VILNIUS UNIVERSITY

LIFE SCIENCES CENTER

DEPARTMENT OF BIOTECHNOLOGY

2nd year master's student of molecular biotechnology study program

DANIEL ADEGUNLE **ADELAKIN**

Determinants of CRISPR nuclease gRNA function

Master thesis

Supervisor

Dr. Stephen Knox Jones Jr

Reviewer

Marijonas Tutkus

Vilnius, 2023

CONTENT

Table of Contents

ABBREVIATIONS

AMN – azidomethylnicotinyl

APS –ammonium persulfate

BAA –bisacrylamide

AsCas12a – *Acidaminococcus sp*. CRISPR-associated protein 12a

BH – Bridge Helix

BpCpf1 – *Butyrivibrio proteoclasticus*

Cas – CRISPR associated

Cas12a – CRISPR-associated protein 12a

Cas9 – CRISPR-associated protein 9

CCR5 - chemokine receptor type 5

CNN - Classifiers convolutional neural networks

Cpf1– CRISPR from *Prevotella* and *Francisella1*

CRISPR – clustered regularly interspaced palindromic repeats

crRNAs – CRISPR- ribonucleic acids

Csx19 – CRISPR-associated protein Csx19

dCas12a – Dead-nuclease CRISPR-associated protein 12a

dgRNA – dual-gRNA

DSB – double strand break

dsDNA – double stranded deoxyribonucleic acid

DTT – dithiothreitol

EDTA. ethylenediaminetetraacetic acid,

FADS - forescence-activated droplet sorting

DMSO – Dimethyl sulfoxide

FnCas12a – *Francisella novicida* CRISPR-associated protein 12a

GFP-HEK – green fluorescent protein- Human embryonic kidney

gRNA – guide RNA

IPTG – isopropyl ß-D-1-thiogalactopyranoside

LbCas12a – Lachnospiraceae bacterium

LKL – loop-lysine helix-loop

MeCP2 – methyl CpG binding protein 2

NPOM - 6-Nitropiperonyloxymethylene

NTS – non-target DNA strand

NUC – Nuclease lobe

PAM – protospacer adjacent motif

PI – PAM interacting domains

R3P – phosphine

PMSF – phenylmethanesulfonyl fluoride

RCA – rolling circle amplification

REC recognition lobe

RNP – ribonucleoprotein complex

RRS – repeat recognition sequence

RuvC – RuvC nuclease domain

SDS – sodium dodecyl sulfate

smFRET – Single molecule fluorescence resonance energy transfer

SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

SpCas9 – *Streptococcus pyogenes* Cas9

ssDNA – single stranded deoxyribonucleic acid

SUMO – Small ubiquitin-related modifier

THPP – tris(hydroxypropyl)phosphine

TnpB – transposon-associated proteins

tracrRNA - Trans-activating CRISPR- ribonucleic acid

TS – target DNA strand

WED – wedge

INTRODUCTION

CRISPR (clustered regularly interspaced palindromic repeats) systems are the most sought-after genome editing tools lately, for therapeutics and diagnostics in the field of biomedicine. CRISPR systems are derived from bacterial adaptive immune systems and are deployed as a defense mechanism by bacteria against invading phages and viruses. Presently, the CRISPR system has been utilized for genome editing (Wang et al., 2019). CRISPR technology has been implemented in gene therapy through genome engineering (base addition, base disruption, base deletion, and base substitution) and in diagnostics (Uddin et al., 2020; Li et al., 2019; Gootenberg et al., 2017). Despite the widespread adoption of CRISPR, additional breakthroughs on how to enhance the technology are being discovered. The technology, meanwhile, has some limitations, such as difficulty in delivering the system into cells for genome editing, precision gene editing, a high rate of off-target effects, and the protospacer adjacent motif (PAM) requirement (Uddin et al., 2020).

A complete CRISPR system consist of CRISPR-associated (Cas) proteins along with a repeat and spacer array. This array encodes RNAs that function as a guide for a Cas effector complex – a ribonucleoprotein (RNP) that digests foreign deoxyribonucleic acid (DNA) to defend against conjugative plasmids, transposable elements, and viral infection (Wong et al., 2015).

During CRISPR-mediated activity, the CRISPR acquisition machinery copy and paste invasive DNA sequences as new spacers at the leading end of the CRISPR array. Cas proteins (mainly Cas9 and Cas12a) transcribe CRISPR arrays and produce mature tiny interfering CRISPR- ribonucleic acids (crRNAs). Once fully processed, a crRNA (sometimes in complex with an additional trans-acting crRNA) acts as a guide RNA (gRNA): a short RNA that contains a scaffold sequence necessary for binding to the Cas nuclease and a variable spacer sequence which can be designed to match a specific target. The gRNA leads the Cas apparatus to complementary DNA flanked by Protospacer Adjacent Motif (PAM) sequence (a short 3-5 base pair sequence: 5′-NGG-′3 in *S. pyogenes* and 5′-TTTN-′3 in *Acidaminococcus sp*, responsible for target sequence identification and cleavage initiation), where it cleaves target DNA in a sequence-specific manner (Das et al., 2010).

Through gRNA, the remarkable programmability of CRISPR-Cas nucleases (effector complex that triggers degradation of the targeted DNA) has found useful purpose in genome editing and targeted nucleic acid detection because the guide RNA sequence can be altered via nucleotide base removal, replacement and addition (Barrangou and Doudna, 2016). However, gene editing with CRISPR nucleases is encompassed with difficulties: the leading nucleases, Cas9 and Cas12a, are unable to

edit certain genes. A main cause of this underlying issue is thought to be guide RNA (gRNA) secondary structures (Fonfara et al., 2016).

Studies (Herring-Nicholas et al., 2023; Konstantakos et al., 2022) have shown that some Cas12a is unable to target some genes in the genome, which might be due to the gRNAs being unable to direct nucleases to their intended DNA targets sites. Finding have suggested that this could be as a result of the gRNA secondary structure i.e. the gRNA architecture and nucleotide composition (Wang et al., 2019). Although gRNA with slight modifications such as nucleotide composition and secondary structure extension has been shown to improve Cas12a nuclease efficiency for gene editing (Kocak et al., 2019). Robust designs of gRNAs with the potential to improve target identification and genome editing are still desired (Riesenberg et al., 2022).

For this research, 12,000 Cas12a gRNA designs with varying secondary structure modifications (such as nucleotide compositions in the stem, loop, and tail regions) and that target cancer genes will be analyzed to show how these features influence transcription, ribonucleoprotein complex formation, and target cleavage. Methodologies were also adopted for the study, like strategies employed in the design of gRNA and library templates. Likewise, modified Cas12a (dCas12a and light-activated Cas12a) gave the study an excellent viewpoint for quantifying the strategies used in this investigation. Adopted microfluidic droplets provide a suitable platform for the performance of thousands of reactions involving each member of the library sequentially. Lastly, at each stage, high-throughput next-generation sequencing was used to characterize the effects of these modified gRNAs' secondary structures.

Objective of the research work:

To investigate how modifications to gRNA secondary structures affect how well gRNAs can be produced from DNA templates in a library, examine how the secondary structures on the gRNAs affect the gRNA's ability to form a ribonucleoprotein complex with Cas12a nuclease and whether the complex can mediate cleavage of DNA library targets.

Aims of the research work:

To investigating how guide-RNA sequence and secondary structures impact:

- gRNA transcription from a DNA library
- Formation of Cas12a protein-gRNA complexes
- Cleavage of DNA templates

Layout of research work:

1. Oncogene guide-RNA library design
2. DNA template preparation: single stranded DNA library to double stranded library
3. Plasmid library preparation
4. Cell library preparation
5. Droplet microfluidic assay for Cas12a cleavage reaction
6. Restriction enzyme digestion of cleavage products
7. DNA fragment size selection
8. Next generation sequencing of template fragments

1. LITERATURE REVIEW

1.1 Background of CRISPR system

CRISPR (clustered regularly interspaced short palindromic repeats) is a prokaryotic (bacteria and archaea) immune response that destroys invading foreign nucleic acids, and was recently adopted as a gene editing tool for therapeutic and diagnostic purposes (Cooper and Hasty, 2020). CRISPR system contain distinctive arrays of short repeated sequences that alternate with short regions of variable sequence (spacers), as well as genes encoding Cas (CRISPR-associated) proteins. Prior infections are encoded as short DNA sequences incorporated into CRISPR loci within the bacterial genome. These earlier infections are stored as spacer sequences that are bordered by repeat sequences. Spacer arrays are transcribed into a pre-crRNA (CRISPR RNAs) before being processed to functional crRNAs (Hoikkala et al., 2021). In prokaryotes, CRISPR-Cas systems are multiplexed by design; organisms can encode one or more CRISPR arrays and produce a plethora of Cas (CRISPR-associated) proteins that aid in the acquisition of additional spacers and the processing of CRISPR arrays (Cong et al., 2013). The crRNAs bind Cas nuclease proteins such Cas9 or Cas12a. Once attached, they direct this 'effector' to complementary DNA or RNA, which the nuclease typically cuts to destroy the invading nucleic acids (McCarty et al., 2020; Mcginn and Marraffini, 2018).

CRISPR-Cas systems are divided into two main classes (class 1 and class 2) based on the architecture of their Cas effector proteins, which provide protection by cleaving foreign nucleic acids. In the class 1 systems, multiple Cas proteins make up the effector modules, which facilitate pre-crRNA processing and CRISPR interference (Kaminski et al., 2021). On the other hand, the class 2 system effector module is made up of a single multidomain Cas protein (Makarova et al., 2020). They are also grouped into six major types (types I–VI) and at least 25 subtypes based on their characteristic Cas gene and their targeting mechanism (Hille and Charpentier, 2016).

Table 1.1.1: Class, types, and subclass of CRISPR/Cas systems. (Table adopted from He, 2023a)

Class	Types	Subtypes	Effector Module	Spacer Integration
Class 1	I	I-A,I-B,I-C,I-G,I-D,I-E,I-F1,I-F3,I-F2,	Cas3″,Cas5,Cas6,Cas7, Cas8, Cas11, Cas10	Cas1,Cas2,Cas4
	III	III-A,III-D,III-E,III-F,III-C,III-B	Cas5,Cas7,Cas10,Cas11, Csx19	Cas1,Cas2
	IV	IV-A,IV-B,IV-C	Cas5,Cas7,Cas8, Cas11	Cas1,Cas2
Class 2	II	II-A,II-B,II-C1,II-C2	Cas9	Cas1,Cas2,Cas4
	V	V-A,V-E,V-B1,V-B2,V-I,V-H,V-C, V-D,V-F1,V-F1(V-U3),V-F2,V-U2, V-U4,V-F3,V-U1,V-G,V-K(V-U5)	Cas12	Cas1,Cas2,Cas4
	VI	VI-A,VI-D,VI-C,VI-B1,VI-B2	Cas13	Cas1,Cas2

Currently, several RNA-guided Cas protein endonucleases have been identified and utilized for gene editing. Cas9 and Cas12a (formerly Cpf1) are the most sorted nucleases (Schubert et al., 2021). The high programmability of the gRNA used by the nucleases to target potential DNA targets during gene editing via Watson-Crick base pairing has made them a favorite choice for scientist. As both nucleases utilize gRNA for target DNA location, the gRNA they use differs in its composition. For the Cas 9 system, a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA) both makes up its gRNA). To reduce the Cas9 gRNA ambiguity, they are now engineered by fusing the crRNA and the tracrRNA together to make a single gRNA (sgRNA). whereas the gRNA utilized by Cas12a nuclease is composed of crRNA only (He et al., 2023a; Schubert et al., 2021).

Cas9 and Cas12a CRISPR systems use a 3 - 4 base pair sequence called the protospacer adjacent motif (PAM) for identifying target DNA locations. The Cas9 (Type II) system recognizes a 5′-NGG-3′ PAM sequence located downstream of the target sequence. While the Cas12a (type V) CRISPR system recognizes a T-rich PAM sequence 5′-TTTN 3′, where N could be either an A, G, or C, the T-rich sequence recognition by Cas12a broadens its adaptation for gene editing (Tang et al., 2021). PAM sequence recognition initiates the unwinding of target DNA, which will start the hybridization between the target strand of the DNA and the gRNA in a process also referred to as "R-loop" formation. Cas9 utilizes a 17–20 nt long gRNA, and after R-loop formation, it generates a blunt end cut on the target double-stranded DNA with the cleavage site being 3 base pairs upstream from the PAM sequence (Konstantakos et al., 2022). While Cas12a uses a 41–44 nucleotide-long gRNA,

which generates a staggered end cut of the target double-stranded DNA with a 5′-overhang (Schubert et al., 2021).

The seed sequence also plays a vital role in hybridization of the target DNA and gRNA; Cas9 uses a 10-nucleotide seed sequence, whereas Cas12a uses 5–6 nucleotides (Fonfara et al., 2016). Target DNA cleavage occurs in catalytic sites upon completion of DNA-RNA hybridization. For Cas9, cleavage of the target strand is done at the HNH domain, while the non-target strand is cleaved at the RuvC domain. In comparison, Cas12a uses only a single nuclease site. Both strands of the target DNA are cleaved at the same site, with the target strand being the first to be cut, followed by the non-target strand; this also explains the occurrence of the staggard DSB produced by Cas12a (Stella et al., 2017; Paul and Montoya, 2020).

1.2. Evolutionary origins of Cas12a

Through evolution, the type II and type V CRISPR systems were derived from separate evolutionary processes; both systems, however, possess certain similarities. Both systems are thought to be derived from a nuclease domain similar to RuvC, which is believed to be a derivative of transposon-associated proteins (TnpB) expressed by autonomous bacterial and archaeal transposons (Safari et al., 2019). However, Cas12a is believed to have been developed from distinct ancestral transposons coupled with the inclusion of various protein domains. These could be the reason why the effector nucleases vary in how they direct cleavage, but they also surprisingly possess similar architecture and capabilities (Swarts and Jinek, 2018).

The CRISPR-Cas12 system has evolved into a bilobed architecture, made up of the recognition (REC) lobe and the nuclease lobe (NUC). The REC lobe of the effector is composed of two domains, the REC1 and REC2 domains. While the NUC lope consists of more numerous domains such as the Wedge (WED), RuvC nuclease domain (RuvC), Bridge Helix (BH), Nuclease (Nuc), and PAM-interacting (PI) domains (Figure 1.2.1 A, B, C) (Swarts and Jinek, 2018; Safari et al., 2019).

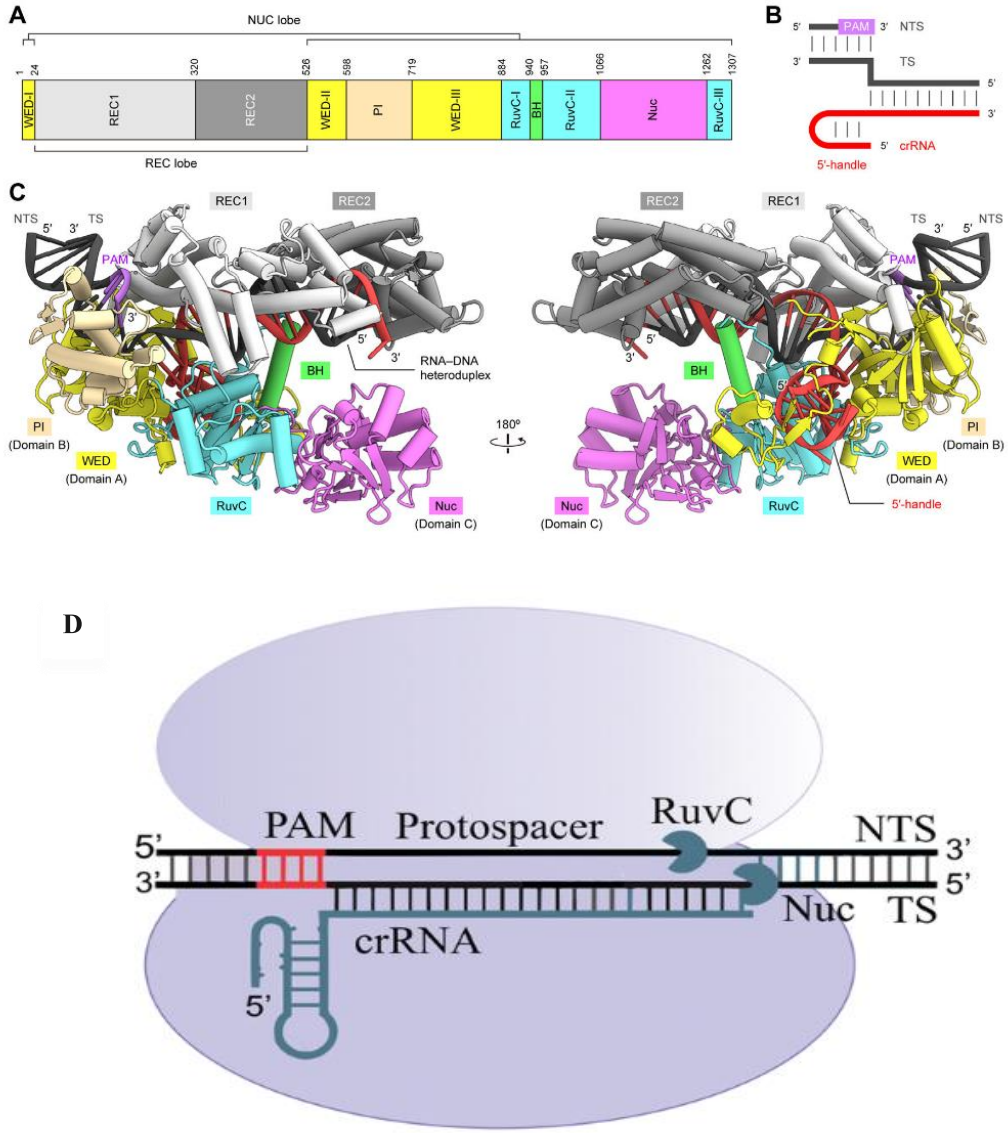


Figure 1.2.1 Domain organization of Cas12a. (**A**) Schematic representation of the Cas12a domain structure organization (figure adapted from Yamano et al., 2016). (**B /D**) Schematic representation of the crRNA and target DNA. TS: target DNA strand; NTS: non-target DNA strand (figure adapted from Yamano et al., 2016; Safari et al., 2019). (**C**) Cas12a domain crystal structure (figure adapted from Yamano, 2016)

1.3. Cas 12a CRISPR RNA Biogenesis

Genomic information needed to target DNA sequences is found in CRISPR RNA (crRNA). The crRNA and the Cas protein form a ribonucleoprotein (RNP) complex. The RNP complex is a cleavage-effecting assembly that utilizes genomic information in the crRNA for target acquisition and the Cas protein to effect target DNA cleavage. A mature crRNA is made up of a repeat section and a spacer section. The mature crRNA uses its repeat section to form connections with the Cas

protein in a structural and sequence-dependent way, while the spacer section mediates target specificity (Hille et al., 2018).

To mediate target DNA cleavage, type II CRISPR-Cas systems utilize the tracrRNA and RNase III for processing the mature crRNA, while mature crRNA processing in type V CRISPR systems does not require the tracrRNA and RNase III. (Stella et al., 2017). For the Cas12a CRISPR system, the genetic information (CRISPR locus) that is transcribed and generates the mature crRNA is transcribed as a single transcript.

The genetic information is made up of 23–25 base pair (bp) long spacer sequences sandwiched between 19–20 bp long repeat sequences. The repeat section transcribed in the formation of the mature crRNA forms pseudoknot structures, or secondary structures. These pseudoknot structures allow the crRNA to bind to the Cas12a protein due to the arrangement of divalent cations like Mg^{2+} or Ca^{2+} on the structures (Paul and Montoya, 2020). In practice, the mature crRNA of the Cas9 CRISPR system is longer than that of the Cas12a CRISPER system. These short lengths of the Cas12a effectors serve a great advantage, as it is easier to transfer the system into target cells and synthesizing the RNAs is cheaper (Behler and Hess, 2020).

