleave the target site of pToxin-EMX1, while not cleaving sites within itself, thus preserving the pgRNA-EMX1 plasmid.

The second part of the off-target screen characterized the ability of SpyCas9:gRNA-EMX1 complex to cut the EMX1 target site on pToxin-EMX1 while in the presence of 15 prominent off-targets and a target with no PAM site via small-scale NGS. After screening, the reduction in individual off-target sequences from the pool quantifies the ability of the pgRNA-EMX1 spacer cassette to recognize and cut similar sequences to the target site. Here, both the plasmid-based screen and the library-based NGS screen show that we can effectively screen simultaneously for SpyCas9:gRNA on and off-target activity.

# METHODS

## PLASMID CONSTRUCTION

### Construction of cassette for genomic integration of off-targets (pTransposon(OT1-6))

A gBlock containing the off-target array of the six most prevalent off-target sites (Integrated DNA Technologies) surrounded by 15 bp of their human genomic context was inserted via HIFI assembly into a NotI and XhoI digested pGRG36 vector (Addgene #16666). This vector contains Tn7 transposon recombination machinery that allows for direct integration of the EMX1 off-target array into the *E. coli* genome. This plasmid construct was named "pTransposon(OT1-6)" (Table 1).

### Construction of self-targeting pgRNA expression plasmids (pgRNA-EMX1$^{OT1}$ and pgRNA-EMX1$^{noPAM}$)

The pgRNA-EMX1 plasmid was PCR amplified to create a vector containing the EMX1 spacer, psc101 origin of replication, kanamycin resistance marker, pL-tetO-1 promoter, and TetR. Ultramer® DNA oligos (Integrated DNA Technologies) containing either the off-target site or the target site with a mutated PAM were inserted into the new pgRNA-EMX1 backbone via HIFI assembly (NEB). These plasmids were named "pgRNA-EMX1$^{OT1}$" and "pgRNA-EMX1$^{noPAM}$", respectively (Fig. 23).

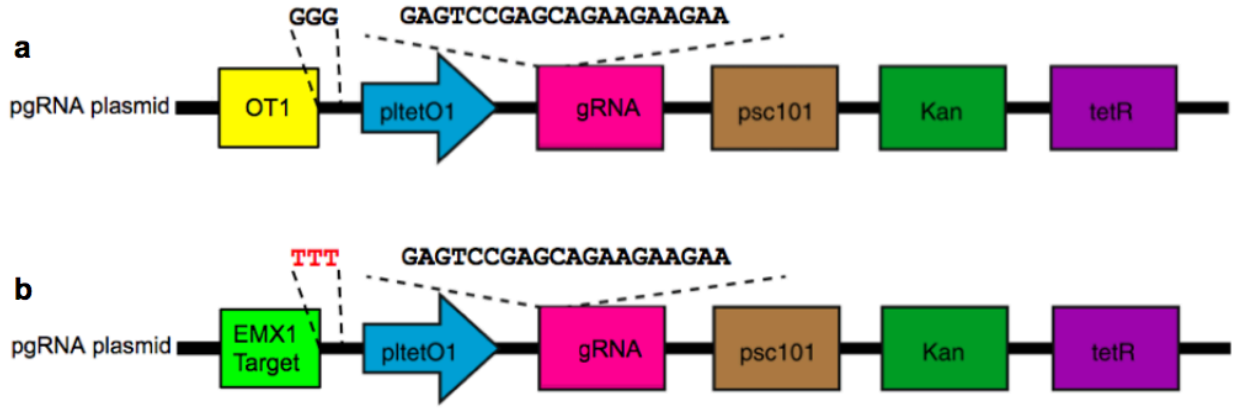



Figure 23. **Expression cassettes of pgRNA-EMX1 variants**. (a) The pgRNA-EMX1$^{OT1}$, with the correct 5'-GGG-3' PAM (b) pgRNA-EMX1$^{noPAM}$ with the mutated 5'-TTT-3' PAM (shown in red).

**Construction of pgRNA expression plasmid library containing an off-target pool (pgRNA-EMX1$^{OTLibrary}$)**

The 15 most prevalent off-target sites in the human genome for the EMX1 gene were previously identified by the Tsai group.[28] The upstream and downstream chromosomal context per off-target site was identified and an oligonucleotide pool containing the 15 off-targets and the true target site with no PAM was PCR amplified and inserted into the pgRNA-EMX1 backbone via HIFI assembly (NEB). This plasmid construct was named "pgRNA-EMX1$^{OTLibrary}$" (Fig. 24).

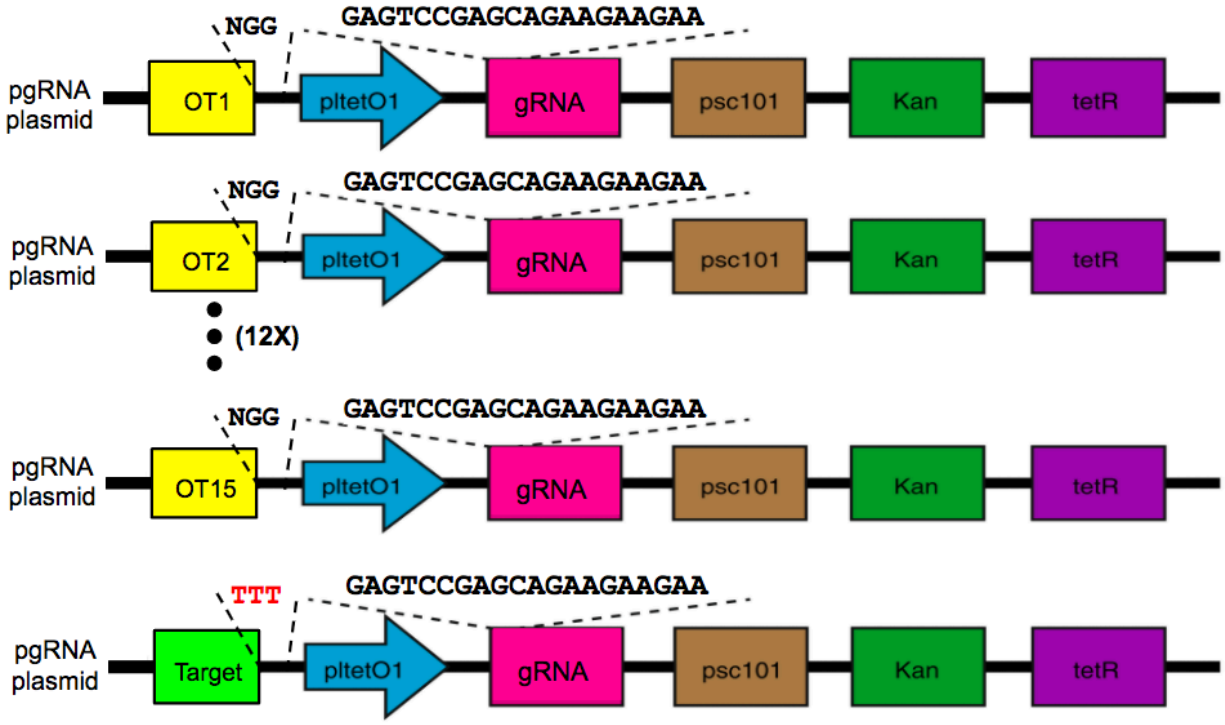



Figure 24. **Expression cassette for the library of 15 off-targets and the target without a SpyCas9 recognizable PAM**. This insertion of these cassettes into the pgRNA plasmid resulted in a library of sixteen plasmids components containing 15 known off-targets for the EMX1 gRNA and one target lacking a PAM as a negative control.

## GENERATION OF BACTERIAL STRAINS

### 1) Plasmid based screening for off-target activity

First, pSpyCas9 and pToxin-EMX1 were transformed into Stbl2, providing a baseline environment of SpyCas9 and target DNA in the *E. coli*, known as strain MR004. After the third transformation each pgRNA variant, pgRNA-EMX1, pgRNA-EMX1$^{Target}$, pgRNA-EMX1$^{OT1}$, and pgRNA-EMX1$^{noPAM}$, was introduced. This resulted in the strains MR005, MR006, MR007, and MR008. Strains were generated via multiple sequential transformations, as shown in Table 4 below.

| Base strain | 1st transformation | 2nd transformation | 3rd transformation | Strain Name |
|---|---|---|---|---|
| Stbl2 | pSpyCas9 | | | MR001 |
| Stbl2 | pSpyCas9 | pToxin-EMX1 | pgRNA-EMX1 | MR005 |
| Stbl2 | pSpyCas9 | pToxin-EMX1 | pgRNA-EMX1$^{Target}$ | MR006 |
| Stbl2 | pSpyCas9 | pToxin-EMX1 | pgRNA-EMX1$^{OT1}$ | MR007 |
| Stbl2 | pSpyCas9 | pToxin-EMX1 | pgRNA-EMX1$^{noPAM}$ | MR008 |
| Stbl2 | pSpyCas9 | pToxin-EMX1 | pgRNA-EMX1$^{OTLibrary}$ | MR009 |

Table 4. **Generation of bacterial strains for SpyCas9 plasmid based and library based off-target activity screenings.**

