VILNIUS UNIVERSITY
LIFE SCIENCES CENTER

Urtė Glibauskaitė

Directed Evolution Studies of a Methylation-Sensitive Cas9 for Human Genome Editing

Master's Thesis

Biochemistry study program

Thesis supervisor

Dr. Stephen Knox Jones Jr.

Consultant

Ieva Jaskovikaitė

Thesis completed at

VU LSC-EMBL Partnership Institute for Genome Editing Technologies

Vilnius 2025

Contents

Abbreviations

4mC – 4-methylcytosine

5mC – 5-methylcytosine

6mA – 6-methyladenine

Ara – arabinose

bp – base pair

Cas – CRISPR-associated

ccdB – control of cell death protein B

CFU – colony-forming unit

CRISPR – clustered regularly interspaced short palindromic repeats

crRNA – CRISPR ribonucleic acid

dCas9 – (catalytically) dead Cas9

ePCR – error-prone polymerase chain reaction

ESC – embryonic stem cell

gRNA – guide ribonucleic acid

HDR – homology-directed repair

HRP – horseradish peroxidase

kb – kilobase

kDa – kilodalton

MTase – methyltransferase

N – A or G or C or T nucleotide

NHEJ – non-homologous end joining

NLS – nuclear localization signal

PAM – protospacer adjacent motif

PID – PAM-interacting domain

pre-crRNA – precursor CRISPR ribonucleic acid

QCM – quick-change mutagenesis

REase – restriction endonuclease

RM – restriction-modification (system)

RNAse III – ribonuclease III

SAM – S-adenosyl-L-methionine

sgRNA – single-guide ribonucleic acid

SR – survival rate

tracrRNA – trans-activating CRISPR ribonucleic acid

WT – wild-type

Introduction

The CRISPR (clustered regularly interspaced short palindromic repeats) locus and Cas (CRISPR-associated) proteins originate in bacteria and archaea, providing immunity against bacteriophages and other mobile genetic elements (Barrangou et al., 2007). The CRISPR-Cas system incorporates short pieces of foreign DNA into the CRISPR locus. Then, this sequence is transcribed and processed to produce a guide RNA (gRNA), which forms an effector complex with a Cas nuclease. The target sequence must be flanked by a short PAM (protospacer adjacent motif) to be recognized by the nuclease (Jinek et al., 2012). Following PAM recognition, the gRNA hybridizes with the matching target sequence, which is then cleaved. The target-complementary part of the gRNA can be engineered to base pair with nearly any target sequence of interest, enabling precise genome editing and diagnostic applications (Hirakawa et al., 2020).

CRISPR-Cas enzymes underly tools to create, detect, and report changes to the genome. Such tools include CRISPR-based epigenome editing platforms that are generally comprised of catalytically dead Cas9 (dCas9) fused with an epigenetic effector domain (McCutcheon et al., 2024). Depending on the type of the epigenetic effector used, these platforms can modify epigenetic marks (e.g., DNA methylation, histone modifications), leading to activation or repression of the targeted genes. However, such platforms cannot: be recruited according to the local epigenetic state of the targeted region, or detect epigenetic changes in a programmable manner. For example, the broadly used CRISPR-Cas9 from *Streptococcus pyogenes* (SpCas9) does not distinguish DNA methylation in the PAM or target sequences (Fujita et al., 2016). However, a CRISPR-Cas9 from *Acidothermus cellulolyticus* (AceCas9) is the only known methylation sensitive Cas9, and it holds potential for use according to the epigenetic state of the targeted region (Das et al., 2020). The activity of this nuclease is affected by the methylation of its PAM sequence 5′-NNNCC-3′. A methyl group on the first, but not second, cytosine of the PAM sequence breaks amino acid contacts in the PAM-interacting domain (PID) of AceCas9 and leads to a loss of cleavage activity. Although AceCas9 can operate in *Escherichia coli* (*E. coli*) (Tsui et al., 2017), its activity in human cells remains unknown. Therefore, this work focused on determining AceCas9 expression and genome editing in human cells. Additionally, as the 5mCpC sensitivity of AceCas9 does not align with the 5mCpG modification prevalent in human cells, we employed a directed evolution strategy to engineer AceCas9 that could detect 5′-NNNCG-3′ PAM sequence.

To assess whether AceCas9 is functional in human cells and make it more suitable for human epigenetic editing approaches, this work had the following objective and tasks:

Objective: Determine AceCas9 expression and genome editing efficiency in HEK293T cells and engineer its PAM specificity through directed evolution for improved compatibility with human genomic targets.

Tasks:

1. Transfect HEK293T cells with a AceCas9-TwinStrep vector and determine if AceCas9 can be expressed in human cells.
2. Perform nuclear fractionation of HEK293T cells to determine if AceCas9 can enter the nucleus.
3. Determine AceCas9 editing efficiency in HEK293T cells for PCSK9 and UBE3A genes.
4. Construct an AceCas9 plasmid library with mutagenized PAM-interacting domains.
5. Perform directed evolution of AceCas9 to select for 5'-NNNCG-3' PAM-recognizing variants.

1. Literature Review

1.1. Role of DNA methylation in prokaryotic and eukaryotic systems

1.1.1. Prokaryotes

In prokaryotes, DNA methylation plays an important role in various cellular processes, such as cell cycle control, gene expression, response to the environment and defense systems against invaders, (e.g., bacteriophages and other foreign genetic elements). Prokaryotic methyltransferases (MTases) transfer methyl groups from cofactor S-adenosyl-1-methionine (SAM) onto adenine and cytosine bases (Cheng, 1995). Known prokaryotic DNA modifications include N-4 and C-5 cytosine methylation (4mC, 5mC) as well as the most predominant N-6 adenine methylation (6mA) (**Figure 1.1**) (Beaulaurier et al., 2019; Sánchez-Romero & Casadesús, 2020). DNA methylation is a reversible process which does not change genetic sequence but rather modulates protein binding and regulates epigenetic functions (Casadesús & Low, 2006). Methyl groups can either physically hinder protein binding or, on the other hand, attract proteins that have specifically evolved to bind methylated DNA (Casadesús & Low, 2006).

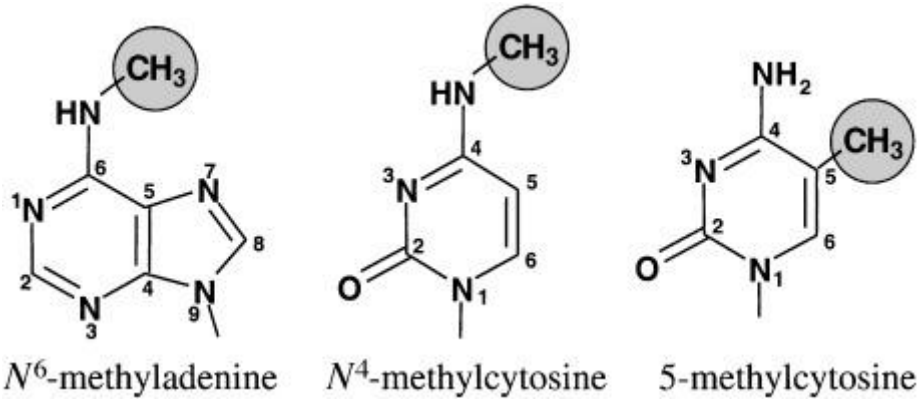

Figure 1.1 Structures of methylated DNA bases found in prokaryotes. Adapted from (Jeltsch, 2002)

Prokaryotic DNA methylation is often associated with restriction-modification (RM) defense systems that protect the host against invasion of viral and other foreign genetic material (Arber, 1974). Generally, RM systems are comprised of two components - a restriction endonuclease (REase) and its cognate MTase. REases recognize short (4-8bp) nucleotide sequences and act as molecular scissors by introducing double-stranded DNA breaks. To avoid destruction of the host's genome, cognate MTases recognize the same sequences of the endogenous DNA and methylate them, preventing recognition and cleavage by intrinsic REases. (Bickle et al., 1978) Therefore, this two-component RM system determines whether the DNA comes from an endogenous or exogenous source and destroys the latter. Depending on sequence recognition, cleavage site, subunit structure and cofactor requirements, RM systems are classified into four main types – type I, II, III and IV. (Roberts et al.,

2003). Type I, II and III RM systems contain both components of restriction and modification (as subunits or separate proteins). However, type IV RM systems contain restriction enzymes only and can target foreign methylated DNA (Roberts et al., 2003).

DNA methylation also plays a role in controlling prokaryotic DNA replication and DNA repair. These mechanisms are best studied in *Escherichia coli* (*E. coli*) DNA-methylating enzyme deoxyadenosine methyltransferase (Dam) system, which methylates the adenine bases in 5'-GATC-3' sequences (Marinus & Morris, 1973). *E. coli*, as well as other γ-proteobacteria, utilize Dam hemimethylation in DNA mismatch repair by discriminating between error-free template strand and error-containing daughter strand (A. L. Lu et al., 1983). This repair happens in a span of 2-4s after DNA replication, when the newly synthesized DNA strand lacks methylation on the Dam sites. The mismatches are recognized by a mismatch repair protein MutS, assisted by MutL and MutH. Once the MutS-MutL-MutH complex assembles at a DNA mismatch, MutH gains endonuclease activity and cuts the phosphodiester bond upstream of the guanine in the nearest 5'-GATC-3' sequence on the unmethylated DNA strand (Hall et al., 1998). Afterwards, UvrD helicase removes MutH from the ternary complex, allowing bacterial exonucleases to degrade the single-stranded DNA. The gap is then filled by DNA polymerase III and the nick is repaired by DNA ligase. Finally, Dam methyltransferase methylates the 5'-GATC-3' site in the repaired strand, producing fully methylated DNA (Modrich, 2016).

The ever-changing environment of prokaryotes requires adaptation, typically involving a reversible switch between different phenotypes. This process is called phase variation and often affects structures such as flagella, pilli or outer membrane proteins (Bayliss et al., 2025; Edwards & Bruner, 1939). Phase variation is well-studied in bacteria, and it is controlled at a genetic level through DNA inversions, slipped-strand mispairing and epigenetic changes (Bayliss et al., 2025; Beaulaurier et al., 2019). DNA methylation is important for the epigenetic control of phase variation, which alters the phenotype without altering the DNA sequence. For example, phase variation controls expression of Agn43 gene, encoding a cell surface protein Antigen 43 (Ag43) in *E. coli* (Henderson et al., 1997). Ag43 is important in mediating cell-cell interactions, promoting biofilm maturation, and contributing to immune evasion through phase-variable expression (van der Woude & Henderson, 2008). The expression of Ag43 is Dam-dependent as its transcription is controlled through methylation of 5'-GATC-3' sequences in the promoter region (Henderson et al., 1997). An oxidative stress regulator (OxyR) binds unmethylated promoter sequence and inhibits transcription of Ag43. If the promoter is methylated, OxyR is unable to bind the promoter and Ag43 is expressed (Henderson et al., 1997). This process works as an epigenetically regulated ON/OFF switch, allowing for population diversity through phase variation.

Unlike Dam, which is constitutively expressed, expression of cell cycle–regulated methyltransferase (CcrM) is tightly regulated. CcrM was first discovered in *Caulobacter crescentus* and is conserved in α-proteobacteria (Reisenauer et al., 1999; Stephens et al., 1996). *C. crescentus* undergoes asymmetric division, creating two disctinct daughter cells - a stalked cell and a swarmer cell (Collier et al., 2007). The asymmetric division is regulated through CcrM-mediated methylation of the origin of replication (*Cori*) (Marczynski, 1999). During the stalked cell phase, CcrM is active and methylates the adenine residues in 5'-GANTC-3' sequences, producing fully methylated *Cori.* Full methylation acts as a signal for binding of DnaA protein, involved in initiation of DNA unwinding and replication (Collier et al., 2007). During the swarmer cell phase, CcrM activity is suppressed, and the DNA remains hemi-methylated, preventing premature replication. To become competent for DNA replication, a swarmer cell must differentiate into a stalked cell, which triggers CcrM activity to fully methylate the DNA, allowing it to replicate (Collier et al., 2007). Overall, CcrM coordinates the timing of DNA replication, ensuring that it occurs in the appropriate phase of the cell cycle and that asymmetric division leads to distinct daughter cells.

The variety of DNA modifications (6mA, 4mC, 5mC) can be introduced by aforementioned RM systems, however, some MTases are thought to have "escaped" from that context, becoming "orphan" MTases. Dam and CcrM are examples of "orphan" MTases that primarily introduce adenine methylation (6mA), functioning mainly as a regulatory modification rather than a defense mechanism. Regardless of the prevalence of 6mA DNA modification, cytosine-methylating "orphan" MTases have been reported in some bacterial and archaeal species. For example, 5mC DNA modification can be introduced by C^5-methylcytosine methyltransferase, also known as Dcm (Militello et al., 2012). Dcm methylates the second cytosine in the sequence 5'-CCWGG-3' (where W = A or T). Existence of 5mC modification is known as a paradox since it creates a risk of mutation due to its tendency to undergo deamination to thymine, forming T:G mismatches during DNA replication (Cherry, 2018). Despite the risks associated with 5mC deamination, it is thought to provide benefits in prokaryotic genomes by contributing to the regulation of gene expression, stress response, and other physiological processes (Kahramanoglou et al., 2012).

1.1.2. Eukaryotes

In contrast to prokaryotes, 5mC DNA modification is known as the most prevalent epigenetic modification in mammalian systems. This modification is established by eukaryotic methyltransferases (MTases) by acting either on nonmethylated (*de novo* methylation) or hemi-methylated substrates (maintenance methylation) (Jeltsch, 2002). *De novo* methylation, which introduces new methylation marks on previously unmethylated regions of the genome, is carried out

by DNMT3A and DNMT3B MTases. In contrast, maintenance methylation, which preserves established DNA methylation patterns following DNA replication, is mediated by DNMT1 (Goll & Bestor, 2005). Eukaryotic DNA methylation plays essential roles in various fundamental biological processes. During early embryonic development, it is involved in the reprogramming of the epigenome to establish cell lineage-specific gene expression patterns (Seisenberger et al., 2013). It is also crucial for X-chromosome inactivation, where one of the two X chromosomes in females is silenced to balance gene dosage (Zamudio et al., 2011). Additionally, DNA methylation regulates genomic imprinting and contributes to proper tissue-specific gene expression throughout development and differentiation (Seisenberger et al., 2013). It also plays a critical role in maintaining genomic stability by silencing transposable elements and repetitive DNA sequences, preventing their mobilization and potential disruption of gene function (Bourc'his & Bestor, 2004).

In mammalian genome context, the 5mC modification is commonly found in CpG dinucleotide clusters called CpG islands (CGIs) (Bird et al., 1985). The "p" between adjacent nucleotides stands for phosphodiester bond joining them together. CGIs are longer than 200bp and occur in around 60-70% of promoter sequences of human genes. (Saxonov et al., 2006). Generally, methylation in promoter-associated CGIs results in gene silencing. Promoter methylation can block binding of transcription factors or recruit repressive proteins such as methyl-CpG-binding domain (MBD) proteins (Cross et al., 1997; Hendrich & Bird, 1998). For example, MBD1 and MBD2 specifically bind to methylated CpG sites and attract enzymes involved in formation of heterochromatin, reducing gene expression levels (C. Zhao et al., 2022; Zhou et al., 2017). CGIs at active promoters are typically unmethylated and stable across tissues. However, CpG island shores (regions up to ~2kb from CGIs) exhibit tissue-specific methylation patterns that correlate with gene expression changes (Irizarry et al., 2009). Genome-wide methylation studies demonstrated that most of the tissue-specific differentially methylated regions (tDMRs) are located in these shores rather than within CGIs (Rakyan et al., 2008; Sugimoto et al., 2009). While promoter CGIs provide a stable, unmethylated platform for gene activation, CpG island shores serve as flexible regulatory elements, modulating expression patterns, necessary for maintaining distinct cellular phenotypes.

Although rarer, CGIs can also be found in gene bodies. Interestingly, methylation of gene body CGIs has the opposite effect than methylation of the promoter sequences and is associated with active transcription (Gutierrez-Arcelus et al., 2013; Lister et al., 2009). This phenomenon is called the "DNA methylation paradox" which gives more insights into a complex relationship between DNA methylation and transcription. Studies suggest that the role of gene body methylation is important in regulation of alternative gene splicing and propose kinetic and recruitment models to explain this (Naftelberg et al., 2015). The kinetic model suggests that DNA methylation in the gene body affects the rate of transcriptional elongation. For example, a faster RNA polymerase II (RNAp II) skips

weaker splice sites and exons while a slower RNAp II includes the weaker splice sites and exons. (Shukla et al., 2011) Unmethylated DNA can be bound by transcriptional repressors, such as CCCTC-binding factor (CTCF), which reduces the elongation rate of RNAp II and facilitates exon inclusion (Shukla et al., 2011). In contrast, CTCF is unable to bind methylated DNA thus, the elongation rate of RNAp II increases and exons are skipped. In the recruitment model, the splicing is controlled by introducing splicing factors such as polypyrimidine tract-binding protein (PTB) and serine/arginine splicing factor 1 (SRSF1) (Pradeepa et al., 2012). Both PTB and SRSF1 can be recruited by a histone H3K36me3, often associated with heterochromatin. Additionally, DNA methylation can induce H3K36me3 binding to heterochromatin protein 1 (HP1) which recruits the serine/arginine splicing factor 3 (SRSF3) to regulate alternative splicing (Yearim et al., 2015). When HP1 binds to an alternative exon, it results in exon skipping, while binding to an intron just upstream of the alternative exon promotes exon inclusion. Gene body methylation is also hypothesized to be involved in regulation of enhancer and insulator sequences as well as tumorigenesis and cell differentiation. (Q. Wang et al., 2022).

DNA methylation works in close coordination with histone modifications to regulate chromatin architecture and gene expression in eukaryotic cells. Heavily methylated DNA regions are often associated with repressive histone marks, such as H3K9me3 and H3K27me3 (Y. Li et al., 2021). These marks promote the formation of heterochromatin, thereby silencing gene expression. For example, DNMT1 and DNMT3a interact with the histone methyltransferase SUV39H1, which deposits H3K9me3, promoting heterochromatin formation (Lehnertz et al., 2003). Additionally, histone modifications guide DNA methylation by recruiting DNA MTases through histone tails. For example, the accessory protein DNMT3L binds to unmethylated histone H3 tails and recruits DNMT3a and DNMT3b to initiate *de novo* DNA methylation (Ooi et al., 2007). Conversely, the active histone mark H3K4me3 inhibits binding of DNMT3a, DNMT3b, and DNMT3L to histone tails, preventing DNA methylation on active promoters (B.-Z. Li et al., 2011). These examples highlight the bidirectional crosstalk between DNA methylation and histone modifications, highlighting their coordinated role in regulating chromatin structure and gene expression.

Although methylation of cytosines at CpG sites is the most common form of DNA methylation in eukaryotes, cytosines can also be methylated when followed by adenine, thymine or cytosine. (CpA, CpT, and CpC). Non-CpG methylation was first described in plants, but it is also abundant in mammalian embryonic stem cells (ESCs) and brain tissue. For example, (Ramsahoye et al., 2000) demonstrated that mouse ESCs exhibit substantial cytosine methylation at CpA and CpT sites while non-CpG methylation is absent in differentiated somatic tissues. Similarly, (Guo et al., 2014) reported that non-CpG methylation comprises up to 25% of total methylation in human ESCs but decreases upon differentiation. The decrease in non-CpG methylation during cell differentiation suggests that

methylation is involved in dynamic regulation during development and cell fate specification. The non-CpG methylation was thought to be lost after differentiation, however, later study revealed that it reappears in mature neurons where it may play a role in long-term gene regulation (Guo et al., 2014). While canonical CpG methylation is largely maintained through DNA replication, non-CpG methylation in neurons is established *de novo* during neuronal maturation (Guo et al., 2014; Ramsahoye et al., 2000). This indicates that non-CpG methylation accumulates after birth, coinciding with critical developmental processes such as synaptogenesis and synaptic pruning (de Mendoza et al., 2021). Additionally, non-CpG methylation exhibits cell-type specificity within the brain. Genome-wide methylation studies reveal that neurons exhibit higher levels of non-CpG methylation compared to glial cells, demonstrating cell type-specific gene regulation and chromatin organization (Lister et al., 2013).

1.2. Overview of CRISPR-Cas systems

1.2.1. Introduction to CRISPR-Cas

Bacteria and archaea are in constant threat of infection by bacteriophages and other mobile genetic elements. This ever-changing environment pressures the development of strategies to defend against these threats, influencing the genetic diversity of microbial populations (Georjon & Bernheim, 2023). The defense strategies can either be inherent (e.g., restriction-modification systems, abortive infection systems) or adaptive (Georjon & Bernheim, 2023). One of the most studied adaptive immune systems is CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins), which provides the ability to "remember" past infections and prevent their reoccurrence (Barrangou et al., 2007; Mojica et al., 2005). The CRISPR-Cas system is generally encoded within the CRISPR locus (**Figure 1.2**), comprised of two main elements:

(1) Cas (CRISPR-associated) genes, which encode Cas proteins, involved in different steps of the adaptive immunity.

(2) CRISPR array – a collection of DNA sequences, called spacers, that keep the genetic information about the past invaders, interspaced with short direct repeats.

In some CRISPR-Cas systems, the CRISPR array is preceded by a leader sequence that regulates the direction of transcription and is involved in acquisition of new spacers into the array (Makarova et al., 2020).

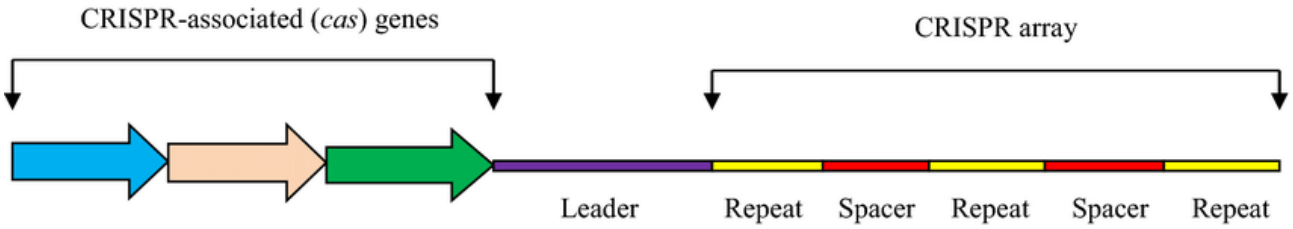

Figure 1.2 Structure of the CRISPR locus. CRISPR-associated (*cas*) genes are located upstream of the CRISPR array. The cas genes and the CRISPR array are linked through the leader sequence. Adapted from (Barman et al., 2020)

The CRISPR-Cas defense system works in three steps (**Figure 1.3**):

A. Adaptation

During the adaptation stage, a foreign genetic material is captured by the CRISPR-Cas system and processed into a short (20-40bp) fragment to form a spacer which is later added to the CRISPR array (Mcginn & Marraffini, 2018). This process involves a highly conserved Cas1-Cas2 integrase complex that predominantly inserts new spacers to the leader end of the CRISPR array. Insertion of spacers on the leader end prioritizes defense against most recent invaders who are likely to pose the biggest threat to the host cell. Ordered integration of spacers into the CRISPR array gives chronological information about past infections and serves as a historical record about oldest and newest encounters (Mcginn & Marraffini, 2018).