1.4. Cas 12a PAM recognition

The protospacer adjacent motif (PAM) is a 3–5 nucleotide sequence recognized by the gRNA in the RNP complex assembly; the gRNA acquisition of the target sequence is initiated by the recognition of the PAM sequence. Upon recognition, the Cas effector, which makes up the other part of the RNP complex, can cleave the target DNA (Krysler et al., 2022). The CRISPR-Cas system utilizes the PAM sequence to distinguish the endogenous genomic deoxyribonucleic acid (DNA) of the bacterium from the genomic DNA of invading nucleic acids. In CRISPR systems, the PAM sequence is part of the vital multistep control process employed by the systems to ensure precise and accurate target cleavage. The PAM sequence in Cas12a system identification also initiates the hybridization of the DNA target with the crRNA, which is degraded in the WED II-III, REC1, and PAM-interacting domains (Figure 1.2.1. C) (Paul and Montoya, 2020).

For effective dsDNA target identification and separation to occur, the conserved loop-lysine helix-loop (LKL) region in the PI domain plays a big role. These domains contain essential lysine amino acids (K667, K671, and K677) as found in FnCas12a. Similarly, prolines in the LKL region of the REC1 and WEB domains also recognize the dsDNA, which helps in inserting the duplex into the

PAM dublex (fig. 1.1.1. C) (Stella et al., 2017). Structural study of the PAM interaction (Stella et al., 2017) showed that the helix is placed at a 45° angle to the dsDNA longitudinal axis, which facilitates the unwinding of the dsDNA into a helical shape. Immediately after the PAM sequence, decoupling of the dsDNA occurs via the Watson-Crick interaction between the base pairs of the dsDNA and crRNA due to the position the three conserved lysines' occupy on the dsDNA. During the decoupling process of the dsDNA strand, the non-target strand (NTS) of the DNA is directed to the PAM-interacting domain. As the DNA strands unzip, this unzipping further allows the hybridization between the target strand (T) and the crRNA (Figure 1.2.1) (Safari et al., 2019; Yamano et al., 2016).

Following an extensive study of 46 varieties of the Cas 12a family protein sequence database containing only essential components needed for target DNA cleavage, 16 Cas12 proteins from the group's PAM sequence were determined, as well as their respective functions in target-sequence identification. Only eight Cas12 protein members, *Prevotella disiens*, *Candidatus methanoplasma termitum*, *Moraxella bovoculi 237*, *Acidaminococcus sp. BV3L6*, *Lachnospiraceae bacterium ND2006*, *Lachnospiraceae bacterium MA2020*, and *Francisella novicida U112*, whose function was determined, were discovered to have the capability to induce active cleavage activity on target DNA when PAM sequences are identified during the initiation of cleavage activity (Safari et al., 2019).

Similarly, it has been revealed that spacer sequences can be effectively targeted by Cas12a following the identification of T-rich sequences. These T-rich PAM sequences are located 5' upstream of the non-target strand, having a sequence of 5′-TTTN-3' for LbCas12a, AsCas12a, and FnCas12a, respectively (Figure 1.4.1). From studies, PAM sequence ending with a C has been shown to be more active in target DNA recognition (Collias and Beisel, 2021).

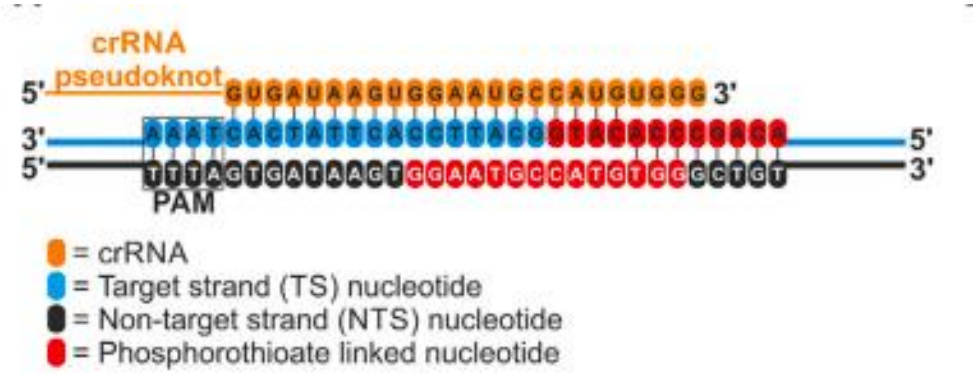


Figure 1.4.1. Showing the Schematic representation of the sequences of TS and NTS DNA Nucleotides. TS: This is the DNA template strand which the guide RNA attaches to. NTS: This is the DNA strand which the PAM is located, and it is complementary to the target strand (Swart, 2017).

The Cas12a protein family has been shown to generate staggered double strand breaks (DSBs) with 5′ overhangs in genomes that have a rich distribution of adenosine and thymine (A,T) sequences. The discovery of the T-rich PAM used by Cas12a has further expanded the range of CRISPR adoption as a gene editing tool when compared to the Cas9 system, which requires a G-rich PAM sequence (Teng et al., 2018).

1.5. Seed segment binding and RNA–DNA duplex hybridization

CRISPR systems employ important process checkpoints to achieve successful cleavage of target DNA sequences; among these is seed sequence pairing, or hybridization. The seed sequence is 5–6 bases downstream of the PAM sequence and shares similar homology with the gRNA spacer sequence and the target DNA. Like most CRISPR systems, Cas12a DNA cleavage depends on the presence of the homologous seed sequence with the gRNA spacer sequence (Yamano et al., 2016).
The seed segment is essential for the gRNA and target strand hybridization (Figure 1.5.1), making the seed segment pairing solely responsible for the formation of the R-loop or heteroduplex, which involves the local unwinding of the target DNA at the PAM proximal end of the target strand followed by base pairing interaction with the homologous segment of gRNA (Forfara et al., 2016, p. 1).
The study of the Cas12a crystal structures (Swarts et al., 2017) shows that crRNA seed segment follows an established A-form helical shape. They also investigated the effect of reducing the entropic penalty connected to the synthesis of the crRNA-target strand duplex and they discovered that the seed segment plays an important role in directing the RNP complex to target identification and cleavage. Other studies were also able to conclude that the presence of mismatches between the seed segment hybridizing pairs could result in a slow or disruption of hybridization, which would have negative effects or hamper target DNA cleavage (Swarts et al., 2017; Bernd Zetsche, 2015; Li et al., 2017), and that the seed region promotes hybridization of the crRNA-target DNA, which ensures target DNA cleavage (Paul & Montoya, 2020).

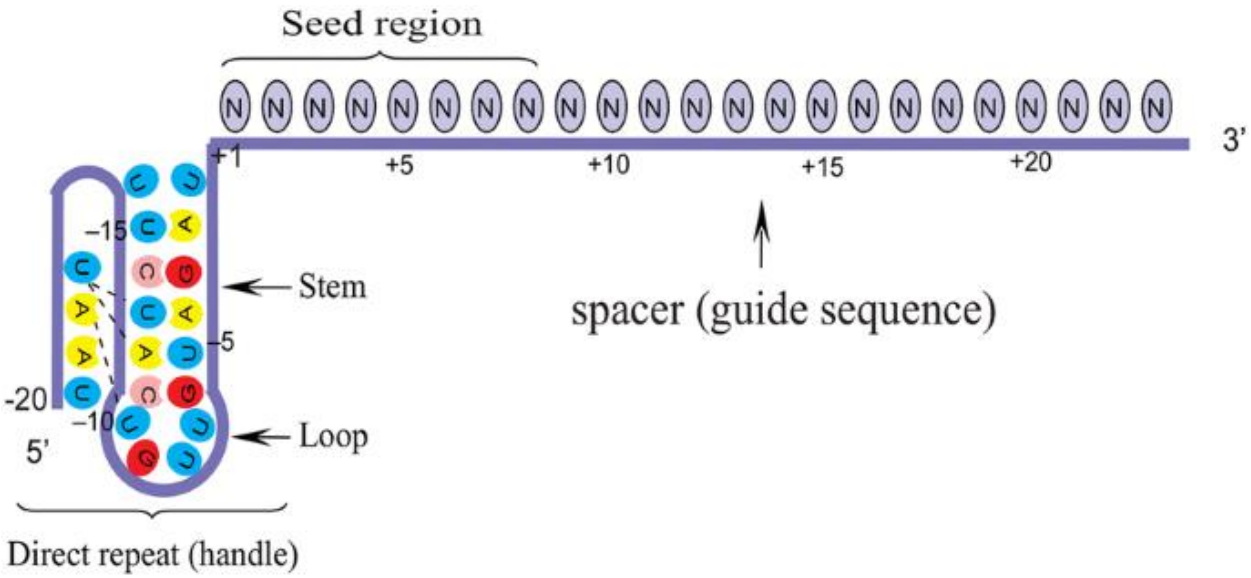


Figure 1.5.1. Schematic representation of the Cas12a guide RNA, showing the spacer region, repeat region and the seed regions (figure adapted from Safari, 2019)

1.6. Cas12a R-loop formation

R-loop formation encompasses the process where the unwound target dsDNA strand hybridizes with the spacer sequence of the crRNA, which is initiated by PAM sequence recognition and seed segment hybridization (Singh et al., 2018). In the Cas12a CRISPR system, the R-loop is formed when the 20 nucleotides of the crRNA spacer sequence base pair with the target strand of the target DNA (Cofsky et al., 2020). For cleavage of the DNA strands by Cas12a to occur, both strands must reach the protein catalytic site and adopt a 5′–3′ polarity before they can be cut (Paul & Montoya, 2020). These were backed up with structural examination via smFRET, which shows the DNA non-target strand having a 5′-3′ polarity, which facilitates its positioning for entry into the RuvC-Nuc catalytic pocket. Similarly, the DNA non-target sites are also positioned in the 5′–3′ polarity (Paul & Montoya, 2020). Because the non-target strands can easily get into the RuvC-Nuc catalytic site via the 5′–3′ polarity, it is difficult for the target strand to get into the catalytic site using the same polarity. Therefore, the REC and Nuc lobes distal ends undergo conformational repositioning to accommodate the target strand into the cleavage site (Swarts & Jinek, 2018). Differences in when both strands enter the catalytic site might explain why the target strand hydrolyzes faster than the target strand, which might also be responsible for the reduction in cleavage rate (Stella et al., 2018; Naqvi et al., 2022).

These findings further confirm that cleavage of the non-target strand is not a result of deliberate strand cleavage but rather because the non-target strand's 5′–3′ positioning makes it an ideal candidate for cleavage in the catalytic site. Compared to the cleavage of the target strand, the REC and Nuc undergo reorientation to cut the target strand in a deliberate manner. Following cleavage of the non-target and target strands, the distal end of PAM dissociates from the R-loop complex while the PAM proximal site stays attached to Cas12a, which facilitated the cleavage of the R-loop (figure 1.6.1) (Stella et al., 2017; Cofsky et al., 2020). Aside from cleaving target dsDNA strands, Cas12a is also capable of degrading ssDNA (Stella et al., 2018).

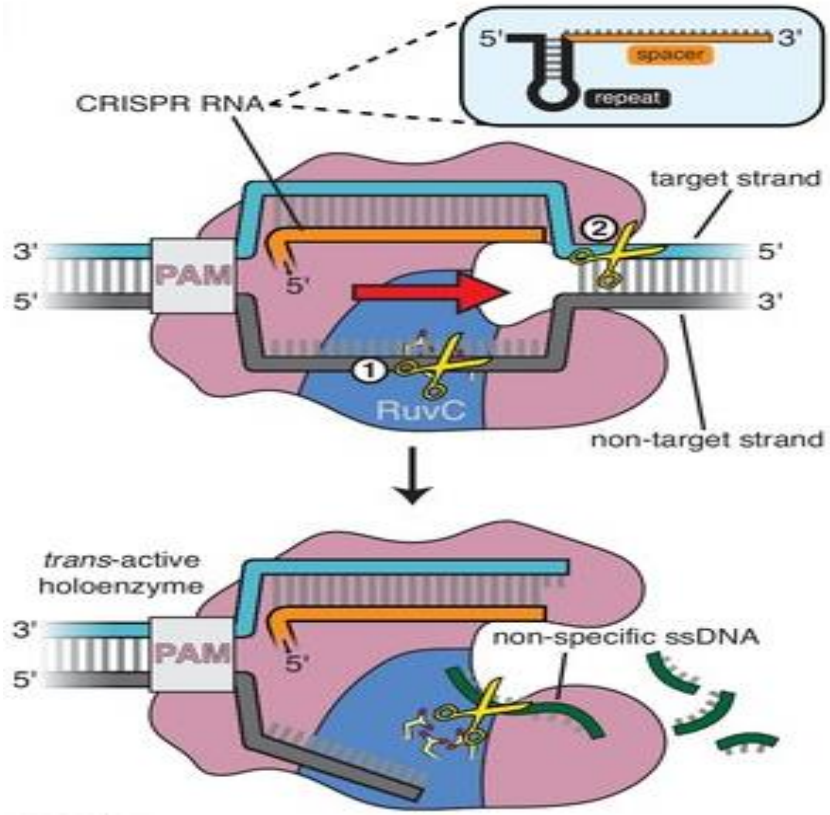


Figure 1.6.1. Illustrating successful R-loop formation and subsequent cleavage of the non-target strand, the target strand in cis, and non-specific ssDNA (in trans). The cis strand cleavage order is indicated by the circled numbers, and the orientation of the R-loop opening is indicated by the red arrow (figure adapted from Cofsky, 2020).

1.7. Cas12a Guide RNA Architecture

The CRISPR/Cas12a system, now regarded as a suitable alternative to the Cas9 system, has been very effective in genome engineering too. These are mostly attributed to the system's straightforward setup and how effectively it can be programmed to target genes of interest. The Cas12a system is known to cause a DSB in their respective target locations, which is caused by the Cas12a nuclease and gRNA, both of which combine to form an RNP effector complex, where the gRNA plays a crucial role in target recognition and binding while the Cas12a performs the sole function of target DNA cleavage (Bernd Zetsche et al., 2015). The gRNA component of the RNP complex recognizes the target sequence via base complementation between both molecules, in a process initiated by the recognition of the PAM sequence (Figure 1.6.1) (Yamano et al., 2016).

The ability of the gRNA to recognize the target site is determined by the variable section of the gRNA (spacer), while the other section (repeat) allows the gRNA to be anchored to the Cas12a protein. The repeat section of the gRNA can be altered to improve DNA targeting and cleavage by the RNP complex. The choice of gRNA is vital when targeting the genome, as it will determine if the target of interest can be cleaved or not. Therefore, the design of suitable gRNA has been given much attention because it is believed to increase target specificity and maximize editing efficiency (Schubert et al., 2021).

Discovery of the CRISPR system as a short tandem repeats and spacer sequences of prokaryotes transcribed from the CRISPR loci, which work synergistically with the nearby Cas protein for the bacterial defense. The CRISPR system component is arranged along the genomic locus in an array of repeating sequences flanked by the spacer sequence similar to the invading genomic element (Jiang and Doudna, 2015). The spacer role is important for the system as it serves as an active memory for the CRISPR system. Transcription and processing of the spacer into mature crRNA designate the CRISPR system in its defense role (Gao et al., 2016). The mature crRNA and the CRISPR-associated (Cas) genes make up the system defense machinery, where multiple Cas-conserved protein-coding genes flank the CRISPR array (He et al., 2023b).

Figure 1.7.1. Showing Schematic of the Cas 12a crRNA-DNA-targeting complex. Cleavage sites are indicated by red arrows (figure adapted from Bernd Zetsche, 2015).

To determine the composition of the mature Cas12a crRNA, RNA sequencing results from *Francisella novicida U112* cultures show that the mature crRNA is 42–44 nucleotides long, while the first 19–20 nucleotides correspond to the repeat sequence and the final 23–25 nucleotides represent the spacer sequence. To validate the results and determine which part of the crRNA is needed for target DNA cleavage, the direct repeat was truncated. The results showed that for cleavage to happen, the repeat must have a minimum of 16 nucleotides, with optimal results gotten when the repeat is made up of over 17 nucleotides. When these results are compared with *Streptococcus pyogenes* CRISPR-associated protein 9 (SpCas9) requirement for cleavage, the result is comparable, as the Cas9 system also requires a minimum of 16–17 nucleotides. In the same study, they reached the conclusion that the gRNA seed region is within the first 5 nucleotides of the spacer sequence at its 5′ end (Bernd Zetsche, 2015).

Further, the direct repeat sequence of *Francisella novicida U112* was compared with other Cas12 families such as *Lachnospiraceae bacterium MC2017*, Lb3Cpf1, *Butyrivibrio proteoclasticus*, BpCpf1, *Smithella sp. SC_K08D17*, SsCpf1, and *Francisella novicida*, FnCpf1. Results showed that they possess different requirements for the repeat sequence, but they all have similar stem-loop structures (Bernd Zetsche, 2015).

1.7.1. Strategies employed to enhance the efficiency of Cas12a guide RNA

The use of the CRISPR system for gene editing is not without its own difficulties. For example, the lower specificity of the Cas12a nuclease, which has been suggested as a replacement for Cas9, limits its use for gene editing (Kocak et al., 2019). The main objectives of most studies are to increase nuclease specificity, and one of the most extensively explored ways to address Cas12a nuclease specificity deficiencies is to try evaluating the significance of guide RNAs given their critical function

in target identification. Despite the effectiveness of the tactics employed, the need to develop straightforward techniques for improving the specificity of various CRISPR systems is still required. Strategies used for gRNA evaluation have several drawbacks as well, such as the need for substantial protein engineering, incompatibility with viral packaging restrictions, and an increase in the number of system components. Published findings on the impact of the nucleotide composition and gRNA secondary structures of the guide are contradictory. Although there is consensus that guides with slight modifications like changing the nucleotide composition and extending the secondary structures can efficiently improve Cas12a nuclease efficiency for gene editing, (Kocak et al., 2019; Liao et al., 2018).

1.7.2. Base composition

Guide RNA engineering and protein engineering of Cas9 or Cas12a effector nucleases (Huang et al., 2022) are two ways that the CRISPR system might be improved as a gene editing tool. Engineering a nuclease protein will be more laborious and time-consuming. When compared to gRNA sequence modification, which can be easily produced, there will also be less concern about protein stabilization during large-scale production (Moon et al., 2019).

The base composition of secondary structures and how they affect the performance of the guide RNAs have been investigated, and several sets of guides with specific designs have been identified to enhance the gRNA's performance. The distribution of bases (A, T, G, and U) in the hairpin region (the tails, stems, and loops) and the occurrence of single repeated bases (TT, GG) and mixed proportions (TGT, GAG, ATA) have been shown to improve how well target DNA is cut (Wang et al., 2019).

1.7.3. Secondary structure

Secondary structure impact was evaluated on GFP-HEK work through secondary structure alteration, a similar modification method was used on additional cell types to see whether it would result in a similar gene editing outcome. The Ai9 mouse primary myoblast gene was edited by guide RNA secondary structure extension. When compared to nonextended RNAs, the experiment results demonstrated that extended secondary structure accomplished a significant amount of gene editing (25-60%). (Park et al., 2018). In another work done by Kocak (2019), they discovered that gRNA

secondary structure modification influences the specificity of the guide by a 55-fold increase when tested on well-known and characterized off-target sites (Kocak et al., 2019).