**2) Library based screening for off-target activity**

Screening of SpyCas9:gRNA with the EMX1 spacer for off-target effects against a library of off-targets containing genomic context was accomplished by introducing pgRNA- EMX1$^{OTLibrary}$ (sequences shown in Table 6) into an *E. coli* environment containing pSpyCas9 and pToxin-EMX1. Survivors of selective and toxin-suppressed conditions were sequenced in duplicate by a small-scale next-generation sequencing (~150,000 reads per experiment) of the PCR products overlapping the off-target sequences performed by Genewiz, Inc.

| Off-target site | Off-target sequence with genomic context | Location in human genome (Chr, start-end) |
|---|---|---|
| On-target | AGGAGGAAGGGCCT**GAGTCCGAGCAGAAGAAGAAGGG**CTCCCATCACATCAA | Chr: 2 73160981-73161004 |

| | | |
|---|---|---|
| 1 | CAAGACAGATTGTCA**GAGTTAGAGCAGAAGAAGAAAGG**CATGGAGTAAAGGCA | Chr: 5<br>45359060-<br>45359083 |
| 2 | ATTCATAGTAGACAA**GAGTCTAAGCAGAAGAAGAAGAG**AGCCACTACCCAACC | Chr: 15<br>45359083-<br>44109769 |
| 3 | TCTTCTGCAAATGAG**GAGGCCGAGCAGAAGAAAGACGG**CGACAGATGTTGGGG | Chr: 2<br>219845055-<br>219845078 |
| 4 | GGGCCAGCATGACCT**GAGGCCGAGCAGGAGAAGAAGAG**GCAGCCTAGAGTCTT | Chr: 8<br>128801241-<br>128801264 |
| 5 | ATAAACATGCTAACA**AAGTCTGAGCACAAGAAGAATGG**TGAGAAGGAATACAT | Chr: 5<br>9227145-<br>9227168 |
| 6 | AGGATGGGTGGTGAG**GAGTCCGGGAAGGAGAAGAAAGG**CTCAGCGCGGCTTGC | Chr: X<br>53467704-<br>53467727 |
| 7 | GAGGAAGGCGCGGGC**GAGCCGGAGCAGAAGAAGGAGGGG**AGGGAGCCAGCCGCT | Chr: 5<br>146833183-<br>146833206 |
| 8 | GCCCAACTCCTGTAG**AAGTCCGAGGAGAGGAAGAAAGG**GTTCTGGAGCTCTCA | Chr: 1<br>23720611-<br>23720634 |
| 9 | GCAAGGGTCTCAGGG**GAATCCAAGCAGGAGAAGAAGGA**GGGAAAAACCACTCT | Chr: 3 |

| | | |
|---|---|---|
| | | 5031597-<br>5031620 |
| 10 | ATACTTTCAGACAAA**AC**GTC**T**GAGCAGAAGAAGAA**T**GGACAGAACTCTGAGGA | Chr: 6<br>9118792-<br>9118815 |
| 11 | GCACAGAGGAGGAGG**GAGTAG**GAGCAG**G**AGAAGAAGG**A**GGAAGATGACCAGGA | Chr: 13<br>27769640-<br>27769663 |
| 12 | ATTTCTTGTGGAAGG**A**AGTCC**CG**GCAGA**G**GAAGAAGGGGCTGAGGCTGTGAGC | Chr: 15<br>100292461-<br>100292484 |
| 13 | AAAGAGTAAGTATGA**TCA**TCC**A**AGCAGAAGAAGAAG**A**GAAGGATTTTGGCAGT | Chr: 3<br>95690179-<br>95690202 |
| 14 | GGGATGCCAGTGAAG**GAGTCTA**AGCAG**G**AGAA**T**AA**A**GGGTCAGGTTCGTGTCT | Chr: 2<br>218378101-<br>218378124 |
| 15 | CATGTCGTATGGCCA**GAGCA**CGAGCA**AG**AGAAGAAGGGAGGCTACCACAACAC | Chr: 10<br>58848711-<br>58848734 |
| Target-<br>noPAM | GAGGAGGAAGGGCCT**GAGTCCGAGCAGAAGAAGAATTT**CTCCCATCACATCAA | N/A |

Table 5. **Sequences of off-target sites and target sites introduced into pgRNA-EMX1 for screening of off-target activity of SpyCas9**. Mismatches in off-target sites are underlined and genomic locations are given.

### 3) Insertion of EMX1 1-6 off-target arrays into NEBDH5α

The insertion of transgene technique designed by McKenzie and Craig was utilized.[43] After confirmation of successful integration of the off-target array into pGRG36 via sequencing, the pTransposon(OT1-6) was transformed into NEBDH5α cells and the plasmid was selected for on LB and ampicillin (100 μg/mL) at 32°C. After selection, three colonies were chosen, streaked out on LB and ampicillin, then inoculated overnight at 32°C without antibiotic to initiate the loss of plasmid. Overnight cultures were plated on LB and grown at 42°C to block plasmid replication, then streaked out on LB again and grown at 42°C to ensure complete loss of plasmid. This loss of plasmid was confirmed by PCR amplification of the genomic segment flanked by the Tn7 attachment site, followed by gel electrophoresis to confirm PCR product size.

## SCREENING FOR OFF-TARGET ACTIVITY

### 1) Plasmid based screening for off-target activity

Each variant of pgRNA was transformed, pgRNA-EMX1, pgRNA-EMX1$^{Target}$, pgRNA-EMX1$^{OT1}$, and pgRNA-EMX1$^{noPAM}$ into electrocompetent strain MR004. The final transformation products, MR005, MR006, MR007, and MR008 were then plated in the same Selective conditions, non-selective conditions, and toxin suppressed conditions as shown in Table 3.

### 2) Library based screening for off-target activity

To screen for the ability of SpyCas9:pgRNA-EMX1 to target and cleave the fifteen top off-targets, 100 ng of pgRNA-EMX1$^{OTLibrary}$ was transformed into electrocompetent strain MR004, resulting in strain MR009, as shown in Table 5. Instead of plating transformants, recovery growth was diluted by a factor of 50, and the transformation product was grown in selective conditions or toxin suppressed conditions (Table 3) in 7 mL of LB media. Overnight growth was miniprepped (NEB), and the resulting plasmids from the screen were PCR amplified with Illumina® adapters for NGS.

## RESULTS AND DISCUSSION

### 1) Off-target arrays integrated into the *E. coli* genome by Tn7 were unstable

Initially, an attempt was made to integrate the off-target array of 6 of the off-targets with the highest known off-target activity into the genome of the NEBDH5α *E. coli* strain. This would allow for a one-step positive and negative selection screen for guide RNA specificity, as a break in the *E. coli* genome would result in immediate cell death and off-targeting guides would be dropped out of the population. Sequencing results showed that the off-targets similarities caused recombinogenic events that resulted in the integration of only a small segment of this array.

The nucleotide sequence of the array that should have been integrated (318 bp):

AGGATGGGTGGTGAGGAGTCCGGGAAGGAGAAGAAAGGCTCAGCGCGGCTTGCATGTATTCCTT
CTCACCATTCTTCTTGTGCTCAGACTTTGTTAGCATGTTTATAAGACTCTAGGCTGCCTCTTCT
TCTCCTGCTAGGACTCAGGTCATGCTGGCCCCCCCAACATCTGTCGCCGTCTTTCTTCTGCTCG
GCCTCCTCATTTGCAGAAGAGGTTGGGTAGTGGCTCTCTTCTTCTTCTGCTTAGACTCTTGTCT
ACTATGAATCAAGACAGATTGTCAGAGTTAGAGCAGAAGAAGAAAGGCATGGAGTAAAGGCA

The actual integration product (53bp):

AGGATGGGTGGTGAGGAGTCCGGGAAGGAGAAGAAAGGCATGGAGTAAAGGCA

The similarity of the sequence highlighted in red suggests this truncated integration product was a result of recombination. The recombinogenic event occurred in NEBDH5α, with the following genotype,

**$F^-$ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17($r_K^-m_K^+$), $\lambda^-$**,

which is known to be recombination deficient. However, since these off-targets had a high degree of similarity, in some cases only differing by one nucleotide, this particular strain was not

able to accommodate the array. To prevent recombination events, the strain Stbl2 (Invitrogen), with the genotype,

**$F^-$ endA1 glnV44 thi-1 recA1 gyrA96 relA1 Δ(lac-proAB) mcrA Δ(mcrBC-hsdRMS-mrr) $\lambda^-$**,

was selected for all subsequent screens involving off-target sites due to the large degree in sequence similarity between experimental components. Stbl2 is known to be a better host for unstable sequences containing direct repeats.

**2) Plasmid-based screening of off-target activity**

Since the integration of off-targets into the *E. coli* genome was unsuccessful, we then incorporated the off-target with the highest number of predicted hits into pgRNA-EMX1. Selective screening of targetability of pgRNA-EMX1 (Fig. 25a,d,g), pgRNA-EMX1$^{noPAM}$ (Fig. 25b,e,f), pgRNA-EMX1$^{OT1}$ (Fig. 25c,f,i), and pgRNA-EMX1$^{Target}$ resulted in widespread colony growth in all conditions except for with pgRNA-EMX1$^{OT1}$ in growth selective conditions (Fig. 25c). This result indicates that in selective conditions the SpyCas9:gRNA-EMX1 complex was able to cut pToxin-EMX1 when the gRNA cassette contained only the EMX1 spacer or contained both the EMX1 spacer and the target with no PAM. In the case where the gRNA cassette contained both the EMX1 spacer and off-target site 1, the plasmid had self-targeting ability and therefore did not cut pToxin-EMX1. The results from non-selective conditions in this assay indicate that when SpyCas9 and the gRNA with the EMX1 spacer are not induced, there is much more *E. coli* survival than expected, as pToxin-EMX1 should not be cleaved in this case. This experiment was carried out twice and the survival results for all pgRNA variations and all conditions were very similar.