B. Expression

The expression stage is critical in CRISPR immune response, where the CRISPR array is transcribed to produce RNA molecules that guide Cas proteins to their respective targets. The host's RNA polymerase starts the transcription from the leader end of the CRISPR array and produces a single precursor CRISPR RNA (pre-crRNA) that carries the spacers and surrounding repeat sequences (Deltcheva et al., 2011). Following the transcription, the pre-crRNA molecule is processed to generate multiple mature CRISPR RNAs (crRNAs) . Depending on the type of the CRISPR-Cas system, the crRNA processing is done by either Cas6 endoribonuclease, endogenous ribonuclease III or by the effector Cas protein that is later involved in the interference process (Makarova et al., 2020). Each crRNA encodes two main elements: (1) a palindromic repeat sequence, that induces formation of crRNA secondary structures necessary for binding with Cas proteins; (2) a spacer sequence that is necessary for targeting foreign genetic material. The mature crRNA molecules are then loaded onto Cas proteins and form an active ribonucleoprotein (RNP) complex that is important in the last stage of the CRISPR-Cas immune response – interference (Marraffini & Sontheimer, 2010).

C. Interference

The interference stage is the last stage of the CRISPR-Cas adaptive immunity response. During this stage, the crRNA guides the active ribonucleoprotein (RNP) complex to recognize and destroy foreign DNA or RNA (Hille et al., 2018). First, the RNP scans for the presence of a short DNA sequence called PAM (protospacer adjacent motif) that flanks the target sequence. The same principle applies to RNA-targeting CRISPR-Cas systems where the PFS (protospacer flanking sequence) serves an equivalent purpose as the PAM sequence (Gleditzsch et al., 2019; Zhang et al., 2024). Both PAM and PFS sequences are crucial for the specificity of the CRISPR-Cas system as they ensure that it can distinguish between self and non-self genetic material. Once the PAM (or PFS) sequence is recognized by the Cas nuclease, the spacer portion of the crRNA base pairs with a complementary sequence of the foreign DNA/RNA. When the crRNA binds its complementary sequence, a Cas nuclease cuts the target which is then degraded by either other CRISPR-associated proteins or endogenous exonucleases. If the infection of the same foreign genetic element recurs, the CRISPR-Cas system uses its "memory" encoded within the CRISPR array to produce respective crRNAs that later guide the nucleases to cut their targets (Hille et al., 2018).

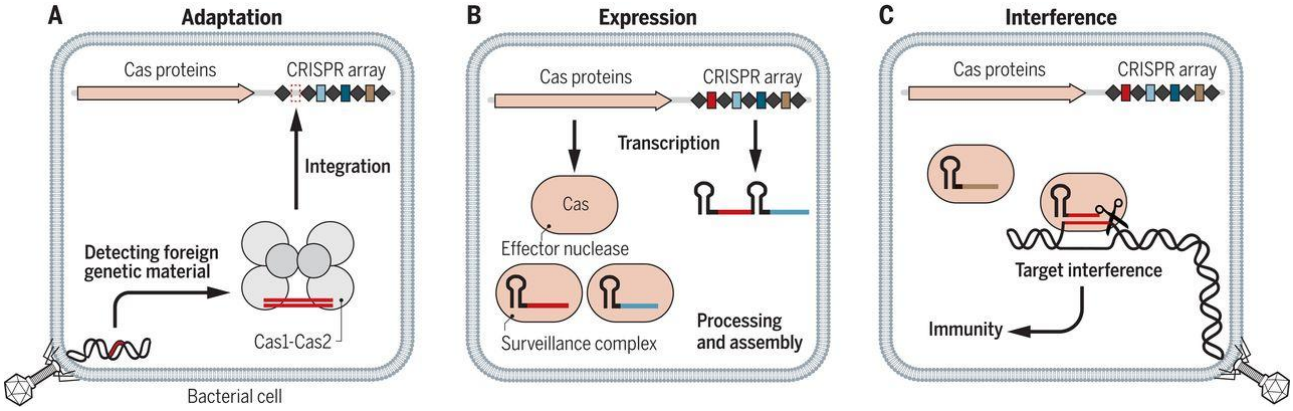

Figure 1.3 Stages of CRISPR-Cas adaptive immunity. **(A)** Adaptation: a foreign genetic material is captured by Cas1-Cas2 and integrated into the CRISPR-array. **(B)** Expression: the CRISPR array and Cas proteins are expressed to form an active surveillance complex. **(C)** Interference: the Cas effector nuclease is guided to bind and target foreign genetic elements complementary to the crRNA. Adapted from (Knott & Doudna, 2018).

The crRNA can be engineered to target any target sequence of interest, therefore, this programmable nature of CRISPR-Cas nucleases holds a potential for precise genome editing technologies. Once the nuclease recognizes and binds the target sequence, a double-stranded break (DSB) is induced (Jinek et al., 2012). If not repaired, DSBs are highly detrimental to the cell as they can result in chromosomal rearrangements, mutations, cell cycle arrest and other genome-destabilizing effects. To avoid this, the cellular machinery repairs DSBs through two main repair mechanisms - non-homologous end joining (NHEJ) or homology-directed repair (HDR) (Ran et al., 2013). In genome editing context, utilization of NHEJ is useful when creating small insertions or

deletions at the break site as the broken DNA ends are ligated back together. In contrast, HDR employs a homologous donor template that allows for precise genome editing outcomes (Ran et al., 2013). However, this process is less efficient than NHEJ as it requires S and G2 phases of the cell cycle for homologous recombination to occur (Saleh-Gohari & Helleday, 2004). Regardless, both DSB repair pathways are already utilized in CRISPR-Cas genome editing technologies to achieve various genetic modifications with applications in research, diagnostics, therapeutics and agriculture (Chavez et al., 2023; Tuncel et al., 2025).

1.2.2. A closer look at Type II CRISPR-Cas systems

Based on the composition of the effector complexes, CRISPR-Cas systems are divided into two major classes (Makarova et al., 2020). Class 1 systems use multi-protein effector complexes while class 2 systems use a single effector protein. Within each class, systems are further divided into types based on signature Cas genes. The signature gene for type II CRISPR-Cas systems is Cas9 which belongs to class 2 of CRISPR-Cas systems (Makarova et al., 2020). Type II CRISPR-Cas systems are subdivided into four subtypes (II-A, II-B, II-C, II-D) based on the similarity of their Cas9 proteins and the presence or absence of additional Cas proteins beyond Cas9, Cas1 and Cas2. All type II CRISPR systems involve a trans-activating crRNA (tracrRNA) that is usually encoded near the CRISPR array. During the expression stage of the CRISPR-Cas adaptive immunity, the tracrRNA hybridizes with the repeat sequences of pre-crRNA and enables full maturation of crRNA molecules (Deltcheva et al., 2011). Type II-A systems include the accessory protein Csn2, while type II-B systems are characterized by the presence of Cas4 (Makarova et al., 2020) (**Figure 1.4**). Both Csn2 and Cas4 are involved in the adaptation stage of CRISPR-Cas immunity, however, their absence is a characteristic of type II-C CRISPR-Cas systems. While ongoing research identifies new variants, the relative abundances of type II-A, II-B and II-C are around 55%, 3% and 41% respectively (Koonin et al., 2017).

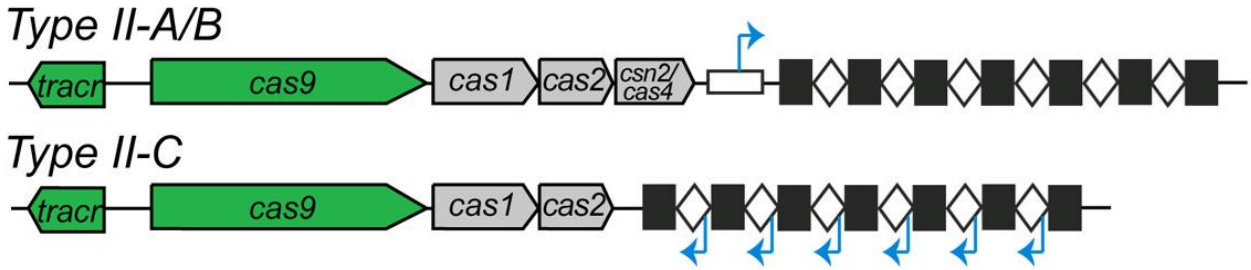

Figure 1.4 Organization of type II CRISPR-Cas loci. Type II-A and II-B systems include accessory proteins Csn2 and Cas4, respectively, which are absent in type II-C. Blue arrows show CRISPR array transcription. Unlike type II-A/B, most type II-C arrays have internal promoters within repeats rather than at the leader end (white rectangle). tracr – trans-activating crRNA, black squares – repeats; white squares – spacers. Adapted from (Mir et al., 2018).

Recent studies identified subtype II-D that differs from the classical II-A, II-B and II-C subtypes by architecture of the genomic locus and Cas9 protein features (Aliaga Goltsman et al., 2022; J. Yang et al., 2025). The distinct genetic composition suggests potentially novel mechanisms of CRISPR-Cas adaptive immunity, however, the functional roles of type II-D systems remain widely uncharacterized.

Type II CRISPR systems involve two RNA molecules – CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA) – which bind together through complementary base pairing to make a guide RNA (gRNA) (**Figure 1.5** A). The crRNA contains the spacer sequence that base pairs with a respective DNA target, while the tracrRNA is essential for processing of the pre-crRNA and formation a stable, functional complex with Cas9 (Deltcheva et al., 2011). To simplify genome editing approaches, the tracrRNA and crRNA can be fused together to make a single-guide RNA (sgRNA), retaining all the essential sequences for formation of a stable Cas9 effector complex and subsequent DNA targeting (Jinek et al., 2012) (**Figure 1.5** B).

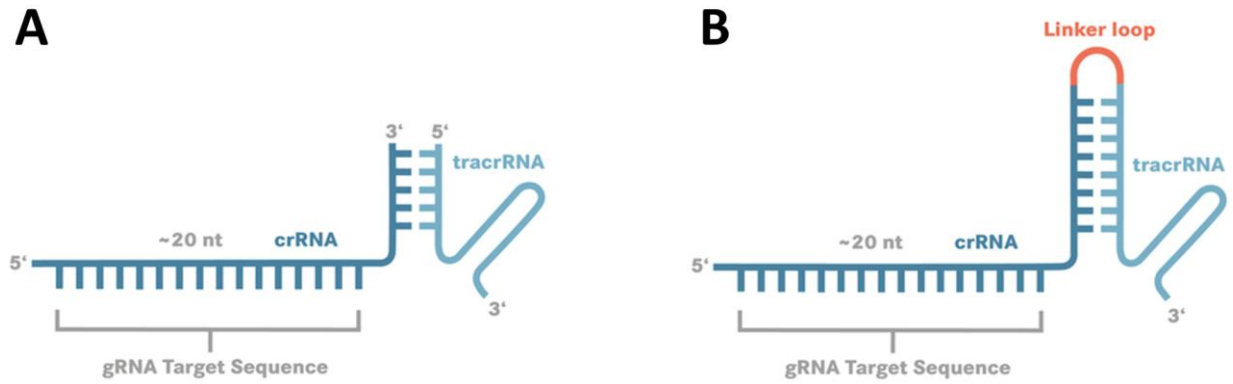

Figure 1.5 Alternative guide RNA (gRNA) variants, accommodated by Cas9 nuclease. **(A)** A naturally occurring crRNA-tracrRNA hybrid. **(B)** An engineered gRNA variant where crRNA and tracrRNA are fused into a single-guide RNA (sgRNA) through a linker loop. Adapted from "LubioScience", 2023.

Cas9, a signature endonuclease of type II CRISPR-Cas systems, shares a similar domain architecture within all subtypes. The best studied Cas9 comes from *Streptococcus pyogenes* (SpCas9), which belongs to the type II-A CRISPR-Cas systems (**Figure 1.6**). SpCas9 is comprised of two major lobes: the REC (recognition) lobe and the NUC (nuclease) lobe, which are connected by an arginine-rich bridge helix (BH) (Jinek et al., 2014; Nishimasu et al., 2014). The REC lobe is responsible for binding both gRNA and target DNA. Once the gRNA is bound, SpCas9 undergoes conformational changes that prepare the complex to accommodate targeted DNA. The PAM-interacting domain (PI or PID), found within the NUC lobe, scans for a protospacer adjacent motif (PAM) near the target sequence (Nishimasu et al., 2014). Once the PAM is identified, the gRNA base pairs with the target DNA and induces conformational changes in the REC lobe. Conformational shifts in the REC lobe

allosterically activate the HNH nuclease domain, which then cleaves one strand of the target DNA. Activation of the HNH domain subsequently triggers conformational changes at the HNH-RuvC interface, leading to activation of the RuvC domain, which then cleaves the other strand of the target DNA (Nishimasu et al., 2014; Sternberg et al., 2015). As a result, the coordinated HNH and RuvC activity produces a double-stranded break (DSB) in the target DNA. After cleavage, SpCas9 remains bound to the DNA and is eventually displaced by cellular mechanisms, allowing it to be recycled (Sternberg et al., 2015).

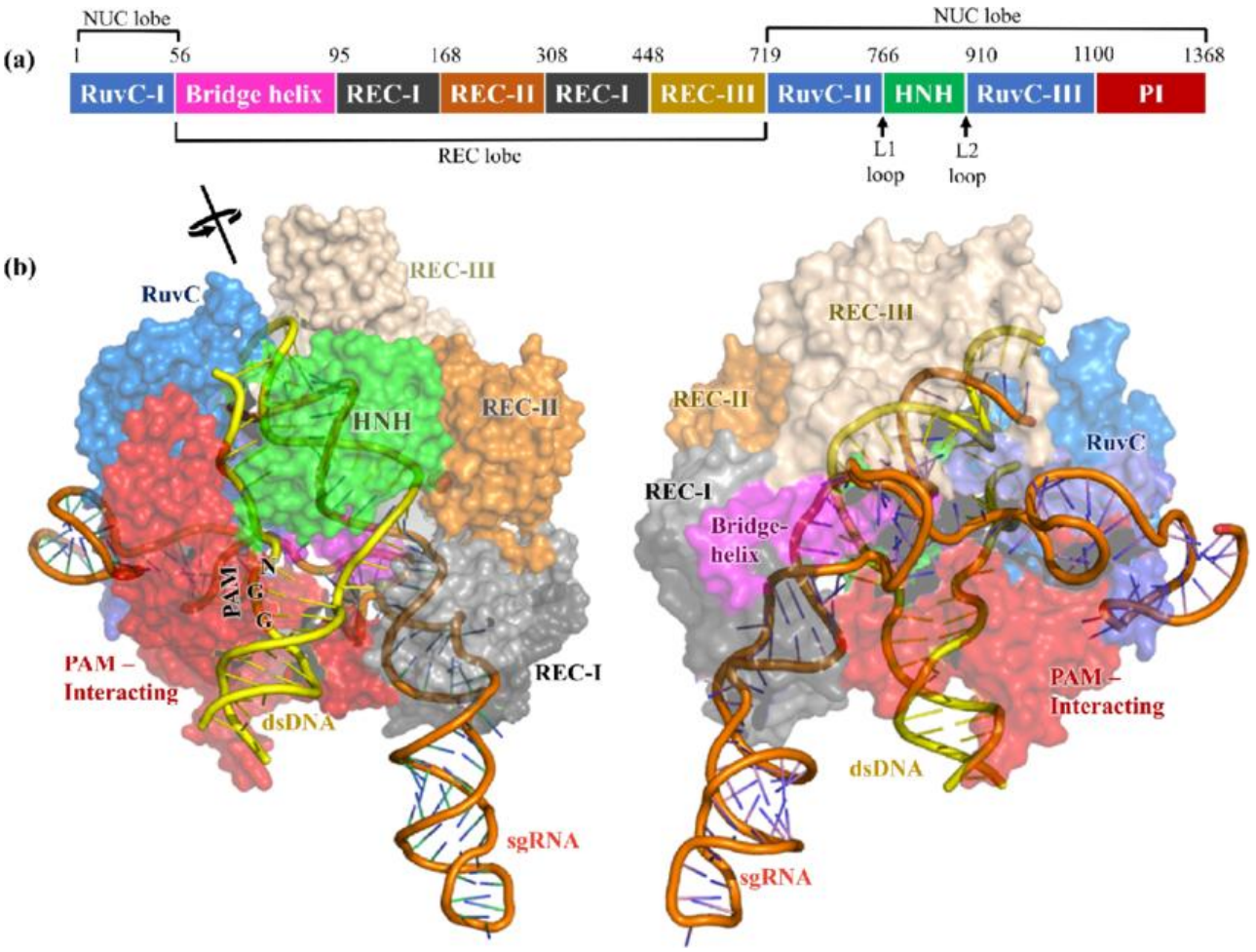

Figure 1.6 Structural and functional organization of SpCas9. **(a)** Domain organization of *Streptococcus pyogenes* Cas9 (SpCas9). SpCas9 is comprised of NUC (nuclease) and REC (recognition) lobes. **(b)** Spatial structure of SpCas9, bound with sgRNA and target DNA. PDB ID 5F9R. Adapted from (Bhattacharya & Satpati, 2023).

Regardless of similar domain organization, type II-A Cas9s are typically larger than type II-C Cas9s. Also, type II-C Cas9s often require longer PAMs and gRNAs, compared to type II-A and II-B (Yamada et al., 2017). Studies suggest that Cas9 proteins from Type II-A systems have greater DNA unwinding and cleavage efficiency, which may contribute to their generally higher genome editing performance compared to type II-C Cas9s (Mir et al., 2018). Type II-B Cas9s vary in size but are generally larger than Cas9s from II-B and II-C types. Recent studies show that type II-B Cas9s

have enhanced specificity due to an extended REC3 domain that senses mismatches and reduces off-target cleavage (Bestas et al., 2023; Hibshman & Taylor, 2024). However, this specificity mechanism is still being investigated, therefore type II-B Cas9s are not yet broadly adopted for routine editing. Variations in Cas9 size, PAM requirements, gRNA length, target specificity, and DNA cleavage efficiency influence the choice of Cas9 variants suitable for genome editing. Smaller Cas9s, like those from type II-C systems, are attractive for delivery in size-constrained vectors, however, their stricter PAM requirements and lower DNA cleavage activity limit versatility. In contrast, type II-A Cas9s have broad PAM compatibility, strong nuclease activity, making them more suitable for a wide range of applications. As a result, type II-A Cas9s remain the most widely used, while the potential of type II-B and type II-C enzymes continues to be explored.

1.3. Epigenetic editing with CRISPR-Cas9

1.3.1. CRISPR-Cas9 epigenetic editing platforms

Early work in epigenome editing technologies focused on utilization of engineered-DNA binding proteins, such as zinc finger nucleases (ZFNs) or transcription activator-like effectors (TALEs), enabling targeted manipulation of DNA methylation (Gersbach et al., 2014). However, the design and assembly of both ZFNs and TALEs can be technically challenging, as targeting new DNA sequences requires modifying the protein structure itself. Conversely, CRISPR-based approaches only require reprogramming of the guide RNA to target different genomic loci, making the system more flexible for epigenetic editing. As a result, CRISPR-based epigenetic editing platforms are now preferred for a wide range of applications in epigenetic research (McCutcheon et al., 2024).

CRISPR-Cas9 epigenetic editing platforms use a catalytically inactive form of Cas9, called dead Cas9 (dCas9). dCas9 is typically fused to various epigenetic effector domains to modulate epigenetic changes at specific genomic loci without cutting the targeted DNA (Qi et al., 2013). Depending on the effector domain attached to dCas9, these platforms can either add or remove epigenetic marks, such as DNA methylation or histone modifications (acetylation or methylation), to activate or repress gene expression. For example, dCas9-Tet1 and dCas9-Dnmt3a fusion proteins were directed to target specific CpG sites in synthetic methylation reporter constructs (X. S. Liu et al., 2016). This study demonstrated targeted methylation by dCas9-Dnmt3a at unmethylated promoters and enhancers that led to gene repression and altered chromatin architecture. Additionally, targeted demethylation by dCas9-Tet1 at methylated promoters activated gene expression (X. S. Liu et al., 2016). Another study reported a more efficient and widespread targeted DNA methylation by using a fused dCas9-Dnmt3a-Dnmt3L complex (Stepper et al., 2017). This complex was targeted at endogenous gene promoters such as EpCAM, CXCR4 and TFRC, enabling precise locus-specific

methylation that induced transcriptional repression. Furthermore, (Lei et al., 2017) have reported on an engineered fusion of dCas9 with the prokaryotic DNA methyltransferase MQ1 (dCas9-MQ1). A mutant version of this fusion protein, containing a Q147L substitution in the MQ1 domain, exhibited greater precision and reduced off-target methylation compared to earlier dCas9-DNMT3a-based systems (Stepper et al., 2017). These studies highlight the versatility of dCas9-based epigenetic editing platforms, demonstrating their potential for locus-specific regulation of gene expression through targeted modification of the epigenome.

1.3.2. Effects of DNA methylation

A study by (Hsu et al., 2013) investigated whether DNA methylation influences the cleavage efficiency of SpCas9. *In vitro* cleavage assays with plasmid DNA methylated at CpG sites demonstrated that SpCas9 efficiently cleaved both methylated and unmethylated DNA, indicating that CpG methylation does not inhibit Cas9 activity. Additionally, they performed *in vivo* experiments by targeting a highly methylated endogenous locus (SERPINB5) in HEK293T cells, showing successful induction of indels and confirming that SpCas9 can effectively edit methylated genomic DNA (Hsu et al., 2013). The results of this study suggested that DNA methylation is not a barrier to CRISPR-Cas9-mediated genome editing. However, it is the chromatin state induced by methylation, rather than the methylation modification itself, that poses a barrier to efficient Cas9 targeting and cleavage. Several studies have demonstrated that nucleosome positioning and chromatin compaction reduce Cas9 accessibility and editing efficiency, with heterochromatic regions being less permissive to Cas9 binding and cleavage (Daer et al., 2017; Fujita et al., 2016; Wu et al., 2014). For example, (Wu et al., 2014) performed genome-wide binding analyses of catalytically inactive SpCas9 (dCas9) in mouse embryonic stem cells to identify potential binding sites. Then, they expressed catalytically active Cas9 with the same sgRNAs and measured cleavage at these sites by targeted deep sequencing. Their results showed that although dCas9 binds extensively to sites with partial sequence complementarity, actual cleavage was highly specific and strongly influenced by chromatin accessibility. Many potential off-target binding sites in closed chromatin were poorly cleaved, emphasizing that chromatin architecture, rather than DNA methylation, governs SpCas9 activity *in vivo* (Wu et al., 2014). Similarly, a study by (Fujita et al., 2016) demonstrated allele-specific genome editing at the p16INK4a locus, where one allele is heavily methylated and transcriptionally silent, and the other is unmethylated but contains a single-nucleotide insertion. This study demonstrated that allele-specific editing and locus binding by CRISPR-Cas9 were influenced by the chromatin state associated with methylation. Notably, allele-specific binding was observed *in vivo* but not *in vitro*, highlighting the critical role of chromatin context in modulating Cas9 accessibility and activity (Fujita

et al., 2016). Together, these findings support the consensus that SpCas9 is not inherently affected by DNA methylation itself. Instead, it is the degree of DNA compaction and accessibility that plays a role in determining SpCas9 binding and cleavage efficiency *in vivo*.