The significance of the secondary structure's hairpin in Cas12 processing was also discovered, using altered repeat sequences with intact stem-loop structures to evaluate how the feature affects target DNA cleavage. The guides were also created with different loop and stem lengths, as well as single nucleotide alterations. The results demonstrated that single nucleotide alterations could not cleave completely or just partly. The findings led to the conclusion that bases in the stem play a role in the cleavage of target DNA. Also, reducing the length of the secondary structure loop had some influence on cleavage, but extending the loop had no noticeable effect. These studies on length and base rearrangement underscore the critical role that Cas12a's secondary structure stem-loop plays (Kocak., et al 2019).

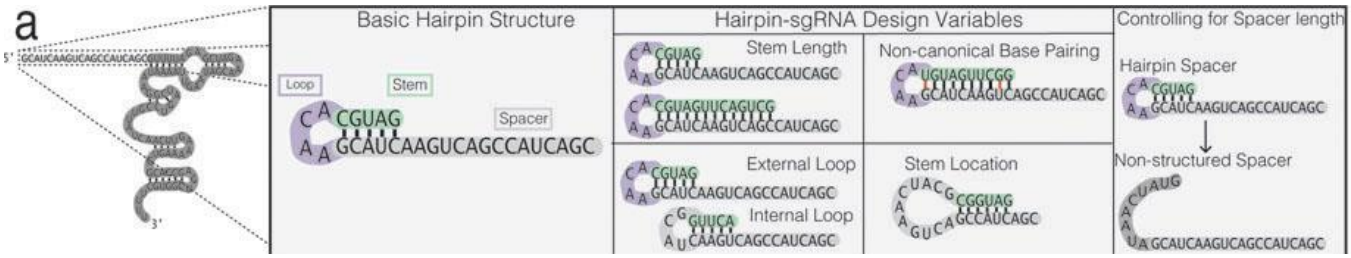


Figure 1.7.3.1 Showing parts of guide RNA hairpin structure and different parts and base modifications . gotten from (figure adapted from Kocak, 2019)

1.8. Guide RNA secondary structure preservation

The gRNA is one of the most critical components of the CRISPR system as it locates target sequences. Its structural characteristics are crucial in its interactions with Cas proteins (Cas9 and Cas12a), and how successfully the gRNA forms an RNP complex is heavily reliant on its structural elements (Riesenberg et al., 2022). Using an in vitro DNA cleavage reaction, modifications to the secondary structures of gRNA have been demonstrated to have variable effects on Cas12a nuclease activity (Riesenberg et al., 2022). The effect of secondary structure was further investigated by testing 5 gRNAs, three with naturally occurring secondary structure and two with modified secondary structure, both of which were utilized to target the mouse MeCP2 gene. The experiment results demonstrated that gRNAs with changed secondary structures have a better cleavage efficiency (Teng

et al., 2019). Kocak et al study also affirms how secondary structure is crucial for modulating Cas12a nuclease activity (Kocak et al., 2019).

1.9. Enhancement of Cas12a guide RNA activity predictability using deep learning

Testing designed gRNAs prior to gene editing research may be unrealistic because it might be labour-intensive, particularly when large pools of gRNA are to be used. To circumvent this problem, in-silico gRNA design has been employed. This method gives a much more dependable method of assessing how successfully the gRNA would perform, which helps in the elimination of several issues related to genome editing using the CRISPR system. Many online tools and programms are currently available to help scientists design effective gRNAs. Despite their effectiveness, these technologies make it impossible to approximate the specificity of the gRNAs generated. Several online tools are available for Cas9-mediated genome editing. This is a result of the nuclease being the first to be developed, approved, and widely adopted for genome editing. When compared to the tools available for Cas12a, there are few of them. Despite all the challenges, many research teams are working to develop new tools compatible with the Cas12 system.

Programs have been developed for predicting the activities of Cas12a; an example is "Seq-deepCpf1' (Kim et al., 2018). Seq-deepCpf1 was developed using a deep-learning-based regression model on a large pool of Cas12a gRNA. They used a total of 19,255 gRNA data sets. The data sets were used to test the activity of lentiviral integrated target sequences in HEK293T cells using a high-throughput approach. From the data, they were able to generate two sets of data sets (HT 1 and HT 2) (Figure 1.9.1), which were used to generate the indel frequencies achieved by the guides.

To further validate the results from Seq-deepCpf1, the results were compared with conventional machine learning algorithms (L1L2-regularised linear regression, L2-regularised linear regression, L1-regularised linear regression, and gradient-boosted regression trees) via Spearman correlation. The machine learning algorithms are the best-performing state-of-the-art programs used to predict the activities of Cas9. From their analyzed result, they arrived at the conclusion that the Spearman correlation of Seq-deepCpf1 in the cross-validation was significantly higher than that of conventional machine-learning-based algorithms when a sufficiently large data set was used (Kim et al., 2018).

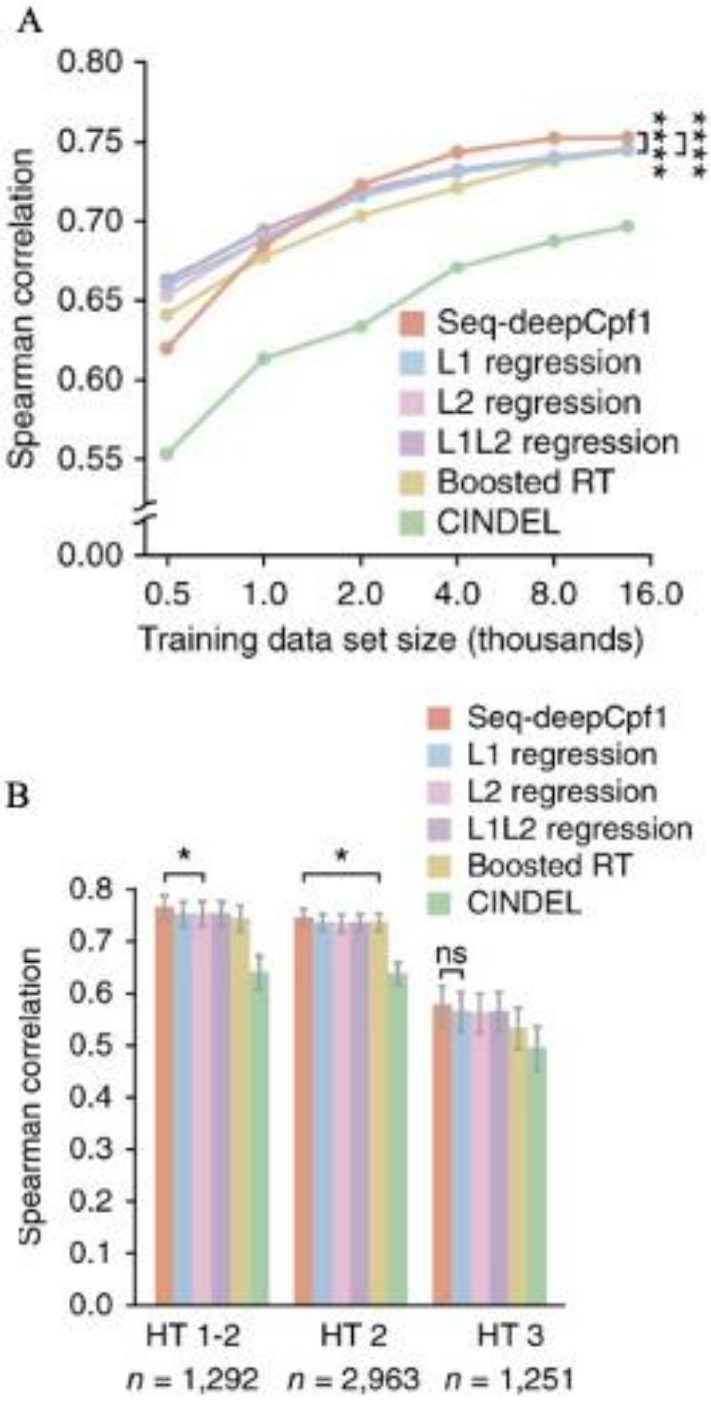


Figure 1.9.1. Prediction of Cpf1 activity based on the target sequence composition between Seq-deepCpf1 and other conventional machine learning models (**A**) Cpf1 activity prediction models trained on different size data sets were nested cross-validated. Each point is the average of the results of ten outer folds. For six different training data set sizes, the Spearman correlation coefficients between empirically obtained indel frequencies and projected scores from Seq-deepCpf1 and other standard machine learning algorithms are displayed. (**B**) Prediction model performance comparison. The Spearman correlation coefficients between observed indel frequencies and estimated Cpf1 activity scores are provided for three separate test data sets (HT 1-2, HT 2, HT 3). (left to right; *P = 0.015, *P = 0.026, and ns = not significant; Steiger's test) (figure adapted from Kim, 2018).

Another deep learning strategy called DeepCpf1 (different from Seq-deepCpf1) has been developed and is used to predict the activity and specificity of Cas12a gRNA. The program is used to capture the importance roles nucleotide position plays in the gRNA. Utilizing two classifiers convolutional neural networks (CNN), which function similarly to the sorting method implementation of image recognition. To get a solid prediction, they also combine the predictions with high-throughput profiling experiments. On-target activity using the matched DNA sequences was predicted by the first classifier, while the second classifier was used to forecast the off-target impacts using the mismatched DNA sequences. To determine the effect of sequence position and sequence compositions, they utilized their training models and were able to discover highly significant combinatorial correlations between the sequence position and sequence compositions of the gRNA. The results of their methods provide a new perspective on how the best Cas12a gRNA can be adopted for experiments and, at the same time, reduce the number of gRNAs that need to be validated as the ideal guide for a targeted gene (Luo et al., 2019).

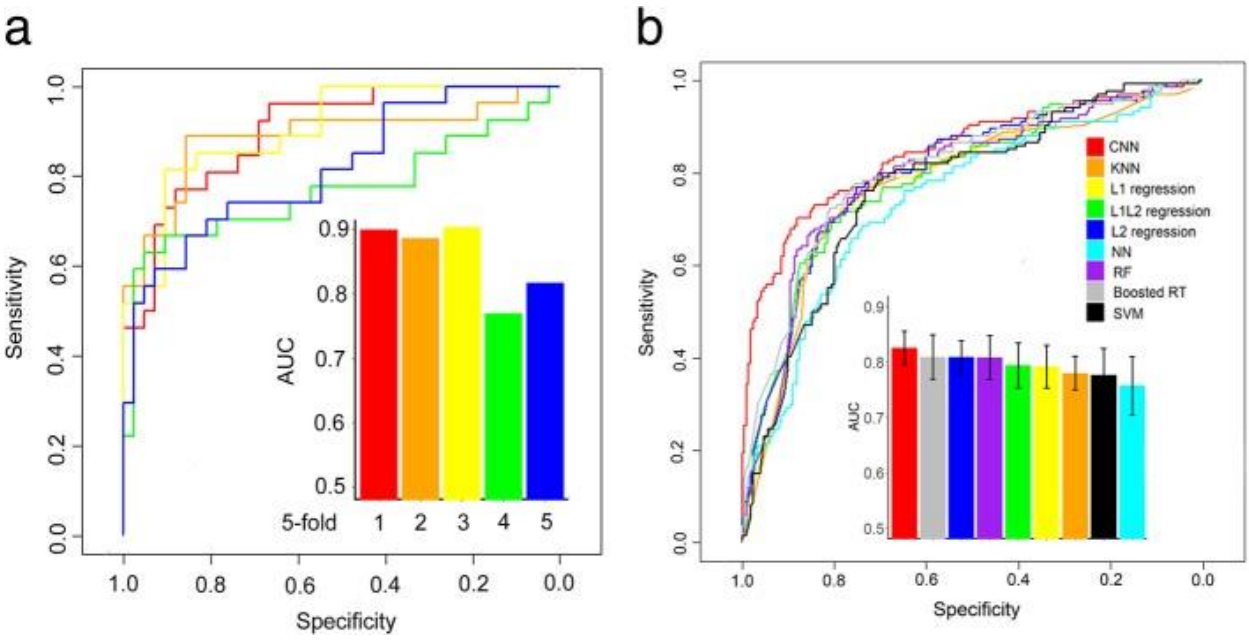


Figure 1.9.2. Showing the use of deep convolutional neural networks for prediction of off-target specificities of Cpf1 guide RNAs. (**A**) DeepCpf1 performance was evaluated using ROC curves and fivefold external cross-validation. (**B**) ROC curves and AUC values comparing CNN and other machine learning algorithms' performance (figure adapted from Luo, 2019).

1.10. Methods and strategies employed for improving Cas9 guide RNAs

To better understand how the CRISPR system works, most of the research done on Cas9 nucleases is being carried out on other nucleases. This way, important strategies that were used for improving the Cas9 system can also be adopted to make other systems used for gene editing more effective. The widespread use of Cas9 for gene editing has sparked much interest in how to enhance it for gen e editing. One of the most popular ways is to investigate the role of gRNA in cleavage efficiency, and the gRNA is being altered to enhance its function (Filippova et al., 2019). Took an interesting turn when the two traditional Cas9 gRNA molecules were engineered into a single molecule by combining the crRNA and tracrRNA to generate the sgRNA. These innovations in the gRNA architecture make it easier to modify (Marx, 2020; Safari et al., 2019).

How effectively Cas9 could cleave targets could be determined by how well it is able to generate a high level indel frequency (Scott et al., 2019). Scott et al. showed this when they created an RNP complex using a dual-gRNA (dgRNA) with an altered tracrRNA for Cas9. Upon using the resulting RNP complex to cleave a target of interest, they discovered that cleavage with the RNP complex having its tracrRNA sequence modified enhanced the activity of the complex, which caused high indel formation across several sites. This was achieved by replacing uridine amino acids in tracrRNAs, the uridine nucleotide changes significantly improved Cas9 RNP knockdown of HIV reporter cell lines. That also tested if similar outcomes could be achieved with other genes using the same strategy as tested on the HIV co-receptor and C-C chemokine receptor type 5 (CCR5). Analysed data corroborate their previous finding about the improved RNP complex activity by uridine replacement, then arrive at the conclusion that the approach could be adopted for research and therapeutic use (Scott et al., 2019).

1.10.1. Light activated guide RNA

Organisms are known to react to light, and similarly, biological processes are sometimes affected by light. Light has now been adopted in experiments to control biological processes, due to its easy control and non-intrusive nature (Brown et al., 2021). The CRISPR/Cas system is one of the areas in which the use of light has been adopted, deploying it as a photoinducer. Complex engineering procedures such as guided evolution, mutant selection, and genetic code scrutiny are done before they can be certified viable (Hemphill et al., 2015). To make a photoinducible gRNA, the spacer sequence

section of the guide is masked from hybridising with the target sequence, which could then be activated via light-induced photolysis to restore target site hybridization (Brown et al., 2021). Hybridization of gRNA: dsDNA has been blocked by the adoption of 6-Nitropiperonyloxymethylene (NPOM) at the spacer region, and hybridization of gRNA:dsDNA can be restored via photolysis of the caging molecule (Figure 1.10.1.1) (Zhou et al., 2020). This is done by inhibiting the ability of the target uridines and guanosine oligonucleotides to pair with their complementary gRNA sequence by including NPOM groups every 5–6 bases of a DNA oligonucleotide. Zhou et al. also tested if Cas9:gRNA complex formation can be inhibited by using NPOM-caged uridines and guanosines; upon introduction, Cas9 and gRNA did not form the RNP complex, which further inhibited the complex's ability to interact with the target DNA. RNP complex formation was restored by decaging the nucleotide with quick light exposure, and the resulting complex was able to interact with targets in cells and animals. NPOM-caged gRNA special features can be included in genome editing strategies, where they will serve as a quick and non-invasive approach that can be used to activate gRNAs before target cleavage. The feature could also be invaluable when introducing RNP complexes into cells and animals without the risk of the effector complex being active before reaching its target site (Zhou et al., 2020).

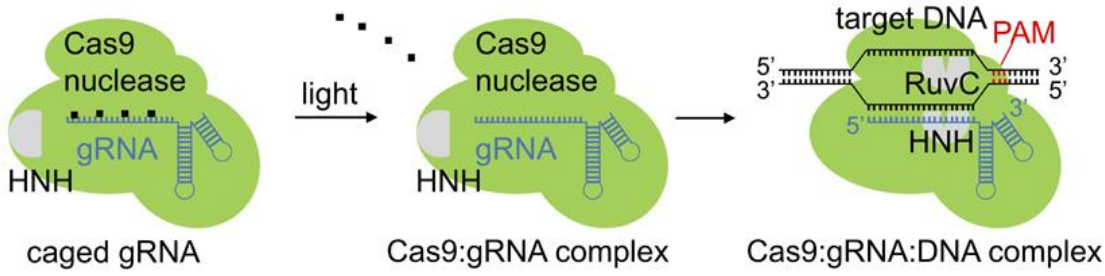


Figure 1.10.1.1. Showing how NPOM-photocaging groups abolish RNP binding to the target dsDNA until they are photochemically cleaved, thereby generating an active Cas9:gRNA complex (figure adapted from Zhou, 2020).

1.10.2. Guide RNA controlled by small molecules through chemical decaging

Like light, small molecule-induced decay has been adopted for integration with the CRISPR system to add an extra function to how the technology can be used. Small molecules have some advantages when used in biological processes, like the ability of the molecules to penetrate deep-lying tissues,

efficient handling without the use of specialized or heavy laboratory equipment, and the user's ability to measure dosages that are needed for specific procedures (Brown et al., 2021).

Habibian et al. in their study, showed that "azide substituted acyl imidazole" can be utilized to mask the activity of gRNA, blocking RNP complex formation. They did this by carrying out aryl azide stadinger reduction on the gRNA's numerous ribose 2'-OH groups, which were randomly acylated by an azide-substituted acyl imidazole reagent. The acetylation restricts the connection between the gRNA and the Cas9 nuclease. To reverse these effects, the gRNA is treated with phosphine (R3P), which eliminates the covalently attached azidomethylnicotinyl (AMN) via Staudinger reduction (Habibian et al., 2020).

1.11. Cells microfluidic platform

Currently, experimental setups entail the use of massive data inputs in high-throughput reactions. The libraries used for the reactions are huge, and it is critical to be able to run these reactions across all members of the library efficiently. Microfluidic droplets have been developed to work as a very effective ultrahigh-throughput screening device to allow numerous reactions in a tiny volume to be done simultaneously. The microfluidic platform allows a high level of sorting that can produce up to 100 variations of droplets per hour, which can be scaled into compartments of different assay volumes for desired procedures. The droplets are water-in-oil emulsion droplets that can be applied to different types of experiments, such as oligonucleotide and cell-based procedures. The platform offers considerable screening capabilities that have been adopted for use with CRISPR experiments (Gielen et al., 2016). Compared to conventional membranes, microfluidic pore size and channel shape, such as the T- and Y-junctions, can be adjusted in the device itself according to user preference. Which allows for the creation of monodisperse droplets of minuscule sizes (Song et al., 2006).

A commonly used type of droplet assay is the fluorescence-activated droplet sorting (FADS), which is a highly efficient microfluidic platform that can be used in the expression of single cells that express a specific biocatalyst (Figure 1.11.1). In the customizable microfluidic FADS system, the cells are compartmentalized in emulsion droplets with a fluorogenic substrate. Due to the system's high flexibility, the droplets can be sorted according to their fluorescence intensity, which is directly related to their enzyme activity. The selection of catalysts can be carried out to characterize the active DNA and determine desirable mutants via sequencing.

The system is very robust, capable of handling 2000 droplets per second, and it can screen up to 100 library clones repeatedly at every circle of evolution. The FADS system showed that microfluidic emulsion droplets can be sorted in a feasible and efficient way for optimized screening of large libraries (Obexer et al., 2016).