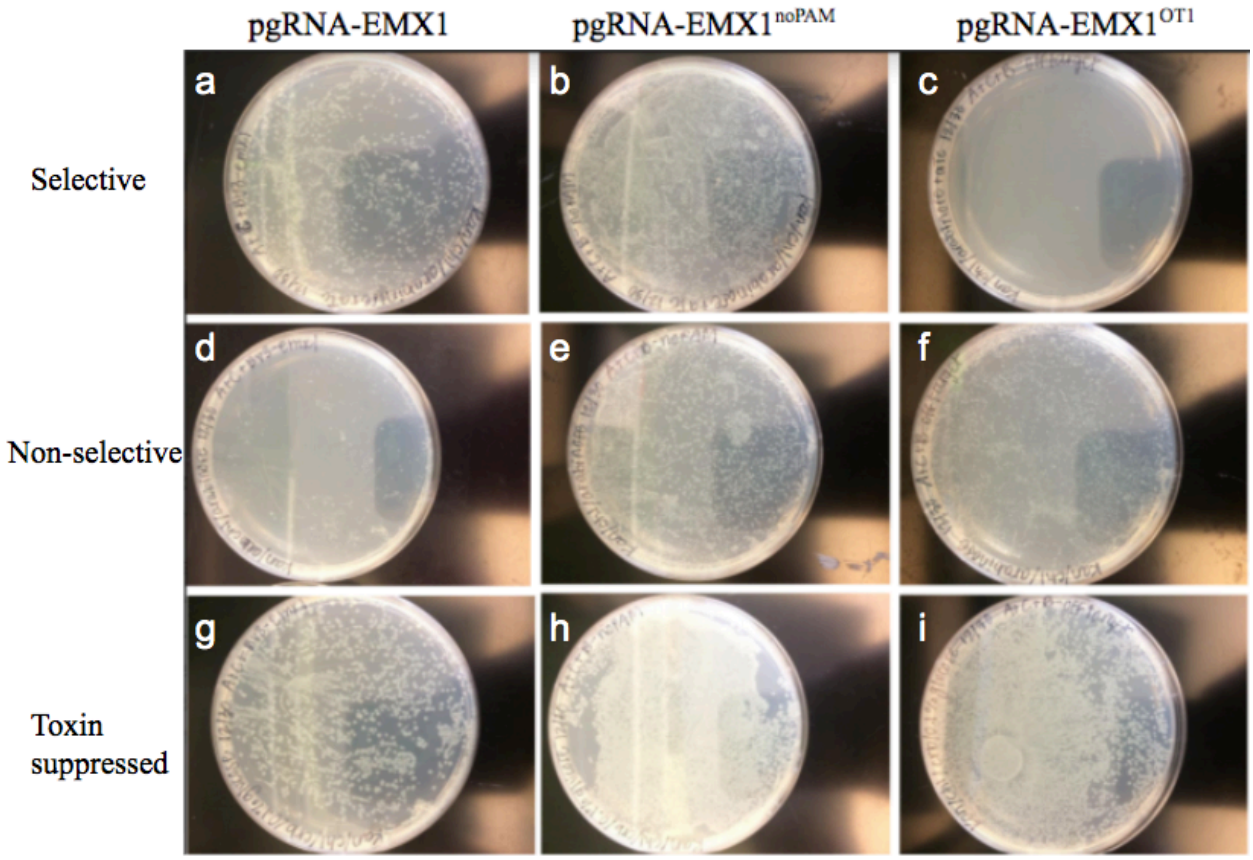



Figure 25. **Colony count assay for off-target activity**. (a,d,g) pgRNA-EMX1 cassette, (b,e,h) pgRNA-EMX1 cassette with EMX1 site without the 5'-GGG-3' PAM, and (c,f,i) the pgRNA-EMX1 cassette with off-target site 1. This assay shows that in selective conditions pgRNA-EMX1$^{OT1}$ targets and cuts itself, resulting in *E. coli* death. Non-selective conditions showed a basal level of *E. coli* survival.

### 3) Library-based screening for off-target activity via small scale Next Generation Sequencing

Results from the screening for relative SpyCas9 activity against all 15 potential off-targets, ranked from OT1 to OT15, with OT1 being the off-target site with the highest known mutational activity from published experiments using human cells and human genomic DNA, normalized to a target without a PAM site showed a strong correlation between experimental duplicates. As expected, transformants that grew overnight in toxin suppressed conditions resulted in a significantly higher count of normalized reads for nearly every off-target site than transformants

that grew overnight in selective conditions, except for the target with no PAM (Fig. 26). Here, the presence of pgRNA-EMX1$^{noPAM}$ provided the control, since the plasmid should not be recognized by SpyCas9. These results indicate that all off-targets were recognized and cut by SpyCas9 and gRNA-EMX1, as the prevalence of their individual sequences were depleted in conditions where SpyCas9 and the gRNA were induced. These results also show how the various off-target sequences have different recognition and scission rates by SpyCas9. Interestingly, off-targets 9, 11, 13, and 14 did not exhibit as much depletion in reads in comparison to the other off-target sites.

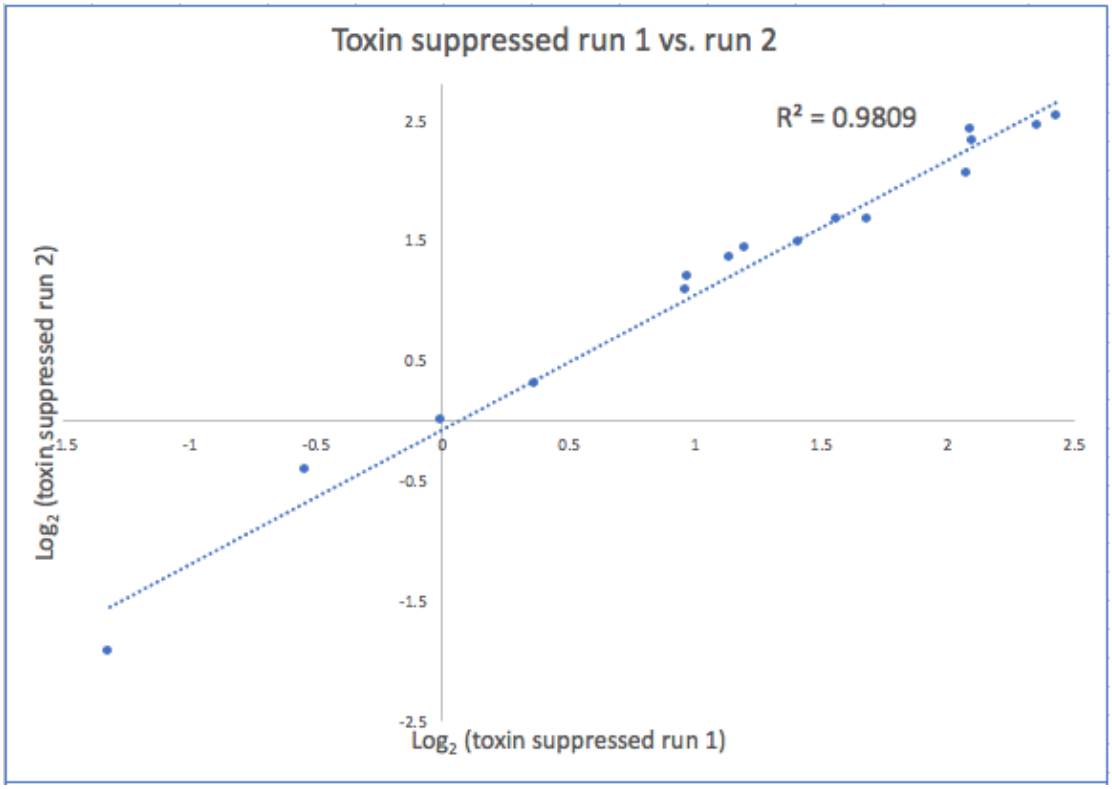



Figure 26. **Correlation between experimental duplicates of screening for off-target activity in toxin suppressed conditions**. The correlation of normalized reads of each off-target sequence present in the toxin-suppressed populations between duplicates is approximately 0.99, indicative of a very strong correlation.