1.3.3. AceCas9 senses DNA methylation

Cas9-based tools are still limited by off-target effects, which can lead to unanticipated epigenetic changes. These off-target effects in CRISPR-Cas9-based epigenetic editing platforms may disrupt gene regulation in unpredictable ways (Hunt et al., 2023). Additionally, as discussed in section 1.3.2, the widely used SpCas9 lacks sensitivity to DNA methylation as it generally cannot distinguish methylated and unmethylated DNA. Consequently, CRISPR-Cas9 epigenetic editing platforms remain prone to off-target activity and cannot be selectively guided to target a genomic region according to its local methylation state. Furthermore, such platforms are not designed to track changes in DNA methylation over time, which makes it difficult to study how epigenetic marks change dynamically.

These limitations can be addressed by using a Cas9 variant derived from *Acidothermus cellulolyticus* (AceCas9). AceCas9 is the only currently known Cas9 that exhibits sensitivity to DNA methylation (Das et al., 2020). AceCas9 recognizes a 5'-NNNCC-3' PAM sequence and is influenced by the methylation status within this motif. The PAM recognition is established by Glu1044, Arg1088 and Arg1091 residues in the PAM-interacting domain (PID) of AceCas9, forming a complex hydrogen bonding network with the first cytosine and its paired guanines (Das et al., 2020) (**Figure 1.7 A**). Methylation on the first cytosine in the PAM disrupts these critical contacts, leading to a loss of cleavage activity on both linear (**Figure 1.7 B**) and plasmid (**Figure 1.7 C**) DNA targets.

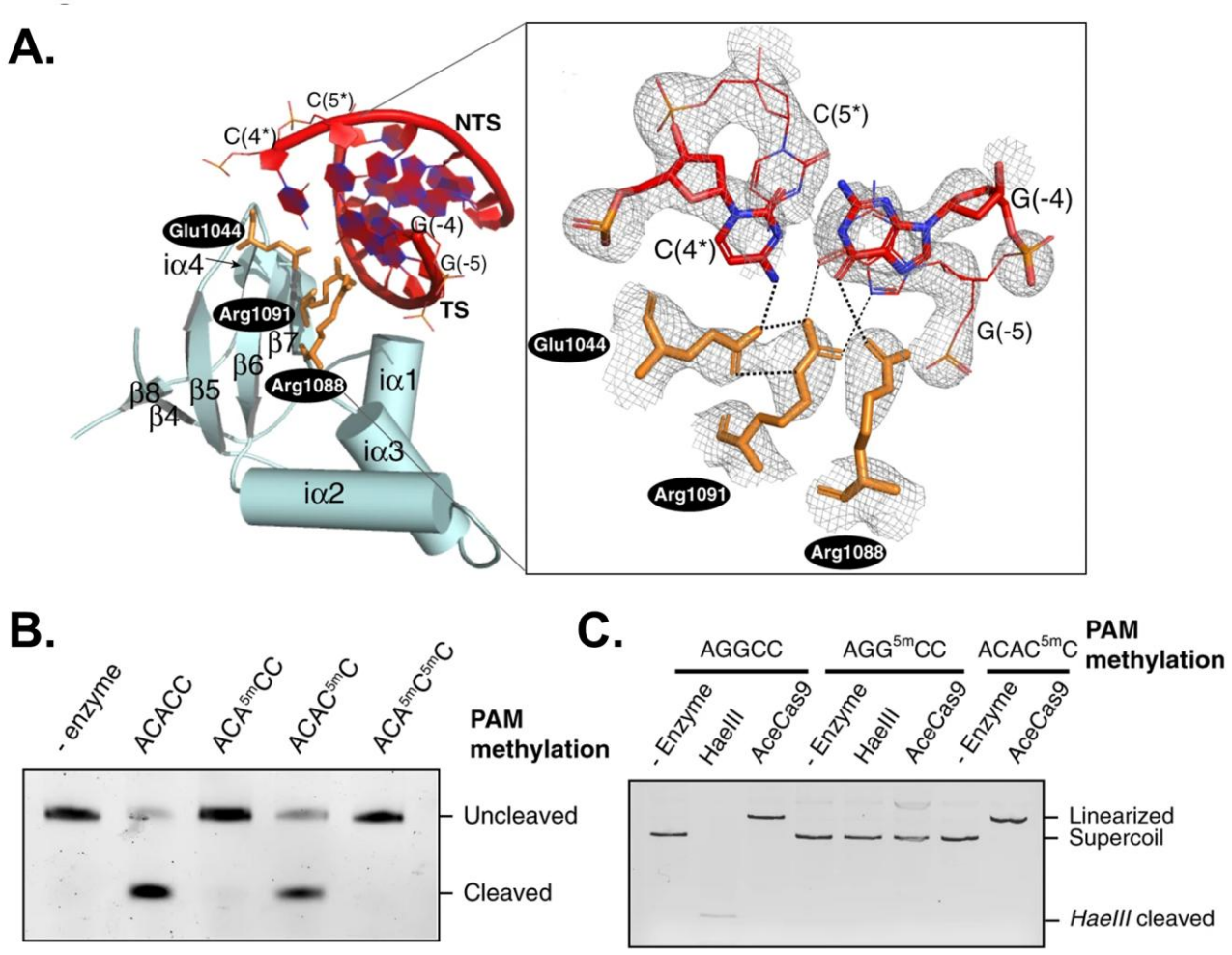

Figure 1.7 Molecular interactions and methylation-dependent activity of AceCas9**. (A)** The architecture of AceCas9 PAM-interacting subdomain and its specific contacts with the 5'-NNNCC-3' PAM sequence. C4* corresponds to the first cytosine in the 5′-NNNCC-3′ PAM sequence, and C5* denotes the second cytosine. The carboxylate group of Glu1044 interacts with the exocyclic amino group of C4* on the non-target strand (NTS) while the guanidinium groups of Arg1088 and Arg1091 form contacts with the exocyclic oxygen atoms of G(−4) and G(−5) on the target strand (TS), respectively. TS is defined as the DNA strand that hybridizes with the gRNA upon PAM recognition. Additionally, Arg1091 forms a salt bridge with Glu1044, reinforcing the interaction network involving Glu1044, C4*, Arg1091, and G(−5). PDB ID 6WBR. **(B)** Cleavage activity of AceCas9 linear DNA substrates containing either non-methylated or methylated PAM sequences. **(C)** Cleavage activity of AceCas9 plasmid DNA substrates containing either non-methylated or methylated PAM sequences. HaeIII restriction endonuclease was used to confirm the methylation on C4*. Adapted from (Das et al., 2020).

AceCas9 is a member of the type II-C CRISPR-Cas systems and originates from thermophilic environments (Mohagheghi et al., 1986; Tsui et al., 2017). AceCas9 has been shown to be active *in vitro*, functioning across a broad temperature range from 25 °C to 60 °C (Tsui et al., 2017). In addition, AceCas9 was also demonstrated to be active in *E. coli* (Tsui et al., 2017), as well as in thermophilic model bacteria *Clostridium thermocellum* (Walker et al., 2020). The optimal guide RNA for AceCas9 contains a 24-nucleotide spacer, which is longer than the 20-nucleotide spacer typically used with the widely studied SpCas9 (Jinek et al., 2012; Tsui et al., 2017). Longer spacer sequence imposes more

stringent target recognition requirements, therefore utilization of AceCas9 provides an opportunity to reduce off-target effects commonly observed with other Cas9-derived tools (Hunt et al., 2023). Additionally, AceCas9 offers more specificity through its unique ability to sense DNA methylation. Together, these features make AceCas9 a promising candidate for precise genome and epigenome editing applications. Regardless, its activity in human cells remains largely unknown and requires further investigation.

1.4. Engineering of CRISPR-Cas9 nucleases

Advances in molecular biology enabled engineering of enzymes with enhanced or altered properties to better suit specific applications. This is typically achieved by two main strategies – directed evolution or rational design (**Figure 1.8**) (Song et al., 2023). Directed evolution replicates the process of natural selection in the laboratory by generating large libraries of enzyme variants and selecting those with desired traits through iterative rounds of mutation and screening. In contrast, rational design relies on detailed knowledge of enzyme structure and function to introduce targeted mutations aimed at enhancing specific characteristics (**Figure 1.8**) (Song et al., 2023).

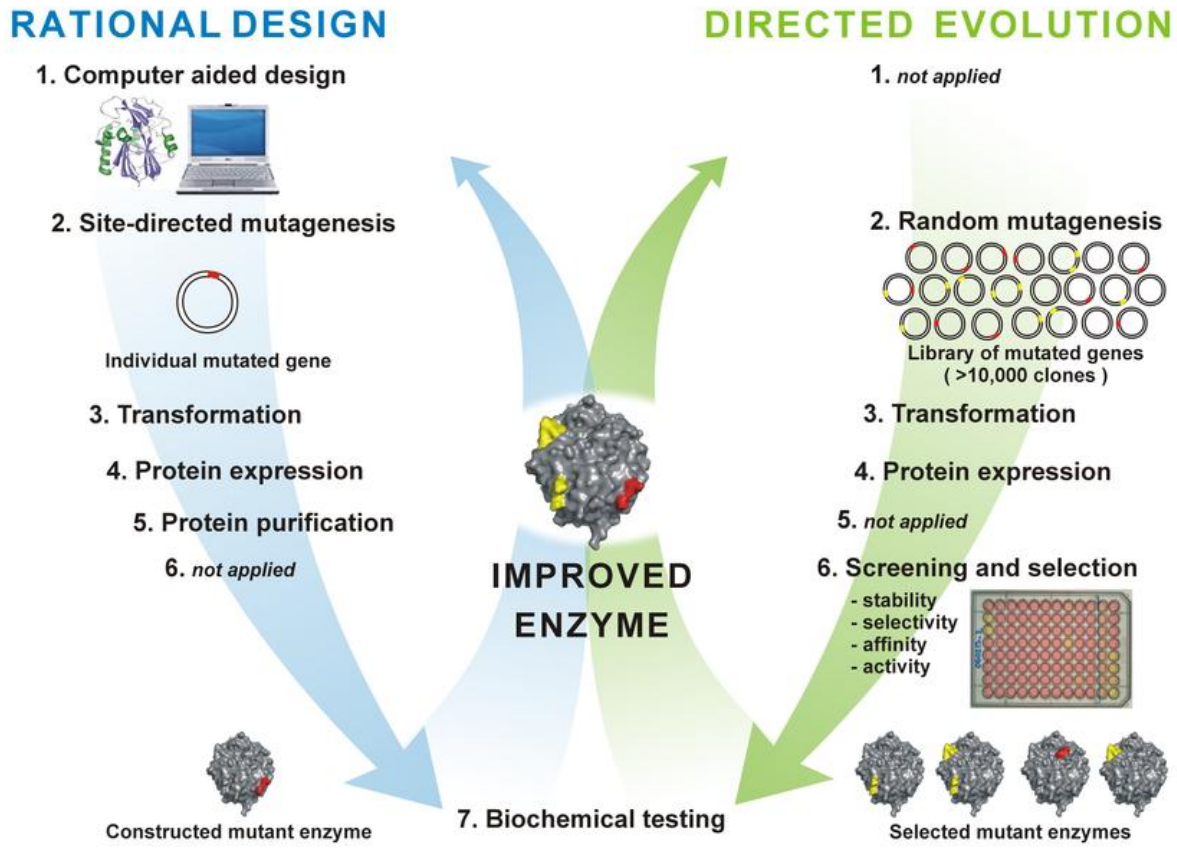

Figure 1.8 Comparison of rational design and directed evolution strategies for enzyme engineering. Rational design uses structure-based insights to introduce targeted mutations, while directed evolution relies on random mutagenesis and selection to evolve enzymes with improved traits. Adapted from (Dvorak, 2007).

Both directed evolution and rational design strategies can be combined into a semi-rational approach, where researchers start with the rational design to guide the mutation process but complement this with directed evolution by screening a library of variants (Xiong et al., 2021). In recent years, artificial intelligence (AI) tools have increasingly supported rational and semi-rational designs by predicting promising mutations and prioritizing variants for experimental testing, thereby reducing screening burden and improving hit rates (Y. Wang et al., 2025). The mentioned enzyme engineering strategies are applicable to a wide variety of enzymes, therefore, they are also used to adapt CRISPR-Cas nucleases to improve their targeting specificity, PAM compatibility, editing efficiency and other properties of interest (Hu et al., 2018; Huang et al., 2023; Kim et al., 2023).

1.4.1. Directed evolution

Researchers have employed a variety of selection methods to drive directed evolution and isolate Cas9 variants with modified properties. For example, a directed evolution platform EPICA (Eukaryotic Platform to Improve Cas Activity) combines yeast auxotrophic selection with mammalian cell reporter assays (Ruta et al., 2024). This platform was utilized to engineer *Campylobacter jejuni* Cas9 (CjCas9), producing UltraCjCas9 – a variant that exhibits up to 12-fold greater activity in mammalian cells while maintaining high specificity. Another successful directed evolution strategy is Phage-Assisted Continuous Evolution (PACE) – a high-throughput directed evolution technique that links protein evolution with phage propagation (Esvelt et al., 2011) PACE was used to modify the PAM recognition of Nme2Cas9 (Huang et al., 2023). This method resulted in variants like eNme2-T.1 and eNme2-T.2, which can recognize a broader range of pyrimidine-rich PAM sequences, thereby expanding the targeting capability of this Cas9 ortholog. PACE was also utilized to evolve SpCas9 with enhanced PAM compatibility, such as xCas9, which recognizes a broader range of PAMs (e.g., NG, GAA, GAT) (Hu et al., 2018). Additionally, the Sniper screen, a bacterial-based selection system, applies simultaneous positive and negative selection in *E. coli* to enrich for variants that retain high on-target activity while minimizing off-target effects (Lee et al., 2019). This method was originally employed to evolve SpCas9 into Sniper-Cas9, a variant that offers enhanced specificity while maintaining its on-target activity. It has since been applied to evolve subsequent variants like Sniper2L, showing its utility as a general directed evolution platform for Cas9 specificity and activity optimization (Kim et al., 2023). Another directed evolution strategy leverages toxic gene-based selection in bacteria, where Cas9 variants are evolved by linking their cleavage activity to survival through targeting a toxic gene such as ccdB. This method was successfully applied to evolve catalytically enhanced Cas9 variants (CECas9) from both AceCas9 and SpCas9 (Hand et al., 2021). The AceCECas9 variant exhibited up to a 4-fold improvement in

catalytic efficiency, while the SpCECas9 showed enhanced performance in homology-directed repair-based gene insertion in human colon cancer cells (Hand et al., 2021). Together, these directed evolution strategies demonstrate the wide range of available approaches for optimizing Cas9 nucleases to meet specific genome editing needs such as enhanced targeting specificity, catalytic activity and expanded PAM recognition.

1.4.2. Rational design

In parallel with directed evolution, rational design was also utilized to evolve CRISPR-Cas9 nucleases by introducing specific, structure-guided mutations to enhance performance or expand functionality. Several rationally engineered SpCas9 variants with altered PAM specificities were described by (Kleinstiver et al., 2015). The reported SpCas9 VQR variant contained D1135V, R1335Q, and T1337R mutations, which enabled recognition of 5′-NGA-3′ PAM sequences instead of the canonical 5′-NGG-3′. A similar variant, called EQR, differs from VQR by having a D1135E substitution and prefers 5′-NGAG-3′ PAM sequences. The VRER variant had rationally designed D1135V, G1218R, R1335E and T1337R mutations, expanding targetable sites to those with 5’-NGCG-3’ PAMs (Kleinstiver et al., 2015). Variants such as Enhanced Specificity Cas9 (eSpCas9) and High-Fidelity Cas9 (Cas9-HF1) were engineered by weakening nonspecific contacts between Cas9 and DNA, reducing off-target cleavage while retaining on-target efficiency (Kleinstiver et al., 2016; Slaymaker et al., 2016). SpCas9-NG is another rationally engineered variant of SpCas9, designed to recognize a broader range of PAM sequences. Unlike wild-type SpCas9 that strictly requires an 5’-NGG-3’ PAM, SpCas9-NG recognizes PAMs with the general motif 5’-NG-3’ (Nishimasu et al., 2018). The Hyper-Accurate Cas9 (HypaCas9) contains mutations in the REC3 domain that increase the conformational checkpoint stringency before cleavage, enhancing specificity by preventing cleavage at mismatched sites while maintaining robust on-target activity (J. S. Chen et al., 2017). These examples illustrate how rational design allows precise modification of Cas9 nucleases, enabling their adaptation to specific genome editing applications.

2. Materials and Methods

2.1. Materials

2.1.1. Bacterial strains

- ***E. coli* BW25141(λDE3)** (Gift from Prof. D. Edgell) *lacI^q rrnB$_{T14}$ ΔlacZ$_{WJ16}$ ΔphoBR580 hsdR514 ΔaraBAD$_{AH33}$ ΔrhaBAD$_{LD78}$ galU95 endA$_{BT333}$ uidA(ΔMluI)::pir$^+$ recA1 λ(DE3 [lacI lacUV5::T7 gene1 ind1 sam7 Δnin5])*

- ***E. coli* DH5α**: *F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK–mK+), λ–*

- ***E. coli* TOP10**: *F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu) 7697 galU galK rpsL (StrR) endA1 nupG λ–*

2.1.2. Plasmids

Table 2.1 provides a brief description of each plasmid used in this work.

Table 2.1 Summary of all plasmids used in this work.

Plasmid name	Description
J.Pl1.58	pET type vector customized for Golden Gate cloning of protein sequences of interest. The cloning site is followed by a C-terminal TwinStrep-tag sequence for protein purification. Contains T7 expression system.
J.Pl1.75	Human codon-optimized AceCas9+gRNA co-expression plasmid. gRNA contains a spacer targeting PCSK9 gene. Gift from the research group of Dr. Hong Li (Florida State University).
J.Pl1.77	Human codon-optimized AceCas9 and its gRNA co-expression vector. gRNA contains a protospacer targeting UBE3A gene. Gift from the research group of Dr. Hong Li (Florida State University).
J.Pl2.02	pUC type vector that encodes XTEN linker followed by TwinStrep-tag sequence. Synthesized by "Azenta Life Sciences".
J.Pl2.26	Human codon-optimized AceCas9-TwinStrep+gRNA co-expression plasmid. gRNA contains a targeting UBE3A gene.
J.Pl2.30	pACYCDuet™-1 vector containing two multiple cloning sites under T7 promoters for cloning of two sequences of interest.
J.Pl2.31	p11-LacY-wtx1 reporter plasmid that encodes the ccdB toxin gene under inducible BAD promoter. Contains a target for AceCas9, flanked by 5'-NNN**CG**-3' PAM sequence. Gift from the research group of Dr. T. Karvelis (Vilnius University), deposited on Addgene by Dr. Huimin Zhao Lab (University of Illinois Urbana-Champaign).
J.Pl2.33	pACYCDuet™-1 vector containing cloned sequence of gRNA of AceCas9, targeting J.Pl2.31 and J.Pl2.39.
J.Pl2.34	pACYCDuet™-1 co-expression vector containing *E. coli* codon-optimized sequence of AceCas9 and its gRNA, targeting J.Pl2.31 and J.Pl2.39.
J.Pl2.37	pET type protein expression vector that includes a sequence for *E. coli* codon-optimized AceCas9 and a TwinStrep-tag fused on C-terminus. The gene block of AceCas9 was synthesized and cloned into J.Pl1.58 by "Azenta Life Sciences".
J.Pl2.39	p11-LacY-wtx1 reporter plasmid that encodes the ccdB toxin gene under inducible BAD promoter. Contains 1 nucleotide difference from J.Pl2.31, encodes a target for AceCas9 flanked by 5'-NNN**CC**-3' PAM sequence.

Specific utilization of each plasmid is elaborated in further sections.

2.1.3. Oligonucleotides & gene-blocks

Oligonucleotides and gene-blocks, listed in **Table 2.2**, were synthesized at "Azenta Life Sciences".