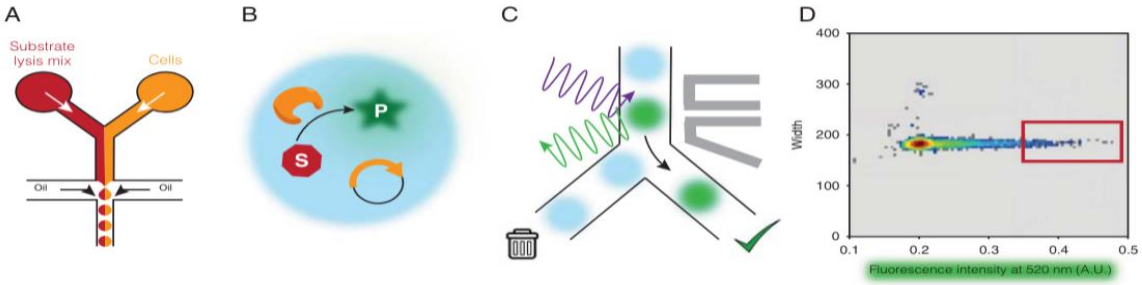


Figure 1.11.1. showing microfluidic-based FADS ultra-high throughput screening. (A) A co-flow device used for the formation of water-in-oil droplets droplets containing encapsulated cells and substrate/lysis. (B) Lysis of cells inside droplet; release of retro-aldolase encoding variant (orange) into the surrounding buffer, followed by the catalyzation of the supplied aldol substrate (red) into fluorescent naphthaldehyde product (green). (C) Sorting of droplets; Droplets are activated with a UV laser (purple) at 375 nm, and the intensity of green fluorescence emission is measured at 520 nm. When a specified threshold is reached, the electrodes (grey) are triggered, and the droplets are deflected into the collecting tube by di-electrophoresis (figure adapted from Obexer, 2016).

Holstein (2021) successfully "developed an enrichment *in vitro* on-chip components addition workflow experiment, where they evolved Savinase, a subtilisin-like protease spontaneously released by the alkalophilic bacteria Bacillus lentus. The workflow starts with (1) droplets filled with multiple copies of Single plasmids of a randomized Savinase library, (2) which is amplified via rolling circle amplification (RCA) together with reagents to increase the number of expression templates, these will lead to more enzyme molecule. (3) the emulsion droplets are incubated off-chip, in vitro transcription and translation (IVVT) reagents were added into the droplet, (4) following the addition of IVTT, the droplets were incubated off-chip for 4 hours at 37 °C, (5) after that, emulsion droplets was pico-injected with a casein tagged with BODIPY (a fluorogenic substrate) that fluoresces after cleavage. The droplets reinjected into a sorting device, and highly fluorescent droplets were collected by fluorescence-activated droplet sorting (FADS), (6) De-emulsification of selected droplets was performed, and the RCA product was recovered before restriction and ligation, (7) plasmids containing active catalyst DNA sequences were transformed into B. subtilis, then rescreened and

sequenced, (8) for further enrichment to uncover better catalysts, iterative selections was carried out, which lead to production of enhanced enzyme variants (Figure 1.11.2)".

Their work showed that, a cell-free ultrahigh-throughput screening platform can be established, likewise DNA amplification, IVTT, and substrate conversion in droplets have laid the groundwork for in vitro experiment in droplets (Holstein et al., 2021).

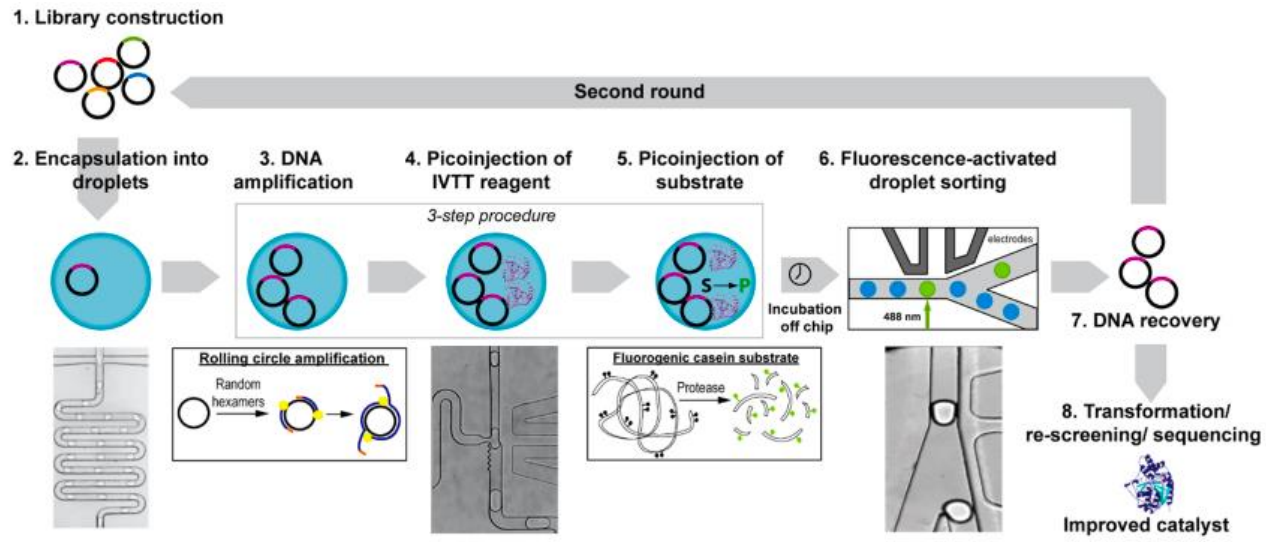


Figure 1.11.2 Showing full illustration of cell free ultrahigh-throughput screening using microfluidic droplets. (1) Bacillus lentus library construct (2) Compartmentalized library in picoliter water-in-oil droplets (3) rolling circle amplification of library. (4) Picoinjection of in vitro transcription and translation reagents steps. (5) Picoinjection of droplets with casein. (6) Droplets sorting via fluorescence-activated droplet sorting. (7) Recovery of plasmids after restriction and ligation. (8) Transformation of plasmid containing DNA sequences of active catalysts into B. subtilis and production of improved catalysts (figure adapted from Holstein, 2021).

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Devices

Table 1. Devices used in this study.

Device	Manufacturer
-4 Refrigerator	LIEBHERR
-20 Freezer	Fisherbrand
-70 Freezer	Fisherbrand
Analytical balance	Fisherbrand
Automatic pipettes	Thermo Scientific
Electronic pipettes	Roth
Balances	Fisherbrand
Bench-top centrifuge	Fisher Scientific
Bench-top mini rocker-shaker	BioSan
DNA electrophoresis system	Invitrogen
Electroporator	BIO-RAD
Fumehood	WALDNER
Gel documentation system	BIORAD, Amersham Typhoon laser-scanner
Heating bath	Fisherbrand
Incubator	Fisherbrand
Interchangeable block Heater	Fisherbrand
Laminar airflow cabinet	ESCO
Large refrigerated high-speed centrifuge	Fisherbrand
Magnetic stirrers	Fisherbrand
Microwave oven	SEVERIN
PCR thermocyler	ProFlex 3 x 32-well
pH meters	BIO-RAD
Protein electrophoresis system	BIO-RAD

Shakers	OHAUS
Spectrophotometers	Fisher Scientific
Speedvac	Fisher Scientific
Ultrasonic homogenizer	Fisherbrand
Ultracentrifuge	Beckman Optima™

2.1.2. Software and databases

Table 2. Software tools and databases that were used to plan and conduct the experiments.

Software	Function
Anaconda	Python programming environment for scientific computing
Biopython	Python programming library tools for computational biology and bioinformatics
Gel analyzer	Electrophoresis gel analysis
NCBI	Biotechnology information database

2.1.3. Chemical substances and commercial kits

Table 3. Materials, reagents, and commercial kits used in the study.

Product	Manufacturer
Glycerol, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, isopropanol,	Fischer bioreagents
Tetramethylethylenediamine (TEMED), bromphenol blue, phenol red	Bio-Rad
CaCl2, KCl, dimethyl sulfoxide (DMSO), NaCl, 2-mercaptoethanol, Imidazole, ethylendiamin-tetraessigsaure (EDTA), aceton,	Carl Roth
Tris(hydroxymethyl)aminomethane hydrochloride (TRIS-HCl)	Fischer Scientific
PVDF filters	FischerBrand
Acrylamide (AA), bisacrylamide (BAA), HCl, NaOH, piperazine-N,N′-bis(2-ethanesulfonic acid) (PIPES)	Fluka
Coomassie Brilliant blue R-250	Merck

Primers	Metabion International AG
phenylmethylsulfonyl fluoride(PMSF)	Serva
Agar-agar, acrylamide, ampicillin, NiCl2, ammonium persulfate (APS), Luria–Bertani (LB) medium, yeast extract, glycine, thymol blue, CH_3COOH, sodium dodecyl sulfate (SDS),	SIGMA Aldrich
PageRule prestained protein ladder, DNA GeneRuler DNA Ladder Mix, GeneJET PCR Purification Kit, dNTP mix, isopropylthio-β-galactoside (IPTG), Phussion DNA polymerase and its buffer, 2x RNA Loading Dye, GeneJET Plasmid Miniprep Kit, T4 ligase and its buffer buffer, BCiVi restriction enzyme, TopVision Agarose, dithiothreitol(DTT), FastAP Thermosensitive Alkaline Phosphatase	ThermoFischer Scientific
99.8 % ethanol	Honeywell

2.1.5. Bacterial strains

DH5a strain used in this study was kindly provided by Vilnius University, Institute of Biosciences, Department of Microbiology and Biotechnology and was used for plasmids transformation.

2.1.6. Recombinant protein expression strains

E. coli BL21 (DE3): It carries the gene for T7 RNA polymerase under the control of a lac UV5 promoter, allowing the expression of the T7 RNA polymerase under induction with IPTG. BL21 does not have Lon and OmpT proteases. The absence of these enzymes decreases the risk of recombinant protein degradation.

2.1.5. Plasmids

pIF502 (Jones et al., 2021) was used as a Cas12a nuclease protein expression vector; the plasmid carries His6, TwinStrep, and SUMO tags.
pIF503 (Jones et al., 2021) was used as a dead Cas12a inactive-nuclease protein expression vector; the plasmid carries His6, TwinStrep, and SUMO tags.

2.1.6. Media

Liquid LB media: 25 g (composition: 5 g/L yeast extract, 10 g/L NaCl, 10 g/L tryptone, pH 7.0) of medium was dissolved in 1 L of deionized water. Media was autoclaved for 20 minutes at 121 ^{o}C at 1 atm pressure.

Solid LB media: 1 % of agar-agar was added to the liquid LB medium. Media was autoclaved for 20 minutes at 121 ^{o}C at 1 atm pressure.

Liquid Terrific broth media: Terrific broth was made from commercial broth powder from ROTH; 50.8 g of powder was used for 1 liter of broth, with 4 ml of autoclaved glycerol per liter of medium.

2.1.7. Solutions

General solutions

AA/BAA (30%): 29.2 g of AA and 0.8 g of BAA were dissolved in a small quantity of deionized water; the solution was diluted to a final volume of 100 mL and filtered through a glass filter. The solution was stored in a dark container at 4 °C.

Ampicillin (50 mg/mL): 0.5 g of ampicillin was dissolved in 10 mL of deionized ampicillin was dissolved in 10 mL of deionized water; the solution was filtered through a sterile PVDF filter and stored at -20 ^{o}C.

APS (10 %):100 mg of APS was dissolved in 1 ml of deionized water. Bradford dye:0.1 g of Coomassie Brilliant blue R-250 was dissolved in 50 mL of 96% ethanol. When the precipitate has dissolved, 100 ml of orthophosphoric acid was added. Dilution to the final volume of 1 L was done with deionized water. The solution was filtered through a glass filter and stored at 4 °C.

Carbenicillin (50 mg/mL): 0.5 g of carbenicillin was dissolved in 10 mL of deionized carbenicillin was dissolved in 10 mL of deionized water; the solution was filtered through a sterile PVDF filter and stored at -20 ^{o}C.

IPTG (1 M): 2.4 g of IPTG was dissolved in 10 mL of deionized water, filtered through a sterile PVDF filter, and stored at -20 ^{o}C.

Loading dye for protein electrophoresis samples (6x) (non-reducing conditions): 300 mM TRIS-HCl (pH 6.8), 0.6 % bromophenol blue, 12 % SDS, and 60 % glycerol were mixed and stored at -20 °C.

PMSF (0.1 M): 1.74 g of solid PMSF was dissolved in 10 mL of ethanol or isopropanol. The solution was stored at room temperature.

Polyacrylamide gel dye: 0.6 g of Coomassie Brilliant blue R250 was dissolved in 113 ml of 95 % ethanol with 23 ml of acetic acid. Dilution to the final volume of 250 mL was done with deionized water. The solution was stored at room temperature.

TAE electrophoresis buffer (1x): 20 mL of TAE (50x) was diluted to 1 L with deionized water.

TAE electrophoresis buffer (50x): 50 mM of EDTA, 2 M of TRIS, and 57.1 mL of acetic acid were diluted to 1 L with deionized water. The solution was stored at room temperature.

TRIS (0.5 M, pH 6.8): 0.5 M of TRIS-HCl was dissolved in a small volume of deionized water. The pH was adjusted to 6.8. Dilution to the final volume was done with deionized water. The solution was stored at 4 °C.

TRIS (1.5 M, pH 8.8): 1.5 M of TRIS-HCl was dissolved in a small volume of deionized water. Dilution to the final volume was done with deionized water. The solution was stored at 4 °C.

TRIS-SDS-glycine protein electrophoresis buffer (10x, pH 8.3): 1.9 M of glycine, 35 mM of SDS, and 25 mM of TRIS were dissolved in deionized water. The solution was stored at room temperature.

Nucleaseq buffer (1x): 20 mM of HEPES, 150 mM of KCL, 10 mM of MgCl2, 2 mM of DTT at pH 7.5. Dissolved in a small volume of deionized water. Dilution to the final volume was done with deionized water. The solution was stored at 4 °C.

NucleaSeq Buffer (10x): 200 mM of HEPES, 1.5 M of KCL, 100 mM of MgCl2, 1M of DTT at pH 7.5. Dissolved in a small volume of deionized water. Dilution to the final volume was done with deionized water. The solution was stored at 4 °C.

2.1.8. Protein purification buffers

Lysis buffer for protein purification by Strep-Tactin: 20 mM Na-HEPES at pH 8.0, 1 M NaCl, 1 mM EDTA, 5% glycerol, 0.1% Tween-20, 2000 U DNase (GoldBio 10700595), 1 HALT protease inhibitor (Fisher Scientific 10516495), and 1 M PMSF were dissolved in a volume of cold deionized water corresponding to approximately 80% of the final intended volume of the solution. The pH was adjusted to 8.0. Dilution to the final volume was done with deionized water. The solution was stored at 4 °C.

Elution buffer for protein purification by Strep-Tactin: 20 mM Na-HEPES at pH 8.0, 1 M NaCl, 5 mM desthiobiotin, 5 mM MgCl2, and 5% glycerol were dissolved in a volume of cold deionized water corresponding to approximately 80% of the final intended volume of the solution. The pH was adjusted to 8.0. Dilution to the final volume was done with deionized water. The solution was stored at 4 °C.

Storage buffer for protein purification by Strep-Tactin: 20 mM HEPES-KOH, 150 mM KCl, 5 mM MgCl2, and 2 mM DTT buffer were dissolved in a volume of cold deionized water corresponding to approximately 80% of the final intended volume of the solution. The pH was adjusted to 8.0. Dilution to the final volume was done with deionized water. The solution was stored at 4 °C.

2.2. Methods

2.2.1. LIBRARY DESIGN

a. Template design

The ssDNA oligonucleotide template library was designed using Biopython. The program created on Biopython was used to iterate through the NCBI database for 12,000 cancer genes of interest.

To search the database, after providing the intended gene of interest, keywords were used to describe the type of data to be searched for in NCBI. The following keywords were used: species = “humans”, database (dp) = “nucleotide”, id = “accession”, return type: “gb”. The “species” keyword indicates that when searching through the human database, the species of interest is human. Dp = “nucleotide” indicates that when searching through the human database, the data of interest is the nucleotide. The “accession” keyword indicates that when searching through the ID of the nucleotide database, the data of interest is the accession numbers of the genes, for which the sequence of each accession is provided. Return type: The “gb” keyword represents the type of file to be returned, which in this case should be a gb file (Figure 3.2.3). A Cas12a nuclease target site in each gene was selected. The gRNAs targeting the target sequences in the library were designed using Benchling (Benchling.com) to include Cas12a secondary structure with different modifications on their stem, loop, and tail regions. ssDNA oligonucleotide libraries were synthesized and purchased from Metabion.

b. Library members

Each library member was designed to perform two primary functions: it produces gRNA using the T7 promoter site via transcription, and it also provides a target for Cas12a nuclease by utilizing the PAM recognition site. The template also has other important sites like the forward and reverse primer sites, the barcode, the constant region, and the BciVI digestive enzyme recognition site.

c. Synthesis

The delivered ssDNA oligonucleotide template library was amplified using Phusion PCR to make a dsDNA template library PCRs were conducted according to the following protocol: initial template denaturation at 98 ^{o}C for 30 seconds, denaturation at 98 ^{o}C for 5 seconds, primer annealing at 50 ^{o}C, and extension at 72 ^{o}C for 10 seconds (denaturation, annealing and extension steps were repeated for 30 cycles). The PCR products were cloned into a linear plasmid using the blunt-end PCR kit. Colony PCR was used to identify successful clones using primer sites on the plasmid and the template and was visualized using a 2 % agarose gel. Bacterial cells from colonies of confirmed positive clone were grown at 32 °C and 225 rpm for 16 h. Plasmid DNA from colonies was extracted from the bacterial cells using the Thermofisher Genejet plasmid miniprep kit and then quantified using a nanodrop spectrophotometer.

Table 4. Primers used to amplify clones inserted during ligation.

Primer	Primer sequence (5’ → 3’)	Description
Template forward primer	CGGATTATGCTGAGTGATATC	Forward primer on the template
Plasmid M16 reverse primer	CAGGAAACAGCTATGACCATG	The primer site is next to the template insert site

2.2.2. Transcription of gRNA from Template DNA

a. PCR of ssDNA template into dsDNA

PCR amplification was carried out on an ssDNA template to make a double-stranded template; the templates were amplified using forward and reverse primers. The template also contains other important sites like a barcode sequence, T7 promoter sequence, guide RNA production sequence, Cas12a PAM recognition and target sites, and the recognition site of BciVI restriction endonuclease endonuclease(Figure 3.3.1). PCR was performed according to the following protocol: initial template denaturation at 98 ^{o}C for 30 seconds, denaturation at 98 ^{o}C for 5 seconds, primer annealing at 60 ^{o}C, extension at 72 ^{o}C for 5 seconds (denaturation, annealing, and extension steps were repeated for 25 cycles).

Table 5. Primers used to make dsDNA from ssDNA template. Underlining of the sequences indicates recognition sequences of BciVI restriction endonuclease.

Primer	Primer sequence (5’ → 3’)	Description
Template forward primer	CGGATTATGCTGAGTGATATC	Forward primer on the template
Template reverse primer	CGTGTACCTATGCATAGTTC	Template reverse primer BCiVI restriction site

b. Template preparation

The resulting products of the PCR reaction were purified using the GeneJET PCR purification kit performed according to manufacturer's protocol. The concentration and quality of the PCR products were determined using the NanoDrop™ 2000 (Thermofisher)

c. Transcription of DNA

Transcription of dsDNA was carried out using TranscriptAid T7 High Yield Transcription Kit from Thermo Scientific. Transcript were cleaned using GeneJET RNA Purification Kit from Thermo Scientific. The concentration and quality of the transcript were determined using the Nanodrop spectrophotometer.

d. Quality and production evaluating

Template DNA quality and product was checked by nucleic acid electrophoresis using a 12% native polyacrylamide gel, while RNA quality and product was checked by nucleic acid electrophoresis using a 12% denaturing polyacrylamide gel.