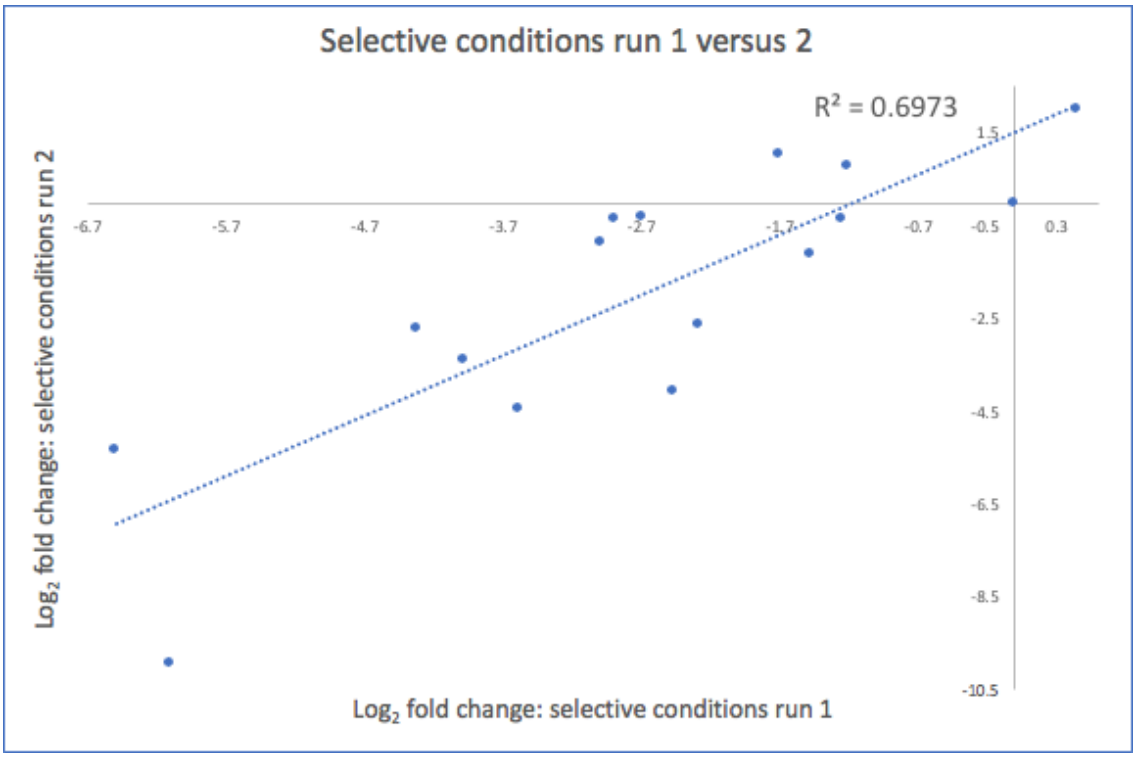



Figure 27. **Correlation between experimental duplicates of screening for off-target activity in selective conditions**. The correlation between the normalized change in the off-target sequence reads across experimental duplicates is approximately 0.70, indicative of a significant positive relationship and a strong correlation.

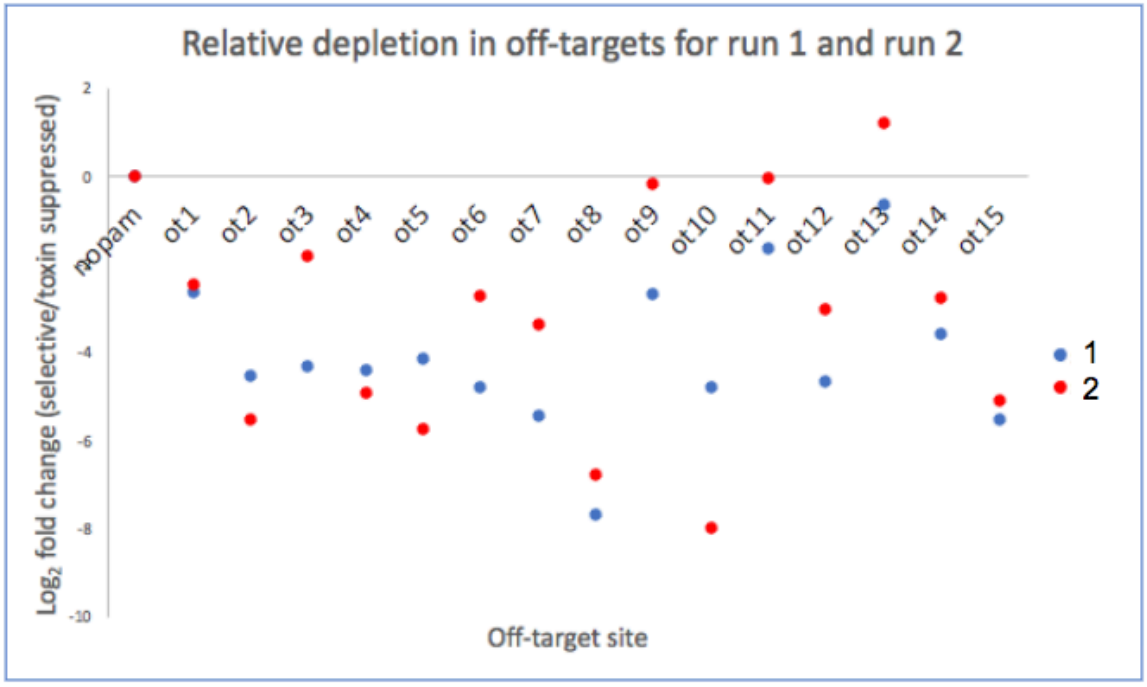



Figure 28. **Depletion in off-target reads before and after screening for SpyCas9:gRNA-EMX1 off-target activity**. Depletion is represented her by the $\log_2$ fold change in reads for all 15 off-targets and the target with no PAM. For both trial 1 and 2, there was a significant reduction in read counts for most of the off-target sites.

Results from this assay indicate that screening for on and off-target activity is possible via the plasmid-based screen. We found that we are also able to eliminate most of the library of prominent off-target sequences from the population relative to the target site with no PAM that should not be recognized at all by the SpyCas9. This means that SpyCas9 and the guide RNA of pgRNA-EMX1$^{\text{OTLibrary}}$ were able to reproducibly recognize and cleave off-targets as expected. If deeper sequencing were available for the library screen rather than the small-scale (~150,000 reads per experimental condition), we would likely have more consistency across replicates and less variation. This illustrates the nearly the full off-target effect profile of SpyCas9 guided by gRNA-EMX1, a system relatively lacking in specificity. In the next chapter, we attempt to enhance the specificity of the gRNA component against off-target site 1 with the introduction of a randomized hairpin library.

# CHAPTER IV: SCREENING OF HAIRPIN LIBRARIES

## AIM

Having created a screen with both positive (no survival without on-target activity) and negative (no survival in the instance of off-target activity) capabilities, we investigated the targeting effect of gRNAs with a PAM distal hairpin extension, hp-gRNAs. The introduction of a randomized library of $4^8$ hp-gRNAs (with an 8 nt extension) into this SpyCas9 system was hypothesized to identify which hairpin structures are associated with high prevalence of only on-target cleavage. NGS reads of hp-gRNAs in selective conditions should provide the optimal hp-gRNA candidates for therapeutic gene editing. To gain insight on the properties that dictate specificity of hp-gRNAs, the surviving hairpins were characterized. Some potential key properties of hp-gRNAs include, secondary structure, free energy, binding to the seed region, purine abundance, and sequence features. The impact of these properties of 5' hairpin extended guides on SpyCas9 DNA targeting can be further understood by identifying hp-gRNA sequences that maintain SpyCas9 on-target activity and weaken off-target activity. Our hairpin library screens enabled the identification of various hairpin structures *in vivo* and *in vitro* that warrant future investigation.

## METHODS

### CONSTRUCTION OF PLASMIDS

#### Construction of hp-gRNA library (pgRNA-EMX1$^{hp1OT1}$)

The pgRNA-EMX1 plasmid was PCR amplified to provide a vector backbone containing the EMX1 spacer, pSC101 origin of replication, kanamycin resistance marker, pL-TetO-1 promoter, and TetR. A gBlock (IDT) containing off-target site 1 along with a randomized 8-nucleotide long

hairpin library was inserted via HIFI assembly (NEB) to link the EMX1 spacer with a 5'-UUCG-3' tetraloop. This plasmid construct was named "pgRNA-EMX1$^{hp1OT1}$" (Fig. 29a).

**Construction of hp-gRNA library with multiple tetraloops (pgRNA-EMX1$^{hp4OT1}$)**

An oligonucleotide pool was ordered from IDT, containing off-target site 1 and a randomized 8-nucleotide hairpin library, in this case linked to the EMX1 spacer with four potential variations of tetraloops (TLs) to promote interactions between the randomized nucleotides and the spacer sequence; 5'-UUCG'3', 5'-GCAA-3', 5'-CUUG-3', or no tetraloop. This oligonucleotide pool was inserted into the same vector described above, also via HIFI assembly. This plasmid construct was named "pgRNA-EMX1$^{hp4OT1}$" (Fig 29b).