Table 2.2 Sequences of oligonucleotides and gene-blocks used in this work. Abbreviations "F" and "R" stand for forward and reverse positions against the PCR templates

TwinStrep cloning primers (5'→3')		
F	J.Pr6.80	CAGCCTCGACTGTGCCTTCTAGTTG
R	J.Pr6.66	GTGTTCCGCCGGACCAAGGATC
F	J.Pr6.67	GATCCTTGGTCCGGCGGAACAC
R	J.Pr6.68	GGCACAGTCGAGGCTGATCAGC
AceCas9 gRNA cloning template and primers (5'→3')		
J.Pr 7.45 (dsDNA gene-block)		CGAACAGAAAGTAATCGTATTGTACACGGCCGCATAATCGTAATACGACTCACTATAGGTACCC TGTTCACCGGTAGCAAAGCTGGGGAGCCTGAAAAGGCTACCTAGCAAGACCCCTTCGTGGGGTC GCATTCTTCACCCCCTCGCAGCAGCGAGGGGGTTCGGCCGGCATGGTCCCAGCCTCCTCGCTGG CGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACCTGCTGCCACCGCTGAGCAATAACT AGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGGCACACGGTCACACTGCT TCCGGTAGTCAATAAACCGG
F	J.Pr7.39	GGTCACACTGCTTCCGGTAGTC
R	J.Pr7.40	GCGGCCGTGTACAATACGATTAC
F	J.Pr7.41	CGAACAGAAAGTAATCGTATTGTACACGG
R	J.Pr7.42	CCGGTTTATTGACTACCGGAAGCAG
AceCas9 cloning primers (5'→3')		
F	J.Pr7.43	CGTATTCAGGCACCATGGGTGGCAGCGAAGTCGGCA (NcoI recognition site)
R	J.Pr7.44	GCAGTAGTCAGTCGACTTACGGGGTGCCGCCACTCCAC (SalI recognition site)
F	J.Pr7.55	GTGATAGCATTGCCGTGGCCG
Quick-change mutagenesis (QCM) primers (5'→3')		
F	J.Pr7.53	GTTCACCGGTAGCAAAGAACCAAGTAAAG
R	J.Pr7.54	GCTAATTTCTTTACTTGGTTCTTTGCTACC
AceCas9 PID library construction primers (5'→3')		
F	J.Pr7.55	GTGATAGCATTGCCGTGGCCG
R	J.Pr7.40	GCGGCCGTGTACAATACGATTAC
F	J.Pr7.62	GTAATCGTATTGTACACGGCCGC
R	J.Pr7.57	CGGCCACGGCAATGCTATCAC
PCSK9 & UBE3A PCR amplification and sequencing primers (5'→3')		
F	J.Pr8.21	GAGCCAGGCAGTGAGACTGGC
R	J.Pr8.22	CGCGAACCTTCCCACTGAATAGC
F	J.Pr7.65	GCTGGTCAGTTTTATCCCTTCAGGG
R	J.Pr7.66	GCACAACAAGCACAACAGAAAGCT

2.2. Methods

2.2.1. TwinStrep-tag cloning

To clone the TwinStrep tag on C-terminus of AceCas9, a HiFi assembly method was utilized, that is similar to Gibson assembly (Gibson et al., 2009) but uses enzymes with higher fidelity. Both TwinStrep-tag sequence and cloning vector were linearized in a PCR using Phusion™ Plus DNA Polymerase ("Thermo Fisher Scientific"). Plasmid J.Pl2.02 (Section 2.1.2) was paired with primers J.Pr6.67 and J.Pr6.68 (Section 2.1.3) to amplify TwinStrep-tag sequence. Similarly, J.Pl1.77 (Section 2.1.2) was paired with primers J.Pr6.80 and J.Pr6.66 (Section 2.1.3) to amplify cloning vector,

containing the sequence of human codon-optimized AceCas9. Both PCR amplicons were purified with DNA Clean & Concentrator-25 kit ("Zymo Research") and analyzed on a 1% (m/w) agarose gel containing 1X TBE buffer (100mM tris base, 100m boric acid, 2mM EDTA disodium salt dihydrate, pH 8.0 at 25°C) and 0.5X SYBR™ Safe DNA Gel Stain ("Thermo Fisher Scientific"). The linear fragments were then mixed with NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) at a molar ratio of 5:1 (0.5 pmol insert to 0.1 pmol vector) and incubated at 50 °C for 15 minutes. The assembly was then transformed into electrocompetent *E. coli* DH5α (Section 2.1.1) (homemade). The electroporation was done using ECM830 electroporator ("BTX Technologies") with the field strength set to 1.8kV/cm. The cells were plated on LB-agar media (10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, 10g/l bacteriological agar, pH 7.0 at 25°C) containing 50µg/ml of carbenicillin and grown overnight at 37°C. Transformants were screened through colony PCR using primers J.Pr6.67 and J.Pr6.68 (Section 2.1.3) and DreamTaq DNA polymerase ("Thermo Fisher Scientific"). Afterward, colonies of interest were inoculated to 4ml of LB broth (10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, pH 7.0 at 25°C), containing 50µg/ml of carbenicillin, and grown overnight at 37°C with 200 rpm^{-1} shaking. The plasmids were then purified with the Monarch® Plasmid Miniprep Kit ("New England Biolabs") and submitted to "SeqVision" for whole plasmid Nanopore sequencing. Sequencing results confirmed successful construction of plasmid J.Pl2.26 (Section 2.1.2) later used for transfection into human cells (Section 2.2.2).

2.2.2. Human cell culturing, transfection and genomic DNA extraction

HEK293T cells ("ATCC") were cultivated in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% of fetal bovine serum and 1% Pen-Strep antibiotic ("Thermo Fisher Scientific"). The cells were plated in a 24-well plate at a density of 1.4×10^5 cells per well. Following one day of growth, 200ng of AceCas9 and its gRNA encoding plasmid J.Pl2.26 (Section 2.1.2) was mixed with reduced serum medium (Opti-MEM) ("Thermo Fisher Scientific") and combined with Lipofectamine 3000 transfection reagent ("Thermo Fisher Scientific") according to the protocol of the manufacturer. The mixture was incubated for 15 min at room temperature and then added to each well. The cells were grown at 37 °C, 5% CO_2 and collected by extensive pipetting the surface of the wells with 1X phosphate-buffered saline (PBS) solution ("Thermo Fisher Scientific") after 27h, 43h and 50h, and pelleted by spinning at 600xg for 5 minutes. The cells were then mixed with a denaturing SDS-PAGE loading dye (50 mM Tris-HCl pH 6.8, 1% (V/V) β-mercaptoethanol, 20g/l SDS, 10% (V/V) glycerol, 2g/l bromophenol blue) and lysed by heating at 98°C for 5 min. The lysates were stored at -20°C until further use for western blot (Section 2.2.4).

To check for AceCas9's entry into the nucleus and editing in PCSK9 and UBE3A genes, the transfection was upscaled to a 6-well plate with 5.6×10^5 cells per well. HEK293T cells were

transfected with 800ng of J.Pl1.75 and J.Pl2.26 (Section 2.1.2) following the same procedure discussed before and grown at 37 °C and 5% CO_2 for 48 hours. The cells were collected by extensive pippeting the surface of the wells with 1X phosphate-buffered saline (PBS) solution ("Thermo Fisher Scientific") and pelleted by spinning at 600xg for 5 minutes. 1/3 of the J.Pl2.26-transfected cells were used for genomic DNA extraction with GeneJet Genomic DNA Purification Kit ("Thermo Fisher Scientific") while 2/3 of the cells were used for nuclear extraction (Section 2.2.3). All of the J.Pl.75-transfected cells were used for genomic DNA extraction.

The extracted genomic DNA was used as a template for PCR to amplify the PCSK9 and UBE3A regions. The PCR was carried out with Phusion™ Plus DNA Polymerase ("Thermo Fisher Scientific") with primers J.Pr7.65 and J.Pr7.66 for UBE3A gene and J.Pr8.21 and J.Pr8.22 for PCSK9 gene. (Section 2.1.3). The PCR amplicons were purified with DNA Clean & Concentrator-25 kit ("Zymo Research") and analyzed on a 1% (m/w) agarose gel containing 1X TBE buffer (100mM tris base, 100m boric acid, 2mM EDTA disodium salt dihydrate, pH 8.0 at 25°C) and 0.5X SYBR™ Safe DNA Gel Stain ("Thermo Fisher Scientific"). The samples were then submitted for Nanopore sequencing at "SeqVision" and Sanger sequencing at "Azenta Life Sciences" with primers J.Pr7.65 and J.Pr7.66 (for UBE3A) and J.Pr8.21 and J.Pr8.22 (for PCSK9) (Section 2.1.3). Nanopore sequencing reads were analyzed with "CRISPResso2" (Clement et al., 2019) while Sanger sequencing reads were analyzed with "Tracking of Indels By Decomposition (TIDE)" (Brinkman et al., 2014).

2.2.3. Nuclear extraction

Following plasmid transfection (Section 2.2.2), the collected HEK293T cells were washed twice with cold 1X phosphate-buffered saline (PBS) ("Thermo Fisher Scientific") and spinned at 600xg for 5 minutes, 4°C. The cell pellets were then resuspended in 1X hypotonic buffer (3mM $MgCl_2$, 10 mM NaCl, 20 mM Tris-HCl, pH 7.4 at 25°C) and incubated on ice for 15 minutes. Afterward, 10% NP-40 ("Thermo Fisher Scientific") was added and the mixture was vortexed for 10 seconds at highest setting to lyse the cells. The lysates were then centrifuged at 3000rpm^{-1} for 10 minutes at 4°C to separate cytoplasmic and nuclear fractions. The supernatant, containing the cytoplasmic fraction, was transferred to a clean microcentrifuge tube and stored at -20°C until further use. The remaining nuclear pellet was resuspended in Cell Extraction Buffer ("Thermo Fisher Scientific") and incubated on ice for 30 minutes, with vortexing every 10 minutes. Afterward, the nuclear fraction was peletted by centrifugation at 14000xg for 30 minutes at 4°C. The supernatant, containing the nuclear fraction, was then transferred to a fresh microcentrifuge tube and stored at -20°C until further use.

2.2.4. Western blot

Western blot was done following similar methods as previously described by (Towbin et al., 1979). The collected HEK293T cells (Section 2.2.2) and nuclear fractionation samples (Section 2.2.3) were mixed with a denaturing SDS-PAGE loading dye (50 mM Tris-HCl pH 6.8, 1% (V/V) β-mercaptoethanol, 20g/l SDS, 10% (V/V) glycerol, 2g/l bromophenol blue) and lysed by heating at 98°C for 5 min. SDS-denatured samples were resolved in 8% SDS-PAGE (AA/BAA ratio 37.1:1) gel ran in 1X SDS-PAGE buffer (25mM tris base, 19.2mM glycine, 3.5mM SDS, pH 8.0 at 25°C). Four pieces of Whatman filter paper and one piece of PVDF membrane were cut according to the size of the gel. The filter paper pieces were submerged into 1X transfer buffer (160mM tris base, 620mM glycine, 155mM tricine, 2.5mM EDTA, pH 8.0 at 25°C). The same was done to the PVDF membrane that was activated in 100% methanol prior. A protein "transfer sandwich" was created by stacking the components as follows: two sheets of wet filter paper, wet PVDF membrane, protein gel, two sheets of wet filter paper. The transfer was done for 15 minutes on Thermo Scientific™ Pierce™ Power Blotter ("Thermo Fisher Scientific") with the current set to 1.3A.

The PVDF membrane was blocked in blocking solution (40g/l skim milk powder, 137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 9mM KH_2PO_4, 0.2% (V/V) Tween 20) and incubated at room temperature with 1:4000 dilution of StrepII-tag Antibody HRP Conjugate ("Sigma-Aldrich") for 1 hour. The membrane was extensively washed in 1X wash solution (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 9mM KH_2PO_4, 0.2% (V/V) Tween 20)) and then exposed to the Thermo Scientific™ SuperSignal™ West Femto Maximum Sensitivity Substrate ("Thermo Fisher Scientific") to start the HRP chemiluminescence reaction. Signal detection was done using the "Bio-Rad Laboratories" ChemiDoc MP Imaging System.

2.2.5. Quick-change mutagenesis

The ccdB expression plasmid used for AceCas9 directed evolution in the ccdB survival assay (Section 2.2.7) contains AceCas9 target, flanked by a 5'-NNNCG-3' sequence. For positive control of the survival assay, one nucleotide in the PAM sequence was adjusted through quick-change mutagenesis (QCM) (H. Liu & Naismith, 2008), producing a plasmid J.Pl2.39 (Section 2.1.2) with the wild-type 5'-NNNCC-3' PAM sequence. J.Pl2.31 (Section 2.1.2) was used as a PCR template with overlapping primers J.Pr7.53 and J.Pr7.54 (Section 2.1.3). The QCM reactions were carried out with Phusion™ Plus DNA Polymerase ("Thermo Fisher Scientific"). To remove the PCR templates, QCM samples were digested with DpnI ("Thermo Fisher Scientific") for 1h at 37°C and incubated at 80°C for 5 min. QCM products were analyzed on a 1% (m/w) agarose gel containing 1X TBE buffer (100mM tris base, 100m boric acid, 2mM EDTA disodium salt dihydrate, pH 8.0 at 25°C) and 0.5X

SYBR™ Safe DNA Gel Stain ("Thermo Fisher Scientific"). Afterward, QCM products were transformed via electroporation (using "BTX Technologies" ECM830 electroporator with 1.8kV/cm field strength) into electrocompetent *E. Coli* DH5α (Section 2.1.1) cells (homemade). Transformants were grown overnight at 37°C on LB-agar media (10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, 10g/l bacteriological agar, pH 7.0 at 25°C), containing 50µg/ml of carbenicillin. A few colonies from the transformation were inoculated into 4ml of LB broth (10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, pH 7.0 at 25°C), containing 50µg/ml of carbenicillin, and grown overnight at 37°C with 200 rpm^{-1} shaking. Plasmid of interest (J.Pl2.39, section 2.1.2) was purified from grown cultures using Monarch® Plasmid Miniprep Kit ("New England Biolabs") and sequence-verified by Nanopore sequencing at "SeqVision".

2.2.6. AceCas9 library assembly

The AceCas9 PID (PAM interacting domain) library was cloned into pACYCDuet™-1 vector (J.Pl2.30, section 2.1.2) containing two multiple cloning sites for co-expression of AceCas9 and its gRNA. The AceCas9 PID library assembly was done in three cloning steps:

(1) HiFi cloning to insert the gRNA sequence, targeting the ccdB plasmid in the survival assay (Section 2.2.7):

Both AceCas9 gRNA sequence and cloning vector were PCR-amplified using Phusion™ Plus DNA Polymerase ("Thermo Fisher Scientific"). The gRNA insert was amplified from a dsDNA gene block (J.Pr7.45, Section 2.1.3) with primers J.Pr7.41 and J.Pr7.42 (Section 2.1.3). Similarly, the cloning vector J.Pl2.30 (Section 2.1.2) was linearized and PCR-amplified with primers J.Pr7.39 and J.Pr7.40 (Section 2.1.3). Both PCR amplicons were purified with DNA Clean & Concentrator-25 kit ("Zymo Research") and analyzed on a 1% (m/w) agarose gel containing 1X TBE buffer (100mM tris base, 100m boric acid, 2mM EDTA disodium salt dihydrate, pH 8.0 at 25°C) and 0.5X SYBR™ Safe DNA Gel Stain ("Thermo Fisher Scientific"). The linear fragments were then mixed with NEBuilder® HiFi DNA Assembly Master Mix ("New England Biolabs") and incubated at 50°C for 15 minutes. The assembly was then transformed into electrocompetent *E. coli* DH5α (Section 2.1.1) (homemade). The electroporation was done using ECM830 electroporator ("BTX Technologies") with the field strength set to 1.8kV/cm. The cells were plated on LB-agar media (10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, 10g/l bacteriological agar, pH 7.0 at 25°C) containing 35µg/ml of chloramphenicol and grown overnight at 37°C. Transformants were screened through colony PCR using primers J.Pr7.41 and J.Pr7.42 (Section 2.1.3) and DreamTaq DNA polymerase ("Thermo Fisher Scientific"). Afterward, colonies of interest were inoculated to 4ml of LB broth (10g/l tryptone, 5g/l

yeast extract, 10g/l NaCl, pH 7.0 at 25°C), containing 35µg/ml of chloramphenicol, and grown overnight at 37°C with 200 rpm^{-1} shaking. The plasmids were then purified with the Monarch® Plasmid Miniprep Kit ("New England Biolabs") and submitted to "SeqVision" for whole plasmid Nanopore sequencing. Sequencing results confirmed successful construction of intermediate co-expression plasmid J.Pl2.33 (Section 2.1.2).

(2) Restriction cloning to insert the wild-type AceCas9 sequence:

The J.Pl2.33 (constructed in the (1) step) was then used as a cloning vector for the sequence of the wild-type AceCas9. J.Pl2.37 (Section 2.1.2) was used as template in a PCR reaction, using Phusion™ Plus DNA Polymerase ("Thermo Fisher Scientific") and primers J.Pr7.43 and J.Pr7.44 (Section 2.1.3) to amplify the sequence of AceCas9 and introduce overhangs for restriction cloning. The PCR product was then analyzed on 1% (m/w) agarose gel containing 1X TBE buffer (100mM tris base, 100m boric acid, 2mM EDTA disodium salt dihydrate, pH 8.0 at 25°C) and 0.5X SYBR™ Safe DNA Gel Stain ("Thermo Fisher Scientific") and purified using DNA Clean & Concentrator-25 kit ("Zymo Research"). The purified cloning insert and the cloning vector (J.Pl2.33, section 2.1.2) were then digested with NcoI and SalI restriction enzymes ("Thermo Fisher Scientific"). The digested vector was gel-purified GeneJET™ Gel Extraction Kit ("Thermo Fisher Scientific") from 1% (m/w) agarose gel containing 1X TAE buffer buffer (40mM Tris-acetate, 1mM EDTA, pH 8.0 at 25°C). The cloning insert was purified with DNA Clean & Concentrator-25 kit ("Zymo Research"). Afterward, the digested vector and insert were ligated with T4 DNA ligase ("New England Biolabs") and transformed into homemade electrocompetent *E. coli* DH5α (Section 2.1.1) by following the same protocol described in the (1) step. The colonies of interest were selected through colony PCR with primers J.Pr7.44 and J.Pr7.55 (Section 2.1.3) and DreamTaq DNA polymerase ("Thermo Fisher Scientific"). Plasmid amplification and purification was done by following the methods described in the (1) step of the library construction. Whole plasmid Nanopore sequencing at "SeqVision" confirmed successful construction of wild-type AceCas9 and its gRNA co-expression plasmid J.Pl2.34 (Section 2.1.2).

(3) HiFi cloning to insert mutagenized PIDs of AceCas9:

The PID region of AceCas9 was amplified from the J.Pl2.34 plasmid using error-prone PCR (ePCR) with primers J.Pr7.55 and J.Pr7.40 (Section 2.1.3). The amplification was performed using the GeneMorph II Random Mutagenesis Kit ("Agilent Technologies"), introducing approximately 9–16 mutations per kilobase. This mutation rate corresponds to the use of 10 ng of template DNA, as specified in the polymerase protocol. J.Pl2.34 was also used as a template in another PCR reaction with primers J.Pr7.57 and J.Pr7.62 (Section 2.1.3) and Phusion™ Plus DNA Polymerase ("Thermo

Fisher Scientific") to amplify the cloning vector for the HiFi assembly. Both PCR products (AceCas9 PID fragment and cloning backbone) were then analyzed on 1% (m/w) agarose gel containing 1X TBE buffer (100mM tris base, 100m boric acid, 2mM EDTA disodium salt dihydrate, pH 8.0 at 25°C) and 0.5X SYBR™ Safe DNA Gel Stain ("Thermo Fisher Scientific") and purified using DNA Clean & Concentrator-25 kit ("Zymo Research"). The linear fragments were then mixed with NEBuilder® HiFi DNA Assembly Master Mix ("New England Biolabs") and incubated at 50°C for 15 minutes. The assembly was then transformed into electrocompetent *E. coli* TOP10 (Section 2.1.1) ("Thermo Fisher Scientific"), using ECM830 electroporator ("BTX Technologies") with the field strength set to 1.8kV/cm. To estimate library size, a 100-fold dilution of the transformed cells was plated on 100mm LB-agar media plate (10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, 10g/l bacteriological agar, pH 7.0 at 25°C), containing 35µg/ml of chloramphenicol. The remaining cells were equally divided within three 150mm plates containing the same growth media and incubated at 30°C for 20 hours. Afterward, the 150mm plates were washed with LB media (10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, pH 7.0 at 25°C) and scraped to collect the grown colonies. This wash and scrape step was repeated three times to ensure thorough recovery of all colonies. The collected cells were then pelleted by spinning at 3500 rcf for 10 minutes and the plasmid library was then purified using ZymoPURE II Plasmid Midiprep Kit ("Zymo Research") according to the protocol of the manufacturer.

2.2.7. ccdB survival assay

Homemade electrocompetent *E. coli* BW25141(λDE3) (Section 2.1.1) were transformed with ccdB-encoding plasmids J.Pl2.31 and J.Pl2.39 (Section 2.1.2) using ECM830 electroporator ("BTX Technologies") with the field strength set to 1.8kV/cm. The transformants were then plated on glucose and carbenicillin-containing plates (20mM D-(+)-glucose, 50µg/ml carbenicillin, 10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, 10g/l bacteriological agar, pH 7.0 at 25°C) and grown overnight at 37°C with 200 rpm^{-1} shaking. The produced two ccdB-harboring *E. coli* strains were then prepared as electrocompetent cells, as previously described by (Hand et al., 2019). The transformation efficiency of these strains was evaluated by transforming them with an empty pACYCDuet™-1 plasmid (J.Pl2.30, Section 2.1.2) that was used for AceCas9 PID library construction (Section 2.2.6).

For the first survival assay, 150µL of the of the electrocompetent ccdB-harboring *E. coli* BW25141(λDE3) cells were transformed with 30ng of the AceCas9 PID library (Section 2.2.5) or J.Pl2.34 (Section 2.1.2) for control. Due to volume constraints of the 1 mm gap cuvettes (max 90 µL), the transformation was performed in two separate electroporations (75µL x 2) using the ECM830 electroporator ("BTX Technologies") at a field strength of 1.8 kV/cm. Following electroporation, the

two samples were combined and resuspended in SOC media (20g/L tryptone, 5g/L yeast extract, 0.5g/L NaCl, 5g/L $MgSO_4$-$7H_2O$, 20 mM glucose, pH 6.9 at 25°C) and grown for 30 minutes at 37°C with 200 rpm^{-1} shaking. Afterward, 10µM of IPTG was added and returned to shaking for additional 60 minutes. 100-fold dilutions of the cells were plated on glucose and chloramphenicol-containing LB-agar plates (20mM D-(+)-glucose, 35µg/ml chloramphenicol, 10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, 10g/l bacteriological agar, pH 7.0 at 25°C) to evaluate the survival rate in the absence of ccdB expression. The remaining cells were plated on arabinose and chloramphenicol-containing LB-agar plates (10mM L-(+)-arabinose, 35µg/ml chloramphenicol, 10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, 10g/l bacteriological agar, pH 7.0 at 25°C). Both arabinose-containing and arabinose-lacking plates were incubated overnight at 37°C. Plates with colonies were photographed using the ChemiDoc Go Imaging System ("Bio-Rad Laboratories"), and colonies were counted using the OpenCFU software.

The second survival assay was upscaled by transforming 400µL of the ccdB-harboring *E. coli* cells with 80ng of either AceCas9 PID library (Section 2.2.5) or J.Pl2.34 (Section 2.1.2). The electroporation was done in 2mm gap cuvette with ECM830 electroporator ("BTX Technologies") and 1.5kV/cm field strength. The following cell recovery, IPTG induction and cell plating steps were done as described in the first survival assay.

3. Results and Discussion

3.1. Human cells express AceCas9

AceCas9 has demonstrated *in vitro* and *in vivo* activity (Tsui et al., 2017). However, its *in vivo* activity was studied in *E. coli* only. *In vivo* AceCas9 studies in *E. coli* identified its optimal gRNA sequence length and confirmed its DNA cleavage activity (Tsui et al., 2017). Later studies utilized *E.coli*-based assays to test mismatch-tolerant AceCas9 mutants (Hand et al., 2018), assess AceCas9 functional regions (Hand et al., 2019), determine AceCas9 PAM methylation sensitivity (Das et al., 2020) and produce catalytically enhanced AceCas9 variants (Hand et al., 2021).

AceCas9 methylation sensitivity holds potential in epigenetic editing approaches, however, its activity in human cells remains unknown. I aimed to establish if AceCas9 can be expressed in human cells. I began with a human expression vector, kindly gifted by the research group of Dr. Hong Li from Florida State University (**Figure 3.1**). The vector encodes a human codon-optimized AceCas9, its gRNA and essential sequences for AceCas9 expression, nuclear import, and subsequent genome editing. To detect AceCas9 expression in human cells, I linearized this vector and combined it with a linear TwinStrep-tag insert in a HiFi cloning reaction (Section 2.2.1). This produced a new TwinStrep-tagged AceCas9 expression vector, useful for AceCas9 detection via Western blotting.