Native polyacrylamide electrophoresis

Native polyacrylamide gel was prepared by mixing 3.6 mL polyacrylamide (29:1), 3.6 mL H20, 1.2 mL 10x TBE, 100 µL APS, and 10 µL TEMED. Gel was allowed to polymerize. When the gel polymerized, it would be transferred to the electrophoresis apparatus with 1xTBE buffer. 5 µL of the DNA samples were loaded in DNA loading dye with 1 µL of 6x Bromophenol blue. The GeneRuler Ultra Low Range DNA Ladder was used to evaluate the size of the DNA bands. The electrophoresis was pre-run at a voltage of 8 V/cm^2 for about 20 minutes. The electrophoresis was performed at a voltage of 8 V/cm^2 for about 90 minutes. After running, the gel was stained with 1 µL of SYBR GOLD gel stain in a container containing 1X TBE, then shaked on a rocker for 30 minutes. The presence of the nucleic acids was determined by scanning the gel with an Amersham Typhoon laser scanner at a 488 nm wavelength.

Denaturing polyacrylamide gel

Denaturing polyacrylamide gel was prepared by melting a mixture of 5.25g of urea, 3.75 mL of polyacrylamide (29:1), 3.75 mL of H_20, and 1.25 mL of 10x TBE to boiling, then chilling before adding 62.5 µL of APS and 12.5 µL of TEMED. Gel was allowed to polymerize. When the gel polymerized, it was transferred to the electrophoresis apparatus with 1X TBE buffer. 5 µL of the DNA samples were dyed with 1 µL of 6x Bromophenol blue DNA Loading Dye. The GeneRuler Ultra Low Range DNA Ladder was used to evaluate the size of the DNA bands. The electrophoresis was pre-run at a voltage of 8 V/cm^2 for about 20 minutes. The electrophoresis was performed at a voltage of 10 V/cm^2 for about 90 minutes. After running, the gel was stained with 1 µL of SYBR GOLD gel stain in a container containing 1X TBE, then rocked on a rocker for 30 minutes. The presence of the nucleic acids was determined by scanning the gel with an Amersham Typhoon laser scanner at a 488 nm wavelength.

2.2.3. gRNA – Cas12a nuclease hybridization

Filter assay : Preparation of the enzyme-gRNA binding

Nitrocellulose is a nitrated derivative of cellulose that has had all its hydroxyl groups substituted. It therefore presents hydrophobic surfaces to protein molecules. The filters have lengthy holes through which the molecules must pass, but they are far bigger than the biological samples that are preserved, therefore the physical foundation of the process is adsorption to the hydrophobic surface. In the past, it has been assumed that this adsorption property is universal and that all proteins, even short peptides with molecular weights >2 kDa (18), will stick, although it is known that they do so with different efficiencies.

Cellulose Nitrate Membrane Filters from Cytiva (Whatman^{TM}) were used; the filters have a pore size of 0.45 µM. Filters are presoaked at room temperature in nuclear-seq buffer for one hour before use. A blunt-end tweezer was used to flip the filter at 30-minute intervals and to ensure the filter was completely wetted. Cas12a gRNA, control gRNA, and Cas12a nuclease protein samples for the reaction was prepared in Nuclear-seq buffer; the solutions was prepared so that the DNA and control DNA solutions concentrations were 40 nM, while the gRNA solution concentration is 25 nM. 900 µL of template DNA, control DNA, and gRNA was each added to a 5 mL tube and mixed using the pipette. The mixture was incubated at 22 °C for 30 minutes to allow the formation of the gRNA-DNA

RNP complex. The incubated mixture was centrifuged before use. The presoaked filter was placed into the filtration manifold and tightly screwed in. 600 µL of sample was then applied to the filter using a sterile syringe. Filtrate from the reaction was collected. The control experiment was carried out like the main reaction; the only difference was that the control experiment mixture contains only control gRNA and template DNA, while Cas12a gRNA was not included. The presence of gRNA-Cas12a RNP was detected using a 12% denaturing polyacrylamide gel.

Bead assay

Preparation of Cas12a nuclease protein binding with Nickel resin beads (HisPur Ni-NTA Resin): The HisPur Ni-NTA Resin beads was chosen for the binding assay between Cas12a gRNA and Cas12a Nuclease protein. The 6-12 histidine residues on the His Tag can anchor a protein onto a surface functionalized with Ni^{2+} NTA motif. Where nitrogen lone pairs on each histidine residue coordinate with the remaining two sites on hexavalent Ni^{2+} ions on the bead held by the NTA chelator group.

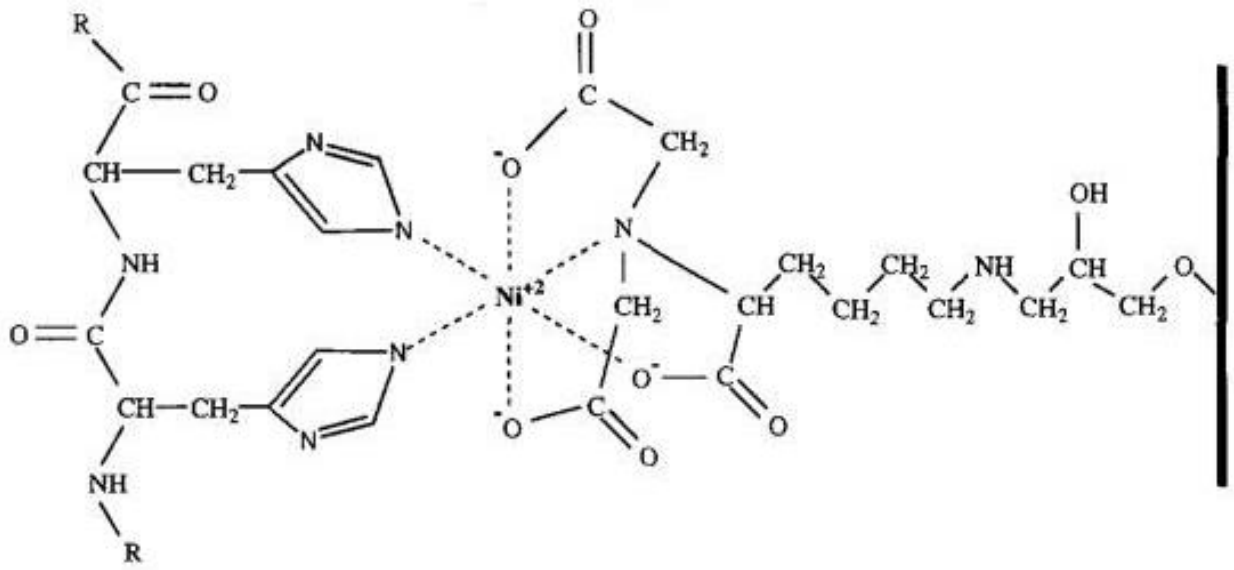


Figure 2.2.3. Interactions between proteins histidine affinity tag and two immobilized Ni^{+2}–NTA affinity chromatography matrices (figure adapted from, Bornhorst and Falke, 2000).

Cas12a gRNA, control gRNA, and Cas12a nuclease protein samples for the reaction were prepared in nucleaseq buffer; the solutions were prepared so that the DNA and control DNA solution concentrations are 40 nM, while the gRNA solution concentration is 25 µM. 200 µL each of template

DNA, control DNA, and gRNA were added into a 1 mL tube and mixed using the pipette. The mixture is incubated at 22 °C for 30 minutes to allow the formation of the gRNA-DNA RNP complex. The incubated mixture is centrifuged before use. 159 µL of hybridised mixture is added into a 1.5 mL tube, 100 µL of bead slurry is introduced into it and mixed, and it is then incubated for 30 minutes on ice while swinging on a rocker. The sample is centrifuged at 4,000 rpm for 5 minutes at 4 °C, and the supernatant is removed. Precipitated sorbent is removed from a new 1.5 mL eppenodorf. 50 µL of Ni-NTA elution buffer is added to the sorbent, the mixture is mixed, and after sedimentation, the buffer with the eluted proteins is removed and prepared for the SDS-PAGE analysis.

Assay analysis

Samples are used for the analysis using the sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) method for protein detection. gRNA detection was done by running the sample and non-filtered hybridized mixture on a 12 % denaturing polyacrylamide gel.

2.2.4. Cleavage reaction of template DNA by Cas12a nuclease

Preparing transformation Competent DH5α E. Coli

Sterile technique was strictly observed when making Chemical competent DH5α E. Coli cells, All centrifuge bottles, tubes, solutions, pipets, etc. were sterile. 4 ml of LB was incubated with DH5a cells stock in glycerol and incubate overnight at 37ºC. The entire 4 ml is transferred into a 1 L erlenmeyer flask containing 500 ml LB and incubate at 37ºC. Cell growth was monitored after 2 hours at regular intervals by measuring OD of an aliquot. Cells were allowed to grow until the OD550 = 0.45 to 0.6. The culture was transferred to a 2 x 500 mL centrifuge bottles and chilled on ice to 4ºC, then Centrifuged at 6,000 g for 5 min at 4ºC. Supernatant was poured off and pipette is used to remove remaining supernatant. Cells were resuspended in 125 ml ice cold 50mM $CaCl_2$. Content of each flask was combined into a single 500 mL centrifuge bottle, Centrifuge at 6,000 g for 5 min at 4ºC. Supernatant is poured off and resuspended in 21.5 mL ice cold $CaCl_2$:glycerol solution. Cells are transferred into 500 µl aliquots in sterile microfuge tubes and snap freeze in liquid nitrogen. Cells are store in -80ºC freezer. Stored cells are viable for 6 months.

Competent cell Testing

Stored chemically competent DH5α E. coli cells from a -80 °C freezer are thawed on ice. The water bath is turned on and set to 42°C. 50 mL of competent cells is transferred into a 1.5 mL Eppendorf tube; the tubes are kept on ice. 50 ng of circular DNA is added to a tube containing chemically competent cells. Then incubated on ice for 10 minutes. A tube containing competent cells and circular DNA is placed into a water bath at 42°C for 45 seconds. Tubes are returned to the freezer for 2 minutes. 1 mL of LB (with no antibiotic added) is added to the tube and incubated for 1 hour at 37 °C. The contents in the tube are microfuge at 2000 x g for about 5 minutes, the supernatant is discarded, and the pellet is resuspended in 200 µL of SOC media. For plating, about 100 µL of the resulting culture is spread (using glass beads) on LB plates with the appropriate antibiotic added (usually ampicillin or kanamycin), then grown overnight. A colony count was done to determine viable bacterial cell number.

Electrocompetent cell

All flasks are autoclaved with water and then discarded before use. All tubes and solutions are pre-chilled on ice, including boxes where cells will be stored. Frozen stock of DH5 is streaked on a LB plate (no antibiotics) and left to grow overnight at 37°C. A starter culture of cells is prepared by picking a single colony from the fresh LB plate, inoculating it in a 10 mL LB starter culture containing Kanamycin antibiotics, and allowing it to grow over night in a shaker incubator at 37°C. 1 liter of LB media is inoculated with 10 mL of starter culture and grown in the shaker at 37°C. OD 600 measurement at regular intervals until the OD 600 reaches 0.35–0.45. The cell culture is then put on ice immediately. The culture is chilled for 20–30 minutes, which are swirled occasionally to ensure even cooling. The cell culture is aliquoted into 4 x 250-mL ice-cold centrifuge tubes, then allowed to chill on ice for an additional 30 minutes. The cells are subjected to a couple of spin circles. In the first step, the cells are centrifuged at 1000g (2400 rpm in the Beckman JA-10 rotor) for 20 minutes at 4°C. The supernatant is decanted, and pellets are resuspended in 200 mL of ice-cold ddH_2O. For the second spin circle, the cells are centrifuged at 1000g (2400 rpm in the Beckman JA-10 rotor) for 20 minutes at 4 °C, the supernatant is decanted, and cell pellets are resuspended with 100 mL ddH_2O. Resuspended cells are combined into two centrifuge bottles (each bottle will now contain 200 mL of suspended cells). Suspended cells are centrifuged at 1000g (2400 rpm in the Beckman JA-10 rotor) for 20 minutes at 4°C during the third spin circle. The supernatant was decanted, and the cell pellet

in each 200-mL tube was resuspended with ddH_2O containing 10% glycerol and transferred into separate 50-mL conical tubes chilled on ice before use. Lastly, in the last circle, the responded cells are centrifuged at 1000g (2100 rpm in the Beckman GH-3.8 rotor) for 20 minutes at 4°C. The supernatant is aspirated using a sterile Pasteur pipette, and the pellet is resuspended in 1 mL of ice-cold ddH_2O containing 10% glycerol. 50 µL of cell suspension is aliquoted into sterile 1.5 mL microfuge tubes pre-chilled on ice prior to use. The cells in the microfuge tubes are frozen with liquid nitrogen. Cells are stored in the -80°C freezer.

Testing of electrocompetent cell

The electroporator (maker and model) is plugged into a power source. The condition for DH5α is set at 25 µFD (Microfarad), 200 W, and 1.8 kV, with the time constant (tau value) set at 3 msec. Electrocompetent DH5 cells in the freezer are thawed on ice. 3 µL of DNA (50 ng) is added into the tube containing 50 µL of electrocompetent cells. The tube is gently flicked to mix the electrocompetent cells and DNA mixture. Then incubated on ice for 10 minutes. The mixture is pipetted into a chilled cuvette, the cuvette is gently tapped on a flat surface to ensure the mixture gets to the bottom of the cuvette, and condensates are wiped off the sides of the cuvette. The cuvette is placed into a pulser and pursed once. Immediately after electroporation, 500 µL of SOC media is added to the cuvette to recover the cells. The mixture is then transferred into a sterile 1.5-mL microcentrifuge tube. The mixture is incubated for 60 minutes at 37°C in the shaking incubator. 50 µL of cells are plated on a LB plate containing kanamycin antibiotic and incubated overnight at 37°C.

2.2.5. Plasmids – Template DNA ligation reaction

Description of the system

The PCR products for the reaction were conducted according to the following protocol: initial template denaturation at 98 ^oC for 30 seconds, denaturation at 98 ^oC for 10 seconds, primer annealing at 65 ^oC for 10 seconds, and extension at 72 ^oC for 10 seconds (denaturation, annealing, and extension steps were repeated for 25 cycles). The resulting amplified products were purified using the GeneJET PCR purification kit. The concentration and quality of the PCR products were determined using the Nanodrop spectrophotometer as well as by nucleic acid electrophoresis using a 12% native

polyacrylamide gel. Plasmid DNA ligation was done using the Zero Blunt® PCR Cloning Kit (thermoscientific).

For the reaction, a 10:1 molar ratio of insert:vector is used to optimize blunt-ended PCR ligation efficiency. In the same way, the amount of vector in the ligation reaction can be reduced to as little as 5 ng. The lower the concentration of the PCR sample, the lower the ability of salt in the mix to inhibit the T4 DNA Ligase. PCR product needed for a 10:1 molar ratio of insert to 25 ng of pCR™-Blunt vector.

$$x \text{ (ng inserts)} = \frac{(10)\ (y \text{ bp PCR product})\ (25 \text{ ng linearized pCR®-Blunt})}{(3500 \text{ bp pCR®-Blunt})}$$

where x (ng) is the amount of PCR product of y base pairs to be ligated for a 10:1 insert:vector molar ratio.

The following setup is used for the ligation reaction: 1 µL of pCR®-Blunt (25 ng), 2 µL of Blunt PCR product, 2 µL of 5X ExpressLinkTM T4 DNA Ligase Buffer, and 1 µL of ExpressLinkTM T4 DNA Ligase (5 U/ µL). Incubate the ligation reaction at room temperature for a minimum of 1 hour. Then Proceeded to transform competent cells.

2.2.6. Transformation of construct into competent DH5a E. Coli

The ligation product was transformed into a DH5α cloning strain by mixing 50 L of pre-prepared electrocompetent cells thawed on ice with 5 µL of the ligation mixtures and incubating the samples on ice for 30 minutes. The mixtures were then transferred into an electroporation cuvette. A cuvette is placed in a pulser and pulsed once. Immediately after electroporation, 500 µL of SOC media is added to the cuvette to recover the cells. The mixture is then transferred into a sterile 1.5-mL microcentrifuge tube. The mixture is incubated for 60 minutes at 37°C in the shaking incubator. The transformed cells were plated on agarose LB medium containing Kanamycin (50 g/mL) and 1M IPTG (1 µL/mL) and grown overnight at 37 ºC.

2.2.7. Colony PCR

7 PCR tubes, each containing 5 µL of ddH_2O, were prepared for each colony on the plateA fresh toothpick tip is used to pick a few cells from each colony, then dipped into a PCR tube and streaked

onto a fresh replicate agar plate using a numbered template, i.e., all 7 colonies on a single agar plate. A new tube for negative control was labelled (use a colony that is negative). A replicate agar plate was inoculated at 37°C for 8 hours for future overnight miniprep cultures. PCRs were conducted according to the following protocol: initial template denaturation at 98 ^{o}C for 30 seconds, denaturation at 98 ^{o}C for 10 seconds, primer annealing at 6 ^{o}C for 30 seconds, extension at 72 ^{o}C for 22 seconds, and final extension at 72 ^{o}C for 10 minutes (denaturation, annealing, and extension steps were repeated for 25 cycles). The presence of the inserted template was verified by nucleic acid electrophoresis using a 1% agarose gel.

2.2.8. Cas12a and dCas12 protein expression

Two plasmids encoding Cas12a nuclease and dead Cas12a inactive nuclease genes were used to transform E. coli BL21 (DE3) cells. Bacterial cultures were grown at 37 ^{o}C in Terrific Broth (TB) medium supplemented with carbenicillin (100 µg/mL) while shaking at 220 rpm until OD600 reached a value of 0.6; afterwards, cultures were chilled at 4 °C for 10 minutes. Recombinant protein expression was induced by the addition of 1 mM of IPTG. Protein expression was continued for 24 hours at 18 ^{o}C while shaking cultures at 220 rpm. The cultured biomass was harvested by centrifugation of the culture for 20 min at 4 °C at 6000 rpm. The cultured cell biomass was stored at -20 ^{o}C.

2.2.9. Protein detection using His-tag beads

Samples of the lysed cells before and after the induction of expression of Cas12a nuclease and dead Cas12a inactive nuclease were prepared for protein expression detection using NTI Ni-His tag beads purification. 350mg of frozen cell biomass is thawed and homogenized in 1.5 mL of Ni-NTA wash buffer by vortexing it until the solution becomes homogenous. The cells are lysed using an ultrasonic homogenizer at 40% amplitude while keeping the sample in ice-cold water. The lysis procedure takes 5 minutes with pulses of 3 seconds each at 3 second intervals. The sample is centrifuged at 13,000 rpm for 10 minutes at 4 °C. The supernatant is separated from the precipitate. The precipitate is discarded. The supernatant containing soluble proteins is mixed with 100 µL of Ni^{+2} loaded sorbent and incubated on ice for 10 min while swinging on a lab rocker. The sample is centrifuged at 4,000 rpm for 5 minutes at 4 °C, and the supernatant is removed. 2 mL of fresh buffer A is added to the centrifuged sorbent; the mixture is mixed and centrifuged at 4,000 rpm for 5 minutes at 4 °C, and the

supernatant is discarded. This is repeated three times. After the last centrifugation, precipitated sorbent is removed to a new 1.5 mL eppendorf, about 1 mL of fresh Ni-NTA wash buffer is added to the centrifuged sorbent, the mixture is mixed and centrifuged at 4,000 rpm for 5 minutes at 4 °C, and the supernatant is discarded. 100–200 L of Ni-NTA elution buffer is added to the sorbent, the sample is mixed, and after the sorbent is sedimented, the buffer with the eluted proteins can be removed and prepared for the SDS-PAGE analysis. Samples are used for the analysis using the sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) method.