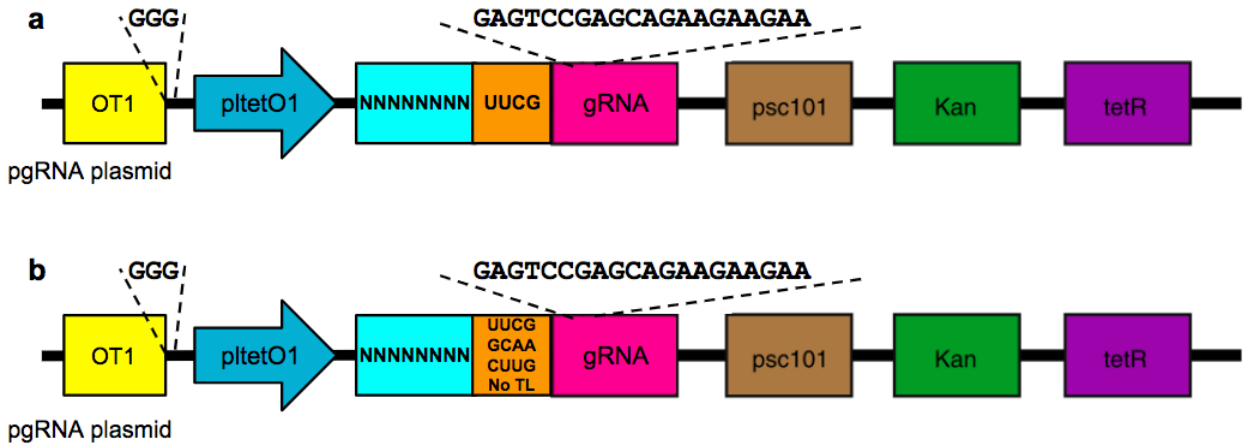



Figure 29. **Cassette for pgRNA-EMX1 hairpin library variants**. (a) the pgRNA-EMX1$^{hp1OT1}$ with an 8-nucleotide hairpin library (turquoise) connected to the spacer via a 5'-UUCG-3' TL motif (orange) (b) the pgRNA-EMX1$^{hp4OT1}$ also contains an 8-nucleotide hairpin library connected to the spacer via the 4 potential TL motifs.

## SCREENING OF HAIRPIN LIBRARIES

| Base strain | 1st transformation | 2nd transformation | 3rd transformation | Strain Name |
|---|---|---|---|---|
| Stbl2 | pSpyCas9 | pToxin-EMX1 | pgRNA-EMX1$^{hp1OT1}$ | MR010 |
| Stbl2 | pSpyCas9 | pToxin-EMX1 | pgRNA-EMX1$^{hp4OT1}$ | MR011 |

Table 6. **Generation of bacterial strains for hairpin library specificity screening**.

### 1) Screening of hairpin library with 5'-UUCG-3' tetraloop

Construct pgRNA-EMX1$^{hp1OT1}$ was transformed into Stbl2 containing pSpyCas9 and pToxin-EMX1, resulting in strain MR010. This strain was then plated in selective, non-selective, and toxin suppressed conditions and grown overnight at 32ºC. After plating, the remaining bacteria were diluted by a factor of 50 and grown overnight at 30ºC in selective conditions or toxin suppressed conditions (Table 3) in 7 mL of fresh LB media. Overnight growth was miniprepped (NEB), and the resulting plasmids from the screen were PCR amplified. The miniprepped overnight growth as well as colonies selected from selective conditions and non-selective conditions plates were PCR amplified for Sanger sequencing.

### 2) Screening of hairpin library with 4 potential tetraloops

Construct pgRNA-EMX1$^{hp4OT1}$ was transformed into MR004. This strain was then plated in selective, non-selective, and toxin suppressed conditions and allowed to grow overnight at 32ºC, as mentioned above. Recovery growth was diluted and inoculated overnight as mentioned above. Miniprepped overnight growth as well as 10 colonies from selective and non-selective conditions were also PCR amplified for Sanger sequencing. The screening of the four potential hairpin libraries was not deeply sequenced.

### 3) *In vitro* determination of specificity of isolated hairpin structures

Sanger sequencing of colonies surviving selective conditions from hairpin library with 4 potential TLs provided three hairpin structures, all with different tetraloop motifs and vastly different secondary structures. Guides with spacers targeting the EMX1 site with these hairpin extensions, along with one guide with no hairpin, were purchased from IDT for in vitro determination of their specificity. Short oligonucleotides containing the EMX1 target site and EMX1 off-target site 1, shown below, with human genomic context were also purchased (IDT).

EMX1-ON target sequence (Chr: 2):

ctatgtagcctcagtcttcccatcaggctctcagctcagcctgagtgttgaggccccagtggctgctctg
ggggcctcctgagtttctcatctgtgcccctccctccctggcccaggtgaaggtgtggttccagaaccgg
aggacaaagtacaaacggcagaagctggaggaggaagggcct**GAGTCCGAGCAGAAGAAGAAGGG**ctccc
atcacatcaaccggtggcgcattgccacgaagcaggccaatggggaggacatcgatgtcacctccaatga
ctagggtgggcaaccacaaa

EMX1-OFF target 1 sequence (Chr: 5):

taacacaaactggtaatagattaacaggaaaaaagggcatacaaatttattatctgcacatgtatgtaca
ggagtcatacaaaatatgaaaactcaaagaaatgcccaatcattgatgcttttataccatcttggggtta
cagaaagaataggggcttatggcatggcaagacagattgtca**GAGTTAGAGCAGAAGAAGAAAGG**catgg
agtaaaggcaatcttgtgcagatgtacaggtagcagccctcagaaaaaataggtgatagtctatggtaaa
tgtttctctgtcagatctta

Using the "Alt-R CRISPR-Cas9 System" (IDT), these selected hp-gRNAs were combined with purified SpyCas9, forming ribonucleoprotein complexes, and then incubated with either the EMX1 on-target and off-target DNA. Sizes of the *in vitro* scission products were then visualized using agarose gel electrophoresis.

## RESULTS AND DISCUSSION

### 1) Results from screening of hairpin library with 5'-UUCG-3' tetraloop

Isolated hairpins with 5'-UUCG-3' tetraloops from Sanger sequencing of selective plates and miniprepped overnight growth as well as non-selective plates are shown in Table 7 and Table 8

below. Interestingly, hairpin extensions from selective conditions have a much greater prevalence of purines. Since the 3' end of the spacer away from the mismatch sites between the target and off-target is purine-rich, we hypothesize that there is a selection for hairpins exhibit weak interactions with the 3' end of the spacer.

Secondary structures and free energies of these hp-gRNAs were predicted using the NuPack analysis tool for nucleic acid systems (Fig. 28-29).[44] Hairpin structures identified under selective conditions exclusively formed hairpins that overlapped the nucleotides that differed between the target and off-target sequences, while the structure of the hairpins identified under non-selective conditions were predicted to form secondary structures that also interacted with the other portions of the spacer.

| hp-gRNA | Sequences from non-selective conditions |
|---|---|
| n1 | UUGCTAUUUUCG**GAGUCCGAGCAGAAGAAGAA** |
| n2 | CCAUCAUGUUCG**GAGUCCGAGCAGAAGAAGAA** |
| n3 | UAUUCAUAUUCG**GAGUCCGAGCAGAAGAAGAA** |
| n4 | GGGUAGGUUUCG**GAGUCCGAGCAGAAGAAGAA** |
| n5 | ACGCCUGUUUCG**GAGUCCGAGCAGAAGAAGAA** |

Table 7. **Isolated hairpin sequences from Sanger sequencing of hairpin library screening in non-selective conditions with tetraloop 5'-UUCG-3'**. The tetraloop sequence is underlined and the spacer is bold.

| hp-gRNA | Sequences from selective conditions |
|---|---|
| s1 | AGAAGGGUUUCG**GAGUCCGAGCAGAAGAAGAA** |
| s2 | GGTNGAGGUUCG**GAGUCCGAGCAGAAGAAGAA** |
| s3 | AGGAGGUUUUCG**GAGUCCGAGCAGAAGAAGAA** |
| s4 | GAGGCACGUUCG**GAGUCCGAGCAGAAGAAGAA** |
| s5 | GGAGGAAGUUCG**GAGUCCGAGCAGAAGAAGAA** |
| s6 | AGGAAAGGUUCG**GAGUCCGAGCAGAAGAAGAA** |
| s7 | GAUAGGGAUUCG**GAGUCCGAGCAGAAGAAGAA** |

Table 8. **Isolated hairpin sequences from Sanger sequencing of hairpin library screening in selective conditions with tetraloop 5'-UUCG-3'**. The character "N" (s2) specifies an ambiguous nucleotide. These hairpins are purine rich, possibly due to interactions with the purine rich spacer.