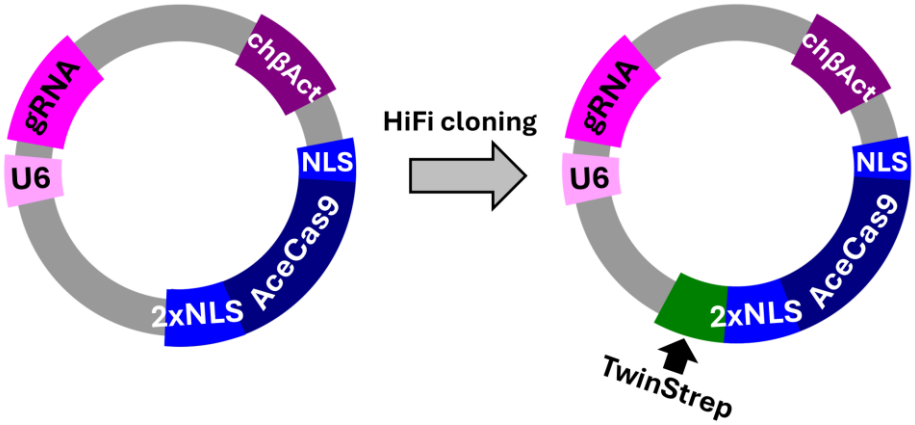

Figure 3.1 Construction of TwinStrep-tagged vector to detect AceCas9 expression in human cells. The initial vector encodes human codon-optimized AceCas9 with nuclear localization signals (NLS) under a chicken β-actin (chβAct) promoter, and gRNA (targeting UBE3A gene) under a U6 promoter. The added TwinStrep-tag enables AceCas9 detection via Western blotting.

After cloning, I used the AceCas9-TwinStrep vector for lipofection into HEK293T cells, a commonly used human embryonic kidney cell line, widely used in CRISPR-Cas research (Boyle et al., 2017; S. Lin et al., 2015) (Section 2.2.2, **Figure 3.2** A). Protein expression in HEK293T cells is reported to peak after around 24 hours following the lipofection (C.-Y. Lin et al., 2015). Transfected plasmid DNA must first enter the nucleus and undergo transcription, after which the resulting mRNA is exported to the cytoplasm for translation. Therefore, I collected transfected cell samples after 27h, 43h, 50h for comparison, as well as 0h for negative control. I then detected AceCas9 expression in

HEK293T cells at different time points using the TwinStrep-tag and Western blot analysis (**Figure 3.2** B, Section 2.2.4). For detection, I employed an anti-Strep-tag II antibody conjugated to horseradish peroxidase (aStrepII-HRP), which binds the TwinStrep-tag (TwinStrep-tag and Strep-tag II share structural similarities, as the TwinStrep-tag is composed of two Strep-tag II sequences connected by a flexible linker) (Schmidt et al., 2013). This western blot strategy produced a chemiluminescence signal corresponding to the size of AceCas9 (~135kDa), confirming that this nuclease can be expressed in HEK293T cells. For negative control, I collected untransfected cells across the same time points and performed the same western blot procedure, which produced no detectable bands.

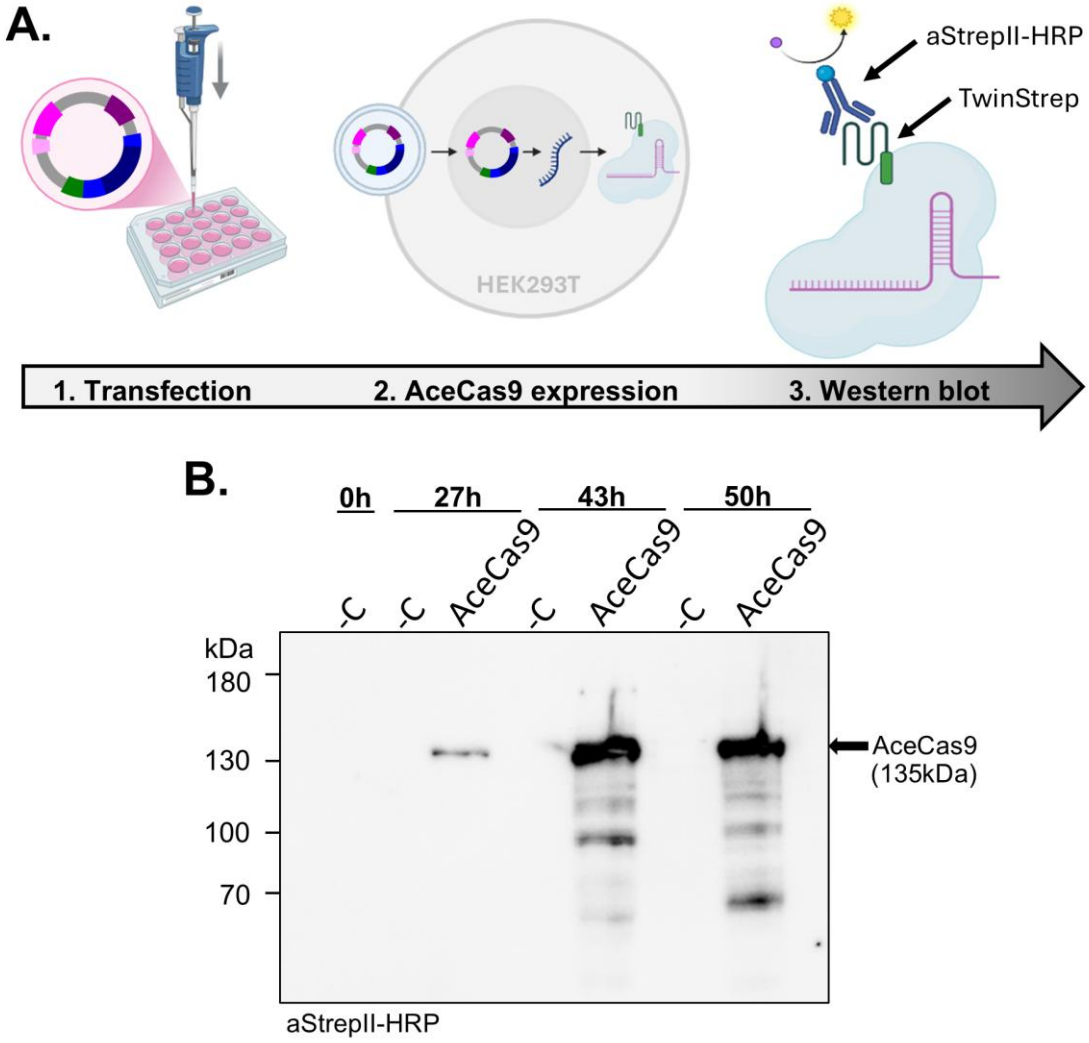

Figure 3.2 AceCas9 expression in HEK293T cells**. (A)** Workflow to assess AceCas9 expression in HEK293T cells. The HEK293T cells were transfected with TwinStrep-tagged AceCas9 expression vector. Following expression, the TwinStrep tag is identified in Western blot with an aStrepII-HRP antibody, showing the presence of AceCas9. **(B)** Western blot, demonstrating expressed AceCas9 (~135kDa) at the 27, 43 and 50 hours post-transfection. Additional bands indicate possible AceCas9 degradation, likely due to issues with thermostability and cellular processing. -C – negative control (untransfected cells), aStrepII-HRP – anti-StrepII-tag conjugated with an HRP (horseradish peroxidase), used for chemiluminescent detection. Protein size standards are indicated on the left. Illustrations made in BioRender.

Additional bands on the blot could result from AceCas9 degradation from its N-terminus, as the TwinStrep-tag is attached to the C-terminus. This is unsurprising, since AceCas9 comes from a thermophilic bacteria, which grows best around 55-60°C and in an acidic pH of ~5.0 (Mohagheghi et al., 1986). This expression experiment was carried out at physiological conditions for humans (37°C, pH 7-7.4), which might have affected the thermostability of AceCas9 or caused inefficient folding that could trigger AceCas9 degradation by cellular proteasomes (Tsai et al., 2024). Although human cells can express AceCas9, this does not guarantee that the protein is functionally active. AceCas9 activity should be evaluated by checking its genome editing capability in human cells, as described below.

3.2. AceCas9 enters the nucleus

In prokaryotic systems, the CRISPR-Cas nucleases are expressed and function in the same cellular compartment, since transcription, translation, and DNA interference all take place in cytoplasm. When applying CRISPR-Cas as genome editors in a eukaryotic context, the involved processes are compartmentalized: transcription of the Cas protein (and its gRNA) occurs in the nucleus, while its expression takes place in the cytoplasm. To enable genome editing, the expressed CRISPR-Cas nuclease must be transported into the nucleus, where the genomic DNA is located. The nuclear envelope is crossed with the help of nuclear importins, which recognize short peptide sequences known as nuclear localization signals (NLS), allowing proteins to enter the nucleus (J. Lu et al., 2021). CRISPR-Cas nucleases, originating from prokaryotic environments, lack NLS sequences. As a result, NLS sequences are added through genetic engineering of CRISPR-Cas expression vectors to enable their import into the nuclei of eukaryotic cells.

All AceCas9 human expression vectors used in this study include both N-terminal and C-terminal NLS sequences. Before evaluating AceCas9's genome editing efficiency in HEK293T cells, I aimed to assess whether these NLS sequences enabled AceCas9's entry into the nucleus. Using the same AceCas9 expression vector as before (Section 3.1), I performed plasmid lipofection into HEK293T cells. I separated nuclear and cytoplasmic fractions (Section 2.2.3) and then analyzed the samples by Western blot (Section 2.2.4). A band with a size corresponding to AceCas9 appears in both cytoplasmic and nuclear fractions, with a stronger signal observed in the nuclear fraction. (**Figure 3.3**).

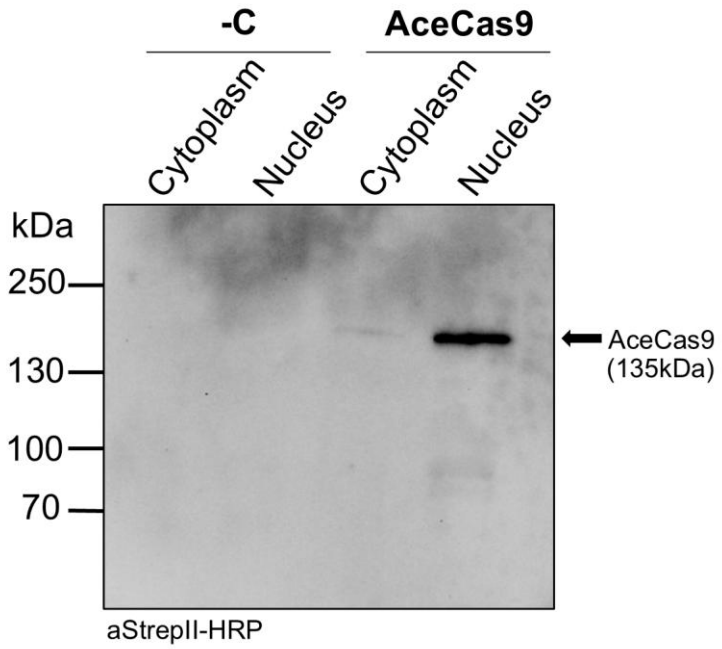

Figure 3.3 Western blot analysis of nuclear and cytoplasmic fractions from HEK293T cells expressing NLS-tagged and TwinStrep-tagged AceCas9. The ~135kDa band observed in the nuclear fraction suggests that the attached NLS sequences are functional and allow AceCas9's entry into the nucleus for possible genome editing. A faint band in the cytoplasmic fraction suggests partial cytoplasmic retention. NLS – nuclear localization signal, -C – negative control (untransfected cells), aStrepII-HRP – anti-StrepII-tag conjugated with an HRP (horseradish peroxidase), used for chemiluminescent detection. Protein size standards are indicated on the left.

This experiment provides preliminary, qualitative insight into nuclear localization of AceCas9. For more accurate quantitative analysis, this blot should be improved by including markers specific to the cytoplasmic and nuclear compartments. Nevertheless, these preliminary results suggest that AceCas9 can enter the nucleus, where it could potentially initiate genome editing.

3.3. AceCas9 does not generate detectable edits in human cells

Following confirmation of AceCas9 expression and nuclear entry into human cells, I proceeded to evaluate its potential genome editing activity. HEK293T cells were transfected with plasmids expressing AceCas9 and gRNAs that target UBE3A or PCSK9 genes (Section 2.2.2). These plasmids were generously provided by Dr. Hong Li (Florida State University). UBE3A and PCSK9 are common target genes for CRISPR-Cas nucleases due to their relevance in medical research. UBE3A is associated with Angelman syndrome, a neurodevelopmental disorder, making it an important target for potential therapeutic interventions (Wolter et al., 2020). PCSK9 plays a key role in cholesterol regulation and is important in the context of cardiovascular diseases (Musunuru et al., 2021).

After the transfection of the UBE3A and PCSK9-targeting plasmids into HEK293T cells, I extracted their genomic DNA (gDNA) and PCR-amplified the potentially edited regions (**Figure 3.4**)

(Section 2.2.2). Each editing experiment was performed in three replicates for target genes to assess variability in editing efficiency. Untransfected cells served as a negative control. The sequences of these PCR amplicons were then verified by either Sanger or Nanopore sequencing. Although Sanger sequencing offers high accuracy, it provides only single-read outputs per reaction. Nanopore yields multiple reads per sample and offers greater sequencing depth, enabling a more comprehensive assessment of editing efficiency but with lower base-calling accuracy.

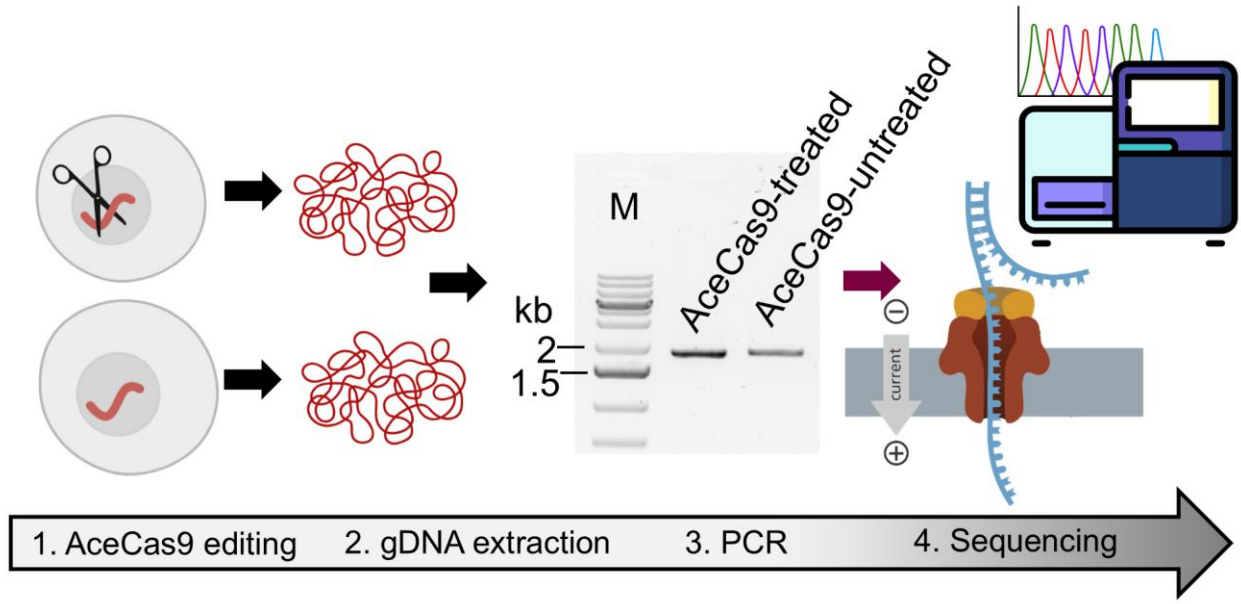

Figure 3.4 Workflow for AceCas9-treated versus untreated HEK293T cells to assess AceCas9 genome editing efficiency. Following AceCas9 treatment, genomic DNA (gDNA) was extracted from HEK293T cells and used as a PCR template to amplify the targeted sequences. PCR amplicons were submitted to Sanger and Nanopore sequencing. AceCas9-untreated samples were prepared as negative controls. M – DNA size marker. Illustrations made in BioRender.

For Sanger sequencing, each PCR product was submitted with two sequencing primers – one forward and one reverse – covering the target site from both directions. Although Sanger sequencing requires only one primer per reaction, I did independent sequencing from both directions to detect AceCas9 editing events more precisely. Sanger sequencing results were then analyzed with the TIDE (Tracking of Indels by DEcomposition) tool (Brinkman et al., 2014). TIDE analyzes Sanger sequencing data of edited and control DNA samples to determine the frequency and spectrum of small insertions and deletions (indels) at a target site. Due to poor-quality forward reads for PCSK9, only the reverse sequencing reads were used in TIDE analysis (**Figure 3.5**). They showed an average editing efficiency of $1.77 \pm 0.35\%$ across three experimental replicates. Forward and reverse reads were analyzed for UBE3A: forward reads indicated an average editing efficiency of $2.10 \pm 1.65\%$, while reverse reads indicated a much lower average of $0.27 \pm 0.15\%$. These values reflect modest editing efficiencies and highlight variability between sequencing directions. All TIDE-analyzed editing efficiencies had p-values equal or above 0.001, indicating a lack of strong statistical support

for most indel calls. In one notable data point, AceCas9 exhibited 4% editing efficiency in UBE3A with 2.4% of the indels detected with a p-value below 0.001. This suggests that a fraction of editing events in this replicate may represent AceCas9 activity.

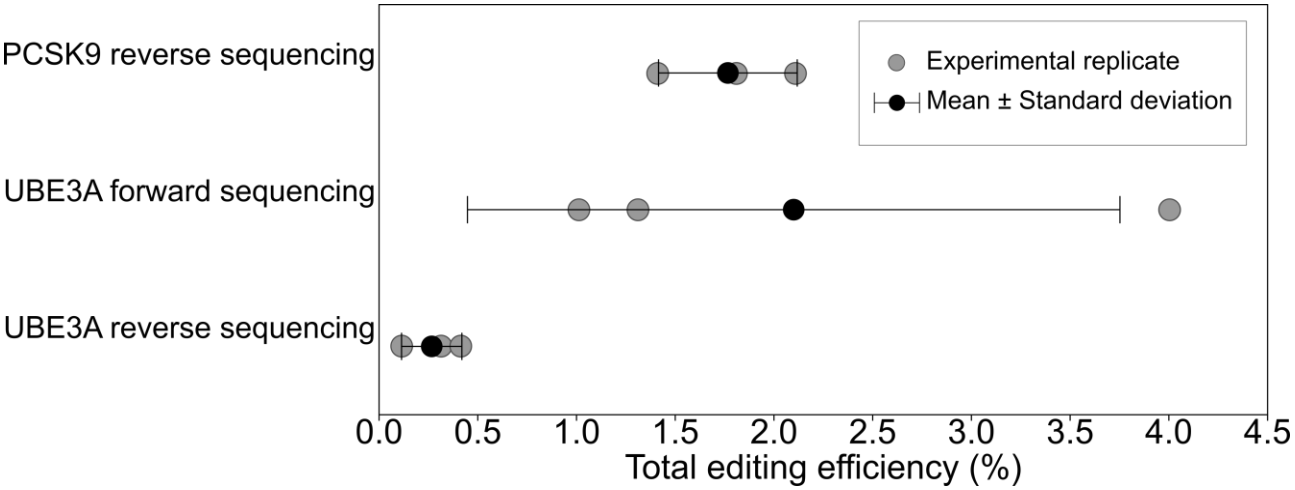

Figure 3.5 Summary of AceCas9 editing efficiencies at PCSK9 and UBE3A loci based on Sanger sequencing direction. Sanger sequencing was performed in both forward and reverse directions to generate independent datasets and improve detection of potential editing events. Total editing efficiencies were calculated using TIDE analysis of Sanger sequencing data and plotted separately for each sequencing read direction.

To more comprehensively analyze AceCas9 editing efficiency, I used Nanopore sequencing on the PCR amplicons. The reads were analyzed using the CRISPResso2 tool, which assesses genome editing outcomes from high-throughput sequencing data (Clement et al., 2019). CRISPResso2 aligns sequencing reads to a reference amplicon (unedited DNA) and identifies indels around the cut site. CRISPREsso2 also reports the proportion of sequencing reads classified as "modified" versus "unmodified", estimating overall editing efficiency. The percentage of modified sequences detected in unedited control samples was very similar to the percentage in the AceCas9-treated samples across all experimental replicates (**Figure 3.6**). For PCSK9, the AceCas9-treated samples showed 2.24 ± 0.23% modified sequences, compared to 2.39% in unedited control. For UBE3A, the AceCas9-treated samples showed 3.65 ± 0.15% modified sequences, while the unedited control exhibited 4.01%. As a result, I could not determine whether AceCas9 exhibits genome editing activity from Nanopore sequencing data. The lack of distinction between edited and unedited samples may be due to Nanopore sequencing errors and/or limited resolution of the experimental setup. Thus, if AceCas9 edits DNA with low efficiency, its levels are below our thresholds of detection.

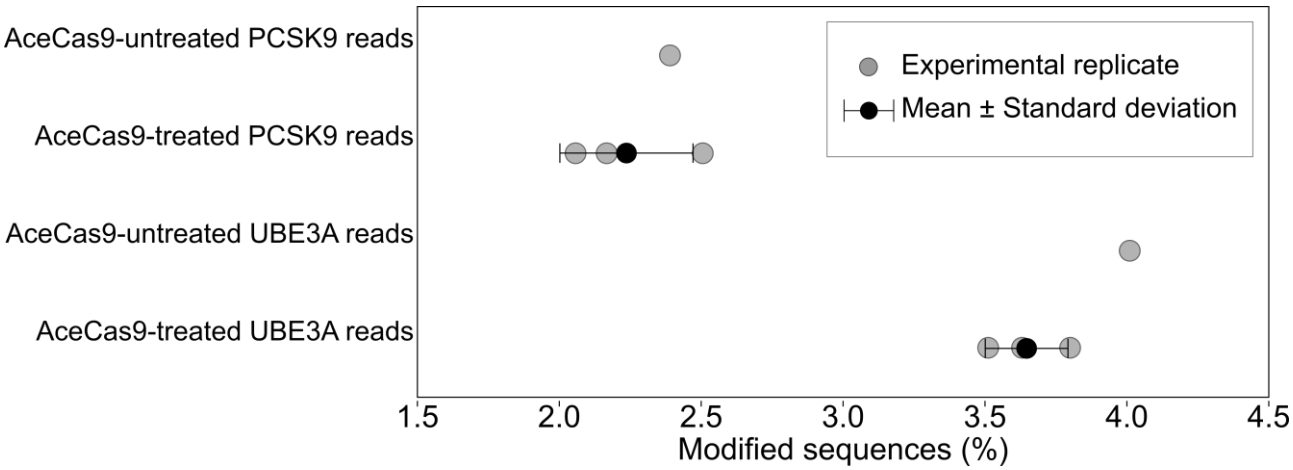

Figure 3.6 Comparison of modified sequence frequencies in AceCas9-treated and AceCas9-untreated (control) samples based on CRISPResso2 analysis of Nanopore sequencing data. CRISPResso2 identifies indels by categorizing "modified" or "unmodified" reads relative to a reference amplicon.