2.2.10. Streptag- column protein purification

Stored cell biomass is lysed by sonication at 4 °C in 35 ml of lysis buffer. The lysate was clarified by ultracentrifugation at 35,000 RCF, applied to a hand packed StrepTactin Superflow gravity column (IBA Lifesciences), and then eluted with elution buffer. The eluate was concentrated to less than 1 ml using a 30-kDa MWCO concentrator (Millipore); SUMO protease was added at 3 M; and then the eluate was incubated overnight on a rotator at 4 °C. The protein was resolved on a HiLoad 16/600 Superdex 200 Column (GE Healthcare), pre-equilibrated with storage buffer, and finally snap frozen in liquid nitrogen and stored in 10 μL aliquots at 80 °C.

Cas9 and Cas12a RNP complexes were reconstituted by incubating a 2:3 molar ratio of apoprotein and RNA (sgRNA and pre-crRNA for Cas9 and Cas12a, respectively) in RNP buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 10 mM MgCl2, 2 mM DTT) at room temperature for 30 min before each experiment.

2.3. Statistical Methods

Statistical methods are not yet being utilized in the study because the molecular research methods are still being designed, tested, and optimized. When the results are obtained, they will be submitted to a variety of statistical methodologies to ensure that the data size are sufficient, that parametric or non-parametric tests are used in accordance with the data's properties, and that statistical tests are selected that correspond with my research questions. For example, I may use the Pearson correlation coefficient to examine the distribution of my data sets as well as correlations within datasets. Data sets may be compared and ranked using Spearman rank-order correlation.

3. RESULTS

3.1. General overview

For this study, we want to show how gRNA sequence and secondary structures influence the use of Cas12a as a broad, reliable, and predictable tool for genome editing. We plan to show how different secondary structure modifications (nucleotide base composition and length) impact transcription of a double-stranded target DNA library, ribonucleoprotein complex formation between the transcribed gRNA and Cas12a protein, and cleavage efficiency on a wide range of gene targets. A designed gRNA will include varying edits along the gRNA's secondary structure, i.e., the stem, loop, and tail regions. Different edits will be made to these regions by purposely changing the base composition within these regions of the gRNA, which will involve, for example, the extension or reduction of the designed gRNA stems, loops, and tail regions. Some edits will also involve combinational base substitution along the secondary structure (repeated bases, e.g., AAA, UUU, CC) by checking for different bases. To see how these sequence alterations influence overall gRNA activity, cancer genes are exported from the NCBI database, and probable Cas12a nuclease target sequences for each gene are identified, which are then utilized to build the cancer gene target library. The gRNAs derived from the library template will include secondary structure modifications, and we will evaluate how they influence the primary role of the gRNA's spacer region. This study will provide a strategy for increasing the specificity of Cas12a gene editing. This will be beneficial in determining how gRNAs influence CRISPR Cas12a DNA cleavage efficiency.

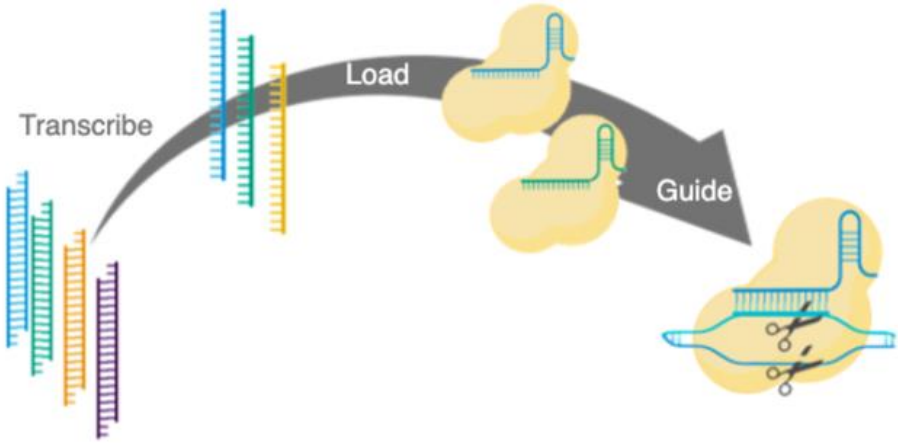


Figure 3.1. gRNA lifetime: transcription of gRNA from DNA template, formation of RNP complex and cleavage of target gene by RNP complex (Made via BioRender).

3.2. Library design

A robust approach is used to generate a library of 12,000 DNAs containing paired gRNA templates and targets that query the role of structure and sequence in gRNA function. Biopython, a Python based biocomputer software manipulating biological data, was used to create the template library. The use of Biopython to create the library offers a lot of versatility in its application as a tool that helps the programme's interaction with other databases, libraries, and modules like Entrez and SeqIO.

Gene target	3' – Spacer sequence	Repeat sequence (edits) – 5'	Edits region
A1BG	TAATTTCTACTAAGTGTAGAT	TAGATGTGAATCATCTTTAATAA	Tail
		TAGATGTGAATCATCTTTAATACA	Tail
		TAGATGGGTGAATCCCATCTTTAAT	Stem
		TAGATGTTGAATACATCTTTAAT	Stem
		TAGATGTGAGGATCATCTTTAAT	Loop
		TAGATGTTTGAATTTCATCTTTAAT	Loop

Figure 3.2.1. Showing different edits made on the stem, loop, and tail sections of the gRNA repeat sequence for the A1BG cancer gene target template in the library

Figure 3.2.2. Biopython logo (https://biopython.org/)

The biopython code was used to search the National Centre for Biotechnology Information (NCBI) database for cancer genes of interest gotten from the "Human GeCKO v2 genome-wide library" from Zheng Zhang Laboratory. While looking for genes, the integration of the two systems allows considerable versatility. The manner in which the genes were searched is an example of such accessibility. Formerly, while searching the NCBI database through the website, just a single gene details (nucleotide bases or translated protein sequences) could be retrieved at a single event. This traditional method of searching is inefficient if there are more inputs to be searched in the database, wasting a lot of time.

Figure 3.2.3. NCBI logo (https://www.ncbi.nlm.nih.gov/)

This issue was solved by using BioPython and MyGene.py library (Lelong, n.d.) to search the NCBI database. So, biopython code may be used to search the database sequentially for various genes or inputs. The search results were used to retrieve thousands (> 4,000) of gene sequences of interest in a single search. To search the database, after providing the intended gene of interest, keywords were used to describe the type of data to be searched for in NCBI (Figure 3.2.3).

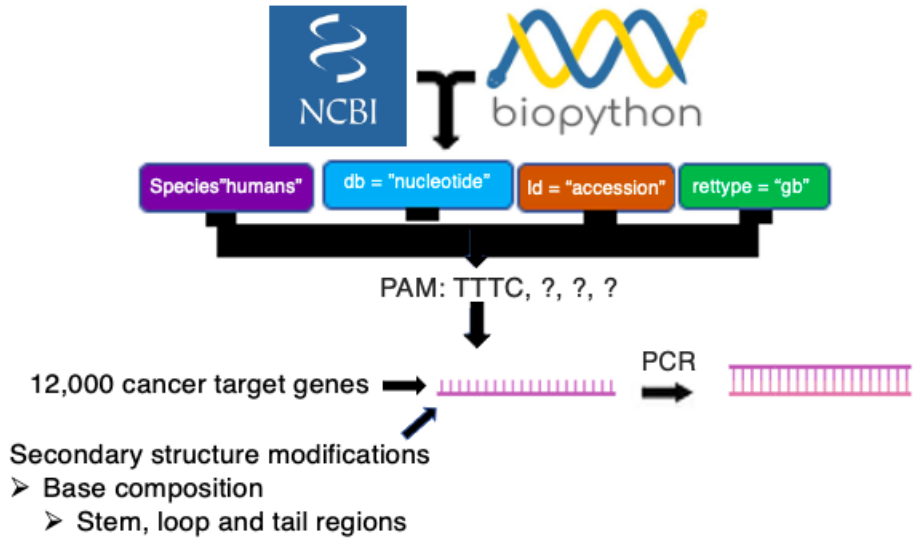
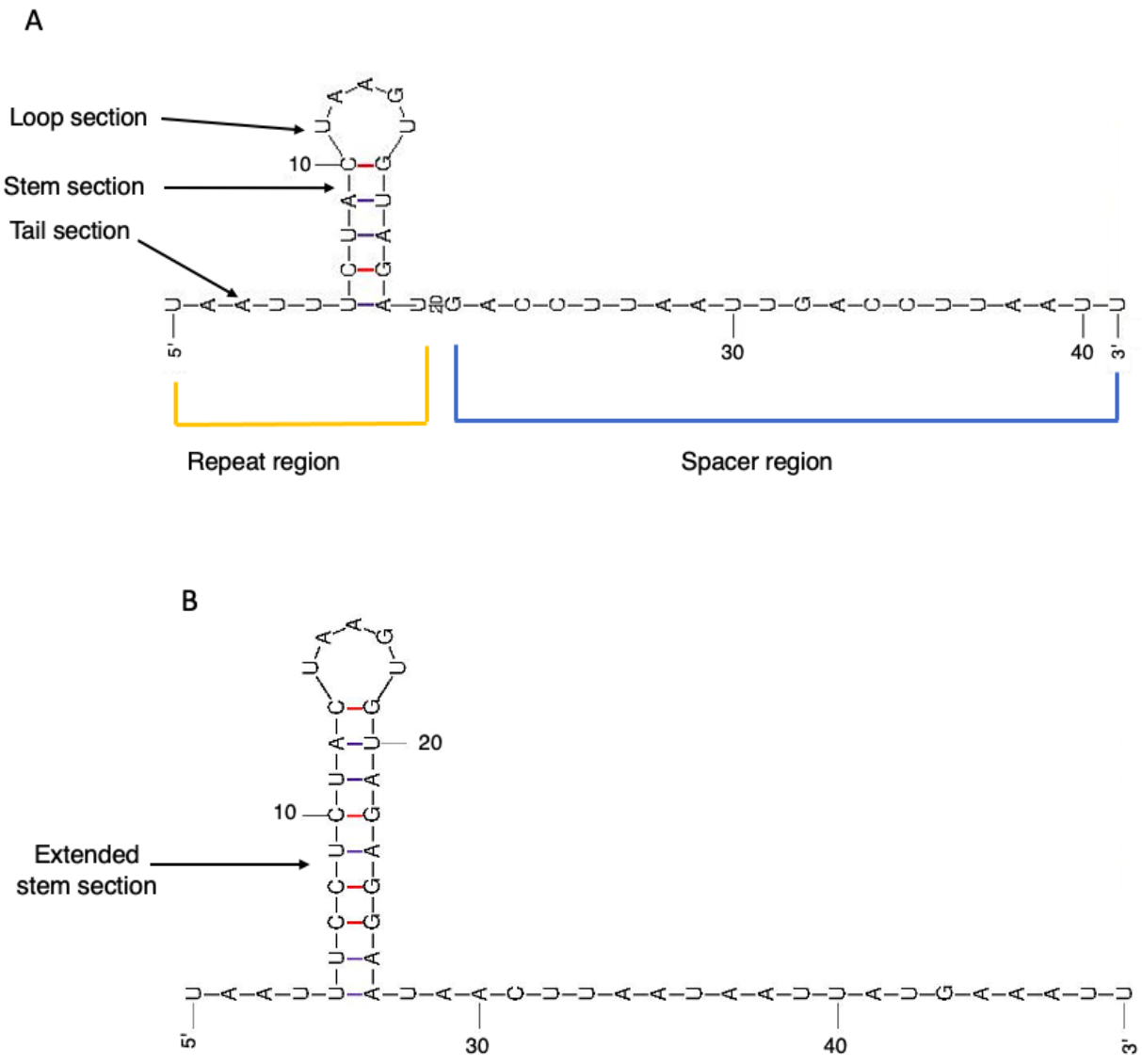


Figure 3.2.4. Pictural illustration of Library design using Biopython. Biopython is used to create a template library by iterating through NCBI database for 12,000 cancer genes. A Cas12a nuclease target site in each gene is selected; these are used to design the library members. Members also have different modifications (nucleotide compositions) within the gRNA secondary structure.

Upon the availability of the cancer target gene sequences, biopython is used to search the sequence for the first Cas12a target sequence in each gene using the PAM (TTTC) sequence as a search query. This result is a return of the specific target within each gene. At this point, the Cas12a targets are

ready for further modifications. The target sequences are DNA sequences, which are then transcribed into RNA sequences across all genes. A constant Cas12a secondary structure, also known as a repeat sequence, is added to the converted RNAs, i.e., the same repeat sequence is used for all RNAs. The combination of both converted RNAs and the constant repeat sequence results in a complete Cas12a gRNA sequence. The repeat sequence of each gRNA sequence is then modified with various secondary structure modifications. The secondary structure's stems, loops, and tail sections were modified (Figure 3.2.5).



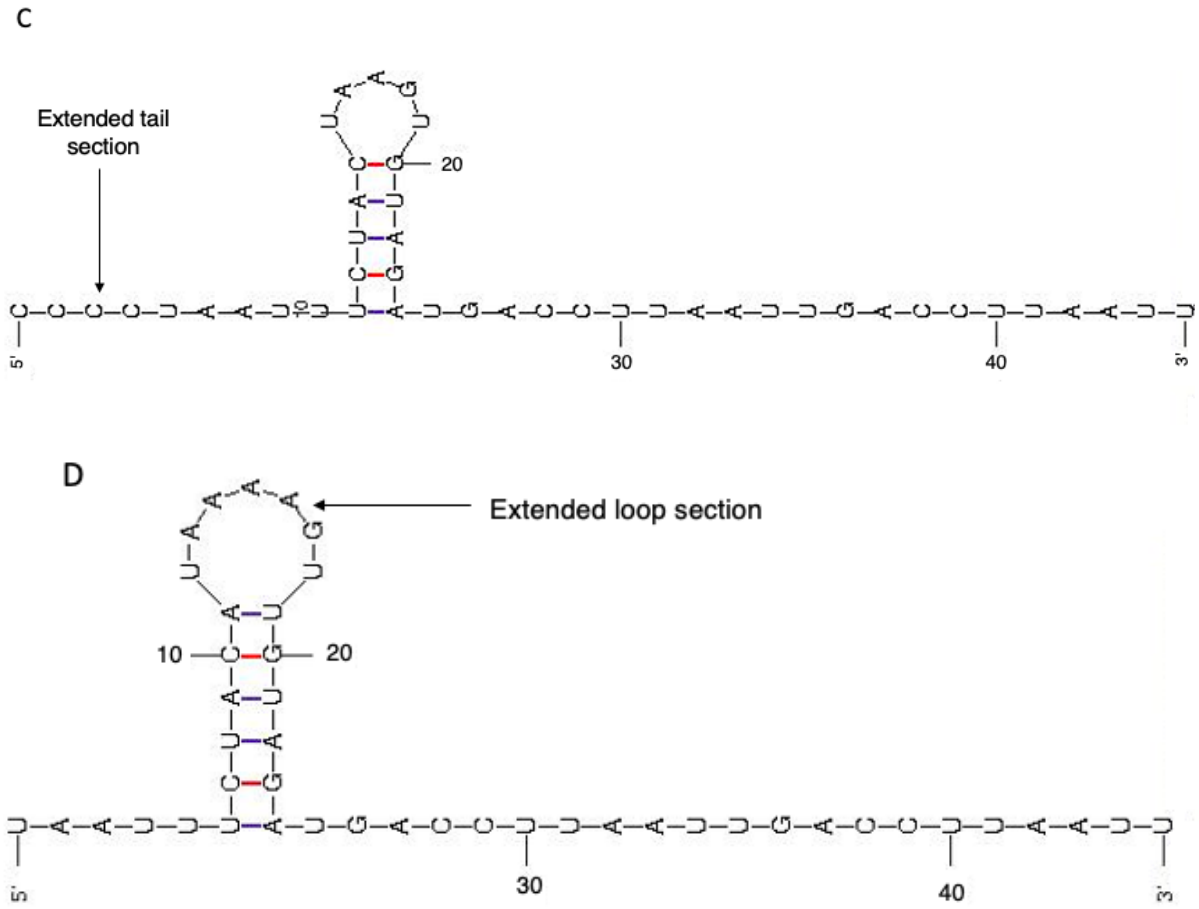


Figure 3.2.5. Image showing a Cas12 gRNA. (A) secondary structure with modifications on its stem section, (B) secondary structure with modifications on its stem section. (C) secondary structure with modifications on its tail section. (D) secondary structure with modifications on its loop section.

At this point, the DNA target sequences and modified gRNAs for each cancer gene are available. The ssDNA target DNA library is built using both gRNAs and DNA target sequences, with additional vital sequences added to fulfil specialized roles. Further components of each library member include forward and reverse primer sites for PCR amplification to produce additional copies of the library.

3.3. Library member design

It was necessary to create a template that is both efficient for gRNA synthesis and a possible target for the Cas12a nuclease. The arrangement of components within the template is critical to achieving the desired purpose. The initial design that was used included numerous components:

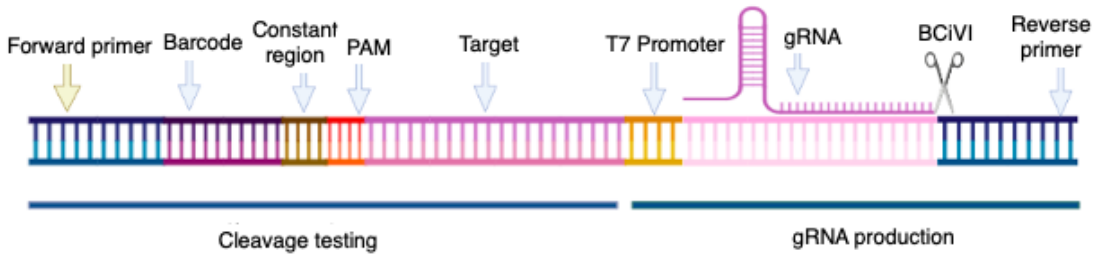


Figure 3.3.1. Each member the library performs two main functions: it produces gRNA using the T7 promoter site via transcription, and it also provides a target for Cas12a nuclease by utilizing the PAM recognition site. The template also has other important sites like the forward and reverse primer sites, barcode, and BciVI digestive enzyme recognition site.

The forward and reverse primer sites are used in PCR amplification to make additional copies of each member of the library. One of the primers is also required for amplification of clone insert detection through colony-PCR after cloning the library template into a plasmid. The barcode sequence will be deployed to characterize fragments following cleavage reactions and NGS. Along with the T7 promoter, the constant region provides a common sequence context for all targets and is shared by all members of the library. The PAM sequence is also present in the library member template, where it acts as a recognition site for target gene recognition. The Cas12a nuclease target sequence is next to the PAM sequence on the template. The T7 promoter sequence is also included in the template design; the T7 promoter is used to initiate gRNA transcription. The gRNA synthesis sequence follows the T7 promoter sequence (repeat and spacer sequences). The template also contains BciVI restriction enzyme recognition and digestion sites used for cutting off the reverse primer region. This allows for run-off transcription of the gRNA by T7 RNA polymerase. To achieve these, the template is cut using the BciVI site incorporated into the template, and the digestion of the template is done prior to the transcription reaction so that when transcription is ongoing, the T7 machinery falls off the template, automatically ending the transcription (Figure 3.3.1). This method is preferred compared to the use of a T7 terminator sequence or a ribozyme: In some cases, the T7 promoter continues with transcription after encountering the terminator sequence. Similarly, ribozymes can cut the template with unwanted extra bases hanging at the end of the cut site.