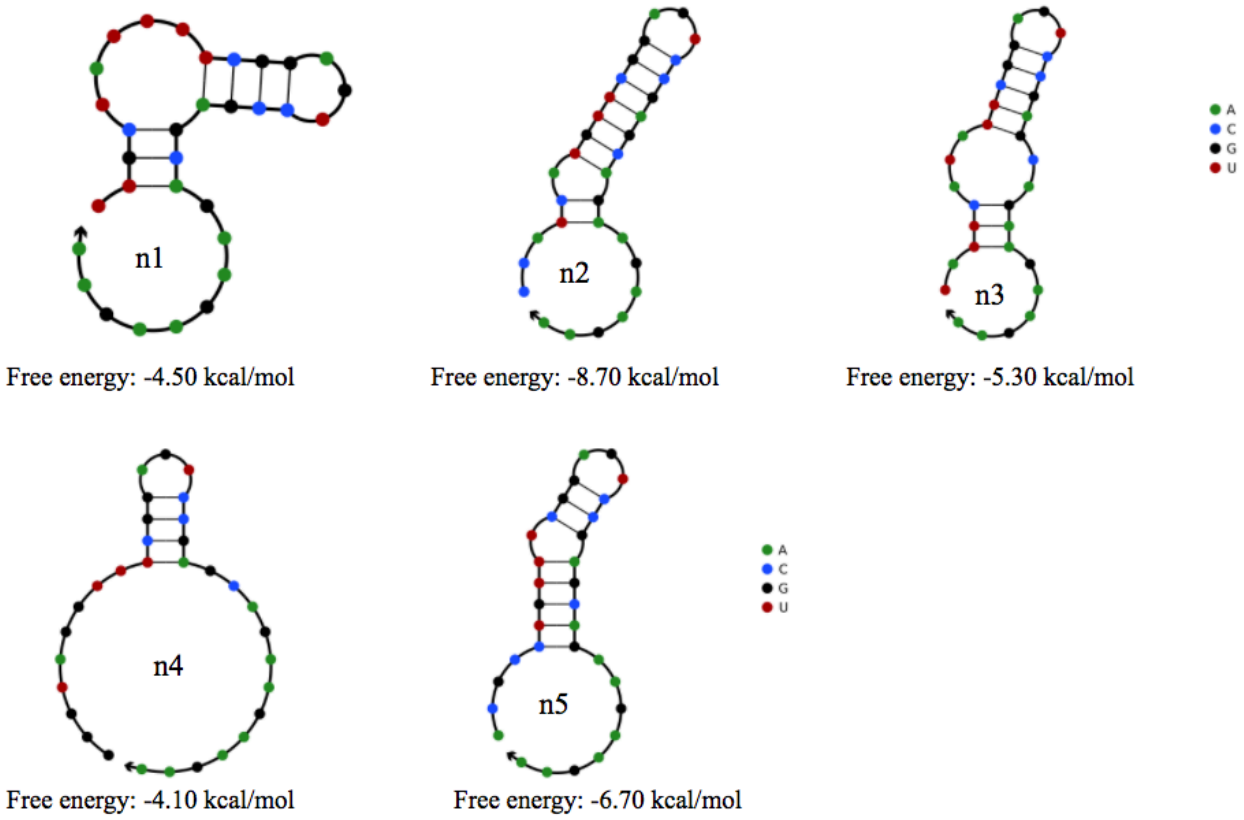



Figure 30. **Predicted secondary structures and their free energies of hp-gRNAs isolated from non-selective conditions**. For most of these hp-gRNA, with the exception of "n4", interact very heavily with the spacer.

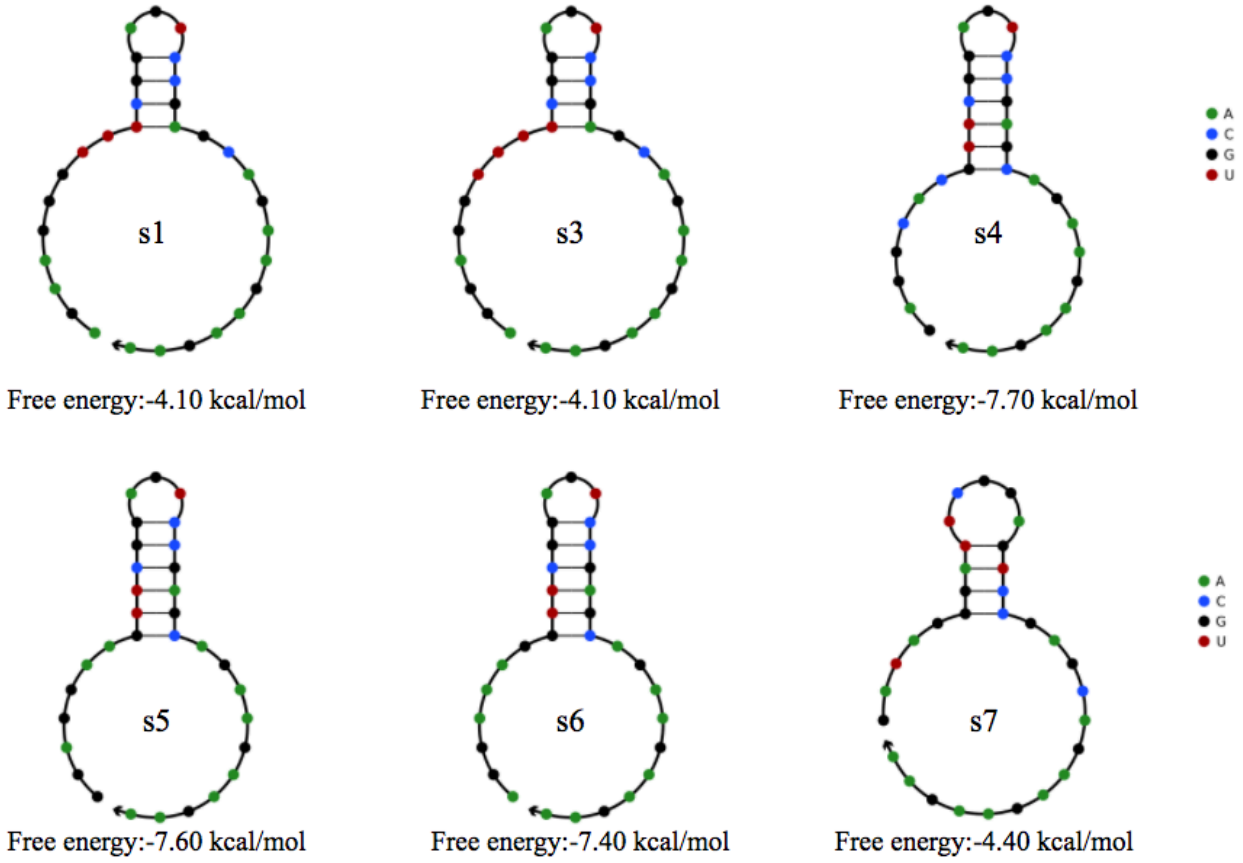



Figure 31. **Predicted secondary structures and their free energies of hp-gRNAs isolated from selective conditions**. These hp-gRNAs formed hairpins that bound the tetraloop to positions 14-18 of the spacer sequence, masking the mismatches in off-target 1 at positions 17 and 18.

From this library screen, we were able to identify and compare hairpin structures and free energies of these structures that survived in non-selective and selective conditions. Most hairpins in non-selective conditions formed secondary structures with a large portion of the spacer region, while selective conditions resulted in hairpins that formed secondary structures that are predicted to have strong interactions with the seed region of the spacer. A screen with hairpin libraries containing other potential tetraloop motifs was necessary, since the 5'-UUCG-3' tetraloop interacted with the region of the spacer containing the two mismatches with the off-target site. Ideally, the tetraloop motif should link the hairpin to the spacer, allowing for hairpin-spacer interactions in the presence of an off-target site.

### 2) Results from screening of hairpin library with library of 4 tetraloops

Ten colonies were picked from both selective and non-selective plating conditions, and three colonies from selective conditions provided clear Sanger sequencing results after colony PCR. None of the resulting hairpins from this sequencing method contained the 5'-CUUG-3' tetraloop motif. Sequences amplified from individual colonies are shown below.

| Hp-gRNA | Sequences from selective conditions |
|---|---|
| s8 | UUGAGUAUUUCG**GAGUCCGAGCAGAAGAAGAA** |
| s9 | CGCUCUAU**GAGUCCGAGCAGAAGAAGAA** |
| s10 | CUUGAGUUGCAA**GAGUCCGAGCAGAAGAAGAA** |

Table 9. **Hairpin extended guide RNAs from screening of hairpin library with 4 potential tetraloops**. Each sequence contains the hairpin extension, the tetraloops (underlined), and the EMX1 spacer segment of the guides (bold).

Secondary structures of these sequences were predicted (Fig. 30). Three very different hairpin-spacer secondary structures were predicted, each with a different tetraloop motif. We then decided to test these three selected hp-gRNAs in an *in vitro* environment to test their effect on SpyCas9 specificity.