Although Sanger sequencing data suggested that AceCas9 may exhibit low editing efficiency at PCSK9 and UBE3A loci, Nanopore sequencing results failed to confirm this, as the percentage of modified sequences in AceCas9-treated samples was similar to unedited controls. Therefore, AceCas9 editing efficiency results remain ambiguous and inconclusive. DNA accessibility and chromatin state of the targeted genomic regions should also be considered. For example, PCSK9 is highly expressed in liver cells but shows low expression in non-liver cell types such as HEK293T (Saitoski et al., 2022). This low expression in HEK293T cells suggests that the gene is likely less accessible to genome editors, potentially limiting editing efficiency at this locus. In contrast, UBE3A is constitutively expressed in HEK293T cells, making it more likely to be accessed by the genome editor (Vihma et al., 2024). However, relying on a single constitutively expressed target gene is insufficient to evaluate whether AceCas9 is functional in human cells. Consequently, future studies should consider targeting alternative, more accessible genes (e.g. DNMT1, ACTB, HPRT) to better understand if AceCas9 operates in human cells.

3.4. Engineering AceCas9 for CpG recognition

The wild-type AceCas9 recognizes a 5'-NNNCC-3' PAM sequence with sensitivity to 5mCpC DNA modification. Since 5mCpC is rare in the human genome compared to the more prevalent 5mCpG modification, we aim to engineer AceCas9 to recognize a 5'-NNNCG-3' PAM, making it more suitable for use in human cells. Altering PAM specificity requires protein engineering, which can be achieved through either a rational design approach or directed evolution, as discussed in the literature review (Section 1.4). In this work, I combined both approaches by rationally altering the PAM interacting domain (PID) of AceCas9, where I introduced mutations in a randomized manner.

To select 5'-NNNCG-3' PAM recognizing AceCas9 variants, we employed a directed evolution method that was first introduced by (Z. Chen & Zhao, 2005) to engineer homing endonucleases. This strategy is based on the survival of *E. coli* cells carrying a plasmid that expresses the DNA gyrase toxin ccdB (control of cell death protein B), which blocks DNA replication and leads to cell death (Bernard et al., 1993). These cells are transformed with another plasmid encoding the engineered variant of the nuclease of interest. Their survival depends on the activity of this nuclease: if the ccdB plasmid is cleaved by the nuclease, the cells can survive on arabinose-containing plates. However, if the ccdB plasmid remains intact, the toxic ccdB protein is expressed upon plating on arabinose, causing the cells with non-functional mutants to die (**Figure 3.7**).

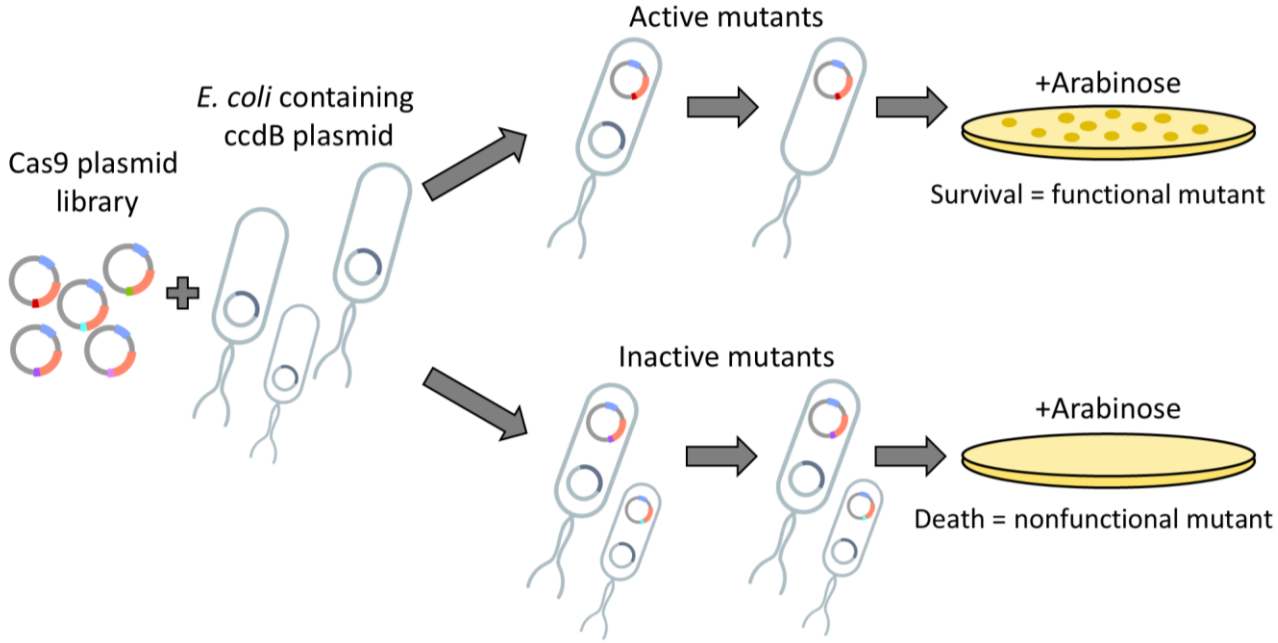

Figure 3.7 ccdB survival assay for directed evolution of Cas9. A Cas9 library is transformed into *E. coli*, harboring a plasmid that expresses the toxic ccdB gene under an arabinose-inducible promoter. Active mutants which cleave the ccdB plasmid enable cell survival on arabinose-containing agar plates. In contrast, incomplete or failed cleavage of the ccdB plasmid leads to cell death in the presence of arabinose. This selection can be performed in iterative rounds to enrich for increasingly active or specific Cas9 variants.

This strategy was utilized for directed evolution studies of AceCas9 to identify functional regions of AceCas9 (Hand et al., 2019) and to generate catalytically enhanced AceCas9 variants (Hand et al., 2021). I used it to identify AceCas9 variants capable of cleaving a DNA target flanked by a 5'-NNNCG-3' PAM sequence. Our initial goal was to alter the PAM specificity of AceCas9 without taking its methylation sensitivity into account. If we succeeded in identifying variants that recognize the 5'-NNNCG-3' PAM, we would proceed to assess their sensitivity to 5mCpG methylation in future experiments or perform further directed evolution to engineer variants with methylation sensitivity.

3.4.1. AceCas9 PID library construction

For the AceCas9 directed evolution assay, I began with library construction of AceCas9 variants with mutations in the PAM interacting domain (PID) region. I first cloned a gRNA sequence containing a spacer sequence targeting the ccdB plasmid (from the previously described survival assay in section 3.4). I cloned the wild-type sequence of AceCas9 into the same plasmid, producing a co-expression vector for AceCas9 and its gRNA (Section 2.2.6). I then used the wild-type AceCas9 sequence as a template in an error-prone PCR (ePCR) to introduce mutations into the PID, generating variants with mutation rates ranging approximately from 9 to 16 nucleotide substitutions per PID sequence (**Figure 3.8**). The mutagenized PID fragments were then cloned into the AceCas9 and its gRNA co-expression plasmid using a HiFi assembly reaction, completing the construction of the AceCas9 PID mutant library (Section 2.2.6). Based on the mutation rate of ePCR and the redundancy of the genetic code, we estimate that each library member contains approximately 6 to 12 amino acid substitutions within the 306-amino-acid PID region of AceCas9.

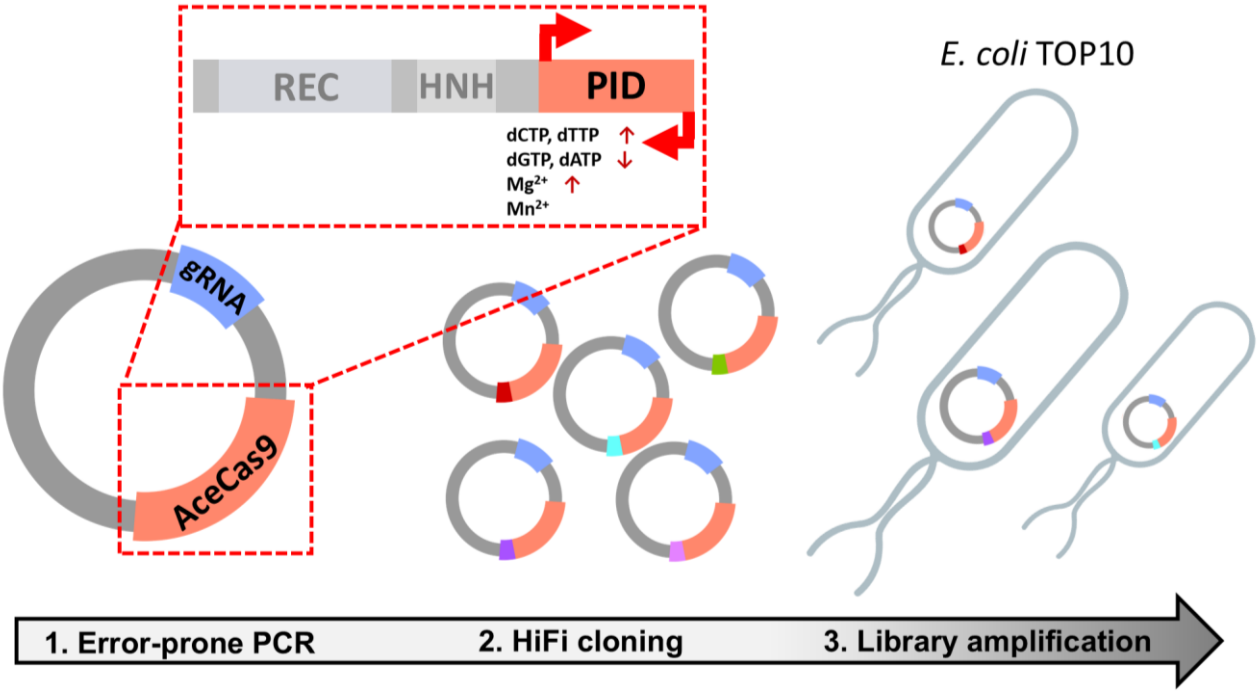

Figure 3.8 Workflow of AceCas9 variant library construction. A co-expression plasmid of wild-type AceCas9 and its gRNA was used as a template for an error-prone PCR (ePCR) reaction to introduce random mutations in the PAM interacting domain (PID) of AceCas9. The mutated PID fragments were cloned back into the same co-expression vector to produce an AceCas9 PID library. Following library construction, it was transformed into *E. coli* TOP10 cells for amplification and subsequent purification.

The assembled AceCas9 PID mutant library was transformed and amplified in *E. coli* using a high-throughput method (Section 2.2.6). We used the amount of colony forming units (CFUs) to estimate the theoretical library size. A 100-fold dilution of the transformed culture was plated,

yielding 116 CFUs (**Figure 3.9**). Based on this, we estimate that the library contains approximately 11600 members.

Figure 3.9 Agar plate containing 100-fold dilution of *E. coli* TOP10 cells transformed with the assembled AceCas9 PID library. The undiluted transformation is estimated to contain approximately 11600 variants. PID – PAM interacting domain.

However, these members might not be unique, as both PCR amplification of the PID sequence and plasmid transformation are subject to sequence-dependent biases (Kebschull & Zador, 2015). Consequently, the actual diversity of the library is likely lower than the theoretical maximum of 11600 variants, due to potential redundancies and similarities among members. Regardless, we use this theoretical library size as a rough benchmark for estimating the number of variants that we need to sample during selection experiments.

3.4.2. Survival assay

After the AceCas9 PID variant library construction, I proceeded with a survival-based selection of AceCas9 mutants that could recognize a 5'-NNNCG-3' (hereafter – CG) PAM sequence. For this assay, I transformed *E. coli* BW25141 (λDE3) cells with two different ccdB-expressing plasmids, generating two distinct strains (Section 2.2.7). Both strains harbor ccdB plasmids that encode identical target sequences, corresponding to the spacer region of AceCas9's guide RNA. The only difference is the PAM sequence: one plasmid contains the test PAM (CG) while the other contains the wild-type PAM 5'-NNNCC-3' (hereafter - CC) and serves as a control.

I prepared the constructed ccdB-harboring strains as electrocompetent cells and then evaluated their ability to take up plasmid DNA by testing their transformation efficiency. The cells were transformed with an empty pACYCDuet-1 vector (starting vector for our AceCas9 PID library). Both CG and CC PAM strains yielded transformation efficiency of 10^6 CFU/µg of DNA, which is on the lower end of the typical range (10^7-10^9 CFU/µg) for electrocompetent *E. coli*. Understanding the transformation efficiency is important as it can limit the total number of unique AceCas9 PID library variants that can be screened in a single survival assay experiment.

The survival assay was performed with three experimental setups that had the following goals:

1. WT AceCas9 + CC PAM: to validate the survival assay.
2. AceCas9 PID library + CC PAM: to assess whether AceCas9 mutants can still recognize the WT PAM.
3. AceCas9 PID library + CG PAM: to select functional mutants, capable of recognizing the CG PAM sequence.

The survival assay is based on *E. coli* survival in the presence of arabinose, which induces the expression of the plasmid-encoded ccdB toxin. To estimate the survival rates, I compared the amount of colony forming units (CFUs) on arabinose-containing (+Ara) versus arabinose-lacking (-Ara) plates. The cells plated on -Ara plates serve as a control, representing the number of colonies that would survive in the absence of ccdB expression. The survival rates (SR) were calculated by the (1) formula:

$$SR = \frac{+Ara\ CFUs}{-Ara\ CFUs\ \times\ -Ara\ plate\ dilution\ factor} \times 100\% \qquad (1)$$

To cover the estimated AceCas9 PID variant library size (~11600 variants), I performed two rounds of the ccdB survival assay (**Figure 3.10**). As expected, both experiments yielded the highest SR under wild-type conditions (28% and 73%). The reported SR under wild-type conditions was 75% ± 6% (Tsui et al., 2017). Although we would expect 100% of survival under WT conditions, the variation of the SR can be attributed to factors such as transformation efficiency, timing of AceCas9 and its gRNA expression or basal expression of ccdB, which may cause premature cell death. Additionally, ccdB expression can affect cell growth, leading to observed smaller colony sizes on +Ara plates compared to -Ara plates (**Figure 3.10**). Both survival assays yielded survival rates of 5% (**Figure 3.10 A**) and 2% (**Figure 3.10 B**) in the conditions where I tested the AceCas9 PID library against a wild-type PAM. These results suggest that some of the AceCas9 mutants retained CC PAM specificity, and that any mutations introduced in the PID region did not affect their ability to recognize the wild-type PAM.

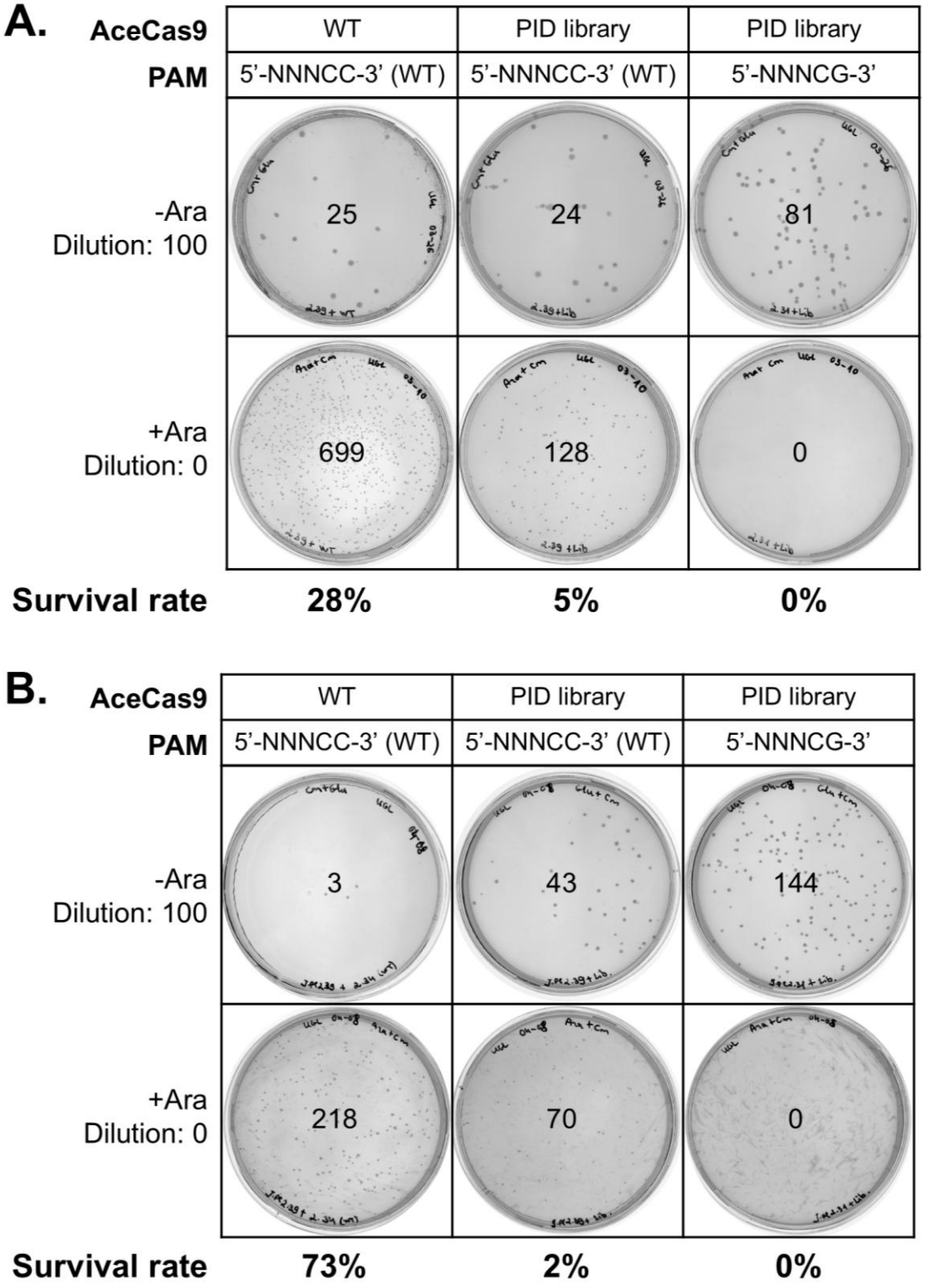

Figure 3.10 ccdB survival assay for directed evolution of AceCas9. This assay was performed to identify AceCas9 variants with mutations in the PAM-interacting domain (PID) that enable recognition and cleavage of targets flanked by a 5'-NNNCG-3' PAM sequence. The assay included controls with the wild-type (WT) PAM 5'-NNNCC-3' to validate the survival assay and determine if the PID mutants are still able to recognize the WT PAM. The arabinose-lacking (-Ara) plates contain a 100-fold dilution of the transformation to represent the amount of colony forming units (CFUs) that would be present in the absence of arabinose. Numbers on the agar plates show counted CFUs that were used for calculating survival rates **(A)** First survival assay, covering around 8100 AceCas9 PID library members against a CG PAM. **(B)** Second survival assay, theoretically covering the whole AceCas9 PID library (11600 variants) against a CG PAM.

The first survival assay experiment yielded 0% survival rate when testing the AceCas9 PID library against the CG PAM (**Figure 3.10 A**). This experiment resulted in 81 CFUs on the 100-fold dilution -Ara plate and covered approximately 8100 variants (assuming that each CFU on this plate represents a unique library member). Since this number is lower than the theoretical size of the whole library (~11600), I conducted another round of the survival assay which resulted in 144 CFUs on the 100-fold dilution -Ara plate, theoretically covering the whole AceCas9 PID library (**Figure 3.10 B**). Nevertheless, both experiments had no surviving colonies on the +Ara plates, concluding that none of the tested 11600 variants were able to recognize a CG PAM.

3.5. Discussion

Although CRISPR-Cas9-based epigenetic editing platforms were successfully utilized for various epigenetic applications, they still face limitations such off-target binding, which may lead to unintended epigenetic changes. However, the methylation sensitivity of AceCas9 offers a potential solution by distinguishing between methylated and unmethylated DNA, thereby increasing the specificity of epigenetic targeting and reducing off-target effects. AceCas9's methylation sensitivity is related to methylation of the first cytosine in the 5'-NNNCC-3' PAM sequence. CpC methylation sensitivity could be partially useful for detecting epigenetic changes in human cells, containing 5mCpC DNA modification. For example, this modification is present in some embryonic stem cells or neurons (Section 1.1.2). Nevertheless, AceCas9 applications in human epigenetic editing are limited to a small subset of genomic sites, as most of the methylated cytosines are found in CpG sites. To broaden its potential use, this study aimed to determine if AceCas9 is functional in human cells and engineer it to recognize a 5'-NNNCG-3' PAM sequence.

AceCas9 expression experiments showed that this nuclease can be expressed in HEK293T cells. Additionally, nuclear fractionation experiment of HEK293T cells demonstrated that AceCas9 is present in the nucleus where it can potentially initiate genome editing. However, genome editing experiments with AceCas9 were inconclusive and could not confirm if this nuclease functions in human cells. Sanger sequencing data suggested that AceCas9 may edit PCSK9 and UBE3A with low efficiency (<2.1% and <4%, respectively). Generally, this editing efficiency is considered low, as SpCas9 can reach indel efficiencies of 20-60% (H. Yang et al., 2013). This difference may be partially explained by the evolutionary origins of these nucleases. AceCas9 is derived from a type II-C CRISPR-Cas system, while SpCas9 originates from a typically more efficient type II-A CRISPR-Cas system (section 1.2.2). However, Nanopore sequencing results were inconclusive in confirming whether AceCas9 exhibited any editing activity in human cells, since edits and sequencing errors could not be distinguished from one another. As a result, we could neither confirm nor rule out

AceCas9-mediated genome editing. To address this, future experiments could include targeting alternative genes (e.g. DNMT1, ACTB, HPRT) and utilization of alternative validation methods such as Illumina sequencing.

AceCas9 activity has been tested at elevated temperatures in the thermophile *Clostridium thermocellum (C. thermocellum)* (Maeder et al., 2019). Since SpCas9 is not active in temperatures above 42 °C (Harrington et al., 2017), this study used thermophilic Cas9 variants (AceCas9, *Geobacillus stearothermophilus* Cas9 (GeoCas9) and *Geobacillus thermodenitrificans* T12 Cas9 (ThermoCas9)) to edit the genome of *C. thermocellum* for improved biofuel processing (Maeder et al., 2019). In the activity assay conducted in that study, ThermoCas9 showed no activity (0%), AceCas9 exhibited moderate activity (37%), and GeoCas9 achieved full activity (100%). The researchers suggested that the lower efficiency of AceCas9 may be due to protein misfolding or ineffective gRNA design. Given that AceCas9 struggles to function in thermophilic environments, it is perhaps not surprising that it shows little to no detectable editing efficiency in human cells. However, it is possible to adapt thermophilic enzymes for function in mammalian systems, offering increased protein stability and resistance to thermal degradation. For example, GeoCas9 was demonstrated to induce similar levels of indels in HEK293T cells as SpCas9 (Harrington et al., 2017). Additionally, directed evolution of GeoCas9 yielded an improved variant, iGeoCas9, exhibiting >100-fold higher genome-editing efficiency in human cells compared to wild-type GeoCas9 (Eggers et al., 2024).