3.4. Library templates implementation

One library member (containing a gRNA and target for the ABCBI cancer gene) was picked to evaluate how well the design (Figure 3.3.1) would perform, including how successfully the forward

and reverse primers could be utilized to PCR amplify library members. This is necessary in order to create additional high-quality copies of the library, and the amplification process indicates whether the ssDNA library (as synthesized) can be properly turned into a dsDNA library. The ABCB1 library member template should be 149 base pairs long as synthesized.

The ABCBI ssDNA was amplified, and the lengths of the PCR reaction product were verified by nucleic acid native polyacrylamide electrophoresis to verify if they matched the theoretically expected lengths of the PCR-amplified ssDNA ABCB1 (Figure. 3.4.1). The PCR electrophoresis profiles showed that a product of the predicted length, and another product twice the size of the expected product, were generated during amplification. Consequently, the results indicated that the ssDNA ABCB1 template was amplified in the reaction, and the existence of another product twice the size of the product indicated that hybridization between two templates likely occurred during the amplification process. The results also showed that PCR amplification can successfully convert the entire ssDNA library to dsDNA, provided that the reaction is optimized to remove the undesired product (Figure. 3.4.1).

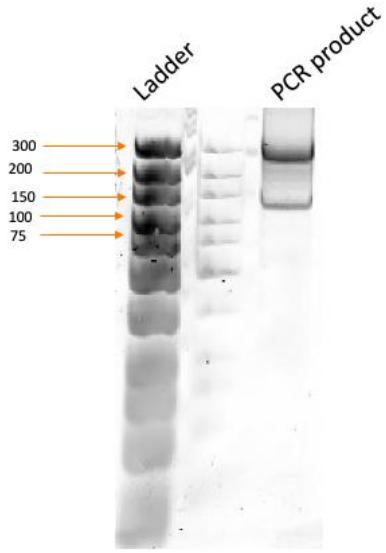


Figure 3.4.1. PCR amplification of ssDNA ABCB1. Sample order: Ladder – Ultra low range DNA ladder, PCR reaction product. 12% Native polyacrylamide, SYBR gold gel stain.

To optimize the results obtained via phusion polymerase, the exact PCR reaction was set up, but this time with different enzyme (DreamTaq polymerase) was used for the reaction, DreamTaq polymerase was used to determine if the unwanted product is a result of the template design or the enzyme. The

outcome was consistent with amplification by phusion polymerase, though the yield of ABCB1 dsDNA was lower when using DreamTag polymerase. This proved that the polymerase employed was a good choice, but it also revealed that the template design is not performing as anticipated due to the creation of an undesired product (Figure 3.4.2).

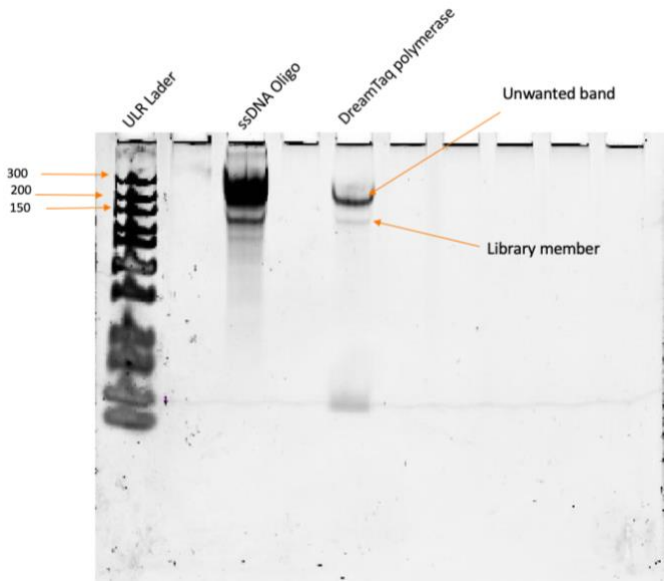


Figure 3.4.2. PCR amplification of ssDNA ABCB1 using DreamTaq polymerase enzyme. Sample order: Ladder – Ultra low range DNA ladder, ssDNA oligo, PCR reaction product. 12% Native polyacrylamide, SYBR gold gel stain.

Another PCR reaction was performed using the ABCB1 ssDNA template library member and its complementary strand. The complementary strand was adopted to generate a dsDNA library member and to assess the effect of the dsDNA template when making more copies of a dsDNA template library. According to the polyacrylamide gel analysis, the PCR amplification with dsDNA ABCB1 (provided with a complementary strand) produced the same unwanted product after amplification. This gel analysis again suggested that PCR amplification of a dsDNA library template does not produce a clean product when using the current template design, likely due to mis-annealing of the template DNAs with one another (Figure 3.4.3).

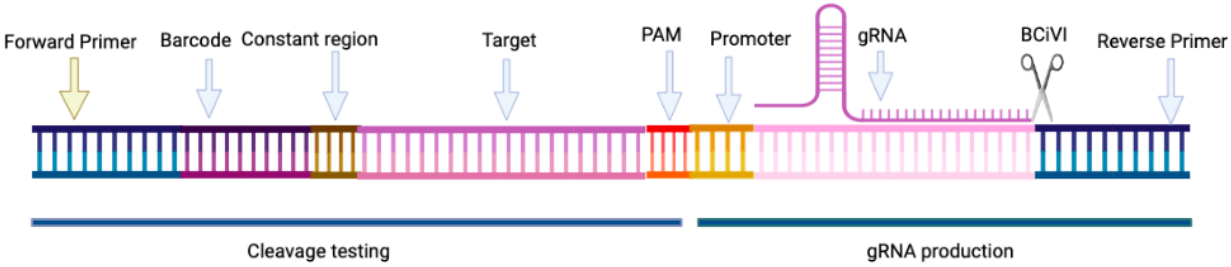


Figure 3.4.3. New design of library member: the PAM and target sequence is flipped towards the 3'-5' direction.

A new ssDNA template design with a flipped PAM and target sequence (Figure 3.5.3) in the 3'–5' direction is employed to evaluate whether it will produce a better amplification product, the old template is in figure 3.3.1. After PCR amplification, the electrophoresis profiles of the newly designed ssDNA template revealed that products of the expected lengths were produced without the unwanted product. As a consequence, the results indicated that more copies of the ssDNA ABCB1 library member were successfully created, confirming that a ssDNA library can be converted to dsDNA utilizing the newly constructed template. This is owing to the template's bases inability to bind with one another during the amplification process as a result of rearrangements of parts on the template (Figure 3.4.4).

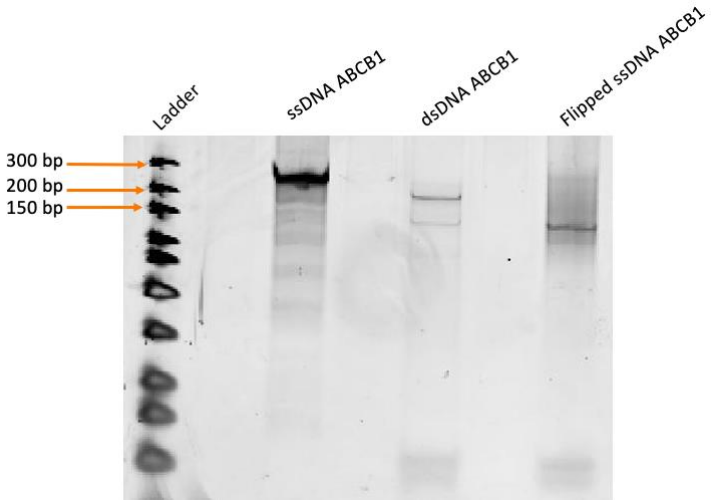


Figure 3.4.4. PCR amplification from ssDNA ABCB1, dsDNA ABCB1 and flipped ssDNA templates. Sample order: Ladder – Ultra low range DNA ladder, ssDNA ABCB1, dsDNA ABCB1 and flipped ssDNA. 12% Native polyacrylamide, SYBR gold gel stain.

3.5. Cloning into a plasmid vector

Previously amplified ABCB1 product gotten from a flipped template design was cloned into a recipient linear vector (pCR™-Blunt). The plasmid have a unique ccdB gene, which kills cells bearing nonligated (linear) plasmids following successful template-linear plasmid ligation and transformation. The ccdB gene guarantees that only cells harbouring circular plasmids survive, which can then be selected using appropriate antibiotics. Seven colonies were tested for successful insertion after transformation using a primer site on the template and another primer from the linear plasmid. Just two colonies from the colony PCR gel analysis exhibited a similarity with the predicted length (~300) after amplification (Figure 3.5.1).

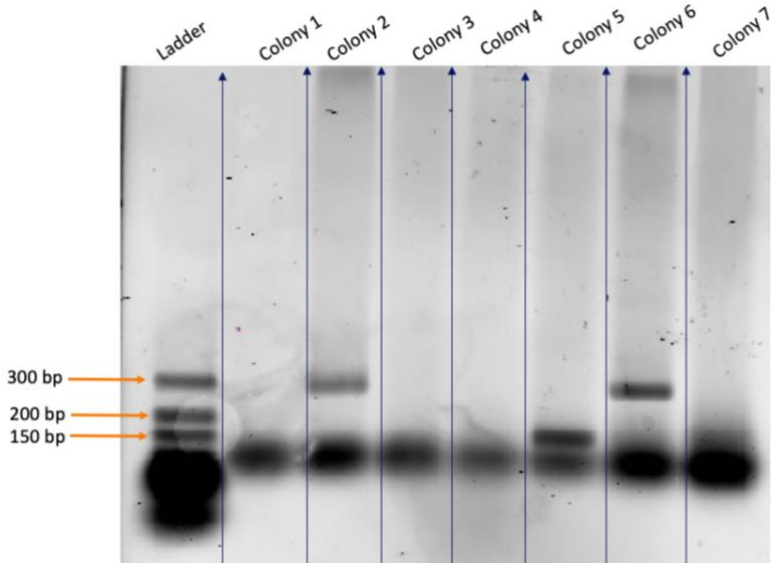


Figure 3.5.1. Colony PCR of cells transformed with the assembled vector containing the ABCB1 template. 2% Agarose gel electrophoresis, Ladder – Ultra low range DNA ladder, SYBR gold gel stain.

To validate the clones in the two colonies, full plasmid sequencing was performed. The sequencing result indicated that the circular plasmid indeed had the template cloned into it with all of the ABCB1 template components intact.

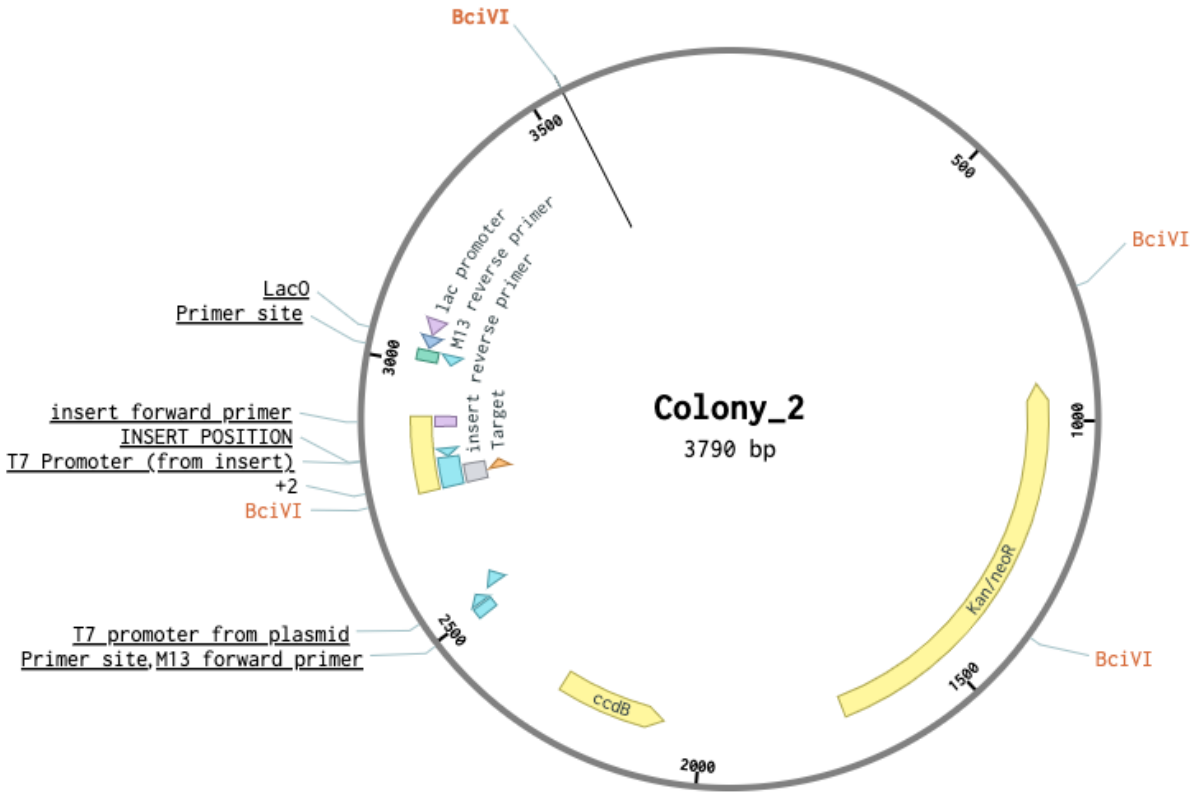


Figure 3.5.2. Map of sequenced full plasmid of colony PCR

Plasmids were isolated from these colonies and used for further experiments. I next needed to test if gRNA could be produced from this template. However, the template is currently ligated to the plasmid, the circular plasmid was digested with the BciVI restriction enzyme to. Based on the position of the BciVI enzyme cut site on the template, the enzyme will cut off the reverse primer sites on the plasmid fragment bearing the cloned template. This condition sets an ideal condition for potential runoff transcription. The expected sizes after digesting are 1.4 kilobase, 1 kilobase, 785 base pair and 600 base pair. The gel analysis revealed that the BciVI enzyme was able to cut the plasmid into multiple fragments and the expected sizes was recorded, one of which contained my gRNA template (Figure 3.5.3).

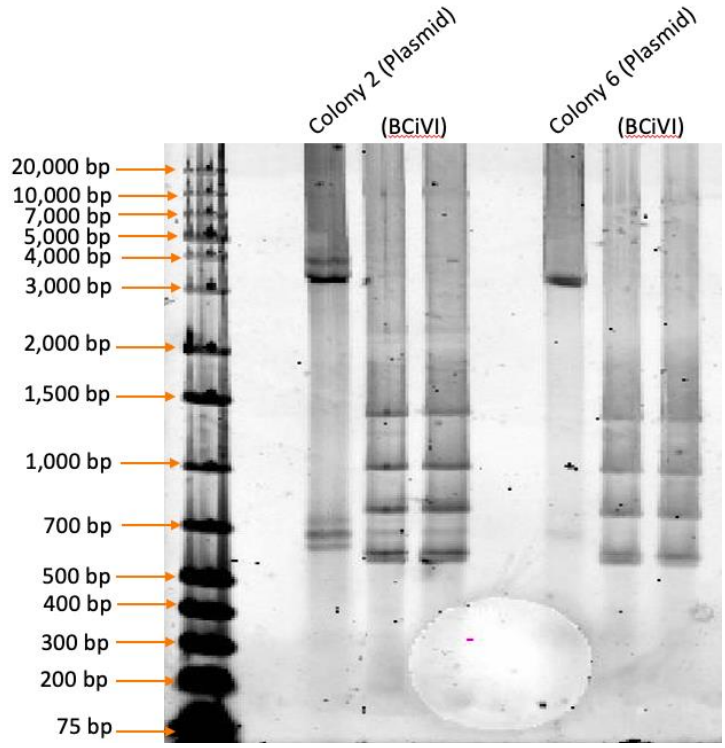


Figure 3.5.3. BciVI digested colony 2 and colony 6 plasmid. 1 % Agarose gel electrophoresis, Ladder – 1 kb plus ready to use Gene ruler, SYBR safe gel stain.

3.6. Transcription of gRNA from a plasmid template

Following the strategy for the experiment on been able to make gRNA from the template created, since the template has been ligated into a plasmid which the fragment has been cut into fragments and the template has been confirmed to be in the fragment. Transcription of the fragment needs to be done by carrying out transcription on the digested plasmid. The gel electrophoresis analysis obtained from the transcription reveals that the observed transcript length (>500 nucleotides) is considerably longer than the expected transcript length (40 nucleotides). This was attributed to the presence of multiple T7 promoter sites on the cloned plasmid; the linear plasmid has a T7 site incorporated into it, and the cloned template also has a T7 promoter site. This led to the production of gRNA with an undesired length (Figure 3.6.1).

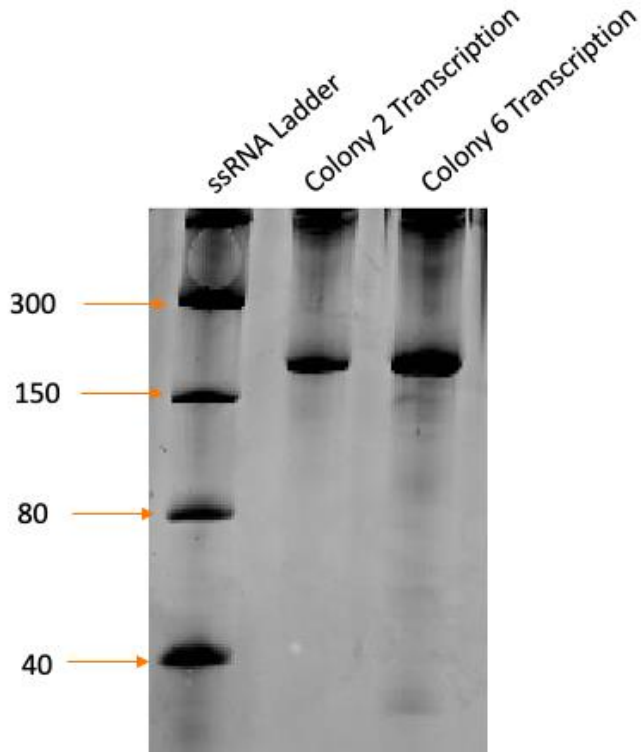


Figure 3.6.1. Colony 2 and colony 6 transcription. 15% Denaturing polyacrylamide gel, Ladder – ssRNA ultra low range ladder. SYBR gold gel stain.

4. DISCUSSION

Our results describe a strategy for creating a pooled library of paired targets and their corresponding gRNAs. We can next evaluate how library members with varying primary sequences and secondary structure modifications perform during transcription from dsDNA, RNP complex formation, and Cas12a nuclease DNA cleavage. The large library's size allowed the inclusion of a diverse set of human cancer genes and, at the same time, improves the quality of our results from NGS characterization. The 12,000 library members also allow a thorough examination of the impact of gRNA secondary structures. The adoption of Python for cancer target gene library development offers a great deal of flexibility and has demonstrated that it is a valuable biocomputing tool that can be utilized to search through other biologic organisms and nucleases databases beyond human and cancer genes in a similar rapid and cost-effective manner.