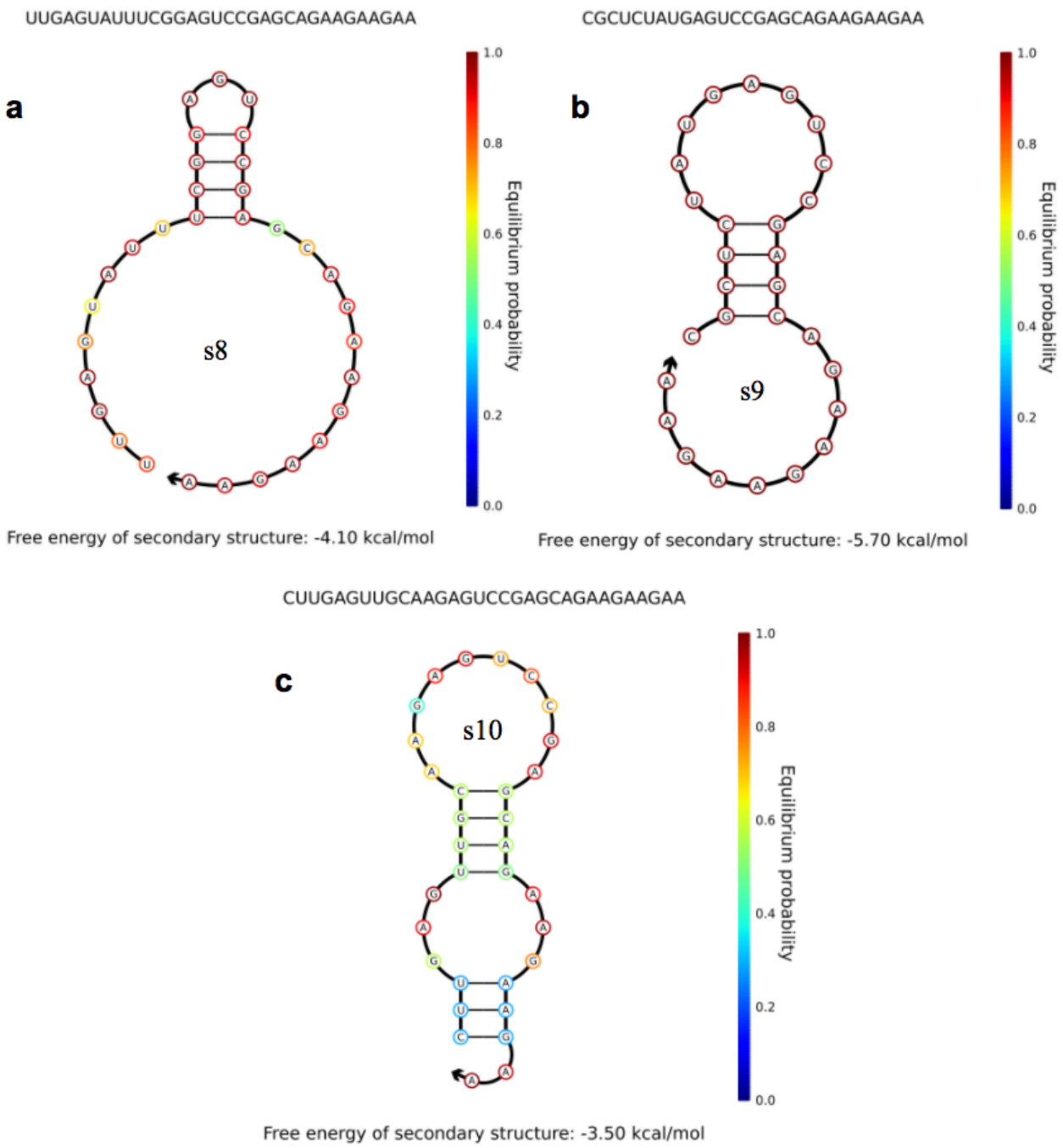



Figure 32. **Predicted of secondary structure of isolated hairpins from screening of hairpin library with library of 4 potential tetraloops**. (a) hp-gRNA with 5'-UUCG-3' tetraloop (b) hp-gRNA with no tetraloop (c) hp-gRNA with 5'-GCAA-3' tetraloop.

The isolation of these hairpins enabled us to investigate the secondary structures of these three significantly different hp-gRNA sequences. Hairpin "s8" resulted in the same tetraloop-spacer binding that resulted from the first hairpin library screen. Hairpin "s9", having no tetraloop, forms a hairpin with positions 11-14 of the spacer and nucleotides in positions 15-20 form a loop. Hairpin "s10" has a predicted secondary structure that weakly binds the hairpin extension to positions 3-5 and 9-12 of the EMX1 spacer. Since these three surviving members of the hairpin screen in selective conditions provided clear Sanger sequencing results, and our hypothesis was that hairpins recovered from the assay provide higher specificity, we chose to further examine these hairpin structures for their specificity *in vitro*. Further optimization of this screening by plating technique will allow for the isolation of more potential hairpins via colony PCR amplification, followed by *in vitro* confirmation of hp-gRNA specificity.

**3) Results from *in vitro* screening of selected hp-gRNAs**

To test hp-gRNA specificity *in vitro*, hairpins "s8", "s9", and "s10", and the normal gRNA targeting EMX1 were synthesized (IDT) and individually complexed with purified SpyCas9, resulting in four different ribonucleoprotein (RNP) complexes. The RNPs were then combined with 300 bp DNA segments containing either the EMX1 on-target or off-target site and the *in vitro* digestion reaction was performed. Post digestion, agarose gel electrophoresis allowed for visualization of the cleaved products, one 200 bp segment and one 100 bp segment. Results from duplicates of the *in vitro* assay show that the three isolated hairpin structures from the screening of multiple hairpin libraries were able to cleave only the EMX1 on-target sites. Here, the positive control, the canonical gRNA with the EMX1 spacer, shows cutting of both the on-target and the off-target site (Fig. 33-34). From this screen, we were successfully able to identify three hairpin structures that limit SpyCas9 ability to cleave the EMX1 off-target site, with maintaining SpyCas9 on-target activity.

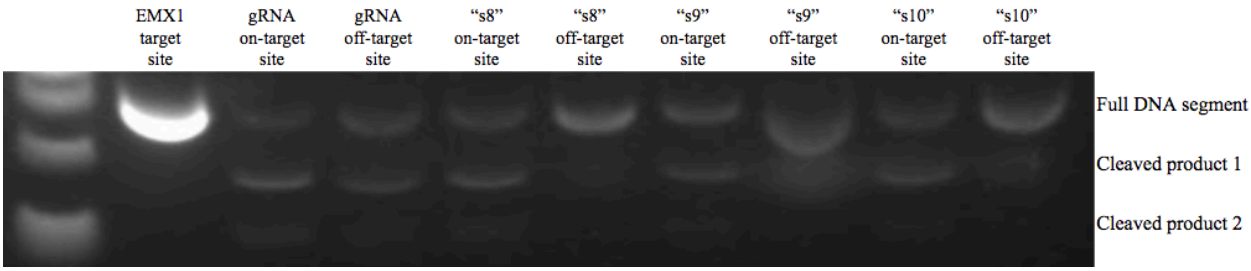



Figure 33. **Gel electrophoresis results from the *in vitro* digestion of gRNA-EMX1, "s8", "s9", and "s10".** The entire EMX1 on-target 300 bp DNA segment is in the first lane. The second and third lanes show gRNA cleavage of both the on-target and off-target sites. Subsequent lanes show the cleavage by "s8", "s9", and "s10" at only the on-target site as no bands cleaved product bands can be seen at off-target sites.

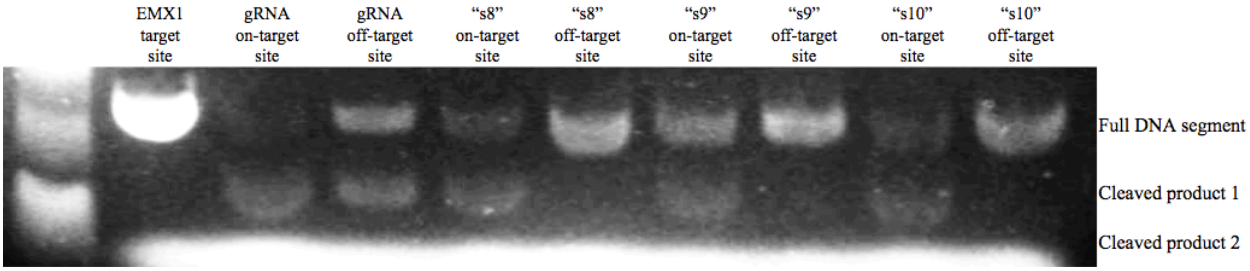



Figure 34. **Replicate of gel electrophoresis results from *in vitro* digestion of gRNA-EMX1, "s8", "s9", and "s10".** Here, we see the same results as the first *in vitro* digestion, confirming the repeatability of the assay.

The results from both hairpin library screens show that we were able to develop and conduct a screen that selects for a few specific hp-gRNAs *in vivo*. We were then able to confirm, with repeatability, that three out of three of these isolated hp-gRNAs from the multiple library screen were effective at maintaining on-target activity while diminishing off-target activity in an *in vitro* environment.

# CHAPTER V: SUMMARY AND FUTURE WORK

This screening technique enables the identification of specific hairpin structures that mitigate unwanted genomic edits for the EMX1 gene, while maintaining on-target activity. Secondary structure prediction and sequence analysis of surviving hp-gRNAs provides crucial insight on hp-gRNA behavior in various environments. Also, *in vitro* analysis of surviving hairpins shows that this screening method enables the identification of some design features for highly effective hp-gRNAs for gene editing purposes. However, to identify design rules of highly specific guides with extended hairpin motifs, we need higher throughput screens.

Further optimization of this screen to identify more highly specific hp-gRNAs will allow for larger datasets, perhaps even identifying all hairpin extensions that enhance specificity. This screen can be optimized by enhancing transformation efficiency of plasmids containing libraries into strains that have been developed to only survive in the presence of contributing members of the hairpin libraries. This can be accomplished by introducing hairpin libraries into larger cell concentrations of the selected *E. coli* strain, or by ensuring all survivors are sequenced via higher throughput NGS. Another potential method to enhance our datasets would be to ensure appropriate levels of all components of the SpyCas9 system.