Our directed evolution experiments aimed to evolve AceCas9 to recognize a 5’-NNNCG-3’ PAM sequence. The constructed AceCas9 PID mutant library contained approximately 11600 members (transformation-efficiency limited), assuming there were no redundant sequences. However, no variants in the library successfully cleaved the target with the 5’-NNNCG-3’ PAM. Interestingly, some mutants still recognized the wild-type PAM sequence (5’-NNNCC-3’), suggesting that some mutations in the PID of AceCas9 did not significantly affect its ability to recognize the wild-type PAM. Further experiments could be optimized by altering the mutation rates within the PID of AceCas9. Additionally, the AceCas9 libraries from previous directed evolution studies typically contained 10^4–10^7 members (Hand et al., 2019), therefore the library size of AceCas9 PID mutants should be increased. Protein2PAM, a recent deep learning model, predicts PAM specificity from Cas protein sequences without requiring structural data (Nayfach et al., 2025). Protein2PAM was trained with over 45000 PAM sequences of CRISPR-Cas nucleases, including AceCas9. Protein2PAM identified AceCas9 as highly engineerable through single amino-acid substitutions in the PID, especially compared to the widely used SpCas9. These findings highlight the potential for engineering AceCas9 to recognize a 5’-NNNCG-3’ PAM, as its PID appears to be flexible and amenable to modification.

An alternative approach to developing CRISPR-Cas tools for human epigenome editing is to engineer existing Cas9 variants to recognize PAM sequences that include a 5'-CG'-3' motif. A 5'-CG-3' PAM is recognized by SpCas9-NG, mentioned in (section 1.4.2; Nishimasu et al., 2018). SpCas9-NG maintains comparable target specificity and genome-editing efficiency to wild-type SpCas9 in human cells, making it a strong candidate for potential applications in epigenetic editing (Nishimasu et al., 2018). To enable methylation sensitivity, SpCas9-NG could be further engineered to become methylation-sensitive through either rational design or directed evolution. Another promising strategy would be to create a chimeric Cas9 enzyme by fusing domains from AceCas9 and SpCas9. Cas9 chimeras are engineered proteins combining domains from different Cas9 homologs to enhance or alter their properties such as PAM recognition, editing efficiency or specificity (Liang et al., 2025; L. Zhao et al., 2023). For example, integrating the PAM-interacting domain (PID) of AceCas9 into the SpCas9 backbone could hypothetically yield a hybrid enzyme with methylation sensitivity and retained functionality in human cells. The structural compatibility of this chimera could first be assessed using AlphaFold protein structure prediction models (Abramson et al., 2024), eliminating the need for initial biochemical or cell-based assays. This chimeric protein would potentially be sensitive to distinguish 5mCpC modifications; further studies would be needed to evolve the enzyme for 5mCpG detection.

To this day, AceCas9 is the only known methylation-sensitive CRISPR-Cas nuclease. However, CRISPR-Cas nucleases are not typically tested for methylation sensitivity, therefore additional methylation-sensitive variants may exist but remain uncharacterized. Given that CRISPR-Cas systems and restriction-modification (RM) systems share functional similarities and often coexist in prokaryotic defense strategies (Oliveira et al., 2014), it is plausible that some Cas nucleases may have evolved to recognize DNA methylation states. This overlooked relationship suggests that other methylation-sensitive CRISPR-Cas nucleases could be discovered. A promising starting point for identifying such variants would be testing close homologs of AceCas9, as these are likely to share similar PAM recognition mechanisms and structural features. This would provide deeper insights into how DNA modifications influence targeting and editing properties of CRISPR-based systems and could offer tools for detecting or manipulating DNA methylation states.

Conclusions

1. AceCas9 can be expressed in HEK293T cells.
2. AceCas9 can enter the nuclei of HEK293T cells for potential genome editing.
3. AceCas9 editing efficiency in PCSK9 and UBE3A genes is inconclusive due to low detected editing levels (<2.1% and <4%, respectively) and lack of clear distinction in sequencing data.
4. Construction of the AceCas9 PAM-interacting domain (PID) variant library yielded approximately 11600 members.
5. Directed evolution of AceCas9 did not yield variants capable of recognizing and cleaving 5'-NNNCG-3' PAM sequences.

Participation in conferences

Urtė Glibauskaitė, Ieva Jaskovikaitė, Stephen Knox Jones Jr. "*In vivo* Expression Studies of a Methylation Sensitive CRISPR-Cas9":

- International Conference of Life Sciences "The Coins 2025", Vilnius, Lithuania, 2025-03-18.
- 5th EMBL Partnership Conference "AI in one health", Heidelberg, Germany, 2025-04-01.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Stephen Knox Jones Jr., for his invaluable guidance, support, and encouragement throughout the course of this work. His expertise, insightful feedback, and patience played a crucial role in the progress and successful completion of this thesis.

I am also deeply thankful to my consultant, Ieva Jaskovikaitė, for her constructive feedback and professional insights, which significantly contributed to the quality and clarity of this work. Her foundational ideas helped shape the initial direction of this project, which I was then able to build upon and continue.

A special thank you goes to my laboratory teammate Uzair Ahmed for generously sharing his knowledge and advice on mammalian cell culture techniques. His guidance on how to grow and handle mammalian cells was incredibly helpful and greatly supported the experimental part of this work.

Finally, I want to thank the rest of the “Jones!Lab” members for their collaborative spirit and positive energy, which made the laboratory a truly inspiring and enjoyable place to work.

VILNIAUS UNIVERSITETAS
GYVYBĖS MOKSLŲ CENTRAS

Urtė Glibauskaitė

Magistro baigiamasis darbas

Metilinimui jautrios Cas9 nukleazės kryptinga evoliucija žmogaus genomo redagavimui

SANTRAUKA

CRISPR (*angl.* clustered regularly interspaced short palindromic repeats) lokusas ir Cas (*angl.* CRISPR-associated) baltymai suteikia adaptyvų imunitetą bakterijoms ir archėjoms, integruodami bakteriofagų ar kitos „svetimos“ DNR fragmentus į CRISPR lokusą. Šie fragmentai transkribuojami į gidinę RNR (gRNR), kuri nukreipia Cas endonukleazę atpažinti ir sukirpti taikinio DNR šalia PAM (*angl.* protospacer adjacent motif) sekos. gRNR galima pritaikyti bet kuriai tikslinei sekai, leidžiant tiksliai redaguoti genomą.

CRISPR-Cas aktyvumui įtakos turi epigenetinės modifikacijos. Nors plačiai naudojama *Streptococcus pyogenes* Cas9 (SpCas9) nėra tiesiogiai veikiama taikinio ar PAM metilinimo, jos aktyvumą gali slopinti DNR kondensacija metilintose srityse. Tačiau *Acidothermus cellulolyticus* Cas9 (AceCas9) yra savaime jautri metilinimui. Metilinus pirmąjį citoziną 5‘-NNNCC-3‘ PAM sekoje, AceCas9 nebekerpa taikinio DNR. Ši savybė potencialiai leistų taikyti AceCas9 specifiškam genomo redagavimui, atsižvelgiant į taikinio metilinimo būseną.

Nors AceCas9 aktyvumas pademonstruotas *Escherichia coli* ir *Clostridium thermocellum*, jos aktyvumas žmogaus ląstelėse yra nežinomas. Todėl šio darbo tikslas buvo nustatyti, ar AceCas9 gali veikti žmogaus ląstelėse. Šiame darbe parodyta, kad AceCas9 susintetinama žmogaus embrioninėse inkstų ląstelėse (HEK293T) ir lokalizuojama branduoliuose. Siekiant įvertinti AceCas9 redagavimo efektyvumą žmogaus ląstelėse, buvo taikytasi į du genus – PCSK9 ir UBE3A. Dėl nustatyto žemo redagavimo lygio ir sunkiai interpretuojamų sekoskaitos duomenų nepavyko patikimai įvertinti AceCas9 genomo redagavimo efektyvumo.

Kadangi citozino metilinimas CpG dinukleotide yra dažniausia DNR modifikacija žmogaus genome, šiame darbe buvo atlikta AceCas9 mutagenezė, siekiant pakeisti PAM atpažinimo seką į 5′-NNNCG-3′. Sukurta AceCas9 variantų biblioteka, kurioje yra 11600 skirtingų AceCas9 variantų su mutacijomis PI (*angl.* PAM-interacting) domene. Atrankai pritaikytas kryptingos evoliucijos metodas, paremtas ccdB toksino raiška ir bakterijų išgyvenamumu, priklausomai nuo Cas9 aktyvumo. Šiuose eksperimentuose nė vienas iš ištirtų variantų nekirpo taikinio su 5′-NNNCG-3′ PAM seka.

Nors ir AceCas9 raiška žmogaus ląstelėse vyksta bei nukleazė lokalizuojama ląstelių branduoliuose, reikalingi tolimesni optimizacijos eksperimentai, siekiant įvertinti šios nukleazės potencialą žmogaus genomo redagavimui. Papildomi eksperimentai reikalingi ir AceCas9 mutagenezei bei evoliucijai, kad AceCas9 atpažintų 5′-NNNCG-3′ PAM seką ir taip būtų praplėsta Cas baltymais paremtų genomo redagavimo įrankių įvairovė.

VILNIUS UNIVERSITY
LIFE SCIENCES CENTER

Urtė Glibauskaitė

Master's thesis

Directed Evolution Studies of a Methylation-Sensitive Cas9 for Human Genome Editing

ABSTRACT

CRISPR (clustered regularly interspaced short palindromic repeats) locus and Cas (CRISPR-associated) proteins provide adaptive immunity in bacteria and archaea by integrating foreign DNA into the CRISPR locus. This sequence is transcribed into a guide RNA (gRNA), which directs a Cas nuclease to recognize and cleave target DNA adjacent to a PAM (protospacer adjacent motif). The gRNA can be programmed to target any sequence of interest, enabling precise genome editing and diagnostic applications.

Epigenetic modifications can influence CRISPR-Cas activity. Though target or PAM methylation does not hinder the widely used *Streptococcus pyogenes* Cas9, DNA condensation in highly methylated region can, making certain genomic sites less accessible for editing. In contrast, *Acidothermus cellulolyticus* Cas9 (AceCas9) senses methylation. Its cleavage activity depends on whether the first (but not the second) cytosine of its PAM sequence is methylated (5′-NNNCC-3′). This property offers potential for methylation-sensitive gene editing.

AceCas9 operates in *Escherichia coli* and *Clostridium thermocellum*, but function in human cells remains unknown. This work aimed to determine this. This study demonstrated that AceCas9 can be expressed in human embryonic kidney (HEK293T) cells and enter the nucleus, which are key processes preceding genome editing. To assess editing in human cells, two endogenous human gene loci, PCSK9 and UBE3A, were targeted with AceCas9. AceCas9-mediated editing in human cells remains inconclusive due to low levels of detected editing and the lack of clear distinction in sequencing data.

This study also aimed to evolve AceCas9 to recognize a 5′-NNNCG-3′ PAM, as cytosine methylation in CpG dinucleotides is the most prevalent DNA methylation in human cells. A library of 11600 AceCas9 PAM-interacting domain variants was constructed and a ccdB-based directed evolution method was employed, linking Cas9 activity with bacterial survival. However, none of the tested variants were able to recognize a target with the 5′-NNNCG-3′ PAM.

These findings suggest that while AceCas9 can be expressed in human cells and localized to the nucleus, it requires further optimization for genome editing in human cells. The inability to evolve AceCas9 to recognize a 5′-NNNCG-3′ highlights our need for continued development of human-specific genome editing tools.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w

Aliaga Goltsman, D. S., Alexander, L. M., Lin, J.-L., Fregoso Ocampo, R., Freeman, B., Lamothe, R. C., Perez Rivas, A., Temoche-Diaz, M. M., Chadha, S., Nordenfelt, N., Janson, O. P., Barr, I., Devoto, A. E., Cost, G. J., Butterfield, C. N., Thomas, B. C., & Brown, C. T. (2022). Compact Cas9d and HEARO enzymes for genome editing discovered from uncultivated microbes. *Nature Communications*, *13*(1), 7602. https://doi.org/10.1038/s41467-022-35257-7

Arber, W. (1974). DNA Modification and Restriction. In W. E. Cohn (Ed.), *Progress in Nucleic Acid Research and Molecular Biology* (Vol. 14, pp. 1–37). Academic Press. https://doi.org/10.1016/S0079-6603(08)60204-4

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, *315*(5819), 1709–1712. https://doi.org/10.1126/science.1138140

Bayliss, C. D., Clark, J. L., & van der Woude, M. W. (2025). 100+ years of phase variation: The premier bacterial bet-hedging phenomenon. *Microbiology*, *171*(2), 001537. https://doi.org/10.1099/mic.0.001537

Beaulaurier, J., Schadt, E. E., & Fang, G. (2019). Deciphering bacterial epigenomes using modern sequencing technologies. *Nature Reviews Genetics*, *20*(3), Article 3. https://doi.org/10.1038/s41576-018-0081-3

Bernard, P., Kézdy, K. E., Van Melderen, L., Steyaert, J., Wyns, L., Pato, M. L., Higgins, P. N., & Couturier, M. (1993). The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase. *Journal of Molecular Biology*, *234*(3), 534–541. https://doi.org/10.1006/jmbi.1993.1609

Bestas, B., Wimberger, S., Degtev, D., Madsen, A., Rottner, A. K., Karlsson, F., Naumenko, S., Callahan, M., Touza, J. L., Francescatto, M., Möller, C. I., Badertscher, L., Li, S., Cerboni, S., Selfjord, N., Ericson, E., Gordon, E., Firth, M., Chylinski, K., … Maresca, M. (2023). A Type II-B Cas9 nuclease with minimized off-targets and reduced chromosomal translocations in vivo. *Nature Communications*, *14*(1), 5474. https://doi.org/10.1038/s41467-023-41240-7

Bhattacharya, S., & Satpati, P. (2023). Insights into the Mechanism of CRISPR/Cas9-Based Genome Editing from Molecular Dynamics Simulations. *ACS Omega*, *8*(2), 1817–1837. https://doi.org/10.1021/acsomega.2c05583

Bickle, T. A., Brack, C., & Yuan, R. (1978). ATP-induced conformational changes in the restriction endonuclease from Escherichia coli K-12. *Proceedings of the National Academy of Sciences*, *75*(7), 3099–3103. https://doi.org/10.1073/pnas.75.7.3099

Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, *40*(1), 91–99. https://doi.org/10.1016/0092-8674(85)90312-5

Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, *431*(7004), 96–99. https://doi.org/10.1038/nature02886

Boyle, E. A., Andreasson, J. O. L., Chircus, L. M., Sternberg, S. H., Wu, M. J., Guegler, C. K., Doudna, J. A., & Greenleaf, W. J. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proceedings of the National Academy of Sciences*, *114*(21), Article 21. https://doi.org/10.1073/pnas.1700557114

Brinkman, E. K., Chen, T., Amendola, M., & van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*, *42*(22), e168. https://doi.org/10.1093/nar/gku936

Casadesús, J., & Low, D. (2006). Epigenetic Gene Regulation in the Bacterial World. *Microbiology and Molecular Biology Reviews*, *70*(3), 830–856. https://doi.org/10.1128/mmbr.00016-06

Chavez, M., Chen, X., Finn, P. B., & Qi, L. S. (2023). Advances in CRISPR therapeutics. *Nature Reviews Nephrology*, *19*(1), 9–22. https://doi.org/10.1038/s41581-022-00636-2

Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., Sternberg, S. H., Joung, J. K., Yildiz, A., & Doudna, J. A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*, *550*(7676), 407–410. https://doi.org/10.1038/nature24268

Chen, Z., & Zhao, H. (2005). A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Research*, *33*(18), e154. https://doi.org/10.1093/nar/gni148

Cheng, X. (1995). Structure and function of DNA methyltransferases. *Annual Review of Biophysics and Biomolecular Structure*, *24*, 293–318. https://doi.org/10.1146/annurev.bb.24.060195.001453

Cherry, J. L. (2018). Methylation-Induced Hypermutation in Natural Populations of Bacteria. *Journal of Bacteriology*, *200*(24), 10.1128/jb.00371-18. https://doi.org/10.1128/jb.00371-18

Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K., Bauer, D. E., & Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology*, *37*(3), 224–226. https://doi.org/10.1038/s41587-019-0032-3

Collier, J., McAdams, H. H., & Shapiro, L. (2007). A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 17111–17116. https://doi.org/10.1073/pnas.0708112104

Cross, S. H., Meehan, R. R., Nan, X., & Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nature Genetics*, *16*(3), 256–259. https://doi.org/10.1038/ng0797-256

Daer, R. M., Cutts, J. P., Brafman, D. A., & Haynes, K. A. (2017). The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. *ACS Synthetic Biology*, *6*(3), 428–438. https://doi.org/10.1021/acssynbio.5b00299

Das, A., Hand, T. H., Smith, C. L., Wickline, E., Zawrotny, M., & Li, H. (2020). The molecular basis for recognition of 5′-NNNCC-3′ PAM and its methylation state by Acidothermus cellulolyticus Cas9. *Nature Communications*, *11*(1), 6346. https://doi.org/10.1038/s41467-020-20204-1

de Mendoza, A., Poppe, D., Buckberry, S., Pflueger, J., Albertin, C. B., Daish, T., Bertrand, S., de la Calle-Mustienes, E., Gomez-Skarmeta, J. L., Nery, J. R., Ecker, J. R., Baer, B., Ragsdale, C. W., Grützner, F., Escriva, H., Venkatesh, B., Bogdanovic, O., & Lister, R. (2021). The emergence of the brain non-CpG methylation system in vertebrates. *Nature Ecology & Evolution*, *5*(3), 369–378. https://doi.org/10.1038/s41559-020-01371-2

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602–607. https://doi.org/10.1038/nature09886

Dvorak, P. (2007). *Methods of Directed Evolution and Their Application for Engineering of Haloalkane Dehalogenases*. https://doi.org/10.13140/2.1.3648.0006

Edwards, P. R., & Bruner, D. W. (1939). The Demonstration of Phase Variation in Salmonella abortus-equi. *Journal of Bacteriology*, *38*(1), 63–72. https://doi.org/10.1128/jb.38.1.63-72.1939

Eggers, A. R., Chen, K., Soczek, K. M., Tuck, O. T., Doherty, E. E., Xu, B., Trinidad, M. I., Thornton, B. W., Yoon, P. H., & Doudna, J. A. (2024). Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. *Cell*, *187*(13), 3249-3261.e14. https://doi.org/10.1016/j.cell.2024.04.031

Esvelt, K. M., Carlson, J. C., & Liu, D. R. (2011). A system for the continuous directed evolution of biomolecules. *Nature*, *472*(7344), Article 7344. https://doi.org/10.1038/nature09929

Fujita, T., Yuno, M., & Fujii, H. (2016). Allele-specific locus binding and genome editing by CRISPR at the p16INK4a locus. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep30485

Georjon, H., & Bernheim, A. (2023). The highly diverse antiphage defence systems of bacteria. *Nature Reviews Microbiology*, *21*(10), 686–700. https://doi.org/10.1038/s41579-023-00934-x

Gersbach, C. A., Gaj, T., & Barbas, C. F. I. (2014). Synthetic Zinc Finger Proteins: The Advent of Targeted Gene Regulation and Genome Modification Technologies. *Accounts of Chemical Research*, *47*(8), 2309–2318. https://doi.org/10.1021/ar500039w

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), Article 5. https://doi.org/10.1038/nmeth.1318

Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., & Randau, L. (2019). PAM identification by CRISPR-Cas effector complexes: Diversified mechanisms and structures. *RNA Biology*, *16*(4), 504–517. https://doi.org/10.1080/15476286.2018.1504546

Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry*, *74*, 481–514. https://doi.org/10.1146/annurev.biochem.74.010904.153721

Guo, J. U., Su, Y., Shin, J. H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G., & Song, H. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience*, *17*(2), 215–222. https://doi.org/10.1038/nn.3607

Gutierrez-Arcelus, M., Lappalainen, T., Montgomery, S. B., Buil, A., Ongen, H., Yurovsky, A., Bryois, J., Giger, T., Romano, L., Planchon, A., Falconnet, E., Bielser, D., Gagnebin, M., Padioleau, I., Borel, C., Letourneau, A., Makrythanasis, P., Guipponi, M., Gehrig, C., … Dermitzakis, E. T. (2013). Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife*, *2*, e00523. https://doi.org/10.7554/eLife.00523

Hall, M. C., Jordan, J. R., & Matson, S. W. (1998). Evidence for a physical interaction between the Escherichia coli methyl-directed mismatch repair proteins MutL and UvrD. *The EMBO Journal*, *17*(5), 1535–1541. https://doi.org/10.1093/emboj/17.5.1535

Hand, T. H., Das, A., & Li, H. (2019). Directed evolution studies of a thermophilic Type II-C Cas9. *Methods in Enzymology*, *616*, 265–288. https://doi.org/10.1016/bs.mie.2018.10.029

Hand, T. H., Das, A., Roth, M. O., Smith, C. L., Jean-Baptiste, U. L., & Li, H. (2018). Phosphate Lock Residues of Acidothermus cellulolyticus Cas9 Are Critical to Its Substrate Specificity. *ACS Synthetic Biology*, *7*(12), 2908–2917. https://doi.org/10.1021/acssynbio.8b00455

Hand, T. H., Roth, M. O., Smith, C. L., Shiel, E., Klein, K. N., Gilbert, D. M., & Li, H. (2021). Catalytically Enhanced Cas9 through Directed Protein Evolution. *The CRISPR Journal*, *4*(2), 223–232. https://doi.org/10.1089/crispr.2020.0092

Harrington, L. B., Paez-Espino, D., Staahl, B. T., Chen, J. S., Ma, E., Kyrpides, N. C., & Doudna, J. A. (2017). A thermostable Cas9 with increased lifetime in human plasma. *Nature Communications*, *8*(1), 1424. https://doi.org/10.1038/s41467-017-01408-4

Henderson, I. R., Meehan, M., & Owen, P. (1997). Antigen 43, a phase-variable bipartite outer membrane protein, determines colony morphology and autoaggregation in Escherichia coli K-12. *FEMS Microbiology Letters*, *149*(1), 115–120. https://doi.org/10.1111/j.1574-6968.1997.tb10317.x

Hendrich, B., & Bird, A. (1998). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Molecular and Cellular Biology*, *18*(11), 6538–6547.