Functionality and performance of this study library template design can be affected by a variety of factors, with the most important being the hybridization between the target sequence on the template and the complementary gRNA spacer sequence produced during amplification. This was overcome by judiciously moving the target sequence further away in a different orientation from the gRNA spacer sequence; this reduced the affinity of both sequences to hybridise. The concept of a multifunctional template was also applied in the work done by Shen, M.W., et al. (2018). Where the template target sequence and gRNA sequence are separated by a linker sequence. Like the library member's main function in this study, gRNA can be produced off their template, and the template was cloned into a plasmid vector via ligation. This finding corroborates the strategy used for this study as feasible. Similarly, Ventura and Ventura (Vidigal and Ventura, 2015) were able to ligate their dsDNA library generated from a ssDNA via PCR amplification into a plasmid successfully for gRNA production.

Cloning results from this study can be better optimized, although successful clones were obtained and verified via sequencing. A more robust approach could be employed, especially in the choice of plasmid used during the ligation. The linear pCR Blunt End plasmid had unwanted components like the T7 promoter, which disrupted other downstream processes during the study. In this case, a customized plasmid could be constructed that will have all the machinery needed for successful ligation while at the same time eliminating unwanted components from the plasmid. Clues could also be adopted from the work done by Shen, M.W., and Vidigal, J. for better plasmid design.

After testing the viability of the template design and protocol finetuning, all reactions for the experiment would be carried out in a microfluidic droplet. In each droplet, we would have transformed cells containing plasmids ligated with each library template member (one template per cell per droplet). The droplet will also contain other regulators and components needed by the CRISPR machinery to function properly, which will be made off the plasmid. Some of the components that will be contained in the droplets will include RNA polymerase, light-activated Cas12 enzyme, cell lysis buffer, nucleaseq buffer containing mg^{2+}, and cleavage stop solution. The droplet will serve as an excellent platform for the experiment strategy that will assign a single gRNA produced during transcription to the Cas12a nuclease for RNP complex formation and to its target, which are all in the droplet. This way, we would be sure each member of the library had what was needed for the whole reaction. The droplet will also allow us to track the progress of the experiment at each step of transcription, RNP complex formation, and cleavage of target DNA. Following the cleavage reaction and disruption of the microfluidic droplets. All content of the droplet will be collected and cleaned up for analysis of fragments obtained during the overall reaction for fragment characterize using NGS.

From the NGS results, we would be able to draw a solid inference of how secondary structure modification design influences how gRNA can be transcribes from a DNA substrate, what edit combination better facilitates the formation of RNP complexes, and which sets of modifications are to be adopted for future experiment designs. Also, how does the spacer sequence of the gRNA in the RNP complex perform when it interacts with the complementary target sequence in relation to the edits of the gRNA repeat sequence? We would know if some modifications in the secondary structure promote RNP complex formation, further encourage the interaction between the gRNA, the target sequence and have any significance in target acquisition, and Whether some combination of bases on the gRNA influences the ability of the Cas12a nuclease to cut the dsDNA.

The findings of this study will be extremely useful to other academics working on Cas12a-related studies. They will be able to use our data to optimize the performance of their gRNA. Likewise, the study library and template design, could be easily adapted for different nuclease variants like the Cas13 family. Aside from making important discoveries on the best designs for gRNA during gene editing using Cas12a nuclease, our discovery can also be combined with other existing gene editing tools like base editing, prime editing for an improved result.

As gene editing has grown enormously in the past decade, and more interesting discoveries are expected in the future. The breakthroughs gained in gene editing technology for therapeutic and

diagnostic purposes will be supplemented by the findings of this study, which will look at how all genes in the genome could be edited using CRISPR. Our findings will provide new tools and approaches that can be applied to all CRISPR systems. Our discoveries will open new avenues for future gene editing research, paving the path for innovative therapeutics based on the technique while also encouraging its application in life sciences and development within the gene editing community.

CONCLUSION

In this study, numerous insights on the importance of the gRNA and how best it can be optimized for genome editing and human therapeutics by Cas12a nuclease have been highlighted. Cas12a, dubbed an alternative to Cas9, has not reached its full potential yet. (Paul and Montoya, 2020). However, these CRISPR systems share significant functions with Cas9. Cas9 and Cas12a are the only members of the CRISPR family that have been extensively used for genome editing thus far. Because of their major similarities and distinctions, these two endonucleases alone have made CRISPR applications extremely diverse (Swarts et al., 2018a). Furthermore, the nuclease guides' small size and the presence of an editable secondary structure on the guides make the Cas12a system vital to the genome engineering field. Efforts are being made to know more about the importance of Cas12a guide RNA, especially the vital roles played by its secondary structures (Strohkendl et al., 2018). This study has added to what we know about the Cas12a gRNA. A critical finding has been made on how bioinformatic tools can be adopted to enhance the way libraries templates can be designed. thereby opening new possibilities for the engineering of these secondary structures. Work done in this study also focuses on finding a better platform where Cas12a and other systems efficiency can be tested in genome editing research applications and facilitating existing applications. The strategy of incorporating the CRISPR machinery on a single plasmid and being able to demonstrate how CRISPR workflow components can be made off this single plasmid for investigating how gRNA modifications influence Cas12a nuclease functions to target cancer genes will prove very important for future experimental setups and potential improvements in the gene editing field.

To determine the effect of Cas12a gRNA secondary structure across of members of the constructed library, microfluidic droplet will be used as it may facilitate library-wide cleavage analysis across our library.

Vilnius University

Department of Biotechnology

Daniel Adegunle **Adelakin**

Determinant of CRISPR nuclease guide RNA function

Master thesis

SUMMARY

CRISPR (clustered regularly interspaced palindromic repeats) systems are sought-after genome editing tools for therapeutics and diagnostics. The use of CRISPR systems for gene editing is not without difficulties: the specificity of Cas12a nuclease, a Cas9 alternative, presents a limit in its use for gene editing due to its inability to successfully edit some genes. The primary cause of this problem is believed to be gRNA underperformance, caused specifically by gRNA secondary structures.

The gRNA plays a critical role in CRISPR systems. Experimental analysis has shown that the use of CRISPR for genome editing can result in off-targeting, causing a high degree non-specific editing at undesired genomic. Strategies used for gRNA evaluation have several drawbacks as well, such as a need for substantial protein engineering, incompatibility with viral packaging restrictions, and an increased number of system components. Published findings on the impact of the nucleotide composition and secondary structure of the gRNA contradict one another. Although, guides with slight modifications like changing the nucleotide composition and extension of the secondary structures can efficiently improve Cas12a nuclease efficiency for gene editing . Several bioinformatic tools help users design guide RNA with high targeting capability. Yet, the effectiveness of gRNAs in guiding Cas12a for genome editing varies substantially. Therefore, the field needs methods for designing highly effective gRNAs that can improve target identification and reduce nonspecific targeting. For this research, 12,000 gRNAs, spanning different nucleotide compositions, targets and secondary structures will be analyzed to show how these features influence transcription, ribonucleoprotein complex formation, and target cleavage. We will apply high-throughput next-generation sequencing to characterize gRNAs at each stage.

This study's logical design of gRNAs is a viable way to meet this demand. Using well- characterized oncogene target locations, we will show that rationally designed RNA secondary structures can boost the efficacy and specificity of gene editing tools like Cas 12a.

LIST OF SCIENTIFIC ACTIVITIES

Presentations at international conferences

Poster presentation at the international scientific conference The COINS 2023 (https://www.thecoins.eu/): Determinant of CRISPR nuclease guide RNA function Daniel A. Adelakin, Stephen K Jones Jr, Ieva sereiva. April 24-27, Vilnius, Lithuania: Summary Book, p. 106. https://thecoins.eu/static/resources/booksofabstracts/COINS2023.pdf.

LITERATURE

Barrangou, R., Doudna, J.A., 2016. Applications of CRISPR technologies in research and beyond. Nat. Biotechnol. 34, 933–941. https://doi.org/10.1038/nbt.3659

Behler, J., Hess, W.R., 2020. Approaches to study CRISPR RNA biogenesis and the key players involved. Methods San Diego Calif 172, 12–26. https://doi.org/10.1016/j.ymeth.2019.07.015

Bernd Zetsche, feng zhang, 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771.

Bornhorst, J.A., Falke, J.J., 2000. [16] Purification of Proteins Using Polyhistidine Affinity Tags. Methods Enzymol. 326, 245–254.

Brown, W., Zhou, W., Deiters, A., 2021. Regulating CRISPR/Cas9 Function through Conditional Guide RNA Control. Chembiochem Eur. J. Chem. Biol. 22, 63–72. https://doi.org/10.1002/cbic.202000423

Cofsky, J.C., Karandur, D., Huang, C.J., Witte, I.P., Kuriyan, J., Doudna, J.A., 2020. CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks (preprint). Biochemistry. https://doi.org/10.1101/2020.02.10.937540

Collias, D., Beisel, C.L., 2021. CRISPR technologies and the search for the PAM-free nuclease. Nat. Commun. 12, 555. https://doi.org/10.1038/s41467-020-20633-y

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339, 819–823. https://doi.org/10.1126/science.1231143

Cooper, R.M., Hasty, J., 2020. One-Day Construction of Multiplex Arrays to Harness Natural CRISPR-Cas Systems. ACS Synth. Biol. 9, 1129–1137. https://doi.org/10.1021/acssynbio.9b00489

Filippova, J., Matveeva, A., Zhuravlev, E., Stepanov, G., 2019. Guide RNA modification as a way to improve CRISPR/Cas9-based genome-editing systems. Biochimie 167, 49–60. https://doi.org/10.1016/j.biochi.2019.09.003

Gao, P., Yang, H., Rajashankar, K.R., Huang, Z., Patel, D.J., 2016. Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. Cell Res. 26, 901–913. https://doi.org/10.1038/cr.2016.88

Gielen, F., Hours, R., Emond, S., Fischlechner, M., Schell, U., Hollfelder, F., 2016. Ultrahigh-throughput–directed enzyme evolution by absorbance-activated droplet sorting (AADS). Proc. Natl. Acad. Sci. 113, E7383–E7389. https://doi.org/10.1073/pnas.1606927113

Habibian, M., McKinlay, C., Blake, T.R., Kietrys, A.M., Waymouth, R.M., Wender, P.A., Kool, E.T., n.d. Reversible RNA acylation for control of CRISPR–Cas9 gene editing. Chem. Sci. 11, 1011–1016. https://doi.org/10.1039/c9sc03639c

He, Y., Yan, W., Long, L., Dong, L., Ma, Y., Li, C., Xie, Y., Liu, N., Xing, Z., Xia, W., Li, F., 2023a. The CRISPR/Cas System: A Customizable Toolbox for Molecular Detection. Genes 14, 850. https://doi.org/10.3390/genes14040850

He, Y., Yan, W., Long, L., Dong, L., Ma, Y., Li, C., Xie, Y., Liu, N., Xing, Z., Xia, W., Li, F., 2023b. The CRISPR/Cas System: A Customizable Toolbox for Molecular Detection. Genes 14, 850. https://doi.org/10.3390/genes14040850

Hemphill, J., Borchardt, E.K., Brown, K., Asokan, A., Deiters, A., 2015. Optical Control of CRISPR/Cas9 Gene Editing. J. Am. Chem. Soc. 137, 5642–5645. https://doi.org/10.1021/ja512664v

Hille, F., Charpentier, E., 2016. CRISPR-Cas: biology, mechanisms and relevance. Philos. Trans. R. Soc. B Biol. Sci. 371. https://doi.org/10.1098/rstb.2015.0496

Hille, F., Richter, H., Wong, S.P., Bratovič, M., Ressel, S., Charpentier, E., 2018. The Biology of CRISPR-Cas: Backward and Forward. Cell 172, 1239–1259. https://doi.org/10.1016/j.cell.2017.11.032

Hoikkala, V., Ravantti, J., Díez-Villaseñor, C., Tiirola, M., Conrad, R.A., McBride, M.J., Moineau, S., Sundberg, L.-R., 2021. Cooperation between Different CRISPR-Cas Types Enables Adaptation in an RNA-Targeting System. mBio 12, e03338-20. https://doi.org/10.1128/mBio.03338-20

Holstein, J.M., Gylstorff, C., Hollfelder, F., 2021. Cell-free Directed Evolution of a Protease in Microdroplets at Ultrahigh Throughput. ACS Synth. Biol. 10, 252–257. https://doi.org/10.1021/acssynbio.0c00538

Huang, H., Huang, G., Tan, Z., Hu, Y., Shan, L., Zhou, J., Zhang, X., Ma, S., Lv, W., Huang, T., Liu, Y., Wang, D., Zhao, X., Lin, Y., Rong, Z., 2022. Engineered Cas12a-Plus nuclease enables gene editing with enhanced activity and specificity. BMC Biol. 20, 1–16. https://doi.org/10.1186/s12915-022-01296-1

Jiang, F., Doudna, J.A., 2015. The structural biology of CRISPR-Cas systems. Curr. Opin. Struct. Biol., Folding and binding/Nucleic acids and their protein complexes 30, 100–111. https://doi.org/10.1016/j.sbi.2015.02.002

Jones, S.K., Hawkins, J.A., Johnson, N.V., Jung, C., Hu, K., Rybarski, J.R., Chen, J.S., Doudna, J.A., Press, W.H., Finkelstein, I.J., 2021. Massively parallel kinetic profiling of natural and engineered CRISPR nucleases. Nat. Biotechnol. 39, 84–93. https://doi.org/10.1038/s41587-020-0646-5

Kaminski, M.M., Abudayyeh, O.O., Gootenberg, J.S., Zhang, F., Collins, J.J., 2021. CRISPR-based diagnostics. Nat. Biomed. Eng. 5, 643–656. https://doi.org/10.1038/s41551-021-00760-7

Kocak, D.D., Josephs, E.A., Bhandarkar, V., Adkar, S.S., Kwon, J.B., Gersbach, C.A., 2019. Increasing the specificity of CRISPR systems with engineered RNA secondary structures. Nat. Biotechnol. 37, 657–666. https://doi.org/10.1038/s41587-019-0095-1

Konstantakos, V., Nentidis, A., Krithara, A., Paliouras, G., 2022. CRISPR–Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning. Nucleic Acids Res. 50, 3616–3637. https://doi.org/10.1093/nar/gkac192

Krysler, A.R., Cromwell, C.R., Tu, T., Jovel, J., Hubbard, B.P., 2022. Guide RNAs containing universal bases enable Cas9/Cas12a recognition of polymorphic sequences. Nat. Commun. 13, 1617. https://doi.org/10.1038/s41467-022-29202-x

Lelong, C.W., Cyrus Afrasiabi, Sebastien, n.d. mygene: Python Client for MyGene.Info services.

Li, B., Zhao, W., Luo, X., Zhang, X., Li, C., Zeng, C., Dong, Y., 2017. Engineering CRISPR-Cpf1 crRNAs and mRNAs to maximize genome editing efficiency. Nat. Biomed. Eng. 1, 0066. https://doi.org/10.1038/s41551-017-0066

Li, Y., Li, S., Wang, J., Liu, G., 2019. CRISPR/Cas Systems towards Next-Generation Biosensing. Trends Biotechnol. 37, 730–743. https://doi.org/10.1016/j.tibtech.2018.12.005

Liao, C., Slotkowski, R.A., Achmedov, T., Beisel, C.L., 2018. The Francisella novicida Cas12a is sensitive to the structure downstream of the terminal repeat in CRISPR arrays. RNA Biol. 16, 404–412. https://doi.org/10.1080/15476286.2018.1526537

Luo, J., Chen, W., Xue, L., Tang, B., 2019. Prediction of activity and specificity of CRISPR-Cpf1 using convolutional deep learning neural networks. BMC Bioinformatics 20, 332. https://doi.org/10.1186/s12859-019-2939-6

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Scott, D., Shah, S.A., Siksnys, V., Terns, M.P., Venclovas, Č., White, M.F., Yakunin, A.F., Yan, W.,

Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J., Barrangou, R., Koonin, E.V., 2020. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. 18, 67–83. https://doi.org/10.1038/s41579-019-0299-x

Marx, V., 2020. Guide RNAs: it's good to be choosy. Nat. Methods 17, 1179–1182. https://doi.org/10.1038/s41592-020-01003-4

Mcginn, J., Marraffini, L., 2018. Molecular mechanisms of CRISPR–Cas spacer acquisition. Nat. Rev. Microbiol. 17, 1. https://doi.org/10.1038/s41579-018-0071-7

Obexer, R., Pott, M., Zeymer, C., Griffiths, A.D., Hilvert, D., 2016. Efficient laboratory evolution of computationally designed enzymes with low starting activities using fluorescence-activated droplet sorting. Protein Eng. Des. Sel. 29, 355–366. https://doi.org/10.1093/protein/gzw032

Riesenberg, S., Helmbrecht, N., Kanis, P., Maricic, T., Pääbo, S., 2022. Improved gRNA secondary structures allow editing of target sites resistant to CRISPR-Cas9 cleavage. Nat. Commun. 13, 489. https://doi.org/10.1038/s41467-022-28137-7

Schubert, M.S., Thommandru, B., Woodley, J., Turk, R., Yan, S., Kurgan, G., McNeill, M.S., Rettig, G.R., 2021. Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. Sci. Rep. 11, 19482. https://doi.org/10.1038/s41598-021-98965-y

Scott, T., Urak, R., Soemardy, C., Morris, K.V., 2019. Improved Cas9 activity by specific modifications of the tracrRNA. Sci. Rep. 9, 16104. https://doi.org/10.1038/s41598-019-52616-5

Singh, D., Mallon, J., Poddar, A., Wang, Y., Tippana, R., Yang, O., Bailey, S., Ha, T., 2018. Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). Proc. Natl. Acad. Sci. 115, 5444–5449. https://doi.org/10.1073/pnas.1718686115

Song, H., Chen, D.L., Ismagilov, R.F., 2006. Reactions in Droplets in Microfluidic Channels. Angew. Chem. Int. Ed Engl. 45, 7336–7356. https://doi.org/10.1002/anie.200601554

Tang, Y., Gao, L., Feng, W., Guo, C., Yang, Q., Li, F., Le, X.C., 2021. The CRISPR-Cas toolbox for analytical and diagnostic assay development. Chem. Soc. Rev. 50, 11844–11869. https://doi.org/10.1039/d1cs00098e

Vidigal, J.A., Ventura, A., 2015. Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. Nat. Commun. 6, 8083. https://doi.org/10.1038/ncomms9083

Wang, D., Zhang, C., Wang, B., Li, B., Wang, Q., Liu, D., Wang, H., Zhou, Y., Shi, L., Lan, F., Wang, Y., 2019. Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. Nat. Commun. 10, 1–14. https://doi.org/10.1038/s41467-019-12281-8

Wong, N., Liu, W., Wang, X., 2015. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system.

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., Fedorova, I., Nakane, T., Makarova, K.S., Koonin, E.V., Ishitani, R., Zhang, F., Nureki, O., 2016. Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell 165, 949–962. https://doi.org/10.1016/j.cell.2016.04.003

Zhou, W., Brown, W., Bardhan, A., Delaney, M., Ilk, A.S., Rauen, R.R., Kahn, S.I., Tsang, M., Deiters, A., 2020. Spatiotemporal Control of CRISPR/Cas9 Function in Cells and Zebrafish using Light-Activated Guide RNA. Angew. Chem. Int. Ed Engl. 59, 8998–9003. https://doi.org/10.1002/anie.201914575

ACKNOWLEDGEMENT

I would like to thank Dr. Stephen Knox Jones Jr., my esteemed supervisor, for allowing me to join his laboratory and for his invaluable supervision, support, and guidance during my Master's programme. In addition, I want to thank Dr. Mindaugas Zeremba for his amazing support, which was important in shaping my research. Likewise, I would like to thank Ieva Sereiva for her generous help in building my target DNA library. My heartfelt thanks go to the Jones Laboratory team for an unforgettable and treasured time spent together in the lab and in social settings.

My thanks also go to my family and friends for their encouragement and support during my master's studies.