In order to establish an empirical method to identify highly specific hp-gRNAs, we began with only one therapeutically important gene, the EMX1 gene. It is likely that design features or design rules will not apply to other therapeutically important genes, due to sequence variance. In order to develop strict design rules for hp-gRNAs that are applicable across a majority of therapeutic target sites, we must repeat this process for all therapeutic genes of interest, also in the presence of all of their respective genomic off-targets. Repeating this method with various target sites will help to evaluate the important design criteria for hp-gRNAs.

The most advantageous length of overlap between the hairpin extension and the gRNA spacer region remains unknown. The optimum length of overlap between hairpin and gRNA spacer may vary drastically for different genetic target, as the stability requirement of the hairpin structure

may differ per protospacer. Screening of various hairpin lengths for their efficacy in targeting various genes will also provide us with more insight to assign design rules.

Finally, enhancing detection limits of this screening may allow for design of highly specific hairpins for allele specific gene editing. Fine tuning of hairpin library screens, potentially by introducing multiple rounds of screens, may result in identifying hp-gRNAs that are specific enough to treat genetic diseases that are autosomal dominant, where only one allele results in the unhealthy phenotype. Once these hp-gRNAs are identified, we will have a deeper understanding of how to design hairpin structures that discriminate between targets and off-target that potentially only differ by one or two nucleotides.

REFERENCES


1 Barrangou R, Doudna JA. Applications of CRISPR technologies in research and beyond. *Nat Biotechnol* 2016;**34**:933–41. https://doi.org/10.1038/nbt.3659.

2 Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 2005;**435**:646–51. https://doi.org/10.1038/nature03556.

3 Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing 2013;**14**:49–55. https://doi.org/10.1038/nrm3486.TALENs.

4 Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014;**157**:1262–78. https://doi.org/10.1016/j.cell.2014.05.010.

5 Knott GJ, Doudna JA. CRISPR-Cas guides the future of genetic engineering. *Science (80-)* 2018;**361**:866–9. https://doi.org/10.1126/science.aat5011.

6 Zhang F, Wen Y, Guo X. CRISPR / Cas9 for genome editing : progress , implications and challenges 2014;**23**:40–6. https://doi.org/10.1093/hmg/ddu125.

7 Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat Rev Microbiol* 2015;**13**:722–36. https://doi.org/10.1038/nrmicro3569.An.

8 Nuñez JK, Lee ASY, Engelman A, Doudna JA. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 2015;**519**:193–8. https://doi.org/10.1038/nature14237.

9 Koonin E V., Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* 2017. https://doi.org/10.1016/j.mib.2017.05.008.

10 Koonin E V., Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* 2017:67–78. https://doi.org/10.1016/j.mib.2017.05.008.

11 Wright A V., Nuñez JK, Doudna JA. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* 2016;**164**:29–44. https://doi.org/10.1016/j.cell.2015.12.035.

12 Whinn KS, van Oijen AM, Ghodke H. Spy-ing on Cas9: Single-molecule tools reveal the enzymology of Cas9. *Curr Opin Biomed Eng* 2019;**12**:25–33. https://doi.org/10.1016/j.cobme.2019.08.013.

13 Luo ML, Jackson RN, Denny SR, Tokmina-Lukaszewska M, Maksimchuk KR, Lin W, *et al.* The CRISPR RNA-guided surveillance complex in Escherichia coli accommodates extended RNA spacers. *Nucleic Acids Res* 2016;**44**:7385–94. https://doi.org/10.1093/nar/gkw421.

14 Briner AE, Donohoue PD, Gomaa AA, Selle K, Slorach EM, Nye CH, *et al.* Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell* 2014;**56**:333–9. https://doi.org/10.1016/j.molcel.2014.09.019.

15 Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science (80- )* 2016;**351**:867–71. https://doi.org/10.1126/science.aad8282.

16 Singh D, Ha T. Understanding the Molecular Mechanisms of the CRISPR Toolbox Using Single Molecule Approaches. *ACS Chem Biol* 2018;**13**:516–26. https://doi.org/10.1021/acschembio.7b00905.

17 Bernheim A, Calvo-Villamañán A, Basier C, Cui L, Rocha EPC, Touchon M, *et al.* Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria. *Nat Commun* 2017;**8**:25–8. https://doi.org/10.1038/s41467-017-02350-1.

18 Maxwell CS, Jacobsen T, Marshall R, Noireaux V, Beisel CL. A detailed cell-free transcription-translation-based assay to decipher CRISPR protospacer-adjacent motifs. *Methods* 2018;**143**:48–57. https://doi.org/10.1016/j.ymeth.2018.02.016.

19 Boyle EA, Andreasson JOL, Chircus LM, Sternberg SH, Wu MJ, Guegler CK, *et al.* High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc Natl Acad Sci U S A* 2017;**114**:5461–6. https://doi.org/10.1073/pnas.1700557114.

20 Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (80- )* 2014;**343**:. https://doi.org/10.1126/science.1247997.

21 Tan J, Zhang F, Karcher D, Bock R. Expanding the genome-targeting scope and the site selectivity of high-precision base editors. *Nat Commun* 2020;**11**:. https://doi.org/10.1038/s41467-020-14465-z.

22 Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Med* 2014;**11**:783–4. https://doi.org/10.1038/nmeth.3047.Improved.

23 Choudhary E, Thakur P, Pareek M, Agarwal N. Gene silencing by CRISPR interference in mycobacteria. *Nat Commun* 2015;**6**:. https://doi.org/10.1038/ncomms7267.

24 Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, *et al.* Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature* 2017;**551**:464–71. https://doi.org/10.1038/nature24644.

25 Hilton IB, D'Ippolito AM, Vockley CM, Thankore PI, Crawford GE, Reddy TE, *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 2015;**33**:510–7. https://doi.org/https://doi.org/10.1038/nbt.3199.

26 Anzalone A V., Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019;**576**:149–57. https://doi.org/10.1038/s41586-019-1711-4.

27 Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, *et al.* Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell* 2013;**152**:1173–83. https://doi.org/10.1016/j.cell.2013.02.022.

28 Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar V V., Aryee MJ, Joung JK. CIRCLE-seq: A highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods* 2017;**14**:607–14. https://doi.org/10.1038/nmeth.4278.

29 Josephs EA, Kocak DD, Fitzgibbon CJ, McMenemy J, Gersbach CA, Marszalek PE. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res* 2015;**43**:8924–41. https://doi.org/10.1093/nar/gkv892.

30 Lazzarotto CR, Malinin NL, Li Y, Zhang R, Yang Y, Lee G, *et al.* CHANGE-seq reveals genetic and epigenetic effects on CRISPR-Cas9 genome-wide activity. *Nat Biotechnol* 2020. https://doi.org/10.1038/s41587-020-0555-7.

31 Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 2017;**550**:407–10. https://doi.org/10.1038/nature24268.

32 Singh D, Wang Y, Mallon J, Yang O, Fei J, Poddar A, *et al.* Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat Struct Mol Biol* 2018;**25**:347–54. https://doi.org/10.1038/s41594-018-0051-7.

33 Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B, Li Y, *et al.* Crystal Structure of Staphylococcus aureus Cas9. *Cell* 2015;**162**:1113–26. https://doi.org/10.1016/j.cell.2015.08.007.

34 Sternberg SH, Lafrance B, Kaplan M, Doudna JA. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* 2015;**527**:110–3. https://doi.org/10.1038/nature15544.

35 Ran FA, Hsu PD, Lin C, Gootenberg JS, Trevino A, Scott D a, *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 2013;**154**:1380–9. https://doi.org/10.1016/j.cell.2013.08.021.Double.

36 Guilinger JP, Thompson DB LD. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* 2014;**32**:577–82. https://doi.org/10.1038/nbt.2909.Fusion.

37 Lee JK, Jeong E, Lee J, Jung M, Shin E, Kim Y hoon, *et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun* 2018;**9**:. https://doi.org/10.1038/s41467-018-05477-x.

38 Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* 2014;**32**:1262–7. https://doi.org/10.1038/nbt.3026.

39 Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 2016;**34**:184–91. https://doi.org/10.1038/nbt.3437.

40 Okafor IC, Singh D, Wang Y, Jung M, Wang H, Mallon J, *et al.* Single molecule analysis of effects of non-canonical guide RNAs and specificity-enhancing mutations on Cas9-induced DNA unwinding. *Nucleic Acids Res* 2019;**47**:11880–8. https://doi.org/10.1093/nar/gkz1058.

41 Kulcsár PI, Tálas A, Tóth E, Nyeste A, Ligeti Z, Welker Z, *et al.* Blackjack mutations improve the on-target activities of increased fidelity variants of SpCas9 with 5â€²G-extended sgRNAs. *Nat Commun* n.d. https://doi.org/10.1038/s41467-020-15021-5.

42 Kocak DD, Josephs EA, Bhandarkar V, Adkar SS, Kwon JB, Gersbach CA. Increasing the

specificity of CRISPR systems with engineered RNA secondary structures. *Nat Biotechnol* 2019;**37**:657–66. https://doi.org/10.1038/s41587-019-0095-1.

43 Mckenzie GJ, Craig NL. Fast , easy and efficient : site-specific insertion of transgenes into Enterobacterial chromosomes using Tn 7 without need for selection of the insertion event 2006;**7**:1–7. https://doi.org/10.1186/1471-2180-6-39.

44 Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, *et al.* Software News and Updates NUPACK : Analysis and Design of Nucleic Acid Systems 2010. https://doi.org/10.1002/jcc.