Hibshman, G. N., & Taylor, D. W. (2024). *Mechanism of type II-B CRISPR-Cas enhanced specificity* (p. 2024.10.22.619592). bioRxiv. https://doi.org/10.1101/2024.10.22.619592

Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., & Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. *Cell*, *172*(6), Article 6. https://doi.org/10.1016/j.cell.2017.11.032

Hirakawa, M. P., Krishnakumar, R., Timlin, J. A., Carney, J. P., & Butler, K. S. (2020). Gene editing and CRISPR in the clinic: Current and future perspectives. *Bioscience Reports*, *40*(4), Article 4. https://doi.org/10.1042/BSR20200127

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9), Article 9. https://doi.org/10.1038/nbt.2647

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., & Liu, D. R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, *556*(7699), Article 7699. https://doi.org/10.1038/nature26155

Huang, T. P., Heins, Z. J., Miller, S. M., Wong, B. G., Balivada, P. A., Wang, T., Khalil, A. S., & Liu, D. R. (2023). High-throughput continuous evolution of compact Cas9 variants targeting single-nucleotide-pyrimidine PAMs. *Nature Biotechnology*, *41*(1), 96–107. https://doi.org/10.1038/s41587-022-01410-2

Hunt, J. M. T., Samson, C. A., Rand, A. du, & Sheppard, H. M. (2023). Unintended CRISPR-Cas9 editing outcomes: A review of the detection and prevalence of structural variants generated by gene-editing in human cells. *Human Genetics*, *142*(6), 705–720. https://doi.org/10.1007/s00439-023-02561-1

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J., Sabunciyan, S., & Feinberg, A. P. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics*, *41*(2), 178–186. https://doi.org/10.1038/ng.298

Jeltsch, A. (2002). Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases. *ChemBioChem*, *3*(4), 274–293. https://doi.org/10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), Article 6096. https://doi.org/10.1126/science.1225829

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E., & Doudna, J. A. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (New York, N.Y.)*, *343*(6176), 1247997. https://doi.org/10.1126/science.1247997

Kahramanoglou, C., Prieto, A. I., Khedkar, S., Haase, B., Gupta, A., Benes, V., Fraser, G. M., Luscombe, N. M., & Seshasayee, A. S. N. (2012). Genomics of DNA cytosine methylation in Escherichia coli reveals its role in stationary phase transcription. *Nature Communications*, *3*, 886. https://doi.org/10.1038/ncomms1878

Kebschull, J. M., & Zador, A. M. (2015). Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Research*, *43*(21), e143. https://doi.org/10.1093/nar/gkv717

Kim, Y., Kim, N., Okafor, I., Choi, S., Min, S., Lee, J., Bae, S.-M., Choi, K., Choi, J., Harihar, V., Kim, Y., Kim, J.-S., Kleinstiver, B. P., Lee, J. K., Ha, T., & Kim, H. H. (2023). Sniper2L is a high-fidelity Cas9 variant with high activity. *Nature Chemical Biology*, *19*(8), 972–980. https://doi.org/10.1038/s41589-023-01279-5

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, *529*(7587), 490–495. https://doi.org/10.1038/nature16526

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., Aryee, M. J., & Joung, J. K. (2015). Engineered

CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, *523*(7561), Article 7561. https://doi.org/10.1038/nature14592

Knott, G. J., & Doudna, J. A. (2018). CRISPR-Cas guides the future of genetic engineering. *Science (New York, N.Y.)*, *361*(6405), 866–869. https://doi.org/10.1126/science.aat5011

Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*, *37*, 67–78. https://doi.org/10.1016/j.mib.2017.05.008

Lee, J., Jung, M.-H., Jeong, E., & Lee, J. K. (2019). Using Sniper-Cas9 to Minimize Off-target Effects of CRISPR-Cas9 Without the Loss of On-target Activity Via Directed Evolution. *Journal of Visualized Experiments: JoVE*, *144*. https://doi.org/10.3791/59202

Lehnertz, B., Ueda, Y., Derijck, A. A. H. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., & Peters, A. H. F. M. (2003). Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology*, *13*(14), 1192–1200. https://doi.org/10.1016/S0960-9822(03)00432-9

Lei, Y., Zhang, X., Su, J., Jeong, M., Gundry, M. C., Huang, Y.-H., Zhou, Y., Li, W., & Goodell, M. A. (2017). Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. *Nature Communications*, *8*(1), 16026. https://doi.org/10.1038/ncomms16026

Li, B.-Z., Huang, Z., Cui, Q.-Y., Song, X.-H., Du, L., Jeltsch, A., Chen, P., Li, G., Li, E., & Xu, G.-L. (2011). Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. *Cell Research*, *21*(8), 1172–1181. https://doi.org/10.1038/cr.2011.92

Li, Y., Chen, X., & Lu, C. (2021). The interplay between DNA and histone methylation: Molecular mechanisms and disease implications. *EMBO Reports*, *22*(5), e51803. https://doi.org/10.15252/embr.202051803

Liang, Z., Wu, Y., Deng, S., Wei, S., Zhang, K., & Guo, Y. (2025). Enhanced Genome Editing Activity with Novel Chimeric ScCas9 Variants in Rice. *Advanced Science (Weinheim, Baden-Wurttemberg, Germany)*, *12*(8), e2411549. https://doi.org/10.1002/advs.202411549

Lin, C.-Y., Huang, Z., Wen, W., Wu, A., Wang, C., & Niu, L. (2015). Enhancing Protein Expression in HEK-293 Cells by Lowering Culture Temperature. *PLoS ONE*, *10*(4), e0123562. https://doi.org/10.1371/journal.pone.0123562

Lin, S., Staahl, B. T., Alla, R. K., & Doudna, J. A. (2015). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife*, *3*, e04766. https://doi.org/10.7554/eLife.04766

Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., Lucero, J., Huang, Y., Dwork, A. J., Schultz, M. D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J. C., Rao, A., Esteller, M., He, C., Haghighi, F. G., … Ecker, J. R. (2013). Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science (New York, N.Y.)*, *341*(6146), 1237905. https://doi.org/10.1126/science.1237905

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322. https://doi.org/10.1038/nature08514

Liu, H., & Naismith, J. H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*, *8*(1), 91. https://doi.org/10.1186/1472-6750-8-91

Liu, X. S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R. A., & Jaenisch, R. (2016). Editing DNA methylation in the mammalian genome. *Cell*, *167*(1), 233-247.e17. https://doi.org/10.1016/j.cell.2016.08.056

Lu, A. L., Clark, S., & Modrich, P. (1983). Methyl-directed repair of DNA base-pair mismatches in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *80*(15), 4639–4643.

Lu, J., Wu, T., Zhang, B., Liu, S., Song, W., Qiao, J., & Ruan, H. (2021). Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Communication and Signaling*, *19*(1), 60. https://doi.org/10.1186/s12964-021-00741-y

Maeder, M. L., Stefanidakis, M., Wilson, C. J., Baral, R., Barrera, L. A., Bounoutas, G. S., Bumcrot, D., Chao, H., Ciulla, D. M., DaSilva, J. A., Dass, A., Dhanapal, V., Fennell, T. J., Friedland, A. E., Giannoukos, G., Gloskowski, S. W., Glucksmann, A., Gotta, G. M., Jayaram, H., … Jiang, H. (2019). Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nature Medicine*, *25*(2), 229–233. https://doi.org/10.1038/s41591-018-0327-9

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., … Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: A burst of class 2 and derived variants. *Nature Reviews Microbiology*, *18*(2), 67–83. https://doi.org/10.1038/s41579-019-0299-x

Marczynski, G. T. (1999). Chromosome Methylation and Measurement of Faithful, Once and Only Once per Cell Cycle Chromosome Replication inCaulobacter crescentus. *Journal of Bacteriology*, *181*(7), 1984–1993. https://doi.org/10.1128/jb.181.7.1984-1993.1999

Marinus, M. G., & Morris, N. R. (1973). Isolation of deoxyribonucleic acid methylase mutants of Escherichia coli K-12. *Journal of Bacteriology*, *114*(3), 1143–1150. https://doi.org/10.1128/jb.114.3.1143-1150.1973

Marraffini, L. A., & Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Reviews. Genetics*, *11*(3), Article 3. https://doi.org/10.1038/nrg2749

McCutcheon, S. R., Rohm, D., Iglesias, N., & Gersbach, C. A. (2024). Epigenome editing technologies for discovery and medicine. *Nature Biotechnology*, *42*(8), 1199–1217. https://doi.org/10.1038/s41587-024-02320-1

Mcginn, J., & Marraffini, L. (2018). Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nature Reviews Microbiology*, *17*, 1. https://doi.org/10.1038/s41579-018-0071-7

Militello, K. T., Simon, R. D., Qureshi, M., Maines, R., Van Horne, M. L., Hennick, S. M., Jayakar, S. K., & Pounder, S. (2012). Conservation of Dcm-mediated Cytosine DNA Methylation in Escherichia coli. *FEMS Microbiology Letters*, *328*(1), 78–85. https://doi.org/10.1111/j.1574-6968.2011.02482.x

Mir, A., Edraki, A., Lee, J., & Sontheimer, E. J. (2018). Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chemical Biology*, *13*(2), 357–365. https://doi.org/10.1021/acschembio.7b00855

Modrich, P. (2016). Mechanisms in E. coli and Human Mismatch Repair. *Angewandte Chemie (International Ed. in English)*, *55*(30), 8490–8501. https://doi.org/10.1002/anie.201601412

Mohagheghi, A., Grohmann, K., Himmel, M., Leighton, L., & Updegraff, D. M. (1986). Isolation and Characterization of Acidothermus cellulolyticus gen. Nov., sp. Nov., a New Genus of Thermophilic, Acidophilic, Cellulolytic Bacteria. *International Journal of Systematic and Evolutionary Microbiology*, *36*(3), 435–443. https://doi.org/10.1099/00207713-36-3-435

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3

Musunuru, K., Chadwick, A. C., Mizoguchi, T., Garcia, S. P., DeNizio, J. E., Reiss, C. W., Wang, K., Iyer, S., Dutta, C., Clendaniel, V., Amaonye, M., Beach, A., Berth, K., Biswas, S., Braun, M. C., Chen, H.-M., Colace, T. V., Ganey, J. D., Gangopadhyay, S. A., … Kathiresan, S. (2021). In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature*, *593*(7859), 429–434. https://doi.org/10.1038/s41586-021-03534-y

Naftelberg, S., Schor, I. E., Ast, G., & Kornblihtt, A. R. (2015). Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annual Review of Biochemistry*, *84*, 165–198. https://doi.org/10.1146/annurev-biochem-060614-034242

Nayfach, S., Bhatnagar, A., Novichkov, A., Estevam, G. O., Kim, N., Hill, E., Ruffolo, J. A., Silverstein, R., Gallagher, J., Kleinstiver, B., Meeske, A. J., Cameron, P., & Madani, A. (2025). *Engineering of CRISPR-Cas PAM recognition using deep learning of vast evolutionary data* (p. 2025.01.06.631536). bioRxiv. https://doi.org/10.1101/2025.01.06.631536

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F., & Nureki, O. (2014). Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*, *156*(5), Article 5. https://doi.org/10.1016/j.cell.2014.02.001

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O. O., Gootenberg, J. S., Mori, H., Oura, S., Holmes, B., Tanaka, M., Seki, M., Hirano, H., Aburatani, H., Ishitani, R., Ikawa, M., Yachie, N., … Nureki, O. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*, *361*(6408), Article 6408. https://doi.org/10.1126/science.aas9129

Oliveira, P. H., Touchon, M., & Rocha, E. P. C. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Research*, *42*(16), 10618–10631. https://doi.org/10.1093/nar/gku734

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C. D., Cheng, X., & Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, *448*(7154), 714–717. https://doi.org/10.1038/nature05987

Pradeepa, M. M., Sutherland, H. G., Ule, J., Grimes, G. R., & Bickmore, W. A. (2012). Psip1/Ledgf p52 Binds Methylated Histone H3K36 and Splicing Factors and Contributes to the Regulation of Alternative Splicing. *PLOS Genetics*, *8*(5), e1002717. https://doi.org/10.1371/journal.pgen.1002717

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), Article 5. https://doi.org/10.1016/j.cell.2013.02.022

Rakyan, V. K., Down, T. A., Thorne, N. P., Flicek, P., Kulesha, E., Gräf, S., Tomazou, E. M., Bäckdahl, L., Johnson, N., Herberth, M., Howe, K. L., Jackson, D. K., Miretti, M. M., Fiegler, H., Marioni, J. C., Birney, E., Hubbard, T. J. P., Carter, N. P., Tavaré, S., & Beck, S. (2008). An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Research*, *18*(9), 1518–1529. https://doi.org/10.1101/gr.077479.108

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P., & Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(10), 5237–5242. https://doi.org/10.1073/pnas.97.10.5237

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), Article 11. https://doi.org/10.1038/nprot.2013.143

Reisenauer, A., Kahng, L. S., McCollum, S., & Shapiro, L. (1999). Bacterial DNA Methylation: A Cell Cycle Regulator? *Journal of Bacteriology*, *181*(17), 5135–5139.

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S. K., Dryden, D. T. F., Dybvig, K., Firman, K., Gromova, E. S., Gumport, R. I., Halford, S. E., Hattman, S., Heitman, J., Hornby, D. P., Janulaitis, A., Jeltsch, A., … Xu, S. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research*, *31*(7), 1805–1812. https://doi.org/10.1093/nar/gkg274

Ruta, G. V., Ciciani, M., Kheir, E., Gentile, M. D., Amistadi, S., Casini, A., & Cereseto, A. (2024). Eukaryotic-driven directed evolution of Cas9 nucleases. *Genome Biology*, *25*(1), 79. https://doi.org/10.1186/s13059-024-03215-9

Saitoski, K., Ryaboshapkina, M., Hamza, G. M., Jarnuczak, A. F., Berthault, C., Carlotti, F., Armanet, M., Sengupta, K., Underwood, C. R., Andersson, S., Guillas, I., Le Goff, W., & Scharfmann, R. (2022). Proprotein convertase PCSK9 affects expression of key surface proteins in human pancreatic beta cells via intracellular and extracellular regulatory circuits. *The Journal of Biological Chemistry*, *298*(7), 102096. https://doi.org/10.1016/j.jbc.2022.102096

Saleh-Gohari, N., & Helleday, T. (2004). Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Research*, *32*(12), 3683–3688. https://doi.org/10.1093/nar/gkh703

Sánchez-Romero, M. A., & Casadesús, J. (2020). The bacterial epigenome. *Nature Reviews Microbiology*, *18*(1), 7–20. https://doi.org/10.1038/s41579-019-0286-2

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(5), 1412–1417. https://doi.org/10.1073/pnas.0510310103

Schmidt, T. G. M., Batz, L., Bonet, L., Carl, U., Holzapfel, G., Kiem, K., Matulewicz, K., Niermeier, D., Schuchardt, I., & Stanar, K. (2013). Development of the Twin-Strep-tag® and its application for purification of recombinant proteins from cell culture supernatants. *Protein Expression and Purification*, *92*(1), Article 1. https://doi.org/10.1016/j.pep.2013.08.021

Seisenberger, S., Peat, J. R., Hore, T. A., Santos, F., Dean, W., & Reik, W. (2013). Reprogramming DNA methylation in the mammalian life cycle: Building and breaking epigenetic barriers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1609), 20110330. https://doi.org/10.1098/rstb.2011.0330

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., & Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, *479*(7371), 74–79. https://doi.org/10.1038/nature10442

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science*, *351*(6268), Article 6268. https://doi.org/10.1126/science.aad5227

Song, Z., Zhang, Q., Wu, W., Pu, Z., & Yu, H. (2023). Rational design of enzyme activity and enantioselectivity. *Frontiers in Bioengineering and Biotechnology*, *11*. https://doi.org/10.3389/fbioe.2023.1129149

Stephens, C., Reisenauer, A., Wright, R., & Shapiro, L. (1996). A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. *Proceedings of the National Academy of Sciences*, *93*(3), 1210–1214. https://doi.org/10.1073/pnas.93.3.1210

Stepper, P., Kungulovski, G., Jurkowska, R. Z., Chandra, T., Krueger, F., Reinhardt, R., Reik, W., Jeltsch, A., & Jurkowski, T. P. (2017). Efficient targeted DNA methylation with chimeric dCas9–Dnmt3a–Dnmt3L methyltransferase. *Nucleic Acids Research*, *45*(4), 1703–1713. https://doi.org/10.1093/nar/gkw1112

Sternberg, S. H., LaFrance, B., Kaplan, M., & Doudna, J. A. (2015). Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature*, *527*(7576), Article 7576. https://doi.org/10.1038/nature15544

Sugimoto, K., Koh, E., Sin, H.-S., Maeda, Y., Narimoto, K., Izumi, K., Kobori, Y., Kitamura, E., Nagase, H., Yoshida, A., & Namiki, M. (2009). Tissue-specific differentially methylated regions of the human VASA gene are potentially associated with maturation arrest phenotype in the testis. *Journal of Human Genetics*, *54*(8), 450–456. https://doi.org/10.1038/jhg.2009.59

Towbin, H., Staehelin, T., & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(9), 4350–4354. https://doi.org/10.1073/pnas.76.9.4350

Tsai, J. M., Nowak, R. P., Ebert, B. L., & Fischer, E. S. (2024). Targeted protein degradation: From mechanisms to clinic. *Nature Reviews Molecular Cell Biology*, *25*(9), 740–757. https://doi.org/10.1038/s41580-024-00729-9

Tsui, T. K. M., Hand, T. H., Duboy, E. C., & Li, H. (2017). The Impact of DNA Topology and Guide Length on Target Selection by a Cytosine-Specific Cas9. *ACS Synthetic Biology*, *6*(6), 1103–1113. https://doi.org/10.1021/acssynbio.7b00050

Tuncel, A., Pan, C., Clem, J. S., Liu, D., & Qi, Y. (2025). CRISPR–Cas applications in agriculture and plant research. *Nature Reviews Molecular Cell Biology*, 1–23. https://doi.org/10.1038/s41580-025-00834-3

van der Woude, M. W., & Henderson, I. R. (2008). Regulation and function of Ag43 (flu). *Annual Review of Microbiology*, *62*, 153–169. https://doi.org/10.1146/annurev.micro.62.081307.162938

Vihma, H., Li, K., Welton-Arndt, A., Smith, A. L., Bettadapur, K. R., Gilmore, R. B., Gao, E., Cotney, J. L., Huang, H.-C., Collins, J. L., Chamberlain, S. J., Lee, H.-M., Aubé, J., & Philpot, B. D. (2024). Ube3a unsilencer for the potential treatment of Angelman syndrome. *Nature Communications*, *15*(1), 5558. https://doi.org/10.1038/s41467-024-49788-8

Walker, J. E., Lanahan, A. A., Zheng, T., Toruno, C., Lynd, L. R., Cameron, J. C., Olson, D. G., & Eckert, C. A. (2020). Development of both type I–B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium *Clostridium thermocellum*. *Metabolic Engineering Communications*, *10*, e00116. https://doi.org/10.1016/j.mec.2019.e00116

Wang, Q., Xiong, F., Wu, G., Liu, W., Chen, J., Wang, B., & Chen, Y. (2022). Gene body methylation in cancer: Molecular mechanisms and clinical applications. *Clinical Epigenetics*, *14*(1), 154. https://doi.org/10.1186/s13148-022-01382-9

Wang, Y., Han, S., Wang, Y., Liang, Q., & Luo, W. (2025). Artificial Intelligence Technology Assists Enzyme Prediction and Rational Design. *Journal of Agricultural and Food Chemistry*, *73*(12), 7065–7073. https://doi.org/10.1021/acs.jafc.4c13201

Wolter, J. M., Mao, H., Fragola, G., Simon, J. M., Krantz, J. L., Bazick, H. O., Oztemiz, B., Stein, J. L., & Zylka, M. J. (2020). Cas9 gene therapy for Angelman syndrome traps Ube3a-ATS long non-coding RNA. *Nature*, *587*(7833), 281–284. https://doi.org/10.1038/s41586-020-2835-2

Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., Cheng, A. W., Trevino, A. E., Konermann, S., Chen, S., Jaenisch, R., Zhang, F., & Sharp, P. A. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology*, *32*(7), Article 7. https://doi.org/10.1038/nbt.2889

Xiong, W., Liu, B., Shen, Y., Jing, K., & Savage, T. R. (2021). Protein engineering design from directed evolution to de novo synthesis. *Biochemical Engineering Journal*, *174*, 108096. https://doi.org/10.1016/j.bej.2021.108096

Yamada, M., Watanabe, Y., Gootenberg, J. S., Hirano, H., Ran, F. A., Nakane, T., Ishitani, R., Zhang, F., Nishimasu, H., & Nureki, O. (2017). Crystal Structure of the Minimal Cas9 from Campylobacter jejuni Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. *Molecular Cell*, *65*(6), 1109-1121.e3. https://doi.org/10.1016/j.molcel.2017.02.007

Yang, H., Wang, H., Shivalila, C. S., Cheng, A. W., Shi, L., & Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas mediated genome engineering. *Cell*, *154*(6), 1370–1379. https://doi.org/10.1016/j.cell.2013.08.022

Yang, J., Wang, T., Huang, Y., Long, Z., Li, X., Zhang, S., Zhang, L., Liu, Z., Zhang, Q., Sun, H., Zhang, M., Yin, H., Liu, Z., & Zhang, H. (2025). Insights into the compact CRISPR–Cas9d system. *Nature Communications*, *16*(1), 2462. https://doi.org/10.1038/s41467-025-57455-9

Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J.-P., Nissim-Rafinia, M., Cohen, A.-H. S., Rippe, K., Meshorer, E., & Ast, G. (2015). HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell Reports*, *10*(7), 1122–1134. https://doi.org/10.1016/j.celrep.2015.01.038

Zamudio, N. M., Scott, H. S., Wolski, K., Lo, C.-Y., Law, C., Leong, D., Kinkel, S. A., Chong, S., Jolley, D., Smyth, G. K., de Kretser, D., Whitelaw, E., & O'Bryan, M. K. (2011). DNMT3L is a regulator of X chromosome compaction and post-meiotic gene transcription. *PloS One*, *6*(3), e18276. https://doi.org/10.1371/journal.pone.0018276

Zhang, Y., Li, S., Li, R., Qiu, X., Fan, T., Wang, B., Zhang, B., & Zhang, L. (2024). Advances in application of CRISPR-Cas13a system. *Frontiers in Cellular and Infection Microbiology*, *14*, 1291557. https://doi.org/10.3389/fcimb.2024.1291557

Zhao, C., Miao, J., Sun, R., Liang, R., Chen, W., Gao, Y., Wang, X., Han, S., Zhao, W., Lei, T., & Huang, C. (2022). MBD1/HDAC3-miR-5701-FGFR2 axis promotes the development of gastric cancer. *Aging (Albany NY)*, *14*(14), 5878–5894. https://doi.org/10.18632/aging.204190

Zhao, L., Koseki, S. R. T., Silverstein, R. A., Amrani, N., Peng, C., Kramme, C., Savic, N., Pacesa, M., Rodríguez, T. C., Stan, T., Tysinger, E., Hong, L., Yudistyra, V., Ponnapati, M. R., Jacobson, J. M., Church, G. M., Jakimo, N., Truant, R., Jinek, M., … Chatterjee, P. (2023). PAM-flexible genome editing with an engineered chimeric Cas9. *Nature Communications*, *14*(1), 6175. https://doi.org/10.1038/s41467-023-41829-y

Zhou, S., Shen, Y., Zheng, M., Wang, L., Che, R., Hu, W., & Li, P. (2017). DNA methylation of METTL7A gene body regulates its transcriptional level in thyroid cancer. *Oncotarget*, *8*(21), 34652–34660. https://doi.org/10.18632/oncotarget.16147